10th Scientific Conference of Young Researchers

SCYR 2010

Faculty of Electrical Engineering and Informatics
Technical University of Košice

Proceeding from Conference

May 19th, 2010
Košice, Slovakia

Sponsors


TECHNICKÁ UNIVERZITA
V KOŠICIACH
FEI


SES


elfa
nové myšlienky nové možnosti


ZSVTS

10th Scientific Conference of Young Researchers
of Faculty of Electrical Engineering and Informatics
Technical University of Košice

Proceeding from Conference

Published: Faculty of Electrical Engineering and Informatics
Technical University of Košice
I. Edition, 381 pages, the number of CD Proceedings: 120 pieces

Editors: prof. Ing. Alena Pietriková, PhD.
Ing. Jana Modrovičová

ISBN 978-80-553-0423-6

Program Committee of 10th Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Košice

Chairman: prof. Ing. Liberios Vokorokos, PhD.

Members:
Prof. Ing. Roman Cimbala, PhD.
Prof. Ing. Dušan Kocur, CSC.
Prof. Ing. Iraida Kolcunova, PhD.
Prof. Ing. Dobroslav Kováč, PhD.
Prof. Ing. Irena Kováčová, PhD.
Prof. Ing. Dušan Krokavec, CSc.
Prof. Ing. Dušan Levický, CSc.
Prof. RNDr. Valerie Novitzka, CSc.
Prof. Ing. Alena Pietriková, PhD.
Prof. Ing. RNDr. Ján Turán, DrSc.
Asoc. Prof. Ing. Ján Bača, CSc.
Asoc. Prof. Ing. Ľubomír Doboš, PhD.
Asoc. Prof. Ing. Miloš Drutarovský, CSc.
Asoc. Prof. Ing. Želmíra Ferkova, PhD.
Asoc. Prof. Ing. Zdeněk Havlice, CSc.
Asoc. Prof. Ing. Anna Jadlovská, PhD.
Asoc. Prof. Ing. Jozef Juhár, PhD.
Asoc. Prof. Ing. Marián Mach, CSc.
Asoc. Prof. Ing. Kristína Machová, CSc.
Asoc. Prof. Ing. Ľuboš Ovseník, PhD.
Asoc. Prof. Ing. Daniela Perduková, PhD.
Asoc. Prof. Ing. Jaroslav Porubän, PhD.
Asoc. Prof. Ing. Branislav Sobota, CSc.
Asoc. Prof. Ing. Iveta Zolotová, PhD.

Organization Committee of 10th Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Košice

Members:
Ing. František Baník
Ing. Radoslav Bučko
Ing. Vieroslava Čačková
Ing. Michal Kravčík
Ing. Jana Modrovičová
Mgr. Jana Petrillova
Ing. Eva Vozáriková
Ing. Miroslav Sabo

Contact address:
Faculty of Electrical Engineering and Informatics
Technical University of Košice
Letná 9
042 00 Košice
Slovak Republic

Foreword

Dear Colleagues,

SCYR (Scientific Conference of Young Researchers) is a Scientific Event focused on exchange of information among young scientists from Faculty of Electrical Engineering and Informatics Technical University of Košice - series of annual events that was founded in 2000. Since 2000 the conference has been hosted by FEI TUKE with rising technical level and unique multicultural atmosphere. The Tenth Scientific Conference of Young Researchers (SCYR 2010), conference of Graduates and Young researchers, was held on 19th May 2010. The primary aims of the conference, to provide a forum for dissemination of information and scientific results relating to research and development activities at the Faculty of Electrical Engineering and Informatics has been achieved. 105 participants mostly by doctoral categories were active in the conference.

Faculty of Electrical Engineering and Informatics has a long tradition of students participating in skilled labor where they have to apply their theoretical knowledge. SCYR is opportunities for doctoral and graduating students use this event to train their scientific knowledge exchange. Nevertheless, the original goal to represent a forum for the exchange of information between young scientists from academic communities on topics related to their experimental and theoretical works in the very wide spread field of electronics, telecommunication, electrotechnics, computers and informatics, cybernetics and Artificial intelligence, electric power engineering, remained unchanged.

10th Scientific Conference of Young Researchers at Faculty of Electrical Engineering and Informatics Technical University of Košice (SCYR 2010) was traditionally organized in the campus of Technical University of Košice. The Conference was opened in the name of dean prof. Ing. Liberios Vokorokos, PhD. by the vicedean of faculty, doc. Ing. Roman Cimbala, PhD. In his introductory address he noted the importance of the Conference as a forum for exchange of information and a medium for broadening the scientific horizons of its participants and stressed the scientific and practical value of investigations being carried out by young researchers.

The program of conferences traditionally includes two parallel sessions (both consist of oral and poster part):

- Electrical & Electronics Engineering
- Informatics & Telecommunications

with about 105 technical papers dealing with research results obtained mainly in university environment. This day was filled with a lot of interesting scientific discussions among the junior researchers and graduate students, and the representatives of the Faculty of Electrical Engineering and Informatics. This Scientific Network included various research problems and education, communication between young scientists and students, between students and professors. Conference was also a platform for student exchange and a potential starting point for scientific cooperation. The results presented in papers demonstrated that the investigations being conducted by young scientists are making a valuable contribution to the fulfillment of the tasks set for science and technology at Faculty of Electrical Engineering and Informatics.

We want to thank all participants for contributing to these proceedings with their high quality manuscripts. We hope that conference constitutes a platform for a continual dialogue among young scientists.

It is our pleasure and honor to express our gratitude to our sponsors and to all friends, colleagues and committee members who contributed with their ideas, discussions, and sedulous hard work to the success of this event. We also want to thank our session chairs for their cooperation and dedication throughout the whole conference.

Finally, we want to thank all the attendees of the conference for fruitful discussions and a pleasant stay in our event.

Liberios VOKOROKOS

May 19th Košice

Content

1st section: Electrical & Electronics Engineering

2nd section: Informatics & Telecommunications

1st section: Electrical & Electronics Engineering

Analysis of renewable energy sources utilization

Ing. Martin Bačko,Ing. Anna Hodulíková

Dept. of Theoretical Electrotechnics and Electrical Measurement, FEI TU of Košice, Slovak Republic

martin.backo@tuke.sk, anna.hodulikova@tuke.sk

***Abstract*—This article deals with possibilities of using renewable energy sources especially wind and solar energy for gradual replacement of traditional energy gained from fossil fuels which will be depleted relatively soon. Renewable energy sources are available in much higher amount than the whole world needs and their usage is imperative for keeping the ecological equilibrium.**

***Keywords*—photovoltaics, renewable sources, solar energy, wind energy**

I. INTRODUCTION

In term renewable sources of energy we understand ecologically clean direct or indirect form of solar energy, which can be transformed by suitable technological solution to electrical energy. Burning of fossil fuels probably caused accumulation of harmful carbon dioxide emissions, which are responsible for greenhouse effect and climatic changes. Direct form of solar energy can be used for direct water heating, or direct transformation to electrical energy using the photovoltaics principle. Indirect form of solar energy is in the form of water and wind energy which again can be with suitable technical solution used for generating the electrical energy.

II. PHOTOVOLTAICS

A. Basic principle

Photovoltaic systems use the principle of semiconductors. Semiconductors are elements from IV group of periodic table such as silicon (Si), germanium (Ge), tin (Sn) which have 4 valence electrons on sphere. Semiconductor systems consist of 2 elements, for example III-V semiconductor gallium-arsenic (GaAs) and II-VI semiconductor cadmium-tellurium (CdTe).

Silicon is the most common material in photovoltaics. It is the second most common element in Earth´s crust, but it is not possible to find it in chemically clean form. It is the fundamental semiconductor of IV group of periodic table, therefore it have 4 valence electrons. Because it wants to keep the most stable electron configuration, 2 electrons from neighboring atoms in crystal lattice form a pair connection. Such pairing (covalent bond) with 4 neighboring atoms provides silicon with stable electron configuration similar to rare gas argon (Ar).

From energetical point of view, the valence zone is fully taken and the conductive zone is empty. By providing additional energy, for example from light or heat source electron moves from valence to conductive zone. Electron can freely move on the crystal lattice. When electron leaves the valence zone, the free space is known as the hole, or so called defective electron. Creation of these defective electrons is responsible for inner semiconductors conductivity. Electrones and holes are always in pairs, so in other words there is the same amount of electrons and the holes.

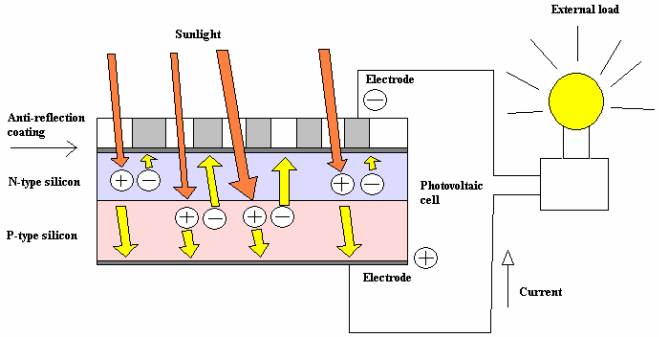


Fig.1 Photovoltaics solar cell

Current flowing through PN crossing can be formulated as algebraic sum of balanced heat flows of electric charge carriers. In state of equilibrium is the sum of the heat flows zero. Sums of electron and hole currents passing through the PN crossing are also zero. Absolute values of electron and hole currents from the N type semiconductor can be marked as I_n^N and I_p^N, from semiconductor type P can be marked as I_n^P and I_p^P.

In equilibrium state:

$$-I_n^N + I_p^N + I_n^P - I_p^P = 0 \quad (1)$$

$$-I_n^N + I_p^N = 0 \quad (2)$$

$$+I_n^P - I_p^P = 0 \quad (3)$$

Illumination will cause the increase in concentration of minority carriers. It will create the I_f current which flows through the PN crossing. When illuminated, Fermi´s level shatters to quasilevels for electrons and holes. Their difference φ resembles the voltage $U_f = \frac{\varphi}{e}$, which was created as the result of illumination.

In stationary state, the current flowing through the PN crossing equals to zero.

$$I_f - I_n^N + I_p^N + I_n^P - I_p^P = 0 \quad (4)$$

Majority carriers currents I_n^N and I_p^P will change because of illumination. Energetic levels are mutually shifted and levels of potential barriers are changed:

$$I_n^N = I_n^P \exp(\frac{\varphi}{kT}) \quad (5)$$

$$I_p^P = I_p^N \exp(\frac{\varphi}{kT}) \quad (6)$$

Using the equations (4), (5), (6) and after adjustments we get:

$$I_f - I_s[\exp(\frac{\varphi}{kT}) - 1] = 0 \quad (7)$$

For photoelectromotoric force:

$$U_f = \frac{\varphi}{e} = \frac{kT}{e}\ln(\frac{I_f}{I_s} + 1) \quad (8)$$

If PN crossing is connected to circuit where current I is flowing, using the (7), (8) equations we get:

$$I_f = I + I_s[\exp(\frac{\varphi}{kT}) -] \quad (9)$$

$$U_f = \frac{kT}{e}\ln(\frac{I_f - I}{I_s} + 1) \quad (10)$$

If PN crossing is connected to resistor $R = \frac{U_f}{I}$, equation (9) will be:

$$I_f = \frac{U_f}{R} + I_s[\exp(\frac{\varphi}{kT}) - 1] \quad (11)$$

In the case of small external resistors when: $I >> Is[\exp(\frac{\varphi}{kT}) - 1]$ we get $I_f \approx I$.

In the case of big external resistors when $I -> 0$, we get $I_f - I_s[\exp(\frac{\varphi}{kT}) - 1] = 0$.

If we connect the source of voltage to PN crossing we get:

$$I_f = \frac{U_f - U}{R} + I_s[\exp(\frac{\varphi}{kT}) - 1] \quad (12)$$

For the solar cell power we use this formula: P=U.I For the maximum output:

$$\frac{d(UI)}{dU} = I_k - I_s + I_s \frac{e}{kT}\exp\frac{eU_m}{kT} = 0 \quad (13)$$

For idle connection:

$$U_0 = \frac{kT}{e}\ln(\frac{I_k}{I_s} + 1) \quad (14)$$

B. Solar cell equivalent scheme, V-A characteristics

One diode enhanced model

Simple equivalent circuit is sufficient for most applications. The difference between measured and calculated values is only few percents. Only enhanced model (fig.2) describes exactly the solar cells behavior, especially in cases where different operational conditions have to be considered. Charge carriers in real solar cell show a voltage loss when passing through the PN crossing. Serial connected resistor R_S allows representation of this voltage loss. Parallel resistor R_P represents inner loses in cell. R_S value for real cells is about few milliohms (fig. 3) and R_P value is mostly higher than 10 Ω (fig. 4).

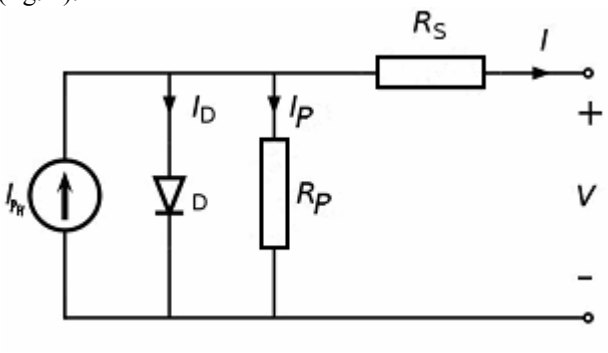


Fig.2 Solar cell equivalent scheme

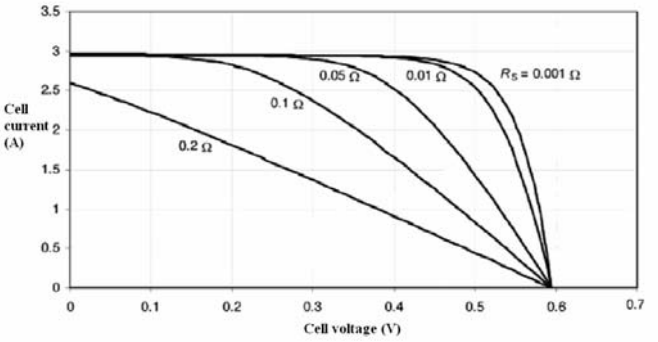


Fig.3 V-A Characteristics of Rs resistor

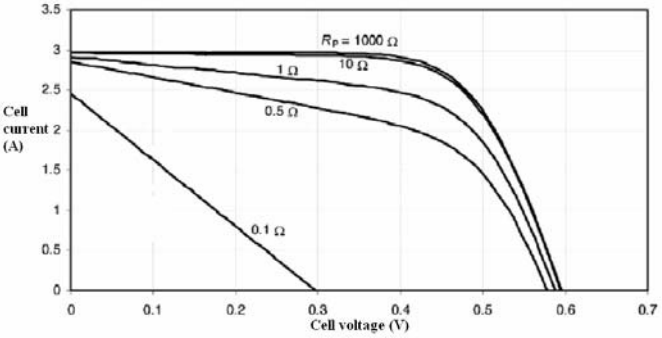


Fig.4 V-A Characteristics of Rp resistor

C. Photovoltaics on KTEEM – module and program

For measurements on department the photovoltaics cell QX6926 (fig.6) and SFH203 infra diode are used. We can calculate the power P [W] because of 4Ω resistor which is connected to cell and simulates the load.

Program (fig.5) was written for the measurement, which can collect the voltage or current values simultaneously from 4 devices. The results are daily written to *.csv file (fig.7) and are sent to specified ftp server at midnight, where they can be further evaluated (fig.8).

Figures 9 and 10 shows the classic bigger photovoltaic module which is commonly used for household or industrial applications and its technological parameters.

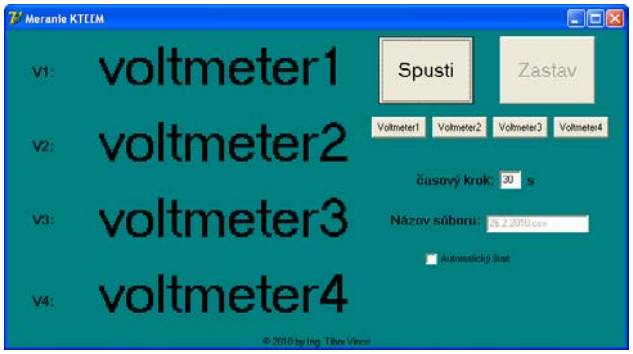


Fig.5 Measuring program main screen

Fig.6 Solar cell QX6926

544	1.3.2010	9:04:01	1,876	V	0,447	V
545	1.3.2010	9:05:01	1,843	V	0,447	V
546	1.3.2010	9:06:01	1,745	V	0,444	V
547	1.3.2010	9:07:01	1,644	V	0,439	V
548	1.3.2010	9:08:01	1,678	V	0,438	V
549	1.3.2010	9:09:01	1,754	V	0,438	V
550	1.3.2010	9:10:01	1,787	V	0,439	V
551	1.3.2010	9:11:01	1,819	V	0,441	V
552	1.3.2010	9:12:01	1,821	V	0,443	V
553	1.3.2010	9:13:01	1,809	V	0,442	V
554	1.3.2010	9:14:01	1,718	V	0,441	V
555	1.3.2010	9:15:01	1,619	V	0,438	V
556	1.3.2010	9:16:01	1,723	V	0,437	V
557	1.3.2010	9:17:01	1,786	V	0,44	V
558	1.3.2010	9:18:01	1,77	V	0,441	V
559	1.3.2010	9:19:01	1,721	V	0,441	V
560	1.3.2010	9:20:01	1,726	V	0,441	V
561	1.3.2010	9:21:01	1,775	V	0,442	V

Fig.7 Example of output file in *.CSV format

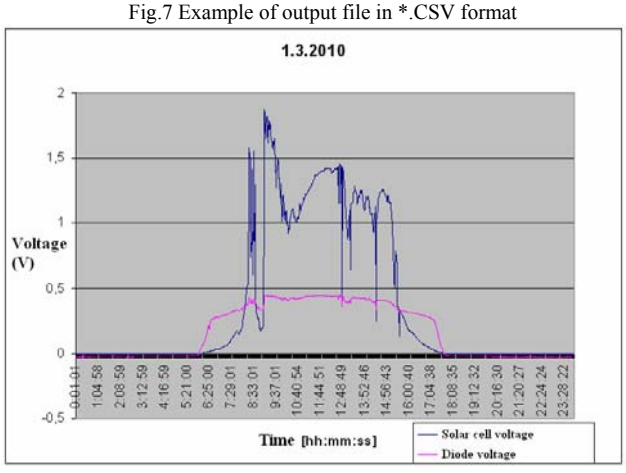


Fig.8 Example of daily graph - Voltage

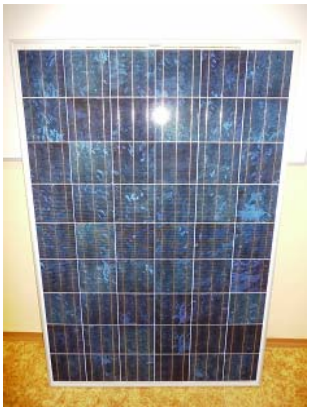


Fig.9 Photovoltaic panel

Model Number		STP200-18/Ud
Rated Maximum Power	(P_{max})	200W
Output Tolerance		±3%
Current at Pmax	(I_{mp})	7.63A
Voltage at Pmax	(V_{mp})	26.2V
Short-Circuit Current	(I_{sc})	8.12A
Open-Circuit Voltage	(V_{oc})	33.4V
Nominal Operating Cell Temp.	(T_{NOCT})	45°C±2°C
Weight		16.8kg
Dimension		1482×992×35(mm)
Maximum System Voltage		1000V
Maximum Series Fuse Rating		20A
Cell Technology		multi-Si
Application Class A		
All technical data at standard test condition AM=1.5 E=1000W/m² Tc=25°C		

Fig.10 Photovoltaic panel – technical parameters

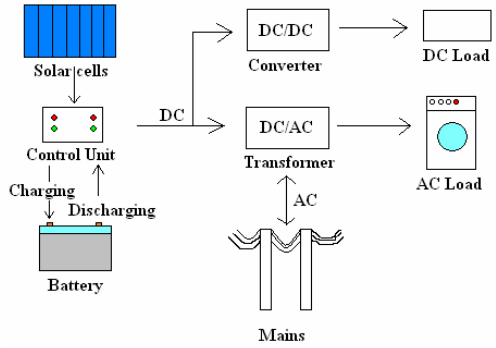


Fig.11 Example of solar cell system

III. WIND ENERGY

Wind energy is the indirect form of solar energy. Solar irradiation causes temperature differences on Earth and these are the origin of winds. Wind can achieve much higher energy concentration than solar radiation (10 kW/m^2 during violent storm and more than 25 kW/m^2 during hurricane, compared to maximum value of solar irradiation 1 kW/m^2). Slow wind speed about 5 m/s however has energy concentration about only 0,075 kW/m^2.

A. Wind from energetic point of view

Kinetic energy W in wind with speed v is equal to:

$$W = \frac{1}{2} m.v^2 \tag{15}$$

Power P of the wind with constant speed v, is:

$$P = W = \frac{1}{2} m.v^2 \tag{16}$$

Density ρ and content V of air determine its weight:

$$m = \rho.V \tag{17}$$

Weight of air with density ρ, which flows through the area S with speed v on trajectory š, can be calculated using this equation:

$$m = \rho.V = \rho.S.š \tag{18}$$

Power P of the wind will be:

$$P = \frac{1}{2}\rho.S.v^3 \tag{19}$$

Wind density ρ is changing due to air pressure p and temperature υ. It is directly proportional to pressure nex to temperature.

Ratio between wind power taken by turbine P_T and total wind power P_0 is called power coefficient C_P:

$$C_P = \frac{P_T}{P_0} = \frac{1}{2}.\left(1+\frac{v_2}{v_1}\right).\left(1-\frac{v_2^2}{v_1^2}\right) \tag{20}$$

Betz calculated ideal wind ratio which is:

$$\frac{v_2}{v_1} = \frac{1}{3} \tag{21}$$

After using the equation (20) we get so called Betz power coefficient:

$$C_{P\_Betz} = \frac{16}{27} \approx 0{,}593 \tag{22}$$

If wind turbine slows the wind from initial speed v1 to one third v_1 (v_2 = (1/3).v_1), then it is theoretically possible to achieve the maximum power which in the case of the wind turbine is 60%.

Real wind generators cannot achieve this theoretical optimum, however good systems have C_P coefficient between 0,4 a 0,5.

Ratio between used wind power P_T and ideal power P_{id} defines efficiency η of the wind generator.

$$\eta = \frac{P_T}{P_{id}} = \frac{C_P}{C_{P\_Betz}} \tag{23}$$

B. *Wind/solar energy system*

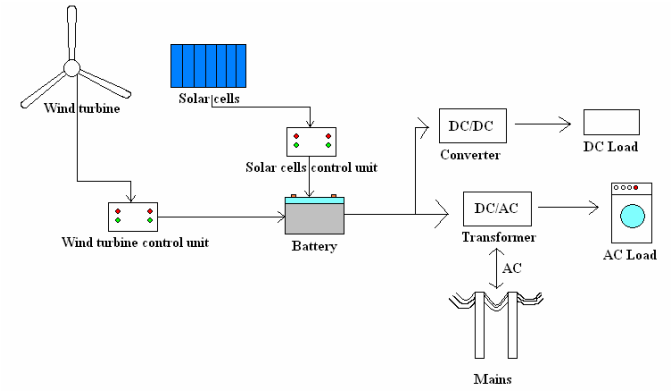


Fig.12 Example of combined solar/wind system

IV. Conclusion

Solar and wind energy can be considered as a real alternative because systems which utilize it already exist (fig.11,12) and achieve good results. Hopefully it will be a matter of few decades until they achieve a wide range utilization in every sphere of industry and household applications, as they are the only way how to get clean and harmless energy. If we would like to determine which energy source is more suitable for Slovak Republic, we must consider that the country has different landscape. Northern part of the country is mountainous, while the southern part is mostly low-country. For northern mountain part, building wind turbines on windward part of the hill is the solution, because relative speed of wind is much higher in mountains, than in low-lands. However in southern low-lands, which have more sunny days than northern mountains, the solar photovoltaic energy is more suitable. Plenty of open mostly flat space for photovoltaic panels with combination of a lot of sunny days makes photovoltaics reasonable choice. The best option however is to use combined wind/solar systems for best utilization of both resources.

Acknowledgment

The paper has been prepared by the support of Slovak grant projects VEGA 1/0660/08, KEGA 3/6386/08, KEGA 3/6388/08

References

[1] *QUASCHNING, Volker: Understanding Renewable Energy Systems, London, 2005, 260s, ISBN 1-84407-128-6*

[2] *TWIDELL, John - WEIR, Anthony: Renewable energy Resources, London, 2006, 597s, ISBN 9-78-0-419-25330-3*

[3] *PIMENTEL, David: Biofuels, Solar and Wind as Renewable Energy Sources, USA, 2008, 504s, ISBN 978-1-4020-8653-3*

[4] *FRERIS, Leon - INFIELD, David: Renewable Energy in Power Systems, 2008, West Sussex, 2008, 283s, ISBN 978-0-470-01749-4*

[5] *KOVÁČOVÁ, Irena - KOVÁČ, Dobroslav: Non-harmonic power measuring. In: Acta Electrotechnica et Informatica. Vol. 8, No. 3 (2008), pp. 3-6. ISSN 1335-8243.*

[6] *MOLNÁR, J. - KOVÁČOVÁ, I.: "Distance remote measurement of magnetic field", Acta Electrotechnica et Informatica, 2007, No.4, Vol.7, ISSN 1335-8243 (in print)*

[7] *MOLNÁR, Ján: Automated oscilloscope measuring. In: SCYR 2009 : 9th Scientific Conference of Young Researchers : Proceedings from conference : May 13th, 2009 Košice, Slovakia. Košice : FEI TU, 2009. s. 66-68. ISBN 978-80-553-0178-5.*

[8] *MOLNÁR, Ján: Telemetric system based on internet. In: OWD 2009 : 11 International PhD Workshop : Wisa, 17-20 October 2009. S.l. : S.n., 2009. p. 38-41. ISBN 83-922242-5-6.*

[9] *KOLLA, I. - KOVÁČ, D.: "Communication interfaces for measuring systems", Acta Electrotechnica et Informatica, 2007, No.4, Vol.7, ISSN 1335-8243 (in print)*

[10] *VINCE, Tibor: Measuring on remote computer through Internet using control Web 2000 and Metex MXD-4660A. In: 4. Doktorandská konferencia a ŠVOS TU v Košiciach FEI : Zborník z konferencie a súťaže, Košice, 13.5.2004. Košice : ETC GRAFO, 2004. s. 113-114. ISBN 80-968395-9-4.*

[11] *VINCE, Tibor - MOLNÁR, Ján - TOMČÍKOVÁ, Iveta: Remote DC motor speed regulation via Internet. In: OWD 2008 : 10th international PhD workshop : Wisla, 18-21 October 2008. p. 293-296. ISBN 83-922242-4-8.*

[12] *VINCE, Tibor: Motor speed regulation via internet and artificial neural network. In: SCYR 2009 : 9th Scientific Conference of Young Researchers : Proceedings from conference : May 13th, 2009 Košice, Slovakia. Košice : FEI TU, 2009. s. 107-109. ISBN 978-80-553-0178-5.*

Modeling and Simulation of the BLDC Motor

Tibor Balogh

Dept. of Electrical, Mechatronic and Industrial Engineering, FEI TU of Košice, Slovak Republic

tiborb@gmail.com

***Abstract*—The paper proposes a model of brushless DC motor considering behavior of the motor during commutation. The torque characteristic of BLDC motor presents a very important factor in design of the BLDC motor drive system, so it is necessary to predict the precise value of torque, which is determined by the waveforms of back-EMF. After development of simple mathematical model of the BLDC motor with sinusoidal and trapezoidal waveforms of back-EMF the motor is simulated in the MATLAB/Simulink environment. Based on analysis of the time responses a comparison study of results of both BLDC motor types is presented.**

***Keywords*—back-EMF, brushless DC motor, modeling, simulation**

I. Introduction

The Brushless Direct Current (BLDC) motor is rapidly gaining popularity by its utilization in various industries, such as Appliances, Automotive, Aerospace, Consumer, Medical, Industrial Automation Equipment and Instrumentation. As the name implies, the BLDC motors do not use brushes for commutation; instead, they are electronically commutated. The BLDC motors have many advantages over brushed DC motors and induction motors [1]. A few of these are:

- Better speed versus torque characteristics
- High dynamic response
- High efficiency
- Long operating life
- Noiseless operation
- Higher speed ranges

In addition, the ratio of torque delivered to the size of the motor is higher, making it useful in applications where space and weight are critical factors [2].

The torque of the BLDC motor is mainly influenced by the waveform of back-EMF (the voltage induced into the stator winding due to rotor movement). Ideally, the BLDC motors have trapezoidal back-EMF waveforms and are fed with rectangular stator currents, which give a theoretically constant torque. However, in practice, torque ripple exists, mainly due to emf waveform imperfections, current ripple and phase current commutation. The current ripple result is from PWM or hysteresis control. The emf waveform imperfections result from variations in the shapes of slot, skew and magnet of BLDC motor, and are subject to design purposes. Hence, an error can occur between actual value and the simulation results. Several simulation models have been proposed for the analysis of BLDC motor [1], [4], [5]. One of the models has a real back-EMF waveform whose appearance is close to a sinusoidal shape. This paper attempts to compare various types of BLDC motor models - with the trapezoidal and sinusoidal back-EMF waveforms.

II. Construction and Operating Principle

The BLDC motor is also referred to as an electronically commuted motor. There are no brushes on the rotor and the commutation is performed electronically at certain rotor positions. The stator magnetic circuit is usually made from magnetic steel sheets. The stator phase windings are inserted in the slots (a distributed winding), or can be wound as one coil on the magnetic pole. The magnetization of the permanent magnets and their displacement on the rotor are chosen in such a way that the back-EMF shape is trapezoidal. This allows the three-phase voltage system, with a rectangular shape, to be used to create a rotational field with low torque ripples. In this respect, the BLDC motor is equivalent to an inverted DC commutator motor, in that the magnet rotates while the conductors remain stationary. In the DC commutator motor, the current polarity is reversed by the commutator and the brushes, but in the brushless DC motor, the polarity reversal is performed by semiconductor switches which are to be switched in synchronization with the rotor position. Besides the higher reliability, the missing commutator brings another advantage. The commutator is also a limiting factor in the maximal speed of the DC motor. Therefore the BLDC motor can be employed in applications requiring high speed [1], [3].

Replacement of a DC motor by a BLDC motor place higher demands on control algorithm and control circuit. Firstly, the BLDC motor is usually considered as a three-phase system. Thus, it has to be powered by a three-phase power supply. Next, the rotor position must be known at certain angles, in order to align the applied voltage with the back-EMF. The alignment between the back-EMF and commutation events is very important. In this condition the motor behaves as a DC motor and runs at the best working point. But the drawbacks of the BLDC motor caused by necessity of power converter and rotor position measurement are balanced by excellent performance and reliability, and also by the ever-falling prices of power components and control circuits.

The simple motor model of a BLDC motor consisting of a 3-phase power stage and a brushless DC motor is shown in Fig. 1.

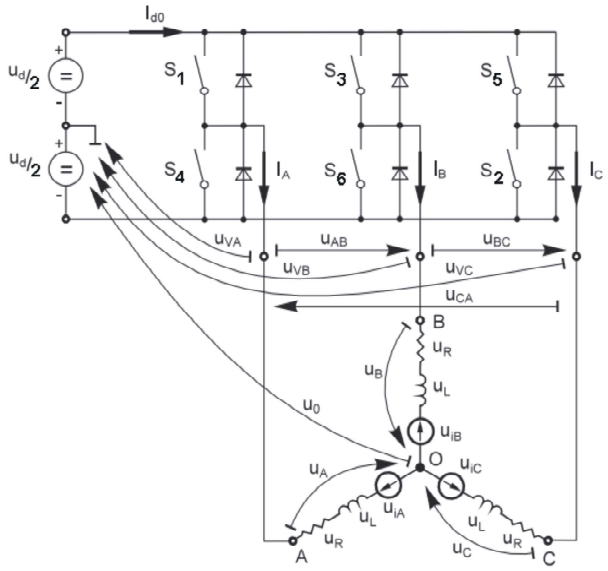


Fig. 1. BLDC motor model [3]

III. Mathematical Model of the BLDC Motor

Modeling of a BLDC motor can be developed in the similar manner as a three-phase synchronous machine. Since there is a permanent magnet mounted on the rotor, some dynamic characteristics are different. Flux linkage from the rotor depends upon the magnet material. Therefore, saturation of magnetic flux linkage is typical for this kind of motors. As any typical three-phase motors, one structure of the BLDC motor is fed by a three-phase voltage source. The source is not necessarily to be sinusoidal. Square wave or other wave-shape can be applied as long as the peak voltage does not exceed the maximum voltage limit of the motor. Similarly, the model of the armature winding for the BLDC motor is expressed as follows:

$$v_a = Ri_a + L\frac{di_a}{dt} + e_a \tag{1}$$

$$v_b = Ri_b + L\frac{di_b}{dt} + e_b \tag{2}$$

$$v_c = Ri_c + L\frac{di_c}{dt} + e_c \tag{3}$$

where
L is armature self-inductance [H],
R - armature resistance [Ω],
va, vb, vc - terminal phase voltage [V],
ia, ib, ic - motor input current [A],
and ea, eb, ec - motor back-EMF [V].

In the 3-phase BLDC motor, the back-EMF is related to a function of rotor position and the back-EMF of each phase has 120º phase angle difference so equation of each phase should be as follows:

$$e_a = K_w f(\theta_e)\omega \tag{4}$$

$$e_b = K_w f(\theta_e - 2\pi/3)\omega \tag{5}$$

$$e_c = K_w f(\theta_e + 2\pi/3)\omega \tag{6}$$

where
K_w is back EMF constant of one phase [V/rad.s^{-1}],
θ_e - electrical rotor angle [° el.],
ω - rotor speed [rad.s^{-1}].

The electrical rotor angle is equal to the mechanical rotor angle multiplied by the number of pole pairs:

$$\theta_e = \frac{p}{2}\theta_m \tag{7}$$

where
θ_m is mechanical rotor angle [rad].

Total torque output can be represented as summation of that of each phase. Next equation represents the total torque output:

$$T_e = \frac{e_a i_a + e_b i_b + e_c i_c}{\omega} \tag{8}$$

where
T_e is total torque output [Nm],
p - number of pole pairs.

The equation of mechanical part is represents as follows:

$$T_e - T_l = J\frac{d\omega}{dt} + B\omega \tag{9}$$

where
T_l is load torque [Nm],
J - rotor inertia [kgm^2],
B - friction constant [Nms.rad^{-1}].

IV. Simulink Model of the BLDC Motor

Fig. 2 shows the BLDC motor SIMULINK model. The model was developed in the rotor reference frame of the BLDC motor using equations shown above [5].

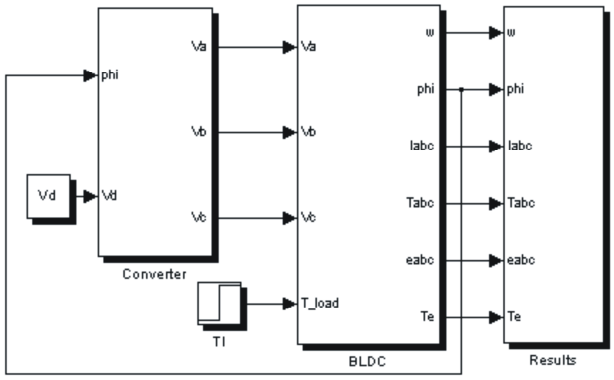


Fig.2. SIMULINK model of the BLDC

Fig. 3. shows the detail of the BLDC motor block. Fig.4. shows SIMULINK diagram of back-EMF. Fig. 4(a) is of trapezoidal back-EMF and Fig. 4(b) is of sinusoidal back-EMF. The trapezoidal functions and the position signals are stored in lookup tables that change their output according to the value of the electrical angle [5].

Unlike a brushed DC motor, the commutation of a BLDC motor is controlled electronically. To rotate the BLDC motor, the stator windings should be energized in a sequence. It is

important to know the rotor position in order to understand

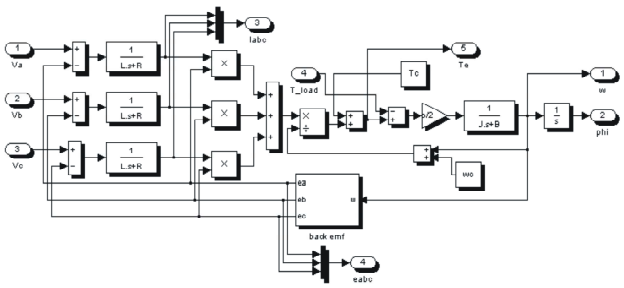


Fig. 3. SIMULINK model of the BLDC

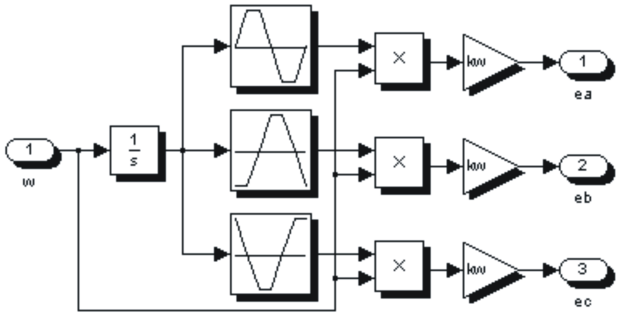


Fig. 4(a). Trapezoidal model of the back-EMF

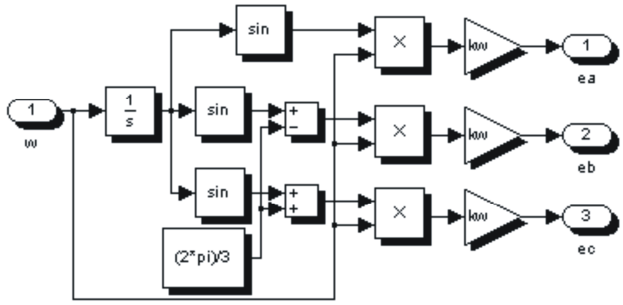


Fig. 4(b). Sinusoidal model of the back-EMF

which winding will be energized following the energizing sequence. Rotor position is sensed using Hall Effect sensors embedded into the stator. Most BLDC motors have three Hall sensors embedded into the stator on the non-driving end of the motor.

Whenever the rotor magnetic poles pass near the Hall sensors, the give a high or low signal, indicating the N or S pole is passing near the sensors. Based on the combination of these three Hall sensor signals, the exact sequence of commutation can be determined.

Fig.5. shows the detail of the converter block. The block was developed using equations below:

$$v_a = (S_1)V_d/2 - (S_4)V_d/2 \quad (10)$$

$$v_b = (S_3)V_d/2 - (S_6)V_d/2 \quad (11)$$

$$v_c = (S_5)V_d/2 - (S_2)V_d/2 \quad (12)$$

Every 60 electrical degrees of rotation, one of the Hall sensors changes the state. Given this, it takes six steps to complete an electrical cycle. Corresponding to this, with every 60 electrical degrees, the phase current switching should be updated.

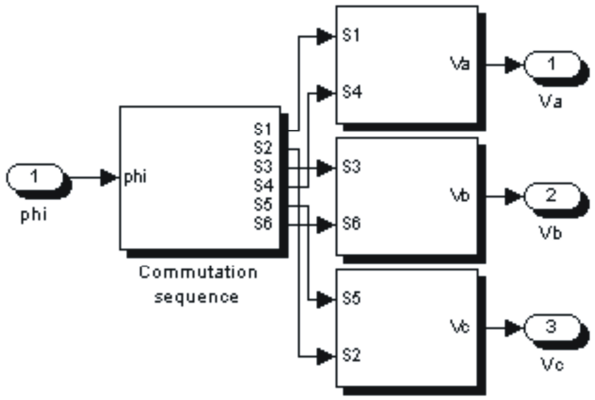


Fig. 5. SIMULINK model of the converter

However, one electrical cycle may not correspond to a complete mechanical revolution of the rotor. The number of electrical cycles to be repeated to complete a mechanical rotation is determined by the rotor pole pairs. For each rotor pole pairs, one electrical degree is completed. The number of electrical cycles/rotations equals the rotor pole pairs [6].

The commutation sequences are shown in Table I.

TABLE I.
ELECTRICAL DEGREE, HALL SENSOR VALUE AND CORRESPONDING COMMUTED PHASE IN CLOCKWISE ROTATION OF THE ROTOR.

Electrical degree	Hall sensor value (ABC)	Phase	Switches
0º - 60º	101	A-C	S1-S2
60º - 120º	001	B-C	S2-S3
120º - 180º	011	B-A	S3-S4
180º - 240º	010	C-A	S4-S5
240º - 300º	110	C-B	S5-S6
300º - 360º	100	A-B	S6-S1

V. SIMULATION RESULTS

Fig. 6 - 9 shows results of simulation of a BLDC motor with the following parameters: Vd = 100V, R = 4,98Ω, L = 5,05 mH, p = 4, J = $15{,}17.10^{-6}$ kgm^2, B = 10^{-3} Nms/rad, kw = $56{,}23.10^{-3}$ V/rad.s^{-1}, Tl = 0.3 Nm. The load time is 0,15 s. Fig. 6 shows source voltage waveforms that are switched according to the commutation sequences in Table I.

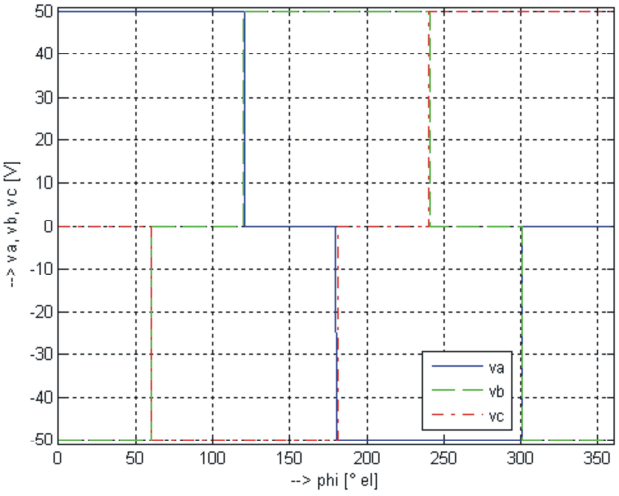


Fig. 6 Voltage source waveforms

The time courses in Fig. 7(a) belong to the BLDC motor with trapezoidal back-EMF and the responses in Fig. 7(b) are valid for the sinusoidal back-EMF BLDC motor. There is a very good correlation with the responses of the BLDC motor models known in the references [5], [6].

Fig. 7 shows speed responses of the trapezoidal and sinusoidal models of BLDC motor. Speed responses of both motor types are in principle the same.

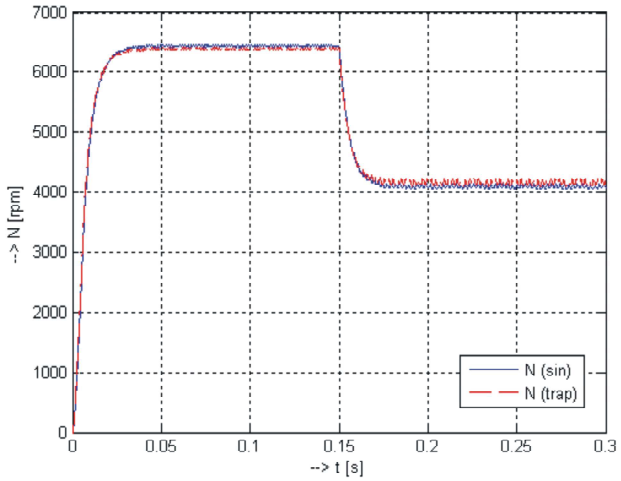


Fig. 7 Speed waveforms of BLDC motor

Fig. 8 shows back-EMF of both models of the BLDC motor. Fig. 8(a) is of trapezoidal back-EMF and Fig. 8(b) is of sinusoidal back-EMF.

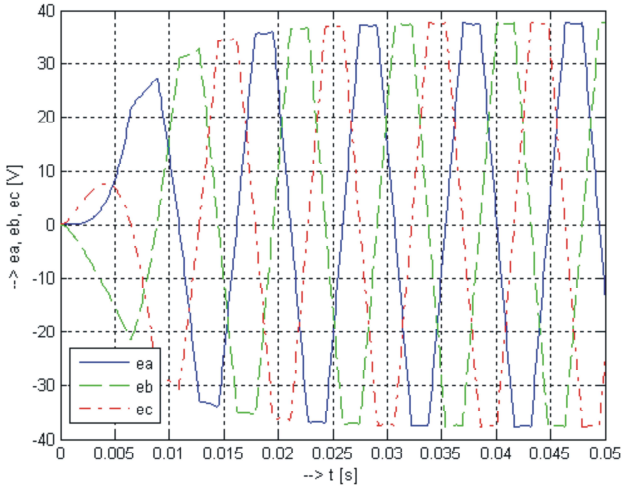


Fig. 8(a) Time courses of trapezoidal back-EMF

Fig. 9. shows the electromagnetic torque of the trapezoidal and sinusoidal BLDC motor models. The torque in both cases is in principle the same.

VI. Conclusion

The modeling procedure presented in this paper helps in simulation of various types of BLDC motor. The performance evaluation results show that, such a modeling is very useful in studying the drive system before taking up the dedicated controller design, accounting the relevant dynamic parameters of the motor.

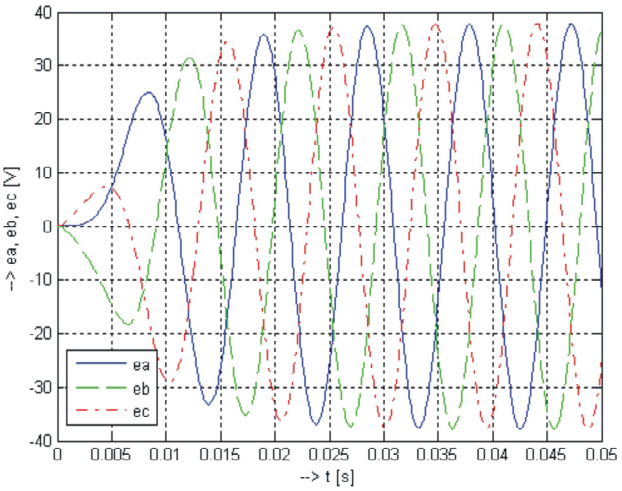


Fig. 8(b) Time courses of sinusoidal back-EMF

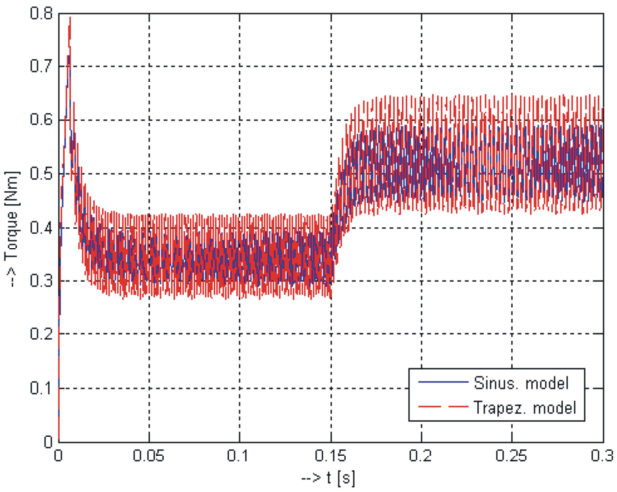


Fig. 9 Electromagnetic torque of BLDC motor

Acknowledgment

This work was supported by Slovak Cultural and Educational Agency of the Ministry of Education of Slovak Republic under the contract KEGA 103-039 TUKE-4/2010 “Students’ Skills Development for Mechatronic Systems Control”.

References

[1] B.Indu Rani, Ashly Mary Tom, “Dynamic Simulation of Brushless DC Drive Considering Phase Commutation and Backemf Waveform for Electromechanical Actuator”, IEEE TENCON 2008, Hyderabad. ISBN: 978-1-4244-2408-5.

[2] Padmaraja Yedamale, “Brushless DC (BLDC) Motor Fundamentals”, Microchip Technology Inc., 2003.

[3] Pavel Grasblum, “Sensorless BLDC motor control using an 8-bit MCU”, Freescale.

[4] Y.S. Jeon, H.S.Mok, G.H. Choe, D.K. Kim and J.S. Ryu, “A New Simulation Model of BLDC Motor with Real Back EMF waveforms”, IEEE CNF. On Computers in Power Electronics, 2000. COMPEL 2000. pp. 217 – 220, July 2000.

[5] S. Baldursson, “BLDC Motor Modelling and Control – A Matlab/Simulink Implementation”, Master Thesis, May, 2005.

[6] W. Hong, W. Lee and B. K. Lee, “Dynamic Simulation of Brushless DC motor Drives Considering Phase Commutation for Automotive Applications”, IEEE International Electric Machines & Drives Conference, 2007. IEMDC 2007, pp.1377-1383, 3-5 May 2007.

Spread Spectrum Steganography with Error Correction

[1]*Vladimír BÁNOCI,* [2]*Gabriel BUGÁR*

[1]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]vladimir.banoci@tuke.sk, [2]gabriel.bugar@tuke.sk

***Abstract*—In this paper, we present a novel method of steganography system based on CDMA (Code Division Multiple Access) approach having regard to perceptibility of stego-image. Also additional extraction algorithms were designed due to enhancing of secret message decomposition. As it will be shown later, suitably imposed features of CDMA techniques are in consonance with imperatives claimed upon steganography systems.**

***Keywords*— Steganography, Discrete Cosine Transformation (DCT), CDMA, PN, PN – Autocorrelation, Code Book**

I. INTRODUCTION

Information hiding, steganography, and watermarking are three closely related fields that have a great deal of overlap and share many technical approaches. However, there are fundamental philosophical differences that affect the requirements, and thus the design, of a technical solution.

Information hiding (or data hiding) is a general term encompassing a wide range of problems beyond that of embedding messages in content. The term *hiding* here can refer to either making the information imperceptible (as in watermarking) or keeping the existence of secret information. Some examples of research in this field can be found in the International Workshops on Information Hiding, which have included papers on such topics as maintaining anonymity while using a network and keeping part of a database secret from unauthorized users [1].

Steganography is a hiding technique that has been mainly used in various applications. The basic principle lies in embedding the secret message into a camouflage media to ensure that an unintended party will not be aware of the existence of the embedded secret in stego-media hence its goal is to hide the presence of communication.

Images are the most popular cover media for steganography. A popular digital steganography technique is so-called Least Significant Bit (LSB) embedding. With the LSB embedding technique, the two parties in communication share a private secret key that creates a random sequence of samples of a digital signal. The secrete message, possibly encrypted, is embedded in the LSBs of those samples of the sequence.

In the last decade, several steganographic schemes have been developed to solve the privacy problem. Among these schemes is an approach that hides a secret message in the spatial domain of the cover image [3][5][6].

In Lee and Chen's method [3], the least significant bit (LSB) of each pixel in the cover image is modified to embed a secret message. In Wang et al.'s method [5], the optimal substitution of LSB is exploited. These schemes have a higher quality stego-image but are sensitive to modification. In Chung et al.'s method [6], singular value decomposition (SVD) based hiding scheme is proposed. In Tsai et al.'s scheme [4], the bit plane of each block truncation coding (BTC) block is exploited to embed a secret message. For such reasons, many imagery steganographic methods have been invented. Here we briefly review research carried out particularly in the Discrete Cosine Transformation (DCT) domain as *JSteg* method that hides information sequentially in LSBs of the quantized DCT coefficients (qDCTCs) while skipping 0's and 1's; *OutGuess* method scatters information into the LSB of qDCTCs [9]. Another method employs the technique of matrix encoding to hold secret information using LSB of qDCTCs in F5. Others researches of using DCT transformation are mentioned in [10][11].

In this paper a proposed steganography method is represented by hiding secret information (*secret-object*) in other regular inconspicuousness information (*cover-object*). Conventional steganography models use the knowledge of cover-object and stego-object to carry out embedding and extraction process. The embedding process is represented by hiding secret image in each block of quantized DCT coefficients of cover image. The acquisition of proposed methods using pseudo-random number (*PN*) sequences is upgrading of security level in steganography systems. Firstly, there is a smaller distortion of secret message, in this manner; the stego-image quality degradation is more imperceptible to the human eye. Secondly, adding the cryptographic element to steganographic algorithm in process of embedding is represented by characteristics of the CDMA. Moreover, CDMA mathematical apparatus applied in our method is instrumental in extraction of secret message ergo it allows lower embedding energy of secret message. Experimental results have demonstrated that applied CDMA has its contribution in hiding communication of steganography system.

II. RELATED WORK

In this section some needed attainments to understanding the proposed method are presented. The first mentioned is the discrete cosine transformation (DCT).

DCT transforms signal or image from spatial domain to frequency domain. In this method all experiments have been done with 2D - DCT, which is given by the following equation [12]:

$$F(u,v) = C(u)C(v)\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} f(i,j)\cos\left(\frac{(2i+1)u\pi}{2N}\right)\cos\left(\frac{(2j+1)v\pi}{2N}\right)$$

Variable *F(u,v)* is block of transformed coefficients of input picture elements *f(i,j)*. Variables *C(u)* and *C(v)* are given by:

$$C(u) = \begin{cases} \sqrt{\frac{1}{N}} & u = 0 \\ \sqrt{\frac{2}{N}} & otherwise \end{cases} \qquad C(v) = \begin{cases} \sqrt{\frac{1}{N}} & v = 0 \\ \sqrt{\frac{2}{N}} & otherwise \end{cases}$$

Energy of transformed frequency domain is centered into coefficients with lower spatial frequencies. The most energy is embodied in coefficient with zero spatial frequency – DC coefficient. The energy is decreasing towards lower frequencies, what usually depends on nature of the cover image. Each DCT transformation coefficient can bear information equal to 1 bit of secret message. If more coefficients are affected in process of embedding secret message, more distortion is generated, also what results to higher undetectability against attacker. This approach underlies the DCT method that is used as reference method for comparison to our proposed methods.

The basic principle of embedding secret object is defined in sense of similarities stego- and cover-object and its measure can be expressed by function of conformity. However, this function of conformity does not have practical use. Hence, the objective measures, which stem from statistical approach, are generally used in practice. Their application is based on measuring the distortion between cover data *f(i,j)* with resolving capacity M x N of picture element and modified data $\hat{f}(i,j)$ with same resolving capacity [11]. The most frequent used criterion of evaluating the quality of reconstructed information is Peak Signal to Noise Ratio (PSNR). The PSNR value is used in all our practical examples as objective simulation model of human visual perception.

The CDMA technique emanates from Direct Spread Spectrum (DSS) approach that stands on mathematical apparatus of orthogonal pseudorandom number sequences (PN Sequence) which allow multiple access on the same medium without interference. This fact is achieved by orthogonality and auto-correlation phenomenon of PN sequences, where auto-correlation represses the influencing of noise on steganographic channel. Direct spreading technique (DSS) by PN sequences incites to use attributes applicable in steganography. Firstly, PN sequences, depending on its type, have subequal quantity of -1 and 1, thus secret data do not directly figure in bulks or other periodical sequences, which can be an incitement to attacker about concealed communication. Secondly, spreading information bit by PN sequence allows using its correlation characteristic. The longer length of PN sequence, better correlation results can be achieved, therefore higher rate of defective bits during the extraction can be admitted.

III. Design of Proposed Algorithm

The secret data has to be accommodated to transfer channel before embedding itself. The preparation consists of two phases. Firstly, selection of the length of *Code word*, which defines the size of secret data stream segmentation. Analogically, the length of Code word b_i, *i=1,2,...,n* linearly determines the number of PN sequences used in the spreading process, ergo CDMA approach, what is shown in Fig. 1. Secondly, the secret key can be used in the process of PN sequences alignment, in which the secret data are spread. In the process of extraction this key denotes position of bits (secret information) in which they are reconstructed from stego-object. However, one noticeable disadvantage of this approach is that this secret key has to be transferred and present on the receiver side, what can possibly lead to exposure of secret communication.

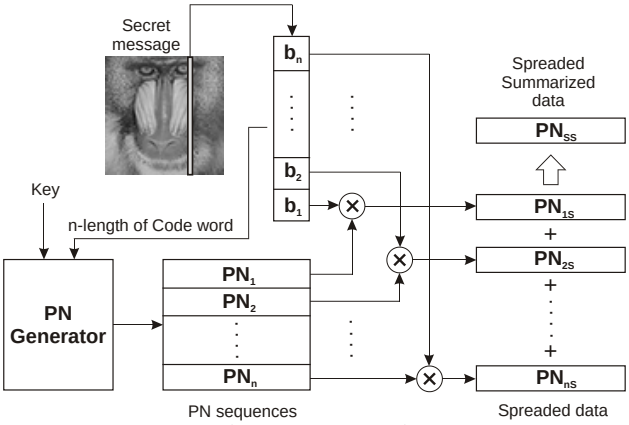


Fig. 1. Algorithm of spreading secret information by using code word

Subsequently, when the secret data are spread, they are embedded to the transformation domain of DCT coefficient with reference to desired attributes of steganography system as imperceptibility and system capacity. One of used parameters is *offset* parameter, which denotes number of DCT coefficient induced in descending order of its frequencies that are modified in process of embedding; another such as *alfa (α)* parameter multiplies secret data. If parameter $\alpha > 1$ then the steganographic system is more robust, hence the secret information is more resistive to distortion in form of channel noise. The aim is to find appropriate balance between human eye imperceptibility on cover object modification with enough capacity to transfer secret information.

The main issue we can encounter on receiver side is the problem of stego object distortion. During transmission the stego object could be affected by noise of the communication channel. The second problem related to extraction of secret information from stego-object is rounding of modified cover-object consists of decimal numbers. These value-added errors are often responsible for false extraction of secret data. For that reason, PN sequences with its autocorrelation attributes significantly improve extraction results, where it is generally known that longer PN sequences are accounted for uplifting auto-correlation attributes, hence better acquired results. This concept is provided by algorithm *Regular Extraction* (RE) , where its results are presented in the next chapter. Moreover additional approach called *Extraction with Error Correction* (EEC) was used in the process of extraction to enhance error correction. This method calculates all possible states (2^n) on output of defined code word could be considered as reference data cell. Thereafter, extracted data are cross-correlated with iteration algorithm with all possible states, where the maximum value determines the most likely transferred spread sequence of each code word. However, incrementing of code word ($n > 5$) requires higher computer power also in relation to

the size of secret information what caused that time for acquiring results would not have be longer in the reasonable limits.

IV. RESULTS

The capacity of proposed method is equal to the method of direct secret data embedding in transformation domain. The capacity of particular method was calculated depending on offset variable. During the practical experiment the offset value is preset to *offset=8*, what practically means that first seven DCT coefficient are not used in process of embedding. The total number of DCT coefficient used in process of embedding represents the capacity in binary form and it does not change with increasing of PN sequences. The difference can be found in measure of energy, which modifies the cover-object. Subsequently, if more PN sequences are used in process of spreading then more energy is embedded to cover-object in transformation domain of DCT transform. That results in degradation of PSNR coefficient, which represents imperceptibility of human perception or distortion of cover-object in general. Tab. 1 and Fig. 2 show relation of PSNR (in dB) to number of PN sequences used in process of spreading with different robust parameter α. The embedding picture in mentioned experiment results is boy69x70.bmp (6118 B) and cover image is lena256x256.bmp (66 614 B) with total 49,4% of method's utilized capacity.

n	PSNR [dB]				
	=0.25	*α= 0.5*	*α= 0.75*	*= 1*	*α= 1.25*
2	58,80	51,94	48,86	46,57	44,72
3	54,07	48,77	45,49	43,07	41,18
4	51,22	45,79	42,38	39,92	38,00
5	48,51	42,78	39,33	36,84	34,91

Tab. 1. The PSNR value (in dB) of the stego-images by using various length code and robust parameter α.

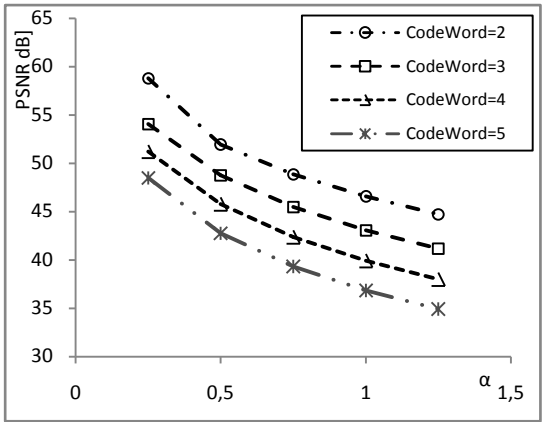


Fig. 2. The PSNR value (in dB) of the stego-images by using various length code and robust parameter α.

However, using of longer PN sequences tends to better extraction results as it is depicted on Fig. 3 - Fig. 6 for different α. Moreover, same figures show that algorithm EEC achieves better results rather than RE algorithm for all variation of α. The differences are even more exposed for longer code words, where better cross-correlation attribute of PN sequences stands its place.

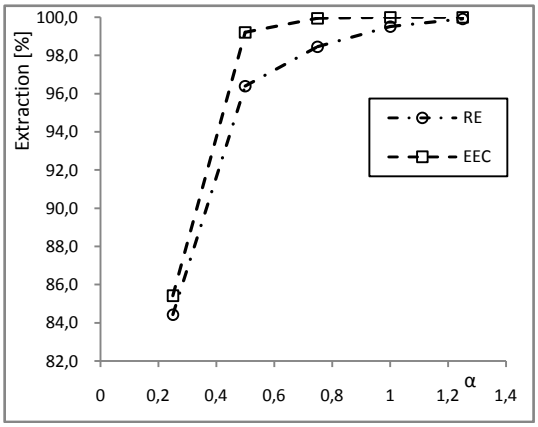


Fig. 3. The comparison of two extraction algorithms with various α and code word of length *n=2*.

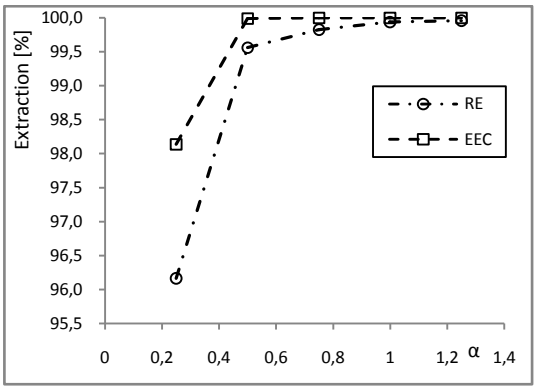


Fig. 4. The comparison of two extraction algorithms with various α and code word of length *n=3*.

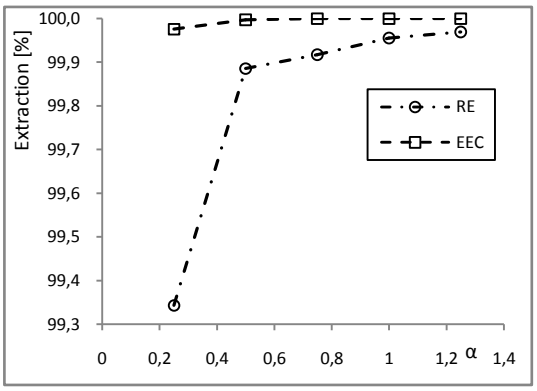


Fig. 5. The comparison of two extraction algorithms with various α and code word of length *n=4*.

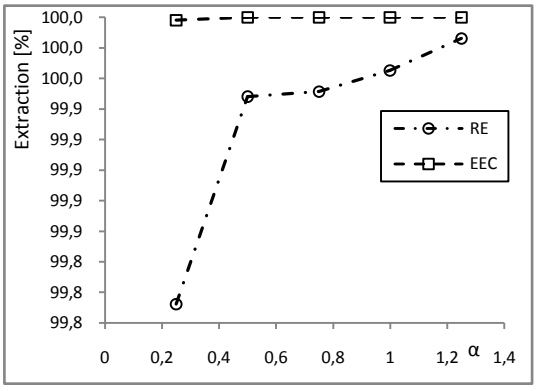


Fig. 6. The comparison of two extraction algorithms with various α and code word of length *n=5*.

V. CONCLUSION

The objectives of proposed method and embedding algorithm were to solve problem of capacity problem of previous research [1], where the incrementing of number PN sequences used in spreading process led to lowering the method's capacity. As it has been shown in this paper, the capacity is not longer depended on number of PN sequences. Moreover the proposed method with EEC algorithm promotes extraction result as it was presented in the previous chapter. Another criterion considered according steganography imperatives is imperceptibility of human observer to modification of cover-object. The PSNR value is decreasing with employment of more PN sequences. This difficulty is solved by decrementing of α parameter by the reason of improved cross-correlation attributes in case of longer PN sequences. The practical application of PN sequences' attributes as autocorrelation and orthogonality were presented in proposed methods. The autocorrelation allowed correct identification of PN sequences and properly decomposition of secret message. Whole steganographic system based on CDMA can be easily modified according to desired requirements by alternation of elected parameters as the secret key *k*, α and *offset* parameters.

ACKNOWLEDGMENT

Research described in the paper was financially supported by Ministry of Education of Slovak republic VEGA Grant No. 1/0065/10, INDECT Grant (7th Research Frame Programme no. 218086) and Centre of Information and Communication Technologies for Knowledge Systems
(project number: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF.

REFERENCES

[1] BÁNOCI V., BUGÁR G., LEVICKÝ D., Steganography system using by CDMA techniques, Radioelectronics conference, 2009, pp. 183-186

[2] COX, I. J., MILLER, M. L., BLOOM, J. A., FRIDRICH, J., KALKER, T. Digital Watermarking and Steganography. USA, 2008. ISBN 978-0-12-372585-1.

[3] LEE, Y.K., CHEN, L. H., High capacity image model. *IEEE Proc. Vision, Image Signal process. 147(3)(2000) 288-294.*

[4] TSAI, P., HU, Y. C., CHANG, C. C. An image hiding technique using block truncation coding. *Proceedings of Pacific Rim Workshop on Digital Steganogarphy.* Kitakyaushu, Japan, July 2002, pp. 54-64

[5] WANG, R. Z., LIN C. F., LIN, J. C. Image hiding by optimal LSB substitution and genetic algorithm. Pattern Recogn. 34 (3) (2001) 671-683.

[6] CHUNG, K. L., SHEN, C. H., CHANG, L.C. *A novel SVD- and VQ-based image hiding scheme.* Pattern Recogn. Lett. 22 (9) (2001) 1051-1058.

[7] DUMITRESCU, S., WU, X., WANG. Z. Detection of LSB Steganography via Sample Pair Analysis. *IEEE Transactions on Signal Processing*, Vol. 51, No. 7, July 2003.

[8] UPHAM, D. Jsteg, Software available at <ftp.funet.f1>, 2000.

[9] PROVOS, N. Defending against statistical Steganalysis. *Proceeding of the 10th USENIX Security Symposium*, 2001, pp. 323–335.

[10] SALLE, P. Model based steganography, in: *International Workshop on Digital Watermarking*, 2003, pp. 154–167.

[11] MIYAKE, K., IWATA, M., SHIOZAKI, A. Digital steganography utilizing features of JPEG images. *IEICE Trans. Fundam.* E87-A, April 2004, pp. 929–936.

[12] HUNG, A. C., MENG, TH-Y. A Comparison of fast DCT algorithms. *Multimedia Systems*, No. 5 Vol. 2, December 1994.

Bi-directional DC/DC Converter controlled by UC3637

[1]*Tomáš Béreš,* [2]*Martin Olejár,* [3] *Ľubomír Matis*

[1,2,3] Dept. of Electrical, Mechatronic and Industrial Engineering, FEI TU of Košice, Slovak Republic

[1]tomas.beres@tuke.sk, [2]martin.olejar@tuke.sk, [3]lubomir.matis@tuke.sk

***Abstract*— Concept of hybrid battery with bi-directional buck-boost DC/DC converter controlled by UC3637 is described in this paper. The first part of the paper is aimed at concept of hybrid battery. Design of power circuit and control circuit with UC3637 of converter is described in the second part of the paper. Experimental results from measuring of converter are mentioned in last part.**

***Keywords*— Converter control, hybrid battery, pulse width modulation (PWM),**

I. INTRODUCTION

The last years are characterized by rapid development of electronic systems, which uses an accumulator as a basic power supply. However, presently the accumulators are the weakest element of the power electronic supply system. It is caused by low dynamics of input power, temperature dependence, short lifetime and a lot of other limitations. The most significant improvement in recent 200 years has been achieved by developing ultra capacitor (UCAP). The ultracapacitor has much better electrical parameters than conventional accumulator. The next table shows comparison of the features of ultra-capacitor, accumulator and classic capacitor.

Available Performance	Accumulator	Ultra – capacitor	Classic capacitor
Charge Time	1 – 5 hrs	0,3 – 30 s	10^{-3} – 10^{-6}s
Discharge Time	0.3 – 3 hrs	0,3 – 30 s	10^{-3} – 10^{-6}s
Energy (Wh/kg)	10 - 100	1 – 10	<0,1
Cycle life	1000	10^6	-
Specific Power	< 1000	10 000	>100 000
Charge/Discharge Efficiency	0.7 – 0.85	0.85 – 0.98	> 0.95

Tab.1. Parameter comparison of ultra-capacitor with accumulator and classic capacitor

At present the low energy density is main disadvantage of ultra-capacitors. One of the possibilities is to fuse the advantages of ultra-capacitors and high energy density of accumulators to a hybrid secondary power source.

II. CONCEPT OF HYBRID BATTERY

Hybrid battery (HB) is a name for an improved topology of secondary voltage power source. Its output power dynamics and lifetime considerably exceed the recent types of accumulators. The hybrid battery is in nature a cascade connection of an ordinary accumulator with an ultracapacitor via a bi-directional DC/DC converter as it is seen in Fig. 1.

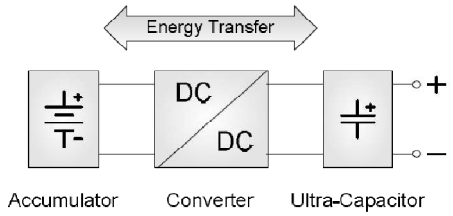


Fig. 1. Block diagram of hybrid battery

High dynamics of input-output power of the hybrid battery is achieved due to the ultra-capacitor. It means that high dynamic parameters of the hybrid battery are given by the parameters of the ultra-capacitor and static parameters by the accumulator. Bi-directional DC/DC converter is a main part of a hybrid battery. The converter has essential influence on the operational properties and the efficiency. Recuperation conditions of the bi-directional DC/DC converter are given by the use of an accumulator in hybrid battery.

III. DESIGN OF DC/DC CONVERTER

The parameters of proposed DC/DC converter are shown in Table 2.

Parameter	**Value**
Input voltage Uin	15-30 V
Output voltage Uout	14/24 V
Max. output voltage ripple ΔUout	5%
Max. output current	10A
Max. current ripple	1A
Switching frequency	50kHz
Efficiency	>80%

Tab. 1. Table of parameters

A. Power circuit of DC/DC converter

Power circuits of the DC/DC converter are in the Figure 2.

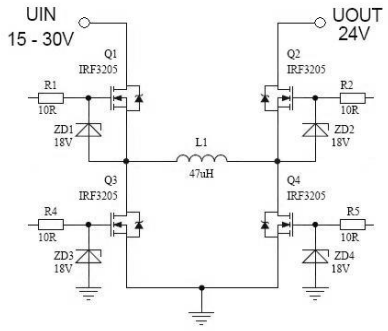


Fig.2.Topology of bi-directional buck-boost DC/DC converter

The bi-directional converter consists of two buck-boost converters connected in cascade. This converters are interconnected through inductance i.e. boost converter with buck converter (Fig.2).

The value of output voltage in general is:

$$V_{OUT} = V_{IN}\frac{D_1}{D_2} \tag{1}$$

where:

$$D_1 = \frac{t_{Q1(ON)}}{T}; \qquad D_2 = \frac{t_{Q2(ON)}}{T} \tag{2}$$

$t_{Q1(ON)}$ and $t_{Q2(ON)}$ indicate the ON time of the MOSFET switches Q_1 and Q_2 respectively, whereas T is the switching period.

Cascaded buck-boost converter can work in three operation modes, which will be introduced below.

a) Buck mode($V_{IN} > V_{OUT}$)

Transistor Q_2 is always ON and Q_4 is always OFF during this mode (D_2 = 1). Only Q_1 and Q_3 are switching synchronously. In this operation mode the cascaded buck-boost converter works as classic buck converter. Then the value of V_{OUT} is for buck mode as follows:

$$V_{OUT} = V_{IN}.D_1 \tag{3}$$

b) Buck-boost mode ($V_{IN} \approx V_{OUT}$)

In this switching mode all four MOSFET's operate during the period. The first path (Q_1, Q_4 are ON) enables charging the inductor, the second path (Q_2, Q_3 are ON) allows the energy stored in the inductor to be delivered to the output capacitor. This way of switching determines the following relation between D_1 and D_2:

$$D_2 = 1 - D_1 \tag{4}$$

By combination of the equations (1) and (4), the following expression is obtained:

$$V_{OUT} = V_{IN}.\frac{D_1}{1-D_1} \tag{5}$$

c) Boost mode ($V_{IN} < V_{OUT}$)

Transistor Q_1 is always ON and Q_3 is always OFF during the period in this mode (D_1 = 1). Only Q_2 and Q_4 are switching synchronously. In this operation mode the cascaded buck-boost converter works as classic boost converter. Then the value of V_{OUT} is for boost mode as follows:

$$V_{OUT} = \frac{V_{IN}}{D_2} \tag{6}$$

B. Control circuit with UC3637

Switched mode controller UC3637 is used for control of this converter. Scheme of control circuit is shown in Figure 4.

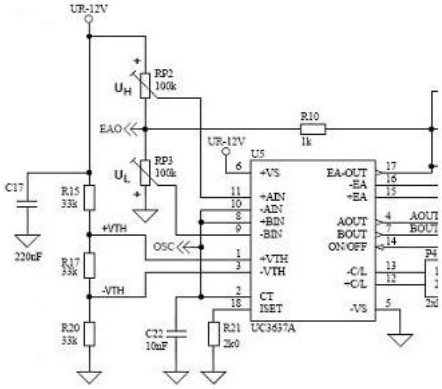


Fig. 4. Control circuit with UC3637

Amplitude of triangle waveform oscillator ($+U_{TH}$; $-U_{TH}$) is set by voltage divider R15, R17, and R20. Value U_H and U_L is set by trimmer RP2 and RP3. Changing modes of converter depend on value U_H and U_L as follows:

If: $U_{EAO}-U_L < U_{-VTH}$ and $U_{EAO}+U_H < U_{+VTH}$
buck mode is set
If: $U_{EAO}-U_L > U_{-VTH}$ and $U_{EAO}+U_H < U_{+VTH}$
buck-boost mode is set
If: $U_{EAO}-U_L > U_{-VTH}$ and $U_{EAO}+U_H > U_{+VTH}$
boost mode is set
where U_{EAO} is output from voltage PI regulator.

For better understanding the function of controller is shown in Figure 5.

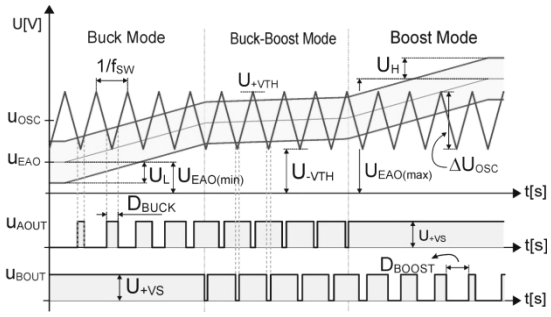


Fig. 5. Function of control circuit

Voltage u_{AOUT} is input for transistors Q_1 and Q_3 driver, voltage u_{BOUT} is input for transistors Q_2 and Q_4 driver. Transistors Q_1 and Q_2 are switched by non-inverted signal and transistors Q_3 and Q_4 are switched by inverted signal.

IV. EXPERIMENTAL RESULTS

The function of the proposed DC/DC converter was verified on the laboratory model. Principle of signal creation with controller UC3637 for drivers is displayed on the following oscillograms.

The first oscillogram (Fig.6) was captured in buck mode of the converter. Transistors Q_1, Q_3 are switched. Transistor Q_2 is always ON, Q_4 is always OFF in this mode. The second oscillogram (Fig.7) shows buck-boost mode operation of the converter. In this mode all transistors are switched in diagonal pairs.

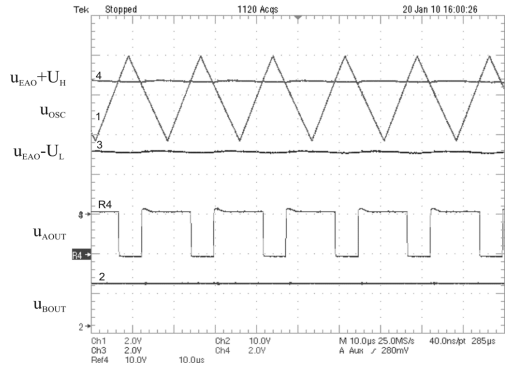


Fig. 6. Control signal creation with UC3637 for drivers in buck mode (U_{IN}=28V,U_{OUT}=24V)

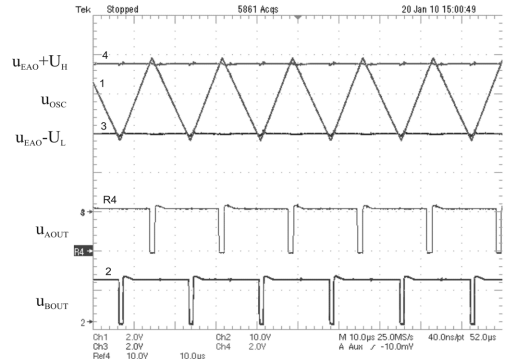


Fig. 7 Control signal creation with UC3637 for drivers in buck-boost mode (U_{IN}=25V, U_{OUT}=24V)

Boost mode operation of the converter is displayed in the third oscillogram (Fig.8). Transistors Q_2, Q_4 are switched. Transistor Q_1 is always ON and transistor Q_3 OFF.

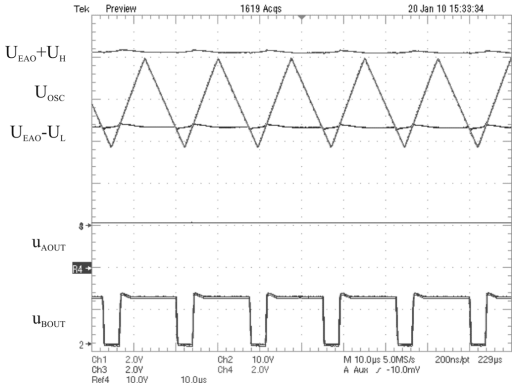


Fig. 8. Control signal creation with UC3637 for drivers in boost mode (U_{IN}=18V, U_{OUT}=24V)

Voltage of transistors Q_1-Q_4 and inductor current i_L in buck-boost mode is shown in Fig.9. Correct function of converter control is shown in this oscillogram. All four transistors are switched on simultaneously.

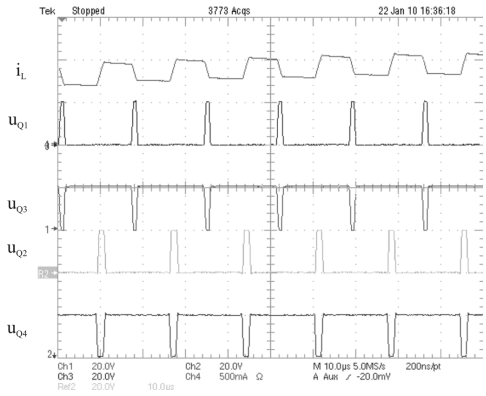


Fig. 9. Voltage of transistors Q_1-Q_4 and current of inductor in buck-boost mode (U_{IN}=20V, U_{OUT}=20V, I_{OUT}=0,6A)

Experimental model of DC/DC converter is in the Fig.10.

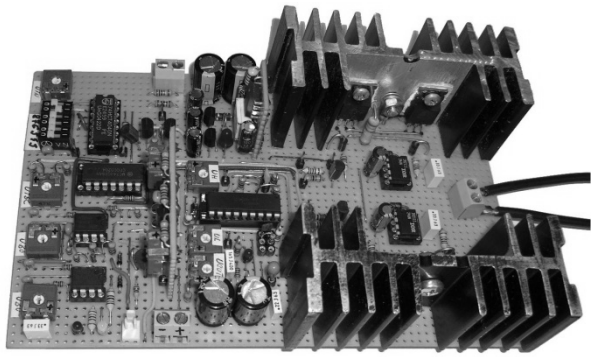


Fig. 10. Photo of DC/DC converter

V. CONCLUSION

New concept of control method for bi-direction buck-boost DC/DC converter is described in the paper. This concept of control decreases power transistor switching losses and thus increases efficiency of converter.

ACKNOWLEDGMENT

This work was supported by Slovak Research and Development Agency under project APVV-0095-07 and by Scientific Grant Agency of the Ministry of Education of Slovak Republic under the contract VEGA No. 1/0099/09.

REFERENCES

[1] Gaboriault M., Notman A.: A High Efficiency, Noninverting, Buck-Boost DC-DC Converter, APEC '04. Nineteenth Annual IEEE Volume 3, 2004

[2] Markel T., Zolot M., Sprik S.: Ultracapacitors and Batteries in Hybrid Vehicles, National Renewable Energy Laboratory (NREL) PR-540-38484, 8.2005, pp.7-15.

[3] J. Hamar, I. Nagy, P. Stumpf H. Ohsaki, E. Masada: New Dual Channel Quasi Resonant DC-DC Converter Topologies for Distributed Energy Utilization

[4] Dudrik J.: High Frequency Soft Switching DC-DC Power Converters. Monograph, ELFA, Košice 2007, ISBN 978-80-8086-055-4

[5] Tereň A., Feňo I. Špánik P.: DC/DC Converters with Soft (ZVS) Switching. Proc. of the Int. Conf. ELEKTRO 2001, section - Electrical Engineering. Žilina 2001, str.82 – 90.

[6] Texas Instruments: Datasheet UC3637 Switched Mode Controller for DC Motor Drive, www.ti.com

Introduction to Single Carrier Frequency Division Multiple Access (SC-FDMA)

[1]*Radovan Blicha,* [2]*Juraj Gazda,* [3]*Šterba Ján*

[1]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]radovan.blicha@tuke.sk, [2]juraj.gazda@tuke.sk, [3]jan.sterba@tuke.sk

***Abstract*— In this paper we present a new scheme in 3GPP Long Term Evolution (LTE) of transmitters systems for the up-link wireless access communication. Single carrier frequency division multiple Access (SC-FDMA) is a novel method of radio transmission under consideration for deployment in future cellular systems. The development of SC-FDMA represents one step in the rapid evolution of cellular technology.**

SC-FDMA uses single carrier modulation at the transmitter and frequency domain equalization at the receiver. This technique has similar performance and the same overall structure as Orthogonal Frequency Division Multiple Access (OFDMA) system. One advantage over OFDMA is that the SC-FDMA signal has lower peak-to-average power ratio (PAPR).

In this paper, we give an introduction to SC-FDMA focused on the physical layer of OSI model.

***Keywords*—Single carrier, wireless, 3GPP, LTE**

I. INTRODUCTION

3rd generation Partnership Project (3GPP) prescribes OFDMA for downlink transmission and SC-FDMA for uplink transmission in the LTE of cellular systems in order to make the mobile terminal power-efficient. One disadvantage of OFDMA is the high PAPR, which raises the cost and reduces the power efficiency of a transmitter's power amplifier. The power amplifiers at mobile terminals with a lower PAPR using SC-FDMA can be simpler and more power-efficient.

OFDMA and SC-FDMA are modified versions of the Orthogonal Frequency Division Multiplexing (OFDM) and Single Carrier with Frequency Domain Equalization (SC/FDE) modulation schemes. All multicarrier techniques employ a discrete set of orthogonal subcarriers distributed across the system bandwidth. To transmit several signals simultaneously, the multiple access techniques assign the signals to exclusive sets of subcarriers.

By using a multicarrier technique which subdivides the entire channel into smaller sub-bands (subcarriers), it is possible to mitigate the frequency-selective fading seen in a wide band channel. OFDM is a multicarrier modulation technique that multiplexes the data on multiple carriers and transmits them in parallel. OFDM uses orthogonal subcarriers, which overlap in the frequency domain. In the frequency domain, since the bandwidth of a subcarrier is designed to be smaller than the coherence bandwidth, each sub-channel is seen as a flat fading channel which simplifies the channel equalization process. In the time domain, by splitting a high-rat€e data stream into a number of lower-rate data stream that are transmitted in parallel, OFDM resolves the problem of ISI in wide band communications [1].

OFDM has its major disadvantages: High peak-to-average power ratio (PAPR), high sensitivity to frequency offset, and a need for an adaptive or coded scheme to overcome spectral nulls in the channel [2, 3].

Since SC-FDMA system uses single carrier modulation, it is characterized by significantly reduced PAPR that allows prolonging the battery life and enhancing the power efficiency of the transmission. This was the fundamental motivation behind implementing SC-FDMA in the uplink of LTE. However it encounters substantial linear distortion manifested as inter-symbol interference (ISI. This can be eliminated by using Frequency Domain Equalizer (FDE) that does not increase the computational complexity at the receiver side.

In this paper, we have evaluated the overview of a single carrier FDMA (SC-FDMA) system, which is multiple access scheme implemented in the uplink of 3GPP Long Term Evolution (LTE). The remainder of this paper gives overviews of SC-FDMA in detail, the SC-FDMA implementation in 3GPP LTE uplink and characterizes the PAPR properties of SC-FDMA signals.

II. SINGLE CARRIER WITH FREQUENCY DOMAIN EQUALIZATION

SC/FDE is a practical technique for mitigating the effects of frequency selective fading. It delivers performance similar to OFDM with essentially the same overall complexity, even for a long channel impulse response [2], [3]. Figure 1 shows the block diagrams of an SC/FDE receiver and, for comparison, an OFDM receiver. These systems have many blocks in common. We can see that both systems use the same communication component blocks and the only difference between the two diagrams is the location of the IDFT block. The both systems have similar performance and spectral efficiency.

In summary, SC/FDE has advantages over OFDM as follows: low PAPR due to single carrier modulation at the

transmitter, robustness to spectral null, lower sensitivity to carrier frequency offset, lower complexity at the transmitter that benefits the mobile terminal in cellular uplink communications.

Single carrier modulation with frequency domain equalization essentially has the same performance and structures as OFDM. Single carrier FDMA is an extension of SC/FDE and provides the multi-user access.

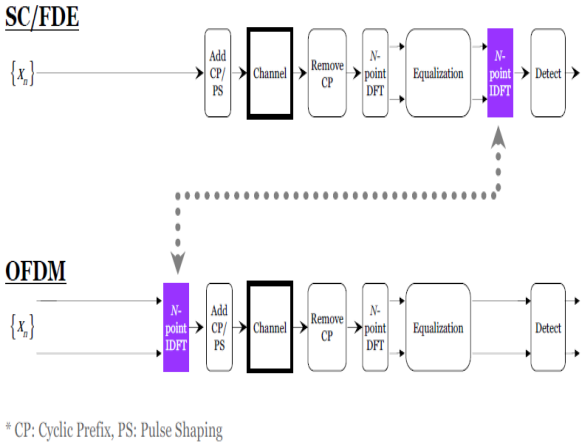


Fig. 1. SC/FDE and OFDM

III. Orthogonal Frequency Division Multiplexing

In 4G wireless communication systems, bandwidth is a precious commodity, and service providers are continuously met with the challenge of accommodating more users with in a limited allocated bandwidth. OFDM is multicarrier modulation technique which provides an efficient means to handle high-speed data streams on a multipath fading environment that causes ISI. To eliminate ISI a cyclic prefix is added. But this phenomenon eventually decreases the bandwidth (BW) efficiency greatly.

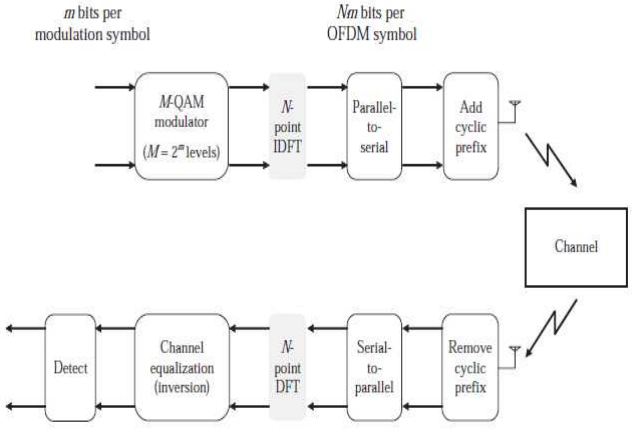


Fig. 2. OFDM signal processing

Figure 2 shows the essential elements of an OFDM transmitter and receiver using digital signal processing technology. The binary input to the OFDM modulator is the output of a channel coder that introduces an error correcting code and cyclic redundancy check to the information signal to be transmitted. The digital baseband modulator, typically performing quadrature amplitude modulation (QAM), transforms the binary input signal into a sequence of complex-valued multilevel modulation symbols. A signal processor then performs an IDFT on a sequence of N modulation symbols to produce one OFDM symbol, consisting of one transformed modulation symbol in each of N frequency sub-bands. The N sub-band samples obtained from the IDFT are transmitted sequentially over a fading channel and the receiver performs a DFT to recover the N time-domain modulation symbols from the received frequency domain representation. The channel inversion operation compensates for the linear distortion introduced by multipath propagation. Finally a detector produces a binary signal corresponding to the original input to the OFDM transmitter [4], [5].

IV. Single Carrier FDMA System

Figure 3 shows a schematic diagram of the conventional SC-FDMA system. At the transmitter, each block of the time domain data symbols is transformed to frequency domain by the application of the discrete Fourier transform (DFT), and then mapped to a subset of the total available subcarriers, which enables OFDMA modulation. As in OFDMA, the transmissions from multiple users remain orthogonal due to the fact that each user is allowed to use a distinct set of the available subcarriers. However, the advantage of this approach as compared to OFDMA is that the overall transmitted signal is a single-carrier signal, which has a lower PAPR as compared to its multicarrier counterpart. After the DFT and IDFT operation, a cyclic prefix CP is added at the end of the resulting signal, and the signal is filtered via a pulse-shaping filter before transmission. Pulse shaping is the process of changing the waveform of the transmitted pulses to make them suit better to the communication channel by limiting the effective bandwidth of the transmission. By filtering the transmitted pulses in this way, the inter-symbol interference caused by the channel can be kept under control [6].

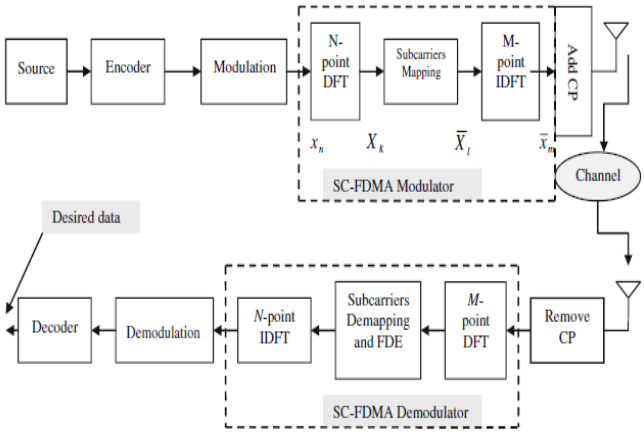


Fig. 3. Transmitter and receiver structures of the f SC-FDMA

At the receiver, the CP is removed from the received signal, and the signal is transformed into the frequency domain via an M-point DFT. Then, FDE and subcarriers demapping are performed. The resulting signal is transformed into time domain via an N-point IDFT. Finally, the demodulation and decoding processes are performed.

Because the SC-FDMA transmitter expands the signal

bandwidth to cover the bandwidth of the channel and time domain data symbols are transformed to frequency domain by DFT before going through OFDMA modulation, SC-FDMA is sometimes denote to as DFT-spread OFDMA. One advantage over OFDMA is that the SC-FDMA signal has lower PAPR because of its inherent single carrier structure.

To prevent inter-block interference (IBI) it inserts cyclic prefix (CP). The cyclic prefix is a repeat of the end of the symbol at the beginning. The fundamental motivation for that is to reduce the effects of the multipath propagation prior to the detection procedure. The length of the cyclic prefix is often equal to the guard interval. The transmitter also performs linear filtering operation pulse shaping in order to reduce out-of-band signal energy [7].

A. Subcarrier Mapping

In SC-FDMA, there are two types of modulation schemes: distributed subcarrier mapping and localized subcarrier mapping and each modulation has different assigned frequency spectrum pattern. Special case of SC-FMDA is a Interleaved FDMA (IFDMA). It is very efficient in that way that the transmitter can modulate the signal strictly in the time domain without the use of DFT and IDFT.

In a distributed subcarrier mapping scheme, a user's data symbols occupy a set of subcarriers distributed over the entire frequency range of the channel and we achieve frequency diversity. In a localized subcarrier mapping scheme, a user's data symbols occupy a set of consecutive subcarriers and we can achieve multi-user diversity by means of channel-dependent scheduling. The two subcarrier mappings also affect the structure of the time domain signal and the peak power characteristics. The two approaches to subcarrier mapping give the network operator flexibility to adapt to the specific requirements of each operating environment [8], [9].

V. Conclusion

Single carrier FDMA (SC-FDMA) which utilizes single carrier modulation at the transmitter and frequency domain equalization at the receiver is a technique that has similar performance and essentially the same overall structure as those of an OFDMA system. SC-FDMA has been adopted as the uplink multiple access scheme in 3GPP Long Term Evolution (LTE) mainly due to its low peak-to-average power ratio (PAPR) which greatly improves the transmit power efficiency. In this paper, we have given the overview of digital modulation techniques and different subcarrier mapping.

The 3GPP-LTE proposed SC-FDMA scheme and a conventional OFDMA scheme are expected to be similar in terms of link level and system level performance. Both schemes provide similar degrees of freedom in system deployment and up-link scheduling without performance degradation.

VI. ACKNOWLEDGEMENT

This work is the result of the project VEGA 1/0045/10 Nové metódy spracovania signálov pre rekonfigurovateľné bezdrôtové senzorové siete. This work has been also funded by Grant Agency SPP Hlavička.

References

[1] R. van Nee, R. Prasad, *OFDM for Wireless Multimedia Communications*, Artech House, 2000.

[2] H. Sari, G. Karam, and I. Jeanclaude, "*Transmission Techniques for Digital Terrestrial TV Broadcasting*," IEEE Commun. Mag., vol. 33, no. 2, pp. 100-109, Feb. 1995.

[3] D. Falconer, S. L. Ariyarisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency Domain Equalization for Single-Carrier Broadband Wireless Systems," IEEE Commun. Mag., vol. 40, no. 4, pp. 58-66, Apr. 2002.

[4] M. Myung, D. Goodman, *Single Carrier FDMA, A New Air Interference for Long Term Evolution*, Wiley, 2008.

[5] H. G. Myung, J. Lim, and D. J. Goodman, "*Single Carrier FDMA for Uplink Wireless Transmission*," IEEE Vehicular Technology Mag., vol. 1, no. 3, pp. 30 – 38, Sep. 2006.

[6] E. Fathi, Abd El-Samie, S. Faisal, Al-kamali, I. Moawad Dessouky, M. Bassiuny Sallam, Farid Shawki, *Performance enhancement of SC-FDMA systems using a companding technique*, Institut TELECOM and Springer-Verlag 2010.

[7] H. Myung, L. Junsung, D.J. Goodman, *Single carrier FDMA for uplink wireless transmission*, IEEE Vehicular Technology Magazine, vol. 1, 2006

[8] Hyung Myung, *INTRODUCTION TO SINGLE CARRIER FDMA*, EURASIP, EURASIPCO, Poznan 2007.

[9] S. Suzuki, O. Takyu, Y. Umeda, *Performance Evaluation of Effect of Nonlinear Distortion in SC-FDMA System*, International Symposium on Information Theory and its Applications, ISITA 2008.

Soft Switching DC-DC Converter with Controlled Output Rectifier

[1]*Marcel Bodor,* [1]*Ľubomír Matis*

[1] Department of Electrical, Mechatronic and Industrial Engineering, FEI TU of Košice, Slovak Republic

[1]marcel.bodor@tuke.sk

***Abstract*—A full-bridge soft switching PWM DC/DC converter with controlled output rectifier is described in this paper. Soft switching is achieved by using controlled output rectifier and lossless turn-off snubber. No circulating current is occurred in the proposed converter with used new control method. The principle of operation is explained and analyzed on prototyp with coaxial transformer.**

***Keywords*—ZCZVS Converter, Pulse Width Modulation (PWM), DC power supply, High frequency power converter, Snubber, Soft switching, DC-DC converter.**

I. Introduction

The conventional phase shifted PWM converters are often used in many applications because their topology permits all switching devices to operate with soft switching by using circuit parasitics such as power transformer leakage inductance and devices junction capacitance. In very used phase-shifted PWM control converters, circulating current flows through the power transformer and switching devices during freewheeling intervals. This circulating current can be eliminated by disconnection of the secondary winding, which can be realized by reverse bias application for the output diode rectifier [1] – [14] or using controlled rectifier [5], [15] – [20].

II. Power Circuits of the Proposed Converter

To improve the properties of the existing converters, the new topology of the energy recovery turn-off snubber was proposed in the following DC/DC converter.

Scheme of the proposed DC/DC converter shown in Fig. 1 consists of full bridge inverter, centre tapped power transformer, controlled output rectifier, output filter and novel type of secondary turn-off snubber.

The converter is controlled by pulse-width modulation of secondary transistors. Soft switching all of the transistors in the converter is reached.

The new snubber circuit eliminates the turn off losses of the secondary transistors. The semiconductor switches T_5, T_6 in the secondary side are used to reset secondary and simultaneously also primary circulating current. The energy of the leakage inductance of the power transformer stored in snubber at secondary switch turn off is transferred to the load.

III. Operation Principle

The switching diagram and operation waveforms are shown in Fig. 2 and operation analysis of the converter is shown in Fig. 3. The DC/DC converter is controlled by pulse width modulation of secondary switches.

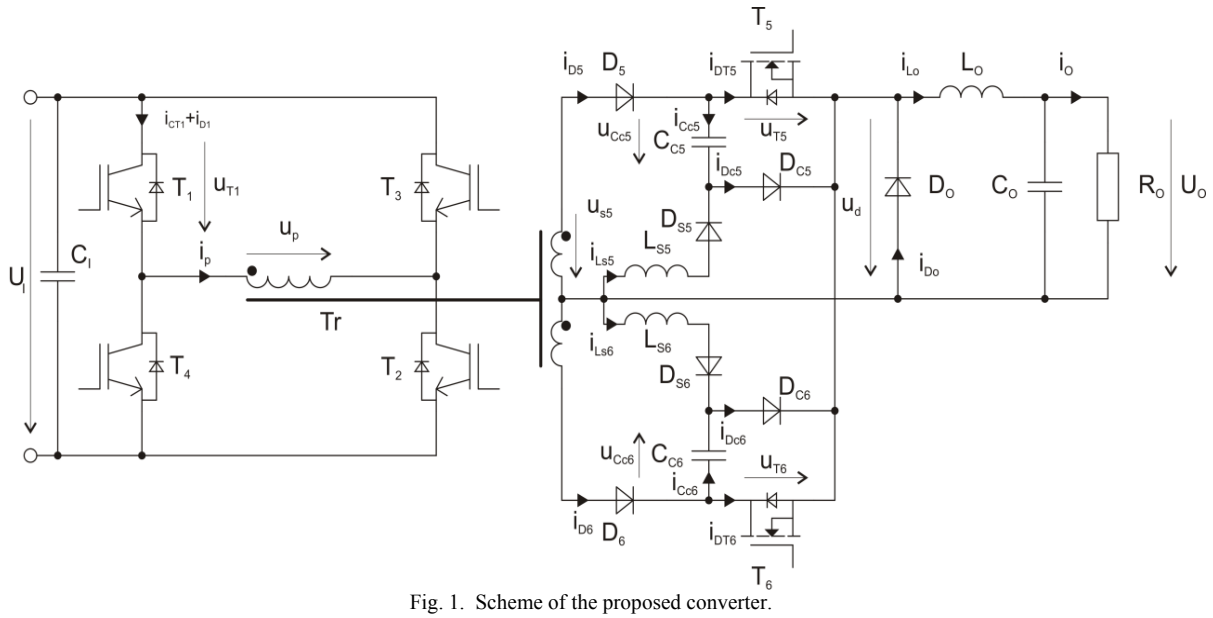


Fig. 1. Scheme of the proposed converter.

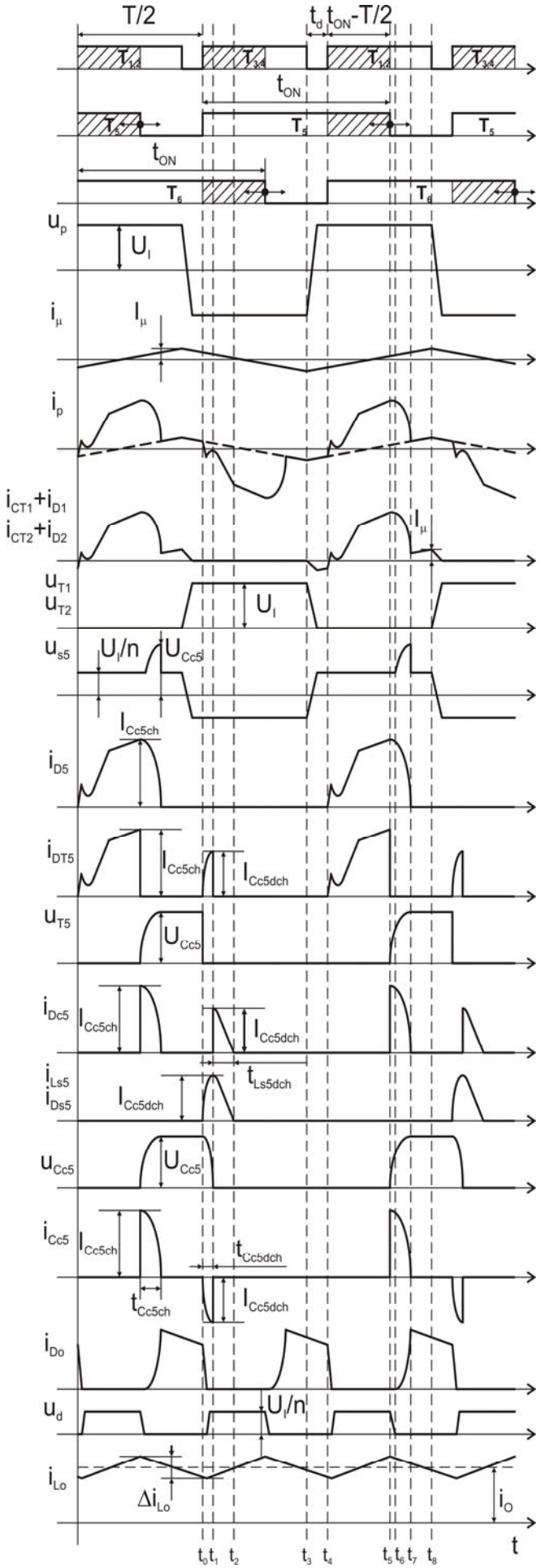


Fig. 2. Operation principle waveforms.

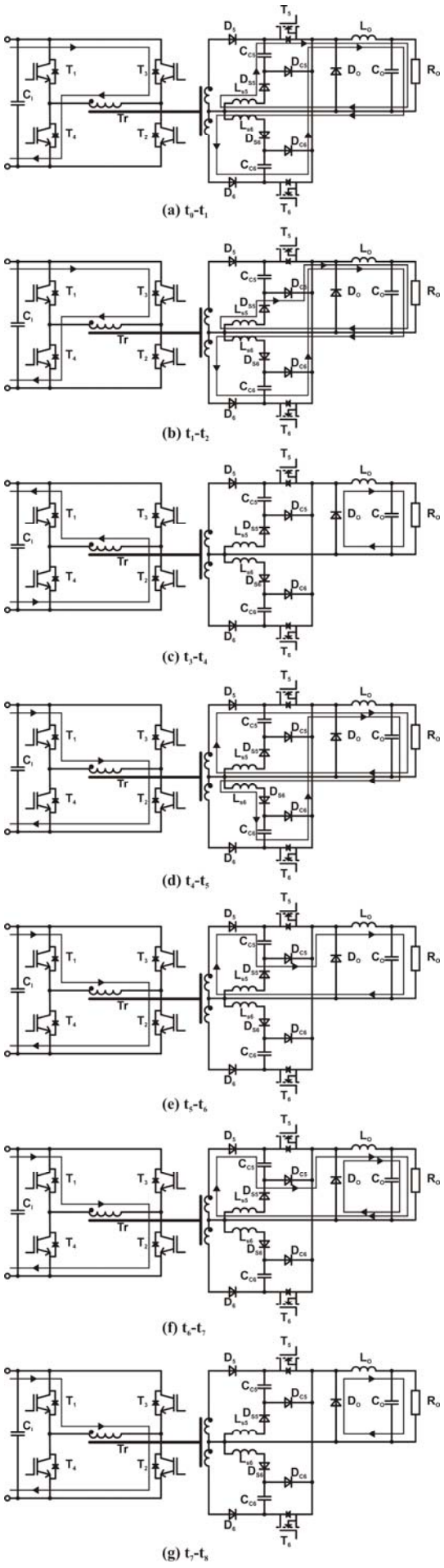


Fig. 3. Operation analysis in the intervals.

A. *Interval t_0-t_1*

The secondary transistor T_5 is turned on at t_0 half period earlier then primary transistors T_1 and T_2. The capacitor C_{C5} starts discharging to the load. The rate of rise of discharging current of this capacitor C_{C5} is limited by the snubber circuit inductance L_{S5}, and thus zero current turn on for the MOSFET transistor T_5 is achieved. In the same time transistors T_3, T_4 are turned on. The current of the primary transistors T_3, T_4 and the current of the secondary transistor T_6 are reduced by the discharging current of capacitor C_{C5}.

B. *Interval t_1-t_2*

The energy stored in snubber inductance L_{S5} is now flowing through D_{C5} to load.

C. *Interval t_3-t_4*

This interval starts with the turn off of the primary transistors T_3 and T_4. The magnetizing current of the transformer T_r discharges the output capacitances C_{OSS} of the transistors T_1, T_2 and charges the output capacitances of the transistors T_3, T_4. The rate of rise the current is limited by the leakage inductance of transformer. Soft turn on for transistors T_1, T_2 is achieved.

D. *Interval t_4-t_5*

The turn on of the transistors T_1, T_2 and T_6 commutations from freewheeling diode D_0 to T_5 transistor occur at t_4. The current of transistor T_5 is reduced by the discharge current of the capacitor C_{C6} and later by the current of the inductance L_{S6}.

E. *Interval t_5-t_6*

At the time t_5 transistor T_5 turns off. Its current commutates on capacitor C_{C5} and diode D_{C5}. Zero voltage turn off of this transistor is ensured because the rate of rise the voltage is limited by the capacitor. The energy of the leakage inductance of the power transformer is absorbed by the snubber capacitance and then by the load.

F. *Interval t_6-t_7*

At t_6 the rectified voltage u_d reached zero and afterwards the waveform of the charging process of the C_{C5} capacitance are changed. In this interval the energy of the leakage inductance is absorbed only by the capacitor C_{C5}. At t_7 the current of the rectifier diode D_5 falls to zero, and the primary current that flows through the transistor T_1 and T_2 drops on value of the magnetizing current because the whole energy was absorbed by the capacitor.

G. *Interval t_7-t_8*

Only magnetizing current flows through the primary winding of the power transformer in this interval. This small magnetizing current is turned off by primary switches and thus zero current turn off is achieved. The current of the smoothing inductance L_O is flowing through the freewheeling diode.

IV. Experimental Results

Laboratory model with components shown in Table I was built to verify the operation principle of the converter. The converter was supplied from DC source with a value of 300V. The rated output power was 1.2kW at switching frequency of 50 kHz. The typical converter waveforms were obtained at output voltage of 40V and output current of 25A.

TABLE I.
Used Components in Converter

T_1-T_4	IRG4PSC71UD
T_5,T_6	IRFP90N20D
D_5,D_6	UFB200FA40
D_{C5}, D_{S5}, D_{C6}, D_{S6}	CSD20060D (SiC diode)
L_{S5}, L_{S6}	24μH
C_{C5}, C_{C6}	33nF
D_O	249NQ135
L_O	44μH
C_O	470μF
Tr	coaxial transformer (turn ratio 6)

Primary transistor collector-emitter voltage and collector current with gate signals of primary and secondary transistor are shown in Fig. 4. After turn off of the secondary MOSFET gate signal the transformer primary current sink to the value of magnetizing current. This small magnetizing current is later turned off by primary IGBT transistors and thus only negligible turn-off losses occur. At the turn on moment of primary IGBT transistors the primary current flows through their freewheeling diodes and rise of current is limited by the leakage inductance of transformer. This ensures zero voltage zero current turn on. Switching trajectory of primary transistor is shown in Fig. 5.

Fig. 4. Primary transistor voltage and current at turn on and turn off.

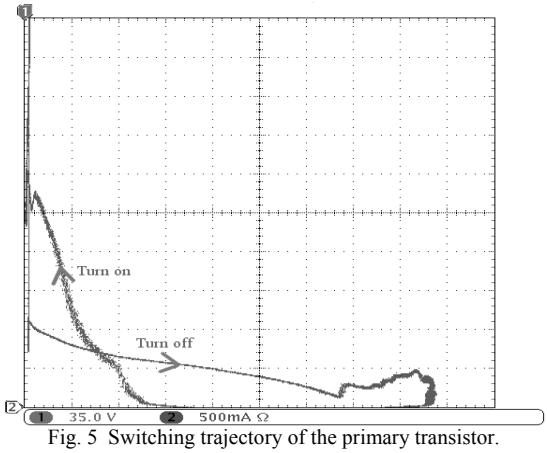


Fig. 5 Switching trajectory of the primary transistor.

Secondary transistor drain-source voltage and drain current at turn on and turn off are shown in Fig. 6. At turn off of this MOSFET its transistor the collector current commutates on the snubber capacitance C_{C5} and thus the transistor voltage rate of rise is reduced. At the turn on of the transistor T_5 the capacitance C_{C5} is discharged through the transistor to the load. The rate of rise of the discharging current is reduced by inductance L_{S5}. Switching trajectory of secondary transistors is shown in Fig. 7. Charging and discharging of snubber capacitors is shown in Fig. 8.

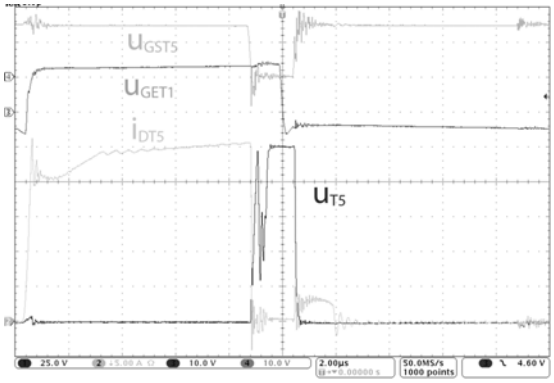


Fig. 6. Secondary transistor voltage and current at turn on and turn off.

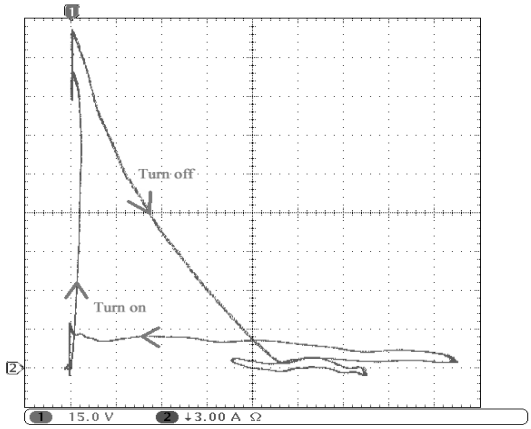


Fig. 7. Switching trajectory of the secondary transistor.

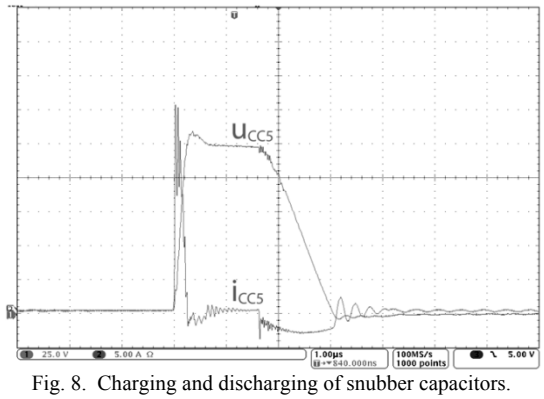


Fig. 8. Charging and discharging of snubber capacitors.

Efficiencies of the converter at various output voltages are shown in Fig. 9.

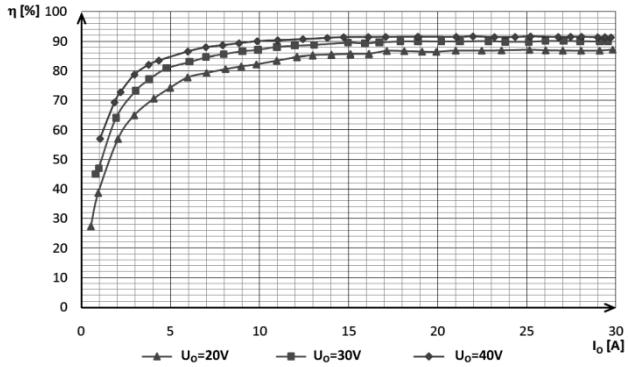


Fig. 9. Efficiencies of the converter.

ACKNOWLEDGMENT

This work was supported by Slovak Research and Development Agency under project APVV-0095-07 and by Scientific Grant Agency of the Ministry of Education of Slovak Republic under the contract VEGA No. 1/0099/09.

REFERENCES

[1] J. Dudrik, J. Oetter, "Soft-Switching PWM DC-DC Converter for High Power Applications", in Record EPE-PEMC 2006, Portorož, Slovenia, ISBN 1-4244-0121-6, CD, pp. 739-744.

[2] J. Dudrik, P. Špánik, D. N. Trip, " Zero Voltage and Zero Current Switching Full-Bridge DC-DC Converter with Auxiliary Transformer", IEEE Trans. on Power Electronics, Vol. 21, No.5, 2006, pp. 1328 – 1335.

[3] J. Dudrik, High Frequency Soft Switching DC-DC Power Converters, Monograph, Elfa, Košice, Slovakia, 2007 (in Slovak).

[4] J. G. Cho, J. W. Baek, Ch. Y. Jeong, and G. H. Rim, "Novel Zero-Voltage and Zero-Current Switching Full Bridge PWM Converter Using a Simple Auxiliary Circuit," IEEE Trans. on Industry Applications, Vol. 35, pp. 15-20, 1999.

[5] S. Moisseev, S. Sato, S. Hamada, M. Nakaoka, "Full Bridge Soft-Switching Phase-Shift PWM DC-DC Converter Using Tapped Inductor Filter", in Record, PESC 2003, pp. 1826-1831.

[6] R. Bojoi, G. Griva, G. Kovacevic, A. Tenconi, "ZVS-ZCS full-bridge DC-DC converter for voltage step-up in fuel cell distributed generation systems", in Record, European Conference on Power Electronics and Applications, 2-5 Sept. 2007, pp. 1 – 8.

[7] V. Ruščin, M. Olejár, M. Lacko, and J. Dudrik, "ZVZCS DC-DC converter with controlled output rectifier", in Record, TRANSCOM 2007: 7-th European conference of young research and science workers: Proceedings: Žilina June 25-27, 2007. Žilina: University of Žilina, 2007. s. 171-174. ISBN 978-80-8070-694-4

[8] J. Dudrik, and J. Šepeľa, "Soft-switching current-mode controlled DC-DC converter with secondary switches", in Record, EDPE 2005: 13th international conference on Electrical Drives and Power Electronics, September 26-28, 2005, Dubrovnik, Croatia. Zagreb: KoREMA, 2005. 4 p.

[9] J. Dudrik, V. Ruščin, and M. Bodor, Nondissipative auxiliary circuit for decreasing of turn-off losses in DC-DC converter with output controlled rectifier, Slovak patent pending No. PP 00033-2008.

[10] Kazurou Harada, Yoshiyuki Ishara and Toshiyuki Todaka, "Analysis and design of ZVS-PWM half-bridge converter with secondary switches", in Record, IEEE power Electronics Specialists Conference, PESC 95/vol.1 18-22 June 1995, pp. 280 – 285.

[11] A. Tereň, I. Feňo, and P. Špánik, "DC/DC Converters with Soft (ZVS) Switching". In Record, ELEKTRO 2001, section - Electrical Engineering. Žilina 2001, Slovakia, pp. 82 – 90.

[12] J. Dudrik, "Soft switching full-bridge PWM DC/DC converter using secondary snubber". In: Journal of Electrical and Electronics Engineering. vol. 2, no. 1 (2009), p. 147-150. ISSN 1844-6035.

[13] J. Dudrik, V. Ruščin, M. Bodor, "Soft switching DC/DC converter using controlled output rectifier with secondary turn-off snubber". In: EDPE 2009: Abstracts & CD Proceedings : 15th International Conference on Electrical Drives and Power Electronics : 4th Joint Croatia-Slovakia Conference: October 12-14, 2009, Dubrovnik, Croatia. Zagreb: KoREMA, 2009. p. 1-5. ISBN 978-953-6037-56-8.

Speech recognition using the classifiers based upon hidden Markov´s models

[1]*Radoslav Bučko,* [2]*Ján Molnár*

[1]Dept. of Theoretical Electrotechnics and Electrical Measurement, FEI TU of Košice, Slovak Republic
[2]Dept. of Theoretical Electrotechnics and Electrical Measurement, FEI TU of Košice, Slovak Republic

[1]radoslav.bucko@tuke.sk, [2]jan.molnar@tuke.sk

Abstract— **This paper describes recognition of spoken speech and problem with recognition, basic part of recognizer and especially classifiers based on statistic methods – called Hidden Markov models. Hidden markov models are described on example**.

***Keywords*—speech recognition, classifiers, hidden markov model.**

I. Introduction

Communication using verbal speech is the most basic, most natural and most important form of information transfer between people. If computer or other device has to be commanded by voice, acoustic signal from speaker, voice recognition and understanding the information has to be technically and algorithmically solved. This problem is being solved from last century and still is not finished. The main reasons are:

a) Speaker´s voice can be different in various conditions. Stress, illness, loudness of voice and even ageing change the voice signal. Coarticulation changes the fonetical properties of the beginning and end of the word, depending upon the context with other words.

b) Different speakers have different voices. Every speaker has different voice colour, accent, speed of speech and more...Voice recognizing is divided into two groups: speaker dependent and speaker independent.

c) Changing environment causes trouble for speech recognition. With increased levels of interference the identification of beginning nad end of the word is more difficult.

d) Recorded voice can be degraded by quality of microphone or by distance from it. [1]

Voice recognition software consist of 2 main parts. First part is the signal processing, which results in sequence of observations (mostly vectors). Second part are the classifiers, which assign the most suitable word from the dictionary to each sequence. Based upon words in the dictionary there are 2 main groups – recognition with small amount of words and with big amount of words.

From the point of applied methods we can divide classifiers into:

a) classifiers, which process the word as a whole and it is assigned to class, which is nearest to the example image – this distance is usually defined by applied methods of dynamic programming.

b) classifiers which use the classification based upon statistical methods – words are modeled using the so called hidden Markov´s models. [2]

II. Hidden Markov´s Models

Markov´s process G with hidden Markov´s model can be expressed using the pentad:

$$G = (S, O, A, B, \pi)$$

(1).

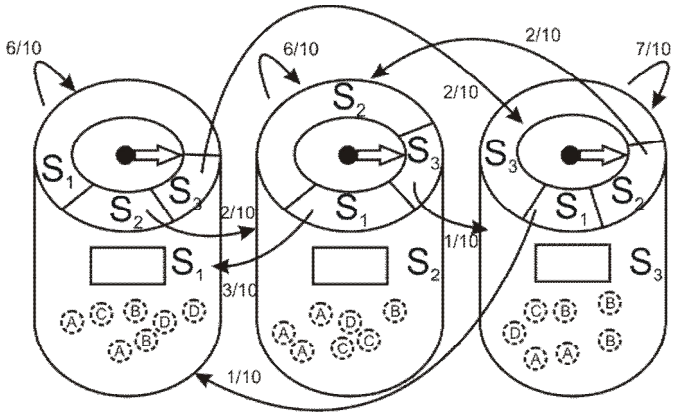


Fig. 1. Markov´s model with 3 states and given probabilities of crossings

Where:

- $S = \{ s_1, s_2, \ldots, s_N \}$ is the file of individual states of Markov´s model,
- $O = \{ o_1, o_2, \ldots, o_L \}$ is the alphabet of L output symbols of vector quantiser,
- $A = [a_{ij}]$ is the matrix of crossings, its elements define the probability of system crossing from state a_i in time t to state a_j in time $t+1$,
  We can say that:
  $$a_{ij} = P\left(s(t+1) = a_j \mid a(t) = a_i\right), \qquad 1 \le i, j \le N,$$
- $B = [b_{jl}] = [b_j(l)]$ is the matrix of generated examples probability. It determines the probability of

generating l entry of the final file of spectral examples when the system is in state a_i,

We can say that:
$b_{jl} = b_j(l) = P(o(t) = o_l \mid s(t) = s_j)$,
$1 \le l \le L, 1 \le j \le N$,

- $\pi = [\pi_i]$ is the column vector of starting state probability,
  We can say that:
  $\pi_i = P(s(1) = s_i), \qquad 1 \le i \le N$.

For parameters π_i, a_{ij} a $b_j(l)$ we can apply these conditions:

$$\sum_{i=1}^{N} \pi_i = 1,$$

$$\sum_{j=1}^{N} a_{ij} = 1 \quad \text{pre i = 1,...,N ,}$$

$$\sum_{l=1}^{L} b_j(l) = 1 \quad \text{pre j = 1,...,N .[3]}$$

We will describe λ as $\lambda = (A, B, \pi)$.

We have 3 enclosed opaque containers with the opening for the arm. These containers will be representing individual states of Markov´s model (set S) s_1, s_2 a s_3. In every container there are 7 balls labelled as *A*, *B*, *C* a *D*. These balls represents the alphabet of four output symbols of vector quantizer and $o_1=A$, $o_2=B$, $o_3=C$, $o_4=D$.[3]

There is a rotary arrow on every container, which points on 3 different parts of circle labelled as s_1, s_2 a s_3. Arrow will represent randomness of container choice, form which the ball will be chosen. Every container has its own arrow because of different probabilities..

Probability of crossing from state s to state s_1 (picking the ball from container s_1) is $a_{11}=0,6$ (60%).
Similarly:
$a_{12}=0,2$, $a_{13}=0,2$, $a_{11}+a_{12}+a_{13}=1$,
$a_{22}=0,6$,$a_{21}=0,3$, $a_{23}=0,1$, $a_{22}+a_{21}+a_{23}=1$,
$a_{33}=0,7$,$a_{31}=0,1$, $a_{32}=0,2$, $a_{33}+a_{31}+a_{32}=1$,

$$a_{ij} = \begin{pmatrix} 0,6 & 0,2 & 0,2 \\ 0,3 & 0,6 & 0,1 \\ 0,1 & 0,2 & 0,7 \end{pmatrix}$$

There is a 100% probability that we we can get somewhere from every state – either by picking the ball or by crossing to another container.

In container s_1 there are 2 balls labeled as *A*, so the probability of picking the ball *A* will be : $b_1(A)=2/7$.
Similarly:

$b_1(B)=2/7$, $b_1(C)=1/7$, $b_1(D)=2/7$, $b_1(A)+ b_1(B)+ b_1(C)+ b_1(D)=1$,
$b_2(A)=3/7$, $b_2(B)=1/7$, $b_2(C)=2/7$, $b_2(D)=1/7$, $b_2(A)+ b_2(B)+ b_2(C)+ b_2(D)=1$,
$b_3(A)=2/7$, $b_3(B)=3/7$, $b_3(C)=1/7$, $b_3(D)=1/7$, $b_3(A)+ b_3(B)+ b_3(C)+ b_3(D)=1$.

$$b_{ij} = \begin{pmatrix} 2/7 & 2/7 & 1/7 & 2/7 \\ 3/7 & 1/7 & 2/7 & 1/7 \\ 2/7 & 3/7 & 1/7 & 1/7 \end{pmatrix}$$

Everytime, only 1 ball is picked up and then it is returned back.

Starting container is chosen and one ball is picked. That ball is then returned back. Sequence of states and pickings is being recorded.

- state: s_1,
- choice: *C*.

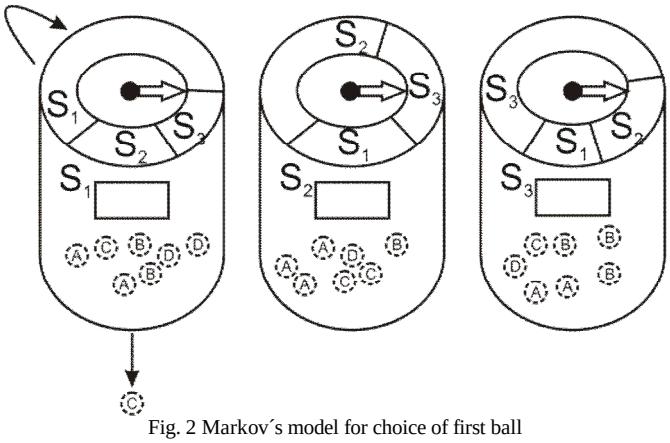


Fig. 2 Markov´s model for choice of first ball

Arrow is spinned and the result will determine if the ball will be picked up, or if we cross to next container. Probability of picking the ball is 3 times higher than probability of crossing to next container so the sequence will be:

- s_1, s_1, s_1, s_1,
- *C, A, C, D.*

Next picking of ball and after spinning crossing to s_2 (fig.3).
Sequence:

- $s_1, s_1, s_1, s_1, s_1, s_1$,
- *C, A, C, D, D, C.*

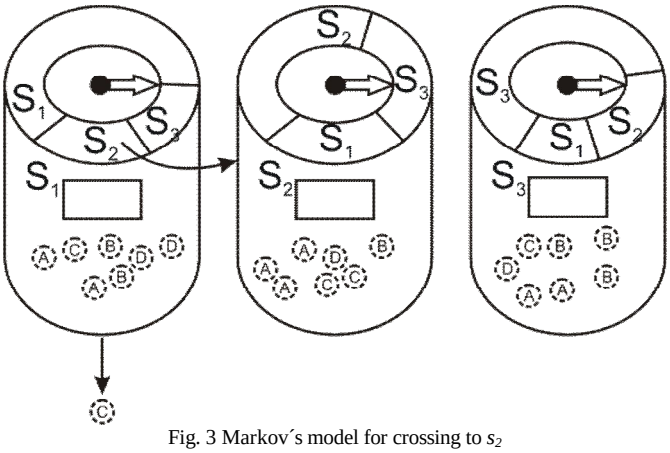


Fig. 3 Markov´s model for crossing to s_2

Choice from z s_2 (fig.4):

- $s_1, s_1, s_1, s_1, s_1, s_1, s_2,$
- $C, A, C, D, D, C, A.$

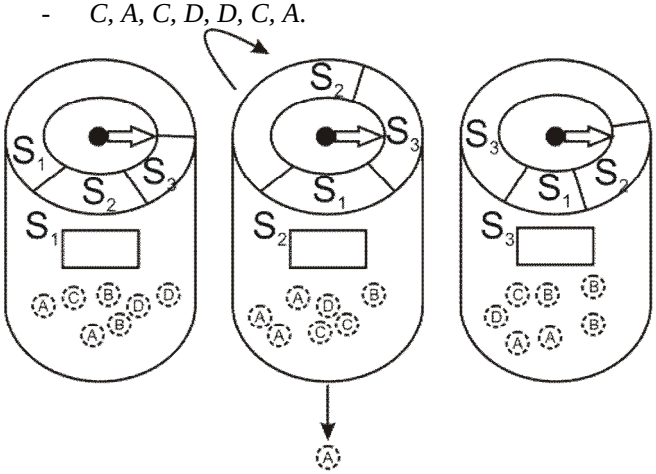


Fig. 4 Markov´s model picking from s_2

Repeating this process we will get to final form of sequence of states and choices:

- $s_1, s_1, s_1, s_1, s_1, s_1, s_2, s_2, s_2, s_2, s_2, s_3, s_3, s_3, s_3, s_3, s_3, s_1,$
- $C, A, C, D, D, C, A, C, C, C, A, A, A, A, A, C, B, C.$

However, if we use hidden Markov´s model we don´t see the sequence of states, only the sequence of pickings:
$C, A, C, D, D, C, A, C, C, C, A, A, A, A, A, C, B, C.$

For this sequence of picking we have to determine the container from which they were picked. We will use Markov´s model for hypothesis, which will most accurately describe probabilities of crossings and pickings. We will choose random hypothesis:

- $s_3, s_2, s_2, s_2, s_1, s_1, s_1, s_2, s_2, s_2, s_1, s_1, s_1, s_3, s_3, s_2, s_2, s_3.$

picking	C	A	C	D	D	C	A	...	C
state	s_1	s_2	s_2	s_2	s_3	s_3	s_3		s_1
P(picking)	3/7	3/7	1/7	2/7	1/7	1/7	2/7		3/7
p(crossing)	2/10	6/10	6/10	3/10	6/10	6/10	2/10		

$P(s_3, s_2, s_2, s_2, s_1, s_1, s_1, s_2, s_2, s_2, s_1, s_1, s_1, s_3, s_3, s_2, s_2, s_3) = 1.6768x10^{-19}$.
We will choose hypothesis which corresponds the reality:

- $s_1, s_1, s_1, s_1, s_1, s_1, s_2, s_2, s_2, s_2, s_2, s_3, s_3, s_3, s_3, s_3, s_3, s_1.$

Picking	C	A	C	D	D	C	A	...	C
State	s_1	s_1	s_1	s_1	s_1	s_1	s_2		s_1
p(picking)	3/7	2/7	3/7	1/7	1/7	3/7	3/7		3/7
p(crossing)	6/10	6/10	6/10	610	6/10	2/10	6/10		

$P_{skut.}(s_1, s_1, s_1, s_1, s_1, s_1, s_2, s_2, s_2, s_2, s_2, s_3, s_3, s_3, s_3, s_3, s_3, s_1) = 8.7350x10^{-16}$.

Probability of hypothesis which corresponds to reality is higher than probability of our chosen hypothesis. If we would like to determine the most probable hypothesis, we would have to count the probability of all hypothesis, which amount is very high. Because we don't want to count the probabilities of every hypothesis, we can use the Viterbi´s algorithm which can determine the most probable option directly. Let´s consider that we have partial sequences created by first n containers, which ends with containers s_1, s_2 a s_3. Let´s consider that we know their probabilities for specific n and for $n+1$ we will count (fig.5):

$ps_3' = ps_3 c.max(\, ps_3.p_{s_3 s_3},\, ps_2.p_{s_2 s_3},\, ps_1.p_{s_1 s_3}\,)$,
$ps_2' = ps_2 c.max(\, ps_3.p_{s_3 s_2},\, ps_2.p_{s_2 s_2},\, ps_1.p_{s_1 s_2}\,)$,
$ps_1' = ps_1 c.max(\, ps_3.p_{s_3 s_1},\, ps_2.p_{s_2 s_1}, ps_1.p_{s_1 s_1}\,)$.

Inductively (fig.6):
$p = max(\, ps_3,\, ps_2,\, ps_1\,)$.

We have to count the sequence, so we will use the branch (fig.7) with highest probability underlined:

$ps_3' = ps_3 c.max(\, ps_3.p_{s_3 s_3},\, ps_2.p_{s_2 s_3},\, \underline{ps_1.p_{s_1 s_3}}\,)$,
$ps_2' = ps_2 c.max(\, ps_3.p_{s_3 s_2},\, \underline{ps_2.p_{s_2 s_2}},\, ps_1.p_{s_1 s_2}\,)$,
$ps_1' = ps_1 c.max(\, ps_3.p_{s_3 s_1},\, \underline{ps_2.p_{s_2 s_1}}, ps_1.p_{s_1 s_1}\,)$.

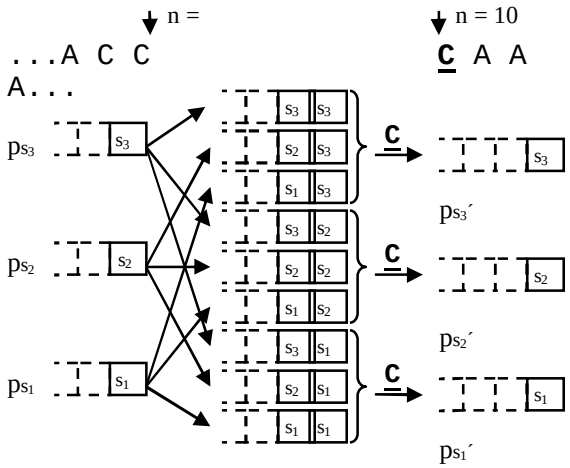


Fig. 5 Viterbi´s algorithm.

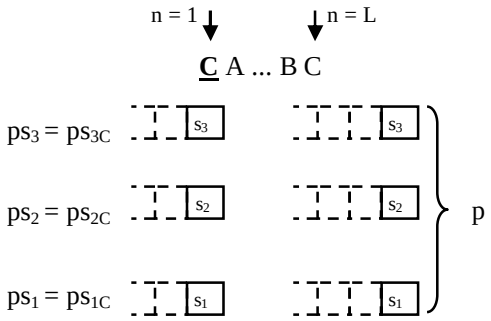


Fig. 6 Viterbi´s algorithm – next step

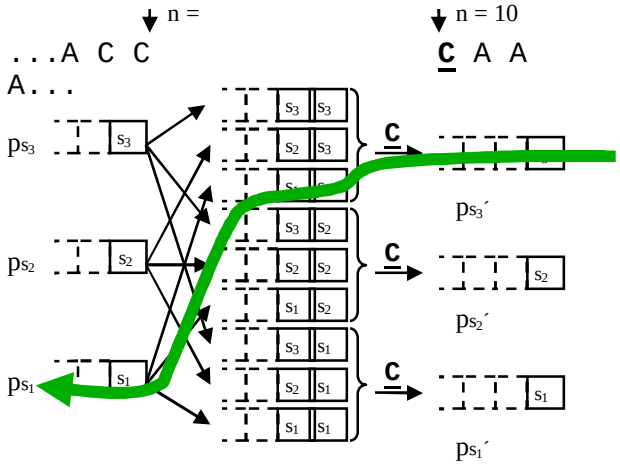


Fig. 6 Viterbi´s algorithm – branch with highest probability

This way we can determine the most probable variant of sequence from our example.

Sequences:

- determined: $s_1, s_1, s_1, s_1, s_2, s_2, s_2, s_2, s_1, s_3, s_3, s_3, s_3, s_3, s_3, s_1, s_1, s_1,$
- real: $s_1, s_1, s_1, s_1, s_1, s_1, s_2, s_2, s_2, s_2, s_2, s_3, s_3, s_3, s_3, s_3, s_1.$

Determined sequence doesn´t have to correspond reality (like in our example) but it is the most probable variant. When using speech recognition, mostly left-to-right Markov´s models are used. These are well suited for modeling processes, which progression is connected with time. [4]

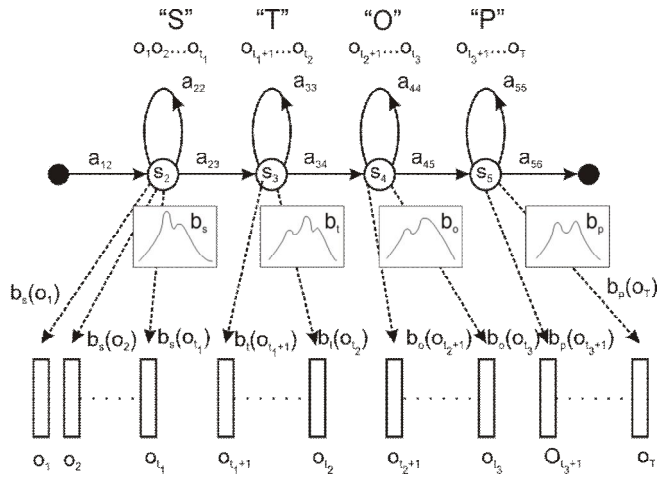


Fig.8 Model of word "STOP"

Figure (fig.8) shows the model of word "STOP" which is constructed from 6 states, from which only states s_2, s_3, s_4 and s_5 are emitting states (states which are able to generate the output vector of observations).

Each of these states corresponds to one articulatory position of speech apparatus during creation of individual phonemes of word "STOP". Every state has its transitional probabilities as well as function of separation of output probability.

For example, for phoneme *[s]* corresponds state s_2, transitional probabilities a_{22} and a_{23} and function of output probability separation $b_s()$. [2]

III. Conclusion

Classifiers based upon hidden Markov´s models are used for speech recognition of isolated words. Their main advantage lies in possibility of recognition of fluent speech when it is possible to model separately input, middle and output part of phoneme, which is very important due to coarticular effects.

Acknowledgment

The paper has been prepared by the support of Slovak grant projects VEGA No. 1/0660/08, KEGA 3/6386/08 and KEGA 3/6388/08.

References

[1] J. Psutka, " Comomunication with PC using spoken speech" ACADEMIA, Praha, 1995.
[2] J. Psutka, "Speaking Czech with computer" ACADEMIA, Praha, 2006
[3] L. R. Rabiner, *"A tutorial a hidden markov models and selected applications in speech recognition"*. Proceedings of the IEEE, 1989.
[4] A. Lúčny, "Skryté markovove modely", http://www.microstep-mis.com/~andy/hmm2.ppt.

Thermal degradation in insulation materials

[1]*Ľudovít CSÁNYI*, [1]*Martin MARCI*

[1]Dept. of Electronics Power Engineering, FEI TU of Košice, Slovak Republic

[1]ludovit.csanyi@tuke.sk, [1]martin.marci@tuke.sk

***Abstract*— Electric power system contains different electrical materials. The thermal stress causes a change in electro physical structures. The IRC analysis (non-destructive diagnostic method) measured this change. The method is used in the laboratory and the practice often. An artificial neural network processes results of these measurements. An artificial neural network based on Back Error propagation is used.**

***Keywords*—isolation, polarization, diagnostic, neural network**

I. INTRODUCTION

The insulation system is a primary component of all electrically systems. The failure of electrical machinery is caused with failure their insulation very often. Ageing of insulation influences on quality insulation very much.

The diagnostic methods have bee developed to determine status of insulation system. The status of system and changes of properties (in insulation underway during aging) are described using these methods.

The some electrical and physical values are monitored and measured. These values show actual status system. Each value shows change of insulation system. IRC analysis (isothermal relaxation current analysis) is one of the new methods. IRC analysis is based on measure of charge currents depending on time at constant temperature. This method is non-destructive.

The results were achieved with measured samples Relanex and with measured insulation machine in operation. These results were further each other evaluated results obtained were transferred to graphic form.

II. DIAGNOSTIC METHODS

The term diagnostics means determination and classification attributes. These attributes show on change parameters during use of equipment. Appropriate diagnostics methods we must choose on following requirement:

- the method must related to property, which is subject of interest,
- distribution of stress must respond to actual,
- non-destructive methods are used preferably,
- the method of measurement should not affect the degradation process,
- the method must be applied in the operating conditions.

III. IRC ANALYSIS

By Maxwell – Wagner model (Fig.1) and the polarized phenomena ongoing in dielectrics has been developed a new method called Isothermal Relaxation Current - Analysis (IRC analysis). This analysis can be used for the evaluation of changes in electro physical structures of material.

A. Microscopic view

Each aging process operates on insulation material. The microstructure and composition of insulation is changes of this

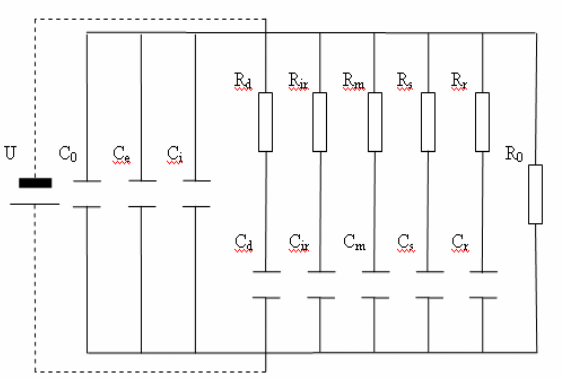


Fig. 1. A replacement scheme dielectric (it occurs in all types of polarization) [2].

process [1]. Characteristic changes result from this process in dielectric relaxation behavior.

IRC analysis observed a decreasing slow polarized process. This decreasing is measured with measurig technique (separate identification of current relaxation amplitudes in the time period more then 100 minutes). Most information of polarization spectrum (Fig. 2) is seen in the range of $10^{-3} \div 10^{5}$ sec..

$$A = \frac{Q(\tau_3)}{Q(\tau_2)} = \frac{1 + \frac{a_2.\tau_2}{a_1.\tau_1}.\left(1 - e^{-\frac{\tau_3}{\tau_2}}\right) + \frac{a_3.\tau_3}{a_1.\tau_1}.\left(1 - \frac{1}{e}\right)}{1 + \frac{a_2.\tau_2}{a_1.\tau_1}.\left(1 - \frac{1}{e}\right) + \frac{a_3.\tau_3}{a_1.\tau_1}.\left(1 - e^{-\frac{\tau_3}{\tau_2}}\right)} \qquad (1)$$

With IRC analysis we can monitoring non-destructive aging system using of aging factor A (1).

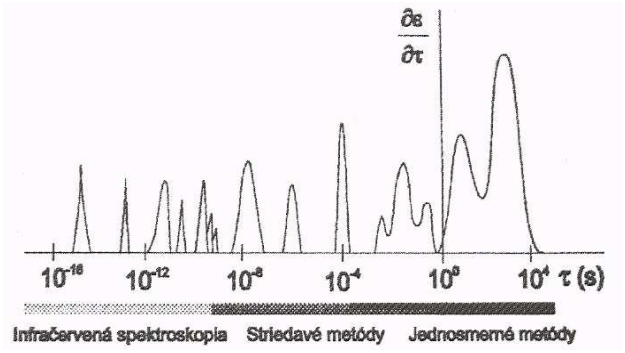


Fig. 2. Polarization spectrum of the insulating material [1], [3]

B. *Macroscopic view*

The main part of polarization spectrum is in the range $10^{-3} \div 10^{5}$ sec.. This range is possible watched using DC methods. The DC Methods are based on voltage and current responses observing.

The process current response to connect DC voltage (charge current and its discharging current) is on (Fig.3).

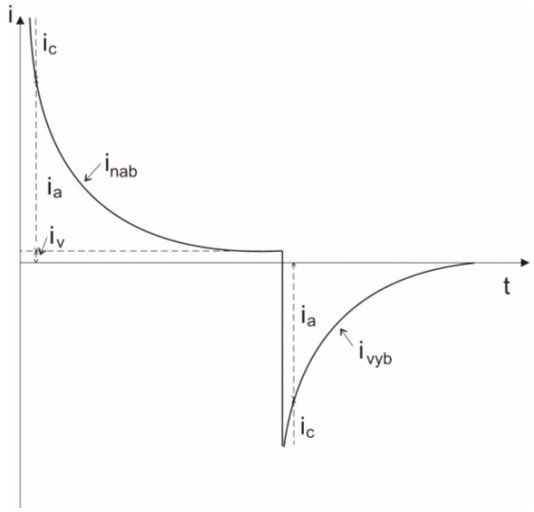


Fig. 3. Current response to an imposed pulse [1], [3], [4]

For the charging current is the relationship:

$$i_n(t) = i_c(t) + i_v + i_a(t) \tag{2}$$

where

$i_c(t)$ - curret from the geometric capacity,

i_v - conductivity current,

$i_a(t)$ - absorption current.

The total current flowing trough the dielectric is expressed as a sum of currents. These currents have exponentially decreasing amplitude. Expressed as:

$$i(t) = \frac{U}{R_i} + \sum_{i=1}^{n} I_{mi} \exp\left(\frac{-t}{\tau_i}\right) \tag{3}$$

where

U - applied DC voltage,

R_i - DC insulation resistance after infinitely long time,

I_{mi} - amplitude i-th elemental components Debye process,

τ_i - relaxation time constant i-th component of the Debye process.

IV. NEURAL NETWORK (NN)

Neural network is defined as follows according to [5]: Neural network (NN) is a massively parallel processor. It has tends to store experimental information and this information use further. The human brain is fake in two aspects:

- evidence collected in the NN learning,
- between neurons are used (synaptic weigh - SV) connection to store knowledge.

The artificial NNs are inspired of biological systems (simulation of the human brain). The NNs are applicated in practice. They are become mean for solution problems in the more areas of practice. They become tools for solving problems in many areas (eg. power electric engineering).

The neuron is a essential element of the NN. The structure of NN is on (Fig. 4).

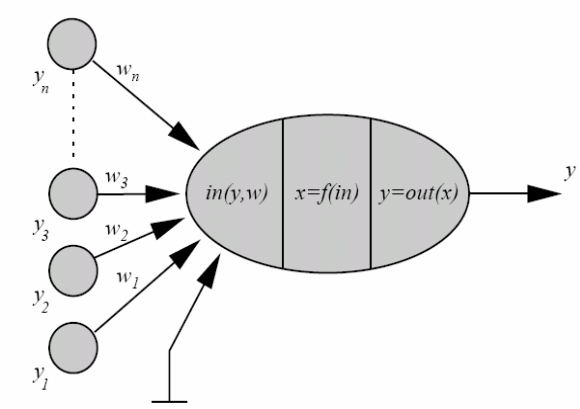


Fig. 4. Structure of the neuron [5]

Neuron consists of the following components [5]:

- entry into the neuron,
- neuron threshold – is value θ_i, she contributes to the input of the external world,
- input function f_in,
- neuron activation function f_a (resulting in a state of neuron x_i),
- output neuron function f_o (results in output y),
- synaptic weights, they are the synaptic connections (synapse), have their direction (beforsynaptic (source) and postsynaptic (target)) and combine

the individual neurons in the NN.

V. EXPERIMENT

The experiment detection degree of thermal degradation sample Relanex (Relanex is using for isolation the electric machine rotating (thermal class F working)) with IRC analysis in laboratory. The samples were then compared with samples measuring on really machines with NNs.

The experiment started with preparing of samples (the average of samples were 89 mm, adjustment were out of measurement electrode). The samples were exposed of thermal stress during determined time intervals in thermal chamber (186°C). We have 11 samples. These samples were exposed different age thermal stress (1 to 0 h, 2 to 12 hours, 3 to 24 hours, 4 to 48 hours, 5 to 96 hours, 6 - 192 hours, from 7 to 384 hours, from 8 to 768 hours, from 9 to 1536 hours, from 10 to 3072 hours, from 11-6144 hours) (Fig.5).

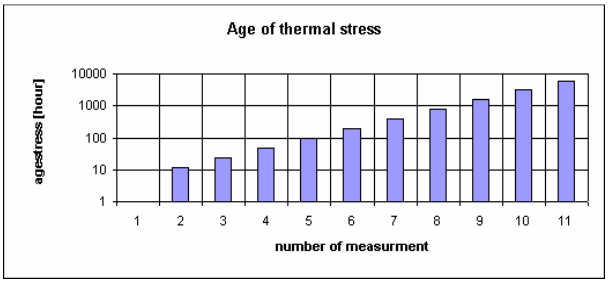


Fig. 5. Duration of heat stress samples in graphical form

The samples were inserted between the measuring electrodes KEITHLEY 8008 RESISTYVITY TEST FIXTURE. On the measuring electrodes test voltage 100 V was applied in electrometer KEITHLEY 617 for 1000 sec. The electrometer KEITHLEY 617 was connected to a computer via the bus. Measured charging currents of each sample were recorded to the computer. The whole measurement was controlled by the program created in Agilent VEE Pro. The diagram measurement is displayed in (Fig. 6).

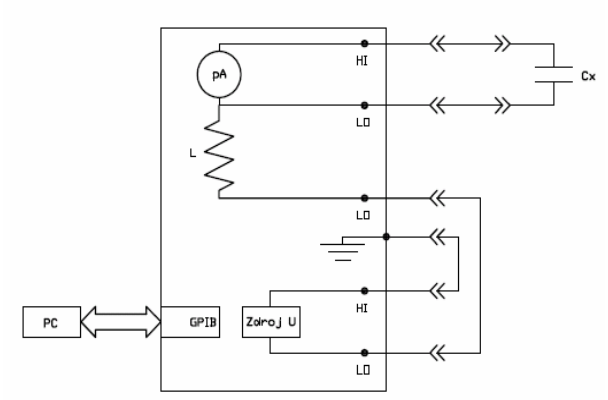


Fig. 6. Diagram measurement [6]

VI. RESULTS

We had identified the training data from measurement of Relanex sample. The date were after filtering normalized on interval <0, 1>. The NN after training had from 90 to 95 % fitting. The accuracy training of NN also depends given parameters learning NN.

After the training we started of test the data. The test data were data measure on the really machine in the practice. Now we normalized train and test data together on interval <0, 1>.

The output of neural network (Tab.1) gives us similarity test sample with train sample. Identified code, which gives us the percentage similarity of the test sample with train data. The graphic demo results are show on (fig. 7 to fig.11).

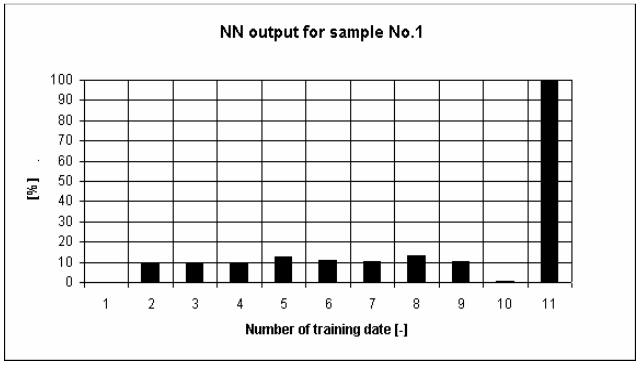


Fig. 7. Graphical output NN motor No.1

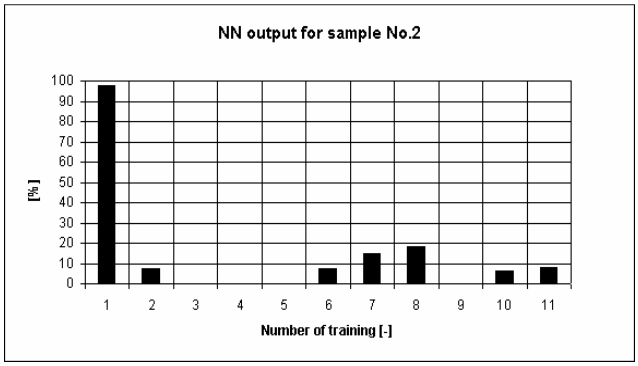


Fig. 8. Graphical output NN motor No.2

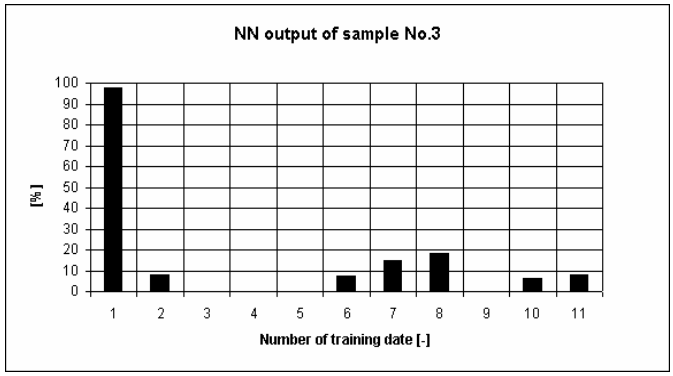


Fig. 9. Graphical output NN motor No.3

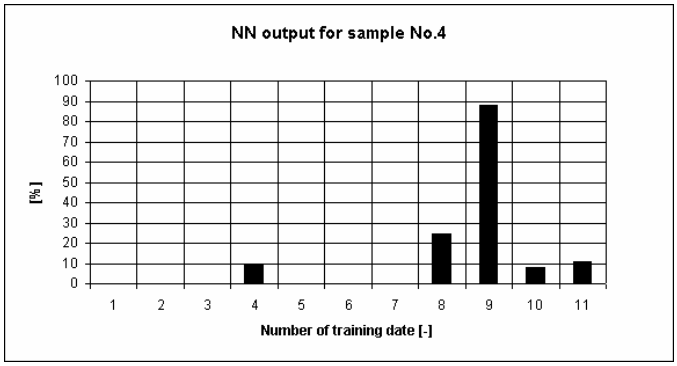


Fig. 10. Graphical output NN motor No.4

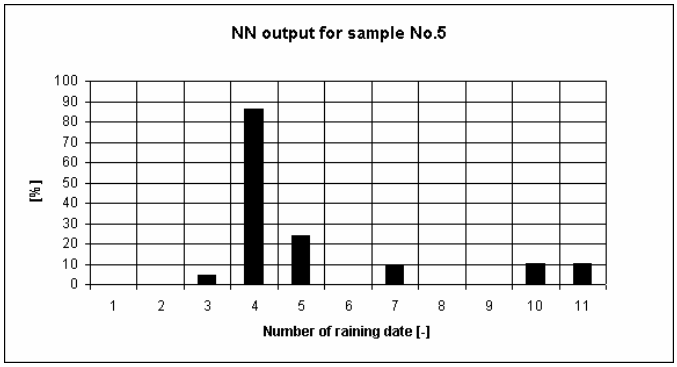


Fig. 11. Graphical output NN motor No.5

On (Table I) and on (fig.7 to fig.11) is clearly seen which training data are included NN to test data (samples).

TABLE I
DIGIT OUTPUT NEURAL NETWORK UNITS FOR MAGNETIC PROPERTIES

	Sample No.1	Sample No.2	Sample No.3	Sample No.4	Sample No.5
0 h	0	**97,46887**	**97,51446**	0	0,014562
12 h	9,379594	7,67498	7,714809	0,000001	0,000559
24 h	9,767261	0,000068	0,000068	0,000083	4,463115
48 h	9,606904	0	0	9,226135	**86,18335**
96 h	12,86703	0,000105	0,000107	0	24,23048
192 h	11,16556	7,439867	7,432626	0,001563	0,004647
384 h	10,42564	14,61370	14,68689	0,000453	9,048901
768 h	13,10272	18,1483	18,2528	24,43087	0,000147
1536 h	10,39069	0	0	**87,72861**	0
3072 h	0,410894	6,50568	6,45879	8,078659	10,02442
6144 h	**99,57049**	8,03123	8,005608	10,92094	10,05610

In the table II are show and compared results (different between the estimate and NN is small).

TABLE II
COMPARISON OF THE RESULTS OF THE IRC ANALYSIS RESULTS NN

Drive number	Estimated by simple insulation condition assessment	Status isolation by NN
M1	Isolation in a vulnerable state	insulation in critical state
M2	insulation in good state	insulation in very good state
M3	Isolation in a vulnerable state	insulation in very good state
M4	insulation in critical state	Isolation in a vulnerable state
M5	insulation in critical state	insulation in good state

The training of NN mainly depends of correct measure Relanex samples.

VII. CONCLUSION

The results of work are in comparison of measurements data and analysis this data. We are compare measure on really machine and measure in laboratory. The comparison we made by using artificial neural network with a controlled type learning of back error propagation. The result of this comparison was assignment the test samples to the degree thermal aging of insulation material. The classification into the classes using neural networks is that the much higher level as previously. The dependency on aging time is not monotonic and that is why their simply comparison is difficult.

The statements of neural network are different from the estimated states only deviations. This deviation we can tolerate. However question remains, which assessment more correspondent with really status.

The most important result of NN is ability to all five test objects proved clearly assign in to single stats aging without significant variance and thus clearly assess their thermal degradation.

The evaluation of using by neural network happens after filling the database by training data are significant instrument by the final evaluation of insulation system.

This system can by extended for more input properties of drivers.

THIS WORK WAS SUPPORTED BY SCIENTIFIC GRANT AGENCY OF THE MINISTRY OF EDUCATION OF THE SLOVAK REPUBLIC PROJECT VEGA NO. 1/0368/09 AND APVV-20-006005.

REFERENCES

[1] R. Cimbala, *Starnutie vysokonapäťových izolačných systémov*. TUKE, 2007.
[2] J. Hassdenteufel, *Elektrotechnické* materiály, ALFA SNTL, Praha, 1978.
[3] P. Semančík, *Tepelná degradácia izolačných* systémov, Dissertation Thesis, Košice, TUKE, 2007
[4] I. Kolcunová, *Diagnostic in power electric*, Lectures for the 5th Year, KEE, Košice, 2006, Available on the Internet: http://web.tuke.sk/fei-kee/predmety/dvee.html
[5] P. Sinčák and G. Andrejková, *Neurónové siete Inžiniersky prístup 1. diel*, ELFA, 1996.
[6] P. Kovaľ, *Degradation of insulation of electrical machines*, Dissertation Thesis, Košice, TUKE, 2008

Degradation Mechanism in Transformers Oil

[1]*Vieroslava ČAČKOVÁ,* [2]*Lýdia DEDINSKÁ,* [3]*Milan KVAKOVSKÝ*

[1] Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic
[2.] Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic
[3] Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]Vieroslava.Cackova@tuke.sk, [2]Lydia.Dedinska@tuke.sk, [3]Milan.Kvakovsky@tuke.sk

***Abstract*—This article is dealt with problem of degradation process in the oil. Attention is devoted of the ageing process in dependence on the temperature.**

***Keywords*— degradation mechanism, ageing process, transformers oil, temperature.**

I. INTRODUCTION

Power transformers belong to the most important component of energy systems, whose reliability is subject to the technical state of all its parts. Trouble-free operation of transformers is depended on the quality of their oil insulation too. Oil insulation fulfils the task electric insulation and medium of heat conduction.

Parameters of transformer oil are deteriorating during service.

It is resulted of combined effect of various influences, especially temperature, solid impurities, water, metal catalyze and electric field. These effects influence to produce organic acids, aldehydes, ketones and polymerization of unsaturated hydrocarbons. All these phenomena are known together as aging transformer oil. The inherence of organic acids in the insulating oil leads to degradation cellulose and oxidation of metals.

If this "aging" is decreased the observed parameters below a critical threshold, that is increased the probability of a major failure of transformer. In order to avoid such undesirable phenomena, on samples of transformers oil are carried out preventive tests and measurements of individual parameters. According to the diagnosis is shown the qualitative indicators of oil samples as longer inconvenient the recommended values, it must be exchanged for new and destroy deteriorate of oil, It can be chosen economically and environmentally preferable option - regeneration of aging oil

II. DEGRADATION MECHANISM

Degradation is non-reversible change in the functional properties of the materials and components as a result of service operations, and leads ultimately to the situation where the component cannot fulfil its task anymore and will fail. The degradation mechanism can be classified in stresses of: [1]:

- thermal,
- electrical,
- mechanical,
- environmental.

The real service influences almost always all together of these mechanism .on the acting equipment and degradation mechanism - ageing is combination of these all mechanisms.

III. INSULATION

Insulation materials are at the most exposed to possible degradation mechanisms, and therefore they have most attention.

Insulator is a substance that contains only a very small amount of free electrical charges and has negligibly small electrical conductivity. Electrical charge carriers are electrons in the insulator, positive and negative ions, or electrically charged colloidal particles. What matters is the amount of free electric charges in the substance and its overall response to the effects of electric field.

Insulation degradation caused by deterioration of insulation material properties, thereby increasing the number of failures.

Kind of insulation

- gaseous insulating materials,
- solid insulating materials,
- liquid insulating materials.

A. Gaseous insulation

Gaseous insulation material is characterized generally by high insulation resistance, low dielectric loss factor and relative permittivity around of value 1 [2] [3]. Gaseous insulators are subjected to aging slightly. It is not occurred above electric strength thresholds breakdown but flashover and restoration of electric strength is restored after the disappearance of electric field and the conditions for a self-maintained discharge. Generation polluting impurities from other parts of electrical equipment are resulted from the arcing degrade gaseous insulator and thus impair its electro - physical properties in general. Particularly it is substantial in the case of sulfur hexafluoride SF_6.

B. Solid insulators

Solid insulators are characterized by higher electrical strength and permittivity insulators such as gas and liquid insulators. In the past, the solid insulating material used mainly asbestos, marble and slate. Gradually asbestos became more remarkable in the various forms. [2] [3].

C. Liquid insulators

Liquid insulators are used widely in electrical engineering. The main reason of using liquid electrical insulators is their

electrical strength, cooling function, impregnation capacity and the ability to suppression discharge. Liquid insulating materials can be divided follows [3]:

- mineral oil,
- vegetable oil,
- synthetic oil,
- other (e.g., magnetic liquid).

Only materials with covalent bonds in molecules can be as liquid insulators. Substances with ionic bonds in the liquid phase (molten salts) are second-class conductors (electrolytes) and substances with metal binding in the liquid phase (liquid and molten metal) are conductors of the first class.

Liquid insulators at form oils are used most frequently as insulating materials of transformers. Natural esters are suitable to lower power transformers.

The trouble-free service is used to guard continually monitor and evaluate of isolation parameters.

IV. PARAMETERS OF TRANSFORMER OILS

The number of monitored parameters in transformer oils is increased with operating voltage levels and power of transformers. The oil transformers with a voltage of 110 kV and higher is measured breakdown voltage, acidity, water content, loss factor, surface tension and content of the inhibitor. For transformers with capacities over 100 MVA is performed also quantitative and qualitative analysis of gases contained in oil by dissolved gas analysis.

Status of oil filling is assessed comprehensively according to these measurements and tests (or other parameters, such as colour, the test for accelerated aging, ... etc.).

By observing the properties of electro-course change may be material to determine trends. For tracking trends is necessary to choose appropriate diagnostic methods.

We have chosen to observation IRC analysis (Isothermic Relaxation Current - Analysis) of heat degradation. We measured the charging currents in the time interval 1000 s, at gradually increasing temperatures from 40°C to 100°C. The charging characteristics of currents can be seen in Fig. 1.

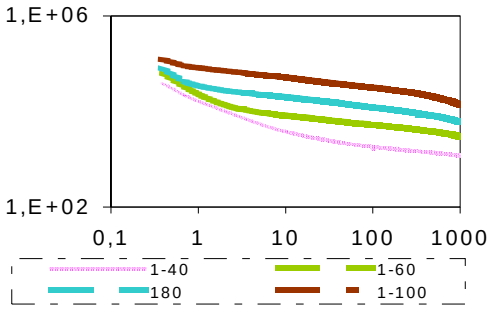


Fig.1 Temperature dependence the charging current of time for the oil with temperature T40°C-T100°C.

With increasing temperature is occurred to increasing of values charging current in the oil, at which the shape of the curve is similar.

The curve of charging current can be spread to exponential function with time constants (τ_1, τ_2, τ_3, τ_4 ...). It is depending to detect number of polarized processes in the dielectric during measurements. Each polarizing constant corresponds to certain polarization action.

Measurement is lasted for 1000s during is occurred markedly on samples seven polarized processes. Each polarization process can be replaced by an equivalent RC element in the substitute model of dielectrics. In our measurements is showed a significant 3 polarization processes. The next approximating values of polarization processes could not be interpreted physically for most samples. We are calculated approximation the value of elementary streams I_{mi} and time constants τ_i for the three polarization processes. We repeat the process for increasing the temperature 1

In Figures 2,3 is presented a graphical representation of the calculated elementary relaxation streams and relaxation time constants depending on temperature according to Table 1

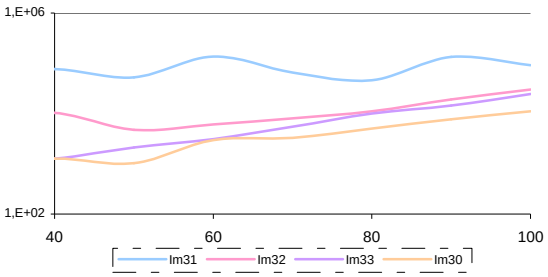


Fig.2 Approximation of currents for the third polarization process

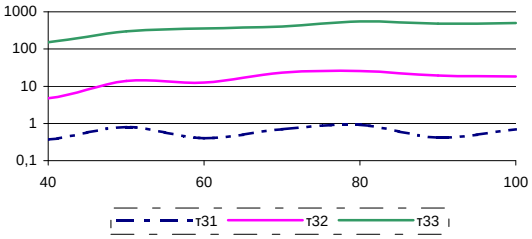


Fig.3 Approximation of time constants for the third polarization process

The values for the elements of the replacement scheme is calculated according to third approximation of the values of elementary streams and the time constants after the formulas

$$R_i = \frac{U_o}{I_m} \quad (1)$$

$$C_i = \frac{\tau_i}{R_i} \quad (2)$$

where I_{mi} τ_i -values from Table 1
U_0 -connected DC 100V

TAB. 1 TEMPERATURE DEPENDENCE THE RESISTIVITES AND CAPACITIES FOR THE OIL WITH TEMPERATURE T40°C-T100°C.

	T °C						
	40	**50**	**60**	**70**	**80**	**90**	**100**
R_{31}	0,001	0,002	0,001	0,002	0,002	0,001	0,001
R_{32}	0,010	0,021	0,017	0,012	0,009	0,005	0,003
R_{33}	0,079	0,047	0,033	0,018	0,010	0,007	0,004
R_{30}	0,079	0,098	0,034	0,030	0,020	0,013	0,009
C_{31}	285,6	420,1	545,7	450,6	421,9	564,5	645,5
C_{32}	483,7	655,9	753,8	1852,1	2829,2	3723,3	5521,1
C_{33}	1948,4	6275,0	10988,4	22029,2	55187,2	69789,3	121904,0

We can make up the equivalent replacement model dielectric based on the calculated values.

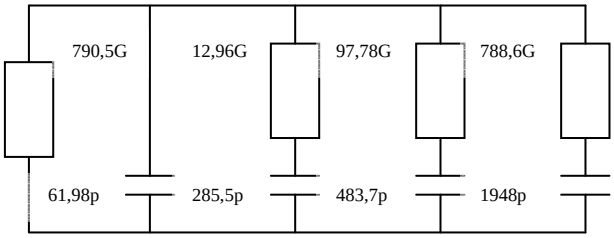


Fig..4 Replacement model at 40°C with the conductivity component

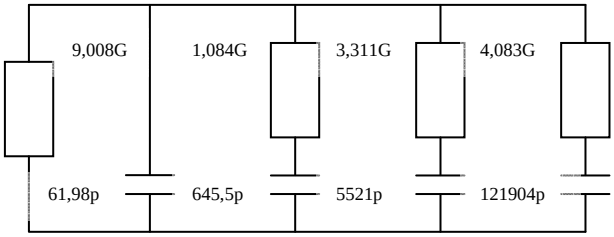


Fig.5: Replacement model at 100°C with the conductivity component

The replacement model is described the behavior of system in the time respectively frequency domain

The replacement model is described the behaviour of system in the time respectively frequency domain.

V. CONCLUSION

Comparing alternate models, we see that with increasing temperature reduces the value of R_i and conversely increase the value of C_i. This fact corresponds to the graphical representation of the approximation of the time constants and elementary streams for the oil, which increases with increasing temperature. Extension of relaxation time constants means prolonging of polarization. These changes can be evaluated the aging process without endangering life, and thus determine the degree of degradation. Important is archiving of measured data for possible after comparison and evaluation.

Disorders are arisen in the system as direct consequence of degradation changes, and their development, aging leads to the complete failure of insulation system. Disorders significantly increase the economic costs service of equipment. Aging is a very complicated process, and therefore the statistical information-processing try to standardize the conditions of investigative methods for determining the service decisions from information on the selected parameters.

ACKNOWLEDGMENT

This work was supported by scientific Grant Agency of the ministry of Education of the Slovak Republic project VEGA No. 1/0368/09 and APVV-20-006005.

REFERENCES

[1] V. K. Agarwal, H. M. Banford, B. S. Bernstein, E. L. Brancato, R. A. Fouracre, G. C. Montanari, J. L. Parpal, H. N. Seguin, D. M. Ryder and J. Tanaka, *"The Mysteries of Multifactor Ageing,"* IEEE Electrical Insulation Magazine, Vol. 11, No. 3, May/June 1995, pp. 37-43.

[2] M ARTON, K. – BANSKÝ, J. – KLUCH, K. – SOMORA, M.: *Elektrotechnické materiály,* ALFA, Bratislava, 1979.

[3] HASSDENTEUFEL, Josef – kol.: *Elektrotechnické materiály,* ALFA SNTL, Praha, 1978.

DC/DC resonant converter for PV system

[1]Erik EÖTVÖS,[2]Marcel BODOR

[1,2]Dept. of Electrical Engineering, Mechatronics and Industrial Engineering, FEI TU of Košice, Slovak Republic

[1]erik.eotvos@student.tuke.sk, [2]marcel.bodor@tuke.sk

***Abstract*— Step-up DC/DC converter, as a part of the photovoltaic (PV) system is described in the paper. The system consists of the mentioned DC/DC converter and single phase inverter, and they together serve for transmission of the energy from PV panels to the grid. The converter is based on the LLC resonant architecture. The task of the converter is to increase and control of the output voltage, which is supply voltage for in the cascade connected inverter. The converter also provides the electrical isolation of PV panels from the grid. It is controlled by an 8-bit microcontroller unit (MCU), which also tracks maximum power point (MPP) to achieve maximum available power from PV. The converter operates at high switching frequency to achieve small size of the power transformer. The main benefit of this converter consists in zero-voltage switching (ZVS) of the primary MOSFETs and zero-current switching (ZCS) of the rectifier diodes over the entire operating range. Laboratory model with maximum 95.5% efficiency was built to verify the properties of the LLC resonant DC/DC converter.**

***Keywords*—LLC resonant converter, ZVS, ZCS.**

I. Introduction

In nowadays the renewable energy resources are used increasingly. The photovoltaics is the most dynamically developing field of the renewable energy sources. It is desired to use the renewable energy sources with maximal efficiency. One of the possibilities how increase the efficiency of a PV system, is to increase the efficiency of an inverter. There are quantities of inverters for PV systems on the market. Some inverters include DC/DC step-up converter, depending on whether the inverter is connected to the string of PV panels with voltage higher than the maximum value of the grid voltage. There is a problem with capacitive currents flowing through the inverter when using thin layer PV panels. One possible solution is to use the DC/DC converter with a transformer.

Described converter should work with input voltage range from 60V to 100V. The required output voltage (input voltage for the inverter) is 400V and maximum output power should be about 600W.

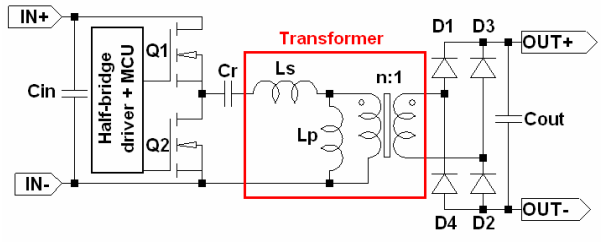


Fig. 1. Principal scheme of the LLC converter.

Considering that it is required to have the high efficiency and an electrical isolation the series resonant LLC half-bridge converter was chosen [1]. Principal scheme of the LLC converter is shown in Fig.1. It consists of the half-bridge inverter, created by power MOSFET switches, from which the resonant tank is supplied.

The resonant tank of the converter comprises of series "L_S" and parallel inductance "L_P" and the resonant capacitor "C_R". On secondary side of the transformer there is a full-bridge rectifier with a filter capacitor.

II. Principle of Operation

The power MOSFETs are switched with variable frequency with fixed 50% duty cycle and no overlapping. The resonant tank has two main resonant frequencies. The higher resonant frequency "f_R" depends on the series inductance and the resonant capacitor and is calculated by using (1). The lower resonant frequency "f_0" of the resonant tank depends on the series inductance, the resonant capacitor and also on the series inductance (2). If switching frequency is higher than "f_R", the converter operates always in the inductive area. It means that the resonant tank current lags input voltage square waveform, and therefore switches work under ZVS condition. Below the "f_0" resonant frequency, the resonant tank behaves as a capacitive load. Therefore resonant tank current leads the input voltage. Switches works under ZCS condition. The area between "f_0" and "f_R" is split by a borderline to the capacitive and the inductive region. The operating point in this area depends on the load of the converter. When the switching frequency is equal to the resonant frequency "f_R", the voltage gain of the resonant tank is 1. It means that converter is load independent. At normal operation condition, the operating point should be placed near to this resonant frequency. Fig. 2. shows voltage gain curves of resonant tank for few load conditions. We can see the capacitive region on the left side of the borderline and the inductive region on its right side.

$$f_R = \frac{1}{2\pi\sqrt{L_S C_R}} \quad (1) \qquad f_0 = \frac{1}{2\pi\sqrt{(L_S + L_P)C_R}} \quad (2)$$

$$M(Q, f_N, \lambda) = \frac{1}{\sqrt{\left(1+\lambda-\frac{\lambda}{f_N^{\,2}}\right)^2 + Q^2\left(f_N - \frac{1}{f_N}\right)^2}} \quad (3)$$

$$Q = \frac{\pi^2 Z_0 P_{OUT}}{8n^2 V_{OUT}^{\,2}} \quad (4) \qquad Z_0 = 2\pi f_R L_R \quad (5)$$

$$\lambda = \frac{L_S}{L_P} \quad (6) \qquad f_N = \frac{f_{SW}}{f_R} \quad (7)$$

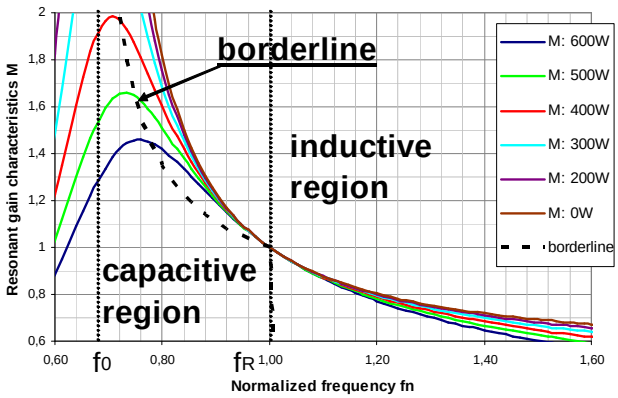


Fig. 2. Voltage gain curves of the resonant tank.

The voltage gain curves are calculated using (3), where "Q" is quality factor (4), "λ" is inductance ratio (6) and "f_N" is the normalized frequency (7).

The „Z_0"is the characteristic impedance (5) and "n" in (4) is the turn ratio of the transformer [2]-[6].

The operation of converter we can explain according to Fig. 3. in the next six phases:

1. The resonant tank current from the previous phase is flowing now through the body diode of Q1. It causes that the voltage across Q1 drops to zero, and creates the zero voltage condition for the lossless turn-on of the switch.
2. Now the current flows through Q1, and has quasi sinusoidal character. Therefore the turn-off current is much smaller.
3. Q1 and Q2 are switched-off. The current of Q1 drops to zero immediately, but the voltage across the switch rises slowly due to the charging of the output capacitance of the MOSFET. It reduces the turn-off losses.
4. Like in the first phase, the resonant tank current flows through the body diode but now of the switch Q2. The voltage of Q2 falls to zero. The switch is turned on.
5. The current flows through Q2 similarly to the second phase.
6. Switches Q1 and Q2 are switched-off again. The current of Q2 falls to zero, but the voltage rises slowly again due to charging of the transistor output capacitance [7].

7. Simulation

The resonant converter was simulated in LTSpice IV program. Components of the resonant tank were calculated by equations presented in chapter II. Working frequency was set to 150 kHz. The collector current and collector-emitter voltage of the switch Q1 are shown in Fig. 4 (upper waveforms). It can be seen there that the switch Q1 starts to conduct when voltage of the switch is zero and thus ZVS is achieved for primary switches. When Q1 is turned-off, the current falls to zero, but voltages rise slowly, because the output capacitance of MOSFET is charging.

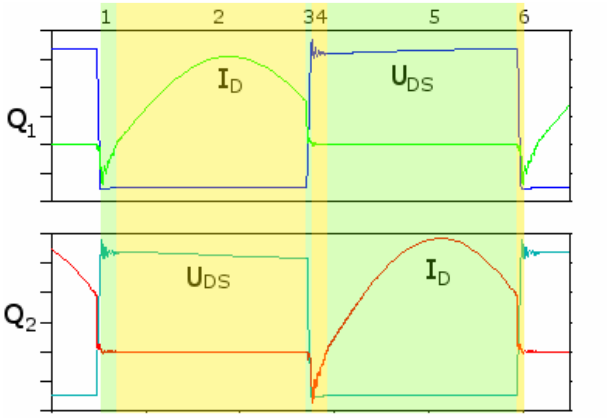


Fig. 3. Characteristic waveforms of collector – emitter voltage and collector current of half-bridge switches.

On the bottom picture there are waveforms of current and voltage of the rectifier diode D1. The current through the diode starts and stops flowing when voltage is near to zero. Therefore the switching losses are minimal.

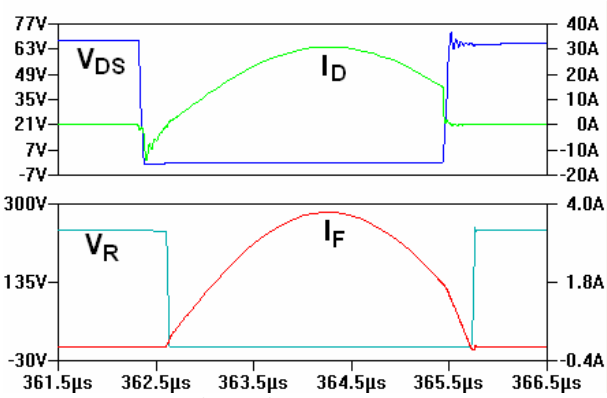


Fig. 4. Simulated waveforms of voltages and currents on switch Q1 and diode D1.

III. Laboratory Model of the Converter

The laboratory model of the resonant converter was built.

Parameters of the model:

Input voltage range V_{IN} = 60-100V.

Output voltage V_{OUT} = 400V.

Output power P_{OUT} = 600W.

When the circuit was designed, there was a need to solve few problems. Because the converter operates at high frequency and in addition with high currents, each part of circuit must by able to withstand this condition. When choosing discrete components, such as MOSFETs and rectifier diodes, we must care; that they should have minimal loses in active mode and short switching times. But there is not problem with the maximum operating voltage, because the voltage stress is very low due to soft switching.

The input and resonant capacitors must handle very high load current at high frequency. Quality capacitors of "KPI" type are suitable to fulfil these conditions.

The transformer can be designed so that it integrates the series and the parallel inductance in one circuit. In classic transformer, the parallel inductance can by replaced by a magnetizing inductance, and series inductance by a primary

leakage inductance. But there is problem with inductance ratio between magnetizing and leakage inductance, because the ratio is very small. One solution is to integrate an air gap into the magnetic circuit of the transformer. By adjusting the air gap; we can control the size of the magnetizing inductance and thereby also the inductance ratio (6). For winding we must use litz-wire to avoid a skin effect in a conductor [8].

High/Low side driver is used for driving of the MOSFETs. The driver is controlled by the 8-bit MCU. The circuit has a voltage feedback with an optocoupler.

IV. Experimental Results

The converter was connected to the DC voltage source, and loaded by an adjustable resistor. Input voltage was set to 90V. Voltages and currents of primary MOSFETs and secondary rectifier diodes were measured by a digital oscilloscope. Result is in Fig. 5. Waveforms of the voltage and the current of the MOSFET Q1 are in the top picture and waveforms of the voltage and the current of the rectifier diode are below. We can see that the current starts to flow when voltage across MOSFET is zero and thus ZVS is achieved. Moreover, the switch-off current is minimal. The current through the rectifier diode starts to flow when the voltage is near to zero. It means small switching loss. The voltage of the diode starts to rise when current falls to zero. The ZCS of the diode is achieved, too. If we look at waveforms of both voltages, we do not see any voltage spikes. It means that there is no voltage stress across MOSFETs and rectifier diodes. Therefore we can use these components with lower break down voltage, but with better other parameters.

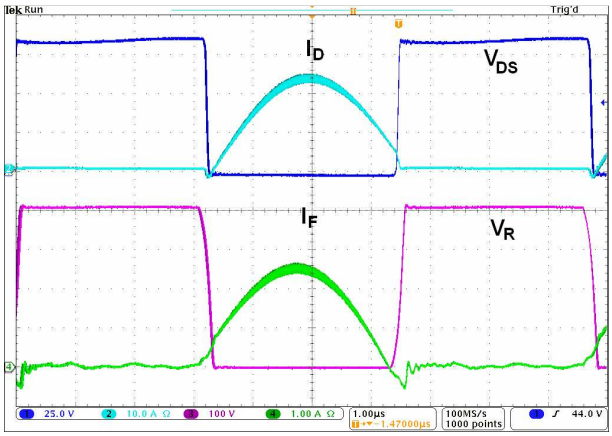


Fig. 5. Measured waveforms of currents and voltages on the primary MOSFET and the rectifier diode.

Next was measured the efficiency of the converter. The converter was loaded from 10% to 100%. Results are in a graph in Fig. 6. As we can see, the efficiency is high in wide range of load, especially over 30% load. In 50% load the efficiency reaches its maximum, and up to 100% load, slowly declines.

V. Conclusion

The resonant converter with LLC topology has many advantages compared to other converters. Due to the high efficiency over the entire operation range, the converter is well suitable for applications such as PV systems.

Next I am going to build the compact version of the converter, implement the MPPT function and connect the converter to the simulated PV panel DC source. Afterwards, connect to the inverter and verify the operation of the whole converter.

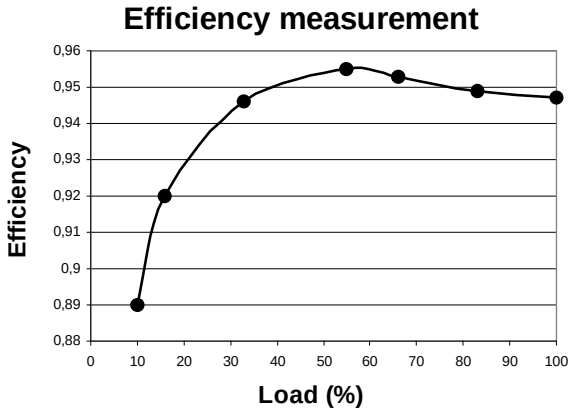


Fig. 6. Measured efficiency of converter.

Acknowledgment

This work was supported by Slovak Research and Development Agency under project APVV-0095-07 and by Scientific Grant Agency of the Ministry of Education of Slovak Republic under the contract VEGA No. 1/0099/09.

References

[1] Ya Liu, "High Efficiency Optimization of LLC Resonant Converter for Wide Load Range", Blacksburg, Virginia 2007
[2] Bo Yang, "Topology Investigation Front End DC/DC Power Conversion for Distributed System", Virginia 2003
[3] ST Microelectronics, "LLC resonant half-bridge converter design guideline", October 2007
[4] Mingping Mao, Dimitar Tchobanov, Dong Li, Martin Maerz, Tobias Gerber, Gerald Deboy, Leo Lorenz, "Analysis and Design of a 1MHz LLC Resonant Converter with Coreless Transformer Driver", 2007
[5] Hang-Seok Choi/Ph.D, FPS Application group, "Half-bridge LLC Resonant Converter Design Using FSFR-series Fairchild Power Switch", 2007
[6] Christophere Basso, "Understanding the LLC Structure in Resonant Applications", 2008
[7] STMicroelectronics, "Simplified Analysis and Design of Series-resonant LLC Half-Bridge Converters", 2006
[8] Fu Keung Wong, B. Eng. and M Phil., "High Frequency Transformer for Switching Mode Power Supplies", Griffith University, Brisbane, Australia, March 2004
[9] Kácsor, G. , Špánik, P., Lokšeninec, I.: Simulation Analysis of a Zero Voltage and Zero Current Switching, DC/DC Converter. In proceedings: Transcom ´03, Žilina, Slovak Republic, 23 – 25 June 2003, pp. 47-50
[10] Hamar, J., Buti, B., Nagy,I.: Dual Channel Resonant DC-DC Converter Family. EPE Journal, Volume 17, Issue 3, Jul-Sep. 2007, pp.5-15.

Slovak Traffic Signs and Their Preprocessing in HSV Color Space

[1]*Martin FIFIK*

[1]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]martin.fifik@tuke.sk

***Abstract*— This paper describes preprocessing in HSV color space. Tested sings are taken from Slovak roads. Complete preprocessing stage with color segmentation is described. Results and potential of described method are discussed.**

***Keywords*—Traffic signs, HSV color space, color segmentation, classification.**

I. INTRODUCTION

The rules for safety traffic are displayed on traffic signs. Traffic signs are designed to show us some rule or warn us before something. If we leave out some traffic sign, we can put us in danger situation or worst; we can be participant in car accident. An automatic road sign detection system will be helpful [2,3].

Traffic sign follow strictly shape and color. For that, can be good recognizing from surrounding environment, while driving. Every government has their laws about this, how the traffic sign must look like. Therefore the solution here presented will follow Slovak road signs with their shape and colors [1,4].

II. SYSTEM DESCRIPTION

A. HSV Color Space

Input is converted in to the HSV color space. Every traffic sign has his dominant color. On Slovak roads most often yellow, red and blue color are used. This means we need to create three binary maps, one for each of these colors.

By analyzing hue component (H), we can identify blue, yellow and red regions in our detected image. For each image pixel, hue-based detection of blue, red and yellow colors is done. For each color one passes one of following equation[1]

$$Y = e^{\frac{-(x-42)^2}{30^2}} \tag{1}$$

$$R = e^{\frac{-x^2}{20^2}} + e^{\frac{-(x-255)^2}{20^2}} \tag{2}$$

$$B = e^{\frac{-(x-170)^2}{30^2}} \tag{3}$$

Equations *Y* gives values close to 1 for yellow regions, *R* gives values close to 1 for red regions and *B* gives values close to 1 for blue regions. In this equations we can see, that H can be from range 0-255. Yellow can be detected near value 42, red near values 0 and 255 and blue value is 170. These equations can be tuned for every color.

Now we need saturation detection value, by exploring the S component. This is described by following equation

$$S = e^{\frac{-(x-255)^2}{115^2}} \tag{4}$$

From equations (1), (2) and (3) we got 3 values. Every value is multiplied with *S* value. This value will be called *D*. From *D* we create D_n, which means *D* normalized. Values close to zero will be discarded. Other will follow next equation (5).

$$D_n = \begin{cases} 0, if\ \det < 0{,}3 \\ \left(\dfrac{\det - 0{,}3}{0{,}7 - 0{,}3}\right)^2 \\ 1, if\ \det \geq 0{,}7 \end{cases} \tag{5}$$

Now the threshold can be created. For this we use Otsu's algorithm on D_n [6]. Then we can create three binary maps[1,5].

Now every binary map must be cleared. Too small regions and too big regions are discarded. In first step we must find first white point in image. Searching is done by rows. After finding first white point, method called seed-fill is used. With this method we are finding regions, and these regions are valued. If valued region has value lower than 400, then is discarded. If region has value more than 16000 then is also discarded. The regions that are left are potential candidates for next processing.

Method seed-fill we can apply only, if all white areas in image have their boundaries. Next condition is that there must be at least one white point. After successful finding white point A (seed), its neighbors are tested (Fig. 1).

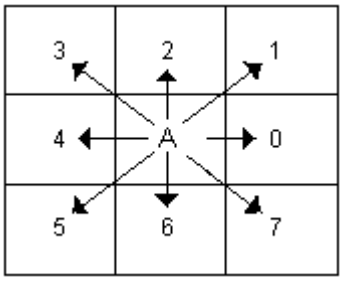


Fig.1 Values 0-7 for neighbors of seed A

This white point is filled and its neighbors' relations are checked. These points are now stored in seed vector. In next step we continue with last filled point. If this point has no unfilled neighbor, then point from seed vector is choosing for next checking. After using all points from seed vector, we have our region that has value. On the end we have selected **R**egion **O**f **I**nterest(s) (**ROI**).

B. Shape Detection

Fig.2 Four perfect shapes

Now with pattern matching method, on that size of ROI are created perfect shapes, like it is show on Fig.2. Here every **ROI** is tested pixel by pixel with every perfect shape, with circle, rhomb, triangle, reverse triangle, filled circle-STOP sign, square /rectangle.

C. Sign Type Classification

TABLE I. EXAMPLE OF TRAFFIC SIGN CLASSIFICATION

Color / Shape	Red	Blue	Yellow
Square/Rectangle	-	Information	-
Circle	Obligation	Prohibition	-
Rhomb	-	-	Highway
Triangle	Danger	-	-
Reverse triangle	Yield sign	-	-
Cut square	Stop sign	-	-

Now when we got shape and color we can classify traffic signs in classes. This is shown in table I.

III. EXPERIMENTS

Experiments were done for HSV color space. Input images were in resolution 640x480 pixels. Tests were done on Pentium PC with dual core processor 2 x 2 GHz.

Preprocessing time was average from 1~2 seconds. Table classification average time was 0,2 seconds.

TABLE II. RESULTS FOR HSV COLOR SPACE

Brightness condition	Number of sings	Detected signs	Rate[%]	Average rate[%]
low	93	80	86,022	
normal	220	211	95,909	
extreme	13	3	23,077	90,18

IV. CONCLUSION

Experiments show us average system recognition rate 90% at HSV color space. Factors which affect images in preprocessing stage was darkness and sun lightning.

All reached results was in 2 seconds per image, which means that this system can be used in real time.

ACKNOWLEDGMENT

This work was partially supported from the grants VEGA, project COST IC0802 and by Agency of the Ministry of Education of the Slovak Republic for the Structural Funds of the EU under the project Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020).

REFERENCES

[1] F. P. Paulo, L. P. Correia, "Automatic detection and classification of traffic signs," Image Analysis for Multimedia Interactive Services, 2007, WIAMIS, 6-8 June 2007.

[2] A. Laika, W. Stechele, "A review of different object recognition methods for the application in driver assistance systems, " Image Analysis for Multimedia Interactive Services, 2007, WIAMIS, 6-8 June 2007.

[3] J. Turán, M. Fifik, Ľ. Ovseník and J. Turán Jr., "Transform based system for traffic sign recognition" 15th International Conference on Systems, Signals & Image Processing IWSSIP'08, Bratislava, Slovakia, 2008

[4] A. Broggi, P. Cerri, P. Medici, P. P. Porta, "Real Time Road Signs Recognition", Intelligent Vehicles Symposium, 2007 IEEE Volume , Issue , 13-15, 981 – 986, June 2007

[5] J. Turán, Ľ. Ovseník, J. Turán, Jr., "Transform Based Invariant Feature Extraction", 13th International Conference on Systems, Signals & Image Processing IWSSIP'06, Budapest, Hungary, September 21-23, 79-82, 2006

[6] N. Otsu. "A threshold selection method from gray-level histogram" IEEE Transactions on Pattern Analysis and machine Intelligence, vol. 25, nº8, August 2003, pp 959- 973

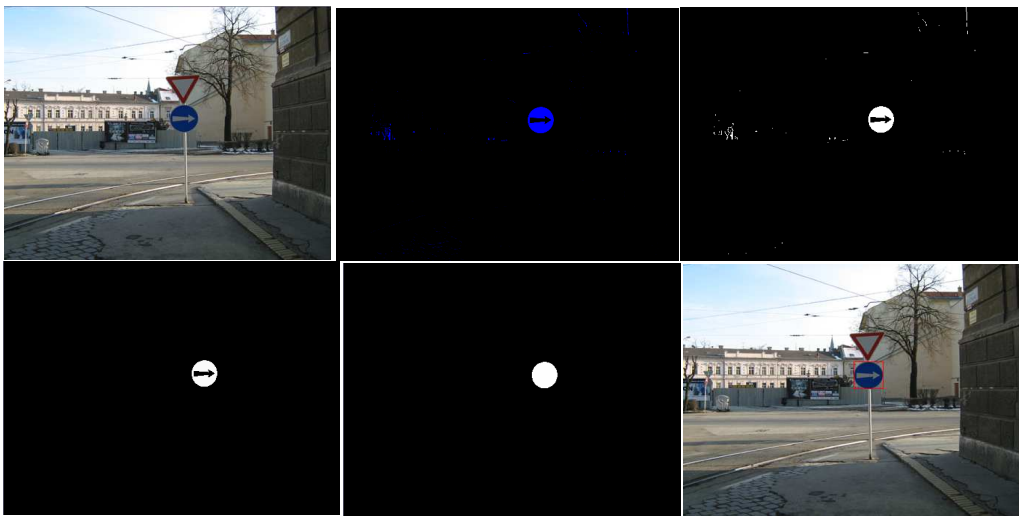


Fig. 3 Preprocessing and ROI extraction from blue color in HSV color space

Iterative receiver for nonlinearly distorted SC-FDMA tranmission

[1]*Juraj GAZDA*

[1]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]juraj.gazda@tuke.sk

***Abstract*—Single Carrier Frequency Division Multiplex (SC-FDMA) is the modulation technique being considered in Long Term Evolution (LTE) proposal announced recently. The reason for this stems from the fact that SC-FDMA envelope fluctuation is considerable lower than that of conventional Orthogonal Frequency Division Multiplex (OFDM) technique, thereby results in less sensitivity to the nonlinear distortion introduced by the High Power Amplifiers (HPA). However, in order to achieve high spectrum and power efficiency of the transmission, it is convinient to look for additional strategies to further improve the SC-FDMA robustness against the nonlinear amplification. Therefore, in this paper we introduce the technique based on the iterative detection of nonlinearly distorted SC-FDMA symbols that aims to tackle the nonlinear distortion in an iterative fashion. The performance analysis show the considerable performance improvements of presented technique performing over frequency selective channels.**

***Keywords*—HPA, iterative detection, SC-FDMA**

I. Introduction

SC-FDMA is recognized as an attractive modulation scheme that has been agreed to be used in the uplink of LTE and according to the 3rd Generation Partnership Project (3GPP) Release 10 Standardisation will also be employed in the upcoming fourth generation (4G) of wireless cellular system named as LTE Advanced [1]. The mayor reason standing behind the application of SC-FDMA instead of OFDM in the uplink of the wireless cellular systems is its considerable lower instantaneous power variation enabling battery efficient operation of the power amplifier without using extremely high input back off (IBO). However, following strict requirements of the 3GPP Release 10, further strategies to improve the performance of SC-FDMA operating particularly over largely nonlinearly distorted environment should be introduced.

Up to date, sensitivity to the nonlinear amplification has received a special interest especially in conventional OFDM. In order to cope with this problem, many strategies based on different approaches have been proposed so far. Besides others, very attractive and promising solutions have been designed by Tellado et. al [2], and Colas et al. [3] where the effect of nonlinear distortion due to HPA is compensated at the receiver side in an iterative manner. It should be noted, that interesting performance improvement reported there can be achieved even after very few iterations.

Unfortunately, the SC-FDMA sensitivity to nonlinear amplification and corresponding nonlinear sensitivity reduction techniques do not receive special attention and one can find only several contributions related to this topic. The theoretical analysis of SC-FDMA based transmission systems undergoing nonlinear amplification has been addressed in [4]. Performance evaluation of nonlinear distortion effects in SC-FDMA has been evaluated in [5] and [6]. In order to alleviate the nonlinear distortion, Deumal et al. have introduced in [7],[8] the iterative decoder based on the solutions in [2],[3] but reformulated to SC-FDMA environment.

The goal of this paper is to provide the extension of work presented in [7],[8] and introduce the modification of presented scheme for coded SC-FDMA transmission. This modification is based on jointly using the original receiver structure presented in [7],[8] concatenated with the channel decoder represented by the hard output Viterbi Algorithm in this paper. The performance results show the reasonable performance improvement that can be gained by applying the modified structure in comparison with the system introduced in [7],[8].

II. System model

In SC-FDMA, the bit sequence is first mapped onto N complex modulation symbols. QPSK, 16-QAM or 64-QAM are employed in LTE according to the channel conditions. Afterwards, the block of N data symbols is applied to a size-N discrete Fourier transform (DFT), this is typically known as DFT-precoding. The output of the DFT is applied to consecutive inputs of a size-M inverse DFT corresponding to the desired part of the overall system bandwidth, where $M \leq N$ and the unused inputs of the IDFT are set to zero. Finally, as in OFDM a cyclic prefix is inserted to each transmitted block.

Let $a_i, i = 0, \ldots, N-1$ be the complex data symbols, then the signal at the output of DFT-precoder can be expressed as

$$S_k = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} a_i e^{-j2\pi ik/N}, \quad k = 0, \ldots, N-1. \quad (1)$$

Consider a baseband OFDM symbol $s(t)$ defined over the time interval $t \in [0, T_s)$,

$$s(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} S_k e^{j2\pi(k+k_0)t/T_s}, \quad (2)$$

where k_0 is the position of the first assigned subcarrier. For the sake of brevity and without loss of generality we assume $k_0 = 0$. If $s(t)$ is sampled at a frequency LN/T_s, where $L = M/N$ is the oversampling factor and N/T_s is the Nyquist rate, the signal at the output of the SC-FDMA modulator is

$$s_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} S_k e^{j2\pi kn/M}, \quad n = 0, 1, \ldots, M-1. \quad (3)$$

By substituting (1) in (3), the SC-FDMA signal is found to be

$$s_n = \frac{1}{N}\sum_{i=0}^{N-1} a_i \sum_{k=0}^{N-1} e^{j2\pi k(n-Li)/M}. \quad (4)$$

By analyzing (4) at multiples of the oversampling factor, $n = Lr$, one observes that the Lr-th sample of the time domain SC-FDMA signal is equal to the data symbol a_r. The samples at positions $n \neq Lr$ describe the transition of the time domain signal between values a_r and a_{r+1}. The presence of these transitions between modulated symbols increase peak-to average power ratio (PAPR) and cubic metric (CM) [4].

III. Advanced detection of nonlinearly distorted SC-FDMA signal

A. Compensation for the realization-varying complex amplification term

Let us model the SC-FDMA system as a stochastic process $\mathcal{S}$ such that each symbol $s^{(m)}(t), \forall t \in [0, T_s)$, is a different realization of $\mathcal{S}$. The corresponding signal at the nonlinearity output is denoted as $s_d^{(m)}(t)$. Conventional receivers only process the part of $s_d^{(m)}(t)$ that corresponds to a linear amplification of $s^{(m)}(t)$, while the remainder is seen as noise. Therefore, without any assumption on the distribution, the output signal can be expressed as [4]

$$s_d^{(m)}(t) = \alpha^{(m)} s^{(m)}(t) + d^{(m)}(t), \quad (5)$$

where $d^{(m)}(t)$ is the NLD and $\alpha^{(m)}$ is the realization-varying complex amplification term (RVCA) that depend on the outcome m of $\mathcal{S}$. The value of $\alpha^{(m)}$ that minimizes the mean square error (MSE) of the unbiased input and output signals is found to be [4]:

$$\alpha^{(m)} = \frac{\langle s_d(t), s(t)\rangle_m - \langle s_d(t)\rangle_m \langle s(t)\rangle_m^*}{\langle |s(t)|^2\rangle_m - \langle s(t)\rangle_m \langle s(t)\rangle_m^*} \quad (6)$$

where $\langle x(t), y(t)\rangle_m = \frac{1}{T_s}\int_0^{T_s} x^{(m)}(t)(y^{(m)}(t))^* dt$ is the inner product of $x^{(m)}(t)$ and $y^{(m)}(t)$, and $\langle x(t)\rangle_m = \frac{1}{T_s}\int_0^{T_s} x^{(m)}(t)dt$ is time average of $x^{(m)}(t)$.

Conventional receivers typically neglect RVCA, instead it is assumed to be constant. However, it can be compensated by means of advanced detection that will be shown in remainder of this paper. For further details of RVCA compensation the reader is advised to refer to [4] and references therein .

B. Coded-aided iterative receiver for nonlinearly distorted SC-FDMA tranmission

In the receiver, the signal undergoing multipath propagation expressed in the the frequency-domain at the output of OFDM demodulator is given by

$$R_k = H_k\left(\alpha^{(\mathbf{a})} S_k + D_k^{(\mathbf{a})}\right) + W_k, \quad k = 0, \ldots, M-1, \quad (7)$$

where H_k is the frequency response of the channel at the k-th frequency position, W_k is the Gaussian noise component and D_k is the nonlinear distortion term (NLD) in the frequency domain. The superscript $(\mathbf{a})$ in $D_k^{(\mathbf{a})}$ and $\alpha^{(\mathbf{a})}$ is used to stress that NLD and RVCA are function of the symbol vector $\mathbf{a} = [a_0, \ldots, a_{N-1}]$.

Let $\mathbf{H} = [H_0, \ldots, H_{N-1}]$ be the in-band channel frequency response, then the amplitude and phase distortion introduced in received vector $\mathbf{R}^{(\mathbf{H})}$ due to the multipath channel is compensated by the Minimum Mean-Square Error (MMSE) frequency domain equalizer (FDE). The FDE complex coefficient vector, $\mathbf{C}$ can be obtained under the MMSE criterion. It is given by [5]

$$\mathbf{C} = \frac{\mathbf{H}^*}{|\mathbf{H}|^2 + \frac{\sigma_n^2}{\sigma_s^2}} \quad (8)$$

where σ_n^2 the variance of the additive noise, and σ_s^2 is the variance of transmitted pilot symbol. However, the IDFT despreading block in the receiver averages the noise over each subcarrier and particular subcarrier may experience deep fading in a frequency selective fading channel. IDFT despreading averages and spreads the deep fading effect, which results subsequently in a unfeasible performance degradation. The choice of the MMSE equalizer in this study is motivated by the fact that it reduces the amplitude of the errors and prevents error propagation during the iterative process, thereby improving overall performance.

Assuming that the first assigned subcarrier is at position $k_0 = 0$, the decision variables at the input of the demodulation stage are computed by taking the size-N IDFT of $\mathbf{R} = [R_0, \ldots, R_{N-1}]$. Let $\mathbf{D} = [D_0, \ldots, D_{N-1}]$ and $\mathbf{W} = [W_0, \ldots, W_{N-1}]$ be the in-band NLD and the in-band Gaussian noise component in the frequency domain. Then, using vector notation the contribution of the additive white Gaussian noise (AWGN) to the decision variables can be expressed as $\mathbf{w} = \mathrm{IDFT}_N(\mathbf{W})$. The contribution of NLD, which depends on both the transmitted symbol vector $\mathbf{a}$ and the channel response $\mathbf{H}$ can be expressed as $\mathbf{d}^{(\mathbf{H},\mathbf{a})} = \mathrm{IDFT}_N(\mathbf{H} \odot \mathbf{D}^{(\mathbf{a})})$, where $\odot$ denotes the Hadamard product (element-wise multiplication). Finally, the decision variables can be written as

$$\mathbf{b} = \alpha^{(\mathbf{a})} \cdot \mathbf{a}^{(\mathbf{H})} + \mathbf{d}^{(\mathbf{H},\mathbf{a})} + \mathbf{w}, \quad (9)$$

where $\mathbf{a}^{(\mathbf{H})} = \mathrm{IDFT}_N(\mathbf{H} \odot \mathrm{DFT}_N(\mathbf{a}))$ is the received symbol vector assuming that the symbol vector $\mathbf{a}$ was transmitted over a multipath fading channel with frequency response $\mathbf{H}$.

Since $\mathbf{w}$ is AWGN the maximum likelihood (ML) sequence detector is

$$\hat{\mathbf{a}} = \arg\min_{\forall \check{\mathbf{a}}} ||\mathbf{b} - (\alpha^{(\check{\mathbf{a}})} \cdot \check{\mathbf{a}}^{(\mathbf{H})} + \mathbf{d}^{(\mathbf{H},\check{\mathbf{a}})})||^2, \quad (10)$$

where $\check{\mathbf{a}}$ is any possible transmitted symbol vector. Obviously, solving (7) is too complex and, therefore, it becomes necessary to find solution with reduced complexity. The sub-optimal solution for uncoded case has been reported in [7],[8] recently. Here we present the reformulated approach intended to be used in coded SC-FDMA based transmission systems performing over highly nonlinear distorted environment.

In practice the receiver does not know $\mathbf{d}^{(\mathbf{H},\mathbf{a})}$and $\alpha^{(\mathbf{a})}$. However, provided it knows the transmit nonlinear function $\mathbf{f}(\cdot)$, they can be estimated from the received symbol vector $\mathbf{b}$ in order to enhace BER performance. This process can be done iteratively by using the following procedure:

1) Compute the received noisy symbols vector

$$\hat{\mathbf{a}}^{(q)} = \left\langle \mathrm{IDFT}_N\left(\frac{\mathbf{R}}{\hat{\alpha}^{(\hat{\mathbf{a}}^{(q-1)})} \cdot \mathbf{C}^{-1}} - \frac{\hat{\mathbf{D}}^{(\hat{\mathbf{a}}^{(q-1)})}}{\hat{\alpha}^{(\hat{\mathbf{a}}^{(q-1)})}}\right)\right\rangle \quad (11)$$

where q is the iteration number.

2) The noisy symbols are fed to the symbol decoder. The symbol decoder aims at denoising the QAM symbols

distorted by AWGN as well as by the nonlinear distortion with the great help of the channel decoder. The three steps of the symbol decoder are then (j) a QAM demapper that is performed with hard decision output and thus, low complexity is ensured. (jj) a channel decoder that is represented as a hard output Viterbi algotithm followed by interleaver with the pre-determined size. The output bits of the symbol decoder are denoted as $\hat{\mathbf{b}}^{(q)}$.

3) Afterwards, the output bits $\hat{\mathbf{b}}^{(q)}$ are propagated backwards to the reverse symbol coder that is represented by (j) an interleaver (jj) convolutional coder and finally (jjj) QAM symbol mapper. The output of the symbol coder presents the refined symbol vector observation in step 1 and replaces the symbol vector $\hat{\mathbf{a}}^{(q)}$
4) Assuming that $\hat{\mathbf{a}}^{(q)}$ is the transmitted symbol vector , compute RVCA as

$$\hat{\alpha}^{(\hat{\mathbf{a}}^{(q)})} = \frac{\langle \hat{\mathbf{s}}_d^{(q)}, \hat{\mathbf{s}}^{(q)} \rangle - \langle \hat{\mathbf{s}}_d^{(q)} \rangle \langle \hat{\mathbf{s}}^{(q)} \rangle^*}{\langle |\hat{\mathbf{s}}^{(q)}|^2 \rangle - \langle \hat{\mathbf{s}}^{(q)} \rangle \langle \hat{\mathbf{s}}^{(q)} \rangle^*} \qquad (12)$$

where here $\langle \mathbf{x}, \mathbf{y} \rangle$ and $\langle \mathbf{x} \rangle$ denote inner product and sample average, respectively. $\hat{\mathbf{s}}$ is the length-M time-domain SC-FDMA symbol vector before cyclic prefix insertion and $\hat{\mathbf{s}}_d = \mathbf{f}(\hat{\mathbf{s}})$.

5) Compute the corresponding in-band NLD from $\hat{\mathbf{a}}^{(q)}$ and $\hat{\alpha}^{(\hat{\mathbf{a}}^{(q)})}$ as

$$\hat{\mathbf{D}}^{(\hat{\mathbf{a}}^{(q)})} = \mathrm{DFT}_N \left(\hat{\mathbf{s}}_d^{(q)} - \hat{\alpha}^{(\hat{\mathbf{a}}^{(q)})} \cdot \hat{\mathbf{s}}^{(q)} \right). \qquad (13)$$

Here DFT_N is used to denote that we are only interested in the N samples of NLD that affect to the decision variables and not in the remaining $M - N$ samples. Note however, that the time-domain SC-FDMA symbol is of length M.

6) Go to step 1 and compute $\hat{\mathbf{a}}^{(q+1)}$.

The iterative process is terminated when the performance impovement of the investigated systems is sufficient from the BER point of view.

In the remainder of this paper, we will call the presented scheme as a coded-aided iterative receiver, due to the coding/decoding processes presented in the iterative backward loop that aims to decrease jointly the effects of nonlinear distortion and AWGN, respectively.

IV. Performance evaluation

In this section, performance impovement capabilities of the proposed coded-aided iterative receiver by means of computer simulations are shown. In order to get a better insight to this, the performance of the original iterative receiver introduced in [7],[8] and conventional SC-FDMA receiver are also showed.

The parameters of SC-FDMA transmitter/receiver and propagation channel attributes considered within the presented scenarions are depicted in the Table 1. In order to illustrate largely nonlinearly distored environment, Saleh model of HPA operating at IBO$= 0, 3dB$ and 64-QAM baseband modulation are considered. As the propagation channel, typical 6-tap ITU Pedestrian A channel is used. Note that new channel realizations are considered for each SC-FDMA symbol in order to model a block-stationarity behavior. Fig. 1 and Fig. 2 show BER of SC-FDMA system undergoing strong nonlinear distortion (IBO=$\{0, 3\}$dB) when a conventional receiver, original iterative receiver and coded-aided iterative receiver are employed for the detection. As we can see, conventional receiver fails in both cases and provide very poor results. As the results presented in Fig. 1 indicate, the original iterative receiver improves the performance in comparison with that of conventional receiver, though reaches only the lower bound at BER=10^{-2}. In this scenario, coded-aided receiver approaches the linear performance even after the 1st iteration and therefore the others iteration are not inevitable. In Fig. 2, the SC-FDMA system is less affected by nonlinear distortion (IBO=$3dB$) and original iterative receiver is able to compensate for the nonlinear distortion very efficiently. However, there is still performance gap of $3dB$ at BER=10^{-4} in comparison with the linear system. Even in this scenario, coded-aided iterative receiver provide almost the same performance as the linear system.

Simulation method	Monte Carlo
Number of subcarriers	64
Baseband modulation	64QAM
Convolutional code	code rate $R = 1/3$
HPA model	Saleh model with IBO = $\{0,3\}$dB
Propagation channel	ITU Pedestrian A
Cyclic prefix length	20% of the SC-FDMA symbol length
Equalization	MMSE
Channel Estimation	Perfect

TABLE I
Simulation parameters

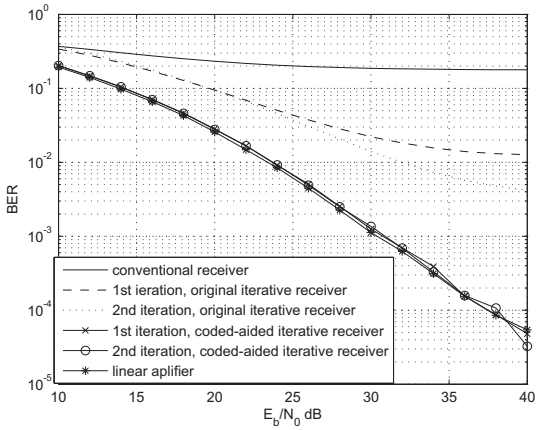


Fig. 1. BER vs. E_b/N_0 of the proposed method, IBO=0dB

V. Conclusion

In this paper, the new coded-aided iterative scheme for the nonlinearly distorted SC-FDMA based transmission system is presented. The presented scheme combines the iterative receiver as proposed in [7],[8] with the symbol encoder/decoder located in the backward loop, in order to provide the better estimation of nonlinear distortion and thereby improving overall performance. By means of computer simulations, the reasonable performance gain provided by the coded-aided iterative scheme has been proved, however it should be noted, that the performance gain is reached at the cost of the higher receiver structure complexity. The application of presented scheme can find its meaning especially in highly nonlinearly distorted LTE uplink transmission without any transmitter modification.

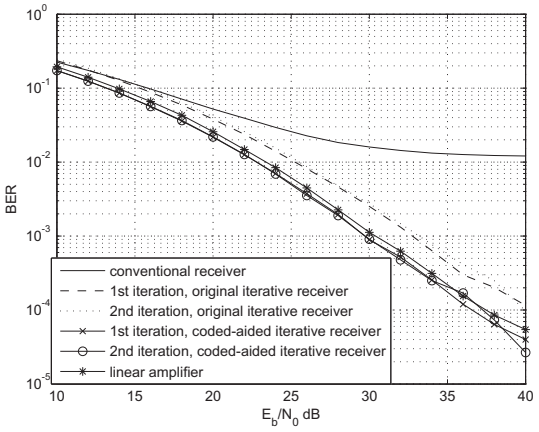


Fig. 2. BER vs. E_b/N_0 of the proposed method, IBO=3dB

ACKNOWLEDGEMENTS

This work is the result of the project implementation Center of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

[1] DAHLMAN, E., PARKVALL, S. 3G Evolution. *HSPA and LTE for Mobile Broadband* . Academic Press, Oxford UK, 2007

[2] TELLADO, J, HOO, L., CIOFFI, J.M. Maximum-Likelihood Detection of Nonlinearly Distorted Multicarrier Symbols by Iterative Decoding. *IEEE Trans. Commun*, 2003, vol. 51, p. 218-228

[3] COLAS, M., GELLE, G., DECLERCQ, D. Turbo Decision Aided Receivers for Clipping Noise Mitigation in Coded OFDM, *EURASIP Journal on Wireless Communications and Networking*. 2008, vol. 2008, pp. 1-10

[4] DEUMAL, M., Multicarrier communication systems with low sensitivity to nonlinear amplification. PhD dissertation thesis, *Enginyeria i Arquitectura La Salle, Universitat Ramon Llull*, Barcelona, 2008

[5] PRIYANTO, B., CODINA, H., et.al. Initial Performance Evaluation of DFT-Spread OFDM Based SC-FDMA for UTRA LTE Uplink. *IEEE Vehicular Technology Conference - Spring*, 2007, pp. 3175-3179

[6] SUZUKI, S., TAKYU, O., UMEDA, Y. Performance Evaluation of Effect of Nonlinear Distortion in SC-FDMA System. *International Symposium on Information Theory and its Applications*, Auckland(Australia), 2008, pp.1-5

[7] DEUMAL, M., GAZDA, J., et.al. Iterative detection of SC-FDMA signals undergoing nonlinear amplification. *EURASIP Journal on Wireless Communications and Networking*, submitted for publication, December, 2009, also available on www.jurajgazda.com

[8] GAZDA, J., DROTAR, et.al. : Simple Iterative Cancellation of Nonlinear Distortion in LDFMA systems. *14th International OFDM Workshop*, Hamburg, Germany, 2009, pp. 133-137

Current fluctuations due to Brownian motion of magnetic domain walls

[1]*Lukáš GLOD*, [2]*Vladimír LISÝ*

[1]Dept. of Mathematics and Physics, Institute of Humanitarian and Technological Sciences,
The University of Security Management, Košice, Slovak Republic
[2]Dept. of Physics, FEI TU of Košice, Slovak Republic

[1]lukas.glod@vsbm.sk, [2]vladimir.lisy@tuke.sk

***Abstract*—The electric current induced by a thermally fluctuating magnetic domain wall is studied. The domain wall motion is described coming from the analogy with the damped Brownian oscillator and is considered as a particle characterized by the mass, position and velocity. The use of an effective method taken from the statistical physics allowed us to convert the linearized stochastic equations for the domain wall into an ordinary differential equation. From the solution of this equation the spectral density of the induced current has been calculated.**

***Keywords*—Magnetic domain wall, motion in wires, fluctuations, Brownian motion.**

I. INTRODUCTION

Spatially localized changes of the magnetization configuration in magnets, the magnetic domain walls (DWs), are of great interest both from the fundamental point of view and with regard to applications, e.g., in the development of high-density magnetic storages such as hard disks [1]. Long ago [2] it has been proposed to describe the behavior of DW as if it would be a mass object that can be characterized by the position and velocity. The mass of a single DW in a ferromagnetic nanowire (of the order of 10^{-23} kg) has been recently determined by detection of its resonance motion induced by an oscillating current [3]. If such a "particle" moves along the wire, its motion is hindered due to the friction. It is also subjected to thermal fluctuations that, according to the fluctuation-dissipation theorem, are coupled to the dissipation [4]. Consequently, the DW behavior can be considered as a motion of a Brownian particle that has been a subject of innumerable investigations [5]. Thus a number of effective methods developed in the studies of Brownian motion of particles can be directly applied to DWs.

The motion of DW can be caused by an external field but also by the always present thermal fluctuations. Due to the time dependent change of the magnetization this motion induces a noisy electric current. The colored noise produced by these fluctuations carries information on the properties of the system. To have the possibility to extract this information from the observed spectrum, one should dispose with the theory that appropriately describes the DW fluctuations. Recently, an attempt to build such a theory was done in the work [6]. Based on that work, in our contribution we calculate the spectral density of the current induced by a rigid DW thermally fluctuating in one dimension. The DW is considered as a Brownian particle with nonzero mass. The particle moves in a harmonic potential due to an extrinsic pinning and experiences a friction force proportional to the particle velocity, an analog of the Stokes force for real particles. An efficient method from the theory of Brownian motion is applied to calculate the mean square displacement (MSD) of the DW. From the MSD, all the time correlation functions, which are necessary to obtain the correlation function for the current can be evaluated. The spectral density of the current is found within the theory of stationary random processes. The obtained spectrum significantly differs from the previous results in the literature [6].

II. DOMAIN WALL FLUCTUATIONS AND THE INDUCED CURRENT

The current induced by the DW motion is given by the expression [6]

$$I(t) \sim \dot{\phi}(t) - \frac{\beta}{\lambda}\dot{x}(t), \tag{1}$$

where ϕ, x and λ are the DW chirality, position and width, respectively, and β is a phenomenological parameter (the sum of the spin transfer torque parameter and non-adiabatic contributions). For the derivation of this expression and the proportionality constant C in (1) see [7, 8]. Up to the linear order in time derivatives C depends on the conductivities of the majority and minority electrons, the length and cross-sectional area of the sample. In (1), the quantities x and ϕ are of stochastic nature. They obey the system of equations for the DW in a parabolic potential. The linearized ($\phi \to 0$) form of these equations is

$$\frac{\dot{x}}{\lambda} = \alpha\dot{\phi} + 2\frac{K}{\hbar}\phi + \sqrt{\frac{D}{2}}\eta_1(t), \tag{2}$$

$$\dot{\phi} = -\alpha\frac{\dot{x}}{\lambda} - 2\omega_{\text{pin}}\frac{x}{\lambda} + \sqrt{\frac{D}{2}}\eta_2(t). \tag{3}$$

In the case of a field-driven DW (not considered in the present work) the right-hand side of (3) contains also a term proportional to the magnetic field. Here α denotes the Gilbert damping, and K is the transverse-magnetic-anisotropy energy.

It is assumed that the pinning force F_{pin} that accounts for the irregularities in the material depends only on the DW position and is described by the potential that is quadratic in x, so that $F_{pin} = -2\omega_{pin}x/\lambda$. As a consequence of the fluctuation-dissipation theorem [4] the stochastic forces η_i with zero mean that describe the thermal fluctuations are at different moments of time statistically independent, $\langle \eta_i(t)\eta_j(t')\rangle = \delta_{ij}\delta(t - t')$, and their intensity is determined by the constant $D = 2\alpha k_B T/\hbar N_w$, where k_B is the Boltzmann's constant, T is the temperature, and N_w is the number of spins in the DW [6].

If the random forces are not considered, from (2) and (3) the deterministic equations for the motion of DW can be derived. The equation for its coordinate,

$$\left(1+\alpha^2\right)\frac{\ddot{x}}{\lambda}+2\alpha\left(\omega_{pin}+\frac{K}{\hbar}\right)\frac{\dot{x}}{\lambda}+4\omega_{pin}\frac{K}{\hbar}\frac{x}{\lambda}=0\,, \tag{4}$$

corresponds to the equation of motion for a noisy harmonic oscillator,

$$m\ddot{x}+\gamma\dot{x}+m\omega_0^2 x = f(t)\,, \tag{4a}$$

where the influence of the thermal fluctuations has been already accounted for by the Langevin force $f(t)$ [9]. The mass of the DW is $m = \hbar^2 N_w/K\lambda^2$ [3].

According to the Wiener-Khinchin theorem, the spectrum (more exactly, the spectral density) of the fluctuating current is determined from the time correlation function $\langle I(t)I(t')\rangle$, as its Fourier transform [4, 10]. If (1) is inserted in this correlator with the use of $\dot{\phi}(t)$ from (3), one has to average the products $x(t)x(t')$, $\dot{x}(t)\dot{x}(t')$, $\dot{x}(t)x(t')$, $x(t)\dot{x}(t')$. Besides them, there will be products of the quantities x and $\dot{x}$ with the random force η and the term ~ $\eta(t)\eta(t')$, which, after the averaging, yields the delta function $\delta(t - t')$.The former products can be omitted since, due to very different scales characterizing the changes of $x(t)$, $\dot{x}(t)$ and $\eta(t)$ in the time, their statistical averages can be put to zero. We thus have to average

$$I(t)I(t')\sim\left(\frac{\alpha+\beta}{\lambda}\right)^2\dot{x}(t)\dot{x}(t')+\left(\frac{2\omega_{pin}}{\lambda}\right)^2 x(t)x(t')$$
$$+2\omega_{pin}\frac{\alpha+\beta}{\lambda^2}\left[\dot{x}(t)x(t')+x(t)\dot{x}(t')\right]+\frac{D}{2}\eta_2(t)\eta_2(t')\,. \tag{5}$$

To do this, one can use the solution of the corresponding Fokker-Planck equation for the probability density of the quantities x and $\dot{x}$ [11]. Here we apply the following efficient method. Let us multiply equation (4a) (with the Langevin force on the right hand side) by $x(t)$, then use the identity $x\ddot{x} = d(x\dot{x})/dt - \dot{x}^2$, the statistical independence of x, $\dot{x}$ and f, and the equipartition theorem $m\langle \dot{x}^2\rangle = k_B T$. Equation (4a) can be then rewritten as

$$m\frac{d^2}{dt^2}\left\langle x^2\right\rangle+\gamma\frac{d}{dt}\left\langle x^2\right\rangle+2m\omega_0^2\left\langle x^2\right\rangle=2k_BT\,. \tag{4b}$$

If we now subtract the equation that follows from (4a) for the correlator $\langle x(t)x(0)\rangle$, we obtain

$$m\ddot{X}(t)+\gamma\dot{X}(t)+m\omega_0^2X(t)+m\omega_0^2\left(\left\langle x^2(t)\right\rangle-\left\langle x^2(0)\right\rangle\right)=2k_BT\,, \tag{4c}$$

where $X(t) = \langle\Delta x^2(t)\rangle = \langle[x(t) - x(0)]^2\rangle$ is the MSD of the oscillator. For stationary random processes the last term on the left hand side of (4c) is equal to zero. We thus come to the Vladimirsky's rule formulated long ago for more general conditions, including non-Markovian processes (the only restriction is the linearity of the system) [12]. The evident initial conditions for the ordinary differential equation (4c) are $X(0) = \dot{X}(0) = 0$. The solution of the final equation

$$m\ddot{X}+\gamma\dot{X}+m\omega_0^2X=2k_BT \tag{6}$$

for the conditions of the experiment [3] (when $1-(2m\omega_0/\gamma)^2 < 0$), is

$$X(t)=\frac{2k_BT}{m\omega_0^2}\left[1-\exp\left(-\frac{\gamma}{2m}t\right)\left(\cos\omega_1 t+\frac{\gamma}{2m\omega_1}\sin\omega_1 t\right)\right], \tag{7}$$

where $t > 0$, $\omega_1 = \left[\omega_0^2-(\gamma/2m)^2\right]^{1/2}$.

III. Spectral Density of the Induced Current

For what follows we need only the autocorrelation function for the velocity,

$$\left\langle v(t)v(0)\right\rangle=\frac{1}{2}\ddot{X}(t)\,. \tag{8}$$

If we define the Fourier transformation of $x(t)$ as

$$x_\omega=(2\pi)^{-1}\int_{-\infty}^{\infty}x(t)\exp(i\omega t)dt\,, \tag{9}$$

the spectral density of the quantity $\langle x^2\rangle$ for the stationary random process is [4]

$$(x^2)_\omega=(2\pi)^{-1}\int_{-\infty}^{\infty}\left\langle x(t)x(0)\right\rangle\exp(i\omega t)dt\,. \tag{10}$$

The spectral density of the current can be easily found using this relation (x is replaced by $I(t)$) without calculating the correlation functions (combinations of the velocity and position of DW) in $\langle I(t)I(t')\rangle$. This is because the spectral density of the current, $(I^2)_\omega$, can be expressed through the densities $(v^2)_\omega$, $(x^2)_\omega = \omega^{-2}(v^2)_\omega$, $(vx)_\omega = -i\omega^{-1}(v^2)_\omega$, and $(xv)_\omega = i\omega^{-1}(v^2)_\omega$, that correspond to the correlation functions $\left\langle\dot{x}(t)\dot{x}(0)\right\rangle$, $\left\langle x(t)x(0)\right\rangle$, $\left\langle\dot{x}(t)x(0)\right\rangle$, and $\left\langle x(t)\dot{x}(0)\right\rangle$, respectively. Then the required $(I^2)_\omega$ will be

$$\left(I^2\right)_\omega\sim\lambda^{-2}\left[(\alpha+\beta)^2+\left(\frac{2\omega_{pin}}{\omega}\right)^2\right]\left(v^2\right)_\omega+\frac{D}{4\pi}\,. \tag{11}$$

The spectrum $(v^2)_\omega$

$$(v^2)_\omega=\pi^{-1}\int_0^{\infty}\left\langle\dot{x}(t)\dot{x}(0)\right\rangle\cos(\omega t)dt \tag{12}$$

for the even extension of the VAF to $t < 0$ is found from (7) and (8) in the following compact form,

$$(v^2)_\omega = \frac{k_B T}{\pi m} \frac{\omega_0^2 - \omega^2}{\left(\omega_0^2 - \omega^2\right)^2 + \left(\gamma\omega/m\right)^2}. \quad (13)$$

After its substitution in (11) we directly obtain the spectrum of the current induced by the moving DW. This spectrum does not depend on whether the DW oscillates in the harmonic potential or its MSD only monotonically approaches the value $2k_BT/(m\omega_0^2)$. It is important that at the frequency ω_0 the spectral density crosses the constant value $D/4\pi \sim T$, which could be examined experimentally. Note that the obtained result significantly differs from that by [6], found by a different method. As distinct from the approach used in [6], our method is much simpler and allows for easy proving.

Finally, let us express the quantities entering the velocity spectrum (13) through the DW parameters. The parameters are obtained by comparison of the linearized equation for the DW position (4) and the equation for the harmonic oscillator (4a). The mass of the DW is given above after (4a), and the friction constant and the oscillator frequency are

$$\frac{\gamma}{m} = \frac{2\alpha}{1+\alpha^2}\left(\omega_{\text{pin}} + \frac{K}{\hbar}\right), \quad \omega_0^2 = \frac{4\omega_{\text{pin}}}{1+\alpha^2}\frac{K}{\hbar}. \quad (14)$$

The numerical values of the parameters can be determined from the cited articles: $\gamma/m = 1/\tau = 0.714\times10^8$ s^{-1} ($\tau = 1.4\times10^{-8}$ s is the relaxational time of the DW), $\lambda = 7\times10^{-8}$ m, $K = 1.8\times10^{-24}$ J, and $\omega_0 = 1.57\times10^8$ s^{-1}. As to the parameters α and β, in the previous calculations [6] $\alpha \sim 0.01$ - 0.1, and $\beta/\alpha = 0$ - 2 were used as the values for typical materials. However, it follows from the experiment [3] that α should be much smaller. Indeed, for small α we have from (14) $\omega_{\text{pin}} \approx \gamma/(2\alpha m) - K/\hbar$. To have the pinning potential positive, α should obey the inequality $\alpha < \gamma\hbar/(2mK)$, which is the value about 0.002 or smaller. We thus can use $\omega_0^2 \approx 4\omega_{\text{pin}}K/\hbar$. Then $\omega_{\text{pin}} \approx 0.36\times10^6$ s^{-1}, which corresponds to $\approx 0.2\times10^{-4}\,K/\hbar$ (the value lower than in [6]). In the calculations we take $\alpha \approx 0.002$. The numerical results for the spectral density of induced current are presented in Figs. 1 - 4. The spectrum is normalized to $C^2k_BT(\alpha+\beta)^2/(\pi m\lambda^2)$ and shifted vertically down by a constant $D/4\pi$. Figures 1 - 3 demonstrate the spectra for three values of β, $\beta = 0$, α, 2α (as in [6]) and nonzero pinning potential. Figure 4 shows the spectrum at $\omega_{\text{pin}} = 0$, when it is independent on β. In all the cases the dependence of $(I^2)_\omega$ on ω is very different from the calculations in [6].

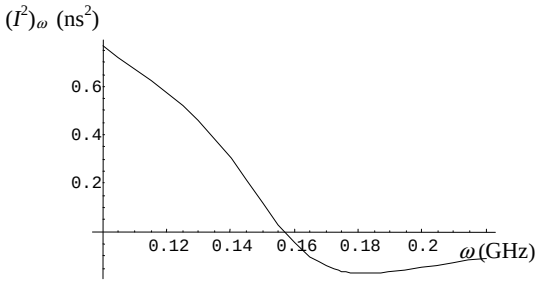


Fig. 1. $(I^2)_\omega$ dependence on frequency ω at $\beta = 0$. The graph is normalized to $10^9\,C^2k_BT(\alpha+\beta)^2/(\pi m\lambda^2)$ and shifted vertically down by a constant $D/4\pi$, see the text.

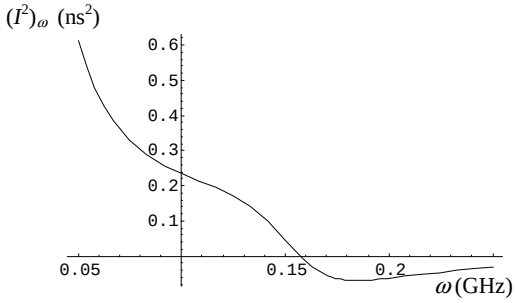


Fig. 2. The same as in Fig. 1 for $\beta = \alpha$.

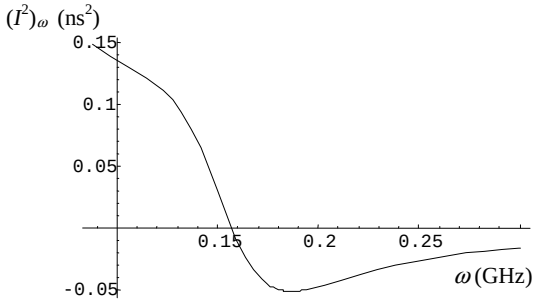


Fig. 3. The same as in Fig. 1 for $\beta = 2\alpha$.

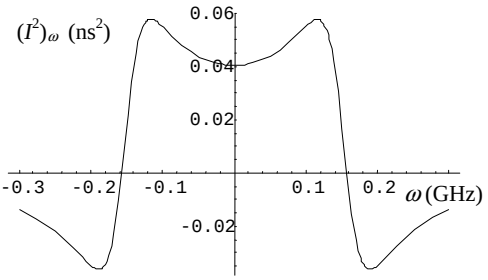


Fig. 4. The same as in Fig. 1 but without extrinsic pinning. The spectral density does not depend on β.

IV. Conclusion

In the present work a particular problem of finding the spectral density of the current induced by a fluctuating magnetic domain wall has been solved. The method of solution was borrowed from the theory of the Brownian motion of particles. We believe that the efficiency of the used approach that could be useful in a number of related problems of the dynamics of domain walls was clearly demonstrated. Moreover, we have obtained new results on the mean square displacement and other time correlation functions, that are easy to verify theoretically and that could be examined experimentally. The calculated spectrum of the fluctuating current significantly differs from the previous results from the literature. It should be however noted that the presented work is restricted to the linear regime of the domain wall dynamics (small chirality). In general, the studied problem is nonlinear and represents a serious challenge. In this case the used method becomes inapplicable. One possibility to make a progress in this class of problems is based on the Fokker-Planck equation for the probability density of the variables describing the domain wall motion. Currently, our attempts are oriented in this direction.

ACKNOWLEDGMENT

This work was supported by the Agency for the Structural Funds of the EU within the project NFP 26220120021, and by the grant VEGA 1/0300/09.

REFERENCES

[1] G. Tatara, H. Kohno, J. Shibata, "Microscopic approach to current-driven domain wall dynamics", *arXiv*:0807.2894v2 [cond-mat.mes-hall] (2008) - an upgraded version of the paper published in *Phys. Rep*. 468, 213-301 (2008).

[2] V. W. Döring, "Über die Trägheit der Wände zwischen Weisschen Bezirken", *Z. Naturforsch*. 3a, 373–379 (1948).

[3] E. Saitoh, H. Miyajima, T. Yamaoka, G. Tatara, "Current-induced resonance and mass determination of a single magnetic domain wall", *Nature* 432, 203-206 (2004).

[4] L. D. Landau, E. M. Lifshitz, *Statistical Physics*, Part I. Moscow: Nauka, 1976 (in Russian); English translation: Oxford e.a.: Reed Educational and Professional Publ., 3rd. Ed., 2000.

[5] R. M. Mazo, *Brownian Motion. Fluctuations, Dynamics, and Applications*. New York: Oxford Univ. Press, 2009.

[6] M. E. Lucassen, R. A. Duine, "Spin motive forces and current fluctuations due to Brownian motion of domain walls". arXiv: 0910.063v1 [cond-mat.mes-hall] (2009), submitted to "Special issue: Caloritronics" in *Solid State Communications*.

[7] R. A. Duine, "Spin pumping by a field-driven domain wall", arXiv:0706.3160v3 [cond-mat.mes-hall]; *Phys. Rev. B* 77, 01440 (2008).

[8] R. A. Duine, "Effects of non-adiabaticity on the voltage generated by a moving domain wall", *arXiv*:0809.2201v1 [cond-mat.mes.-hall]; *Phys. Rev. B* 79, 14407 (2009).

[9] W. T. Coffey, Yu. P. Kalmykov, J. T. Waldron, *The Langevin Equation. With Applications to Stochastic Problems in Physics, Chemistry and Electrical Engineering*. New Jersey e.a.: World Scientific, 2nd Ed., 2005.

[10] N. G. van Kampen, "Fluctuations in nonlinear systems". Chapter 5 of *Fluctuation Phenomena in Solids*. Edited by R. E. Burgess. New York: Academic Press, 1965, pp. 139–177.

[11] H. Risken, *The Fokker-Planck Equation. Methods of Solution and Applications*. Berlin e.a.: Springer-Verlag, 2nd Ed. 1989.

[12] V. V. Vladimirsky, "To the question of the evaluation of mean products of two quantities, related to different moments of time, in statistical mechanics". *Zhur. Ekper. Teor. Fiz*. 12, 199-202 (1942) (in Russian).].

Utilization of elastomagnetic effect in pressure force measurement.

Ing. Anna Hodulíková, Ing. Martin Bačko

Dept. of Theroretical Electrotechnics and Electrical Measurement, FEI TU of Košice, Slovak Republic

anna.hodulikova@tuke.sk, martin.backo@tuke.sk

Abstract—The paper describes utilization of elastomagnetic effect for the measurement of force and presents partial results of the institutional research solved at the Department of Theoretical Electrotechnics and Electrical Measurement at the TU Košice.

***Keywords*—sensor, elastomagnetic, Villari effect, force measurements.**

I. INTRODUCTION

Sensor consists of sensitive input part labelled as sensor and output part labeled as converter which transforms change of sensor´s inner state to electrical quantity. Sensors are the most important part of measurement chain.

Discovery of sensors using the elastomagnetic phenomenom revolutionized the field of force sensors, pressure sensors or torque sensors used mainly in difficult working conditions.

The most important criterion when designing such sensor is its utilization in industry – measurement of big pressure powers, protection against mechanical overload, measurement of rolling powers, measurement of forces in bridge consoles. Development of electronics and discovery of new magnetic materials led to intelligent sensors.

II. ELASTOMAGNETIC SENSOR

A. *Elastomagnetic effect*

Elastomagnetic effect was observed by Villari in year 1865. In ferromagnetic material, the effect is: when force affects the material, nucleus is deformed. Because of this deformation, mutual distances of atoms in crystal lattice change, which results in change of forces which cause spontaneous magnetization in individual domains of ferromagnetic material. This fact presents itself like change of magnetic polarisation (change of magnetic induction when intensity of magnetic field which affects ferromagnetic material is not changed), so the magnetic properties are subject of change. If the material was isotropic before the force affected it, then it becomes anisotropic. If the material was anisotropic, then its anisotropic properties will change. Magnetic properties are represented by permeability, which will change due to the force which affects the material. Change of permeability $\Delta\mu$ because of mechanical tension σ is described by following equation which is derived from thermodynamic equilibrium in ferromagnetic material:

$$\Delta\mu = \frac{2\lambda_{ms}\mu^2}{B_{sef}^2}.\sigma = k_M.\sigma \qquad (1)$$

where λ_{ms} is expected value of magnetostriction coefficient when fed, μ is permeability of ferromagnetic material if the affecting force is equal to zero, B_{sef} is effective value when filled, k_M is the material coefficient. Equation (1) is valid in the case that the directions of mechanical tension and magnetic field are collinear. We can see from equation (1), that it is more convenient to use materials, which have high permeability μ, high value of magnetostriction coefficient when fed λ_s and simultanneously low value of magnetic induction when fed B_s.

B. *Elastomagnetic sensor of pressure force*

First elastomagnetic sensor (EMS) of pressure force was designed and manufactured in year 1954 by swedish company ASEA (ABB nowadays). It consists of ferromagnetic core and vertical primary and secondary winding (fig.1). Elastomagnetic sensors of pressure force designed on Department of Theoretical Electrical Engineering and Electrical Measurement (KTEEM), FEI TU Košice have windings stored in one line oriented vertically to the affecting force. Windings are transformer type. They were manufactured to measure pressure forces from 1,2 kN to 12 MN (fig.2). The biggest focus was on improving the metrological parameters of EMS 120 kN and 6 MN. Operation of elastomagnetic pressure force sensor is based on existence of elastomagnetic phenomenom, which occurs in objects made from ferromagnetic materials when they are affected by mechanical tension caused by external force.

Purpose of this work was the derivation of output voltage of elastomagnetic sensor from the affecting force. The result is the equation (2), which is a form of linearization and attempt for approaching the reality when the input parameters are constant:

$$\Delta U_V = m.n.\frac{2fN_1N_2I_1\lambda_s\mu^2}{B_s^2\pi b}.\ln\frac{b}{a}.F \qquad (2)$$

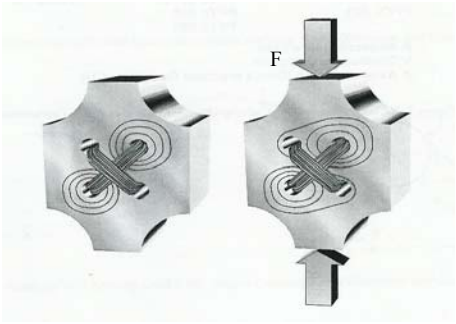
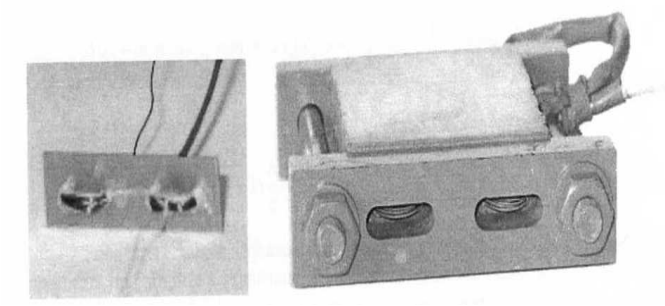
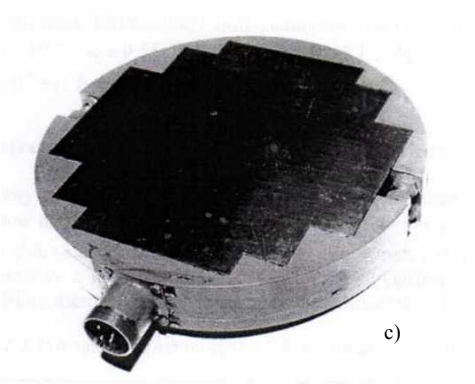


Fig. 1 Elastomagnetic sensor from ABB company



a) b)



c)

Fig. 2 Elastomagnetic sensors manufactured on KTEEM
a) 1.2 kN, b) 120 kN, c) 12 MN

III. ELASTOMAGNETIC SENSOR OF FORCE 120 KN

Practical fabrication of elastomagnetic sensor of pressure force 120 kN, which correspond to 100 MPa pressure, is on figure (fig 2b) and geometric computer model created in SolidWorks is in figures 3 and 4 (Fig.3,Fig.4). The core of sensor is made of 50 lamelles of transformer plate Et 2,6 0,5mm thick. Lamelles are glued and screwed together. The primary (N_1=10, 0.35mm thick copper wire) and secondary winding (N_2 = 8, 0.25-0.3 thick copper wire) are stored in the same direction in parallel way. Circular slots with 1 mm semidiameter used for windings are in one line 12 mm far from each other and their number is even, because of best possible utilization of winding. Optimum working conditions for EMS - 120 kN sensor is the feeding current (effective value) I_{ef} = 0,7 A, frequency 400 Hz and temperature 23°C .

Simplified scheme of sensor´s core is on figure (fig.5) and windings placement in the core slots is on the figure (fig.6)

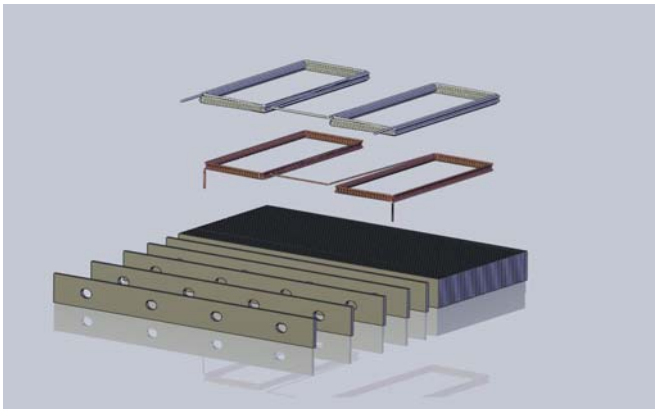


Fig. 3 Design of lamelles and windings

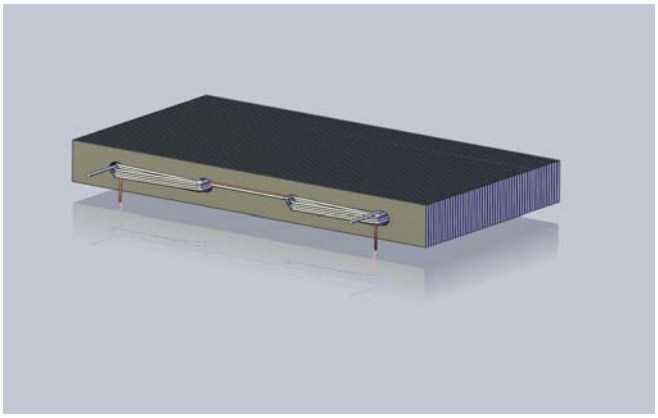


Fig. 4 Design of whole elastomagnetic sensor

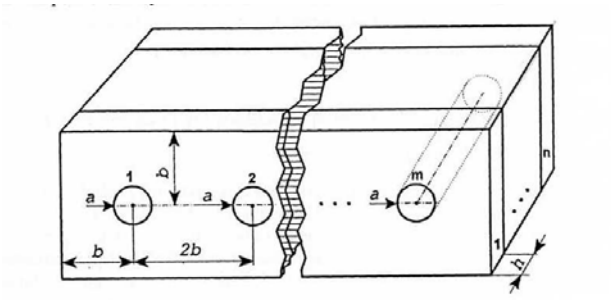


Fig. 5 Simplified scheme of sensor´s core

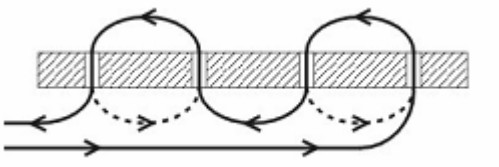


Fig. 6 Placement of windings in the elastomagnetic sensor´s core

Every sensor can be defined as a device used to process information related to some kind of energy. In case of elastomagnetic sensor is pressure the measured quantity and secondary information carrier is the output voltage U_v. Output signal from sensor is related to change of permeability resulting to deformation of magnetic field.

A. *Relationship between input and output signal of elastomagnetic sensor of pressure force*

Input signal for sensor is the external pressure force, output signal is the effective value of voltage, which is induced in secondary winding of sensor and is measured by voltmeter. When defining the voltage on the sensor´s output we use the induce law:

- In case, that the sensor is unaffected by external pressure force, induced voltage $u_V(t)$ will be:

$$u_V(t) = -N_2 \frac{\partial}{\partial t}\left[\int_S \overline{B}(t).d\overline{S}\right], \qquad (3)$$

where N_2 is the number of threads of secondary winding, $\overline{B}$ is the magnetic induction in the sensor´s core unaffected by force in time t, $d\overline{S}$ je normalcy of surface element dS, S is the area of cross-section of sensor´s core.

- In case that sensor is affected by external pressure force F, induced voltage $u_V^F(t)$ will be:

$$u_V^F(t) = -N_2 \frac{\partial}{\partial t}\left[\int_S \overline{B}^F(t).d\overline{S}\right], \qquad (4)$$

where $\overline{B}^F$ is magnetic induction in sensor´s core when affected by force in time t.

From metrological point of view it is suitable, that output signal should be related to input signal. Because of zero force, the output signal will be nonzero, it will be better to determine only the difference in voltage $\Delta u_V(t)$ between voltage when the sensor is loaded $u_{V(p\neq 0)}^F(t)$ and when the sensor is idle $u_{V(p=0)}(t)$. If the magnetostriction coefficient of chosen ferromagnetic material will be positive, output voltage with increasing pressure will begin to fall and $\Delta u_V(t)$ (useful signal) will become negative, therefore we need to adjust the equation:

$$\Delta u_V(t) = \left|u_V^F(t) - u_V(t)\right| =$$
$$= \left|-N_2\left\{\frac{\partial}{\partial t}\left[\int_S \overline{B}^F(t).d\overline{S}\right] - \frac{\partial}{\partial t}\left[\int_S \overline{B}(t).d\overline{S}\right]\right\}\right|, \qquad (5)$$

Because magnetic field in the sensor´s core is created by harmonic current $i_1(t) = I_{m1}\sin\omega t\ [\mathrm{A}]$, where I_{m1} is the maximum value of current $i_1(t)$ and $\omega = 2\pi f$ is the angle frequency of time changes of awaking current. Because of this, magnetic induction and output voltage have timechanging behaviour. The question is, which from the values from timechanging behaviour shall we measure. Practical measurements showed, that it will be effective value of output voltage (6):

$$\Delta U_{Vef} = \sqrt{\frac{1}{T}\int_0^T \Delta u_V^2 dt} \qquad (6)$$

When expressing effective value of sensor´s output power it will be convenient to consider that, that behaviour of sensor´s magnetic induction is harmonic too.

B. *Change of magnetic permeability*

When sensor´s core made from transformer plates is affected by force, then magnetic induction is changing, but the intensity of magnetic field remains unchanged. This leads to change of magnetic properties of sensor´s core. Magnetic properties of specified environment are characterized by magnetic permeability μ, which is function of magnetic induction $\mu = \mu(\overline{B})$.

By measurements and calculations we were able to get the values for graphical representation of dependency μ from size of B when the core is affected, or unaffected by external force at 400 Hz (fig.7, fig.8, fig.9).

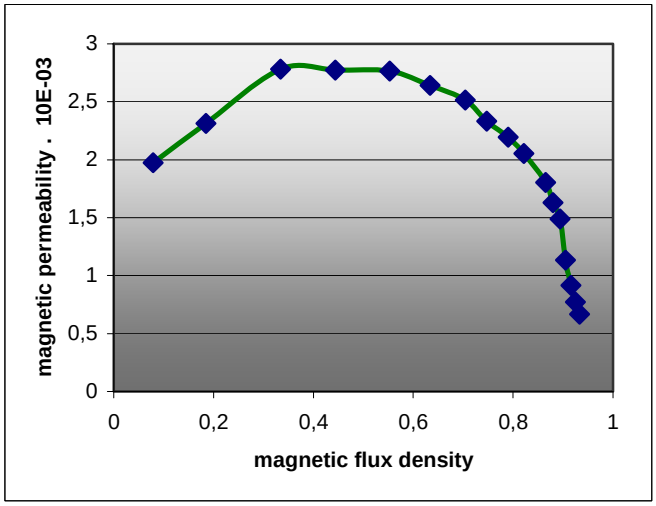


Fig.7. Dependence μ = μ (B), F = 0 kN

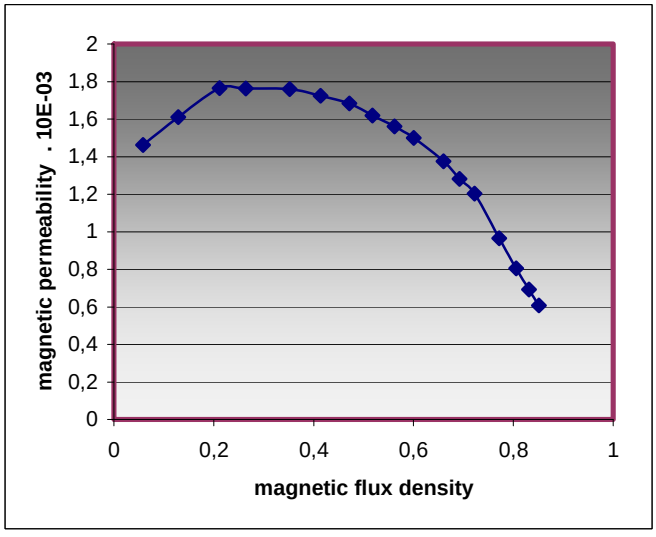


Fig. 8. Dependance μ = μ (B), F = 60 kN

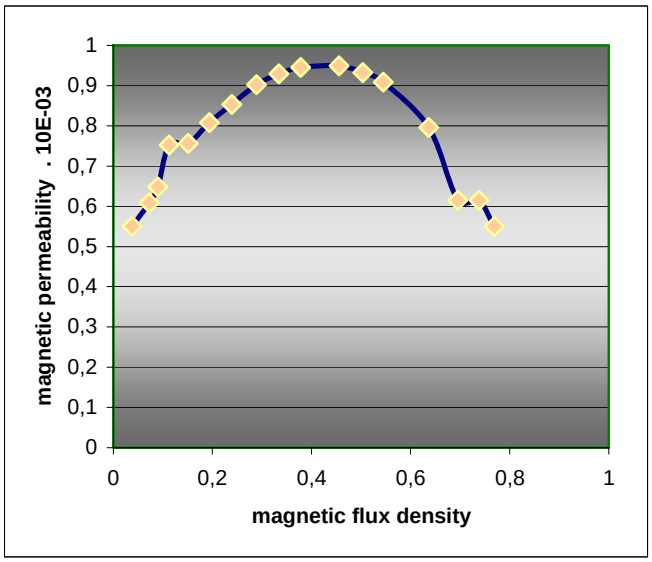


Fig. 9. Dependance $\mu = \mu(B)$, $F = 120$ kN

IV. CONCLUSION

Elastomagnetic sensors of force, when compared with tensometric sensors, have the following advantages: high sensitivity, very high reliability, performance stable in time, resistance against overloading of several orders, mechanical durability. Disadvantages of these sensors are: high inner consumption, hysteresis, non-linearity.

The accuracy of sensors can be improved by higher demands on the quality of their production or by including a circuit into a measurement chain which would modify the output characteristics of sensor.

ACKNOWLEDGMENT

The paper has been prepared by the support of Slovak grant projects VEGA 1/0660/08, KEGA 3/6386/08, KEGA 3/6388/08

REFERENCES

[1] M. Mojžiš, A Contribution to the Solution of Problems of Pressure Measurement Based on the Principle of Elastomagnetic Phenomen, PhD Thesis, Košice, 1979.

[2] I. Tomčíková, D. Špaldonová, Elastomagnetic Sensor Field Determinati Using Matlab, Acta Electrotechnica et Informatica,Vol.. 7,No. 3, 2007, pp. 74-77, ISSN 1335-8243.

[3] I. Tomčíková , Determination of Magnetic And Stress Field Interaction in Elastomagnetic Pressure Force Sensor, X International PhD Workshop OWD, Poland, 2008.

[4] M. Mojžiš, Metrological properties of elastomagneticsensor , IEEE, Vol. 48, 1996, No. 1-2, pp. 46-49.

[5] M. Mojžiš, et al, Pressure Force Sensor. In: proceedings of the II. Internal scientific conference, TU FEI Košice, 2001

[6] J. Vojtko, et al., Utilization of Elastomagnetic effect in Force and torque measurements. Proceedings of the 2nd Conference, MWT TU FEI Košice, 2004.

[7] A. Hodulíková, Measurement of Force Using Elastomagnetic Effect, 5th PhD.Student Conference, TU FEI Košice, 2005.

Model of Production Line with Multi-Motor Drive

Matúš HRIC

Dept. of Electrical, Mechatronic and Industrial Engineering, FEI TU of Košice, Slovak Republic

matus.hric@tuke.sk

***Abstract*— Production and finishing industrial lines occurs in the industrial plants dealing with continuous production of materials of various mature that are in form of a web, strip, fiber etc. From control point of view they belong among complex multivariable mechatronic systems where quality of output products depends on quality of the control system. The main variables to be controlled are usually the speed of the processed material and of working machines. The contribution describes a simplified physical model of a production line and its control part. The model is used for teaching and debugging control algorithms of complex systems that are presented by multi-motor drive in this case.**

***Keywords*—physical model, PLC, multi-motor drive**

I. INTRODUCTION

Continuous production and finishing lines create essential technological equipment in manufacturing production. They are distinguished by mechanical coupling of individual driving units in multi-motor drive through a moving web. For teaching, explanation of mutual couplings and interactions in the line, for practical realization of the control algorithms by PLCs and for performing other laboratory experiments it is advantageous to use physical models that enable to understand easier principles of control strategy of such complex drive systems [1].

II. PRODUCTION LINE WITH MULTI-MOTOR DRIVE IN PRAXIS

The term of multi-motor drive is used to describe all main drives in the technological process used for transport of the processed web which mechanically couples the drives [2]. One group of multi-motor drives is presented by drives of continuous production and finishing lines. The lines are used to perform various technological operations on a moving of a metalic or non-metalic substance.

Typical arrangement of the production line is shown in Fig.1. The mechanical energy of traction cylinders for moving the web comes from electrical drives equipped by appropriate control circuits (current-, speed- and position control). In the most cases, the moving web makes elastic mechanical couplings which leads to complexity of the system to be controlled.

A generalized production/finishing line in Fig.1consists of:
a) The input section ensuring constant material flow – having unwinder, drawing (tractive) cylinders, binding machine (welder), input cartridge (to match with discontinuous material inflow caused by coil exchange on the coiler spindle),
b) The technological section, with technological machines for processing of the material and with combination of tractive S-cylinders and working cylinders, material loops, etc.,
c) The output section, represented by output cartridge, shearing machine to cut the material, and finally the winder.

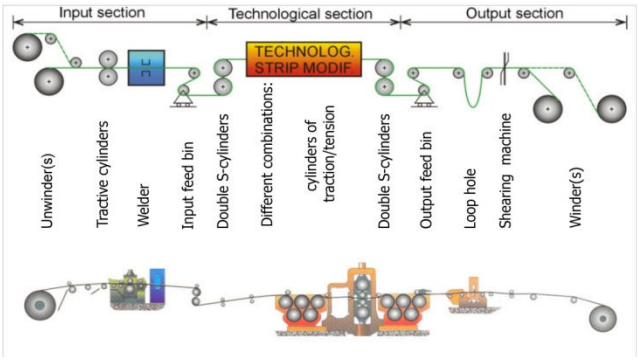
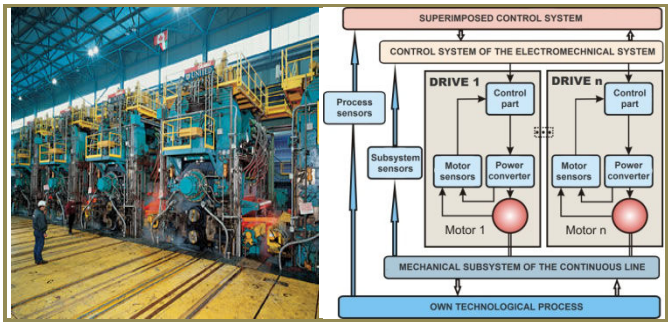


Fig. 1. Typical arrangement of a continuous production line and its parts

The mechanical subsystem of the technological part is driven by a multi-motor drive. Control system of the electromechanical subsystem ensures a correct collaboration of the drives satisfying technological requirements (constant tension in the material independently from its speed, position of the loop, etc.). It also contains some features of artificial intelligence (automatic identification of material properties and of other parameters, automatic tuning - adjusting the controllers parameters to ensure quality of the final product, etc.), [1]. As a supervisor here acts a superimposed control system ensuring contact with a user: data input and collection of technical and economical data from the technological process. From system point of view one recognizes all basic mechatronic subsystems here: the mechanical, electrical and information ones (Fig. 2).



a) b)

Fig. 2. Typical look-out of a production line (a) with multi-motor drive on example of a hot strip mill and its generalized model block diagram (b)

III. Physical Model of a Continuous Line

From the control point of view, the described system of the continuous production line presents a complex MIMO (multi-input, multi-output) system in which calculation and setting the controllers parameters for high control quality is enough difficult and complex task. For teaching and explanation of phenomena occurring in the system, it is highly advantageous to use eLearning instruments [2], [3] where phenomena can be animated and slow down in order to analyze add understand them.

Finally, verification of correct settings of the controller parameters causes a problem - they should be verified on a real system. In industrial environment it is impossible to perform any experiment from technical, safety and economical reasons. Here, the simplified laboratory model of a generalized industrial line (Fig. 3) presents a suitable solution.

The mechanical part of the model consists of unwinder, 3 cylinders and winder: Altogether there are 5 drives to control. The web consists of a film strip from celluloid material. In the loops among the machines there are placed swinging arms that are tensioned by springs enabling to set up a tension in the web. There are also position sensors sensing the angle of the arm. This arrangement enables the operator to check by sight whether the web tension in sections among the neighboring cylinders corresponds to the preset reference value. In the bottom on the left side there is a control panel with the actuating buttons enabling to change modes of the line.

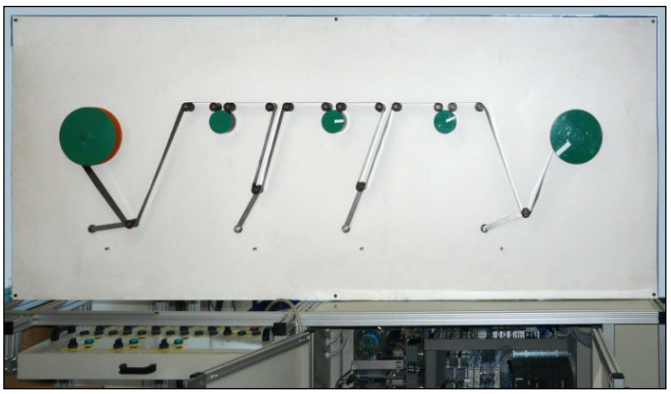


Fig. 3. Physical model of a continuous production line

The driving cylinders in the line are driven by DC motor drives that are supplied by the analogous Allan Bradley converter. For the outer speed controller originally there was used a microprocessor system MS-80 with A/D D/A transducers and network of transputers, where several modern control algorithms were debugged, [4]. Gradually, in line with development of the IT and automation technologies, the systems and its control part have been changed gradually. After modernization only analogous AC/DC converter and the mechanical part remained.

IV. Innovated Control of a Physical Model

Control of electrical drives of a line requires using an underimposed current controller for each drive implemented in each converter module that are connected with an overimposed speed controller.

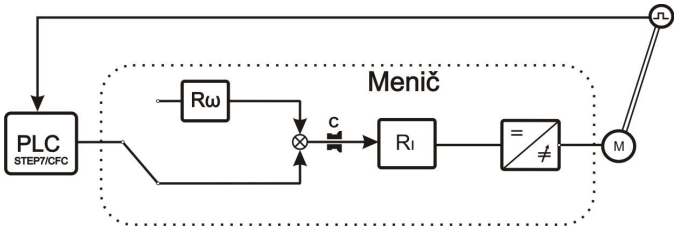


Fig. 4. Principal diagram of drive units control in the production line

During reconstruction of the model we replaced original analogous speed sensors by incremental sensors and from this reason also original analogous control was replaced by digital control where the digital controller is programmed in the PLC. Fig. 5 shows arrangement of the apparatus in the distribution box (power converter, PLC, supply sources for IRC, auxiliary switching relays and breaker apparatuses).

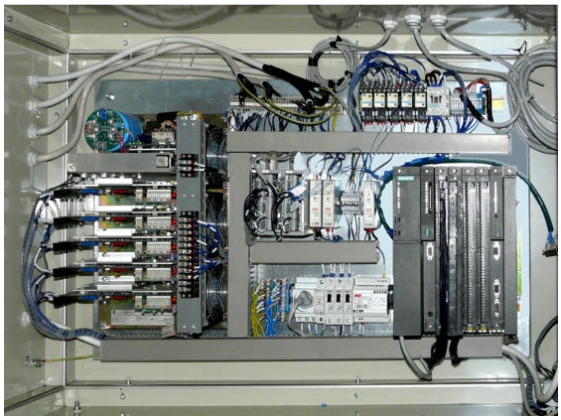


Fig. 5. A view into reconstructed distributing box of the physical model

The superimposed speed controller was realized by the PLC Siemens S7- 400 with the technological card FM 458 that in the programming environment S7 CFC (Continuous Function Chart) from Siemens gives a possibility to interconnect and set up common, or user's function blocks - similarly like it is done in the Matlab-Simulink program. It should be noted that the technological card FM-458 is used frequently in many technological solutions in industry.

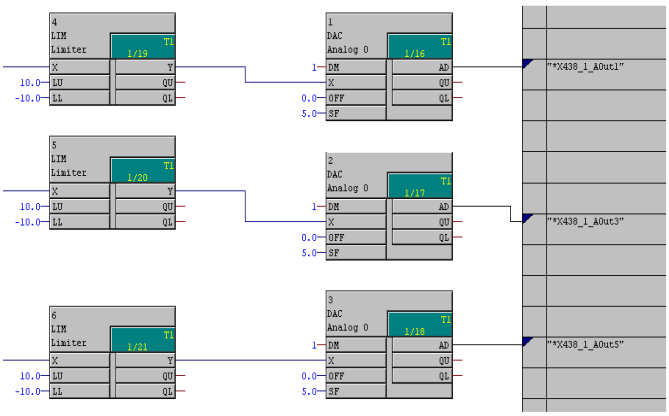


Fig. 6. Output of controlling signals through limiters and D/A transducers and their following addressing and input into a superimposed circuit

Fig. 7 shows scheme of a part of logical commanding that was developed in PLC in the programming language LAD.

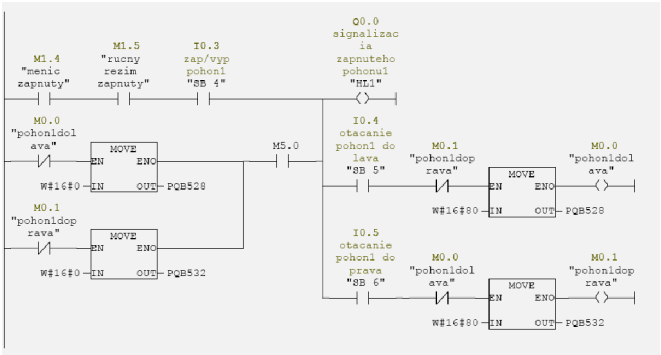


Fig. 7. A Part of logical commanding of drives in the program STEP 7 and programming language LAD

V. Practical Utilization

The physical model is used for laboratory experimentation in several subjects of study, like: Models of Dynamical Systems, Controlled Drives, Mechatronic Production Systems, Control of Assembling Lines by PLCs, Controlling and Visualisating Systems, and Control of Mechatronic Production Systems that are taught at the author's department at Technical University of Kosice.

The students practically verify theoretical knowledge from system dynamics fundamentals, control of electrical drives, design methods for drive controllers and their realization by programming the PLC. They learn deeply the case studies from field of multi/motor drives control and compare obtained time responses from simulation with those got from the laboratory model. Moreover, they have a chance to tune the controllers and follow behavior of the systems at changed system parameters.

A common example of the physical model utilization is in case of teaching subject of Controlled Drives. Fig. 8 shows the speed response to a small reference value step. It shows that the speed controller is tuned properly and its time response corresponds to the simulated time course (on the right side). Fig. 9 shows the speed response and large value of the reference signal without using ARW connection. The students have chance to modify the control program, to change the controller parameters and to observe the system behavior and responses to the changes.

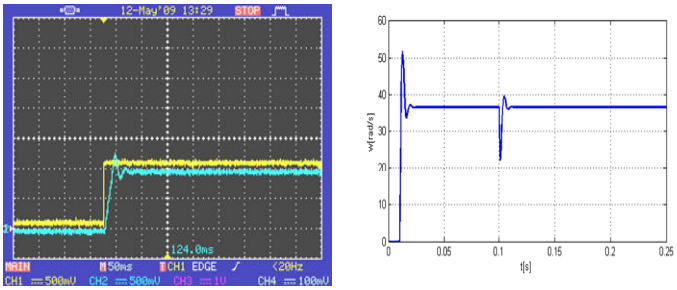


Fig. 8. Speed response and its simulated time course in the linear region at the reference signal step corresponding to 5% of nominal value

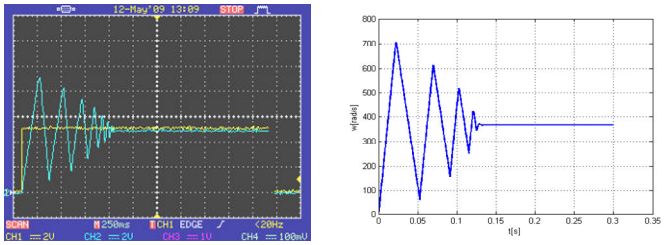


Fig. 9. Speed response and its simulated time course in the linear region at the reference signal step corresponding to 5% of nominal value

VI. Conclusion

The presented physical model of a continuous production line contains all features of mechatronic systems and it offers a unique chance for students to understand control principles of complex drive systems with multi-motor drives. Moreover it offers a safe experimentation on modern control systems presented by PLCs, [5]. At their programming the students performs experiments on the most modern automatic equipments that causes a secondary effect for the graduates from the concerned study program – increasing of their value and competiveness on the labor market.

The original model was recently substantially renovated and we continue in this renovation by exchanging of the gears among the motors and driving cylinders. The previous gears having large backlash will be exchanged for planed gear with the backlash of 2' which will give chance to employ a more precise control and thus to implement more complex algorithms .The physical model also suitable for research at design and verification of new control structures and algorithms for HIL (Hardware-In-the-Loop) simulation that have recently become a standard at verification of mechatronic systems control

References

[1] *BRANDENBURG, G. – WOLFERMANN, W.: State observers for multi-motor drives in processing machines with continuous moving webs," in Proc. Power Electronics and Applications Conf. (EPE'85), Brussels, Belgium, 1985, pp. 3.203–3.210.*

[2] *JEFTENIC, Borislav – BEBIC, Milan – STATKIC, Sasa: Controlled Multi-Motor Drives. In: International Symposium on Power Electronics, Electrical Drives, Automation and Motion, SPEEDAM 2006. 23. - 26.5.2006, pp. 1392 – 1398.*

[3] *FEDÁK, Viliam – FETYKO, Ján – REPIŠČÁK, Martin: Computer Supported Education for Industrial Mechatronics Systems, In: Proc. of Computer Based Learning in Science Int. Conf., CBLIS 2005, Zilina, 2. - 6.7.2005. ISBN 9963-607-63-2, pp. 19-27.*

[4] *FEDOR, Pavol – Perduková, Daniela – Timko, Jaroslav: Study of Controlled Structure Properties with Reference Model. Acta Technica, ČSAV 46, 2001, pp.167-179. ISSN 0001-7043.*

[5] *Virtual Laboratory of Mechatronic Systems Control: http://andromeda.fei.t uke.sk/*, (in Slovak).

Filtration after Band-Pass Sigma Delta modulation

[1]*Marián CHOVANEC,* [2]*Martin SEKERÁK*

[1,2]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]marian.chovanec@tuke.sk, [2]martin.sekerak@tuke.sk

***Abstract*—In this paper, we focus on filtration behind BP SDM. Filtrations as well as converters are important components in electronics. Filters presented on this paper can be used to obtained better properties, especially noise shaping. The SINAD criterion is used to test the quality of filtration. In this paper, we also investigate the decimation as a important part of filtration, which is important component of filtration. These results are a preliminary step to a more complex investigation, where instead of input sinusoidal signal, the AM signal will be used. In such a situation, the filtrations will not only suppress the noise, but also carrier frequency. The results of this study can be used in sensor systems [2].**

***Keywords*—filters, sigma delta, ADC**

I. INTRODUCTION

Converter is very critical component in nowadays electronics systems. Different architectures of radio-frequency (RF) receiver have different specifications on their analog-to-digital-converters (ADCs). Application of filter in signal processing in sigma delta converter sequence isn't trivial task for designer. My observing in filters study devotes to FIR and IIR filters that were engaged in signal processing as shown on Fig. 1. Only the filter on upper leg was modified. There were used only the LP filters by reason of the purpose.

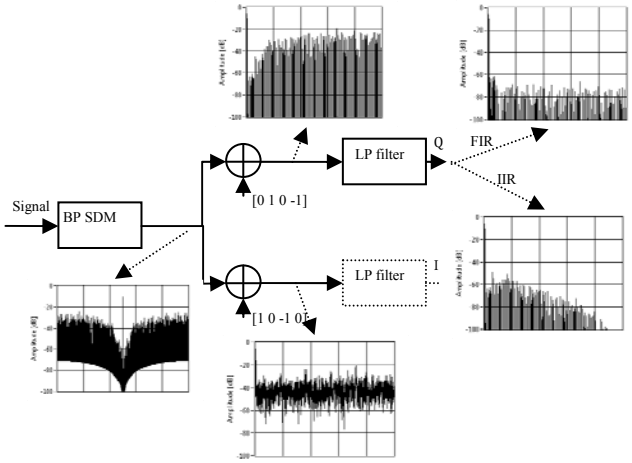


Fig. 1. Signal processing scheme

Figure 1 refers to signal spectrum in concrete processing sections, which are: spectrum behind BP SDM block, spectrum behind digital demodulator at each branch and spectra behind upper branch filter, that are displayed for both of FIR and IIR filters. There were used for work following filters:

FIR with window functions - Rectangular
Hanning
Hamming
Kaiser
Gaussian

IIR this types - Bessel
Chebyshev
Butterworth

A. FILTRATION EFFECT (FIR) WITH DECIMATION ON NOISY SIGNAL

Increased interest was given to filtration effect solving FIR filters with decimation on noisy signal. Decimation in sigma delta transfers can be realized by several forms. Usually decimation is combined with filtration process. Realization of that is comb filter described with equation:

$$H(z) = \left(\frac{1}{M} \cdot \frac{1-z^{-M}}{1-z^{-1}}\right) = \sum_{i=0}^{M-1} z^{-i} \qquad (1)$$

where M represent number of filter coefficients and even decimation coefficient.

The other form of the structure is a FIR filter with individually realized decimation where the filtration is done according formula (3) and the decimation is cancelation of samples according parameter M. Function in figure is an example of sample cancelation and it is a demonstration of a fact that delta t between samples is increasing and therefore the sample frequency is decreasing.

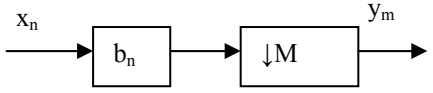
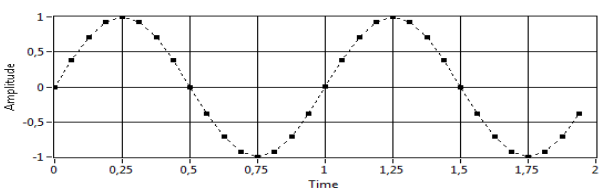


Fig. 2. Filtration and decimacion

M=1



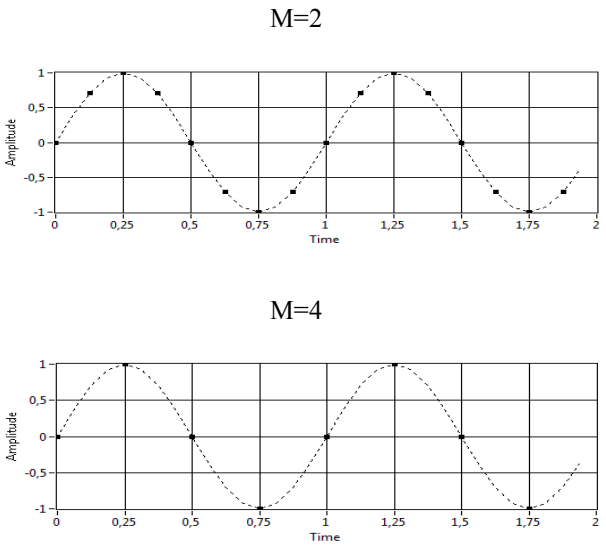


Fig. 3. Example of decimation of sinus signal samples (dashed line)

The effect of changes in sampling frequency depending on the parameter of used decimation was verified in simulation program.

In simulation the signal contained only the DC part of the signal and noise (Gaussian noise), which value was defined in standard deviation. The examination was done for confirmation of attenuation of noise for decimation. The simulation was done using floating FIR filter and FIR filter with decimation. The floating FIR filter was a comb, which all coefficients where of equal value and described by formula(2).Filter with decimation was the same as previous one, but it was application decimation of samples on output. Selected decimation value was equal to number of coefficients of FIR filter. The result of application of FIR filter coefficients was one value, so from one window only one value is extracted on selected size of window of input samples.

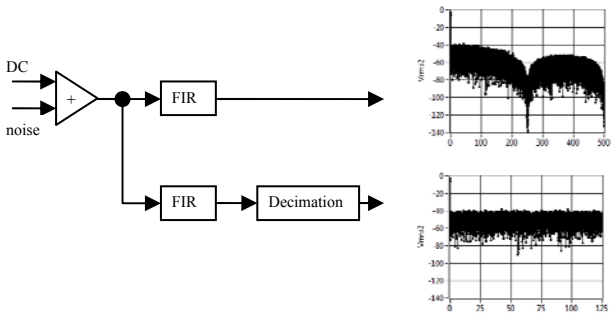


Fig. 4. Realization of examined filtration by FIR filter (M=4)

Floating FIR filter:

$$y_i = \sum_{j=0}^{N_b-1} b_j \cdot x_{i-j} \qquad (2)$$

Where y is a filtrated input x, Nb is number of FIR coefficients, bj is j-th FIR coefficient

Copulation floating filtration with decimation prepares formula (3).

FIR with window function and decimation is:

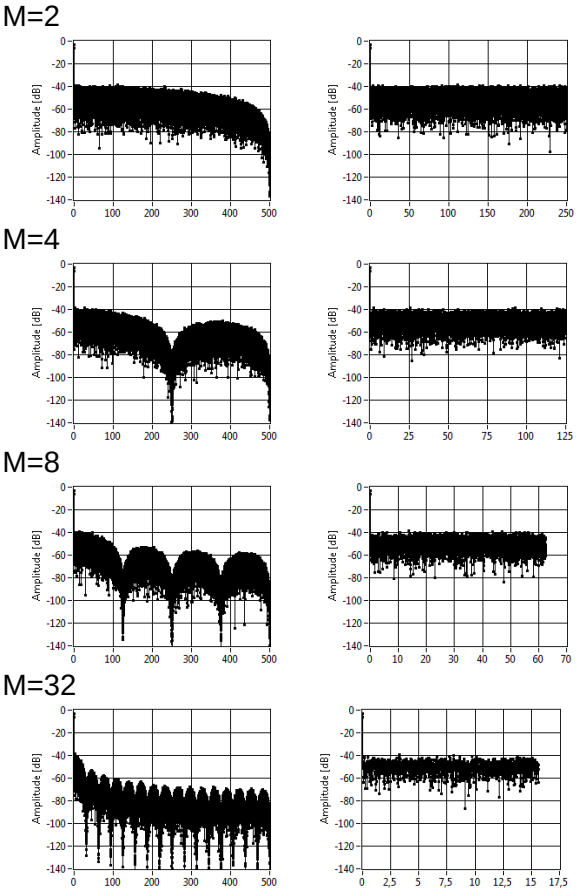
$$y(k) = \sum_{j=0}^{N_b-1} b(j)x(L.k-j) \qquad (3)$$

$k = 0,1,2....$

bj is j-th FIR coefficient

If L=1 is the moving window FIR filter without decimation. If the L=Nb the output signal is being decimated by Nb times. The effect of coefficients and decimation factor are shown on next figures, which are showing us also the power spectrums.

Spectrum after floating FIR and Spectrum after decimation floating filter FIR.



B. *Outcomes with filter FIR and IIR*

The simulation of these filters were verifying only with using equal setting for it an also only for constant amplitude of input signal. Every filter has different behaviour and therefore is not ideal to have for every filter the same boundary frequency. The comparative parameter for interpretation results is used parameter SINAD, which is suitability for that is described in literature [7]:

$$SINAD = \frac{P_{SIGNAL} + P_{NOISE} + P_{DISTORTION}}{P_{NOISE} + P_{DISTORTION}} \qquad (4)$$

Score construction figure 1. is writing down.

SINAD	Filter\order	2	4	8	16	32	64
FIR	Rectangular	4,81	9,42	15,43	22,05	29,44	51,70
	Hanning	4,81	4,81	12,13	23,00	41,15	51,47
	Hamming	4,81	5,97	13,46	25,17	42,01	51,77
	Kaiser	4,81	9,42	15,43	22,05	29,44	51,70
	Gaussian	4,81	7,14	13,93	24,79	41,48	51,32
FIR decimovane	Rectangular	4,94	9,84	15,09	22,71	29,77	50,99
	Hanning	4,94	4,50	12,41	22,93	41,01	50,60
	Hamming	4,94	5,62	13,70	25,21	41,64	50,85
	Kaiser	4,94	9,84	15,09	22,71	29,77	50,99
	Gaussian	4,94	6,79	14,11	24,90	41,14	50,47
IIR	Bessel	32,74	43,82	45,86	44,97		
	Chebyshev	35,68	50,26	49,90	45,84		
	Butterworth	36,55	49,37	50,47	50,49		

II. Conclusion

Based on the results shown in a table, we came to a conclusion that for filtration with FIR filters the filtering with Kaiser and Rectangular windowing obtains best results. On the other hand, from the IIR filters the Butterworth filter is the most suited, as not only it has the best results in the group of IIR filters, but also because of basic advantage of IIR filters over FIR filters. That is, for the same performance the IIR filters requires less coefficients as the window size with FIR filters. For the 2. order, there are no results present in the table for the FIR filter, as the window is in this case too small. Also, the IIR filter was not tested for higher numbers of coefficients, as the structure of the filter would be too complicated and the filter could be unstable. The differences between the results of filtration with and without decimation are of no importance, as they have been heavily influenced by the pass in which SINAD was measured, and also by the settings of lower frequency of the filter.

References

[1] G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.

[2] J. Haze, R. Vrba, L. Fujcik, J. Forejtek, P. Zavoral, M. Pavlik, L. Michaeli, " A Novel Band-Pass Sigma-Delta Modulator for Capacitive Pressure Sensing" Second International Conference on Systems (ICONS'07)

[3] Linus Michaeli, Ján Šaliga, Martin Kollár, "Parameters of band pass RD-ADC and the comparison with the standard ones", Elsevier, Pages 6, Model 3+, January 2007

[4] C. I. Lao, H. L. Leong, K. F. Au, K. H. M and S. P. U, "A 10.7-MHz bandpass sigma-delta modulator using double-delay single-opamp SC Resonator with double-sampling", IEEE Circuit and System, May 2003, pp 1061-1064.

[5] R.W. Erickson, Filter Circuits. ECEN2260, Nov. 10, 1997

[6] DYNAD, "Methods and draft standards for the DYNamic charakterisation and testing of Analogue to Digital converters", (http://www.fe.up.pt/~hsm/dynad)

Planning of path of robots

[1]*Michal KAĽAVSKÝ*, [2]*Miroslav ŤAHLA*

[1]Dept. of Electrical Drivers and Mechatronics, FEI TU of Košice, Slovak Republic
[2]Dept. of Electrical Drivers and Mechatronics, FEI TU of Košice, Slovak Republic

[1]michal.kalavsky@tuke.sk, [2]miroslav.tahla@tuke.sk

***Abstract* — The following paper presents planning of path of robots and his methods. More attention will be devoted method potential field. This method belongs in class of methods planning on grid. Method potential field is one from the most used of methods for searching path in space with obstacles.**

***Keywords*—planning, potential field.**

I. INTRODUCTION

In general, planning is term that means different things to different groups of people, it depends on branch where is used and what is result of planning. We can say that result of planning in robotics is to find algorithms or methods that make safe the certain required move of robot with specific ability.

For planning of path of robot are used several methods of planning that are based on different principles. Of course, each method has his advantage but also his disadvantages. Application of particular method for a certain goal depends on complexity of task that we would like to solve and next application.

II. PLANNING, CONTROL AND LOCALIZATION

As in introduction was considered, the goal of planning path of robot in robotics is to find algorithms or methods that ensure a certain move of robot that if we would like to address him some initial position where he will start and a some goal position where he should have to come in space (in ordered or disordered), so robot will come to goal position. However, in space may be obstacles, so algorithm must generate path without collision with obstacles in order that robot safely could arrive to the goal position. These algorithms are oftentimes supplemented with different approaches of control.

We can disjoin control on reactive control and heuristic control. The basic control parameters are speed of robot and angle of swing out. Reactive control is based on mapping space by various sensorial devices (such as sonar, infrared sensors and laser sensors). This means that robot is fitted with some sensorial devices that are able to map space and obtained information to use at other move of robot. Heuristic control is based on observation of line trajectory computing aberrances. These aberrances are used at determination of value angle of swing out and speed of robot.

The next important component of planning path of robot is his localization in space, actual position of robot. For estimation of position are used above considered sensors. For adjusted localization are used precise information, observation of position by accumulation uncertain information is not sufficient. Notions connected with localization are separated after [6] to the two groups of problems: estimation of global position and position tracking. Estimation of global position expresses ability to find position of robot in known map. Position tracking expresses running update primal of known position. At localization is very essential to find a failure that here is originating. For elimination this failure are used several methods, the best known method is interpretation of probabilistic model solution. This method allocates to each measuring his certain failure. Measured distance d is then expressed as $d \pm r$ that is expected failure with Gaussian distributing with average zero and dispersion r. Then by measuring is possible to estimate position of robot and failure of this estimation.

III. METHODS OF PLANNING OF PATH

In recent years are known several methods for planning of path of robot. This section presents the short overview of basic categories methods for planning of path of robot.

The first category introduces Bug algorithms. Their goal at planning path of robot is to find path from an initial position S into goal position T. Controlling automaton at command has bounded low memory with approximating sensor, acquaintances of his 2D position data, acquaintances of position robot and position goal with it that in space may be local finite number of obstacles. Exist several versions of this algorithm, Bug 1, Bug 2, Bug 1+2, VisBug.

The second category introduces exact planning. This kind of algorithms is applicable just for primitive function, he is not using approximation, if path exists then finds her, in opposed case checks that is not exist. Graph of visibility, Voronoi diagram, trapezoidal decomposition belong in exact planning.

The third category introduces planning on grid. Raster maps, potential field, Dijkstra`s algorithm, A* (pronounced "ay star") and Pseudo-Voronoi diagram belong in planning on grid. Raster maps represent environment by grid where every cell introduces free space or obstacle, size of raster defines exactitude of computing. Raster maps are used mainly in navigation. Dijkstra`s algorithm are used for finding of the shortest path in edge-priced graph. Algorithm is sorting tops under their distance. This algorithm comes through too much cells unnecessarily because always finds all possible cells

from which is possible to get to the goal. However, this is very time-consuming and this is the mainly disadvantage of method Dijkstra`s algorithm. A* algorithms belong to scanning algorithms using heuristics. The principle of A* algorithms is to move in state space that we will traverse into following states that have bigger evaluation than actual state. A* method never will be stuck in local extreme and it is the mainly advantage of this method. Method potential field will be explained further in more detail in section IV.

The fourth category introduces neuron networks at planning. Their basic advantage is ability to elaborate failed or incomplete sensor information. The principle consists in configuration of neurons into specific grid that introduces raster map of given space. The result is grid with value 0 (neuron introduced obstacle) and 1 (neuron introduced free path) introducing potential field. Neurons networks are the most used in navigation of mobile robots.

The fifth category introduces probabilistic planning. This method of planning is based on probability. If we want to describe probabilistic planning, we have to define certain configuration space denoted as Cspace. Cspace is n-dimensional space where n is number of parameters expressly defining position whether configuration of robot. Number of parameters is important for us because are describing configuration of robot. Algorithm is based on two phases. The first is generation of path and the second is finding of path in graph. In first phase are sought-after all possible places in space that belong in the Cspace. The second phase begins finding path in graph. If path is not exist then really is not exist or in the first phase is insufficient covering. Probabilistic planning is efficient to manage with not convex robot (vertiginous robot), with robots with constraint move (car) and with dynamics of move represented with inertia and with constraint acceleration.

IV. METHOD POTENTIAL FIELD

In this chapter will be shortly explained entity of this method and weakness of this method that is needed to eliminate.

Potential field is force field in space which properties are not depending on arrived distance between two points in the space but they are just depending on the position of these two points. Size of absorbed energy is always the same. Potential fields are abstract expression of real surroundings. They are based on structure of virtual maps which simulate the real surroundings. These maps are divided to certain areas and each area will have certain degree of repulsion. For understanding we can imagine maps of certain surroundings which are divided to certain areas or cells. To each cell will be to assign certain value. By this value we will decide later in which direction the path will be constructed. If we enter any value to each cell we have to enter value cell of start and cell of goal. Start will have the lowest value and goal the highest or it can be also in reverse. We can see sample of potential field in Figure 1.

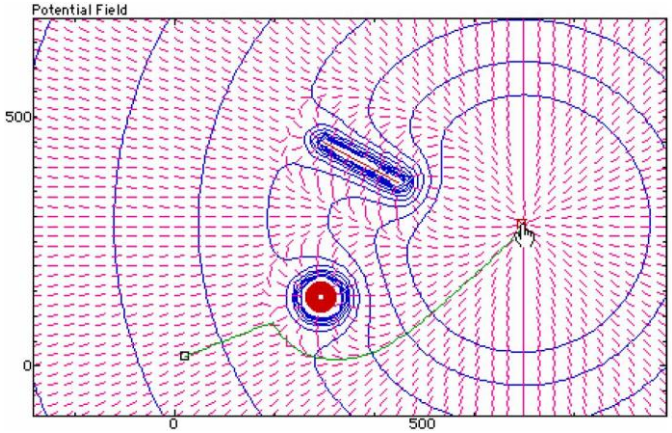


Figure 1: Sample of potential field at circumvention obstacles in space.

For planning path of robot where is needed to determine obstacles in space by sensor of robot is needed to use modified version of potential field after [3]. Potential field has effect at elements with forces and thereby determines move of this elements. The basic forces that are used at planning of path of robots are attractive force and repulsive force which is placed vertically from obstacles. By these forces we can define repulsive and attractive function. Both these functions are defined as convex functions of distance between two points. Repulsive force defines distance between robot and obstacle and if distance is reduced then repulsive force is increased. Attractive force defines distance between robot and goal and if distance is reduced then attractive force is too reduced. The sum of these two functions is implicit potential field with only one minimum in goal point. It follows that potential field U(x) is originated from potential field in direction to goal $U_{goal}(x)$ and to sum of potential fields from obstacle $U_{obstacle\_i}(x)$:

$$U(x) = U_{goal}(x) + \sum_{i \in I} U_{obstacle\_i}(x) \quad (1)$$

where I is ensemble of all obstacles. The final force F after which we are planning move of robot is equal:

$$F(x) = -\nabla U(x) = -\begin{pmatrix} \frac{\partial U}{\partial x} \\ \frac{\partial U}{\partial y} \end{pmatrix}, \quad (2)$$

what is possible to express as:

$$\begin{aligned} F_{attractive}(x) &= -\nabla U_{goal}(x), \\ F_{repulsive}(x) &= -\nabla U_{obstacle}(x), \\ F(x) &= F_{attractive}(x) + F_{repulsive}(x). \end{aligned} \quad (3)$$

Consequently is possible to express attractive final potential as:

$$U_{goal}(x) = \alpha . distance(x, goal)^2, \quad (4)$$

where α defines size of force.

Consequently after [4] is possible to express concretized attractive (final) potential:

$$U_{goal}(x) = \frac{1}{2} k_p (x - x_c)^2, \quad (5)$$

where k_p is constant and x_c is goal position.

Repulsive (obstacle) potential is possible after [4] to express as:

$U_{obstacle}(x) = \frac{1}{2}\eta\left(\frac{1}{\rho(x)} - \frac{1}{\rho_0}\right)^2$, if $\rho(x) \leq \rho_0$ and
$U_{obstacle}(x) = 0$, if $\rho(x) > \rho_0$, (6)

where is just expressed one obstacle, where $\rho(x)$ expresses the shortest distance at given obstacle and ρ_0 distance of impact repulsive force to obstacle O.

By formulation of equations for attractive and repulsive potential is possible determine repulsive and attractive force.

Repulsive force is expressed as:

$F_{repulsive}(x) = -\nabla U_{obstacle}(x) = \frac{1}{2}\eta\left(\frac{1}{\rho(x)} - \frac{1}{\rho_0}\right)\frac{1}{\rho^2 x}\frac{\partial\rho}{\partial x}$, ,
if $\rho(x) \leq \rho_0$ and
$F_{repulsive}(x) = -\nabla U_{obstacle}(x) = 0$, (7)
if $\rho(x) > \rho_0$.

Attractive force is expressed as:

$F_{attractive}(x) = -\nabla U_{goal}(x) = -k_p(x - x_c)$ (8)

If the space was represented with several obstacles so vector of repulsive forces will be equal at sum of partial repulsive forces from particular obstacles.

Method potential field has one weakness which is needed eliminate. This weakness is danger of creation local minimum as is depicted in Figure 2.

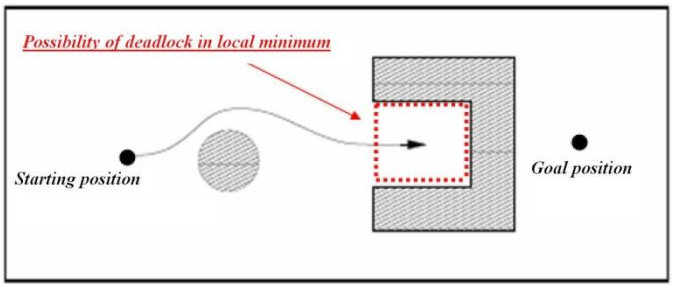


Figure 2: Sample of feasibility to enmesh in local minimum at method potential field.

From Figure 2 results that robot would sidle in local minimum in this critical space in effort to get into goal position. If that happen robot will be circular and then will not get into goal position. This problem is possible to solve using of method harmonic potential field where just exists one minimum scilicet global minimum represented by goal position.

Harmonic potential field defined after [3] on ensemble Ω is function that fulfills Laplace`s equation:

$\nabla^2 U(x) = 0$. (9)

For modification this method are borders of ensemble Ω defined as borders of all obstacles and goal position. Searched potential field must fulfill these boundary conditions:

$U|_{\partial\Omega} = 0$, (10)
$U_{(ciel)} = 0$. (11)

For better understanding of principle harmonic potential field consider the situation that robot is on some startup position which has some value and goal position has value zero. Robot will come through from his actual position on position that value is lesser. If robot will get at zero position so he is in goal position, goal position is global minimum. However, if we want to have guarantee of finding existing path sometimes algorithms of this method must research all state space.

V. Conclusion

The goal of this paper was to introduce methods of planning of path specifically method potential field. Each method has his advantage and weakness. The choice of method depends on task that is needed to solve.

The main goal of my PhD. work is to find algorithm for simulation of move of electrical car in space with obstacles. The task of electrical car will be arrived this space without collision with any obstacle, thus planning of his path in space. For the purpose of my work I have chosen the method of potential field. This method belongs to the most using methods of finding of path in space with obstacles and is based on construction of virtual maps that simulate physical space.

References

[1] J.Vaščák, Utilization of potential field in navigation of mobile robots, Dept. of Cybernetics and Artificial Intelligence, FEI, TU Košice, 2008
[2] S.M. LaValle, Planning Algorithms (Book style), University of Illinois: Cambridge University Press, 2006.
[3] M. Vacek, Intelligent multi-agent parking system, Dept. of Cybernetics and Artificial Intelligence, FEI, TU Košice, Graduation theses, 2004
[4] R. Court – D. Hilbert, Methods of Mathematical Physics, I. Interscience, New York, N.Y 1953.
[5] http://www.atpjournal.sk/casopisy/atp_03/pdf/atp-2003-05-36.pdf
[6] http://rudy.negenborn.net/kal_loc/thesis.pdf
[7] http://en.wikipedia.org/wiki/Motion_planning

Dynamic phenomena on the external power line conductors

[1]*Matúš KATIN*

[1]Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]matus.katin@tuke.sk

Abstract— This article discusses of the dynamic effects incipient on the external power lines which arise as a consequence of different weather influences affecting the line.

***Keywords*—External power line, ice formation, galloping, conductor swing.**

I. INTRODUCTION

The external power lines are one of the most important elements of the electricity supply system. The ability of lines to transmit electricity even during the most adverse weather conditions affects the reliability of the main system considerably.

Nowadays, when is liberalization of the electricity market and the valuation of non-delivered electricity are being introduced into the practice, the reliability of electricity supply is emerging. Dynamic phenomena affecting the external lines are the main aspects that influence the reliability of electricity supply from external power lines.

II. CAUSES OF THE DYNAMIC PHENOMENA ON THE EXTERNAL LINES

If external power line passes through an environment, it influences it a lot. The affects of winds, storms, lightning, ice, temperature changes, altitude and other climatic factors affect external power line. These climatic conditions depend mainly on the geographical location where the external power line is placed so they are different in every country. The influence of weather is determined on the basis of the long-time meteorological observations. These climatic conditions mainly affects the maximum load of the conductor, fittings and construction of spars.

They can be divided into:

- permanent – static
- random – dynamic (mostly short-term)

Dynamic strain of spars, fittings, and conductors are caused by the sudden change of the maximum load (release of stored energy).

The reasond of such change are mostly:

- ice falling of the conductors
- change of the pressure or wind direction
- existence of the galloping

A. Ice Formation

Ice formation is a metrological phenomenon which creates a sediment of ice over the conductor. Those sediments of ice are not only the cause of possible dynamic phenomena but, in the worst case, it can lead to the breakage of the wire or mechanical damage of the spar. Expected sediment of ice over the external power line conductors can be determined from the map of icing area. According to the European standard STN-EN 50341-3, which has the status of Slovak technical standard are icing areas divided into the N0, N1, N3, N5, N8, N12, N18, NK, where the numbers indicates the areas of ice mass in $kg.m^{-1}$.(Fig.1)

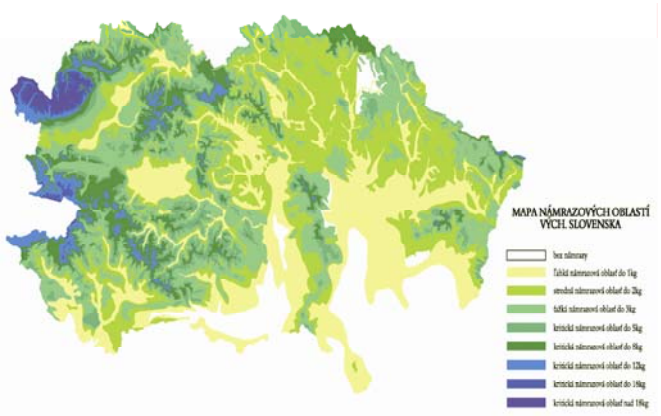


Fig. 1. Icing areas for the territory of Eastern Slovakia

III. DYNAMIC PHENOMENA ON THE EXTERNAL POWER LINES CONDUCTORS

A. Oscillation (Vibration) of Conductors

Besides overloading and conductor deviation, the wind causes even more important phenomenon, which is oscillation (vibration) of conductors in the vertical plain (seiche). The vibration of conductors is the result of aerodynamic affect of the wind with low speed. Air swirls rise behind the conductor and their emerging and expiration as well as the speed of flow above and below the conductor is changing. Therefore, the pressure is altering. This phenomenon is followed by weak dynamic shocks in the vertical plain, affecting regularly and periodically (Fig.2).

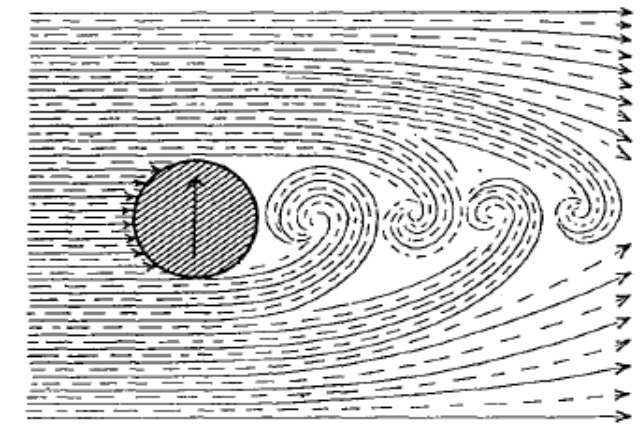


Fig. 2. Air swirls rising behind the conductor

IT was found experimentally, that the frequency of vibration varies from 10 to 20, or up to 50 Hz (when using thinner conductors at the limit of hearing). Wave length varies from 1 to 20 m and the amplitude of this oscillation is few cm. The result of the vibration is an additional dynamic stress in the conductor, that can cause fatigue break of the conductor. This phenomenon rises mainly at by catching of the conductor in the clip, connected to the insulator string.

Forced frequency is given by equation:

$$f_{vn} = 0{,}2\frac{v}{d} \tag{1}$$

Where:
v – speed of wind ($m.s^{-1}$)
d – diameter of the conductor (m)

Vibrations the conductor is seiche and can be expressed by the equation:

$$u = 2.u_0.\cos\frac{2.\Pi.x}{\lambda}.\sin\omega.t \tag{2}$$

Where:
u – immediate value of the amplitude in the given position (m)
u_0 – maximum amplitude (m)
x – distance of the given point from fixation point (m)
λ – length of wave (m)
ω – circular frequency (s^{-1})

Against these unacceptable effects, two types of protection are used:
Passive protection: Reducing the static pull in the conductors
Festons (Ropes hanged on te both sides of the clip, fixed to the conductor at several places)
Use of the oscillation clips repeating the motion of the conductor

Active protection: Use of the anti-vibration ropes
Use of dumpers (device evoking forces that are phase-shifted after conductor motions and thereby working against the vibration of the conductor by the rise of the vibration)

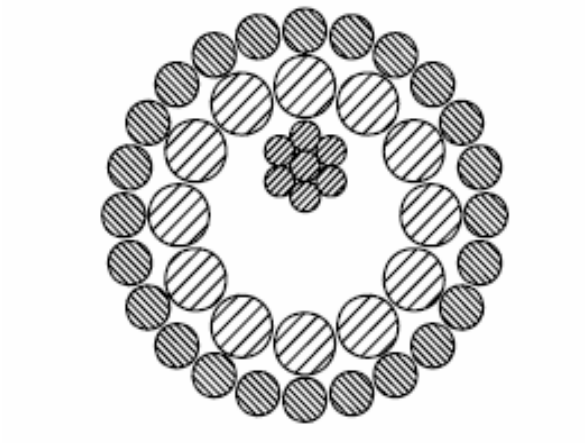


Fig. 3. Anti-vibration rope

B. Swing up of Conductor

Except overloading of the conductor hoar-frost can also cause swing up of the conductor. Due to sudden ice falling off the conductor it swings up and reaches its stability after few strongly damped oscillations (Fig. 4). The impulse for ice falling away is caused by strong collision wind or increase of temperature. In case, that external power line conductors are arranged on the spar one above the other (spar type“ barrel") it may result into the contact of the conductors or to dangerous approach and subsequently to double-phase short circuit. The analysis of difficulties and drop outs of external power lines shows, that from 30 to 40% of them is caused by hoar-frost (in winter).

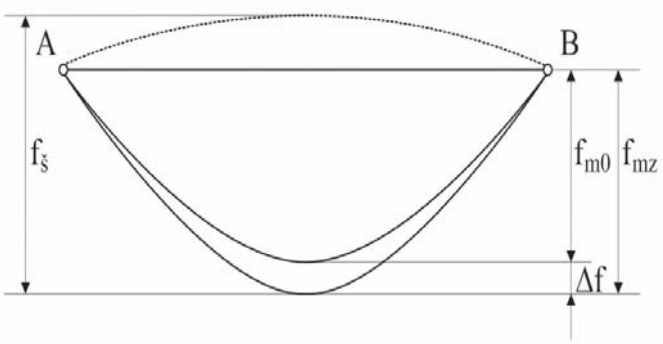


Fig. 3. Swing up of conductor

The size of the oscillation amplitude is more important than the mechanism of creation of unstable oscillations itself. Maximum amplitude of swing can be determined using methods based on the energy balance of the conductor before falling of the ice frost and after it (4).

$$\Delta f = f_{mz} - f_{m0} \tag{3}$$

$$f_{š} = 2.\Delta f \tag{4}$$

Where:
Δf – difference of flexures in the middle of span (m)
f_{mz} – maximum flexure of frozen conductor in the middle of span (m)
f_{m0} – maximum flexure of non frozen conductor in the middle of span (m)

The aforementioned methodology reviews the process of swinging the conductor quasi-steady, ignores the dynamic and it assumes the parabolic shape of the conductor during the whole period of oscillation. As can be seen in Fig. 4 in case of the real movement the middle part of the conductor is in span speeded up more than marginal parts of the conductor closer to the points of gripping. On this basis we can point out that the equation mentioned above can serve only for informative estimate of conductor swing. More precisely results can give us real simulations in real conditions with simulation program.

Minimalization of the negative impacts of faling-of the ice can be realized in practice in two available ways. The first one is to change the configuration of spar heads and the transition from vertical to horizontal layout of conductors, by extending the interphase distance. However this method brings several complications. Reconstruction of line takes a long time without putting it into operation and will cause increase of the land because of line protection zones. The second possibility is the installation of interphase separators between the vertically arranged conductors. The advantage of this method of minimalization of the consequences of swing conductor is time-saving installation of separators, the line is able to operate in the short period since the start of separators installation. Their use does not increase the demands for the land.This phenomenon can be analyzed in 2D dimension, or if the crosswind affects the conductor, in 3D dimension.

IV. THE CONCLUSION

This article discussed of the dynamic phenomena rising at the external power lines. My upcomming work will be dealing with falling-of the ice simulations from the external power lines conductors in real terms using simulation program cosmos/m

REFERENCES

[1] L. Varga, - S. Ilenin, *Elektrické siete* .Košice,Technická univerzita: 2006, ISBN 978-80-8073-856-3.

[2] L. Varga, - S. Ilenin, - R. Hudák, *Dynamické javy na vodičoch vonkajších silových vedení pri opadávaní námrazy.* Stará Lesná, Elektroenergetika*:* 2005.

[3] V. List, - K. Pochop, *Mechanical design of overhead transmission lines.* Praha, SNTL: 1963

[4] L. Varga, - S. Ilenin, - R. Hudák, *Dynamické javy na vodičoch vpnkajších silových vedení pri opadavaní námrazy.* Časopis EE, 12: 2006, str. 6 – 8

[5] L. Varga, - S. Ilenin, *Modelovanie pohybu vodičov vonkajších silových vedení (VSV) pri opadávaní námrazy*. Elektrotechnika v praxi, 3 – 4: 2008. str. 162 - 166

Objects Detection in Video Surveillance System

Anna KAŽIMÍROVÁ KOLESÁROVÁ

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

anna.kolesarova@tuke.sk

***Abstract*— More than ever before, it is important to maintain the safety and security of citizens, public infrastructure, buildings. This paper is concerned with video surveillance systems. With the growing quantity of security video, it becomes vital that video surveillance system be able to support security personnel in monitoring and tracking activities. In this paper is described new video surveillance system.**

***Keywords*— video surveillance systems, security personnel, detection of removed luggage and abandoned luggage.**

I. INTRODUCTION

Video surveillance is an active area of research. Object detection and tracking in video surveillance systems are commonly based on background estimation a subtraction. The primary focus of today's video surveillance systems act is the application of video compression technology to efficiently multiplex or store images from a large number of cameras onto mass store devices (video tapes, discs) [4].

From the perspective of real-time threat detection, it is well know that human visual attention drops below acceptance levels, even when trained personal and assigned to the task of visual monitoring [9]. On the other side, video analysis technologies can be applied to develop smart surveillance systems that can be aid the human operator in real-time threat detection [1]. Specifically, multiscale tracking technologies are the next step in applying automatic video analysis to surveillance systems.

Application of visual surveillance include car and pedestrian traffic monitoring, human activity surveillance for unusual activity detection, people counting, ect. A typical surveillance application consists of three buildings blocks: moving detection, object tracking and higher level motion analysis.

Several video surveillance products are available on the market for office and home security as well as remote surveillance. They monitor a home, an office, or any location of interest, capturing motion events using webcams or camcorders and detect abnormalities [7]. In the case of webcams, the visual data is saved into compressed or uncompressed video clips, and the system trigger various alerts such as sending an e-mail.

II. VIDEO SURVEILLANCE SYSTEM DESCRIPTION

After classifying an object, we want to determine what it is doing. Understanding human activity is one of the most difficult open problems in the area of automated video surveillance. Detecting and analyzing human motion in real time from video imagery has only recently become viable with algorithms. These algorithms represent a good first step to the problem of recognizing and analyzing humans, but they still have some drawbacks. Therefore the human subject must dominate the image frame so that the individual body components can be reliably detected [6].

Tracking accessibility of people to the desired rooms, where there is "Employees only!". At airports, stations, schools and etc., the security is very important for prevention of employees and all others.

We designed system (see Fig. 2) that works follows. We have video output from CCD camera. This video output is divided to video sequences that are input for process called preprocessing. To recognition moving objects on the background, head detection and luggage detection we using the tracker. *Tracker* contained following blocks*: Motion Detector*, *Head Detector*, *Shape Tracker* and *Region Tracker*. Tracking output is recognized in recognition block.

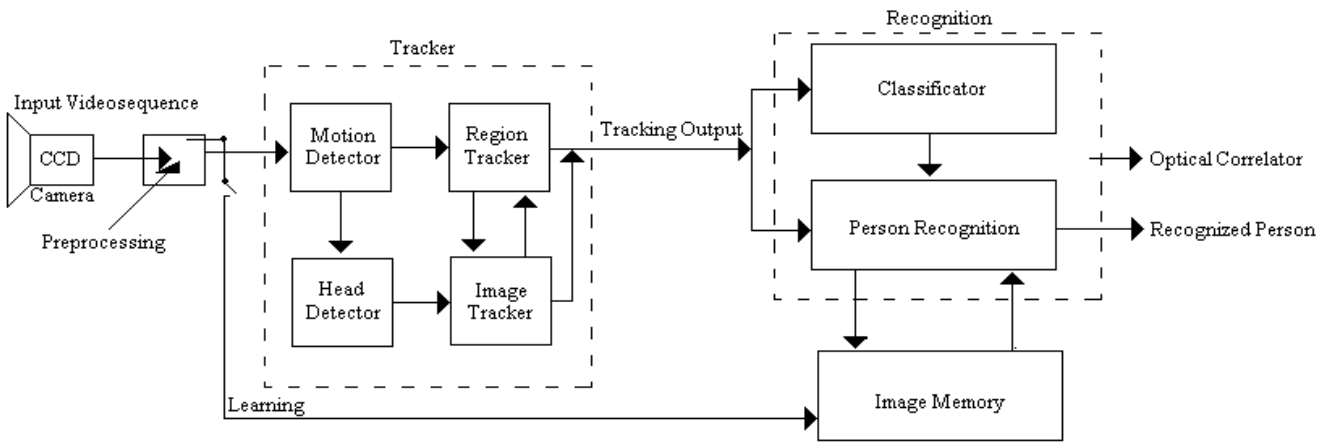


Fig. 1 System block diagram

A Motion Detector detects moving pixels in the image. It models the background as an image with no people in it. Simply subtracting it pixel wise from of the current video image and thresholding the result yields the binary Motion Image. Regions (bounding boxes) with detected moving blobs are then extracted and written out as the output from this module.

Main features of a Motion Detector are:

- simple background image subtraction,
- image filtering (spatial median filter, dilation) depending on available CPU time,
- temporal inclusion of static objects into the background,
- background modelling using a speed-optimized median filter,
- static regions incorporated into background (multi-layer background).

A Head Detector makes rapid guesses of head positions in all detected moving regions.

Main features of a Head Detector are:

- works in binary motion image,
- looks for peaks in detected moving regions,
- vertical pixel histogram with low-pass filter,
- optimized for speed not accuracy.

A Image Tracker uses a deformable model for the 2D outline shape of a walking pedestrian to detect and track people. The initialization of contour shapes is done from the output by the Region Tracker and the Head Detector.

Main features of a Image Tracker are:

- local edge search for shape fitting,
- initializing of shape from Region Tracker, Head Detector and own predictions,
- occlusion reasoning.

A Region Tracker tracks these moving regions over time. This includes region splitting and merging using predictions from the previous frame.

Main features of a Region Tracker are:

- region splitting and merging using predictions,
- adjust bounding box from Shape Tracker results,
- identify static regions for background integration.

Recognition block contained two blocks: *Classificator* and *Personal Recognition*. Data from recognition output are compared with data from *Image Memory*.

Image memory is database of static images of human faces, that have guarded enter to this room (employees faces). ***Learning*** is a process of personal identities creation.

An Optical Correlator is a device for comparing two signals by utilizing the Fourier transforming properties of a lens. It is commonly used in optics for target tracking and identification. The correlator has an input signal which is multiplied by some filter in the Fourier domain.

An optical correlator automatically recognizes or identifies the contents of an image by combining an incoming image with a reference image, and the degree of correlation after combining the images determining the intensity of an output light beam.

First task for the optical correlator is to link together person with his luggage, case or package. This is then monitored if this person leaves guarded room with the same luggage, case, etc.

Second task for optical correlator is to compare faces from tracker with database of known faces that have guarded access to the specific room.

III. RESULTS

A new robust and efficient analysis method of video sequence allows the extraction of foreground objects and the classification of static foreground regions as abandoned or removed objects.

As a first step, the moving regions in the scene are detected by subtracting to the current frame a background model continuously adapted. Then, a shadow removing algorithm is used to extract the real shape of detected objects.

Finally, moving objects are classified as abandoned or removed by matching the boundaries of static foreground regions.

Figures 2 and 3 show two examples of abandoned luggage and removed luggage.

a) b)

Fig. 2 a) Detection of one luggage b) Detection of one abandoned luggage

c) d)

Fig. 3 a) Detection of two luggage, b) Detection of one luggage removed

IV. CONCLUSION

Real-time video analysis provides surveillance systems with the ability to react to an activity in real time, thus acquiring relevant information at much higher resolution [3]. The long-term operation of such systems provides the ability to analyze information in a spatial-temporal context.

Despite the importance of the subject and the intensive research done, background detection remains a challenging problem in applications with difficult circumstances, such as changing illumination, waving trees, water, video displays, rotating fans, moving shadows, inter-reflections, camouflage, occasional changes of the true background, high traffic, etc [2].

The problem of remote surveillance has received growing attention in recent years, especially in the context of public

infrastructure monitoring for transport applications, safety of quality control in industrial applications, and improved public security. The development of a surveillance system requires multidisciplinary expertise, including knowledge of signal and image processing, computer vision, communications and networking pattern recognition and sensor development and fusion [3].

Our system is preventing before entering forbidden person and leaving the suspicious luggage into the guarded room. In this luggage or package could be bomb, gun, drugs, etc. On the other side, big task is checking if some person steals the luggage, package or the other things.

Our system could increase security employees and the other people in schools, stations, airports, etc.

ACKNOWLEDGMENT

This work was partially supported from the grants VEGA No. 01/0045/10, project COST ICO802 and by Agency of the Ministry of Education of the Slovak Republic for the Structural Funds of the EU under the project Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020).

REFERENCES

[1] ARAKI, A. - MATSUOKA, T. - YOKOYA, N. – TAKEMURA, H.: *"Real-Time Tracking of Multiple Moving Object Contours in a Moving Camera Image Sequence"*, IEICE Trans. Inf. &Syst., vol. E83-D, no. 7, pp. 1583 – 1591, July 2001.

[2] BERAN, V. - HEROUT, A. – ŘEZNÍČEK, I.: „Video-Based Bicycle Detection in Underground Scenarios“, In: Proceedings of WSCG'09, Plzeň, CZ, p. 4, 2009.

[3] BHARGAVA, M. – CHEN, CH. - RYOO, M. S. - AGGARWA, J. K.: „Detection of object abandonment using temporal logic“, Springer Berlin, pp. 271-281, January 2009.

[4] BOJKOVIČ, Z. - SAMČOVIČ, A. – TURÁN, T.: “Object Detection and Tracking in Video Surveillance systems”, COST 276 Workshop, Trondheim, Norvegia, 113 – 116, May 25 – 26, 2005.

[5] BOULT, T. - et al.: "Into the Woods: Visual Surveillance of Noncooperative and Camouflaged Targets in Complex Outdoor Settings", in Proceeding of the IEEE, vol. 89, no. 10, Oct. 2001.

[6] COLLINS, R. - et al.: "A System for Video Surveillance and Monitoring", CMU-RI-TR-00-12, 2000.

[7] HARITAOGLU, H.: "Hartwood and Devis, W4: Real Time Surveillance of People and their Activities", IEEE Trans. Pattern Anal. Machine Intell., vol. 22, no. 8, pp. 809 – 830, Aug. 2000.

[8] McKENNA, S. - et al.: "Tracking Groups of People", CVIU 80, pp. 42 - 56, 2000.

[9] RAO, K. R. - BOJKOVIČ, Z. S. - MILOVANOVIČ, D. A.: "Multimedia Communication Systems: Techniques, Standards and Networks", Prentice-Hall PTR, New Jersey, 2002.

Study of the Rheological Behaviors of Solder Pastes

[1]*Michal KRAVČÍK,* [2]*Igor VEHEC*

[1]Dept. of Technologies in Electronics, FEI TU of Košice, Slovak Republic
[2]Dept. of Technologies in Electronics, FEI TU of Košice, Slovak Republic

[1]michal.kravcik@tuke.sk, [2]i.vehec@tuke.sk

***Abstract*—Solder paste is a homogeneous, stable suspension of solder powder particles suspended in a flux binder, and is one of the most important process materials today in surface mount technology (SMT). By varying the solder particle size, distribution and shape, as well as the other constituent materials, the rheology and printing performance of solder pastes can be controlled. Paste flow behavior is very important in defining the printing performance of any paste. The purpose of this paper is to study the rheological behavior of SAC (Sn-Ag-Cu) solder paste used for surface mount applications in the electronic industry.**

The reason why the rheological tests are presented in this paper are two critical sub-processes: aperture filling and paste withdraw. In this paper, we report on the investigation of the rheological profiles, the serrated cone-to-plate system was found as effective in parameter minimizing the wall-slip effect (Fig. 1).

***Keywords*— rheology, solder paste, thixotropy.**

I. INTRODUCTION

The solder paste is used for connecting the terminations of integrated chip with land patterns on the PCB. The paste is applied to the lands by printing the solder paste using a stencil, while other methods like screening and dispersing are also used. A majority of defects in mount assemblies are caused due to the issues in printing process of due to defects in the solder paste. An electrics manufacturer needs to have a good idea about the printing process, specifically the paste characteristics, to avoid reworking costs on the assemblies. Characteristics of the paste, like viscosity and flux levels, need to be monitored periodically by performing in-house tests. One approach currently adopted by the industry is to reduce the solder alloy particle size to facilitate paste flow through the very small stencil apertures.

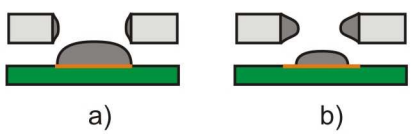


Fig. 1. a) Solder paste is release correctly (no wall-slip effect). b) Solder paste isn't release correctly (wall-slip effect).

However, reducing the particle size in this way has been shown to radically affect the paste rheology and consequently the printing behavior. In this paper we address the need for characterizing 3 solder paste formulations; and present a procedure for evaluating solder pastes being developed for a application using the stencil printing process.

Solder paste is one the most widely used interconnection material in the electrical bond between electronic components and the substrate. Solder paste can by categorized as a homogeneous and dense suspension of solder alloy particles suspended in flux medium. For typical solder paste, the typical metal content is between 88 to 91% by weight, and about 30-70% by volume. The most commonly used lead free solder alloy based SAC. The main constituent of flux medium is a naturally occurring rosin or chemically made rosin. Rosin is used to remove impurities and clean up soldered surfaces and help to solder alloy joint components and metal pads on printed circuit board. A number of different ingredients including solvents, activators, thickeners, thixotropic agent, and tackifiers are added to the flux to provide the desired rheological properties to the solder paste [2].

II. BASIC CONCEPTS OF RHEOLOGICAL PROPERTIES

A. Viscosity

The thixotropy behavior was investigated through rheological test based on shear rate test. In the steady shear rate test, the materials were subjected to a linear rising shear rate from 0 to 32 s^{-1} for a period 600 seconds. To measure the viscosity of liquids required firstly the definition of the parameters which are involved in the flow. Then one has to find suitable test conditions which allow the measurement of flow properties objectively and reproducibly.

Isaac Newton was the first to find basic law of viscosimetry describing the flow behavior of an ideal liquid. He defined the viscosity η as relation shear stress τ over shear rate D (1) [3].

$$\eta = \frac{\tau}{D} \quad [\text{Pa.s}] \qquad (1)$$

τ - shear stress [Pa]

η – viscosity [Pa.s]

D – Shear rate [s^{-1}]

The parallel cone-to-plate model helps to define both shear stress and shear rate Fig. 2. Cone-to-plate system has been chosen for his best results with solder paste.

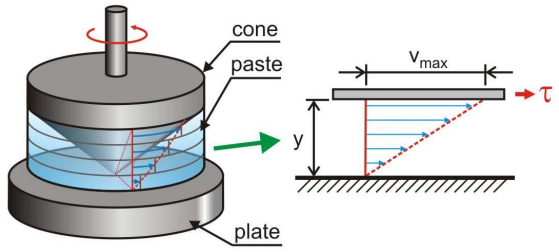


Fig. 2. Principle of cone-to-plate model. a) Cone-to-plate system for measuring solder paste. b) detail for calculation relation of shear stress and shear rate. V_{MAX} is a flow speed, τ is a shear stress, y is gap size.

Shear stress

A force F applied an area being the interface between the upper plate and the liquid underneath leads to a flow in the liquid layer. The velocity of flow that can be maintained for a given force will be controlled by the internal resistance of the liquid, i.e. by its viscosity (2) [3].

$$\tau = \frac{F(force)}{A(area)} = \frac{N(newton)}{m^2} = Pa \text{ (Pascal)} \qquad (2)$$

Shear rate

The shear stress τ causes the liquid to flow in a special pattern (Fig. 2 b). A maximum flow speed 'v_{MAX}' will be found at the upper boundary of plate moved in direction of τ. The speed drops across the gap size 'y' down to 'v_{MIN}=0' at the lower boundary contacting the stationary plate. Laminar flow means that infinitesimally thin liquid layers slide on top of each other, similar to cards in a deck-of-cards. One laminar layer is then displaced with respect to the adjacent ones by a fraction of the total displacement encountered in the liquid between both plates.

In the general form the shear rate *D* is defined by a differential (3) [3]:

$$D = \frac{dv}{dy} \qquad [s^{-1}] \qquad (3)$$

In the case of linear speed drop across the gap the differential in the equation above can be approximated by

$$D \approx \frac{v_{MAX}}{y}[s^{-1}] \qquad (4)$$

B. Thixotropy and rheology of solder paste

Thixotropy is defined as: 'Memory' property of a fluid (especially of solder paste) where by its viscosity (resistance to flow) depends on its recent history of flow and not just on the force applied to it. This idea influenced from the Fig. 4.

Rheology is defined as a term describing the viscosity and surface tension properties of solder pastes or adhesives.

Typically viscosity for solder pastes is in range from 10 to 1000 Pa.s. Solder paste exhibits non-Newtonian and thixotropic behavior when subjected to a shearing stress. The viscosity of a material can be defined as the ratio of shear force to shear rate (1). Comparison of flow as well as viscosity for Newtonian and Non-Newtonian liquids implicit from Fig. 3.

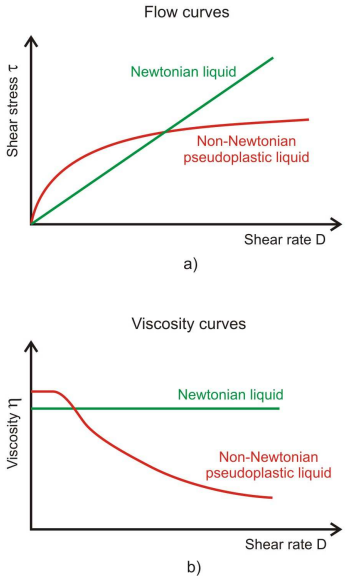


Fig. 3 a) Flow curves: dependency of shear rate stress on shear rate. Curve 1 is Newtonian liquid; curve 2 is Non-Newtonian liquid b) Viscosity curves: Dependency viscosity on shear rate. Curve 1 is Newtonian liquid; curve 2 is Non-Newtonian liquid.

Materials made up of complex organic molecules with a range of organic functional groups are capable of intermolecular interactions that lead to an inherent steady state structure in the material. This phenomenon can also occur as a result of intermolecular interactions between particles, such as solder spheres. Diagram describing thixotropy usually named rheogram is in Fig. 4. Other notable features of the rheogram are the initial increase in viscosity, as the as the shear stress increases without significant shear rate increase, followed by the paste 'yielding' and then undergoing shear thinning. Shear tinning is defined as the property of a fluid (usually solder paste) where the viscosity (that is, the resistance to flow) reduces temporarily as the fluid is subjected to an increased shear force, (for example by a squeegee during the print process). The above behavior is desirable and necessary for satisfactory printing and antislump proprieties. A paste is subjected to a wide range of shear rate during various phases of the printing process (Fig. 5). Theses are classified as mixing, rolling and stencil printing (Fig. 6) [3].

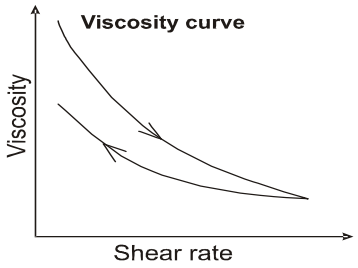


Fig. 4 Diagram describing Thixotropy.

The organic chemicals cream is referred to as 'flux' and is generally a trade secret and/or covered by patents. The purpose of flux is to give the solder paste its cream-like texture and to enable formation of metal joints by ensuring that the metal surfaces are 'clean' of oxides at the time the

metal joint are formed. Rheological behavior of solder paste during stencil printing implicit from the Fig. 6.

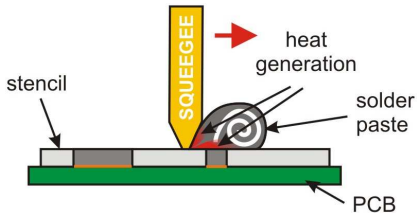


Fig. 5 Stencil printing process. Rubbing squeegee on stencil can produce some heat, and warm up solder pastes.

III. Experimental Procedure

The work report in this work is concerned on rheological characterization of solder pastes designed for pine pitch application to the stencil printing process. The first part of this study deals with solder paste samples rheological characteristics. Two different rheological tests including viscosity sensitivity of temperature and Thixotropy test depend on measuring time duration [1]. We published out only a small part of what we tested within the stencil printing process.

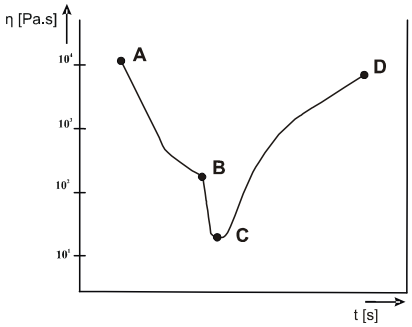


Fig. 6 Viscosity of solder paste during printing process. Point A is beginning of printing process, B is filling up an apertures with solder paste, C is contact point of solder paste and substrate and D is the end of printing process.

A. Solder paste samples

Three commercially available lead-free solder pastes (P1, P2 and P3) were used in the experimental studies report in this paper. All of them are classified as SAC based, no-clean and halide free. All three samples have to same particle size distribution (25-45µm) and metal loading 88.5 % by weight. P1 and P2 sample have melting point range at 217-220°C and P3 has melting point range at 217-234°C. The details of these samples are provided in Table 1 [4][5][6].

Table 1. Solder paste properties.

Paste	*Particle size distribution [µm]*	*Metal loading [% by weight]*	*Melting point range [°C]*	*Flux type*	*Alloy*
P1	25-45	88 ±0.5	217-220	F1	96.5Sn-3Ag-0.5Cu
P2	25-45	88.5	217-220	F2	96.5Sn-3Ag-0.5Cu
P3	25-45	88.5	217-234	F3	99Sn-0.3Ag-0.7Cu

B. Rheological Measurements

All rheological measurement were conducted using a HAAKE rotovisco system comprises of sensor system PK100B and measuring system RV20. Principle of measuring cone-to-plate system at this system is in Fig. 2 Special care was taken while loading the solder paste sample onto measuring geometrics. For every measuring were used new sample of solder paste from the container. This step was repeated because solder paste was embossed out from the "between plate area" during repeated measuring or long time measuring. Also we tried to ensure the same conditions of measuring. Before starting the test, the sample was allowed to rest for the period at least 1 minute to allow the sample to relax and to reach a required temperature. Ideal loading procedures were followed for all the tests. Temperatures during the tests were hold at same level from the beginning to the end of test.

IV. Result and Discussion

A. Viscosity test results

Conditions for viscosity tests are in the Table 2.

Table 2 Options of measuring for viscosity test.

Angle of conic plate [°]	*Diameter of cone [mm]*	*Shear rate [s^{-1}]*	*Temperature range [°C]*	*Time duration [s]*
1	20	0-30	20-30	60

As show at Fig. 7, all of solder pastes samples have decreasing viscosity when temperature is increase. Highest decreasing of viscosity has sample P3, that's mean this type of solder paste is most sensitive for temperature. Sample P2 has lowest dependability on temperature rising. Viscosity of P2 can be more stable during stencil process, because rubbing squeegee on stencil can produce some heat, and warm up solder pastes (Fig. 5).

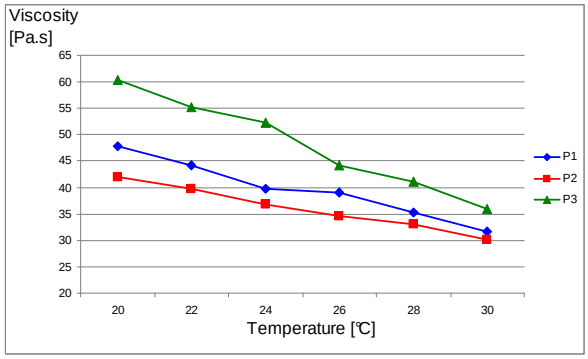


Fig. 7 Viscosity as a function of temperature for different solder paste samples. Shear rate was set up to 30 s^{-1}.

Study of the rheological behaviors of the solder paste helps us better understand behaviors of solder paste in stencil procedure. For our experiment we chose P1 sample.

Dependency of rising shear rate on decreasing viscosity is shown in the Fig. 8.

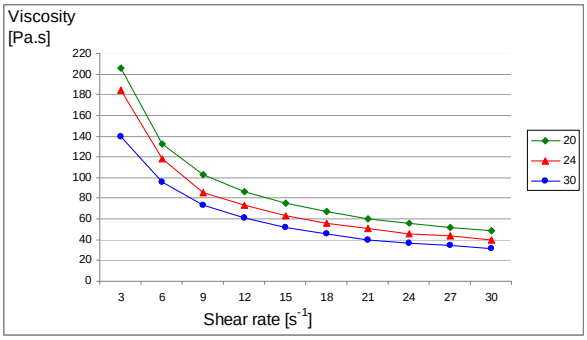


Fig. 8 Viscosity of solder paste sample P1 as a function of shear rate. Set ups of temperature were 20°C, 24°C and 30°C.

We can see were small influence of rising temperature on this effect is more remarkable in Fig. 9. Small shear rate (comparable with stencil printing) decrease has much more effect when temperature decrease. Viscosity decreasing more than 30 % when temperature rising from 20°C to 30°C at shear rate 3 s^{-1}.

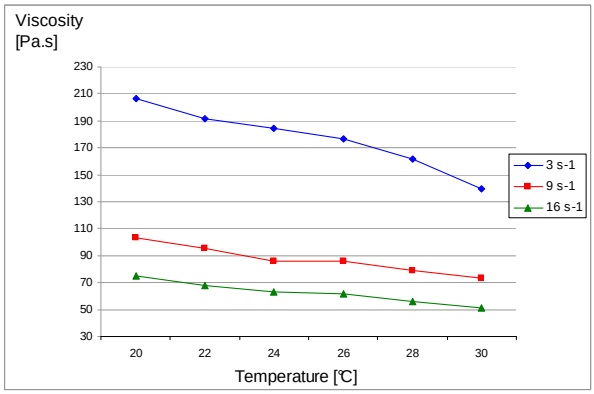


Fig. 9 Viscosity of solder paste sample P1 as a function of temperature. Measured were shear rates 3 s^{-1}, 9 s^{-1} and 16 s^{-1}.

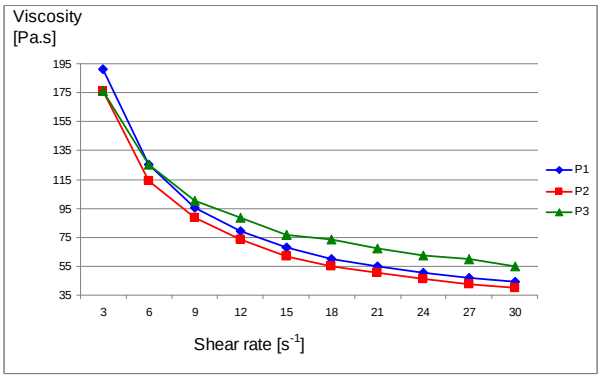


Fig. 10 Viscosity as function of shear rate for different solder paste samples. Temperature was set up to 24°C.

Fig. 10 shows the flow curves obtained for samples P1, P2 and P3, which show functionality viscosity on shear rate. All samples have decreasing trend during increasing of acceleration of measuring plate rotation. P1 sample has produced the highest maximum viscosity followed by P3 and P2 at low shear rate, but at higher shear rate sample P3 has lowest decreasing of viscosity. The differences in maximum viscosity for the solder paste samples can be attributed to the differences in flux systems. We can predict that solder paste type P3 has lowest dependability on printing speed and it can be more stable than samples P1 and P2.

B. Thixotropy test result

Conditions for thixotropy tests are in the Table 3.

Table 3 Option of measuring for thixotropy test.

Angle of conic plate [°]	*Diameter of cone [mm]*	*Shear rate [s^{-1}]*	*Temperature [°C]*	*Time duration [s]*
1	20	0-30	24	60 and 600

The results from the Thixotropy tests are presented in Fig. 11 at solder paste sample P1. Test shows the resulting viscosities as function of time(60 and 600 seconds). Both curves have decreasing course, but "600 seconds" curve is lower as expected. Long time shear rate action make solder paste sample easily to flow, that's mean viscosity of solder paste is depended also on time duration of the shear rate. This experiment confirm of memory properties of solder pate fluid.

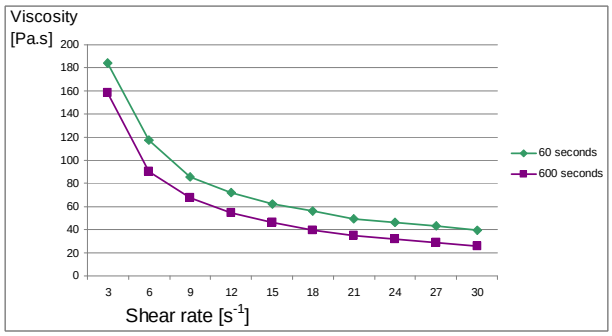


Fig. 11 Thixotropy test at solder paste sample P1. Temperature was set up to 24°C.

V. CONCLUSION

The solder pastes have been reported to be thixotropic, shear-thinning, and to possess a yield stress. The viscosity of solder pastes decrease with increasing temperature as well as with increasing shear rate.

Obtaining accurate rheological measurements on solder paste, help us to obtain the true rheological properties of solder pastes and the difficulties that arose in quantifying the viscometric parameters.

References

[1] Mallik, S., Thieme, J., "Study of the Rheological behaviors of Sn-Ag-Cu Solder paste and their Correlation with Printing Performance", Electronics Packaging Technology Conference, 2009.

[2] Ekere, N.N., Marks, A.E., Mallik, S., Seman, S., "Modeling the Structure Breakdown of Solder Paste Using the Structural Kinetic Model", Jurnal of materials Engineering and Performance, February , 2010.

[3] Schramm, G., "Introduction to Practical Viscometry", HAAKE Viscometers ,1981…

[4] Ecorel Free 305-6 Datasheet, http://www.inventec.dehon.com/

[5] ALPHA OM-338 Technical Bulletin, www.alphametals.com/products/

[6] ALPHA CVP-360 technical Bulletin, www.alphametals.com/products/

Comparison of the calculation of short-circuit currents in the various programs

[1]*Vladimír KRIŠTOF*, [2]*Stanislav KUŠNÍR*

[1]Dept. of Power System Engineering, FEI TU of Košice, Slovak Republic
[2]Dept. of Power System Engineering, FEI TU of Košice, Slovak Republic

[1]vladimir.kristof@tuke.sk, [2]stanislav.kusnir@tuke.sk

***Abstract*— There are a number of reasons why it is important to analyze short-circuit conditions in networks. One of the most important aspect is the safety and reliability of operation. The Most interesting are the minimum and maximum values of short circuit currents. Many of software equipments are used to calculate these values in practise. This article deals with comparing and verifying the results of short circuit currents calculation in the various programs.**

***Keywords*—Short-circuit current, short-circuit calculation, deviation.**

I. INTRODUCTION

It is necessary to know the short-circuit conditions in operating power system. It is important for a safe and reliable control and operation of the power system. Most interesting are the maximum values of short-circuit currents (for dimensioning of equipment) and minimal values of short-circuit currents (for setting-up of protection relays). The solved power system (transmission and distribution system) is usually very vast and complex system,it means that manual calculation would be very time-consuming and computationally intensive. Therefore for this reason a lot of software products (for example GLF, Daisy, Matlab, etc..) are used.

Each of the used programs has some accuracy solutions. STN IEC 60909 standard allows only a permissible deviation of the results (maximum 5%). The aim of this paper is to compare the results of calculating short-circuit current across the selected programs (GLF, DAISY, PSLF, NEPLAN) used in practice, and their comparison with the manual calculation of a simple network from literature [1].

II. MANUAL CALCULATION

Consider the following case. Power plant supplies the system, according to Fig. 1. 3-phase fault occurred in point F. The parameters of the system are:

Generator: S_{rG} = 400 MVA, U_{rG} = 21 kV, cosφ = 0.8, x''_d = 0,25, p_G = 0,05

Transformer:S_{rT} = 400 MVA, U_{THV} /U_{TLV} = 230 / 21 kV, u_k=15%

System: U_{qmin}=230 kV, U_{nq}=220 kV, c= 1,1

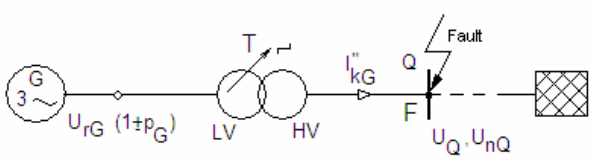


Fig. 1. Short-circuit contribution from power plant to point of fault (F).

The corrective factors according to STN IEC 60909 are considered in calculation [3,4,5].

For the short-circuit impedance calculation in electrical blocks with branches of the load the following relation is used:

$$\mathbf{Z_S} = K_S\left(t^2 \cdot \mathbf{Z_G} + \mathbf{Z_{THV}}\right) \quad (1)$$

Where Ks is corrective factor :

$$K_S = \frac{U_{nQ}^2}{U_{rG}^2} \cdot \frac{U_{TLV}^2}{U_{THV}^2} \cdot \frac{c_{max}}{1+\left(x''_d - x_T\right)\sin j_{rG}} \quad (2)$$

$$K_S = \frac{220.230}{21^2} \cdot \frac{21^2}{230} \cdot \frac{1{,}1}{1+(0{,}25-0{,}15)\sin 36{,}87^\circ} = 0{,}9926$$

Short-circuit impedance of generator :

$$Z_G = x''_d \frac{U_{rG}^2}{S_{rG}} = 0{,}25\frac{21^2}{400} = 0{,}276\ \Omega \quad (3)$$

Short-circuit impedance of transformer :

$$Z_{THV} = u_k \frac{U_{THV}^2}{S_{rT}} = 0{,}15\frac{230^2}{400} = 19{,}836\ \Omega \quad (3)$$

Short-circuit impedance of electrical block:

$$\mathbf{Z_S} = K_S\left(t^2 \cdot \mathbf{Z_G} + \mathbf{Z_{THV}}\right) = 0{,}9926\left(\left(\frac{230}{21}\right)^2 \cdot 0{,}276 + 19{,}836\right) = 52{,}552\ \Omega \quad (4)$$

Initial symmetrical three-phase short-circuit current:

$$I''_{kG} = \frac{c \cdot U_n}{\sqrt{3} \cdot Z_S} = \frac{1{,}1 \cdot 220}{\sqrt{3} \cdot 52{,}552} = 2{,}659\ kA \quad (5)$$

III. CALCULATION IN PROGRAMS

A. GLF

GLF (Graphical Load Flow) is simple and overview program. It is intended to solution mainly high voltage, very high voltage, ultra high voltage networks. It is used for:

- steady state calculation
- analysis of voltage conditions
- analysis of short-circuit conditions
- checking the reliability of network operations by the criterion (n-1)

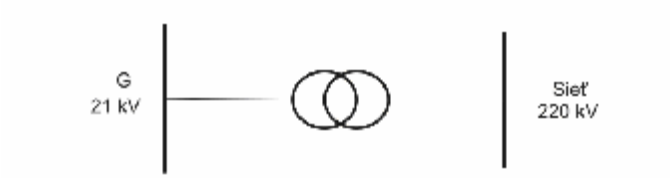


Fig. 2. Network model in GLF

GLF uses Fortescue method of symmetrical components to short-circuit calculation. The short-circuit current is 2,65 kA.

B. PSLF

Software package PSLF (Positive Sequence Load Flow) is a suite of programs for the analysis grid (transmission system). It allows the calculation of the steady state, as well as transitional phenomena (short circuit, ground connections or dynamic stability).

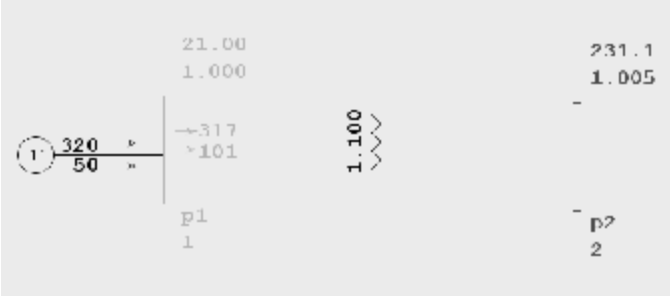


Fig. 3. Network model in PSLF

The short-circuit current is 2, 713 kA.

C. Pass Daisy (Bizon)

Daisy is package of programs used in preparation of operations, planning for further development, design, evaluation and operation of networks. It is characterized by enhanced supply of calculation method.

Fig. 4. Network model in Pass Daisy

Daisy offers the option to choose the method of calculating short-circuit currents:

- calculation according to STN IEC 60909
- calculation according to ČSN
- calculation according to Daisy

Method by STN IEC 60909 was used for this calculation . The short-circuit current is 2,79 kA.

D. Neplan

Neplan is very user-friendly program, serving on the planning and calculation of electric, gas or water supply networks.

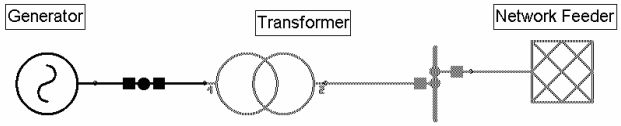


Fig. 5. Network model in Neplan

Neplan also offers the option to choose the method of calculating short-circuit currents:

- calculation according to STN IEC 60909
- calculation according to Neplan
- calculation according to ČSN

The short-circuit current is 2,723 kA.

.

IV. COMPARISON OF RESULTS

Obtained results are listed in Table 1. As the reference ("exact") value the short-circuit current value calculated manually was chosen. Deflection were calculated by the following pattern :

$$D = \frac{I_{kB} - I_{kA}}{I_{kA}} * 100\%$$

Where:

- I_{kB} is calculated value in a program
- I_{kA} is the reference value

	I''_{k3} [kA]	Deviation [%]
Manual calculation	2,659	-
GLF	2,65	-0,33
Pass Daisy	2,79	4,9
PSLF	2,713	2,03
Neplan	2,723	2,4

Tab. 1. Comparision of results of short-circuit calculations in various programs

V. CONCLUSION

Calculation and comparison of the results was focused on three phase short-circuit current value because in most cases that value takes into account for dimensioning equipment and setting up protection relays. A deviation with vaule (+/-) 2% can be neglected. The maximum possible deflection allwos (+/-) 5%. Deviation are mainly due to the following facts, that none of the programs consider corrective factors, which were considered in manual calculations (according to STN IEC 60909). STN IEC 60909 standard is conservative, which means that sets strict conditions for calculating short-circuit current through the correction factors. With these corrective factors is achieved better dimensioning and also setting up of protection relays. That is very important for safety and reliable operation of power system.

REFERENCES

[1] Mešter, M., –Výpočet skratových prúdov v trojfázových striedavých sústavách. ABB-elektro, s.r.o., 2005. ISBN 80-89057-10-1

[2] Krištof, V.,– Výpočet skratových pomerov podľa STN IEC 60909. Diplomová práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2009.

[3] STN IEC 60909-0: Skratové prúdy v trojfázových striedavých sústavách. Časť 0: Výpočet prúdov. Slovenský ústav technickej normalizácie, apríl 2003.

[4] STN IEC 60909-1: Výpočet skratových prúdov v trojfázových striedavých sústavách. Časť 1: Súčinitele na výpočet skratových prúdov v trojfázových striedavých sústavách podľa IEC 60909. Slovenský ústav technickej normalizácie, august 2000.

[5] STN IEC 609009-2: Elektrické zariadenia. Časť 2: Údaje na výpočet skratových prúdov podľa STN IEC 60909. August 2000.

[6] MITOLO, Massimo: SHORT-CIRCUIT CALCULATION METHODS, 10/2004 Dostupné na internete: <http://ecmweb.com/mag/electric_shortcircuit_calculation_methods/>.

Power flows control in electric power systems

[1]*Stanislav KUŠNÍR*, [2]*Vladimír KRIŠTOF*

[1]Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic
[2]Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]stanislav.kusnir@tuke.sk, [2]vladimir.kristof@tuke.sk

***Abstract*— The paper presents the knowledge of the regulation of flows performance in the electricity systems. Modeling flows performance using PST transformers presents the practical example, which is created in the GLF/AES program.**

***Keywords*— power flow, phase shift transformer, transmission lines**

I. INTRODUCTION

The problem of regulation of load flow the power system is currently becoming more and more professionals in the topics discussed. There is a need to regulate the load flow on the lines, which are linked by various electrification systems; this is associated with the gradual liberalization of the electricity market. Since commercially negotiated power flows are significantly different from the actual flows of trade laws and do not exceed the laws of physics, the question arises how to bring near these laws.

As a result, of electricity trading has been increasing interstate transfers, often because of what some of the lines to the state, which are surcharge, while others are not fully utilized. Interstate lines were not constructed for the purpose of trading with electricity, but for mutual emergency assistance, increase in operative security, reducing the necessary power reserves and improvement conditions regulation frequency. In extreme cases, it could happen that the line operating near their limits, will be exceeded this limits and disconnected action protections. Turn off these lines means that lines were not fully utilized and will become congested and their disconnection. This may result a power outages and great economic losses in the given area. To arrive to an effective operation of lines and prevent congestion interstate lines we can with use devices designed to power flows control.

II. PRESENT SITUATION IN SLOVAK REPUBLIC

In the event certain modes the conditions (when you shut down the profile between the APG and CEPS), there is a bottleneck between SEPS - MAVIR.

The problem of bottlenecks of own line only profile SEPS - MAVIR, but it is a Pan-European problem, which solves research teams for nearly all electric systems.

Export possibilities of some European countries now exceed the line capacity, which may result in the occurrence of other danger bottlenecks. By one of possible solution how to removal of the bottlenecks is construction of new lines. It is however time consuming, and the pace of construction is lower than the pace of increase in trade. This forces operators of transmission lines approach to other solutions. Offered solutions are in regulation, respectively, influence load flow by using classic performances, but also technical advanced options. Available options are based on the classic transformers or elements power electronics. For reasons given results that the in following years an increase need for application of special technical means intend for influenced the size and direction of the transmitted flows in the network.

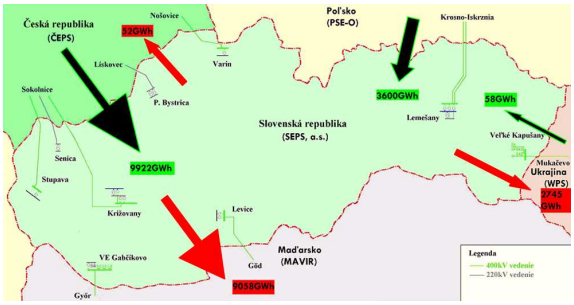


Fig. 1. Measured cross-border transfers of electricity (year 2007)

From Fig. 1 is evident, that the transmission lines of Slovakia are loaded international transits, especially direction from north to south. For this direction of transfer lines were not built in the past. The direction of the transmitted power is focused more on the direction from east to west. Slovakia along with Poland and the Czech Republic had a surplus of electricity which could be transported to south, to countries that are not energy self-existent.

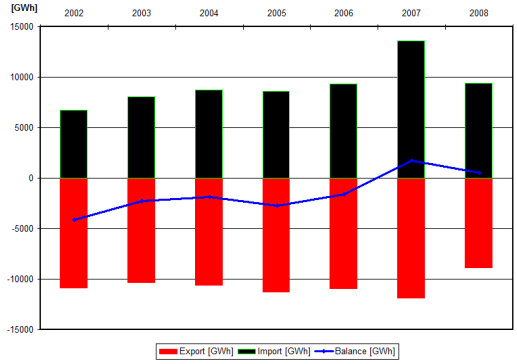


Fig. 2. Import, export and balance

In the last quarter of 2008 is already partially started to show effects of the economic crisis. With the gradual fall in economic performance of the Slovakia occurred also a decline in electricity consumption.[4]

III. Expectation further development of interstate exchanges in Europe

In the following years, it can be assumed that in central Europe the interstate exchange of electricity will have increasing trend. One of the main reasons for this trend is the pressure of traders onto realization of deliveries from sources strong areas with lower price level.

Future developments in the electricity supply will be affected following factors and risks:

- Growth in electricity consumption after the end economic and financial crisis.
- Availability fuels and their price developments at world markets.
- Price developments at the electricity markets.
- Development of price increases in the sphere of new produce technologies.
- Uncertainties associated with setting charges for emissions, especially CO_2.
- Long-term return of investment in realization of projects in electricity sector.
- Pressure to increase the share of wind and solar power plants on covering the graph load.

IV. Means for power flow control in electric power systems

A. *Present means to regulate the power flow*

Their use is not such effective as to using of new resources. However, in electric power system Slovak republic are the only present resources using for power flow control.

Present resources are:

- Influence working sources,
- control consumption,
- changing network topology,
- severance areas supply

B. *New means to regulate the power flow*

With yearly growth of interstate transits of electric energy, the present means will be necessary to replace newer and more efficient devices, whether based on semiconductor components or on special transformers.

New means are [1]:

- **HVDC** - High Voltage Direct Current
- **FACTS** - Flexible Alternating Current Transmission System
- **PST** - Phase Shifting Transformer
- **TPR** - Transformer with cross regulation

V. Simulation PST at the 400kV lines between Slovak Republic and Hungary

In the event certain modes the conditions resorts to the bottleneck on the profile between SEPS - MAVIR. One of possibilities to eliminate this bottleneck is to install devices on power flow control.

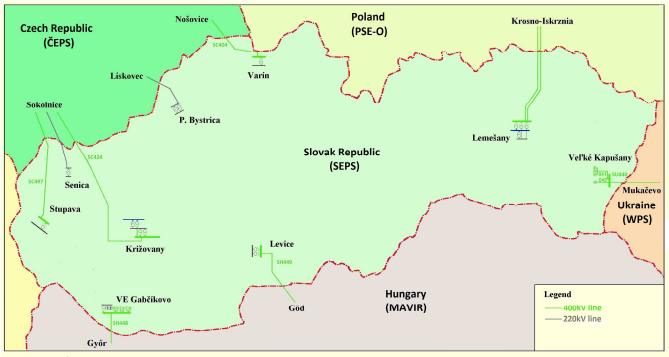


Fig. 3. Interstate lines of the Slovak Republic

A. *Phase shift transformer was connected in line SH448*

When PST_1 was involved to the line SH448 reached to the change the power flow by individual lines. The biggest change of power flow was on line, in which was connected PST and on line SC497. Raising transformer taps that generated line SH448 was derated power and contrary bucking generated to bigger load. By controlling load flow was reached increase transmission losses throughout electric power system.

TABLE. 1
POWER FLOWS DEPENDING UP SETTINGS TRANSFORMER TAP PST_1

	P [MW]							
Hran. vedenie	SH448	SH449	SU440	SP477	SP478	SC404	SC424	SC497
Bez PST	-419,6	-329,5	-16,7	248,2	249,4	482	203,7	288,9
Nast. odbočka PST_1								
PST -6	-501,3	-305,6	-8,8	248,1	249,2	487,3	217,5	319
PST -5	-468,1	-316,7	-12,2	248,1	249,3	485,2	212,2	307,2
PST -4	-434,7	-326	-15,6	248,2	249,3	483	206,9	295,4
PST -3	-401,2	-335,5	-19	248,2	249,4	480,9	201,6	283,5
PST -2	-367,7	-344,9	-22,4	248,3	249,4	478,8	196,2	271,7
PST -1	-334,3	-354,3	-25,8	248,4	249,5	476,7	190,9	259,8
PST 0	-300,8	-361,5	-29,2	248,4	249,6	474,6	185,6	247,9
PST +1	-267,5	-373,1	-32,6	248,5	249,7	472,6	180,3	236,1
PST +2	-234,3	-382,4	-35,9	248,6	249,8	470,5	175,1	224,3
PST +3	-201,3	-391,7	-39,3	248,7	249,9	468,5	169,9	212,7
PST +4	-168,6	-400,9	-42,6	248,8	249,9	466,5	164,7	201
PST +5	-136,1	-410,1	-46	248,9	250	464,6	159,6	190,1
PST +6	-104	-419,1	-49,2	249	250,1	462,6	154,5	178,1

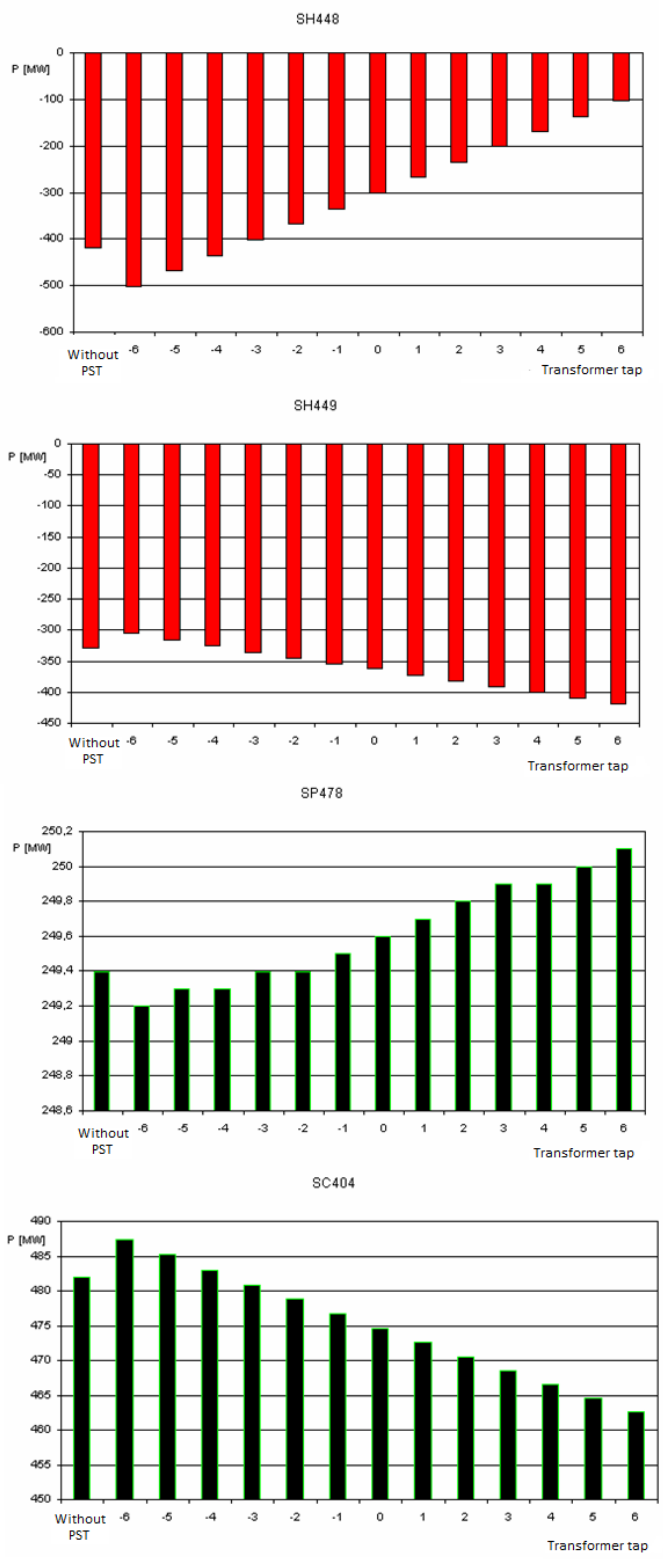


Fig. 4. Power flows in lines SH448, SH449, SP478 and SC404

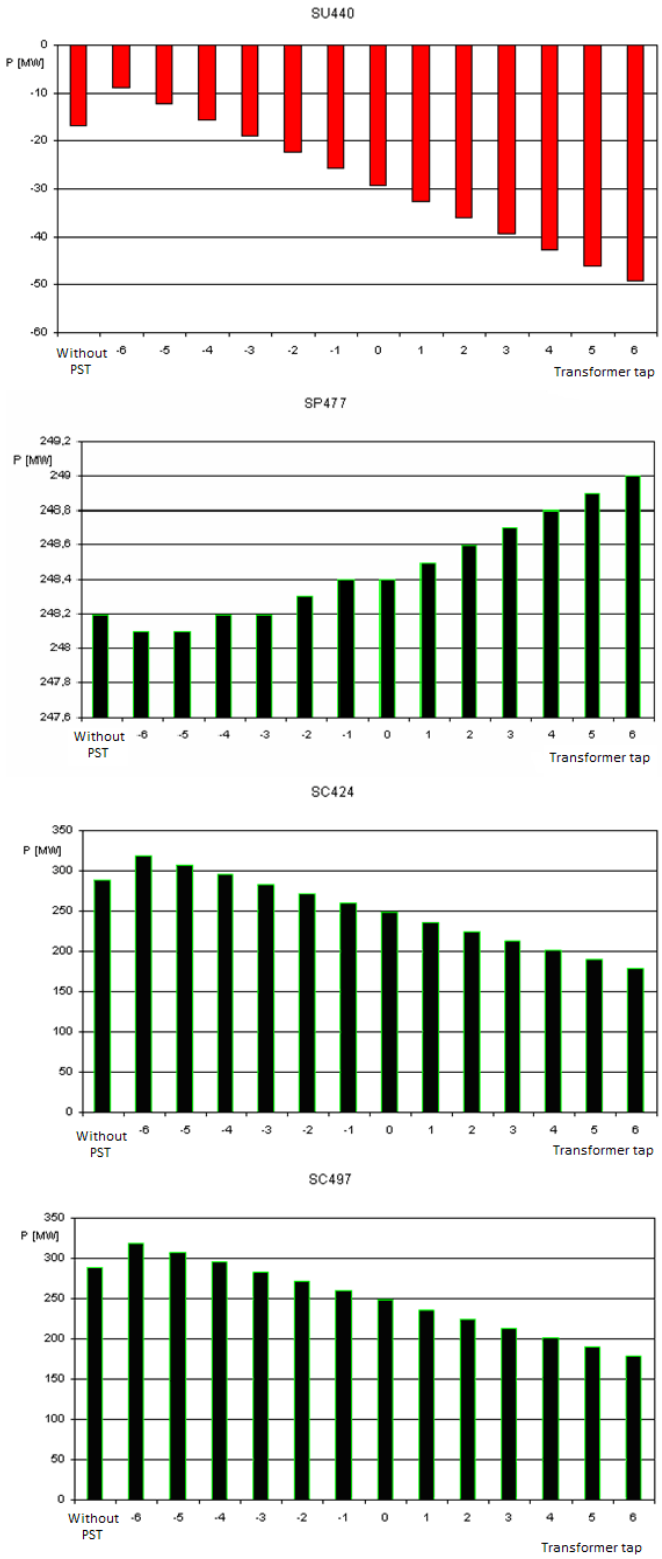


Fig. 5. Power flows in lines SU440, SP477, SC424 and SC497

B. Phase shift transformer was connected in line SH449

When PST_2 was involved to the line SH449 reached to the change the power flow by individual lines. To the biggest change power flow was on lines SH449 and SU440. Raising transformer taps that generated line SH449 was derated power and contrary bucking generated to bigger load. By controlling load flow was reached increase transmission losses throughout electric power system.

TABLE.2.
POWER FLOWS DEPENDING UP SETTINGS TRANSFORMER TAP PST_2

	P [MW]							
Hran. vedenie	SH448	SH449	SU440	SP477	SP478	SC404	SC424	SC497
Bez PST	-419,6	-329,5	-16,7	248,2	249,4	482	203,7	288,9
Nast. odbočka PST_2								
PST -6	-386,1	-436,2	29,4	246,5	247,6	493,4	217,4	290
PST -5	-395,9	-404,9	16	247	248,1	490,4	213,4	289,6
PST -4	-405,7	-373,6	2,6	247,4	248,6	486,8	209,4	289,3
PST -3	-415,5	-342,3	-10,9	247,9	249,1	483,5	205,4	289
PST -2	-425,3	-311	-24,3	248,4	249,6	480,2	201,5	288,7
PST -1	-435,1	-279,8	-37,7	248,9	250,1	476,9	197,5	288,4
PST 0	-444,9	-248,7	-51,1	249,5	250,6	473,7	193,6	288,1
PST +1	-454,6	-217,7	-64,5	250	251,1	470,4	189,7	287,8
PST +2	-464,3	-187	-77,8	250,5	251,7	467,3	185,8	287,5
PST +3	-474	-156,4	-91	251	252,2	464,1	182	287,2
PST +4	-483,5	-126	-104,2	251,6	252,7	461	178,2	287
PST +5	-493	-95,9	-117,2	252,1	253,3	457,9	174,4	286,8
PST +6	-502,4	-66,2	-130,1	252,6	253,8	454,8	170,7	286,5

C. Cooperation PST_1 and PST_2

Cooperation PST_1 and PST_2 was occurred to expressive changes in load flow at almost all interstate lines, with the exception of SP477 and SP478, unlike previous cases where the actual PST_1 varied mainly flows of lines: SH448, SH449, SC424 and SC497, and the actual PST_2 varied mainly flows of lines: SH448, SH449, SU440, SC404 and SC424.

VI. CONCLUSION

The paper was dealt possibilities of power flows control in the electric power system of the Slovak Republic. Phase shifting transformers were analyzed and their impact to power flow. When deciding on the appropriateness of installing such devices would be based on detailed economic and technical analysis from the perspective of the future.

PST transformers are cheaper, easier to operate compared with FACTS devices, but on the other hand FACTS devices are flexible and contribute to the improvement of static and dynamic stability. Special transformers are manufactured only by specific customer requirements. The price depends on the requirements and on the power.

REFERENCES

[1] Ptáček, J.: *Regulace výkonů v propojených elektrizačních soustavách*. [Dizertačná práca]. Brno : FEKT VUT v Brne. 201 s. 2004.

[2] Rusnák, J.: Použitie nových prostriedkov v riadení prevádzky elektrizačnej sústavy. In: Elektroenergetika 2003 – zborník prednášok II. Medzinárodného vedeckého sympózia, Vydavateľstvo: Smékal Publishing, 2003, ISBN 80-89061-80-X

[3] Rusnák, J., Kolcun, M., Mészáros, A.: Transformátor s uhlovou reguláciou – nástroj pre reguláciu toku výkonov v elektrizačnej sústave In: *EE - Odborný časopis pre elektrotechniku a energetiku*, roč. 9, mimoriadne číslo, 2003. Bratislava : Spolok absolventov a priateľov FEI STU (EF SVŠT) v Bratislave, 2003, s. 18-19. ISSN 1335-2547.

[4] Správa o výsledku monitorovania bezpečnosti dodávok elektriny. [Online]. Available on the Internet: <http://www.economy.gov.sk/sprava-o-vysledku-monitorovania-bezpecnosti-dodavok-elektriny--jul-2009-/130669s

Influence of Grounding Point of Coil to Formation of Surface Discharges

[1]*Milan KVAKOVSKÝ*, [1]*Lýdia DEDINSKÁ*, [1]*Vieroslava ČAČKOVÁ*

[1]Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]milan.kvakovsky@tuke.sk, [2]lydia.dedinska@tuke.sk, [3]vieroslava.cackova@tuke.sk

***Abstract*—One of the methods of diagnosis, indicating the quality of the stator winding insulation of electrical rotating machines called the measurement method of partial discharges. The measurements of electrical machines in operation we can obtain the distribution phase partial discharges and the value of apparent charge, under which it is necessary to assess the quality of the equipment under test. In the laboratory condition performing measurements of partial discharges on stator windings models with various disorders insulation system for assessing the quality of insulation system. The article highlighted the impact of potential natural connection point for the coil to rise to charges on the surface coil insulation.**

***Keywords*—partial discharges, coils, electrical rotating machines.**

I. INTRODUCTION

The most important part of the grid is generators that produce electricity at voltage levels ranging from 3.15 kV to 15.75 kV. Electricity produced by the voltage HV is transformed into voltage by the network, which provides long distance transmission [1].

Stator insulation is an important part of rotating electrical machines. The most common failure is the failure of the machine insulation system. This requires regular measurements of diagnostic equipment. Measurement of partial discharges is one of the measurements, which can assess the condition of insulation and the whole system.

For the power supply is reliable it is necessary also to limit the disturbances that arise in the actual generator. With the increasing performance in the power system there are greater demands for electrical machines in it working. This increases the quality and durability of insulation depending on operating conditions and production technology from [2, 3]. The most stressed part of the generator is the stator insulation. During operation is not exposed only to electrical stress but also mechanical, thermal and chemical stress. These degradation effects are caused by deterioration of electrical insulation properties and mechanical equipment to malfunction and failure of the machine operation. To prevent this emergency, it is necessary to limit the effects of degradation on the isolation and track changes in the insulating state at regular intervals.

Regular maintenance of generators and diagnostic measurements can help prevent malfunctions and to extend the life of machines.

Fig. 1. Location of the test coil to measure partial charges

II. MATERIAL AND METHODS

Measurement of partial discharges in insulation of stator windings of electrical machines is one of the diagnostic methods pointing to the quality of insulation system. Insulation system failure modeling and measurements in laboratory conditions it is possible to obtain phase distribution discharge activities that facilitate the detection of failures in service.

To be credible the results of measurements should be thoroughly investigated to know the object place of work the partial discharges and correct modeling of the faulty story. In the case of stator insulation can occur as internal as well external partial discharges. The external discharge includes the discharge outlet of the groove and the coil discharges in stator slot. The operation of these discharges often occurs simultaneously.

To study the development of partial discharges developing at the outlet of the reel slot was used new 6 kV stator coil coated with a conductive protection it is a part of the coil that is inserted into the stator slot. The ends of the windings are interconnected and were placed on high voltage potential. Natural potential is fed to the point of interconnection of two sides with conductive coil protection. It was tested five different grounding involved. Place earth is changed as follows:

- Coil grounded at the top (the farthest from the high voltage electrodes),
- Coil grounded 3.5 cm from the upper edge,
- Coil grounded in the central part (see Fig.1),
- Coil grounded 3.5 cm from the bottom edge,

- Coil grounded all the way down (closer to the high voltage electrode).

Galvanic direct method was use to measurement of partial discharges. Block diagram of involvement is used in Fig. 2. Oscilloscope monitors the shape and location of partial discharges pulses. The advantage of this involvement is that in the case of breakdown of measured object does not damage the measuring instrument.

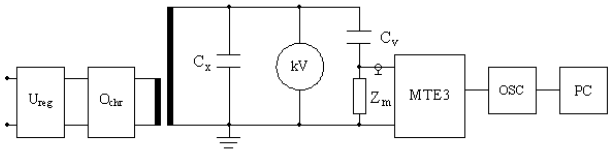


Fig. 2. Diagram for the direct method

U_{reg} - adjustable voltage source,
U_x - replacement of measured samples,
C_V - binding capacity,
Z_m - impedance.

III. RESULTS AND DISCUSSION

The measurement was carried out under laboratory conditions. The test coil was suspended on rope in isolation Faraday cage. The ends of the windings are interconnected and were placed on high voltage potential using spherical electrodes.

The test voltage is gradually increased until the emergence of early discharges in this time was launched a program to record the results of measurements and made the first measurements of partial discharges. Further measurements were made in increasing the voltage step to 200 V to the nominal value 6 kV. Each measurement lasted 3 minutes which was recorded for more than 900 periods of applied voltage.

By increasing the voltage in step 200 V from the initial value of partial charges to the nominal value of 6 kV was obtained voltage dependence of the characteristic parameters of partial discharges for each type of conductive grounding point (see Fig. 3).

The results were processed evaluated and stored by a computer program.

In order to compare the emergence and development of surface discharges depending on the place of earth coils were observed following features incremental charges:

- Maximum apparent charge of partial discharges,
- Mean apparent charge of partial discharges,
- Number of partial discharges,
- summing charge.

The following table shows the maximum value of apparent charge depending on the accompanying voltage for different places of the earth with conductive coil protection where:

- a - Coil grounded at the top (the farthest from the high voltage electrodes),
- b - Coil grounded 3.5 cm from the upper edge,
- c - Coil grounded in the central part (see Fig.1),
- d - Coil grounded 3.5 cm from the bottom edge,
- e - Coil grounded all the way down (closer to the high voltage electrode).

TABLE I
COIL GROUNDED AT THE TOP

U [kV]	Qmax [pC]	Nstr [-]	φ+ [°]	φ- [°]
3,6	2800	0,15	30-80	230-270
3,8	3500	0,3	30-80	230-270
4,0	4000	0,25	50-80	230-270
4,2	5500	0,6	50-80	225-270
4,4	6000	0,8	40-80	220-270
4,6	7000	1,5	30-80	210-280
4,8	7000	1,2	30-80	210-280
5,0	7000	1,8	30-80	210-280
5,2	9000	1,4	30-80	210-280
5,4	9000	1,9	10-80	200-280
5,6	13000	0,9	30-70	210-280
5,8	13000	1,2	30-70	210-310
6	13000	1,4	30-70	200-310

TABLE II
COIL GROUNDED 3.5 CM FROM THE UPPER EDGE

U [kV]	Qmax [pC]	Nstr [-]	φ+ [°]	φ- [°]
3,6	2500	0,08	50-60	225-250
3,8	3000	0,3	30-70	220-260
4,0	6000	0,25	30-70	220-270
4,2	4000	0,32	50-70	220-270
4,4	5500	0,6	30-60	210-250
4,6	6500	0,8	40-70	210-270
4,8	7500	1	40-70	210-270
5,0	8000	1,1	30-80	210-260
5,2	9000	0,5	-	210-270
5,4	8000	0,6	-	210-270
5,6	9000	0,6	-	210-300
5,8	9000	0,7	-	210-300
6,0	9000	0,7	-	200-290

TABLE III
COIL GROUNDED 3.5 CM FROM THE UPPER EDGE

U [kV]	Qmax [pC]	Nstr [-]	φ+ [°]	φ- [°]
3,4	120	0,1	10-70	210-290
3,6	400	0,15	30-60	210-250
3,8	400	0,6	30-70	220-260
4,0	900	0,4	30-70	220-260
4,2	1100	1	30-70	220-260
4,4	2700	0,5	30	220-280
4,6	3500	0,7	30-60	220-270
4,8	3500	1,6	30-60	210-270
5,0	4000	0,2	-	220-270
5,2	4000	0,17	-	220-270
5,4	4000	0,15	-	220-260
5,6	5000	0,25	-	220-260
5,8	11000	0,35	-	210-280
6,0	9000	0,7	-	210-270

TABLE IV
COIL GROUNDED 3.5 CM FROM THE UPPER EDGE

U [kV]	Qmax [pC]	Nstr [-]	φ+ [°]	φ- [°]
3,4	350	0,08	30-70	230-250
3,6	2000	0,12	50-70	220-250

3,8	2500	0,15	40-70	220-250
4,0	2700	0,4	20-70	220-270
4,2	4000	0,25	40-70	220-280
4,4	4000	0,25	60	210-300
4,6	8000	0,4	20-70	220-280
4,8	13000	0,3	30	220-300
5,0	14000	0,45	30	220-280
5,2	13000	0,6	20-50	210-260
5,4	13000	0,6	30-60	210-260
5,6	15000	0,6	20-60	200-300
5,8	15000	0,7	20-60	200-290
6,0	15000	0,6	20-80	190-300

TABLE V
COIL GROUNDED ALL THE WAY DOWN

U [kV]	Qmax [pC]	Nstr [-]	φ+ [°]	φ- [°]
3,6	3500	0,2	30-70	210-300
3,8	3500	0,2	30-80	220-280
4,0	3500	0,2	30-80	210-300
4,2	7000	0,25	30-90	220-300
4,4	10000	0,6	20-100	210-280
4,6	15000	0,25	30-70	220-290
4,8	18000	0,3	30-70	220-310
5,0	20000	0,3	20-60	210-300
5,2	20000	0,6	10-80	210-300
5,4	18000	0,7	30-80	210-300
5,6	18000	1	20-80	210-300
5,8	20000	1,3	20-90	210-320
6,0	20000	0,8	20-100	210-300

IV. CONCLUSION

The measurement results show that the initial discharge voltage levels resulting from the 3.4 and 3.6 kV depending on the place of grounding the coil. When voltage was 3.4 kV Discharge activity recorded unstable. One is likely to discharge internally which will be activated in isolation tubes. The value of the voltage 3.6 kV has created a stable surface discharge which further increases grow around the surface coil and the value of 5 kV is possible to capture the sound of corona [4, 5].

When comparing the results obtained at 3.6 kV voltage level can be said that the maximum apparent charge and the frequency of discharges are greater in the negative half-wave applied voltage. The high-voltage coils stored in the stator grooves where the groove part is grounded and the pin is brought high voltage discharges occur at the exit from the grooves or coils. Therefore the discharges in the negative voltage half-wave are greater than positive. The lowest maximum value of 400 PCs apparent charge (when the voltage level of 3.6 kV) were recorded for grounded coils in the middle part. Proximity to high voltage grounding point electrode increases the amplitude of apparent charge on the value of 2000 PCs (coil grounded at 3.5 centimeters from the bottom edge) and proximity to other high-voltage electrode roll down completely grounded increased amplitude of apparent charge on the value of 3500 PCs. The earthed coils 3.5 cm from the upper edge of the amplitude of apparent charge, reaching 2500 for PCs and grounding at the top value in 2800 PCs.

Situation of 6 kV in terms of phase distribution remained similar as in the case of 3.6 kV voltage levels. Discharges activity increased in negative half-wave than positive half-wave. The grounded coils in the upper part of the amplitude of discharges reached 13,000 PCs. The grounding of 3.5 cm below it was 9000 PCs. Amplitude of discharges in the coil which was grounded in the middle was 9000 PCs. Amplitude of discharges increased for coil grounded 3.5 cm from the bottom edge of the value of 15,000 PCs and followed a further increase to 20,000 PCs for coil grounded in the bottom of coils.

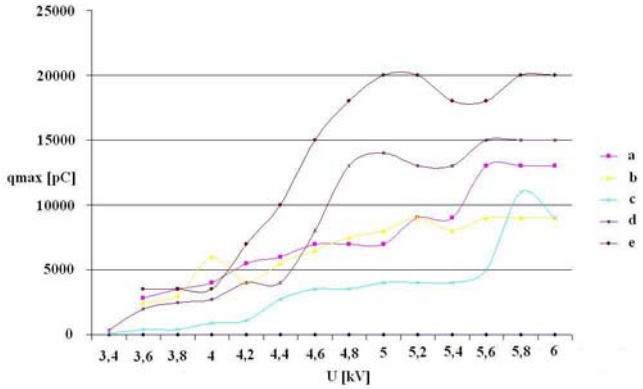


Fig. 3. Dependence of the maximum apparent charge of the voltage.

V. ACKNOWLEDGMENT

This work was supported by scientific Grant Agency of the ministry of Education of the Slovak Republic project VEGA No. 1/0368/09 and APVV-20-006005

REFERENCES

[1] M. Kolcun – V. Chladný – M. Mešter – R. Cimbala – J. Tkáč – M. Hvizdoš – J. Rusnák, "Elektrárne," Technická Univerzita v Košiciach. 2006. ISBN 80-8073-704-5

[2] K. Záliš, "Evaluation of Partial Discharge Activity by Expert Systems," In: 11th International Symposium of High Voltage Engineering, ref. 5.344.P5, London, Sept.1999.

[3] P. Toman, "Simulation of Instrument Transformer Properties in Power Systems," In:Electric Power Engineering, Brno, Czech Republic, 2006.

[4] KOLCUNOVÁ, Iraida - KURIMSKÝ, Juraj - BALOGH, Jozef: Meranie čiastkových výbojov na vn cievkach. In: Diagnostika '07 : Mezinárodní konference, Nečtiny 11.-13. září 2007. Plzeň : Západočeská univerzita, 2007. p. 34-37. ISBN 978-80-7043-557-1.K. Záliš, "Evaluation of Partial Discharge Activity by Expert Systems," In: 11th International Symposium of High Voltage Engineering, ref. 5.344.P5, London, Sept.1999.

[5] PETRÁŠ, Jaroslav - DŽMURA, Jaroslav - BALOGH, Jozef: Analýza signálov získaných z merania akustickej emisie čiastkových výbojov. In: Starnutie elektroizolačných systémov. č. 4 (2008), s. 21-23. Internet: <http://web.tuke.sk/fei-kee/jses/uploads/File/jses-04-2008.pdf> ISSN 1337-0103.K. Záliš, "Evaluation of Partial Discharge Activity by Expert Systems," In: 11th International Symposium of High Voltage Engineering, ref. 5.344.P5, London, Sept.1999.

LOAD TORQUE EMULATOR BASED ON INDUSTRIAL CONVERTERS

Karol KYSLAN

Dept. of Electrotechnics, Mechatronic and Industrial Engineering, FEI TU of Košice, Slovak Republic

karol.kyslan@tuke.sk

***Abstract*— Technology of hardware-in-the-loop simulation becomes a standard in development of control algorithms for mechatronic systems. The article describes HIL simulator consisting of RT-LAB system with MATLAB/Simulink interface and commercial drive converters. This arrangement is used for control of load torque emulator. Emulator enables testing of an electrical drive without real mechanical load where the load and inertia influence of mechanics are emulated through a load torque.**

***Keywords*— hardware–in–the-loop, industrial drives, load torque emulator, RT-LAB**

I. INTRODUCTION

Standard verification of control structures by the mathematical modelling is improved when using one or several actual devices instead of their simulation models. The other parts of the process are simulated in appropriate real-time system. Hardware running real-time system equipped with DAQ interface boards in conjunction with actual devices creates HIL simulation, which is nowadays more and more used to develop new structures and components in many fields. Methodology of HIL yields exhaustive testing of a control system to prevent costly and damageable failures. Moreover, it reduces development time and can enable more tests than on the actual system. Three different kinds of HIL simulation for electrical drives have been proposed in [1]: signal-level, power-level and mechanical-level simulation. HIL simulator controlling load torque emulator represents the last one kind of hardware-in-the-loop simulation.

II. MECHANICAL-LEVEL HIL SIMULATION FEATURES

Assuming an electrical drive decomposed into the process control, the power electronic set, the electrical machine and the mechanical load to load the electrical machine. All the drive excluding load is tested for widely range of operating conditions.

In order to simulate the behaviour of the various mechanical loads, the shaft of the tested electrical machine is connected to another electric machine (load machine) supplied by its own power electronic set. A second controller board in a form of real-time simulation platform is required to control the load machine. Advisable arrangement of model in real-time simulation and DAQ interface provides way to impose required mechanical quantities into common rigid shaft. Quantities are mechanical and so method can be called „mechanical–level“ HIL simulation [1]. In Fig. 1 the schematic form of mechanical-level HIL simulation is depicted. This kind of HIL simulation enables intensive tests on a static experimental benchs for testing of control of single motor and multimotor electrical drives, evaluation of vehicle and hybrid vehicle components, railway traction systems and robotics. Moreover, it can be used for teaching and educational applications.

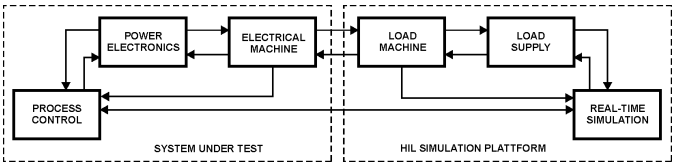


Fig. 1. Mechanical-level HIL simulation

III. HIL SYSTEM STRUCTURE WITH LOAD TORQUE EMULATOR

There are several platforms providing HIL simulation (Real-Time Toolbox of Matlab, dSpace products, Opal RT products etc.). Decentralized Opal RT simulation platform RT-LAB was used because of its possibility to run on the common PC's.

The emulator consists of a motor twin pair where the motors are coupled by a rigid shaft. The DC motor presents a real tested drive that is supplied by the thyristor rectifier of Simoreg DC Master 6RA70 (Siemens). The induction machine emulating behaviour of the load is supplied by the converter Simovert Master Drives 6SE70 (Siemens). Both converters are 4Q. Communication between converters uses the fast optical industry communication tool SIMOLINK. The Simovert frequency converter communicates with superimposed control system through the CAN bus.

Computer equipment of the emulator [2], [3] consists of two personal computers running RT-LAB: host and target. The host PC provides a model development under Matlab/Simulink tools. During the simulation the host PC acts as a console and provides data visualisation and eventually parameter changes. Compiled code is then transferred to the target PC. Real – time simulation runs on the target PC under QNX real time operation system. QNX is type of UNIX OS

and it is unfailing operation system for control of industrial and technological applications and even real-time processes and embedded systems. The target PC can be equipped by interfaces to real environment: in our case a communication card CAN-ACx-PCI for CAN bus was applied (Softing).

Presented emulator is based on a pre-control for the testing drive [4], [5] that is easier realizable than inverse transfer function of the dynamic equation usually used in emulators [6]. The emulator forces the dynamics to the tested drive but it is done through the load torque and not through the control circuits like it is used in forced dynamics systems. The HIL system structure with load torque emulator is shown in Fig.2.

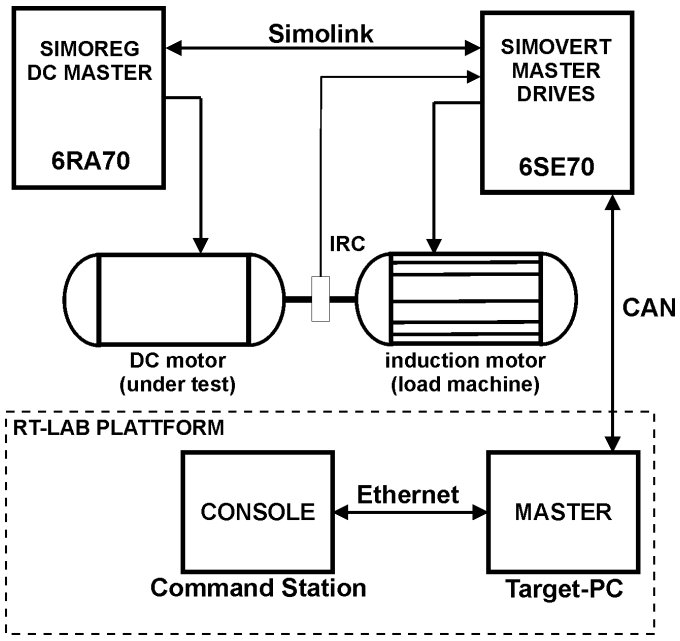


Fig. 2. HIL system structure of load torque emulator

IV. SIMULATION AND EXPERIMENTAL RESULTS

At first simulation model of load torque emulator was designed (Fig. 3). Torque loops of real converters were replaced by their substitutional torque loops in the form of first order lag elements. Simulation model was designed in such a way that it is possible to include and connect different kinds of emulated loads with minimal modifications only. Furthermore, emulator was designed in p.u. variables with regard to control circuits of the most of commercial converters operating with p.u. variables. That is the reason why they are the same for the whole power range of the specific type (e.g. from hundreds of W to several MW).

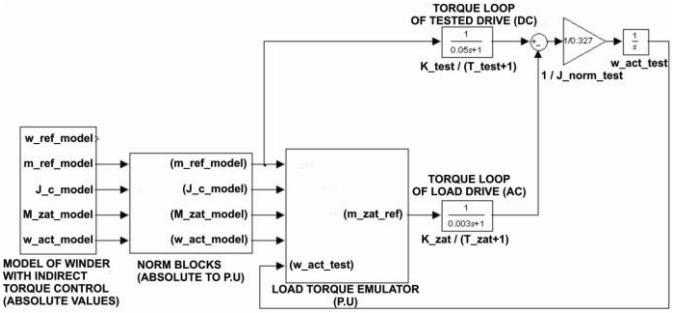


Fig. 3. Simulation model of load torque emulator

Fig. 4 shows the simulated time responses of the speeds and torques for the emulator of the winder drive. The simulated technological process starts at *t=8s* by tensing the steel strip to the constant value and the own winding process starts at *t=15s* by starting of the winder shaft rotation. After reaching the maximum speed (*t=20s*) the winder angular speed is decreasing due to the coil diameter increasing at constant line speed what causes that at constant strip tension the winder torque (m_{test}) increases. The emulator load torque (m_{load}) in reality acts in opposition with the winder motor torque, but it is shown here for better comparison in same direction as winder torque. From Fig. 4 it is also seen an accelerating torque, which is obvious when steeply accelerating or decelerating.

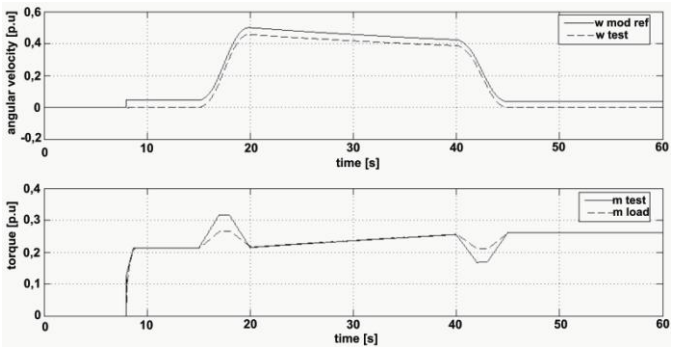


Fig. 4. Simulated time responses of winder drive

When preparing real experiment the norms of angular velocities and torques have to be defined because of using real motors with different rated values as well as normalized moment of inertia for tested drive has to be recounted [7].

Simple experiment of norms' verification is shown in Fig. 5. Experiment starts at the time *t=1s* when 30% step of angular velocity reference ($\omega_{ref\ test}$) is applied (60% in *t=3s*). At the time *t=8s* 40% of load torque ($m_{ref\ load}$) is applied (60% in *t=11s*). It is evident, that applied load torque causes equal torque response of tested drive ($m_{act\ test}$), but when applied 60% of ($m_{ref\ load}$), actual angular velocity of tested drive $\omega_{act\ test}$ is decreasing. It is because of load torque norm is exceeded.

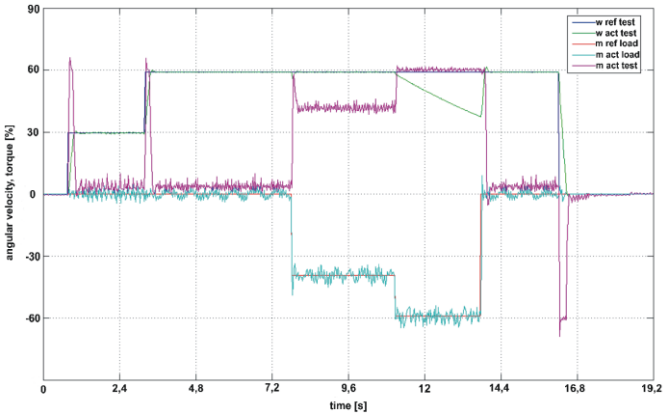


Fig. 5. Simple loading experiment

Simulation results shown in the Fig. 4 was verified by experiment displayed in the Fig. 6. It shows experimental time responses when emulating real industrial winder drive with nominal power 315 kW, nominal speed 1500 rpm and moment of inertia 230 kgm^2. Experiment was done on DC motor as

tested drive with nominal power 4,2 kW, nominal speed 1000 rpm and moment of inertia common with AC load machine equals 0,1 kgm^2. Nominal power of load machine was 7,5 kW and nominal speed was 1500 rpm.

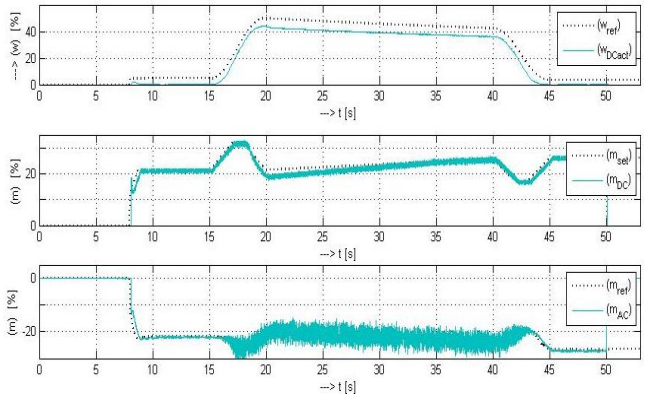


Fig. 6. Experimental time responses of winder drive

V. CONCLUSION

The paper shows load torque emulator used for testing industrial electrical drives under variable conditions. The emulator ensures the same load torque and moment of inertia like they would have in the real application, not taken into consideration only the own mechanics of the drive but also influence of the mechanically coupled drives and properties of the processed material. From the time responses it is obvious that the emulator is able to simulate the behaviour of the high power drive with large moment of inertia on low power on laboratory equipment. This approach enables to develop and test new and complicated control algorithm in the laboratory condition which gives better possibilities to verify various modification of the control algorithm for the industrial drive without real plant restrictions. The next one application of the emulator is developing nowadays - testing of electrical drive components for small vehicles.

Further development of emulator is based on testing of various types of loads and putting them more precisely.

ACKNOWLEDGMENT

This work was supported by the KEGA Project 103-039TUKE-4/2010 "The Development of Students' Skills in Controlled Mechatronical Systems".

REFERENCES

[1] A. Bouscayrol, " Different types of Hardware-In-the-Loop simulation for electric drives," In: *IEEE International Symposium on Industrial Electronics ISIE 2008*, Cambridge. Pp.2146-2151, July 2008.

[2] F. Ďurovský, V. Fedák, K. Kyslan, J. Fetyko, "Emulation of a Winder Drive," In: *15th International Conference on Electrical Drives and Power Electronics EDPE 2009*. Zagreb: KoREMA, 2009. ISBN 978-953-6037-56-8. pp. 1-5.

[3] K. Kyslan, P. Keusch, "HIL simulácia mechatronických systémov s využitím komerčných meničov," In: *Technical Computing Prague 2009*: 17th Annual Conference Proceedings. 2009. P.63.

[4] F. Ďurovský, J. Fetyko, V. Fedák, "Emulátor záťažového momentu so zlepšenou dynamikou," In: *Automatizácia a riadenie v teórii a praxi ARTEP 2009*. Stará Lesná, SR. 3-2009, pp. 82-1-82-13. ISBN 978-80-553-0164-4.

[5] F. Ďurovský, J. Fetyko, V. Fedák, "Testovanie pohonov s emulátorom záťažového momentu," In: *Strojárstvo EXTRA* (cd rom). nb.5 (2009). pp. 8/1-8/5. ISSN 1335-2938.

[6] R. Macko, M. Žalman, M. Uhríček, "Programovateľný emulátor mechanických záťaží pre motory," *AT&P journal 2/2005*. ISSN 1335-2237. pp. 92-95.

[7] K. Kyslan, "*Load Torque Emulator*," Diploma thesis, Technical University of Košice, FEEI, 2009, (in Slovak).

Electric breakdown strength measurement of liquid dielectric samples exposed to the weather effect

[1]*Martin MARCI*, [2]*Ľudovít CSÁNYI*

[1, 2] Dept. of Electric Power Engineering, FEI TU of Košice, Slovak Republic

[1]martin.marci@tuke.sk, [2]ludovit.csanyi@tuke.sk

***Abstract*— In the beginning of the 21st century, the effort of putting ecology principles into all industry branches, including power engineering, had become one of the biggest phenomenon. One of the most important parts of any electrical device is insulation, which is mostly a mixture of chemicals that fulfill a variety of requirements, especially in technical and economic terms, but sadly not environmental. The manufacturing of high-voltage transformers involves daily production, storage, transporting, put in practice and liquidation of hundreds of tons of oil, most commonly mineral oil. Commonly used mineral oils contain a lot of toxic inhibitors to improve their properties. In conjunction with the oil reserves reduction, it is necessary to find a new source of liquid electroinsulants. Vegetable oils seem to be an appropriate solution. The research of vegetable oils as an electroinsulating medium has not reached unambiguous results yet. Because of these facts, this paper deals with the possibility of using vegetable oils as an alternative to the commonly used mineral oils in terms of the breakdown strength. The breakdown strength is one of the parameters for electroinsulating oils quality appraisal.**

***Keywords*—Liquid insulant, breakdown voltage, breakdown strength**

I. INTRODUCTION

The insulation quality greatly affects the period of service, but also the price of the insulant. The requirements for the devices insulation are high, but are often influenced by the financial possibilities of the customer. It is important, that the initial costs will not increase due to overaging of the insulating system, which in such a case is necessary to renovate or to change, or worse, may lead to complete device devastation. In service condition as well as at failure of the device, the operator is responsible for the leakage of an insulating medium and its removal, which is expensive. The problem solution is to use an electroinsulating medium, that is non-aggressive towards the environment and which is not subject to aging or to outdoor weather influences such as temperature, humidity and oxygen access in a great measure. Usage of vegetable oils may be a solution in terms of environmental perspective. Concerning their use in technical terms, there are some papers devoted to this problematic. See e.g. [1]. This paper is aimed to examine the effects of weather exposure on several samples of electroinsulating oils in terms of breakdown voltage and electrical breakdown strength values.

II. PROBLEM OVERVIEW

A. Breakdown in Liquid Dielectrics

Electric breakdown in liquid dielectrics is a phenomenon in which a bridging of distance between electrodes occurs with consequent decreasing of voltage on electrodes and high value of current flowing through. This means, that the dielectric loses its insulating properties. Deterioration of insulating characteristics of liquid dielectrics is only temporary. [1]

B. Breakdown Voltage

Breakdown voltage presents the degree of ability of oil to resist electric stress. It is the minimal voltage value, which causes electric conductivity to rise to a level that causes an electric breakdown. High value of current abounding through a breakdown area causes mechanical, thermal and chemical processes that change dielectric characteristics. These changes are so significant, that the dielectric is not able to completely return to its regular condition, because of solid particles and chemical compounds in liquid or vapor consistency, that arise due to electric discharge activity. Free water, gas bubbles and solid particles are inclined to migrate into the areas with enhanced stress and they cause lower value of breakdown voltage. For this fact, the breakdown voltage can be used as an indicator of oil pollution. [1]

C. Electric Breakdown Strength

Electric breakdown strength is one of the basic qualitative characteristics of dielectrics in addition to polarization and dielectric loss. In an electric field, dielectric keeps its insulating characteristic only up to specific values of electric field intensity. After reaching this boundary (critical) field intensity, resistance of dielectric decreases rapidly to a resistance level of conductive materials. [2]

In case of a homogeneous electric field, the field intensity is the same at full length of the breakdown trajectory; therefore electric breakdown strength can be calculated using the pattern:

$$E_p = \frac{U_p}{d} \quad (1)$$

E_b	[kV.mm^{-1}]	electric breakdown strength,
U_b	[kV]	breakdown voltage,
D	[mm]	interelectrode distance

Electric breakdown strength is a parameter most commonly used for oil quality arbitration. Ranking electric breakdown strength correctly in liquid dielectric is more difficult as in solid or vapour dielectrics. The main reason is the fact that the breakdown process in insulant liquid and the value of breakdown voltage depends on a number of random factors. Among the most important belongs water content, solid impurity content, electrode configuration and applied voltage period.

1) Water content

Water content strongly affects the service period of a transformer due to aging of insulation, whether it is oil, or solid insulation. The increase of the water content causes a decrease of breakdown voltage values and degradation of the insulant. Water can occur in three forms in the oil insulation: as dissolved, emulsified and free. [3]

The two main sources of water in oil transformers and oils as such is humidity penetration from the outside atmosphere and degradation of oil and cellulose. Water has much higher permittivity than oil and therefore colloid water particles are pulled into the areas with the greatest electric field intensity. By forces of the field, particles are deformed and lengthened and create strings, along which a breakdown occurs. To create these strings, low level of water content is sufficient (0.01 - 0.02%). The increase of water content causes parallel strings creation, but these do not affect breakdown strength value no more. Determination of breakdown strength value is an indirect method for water content evaluation. Another possibility is a direct method using special equipment designed for this purpose.

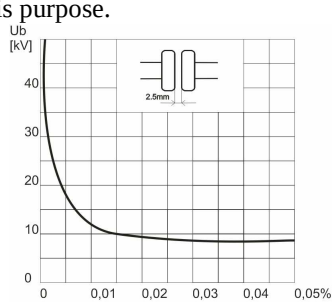


Fig. 1. Water content and breakdown voltage relation in transformer oil. [4]

2) Solid impurities and other foreign particles

In addition to the water, solid particles and gas bubbles have considerable effect on the electrical breakdown process.

The presence of impurities in the liquid dielectric reduces electric strength. The impurities mold is very various and their impact is diverse. Colloidal particles and macroscopic impurities are characterized by high mobility. Their arrangement in liquid can be changed under the electric field influence and this can change the local electric fields status. Gas bubbles are the next impurity type. They are created by liquid heating, under the effect of electric field and can be situated directly in liquid, or on the surface of the electrodes, where cause creating of a layer with lower breakdown strength value. [2]

Solid impurities arise directly in the transformer container during operation too. These impurities are created mainly as a by-product of oil oxidation and aging, partial discharge activity and cellulose pieces release under influence of insulation aging. These solid particles have higher value of permittivity than oil and therefore they are pulled into the areas with the higher field intensity and create strings, which can reduce the breakdown strength of non-conducting medium markedly.

III. Measure Station Arrangement

A. Description and preparation of measurement

Measurement of electrical breakdown strength was made on five oil samples. Specifically, new mineral oil ITO 100, aged mineral oil ITO 100 (oil used in lab for research purposes, this means it was subject to oxygen activity and operation stress), silicone oil Lukosoil M200, vegetable colza-oil Raciol and vegetable sunflower oil Vénusz. All the samples were subject to outdoor weather factors such as low temperature, humidity and oxygen access during one month (12.03.2009 - 14.04.2009). Prevention of infiltration of foreign solid components was performed by multiple layer of gauze.

Measurement was made on the TuR Dresden (Fig. 2) - test device aimed at this purpose, which assures continuous voltage increase during measurement.

Fig. 2. TuR Dresden – test device

B. Measuring Accuracy Procedure

The oil sample was placed into the measuring container (Fig. 3) of the test device so that oil was poured on one of the electrodes due to minimize of immixture of oil with air, what could affect the value of electric breakdown strength. After oil pouring a 30 minute pause was made to settle any solid impurities and also to allow the air gaps to escape.

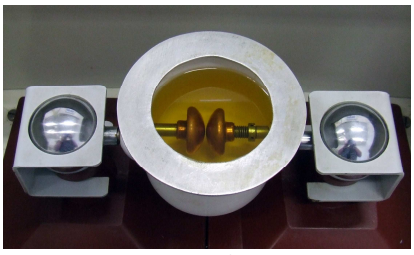


Fig. 3. Measuring container of test device TuR Dresden

The first breakdown was executed at the preset interelectrode distance of 0.3mm. After the breakdown a three minute pause followed due to insulant oil characteristics recovery. This procedure was repeated five times for the actual interelectrode distance, after that, the interelectrode distance was increased continuously on all of distances of 0.6mm; 0.9mm; 1.2mm; 1.5mm; 1.8mm; 2.1mm; 2.4mm; 2.7mm; 3mm. After the measurement, measuring container was purified by petrol and hot water and left to dry for 24 hours.

IV. Evaluation Of Breakdown Voltage Measurement

After obtaining all breakdown voltage values, for one interelectrode distance, the highest and the lowest values were stroked off. Remaining three values of breakdown voltage were used for calculating an average value of breakdown voltage U_b. This value was used as a base for electric breakdown strength determination for every interelectrode distance, according to pattern (1). This procedure was applied for every oil sample. All the calculated values were entered in the tables. On the base of these tables, graphs for electric breakdown strength and breakdown voltage were made. As already mentioned, breakdown is a random event, therefore the join-line of values does not befit an ideal curve (Fig. 4) Because of this fact, the values of breakdown voltage and breakdown strength in graphs are represented just by points and join-line is represented by linear approximation curve for breakdown voltage and logarithmic approximation curve for breakdown strength. The minimal distance of 0.3 mm is chosen deliberately, because of problems with electric arch in the interelectrode area for shorter distances.

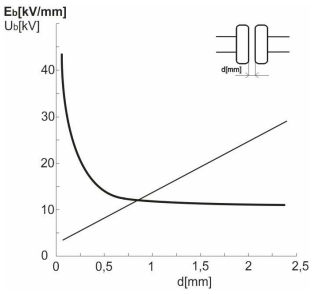


Fig. 4. $E_p = f(d)$, $U_p = f(d)$ relations

A. New mineral oil ITO 100

TABLE I
U_B And E_B Values For New Mineral Oil ITO 100

d[mm]	0,3	0,6	0,9	1,2	1,5	1,8	2,1	2,4	2,7	3
U[kV]	5,33	9	11	14	18,33	22,33	25	27,33	28,33	31
Ep[kV/mm]	17,76	15	12,22	11,66	12,22	12,04	11,9	11,39	10,49	10,33

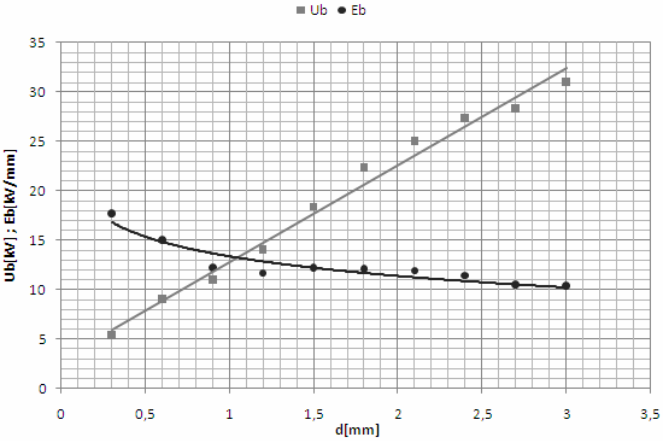


Fig. 5. $E_b = f(d)$, $U_b = f(d)$ characteristics for new mineral oil ITO 100

B. Aged Mineral Oil ITO 100

TABLE II
U_B And E_B Values For Aged Mineral Oil ITO 100

d[mm]	0,3	0,6	0,9	1,2	1,5	1,8	2,1	2,4	2,7	3
U[kV]	5	6,33	9	13	17	20,33	20,33	22,33	26	27
Ep[kV/mm]	16,7	10,56	10	10,83	11,33	11,29	9,68	9,3	9,62	9

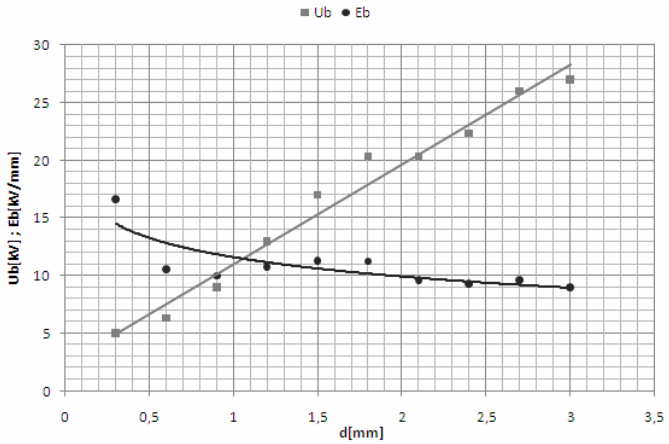


Fig. 6. $E_b = f(d)$, $U_b = f(d)$ characteristics for aged mineral oil ITO 100

C. Silicone oil Lukosoil M200

TABLE III
U_B And E_B Values For Silicone Oil Lukosoil M200

d[mm]	0,3	0,6	0,9	1,2	1,5	1,8	2,1	2,4	2,7	3
U[kV]	8	13,33	17	20	23,67	25,33	29,67	33	37,67	41,33
Ep[kV/mm]	26,67	22,22	18,88	16,67	15,78	14,07	14,13	13,75	13,95	13,77

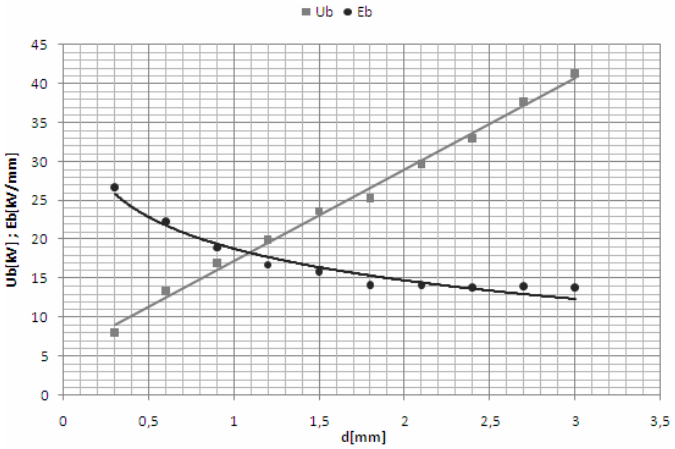


Fig. 7. $E_b = f(d)$, $U_b = f(d)$ characteristics for silicone oil Lukosoil M200

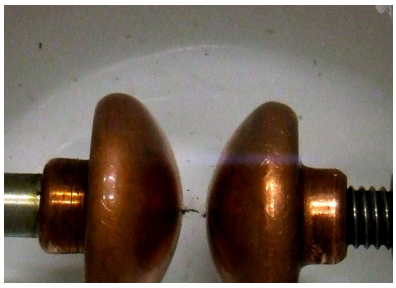


Fig. 8. Creation of soot after breakdown in silicone oil Lukosoil M200

D. Vegetable Colza Oil Raciol

TABLE IV
U_B And E_B Values For Vegetable Colza Oil Raciol

d[mm]	0,3	0,6	0,9	1,2	1,5	1,8	2,1	2,4	2,7	3
U[kV]	16	17	19	23	25,67	27	29,67	34,67	36,67	41
Ep[kV/mm]	53,33	28,33	21,1	19,17	17,11	15	14,12	14,44	13,58	13,67

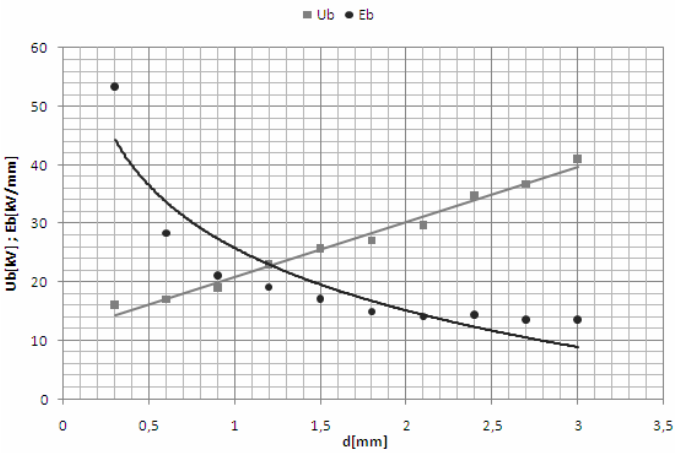


Fig. 9. $E_b = f(d)$, $U_b = f(d)$ characteristics for vegetable colza oil Raciol

E. *Vegetable Sunflower Oil Vénusz*

TABLE V
U_B AND E_B VALUES FOR VEGETABLE SUNFLOWER OIL VÉNUSZ

d[mm]	**0,3**	**0,6**	**0,9**	**1,2**	**1,5**	**1,8**	**2,1**	**2,4**	**2,7**	**3**
U[kV]	16	17	19	23	25,67	27	29,67	34,67	36,67	41
Ep[kV/mm]	53,33	28,33	21,1	19,17	17,11	15	14,12	14,44	13,58	13,67

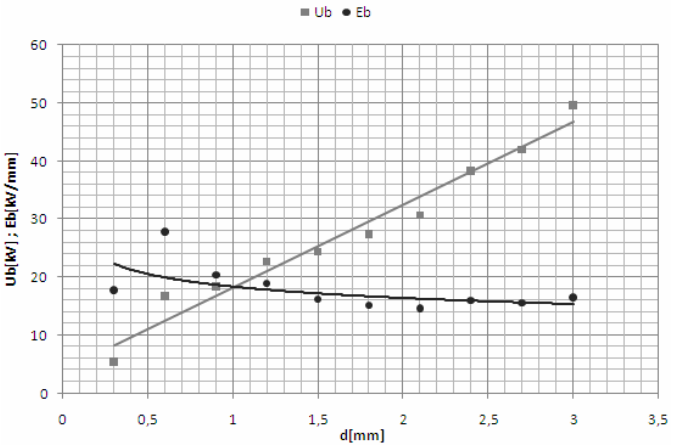


Fig. 10. $E_b = f(d)$, $U_b = f(d)$ characteristics for vegetable sunflower oil Vénusz

V. SUMMARY AND COMPARISON OF RESULTS

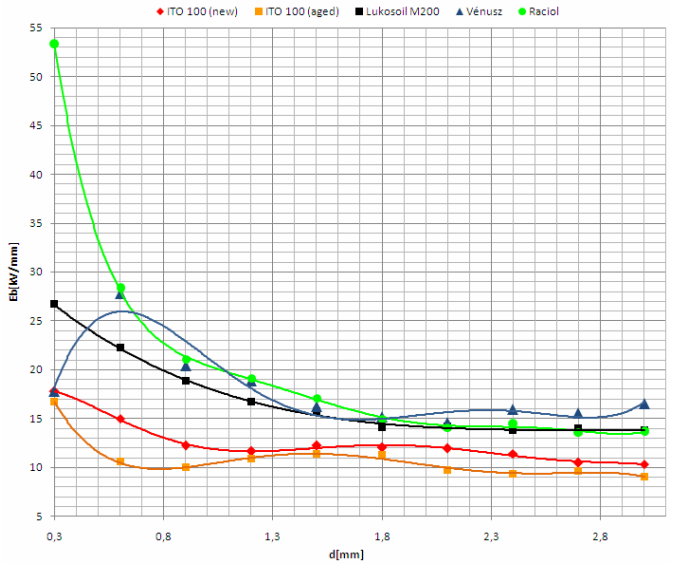


Fig. 11. Comparison of $E_p = f(d)$ characteristic for all oil samples

As shown on the graph (Fig. 11.), the highest value of breakdown strength was reached by vegetable oil samples of Raciol and Vénusz, the lowest by mineral oil samples, new and aged ITO 100. Silicon oil reached average values of breakdown strength, but in terms of behavior during measuring and breakdown process it appeared to be the worst from the measured samples. In contrast, the most stable behavior during measuring was observed with vegetable oils. Mineral oil samples behaved more stable than silicone oil, but they showed dependence of breakdown strength and oil stability on breakdown quantity. The biggest drawback of the vegetable oil samples is a possibility of the creation of electric arc with very low distances (0.3mm), which affected the value of electric strength of sunflower oil sample Vénusz significantly.

VI. CONCLUSION

The task of this paper was to verify the possibility of using of vegetable oils as an electroinsulating medium in terms of electric breakdown strength, which is one of the most important parameters of the electroinsulating liquids. Results of the practical experiment, which consisted of breakdown voltage measuring and breakdown strength calculation, clearly shows that in terms of these parameters, vegetable oils are suitable for use in power system devices. Their use could be beneficial particularly in areas, where weather effect strongly influences the device and in areas with high level of environmental precaution.

ACKNOWLEDGMENT

This work was supported by scientific Grant Agency of the ministry of Education of the Slovak Republic project VEGA No. 1/0368/09 and APVV-20-006005.

REFERENCES

[1] MARCI, M.: Meranie výbojovej činnosti v kombinovanej izolácii olej – papier, diplomová práca, Košice, 2009.
[2] Poljak, František: Izolanty a dielektriká, ALFA, 1983.
[3] KOLCUNOVÁ, I.: Diagnostika elektroenergetických zariadení metódou čiastkových výbojov , Technická univerzita v Košiciach, Košice, 2008, ISBN 978-80-553-0031-3.
[4] HASSDENTUEFEL, Josef – DUBSKÝ , Jan – RAPOŠ, Michal – ŠANERA, Jozef: Elektrochemické materiály, SNTL, Praha, 1978

Induction Heating of Ferromagnetic Charge

[1]*Dušan MEDVEĎ*, [2]*Muhamed Abdulla MUHAMED*

[1] Department of Electric Power Engineering, FEI TU of Košice, Slovak Republic,
[2] Higher Institute of Comperhensive Occupations in Bani Walid, Libya

dusan.medved@tuke.sk, mrabett@yahoo.com

***Abstract*—Induction heating belongs to category of electric heating, where the physical principle comes from Joule law. The heat is generated directly in the charge, where the electromagnetic and thermal fields are distributed non-uniformly and they are mutually influenced. This fact can complicate the analysis of induction heating. In this article will be introduced results of heating analysis of planar board with the respect of relative magnetic permeability change. Given analysis was processed by numerical method of finite differences.**

***Keywords*—Induction heating, relative permeability, numerical modeling, finite difference method.**

I. INTRODUCTION

Recent thermal technologies commonly belong to power demanding processes. The essential part of them uses electric heating for thermal treatment of conductive materials. Because of high electricity costs there is permanent interest of decreasing energy consumption of mentioned technologies. One option is the rationalization of energy consumption using the analysis of thermal field of heated charge.

II. THEORETICAL PROBLEM

Solution difficulty of thermal and electromagnetic field of ferromagnetic materials up to Curie temperature is deeply influenced by material quantities change, that are dependent as on temperature as on external magnetic field. One of the quantities, that essentially influences heating curve, is relative magnetic permeability. Its nature is rapidly changed by temperature up to structural change, so-called Currie point (temperature). It is possible to get very accurate results that can rationalize energy consumption using detailed analysis. One of those methods is to solve the problem as a deeply dependent problem, where the mutual influence of electromagnetic and thermal field is considered.

III. THERMAL AND ELECTROMAGNETIC FIELD MODELING

Mathematical modeling is one of the major factors in the successful design of induction heating systems. Theoretical models may vary from a simple hand-calculated equation to a very complicated numerical analysis which can require several hours of computational work using modern computers. The choice of a particular theoretical model depends on several factors, including the complexity of the engineering problem, required accuracy, time limitations and cost.

A. Electromagnetic Field

The calculating technique of electromagnetic field depends on the ability to solve Maxwell's equations. For general, time-varying electromagnetic fields, and using by some procedures of vector algebra it is possible to get equations in form as

$$\frac{1}{\gamma}\cdot\nabla^2\boldsymbol{H} = j\cdot\omega\cdot\mu_\mathrm{r}\cdot\mu_0\cdot\boldsymbol{H} \tag{1}$$

$$\frac{1}{\mu_\mathrm{r}}\cdot\nabla^2\boldsymbol{E} = j\cdot\omega\cdot\mu_\mathrm{r}\cdot\mu_0\cdot\boldsymbol{E} \tag{2}$$

$$\frac{1}{\mu_\mathrm{r}\cdot\mu_0}\cdot\nabla^2\boldsymbol{A} = -\boldsymbol{J}_\mathrm{z} + j\cdot\omega\cdot\gamma\cdot\boldsymbol{A} \tag{3}$$

where $\boldsymbol{E}$ is electric field intensity, $\boldsymbol{H}$ is magnetic field intensity, $\boldsymbol{J}_\mathrm{z}$ is source current density and γ is electric conductivity. Symbol ∇^2 is the Laplacian, which has different forms in Cartesian and cylindrical coordinates.

Using by (1) to (3) with the correspondent boundary conditions it is possible to solve time-varying harmonic field (for quantities $\boldsymbol{H}$, $\boldsymbol{E}$ or $\boldsymbol{A}$). Equation (1) to (3) can determine required parameters of induction system such as current in coil; power; induced current density in charge and so on.

However, there is important to determine the problem as three-dimensional, but in many cases it is possible to simplify the problem to two- or one-dimensional. Problem simplifying is possible to apply for example to tasks with solenoid coil and where the direction of vectors $\boldsymbol{A}$ and $\boldsymbol{E}$ in longitudinal cross-section has only one component, that is *z*-axis direction. Vectors $\boldsymbol{H}$ and $\boldsymbol{B}$ have also only one component in crosswise section. This fact allows simplifying three-dimensional field to two-dimensional field. For example in the case of magnetic vector potential $\boldsymbol{A}$, the equation (3) can be expressed in two-dimensional orthogonal axis system as follows

$$\frac{1}{\mu_\mathrm{r}\cdot\mu_0}\cdot\left(\frac{\partial^2\boldsymbol{A}}{\partial x^2}+\frac{\partial^2\boldsymbol{A}}{\partial y^2}\right) = -\boldsymbol{J}_\mathrm{z} + j\cdot\omega\cdot\gamma\cdot\boldsymbol{A} \tag{4}$$

The boundary of area is determined so, that the magnetic vector potential $\boldsymbol{A}$ is equal to zero along the boundary (Dirichlet condition) or its gradient should be negligibly small along the boundary in comparison to its value somewhere in the area (very close to boundary) (Neumann condition $\frac{\partial \boldsymbol{A}}{\partial t} = 0$). Because of that fact, the heat transfer equation (5), that is detailed described in next chapter, together with equation (4) and their initial and boundary conditions can completely describe the electroheat processes in very commonly used applications of induction heating of

cylindrical charge. Therefore, any electromagnetic problem of induction heating can be determined by terms for ***A***, ***E***, ***B*** or ***H***.

B. Thermal Field

In general, the transient (time-dependent) thermal field in heat transfer processes in a metal workpiece can be described by the Fourier equation:

$$c \cdot \rho \cdot \frac{\partial \vartheta}{\partial t} + \nabla \cdot \left(-\lambda \cdot \nabla \cdot \vartheta\right) = q_e \qquad (5)$$

where ϑ is temperature, ρ is the density of the metal, c is the specific heat, λ is the thermal conductivity of the metal and q_e is the heat source density induced by eddy currents per unit of time in a unit volume (heat generation). This heat source density q_e is obtained by solving the electromagnetic problem.

Equation (5), with the suitable boundary and initial conditions, represents the three-dimensional temperature distribution at any time and at any point in the workpiece.

If the heated body (charge) is geometrically symmetrical along the symmetry axis, the Neumann boundary condition can be formulated as

$$\frac{\partial \vartheta}{\partial n} = 0 \qquad (6)$$

The Neumann boundary condition implies that the temperature gradient in a direction normal to the axis of symmetry is zero. In other words, there is no heat exchange at the axis of symmetry. This boundary condition can also be applied in the case of a perfectly insulated workpiece.

In the case of planar board heating, equation (5) can be rearranged to form as

$$c \cdot \rho \cdot \frac{\partial \vartheta}{\partial t} = \frac{\partial \vartheta}{\partial x} \cdot \left(\lambda \cdot \frac{\partial \vartheta}{\partial x}\right) + \frac{\partial \vartheta}{\partial y} \cdot \left(\lambda \cdot \frac{\partial \vartheta}{\partial y}\right) + \frac{\partial \vartheta}{\partial z} \cdot \left(\lambda \cdot \frac{\partial \vartheta}{\partial z}\right) + q_e \qquad (7)$$

Equation (7), together with boundary conditions, is the mathematical model of thermal fields with heat source in planar board.

IV. Calculation of One-Dimensional Coupled Problem in Planar Board Respecting Relative Magnetic Permeability Change (μ_R = f(ϑ), γ= f(ϑ), C = konšt., λ = konšt.)

Coupled problem can be solved for the simple shapes of heated body by analytical methods, or there can be used some numerical method. In the next part of this article there will be introduced some analysis results of thermal and electromagnetic field distribution (calculation of one-dimensional problem) in planar board with the respect of relative magnetic permeability change. There was used numerical method of finite differences, where differential equations (4) and (9) were replaced by difference equations.

In this coupled problem, there was considered that the planar board charge has to be heated, and the heat is spread only in x-axis direction. Material charge properties were dependent on temperature (μ_r = f(ϑ), γ= f(ϑ), c = const., λ = const.). Given input data:

- ***charge parameters***:
  board thickness d_2 = 10 cm, ($d = \frac{d_2}{2} = 5$ cm)
  material density ρ_m = 7700 kg.m^{-3},
  thermal conductivity λ = 14,88 W.m^{-1}.K^{-1},
  specific heat capacity c = 510 J.kg^{-1}.K^{-1},
  initial charge temperature ϑ_0 = 20 °C,
- ***inductor parameters***:
  current in inductor I_1 = 2050 A,
  inductors turns number per 1 m of length N_{11} = 49,
  current frequency in inductor f = 50 Hz,
- ***parameters for determination of relative magnetic permeability and electric conductivity***:
  relative permeability μ_r calculated in every step (see [4]),
  magnitude of magnetic field intensity in saturation H_S = 150 A.m^{-1},
  magnitude of magnetic flux density in saturation B_S = 0,4 T,
  Currie temperature ϑ_C = 768 °C,
  inclination constant of hyperbolic function c = 10 [4],
  electric conductivity γdetermined by spline function
- ***parameters for boundary conditions***:
  surrounding temperature ϑ_{pr} = 20 °C,
  heat transfer coefficient α = 150 W.m^{-2}.K^{-1},
- ***parameters of calculation step***:
  number of charge dividing: n = 50,
  calculation time step satisfying stability condition: Δt = 0,05 s
  heating time t_k = 540 s

It is possible to get series of heating curves after entering input parameters and applying forward finite difference method in MATLAB.

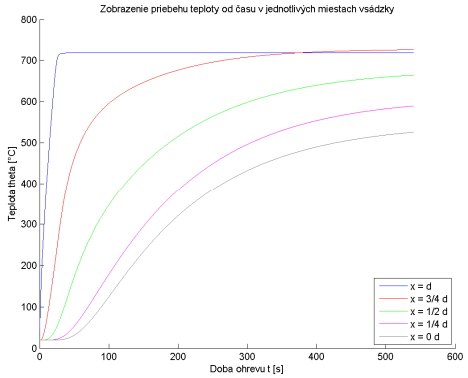


Fig. 1. Dependence of temperature arrangement on heating time in particular locations of charge (μ_r = f(ϑ), γ= f(ϑ), c = const., λ = const.)

Characteristics of temperature change during the heating time in particular chosen places of a charge are presented in the graph on Fig. 1. One can see the rapid temperature change close to boundary layer ($x = d$) already at t = 50 s, where the magnetic material becomes “non-magnetic”. Transformation of ferromagnetic material to paramagnetic is caused by structural change and it express as sudden decreasing of relative magnetic permeability μ_r to value approximately equal to 1. This change is not so rapid in other observed places in the charge.

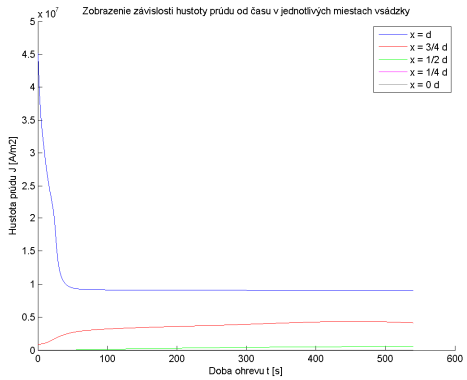


Fig. 2. Dependence of current density dissipation on heating time in particular places in the charge (μ_r = f(ϑ), γ= f(ϑ), c = const., λ = const.)

Characteristics on Fig. 2 show the electromagnetic field distribution (current density) in chosen particular places in the charge during heating time. Similarly as on Fig. 1, also here, close to boundary layer ($x = d$) it is visible rapid change at time approx. $t = 50$ s, where the magnetic material becomes non-magnetic. After this conversion point the current density is then dependent only on electric conductivity change and because of decreasing of electric conductivity by temperature, this shape is also slightly decreased.

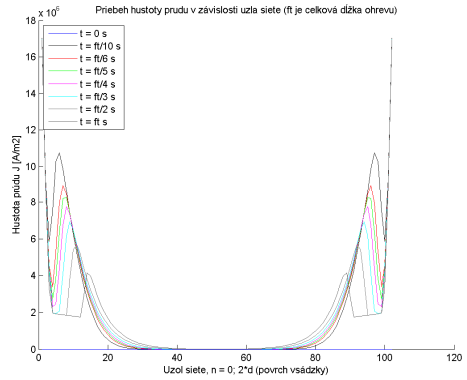


Fig. 3. Dependence of current density dissipation in particular locations in charge (μ_r = f(ϑ), γ= f(ϑ), c = const., λ = const.)

Another graph on Fig. 3 presents the dependence characteristics of current density J on distance from charge surface (denoted by node number) in particular chosen times, where charge symmetry was considered (symmetrical heating from both sides). From the correspondent curves it is visible again the current density change in locations where the ferromagnetic material becomes paramagnetic. These curves are growing exponential towards from charge center to surface and at the Currie point the current value fall down very rapidly. Package of all these local extremes (maximums – current density particular rapid changes) create the curve that express conversion of magnetic material to non-magnetic.

Current density magnitude in particular charge layers is conditioned also by charge temperature. Material properties of heated charge are deeply dependent on temperature and because of that, the current density is not linear during the heating time. Current density decreases by increased temperature from the layer to layer (from surface to center).

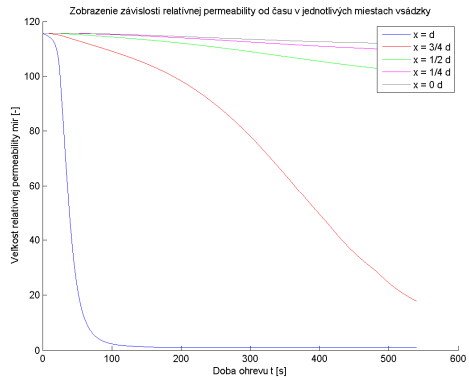


Fig. 4. Dependence of relative permeability on heating time in particular charge locations (μ_r = f(ϑ), γ= f(ϑ), c = const., λ = const.)

Since the charge is "two-layer" (non-magnetic in the middle of charge and magnetic close to charge surface) during the heating time, the main material quantity, which describes this process, is magnetic permeability. Illustration of relative permeability in particular chosen charge places in dependency on heating time is on Fig. 4.

V. Conclusion

Presented results are the part of analysis of characteristics determination of ferromagnetic material during the heating process. They have to show the influence of material parameters change during the induction heating. Rapid change in characteristics, especially of current density and temperature, are essentially visible by respecting as the relative magnetic permeability μ_r as the electric conductivity γ change on temperature. This sudden change is caused by conversion of magnetic material to paramagnetic, where the value of relative permeability is suddenly decreased to value approximately equal to 1. More detailed results are published in [4].

An advantage of designed mathematical model and used program is the analysis of material quantities influence and the phenomena performed during an induction heating process are possible to refine and get more accurate results that are suitable for another treatment.

Presented method is suitable for various coordinate systems and charge shapes. Obtained results confirm the empirical experiences from solution of both fields.

Acknowledgment

This work was supported by project VEGA SR No. 1/0166/10.

References

[1] Medveď, D.: *Numerical solution of induction heating in 2D.* In: SCYR 2009: 9th Scientific Conference of Young Researchers: Proceeding from conference: May 13th, 2009 Košice, Slovakia. Košice: FEI TU, 2009. s. 1-4. Internet: <http://web.tuke.sk/scyr/> ISBN 978-80-553-0178-5.

[2] Medveď, D – Novák, P.: Ohrev rovinnej dosky s uvažovaním zmeny permeability. In: *EE časopis: Odborný časopis pre elektrotechniku a energetiku*. roč. 14, č. 6 (2008), s. 33-36. ISSN 1335-2547.

[3] Sadiku, M. N.: *Numerical Techniques in Electromagnetics*. CRC Press: 2000. 760 s. ISBN: 978084931

[4] Medveď, D.: *Ohrev feromagnetických materiálov do Curieho teploty indukčnou metódou*. Dizertačná práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2008. 170 s.

Telemetric system in automobile based on internet

[1]*Ján MOLNÁR,* [2]*Radoslav BUČKO*

[1]Dept. of Theoretical Electrotechnics and Electrical measurement, FEI TU of Košice, Slovak Republic
[2]Dept. of Theoretical Electrotechnics and Electrical measurement, FEI TU of Košice, Slovak Republic

[1]jan.molnar@tuke.sk, [2]radoslav.bucko@tuke.sk

***Abstract*— The paper deals with the proposal of the measuring chain assigned for remote measuring in the automobile industry. The designed solution should be able to do automatic measurement of all required parameters and to send obtained data to the remote centre, where they could be analyzed by telemetric expert system.**

***Keywords*—measuring, automatic, GSM (Global System of Mobile communication), WiFi (Wireless Fidelity), Expert system**

I. Problem Description

Due to big progress of automotive industry the remote measuring system could help to solve many problems. Various automobile failures could be eliminated by prevention, early malfunction detection or failure detection. The detection and prevention would be based on selected data measuring, data collecting and transporting to the remote center for its further analyzing and evaluation.

Requirements for such a measure chain depend on the data type, which have to be measured. For instance, it is a big difference between a battery voltage and engine revolution measuring. In first case, it is sufficient to collect data few times per hour. In the other case, there is necessary to take a data few times per second. The sampling time is then some tens of miliseconds. In the case of automobile measuring, there exist various data of various types to be measured.

Another problem is created by the information transport to the remote center, where the data are analyzed, evaluated and also the decision about the failure state is made. Today's situation is based on preventive service inspections at regular intervals, where the car is connected to the PC and all diagnostic methods are done. It would be more suitable, if it would be possible to do at the moment of optimal value deviation or in any failure detection. All required data would be transferred to the center and evaluated by the expert system.

Expert system should decide, if the failure state exists or not. In the case of failure state appearance, the driver should receive the information about failure from center. Measurement should be running during the automobile performance. Such a way discovering of incidental and a periodic failures will be easier, because such failures are the most difficult to identify. By designed method should be possible to transport measured data at the moment, when the failure appears.

II. Solution Desing

Two types of net seem to be the most appropriate for data transport. The first is GSM (Global System of Mobile communication) and the second one is WiFi connection (Wireless Fidelity). Both nets have own advantages and disadvantages. The most important advantage of WiFi (compared with GSM) is bandwidth. It means that the bigger data amount in shorter time period can be transferred by WiFi using. The most important advantage of GSM (compared with WiFi) is that the great area is covered by GSM signal. Except that, it is possible to use various tools of GSM communication such a SMS etc. For remote measurement problem solving it is possible to use both types of the net. The nets could complement each other, depending on situation.

In praxis, if every automobile would be connected to the net and every one car would continuously sending a data, it would be inefficient and the net would be very soon overloaded. The net should be used only for diagnostic and optimization case.

During diagnostic, there would be monitoring of all automobile processes and values. Measured date would be compared with optimal values. Evaluation could be done by some local expert system or using artificial neural network etc. In case of deviation from standard state, automobile should connect the expert system and measured data would be transferred to the remote expert system. Remote expert system should be more sophisticated and frequently updated, so the evaluation by remote center should be more exact. For such communication a GSM network would be sufficient. Similar local expert system is nowadays implemented in advanced automobiles of higher class. But connection with central expert system will enable much complex and reliable diagnostic. It is much effective to update the central expert system than individual local systems in each automobile.

If the car will be inside the area covered by WiFi signal and transfer capacity is sufficient, then the optimizing of engine setup, the automobile software update and various modifications would be enabled which were done only in car service yet. The basic condition of such modification is given by sufficiently fast connection existence, which would be able to transfer the fast changed parameters (such engine revolution, gas consumption etc).

III. Architecture Proposal

Architecture proposal of such connection is displayed in figure Fig.1. Architecture doesn't deal directly with

measuring. Measuring problems are mostly solved nowadays. Architecture deals with problem of data transport. The "bus" of the measure chain is responsible for data transport to the evaluation system.

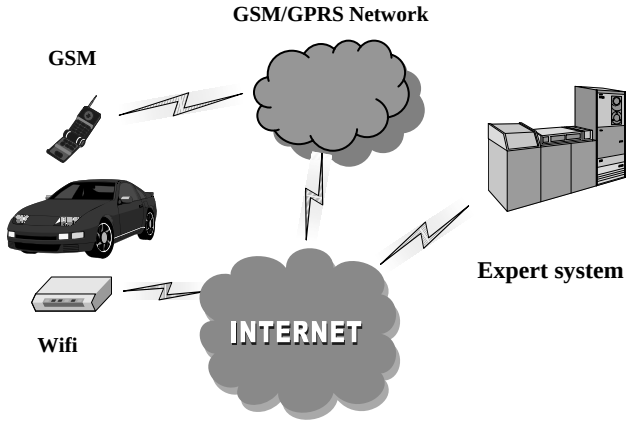


Fig. 1. Architecture proposal

As the figure shows, the automobile should have integrated GSM modem and also device for WiFi net connection. Using the GSM or WiFi net, the automobile would have access to the Internet. Via Internet the car would be connected with remote expert system.

Local expert system would not just watch the automobile parameters, but it could also control the communication. It would decides, which data and how often would be send to the remote system in the case of failure state. It would also decide which data would be sending via GSM and which via WiFi. It would be also responsible for optimizing data transfer.

Expert system would offer suitable solution for driver. In case of failure, it would recommends the driver to stop the car immediately, drive the car to the next service or ignore the problem (problem is temporary). It would be possible also give attention to the driver, in the case when he often uses the high engine revolution etc. Connection to the Internet would have much more advantages like as an on-line help (how to change electrical fuse – and which one, change a wheel, weather forecast). But the most important is possibility of emergency call. Using GSM net it would be possible to find out the car location etc.

Such a way created system, the part of which is remote measurement chain utilizing the Internet and connection to the expert system, would be great benefit for automobile reliability and safety.

IV. Measuring and Communication System

The root of the whole communication and measurement system is embedded system. It ensures communication using GSM modem or Wifi. Individual equipments are directly connected to embedded system using USB interface. Communication itself controls the program which surveys accessible networks. According to the network the program chooses the device used to connect the Internet and data transfer. 12bit converter handles with the analogue signals and changing them to the digitals. The converter is directly connected to a one of the I/O interfaces of embedded system. A/D converter and embedded system communication ensures the program, which control a sampling rate and save measured data to a text file on memory medium. The memory medium has also a backup function in case of no accessible network.

All saved data are transferred to a server via Internet. Measured data are saved and analyzed on the server. Block diagram of the measure chain is on the following figure Fig. 2.

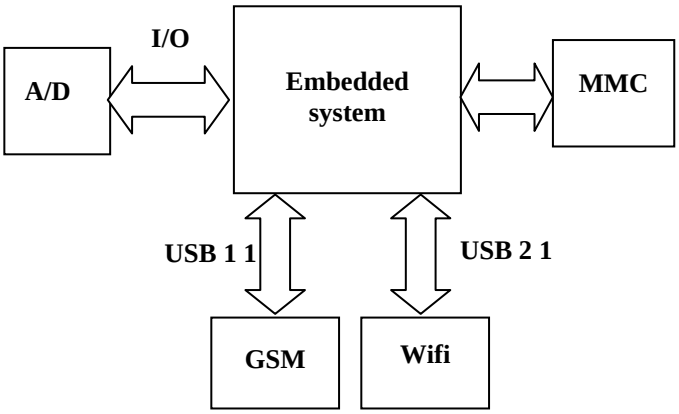


Fig. 2. Block diagram of the measure chain

V. Embedded System

Embedded system runs a real Linux operating system (not a µCLinux) on an ETRAX 100LX microprocessor a 100MIPS RISC CPU made by has two main field applications:

- As a stand alone device to build a micro web server or other network devices as proxy, router, firewall, etc.
- As a core engine to plug onto the PCB of a user application board instead of a simple microcontroller.

Two USB 1.1 host interfaces can be connected to USB memory stick, hard disk, webcam, modem, Wi-Fi or Bluetooth dongle, ADSL adapter, Serial converter, etc. Through the 10/100 Ethernet interface it is possible to have access to the internal Web server, FTP server, SSH, Telnet and the complete TCP/IP stack.

Compilation of simple program can be done by web-compiler from the manufacturer web site. For more complicated programs is possible to download whole system core. It is possible to install also in desktop computer as a virtual application. Using this application it is possible to compile whole Linux core and upload it to the embedded system.

Individual programs handling whole measurement chain are located in memory of embedded system. In case of more memory requirement, it is possible to use on of more memory media connectable to the embedded system.

Communications with embedded system itself, user can establish with any software which communicates with telnet server build in embedded system for example PuTTY, HyperTerminal etc.

VI. Data Analysis

Measured data transferred to the server are inspected by data consistency test. Then the data are analyzed by expert system. The system displays received data or further processed. Sample of measured data shows Fig 3.

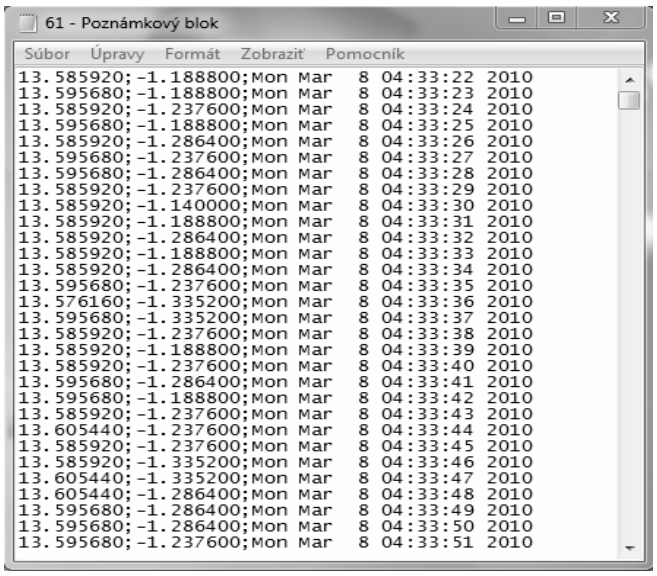


Fig. 3. Sample of received data

Further processing represents graphical display of measured data. Many failures of inspected systems are more clear in graphical representation of the data. On following figure is sample of graphical representation of measured data (Fig.4)

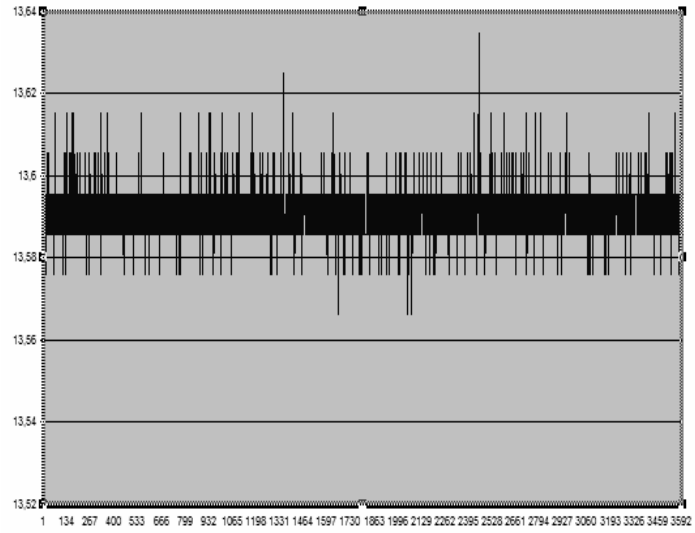


Fig. 4. Graphical representation of measured data

As seen from the graph of measured data, different anomalies or malfunctions which may occurs in car operation could be discovered and determined. Because of long term data collection, many statistics methods may be used. These methods may precisely detect incurred failures. Expert system may use different tools of artificial intelligence, for instance artificial neural network, fuzzy logic, genetic algorithm, which may represent logic core of whole system. After failures and problems discovery may Expert System suggest new settings of engine parameters according to determined problems of specific car. Parameters update may me done automatically or manually in next regular technical inspections.

VII. Conclusion

Using of worldwide network – Internet seems to be optimal for remote data measurement inside the car. The GSM and WIFI connection is suitable for Internet connection. The biggest advantage of WiFi is bandwidth. It is able to transfer data faster than GSM. The biggest advantage of GSM net is that the GSM signal covers majority of area. Realization of such a remote measure chain will serve for future exploration of remote measure chain using Internet as a bus, also in real-time applications. That way would increase possibilities to integrate various measurement systems in the whole complex system.

Acknowledgment

The paper has been prepared by the support of Slovak grant projects VEGA 1/0660/08, KEGA 3/6386/08 and KEGA 3/6388/08.

References

[1] Kováč, D., Kováčová, I., Molnár, J.: Electromagnetic Compatibility - measurement, TU Košice publisher, 72 pages, ISBN 978-80-553-0151-8.

[2] Kováč, D., Kováčová, I., Vince, T.: Electromagnetic Compatibility, TU Košice publisher, 138 pages, ISBN 978-80-553-0150-1.

[3] Tomčík, J., Tomčíková, I.: IT Security of automation and SCADA systems (Part 1). In: AT&P Journal, Vol. 13, No. 4 (2006), pp. 50 – 54, ISSN 1336-5010.

[4] Molnár, J., Kováčová, I.: Distance remote measurement of magnetic field. In: Acta Electrotechnica et Informatica, Vol. 7, No. 4 (2007), pp. 52-55, ISSN 1335-8243

[5] Kováčová, I., Kováč, D.: Modelling and Measuring of Electronic Circuits, textbook FEI TU Košice, ELFA s.r.o. Publisher, 1996, 92 pages, ISBN 80-88786-44-4.

[6] Kováčová, I., Kováč, D.: EMC Compatibility of Power Semiconductor Converters and Inverters, Acta Electrotechnica et Informatica, 2003, Vol.3, No.2, pp.12-14, ISSN 1335-8243.

[7] Molnár, J.: Automatic measurement of magnetic field via internet. 7th PhD Student Conference and Scientific and Technical Competition of Students of Faculty of Electrical Engineering and Informatics Technical University of Košice p. 57-58. ISBN 978-80-8073-803-7.

[8] Vince, T., Molnár, J., Tomčíková, I.: Remote DC motor speed regulation via Internet. In: OWD 2008 : 10. international PhD workshop : Wisla, 18-21 October 2008, p. 293-296. ISBN 83-922242-4-8

[9] Molnár, J.: Automatic measurement of magnetic field via internet. 7th PhD Student Conference and Scientific and Technical Competition of Students of Faculty of Electrical Engineering and Informatics Technical University of Košice p. 57-58. ISBN 978-80-8073-803-7.

[10] Vince, T.: Artificial neural network in remote DC motor speed regulation via internet. In: OWD 2009: 11th International PhD Workshop: Wisla, 17-20 October 2009, p. 419-422. ISBN 83-922242-5-6.

[11] Tomčíková, I., Molnár, J., Vince, T.: Interaction of magnetic field and tension for elastomagnetic sensor of pressure force (In Slovak). In: Elektrorevue. no. 34 (2008), pp. 34-1-34-11. ISSN 1213-1539.

[12] Vince, T., Kováčová, I.: Distance control of mechatronic systems via Internet. In: Acta Electrotechnica et Informatica, 2007, No.3, Vol.7, pp. 63-68, ISSN 1335-8243

Electric Power System of Libya and its Future

Ing. Maher NASR

Department of Electric Power Engineering, FEI TU of Košice, Slovak Republic

maher.nasr@tuke.sk

***Abstract*—This article deals with electricity system of Libya. The current transmission lines are very loaded and there is some new calculation for building of new ones. The main energy strategy is spreading of interconnection between other countries. In this article will be presented also what is the expected peak load in the future and the solution for ensure the sufficient amount of electric energy.**

***Keywords*—Electricity system, transmission lines, energy of Libya.**

I. Introduction

Today, the electricity system of Libya is a state owned vertically structured power utility company and is responsible for generation, transmission and distribution of electric energy. The installed generation capacity was around 6612 MW while the peak load was 4756 MW in 2008. The transmission system consists mainly of 220 kV lines, but the expected load will increase so there are some new calculations, that you can see in the next chapter.

II. The Actual Electricity System of Libya

The actual electricity system of Libya is controlled by state owned company GECOL. This company operates 30 electricity generation plants, mainly steam and simple-cycle gas-turbine units and diesel generators in rural areas. The company is also the sixth largest operator of water desalination plants in the world. More than $1,000 million will be spent on new desalination units over the next ten years consist mainly from based mainly on oil.

The overall electricity statistics in 2007 was as follows:

- Generated energy: 28666 GWh
- Peak demand: 4756 MW
- 400 kV transmission system: 442 km
- 220 kV transmission system: 13677 km
- 33 and 66 kV sub-transmission system: 22556 km
- 11 kV distribution system: 50000 km
- Number of customers: 1224193
- Average consumption per 1 customer: 4 271 kWh
- Installed capacity of desalination system: 22560 m^3/day.

The total energy that plants produced during the year 2008 was 28,666 TWh, what is the energy rate growth of 12,8 % compared to 2007. The annual electricity generation according to years 2001 to 2008 shows the next table.

TABLE I
The Annual Electricity Generation

Year	2001	2002	2003	2004	2005	2006	2007	2008
Generation [TWh]	16,111	17,531	18,943	20,202	22,450	23,992	25,415	28,666

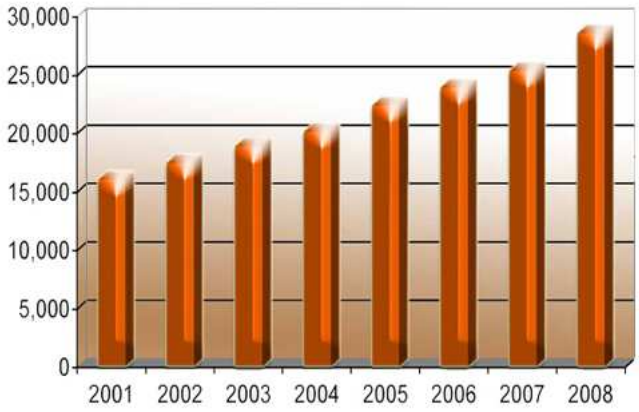


Fig. 1. Graphical representation of annual electricity generation

This generation increased every year for increased electricity demands. It was followed by spreading of electricity transmission lines.

The total portion of electricity transmission lines according to years 2000 to 2008 shows the next table.

TABLE II
The Annual Portion of Transmission Lines

	400	**220**	**66**	**30**
2000	–	60	152	267
2001	–	62	156	268
2002	–	62	163	277
2003	–	62	166	281
2004	–	62	167	286
2005	2	64	169	302
2006	2	70	169	321
2007	3	70	175	355
2008	3	71	178	373

The load of the network during the year 2008 is 4,756 GW, compared with the year 2007 was 4,420 GW, which was a growth rate of 7,6 %. The other loads are interpreted in the next table.

TABLE III
The Annual Electricity Load

Year	2001	2002	2003	2004	2005	2006	2007	2008
Generation [GW]	2,934	3,081	3,341	3,612	3,857	4,005	4,420	4,756

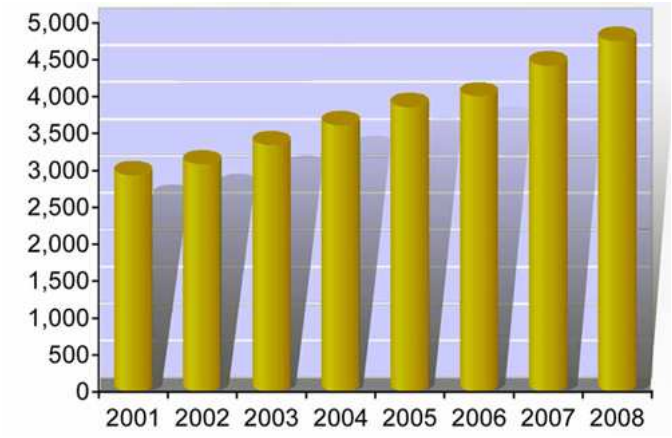


Fig. 2. Graphical representation of annual electricity load

The energy statistics of produced energy during the year 2008 according to type of generation is presented in following table.

TABLE IV
THE ENERGY STATISTICS IN 2008

Energy produced [GWh]	Type of Generation
7264,304	Steam
9882,786	Nature gas
11519,051	Double cycle
28666,141	Total

The total number of substations shows the next table.

TABLE V
THE PORTION OF SUBSTATIONS (TRANSFORMER CAPACITY)

[kV]	400	220	66	30
Number of stations	3	71	178	373
[MVA]	2400	13308	3679	10404

The total length of transmission lines shows the next table.

TABLE VI
TRANSMISSION LINES

Item	1993	2000	2008
400 kV lines [km]	–	–	442
220 kV lines [km]	12640	12887	13667
66 kV lines [km]	10568	12475	13973
30 kV lines [km]	6166	6986	8583
Total [km]	29374	32348	36665

III. THE FUTURE PLANS OF ELECTRICITY SYSTEM OF LIBYA

The main master plan of electricity system of Libya concludes objectives:

- securing and guarantee the electrical power supply to growing demand of electrical energy to all sectors in the country,
- increasing the level of security and adequacy of supply,
- reducing cost by improving service quality and efficiency,
- reducing technical and non technical losses,
- reinforcing international interconnections.

The long term load forecast:

- the peak load of electrical power in Libya is continuously increasing with a relatively high growth rate 8 % per annum,
- recent studies have shown that the expected peak demand in Libya in the year 2008 will be about 4670 MW and the figure is expected to reach 5450 MW by year 2010 and approximately 8000 MW in 2015.

TABLE VII
THE LONG TERM LOAD FORECAST (IN MW)

Year	2008	2009	2010	2011	2012	2013	2014	2015
Total	4671	5045	5458	5884	6355	6863	7412	8005

These loads will require building new power capacities:

TABLE VIII
NEW PROJECTS (IN MW)

Year	2008	2009	2010	2011	2012	2013	2014	2015
Total	164	1937	3386	4192	5047	5669	5779	5889

TABLE IX
THE TOTAL EXPECTED PEAK LOAD (IN MW)

Year	2008	2009	2010	2011	2012	2013	2014	2015
Total	4835	6982	8834	10076	11402	12532	13191	13894

There is around 11000 MW generation capacity needed to be added during the period 2005 – 2015. Nearly 4000 MW are needed by the horizon 2010 with a mixed generation options based on latest technology (steam and combined cycle) using natural gas. Also, under construction there is approximately 1500 MW of new power plants (750 MW in Benghazi C. C, and 750 MW in Misurata C. C). The other awarded contracts for building of new power plants are in West Mountain Ext. (312 MW), Zwitina gas plant (500 MW), Srir west gas plant (750 MW), Gulf steam plant (1400 MW), Tripoli West steam plant (1400 MW) and under contract is Sebha gas plant (750 MW).

Fig. 3. Planned generation projects

During the horizon 2010 the main infrastructure for a strong, reliable, flexible and adequate backbone of the new 400 kW grid will be executed. Therefore, the future plan of the Libyan transmission system is concentrated on 400 kV grids. There are 2 400/220 kV substations and 442 km of 400 kV lines now. Under construction are 7 substations and 2720 km of new 400 kV transmission lines.

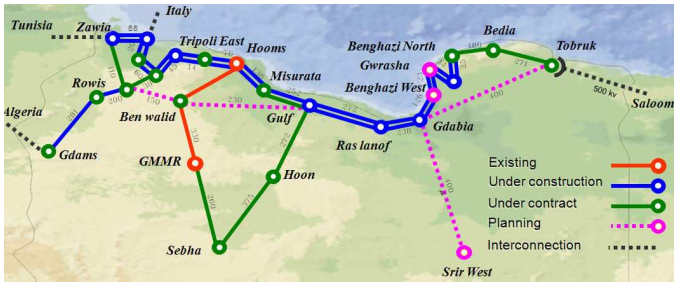


Fig. 4. The new Libyan 400 kV network

The future plan of the Libyan transmission system expansion includes also the infrastructure of:

- 256 km (1000 km cables core length) of 220 kV cables in cities, i.e. Tripoli, Benghazi, Zawia and Musrata,
- 71 substations of 220 kV in different locations in the country,
- Around 2000 km of new 220 kV of transmission lines.

IV. RENEWABLE ENERGY IN LIBYA

The share of renewable energy technologies in Libya up to now hold only a small contribution in meeting the basic energy needs, it is used to electrify rural areas for sustainable development, supply microwave repeater station, and in cathodic protection. A setup plane was planned for implementing renewable energy sources is to contribute a 10% off the electric demand by the year 2020. The short plane for renewable energy is to invest 500 million euro in the next five years. During the past three decades, photovoltaic is the most technology which has been used in rural applications, particularly for small- and medium- sized remote applications with proven economic feasibility, several constraints and barriers, including costs exist. The experience

Raised from PV applications indicates that there is a high potential of building a large scale of PV plants in the south of the Mediterranean. There is a great potential for utilizing, home grid connected photovoltaic systems, large scale grid connected electricity generation using Wind farms, and solar thermal for electricity Generation, with capacities of several thousands of MW. The high potential of solar energy in Libya may be considered as a future source of electricity for the northern countries of Mediterranean.

V. ELECTRICAL INTERCONNECTIONS OF LIBYA

The internal interconnection of the Libyan transmission system has been achieved since 1993 creating the high voltage 220 kV Libyan national grid. The electrical networks of Libya and Egypt are interconnected on 220 kV voltage level since May 1998. By the end of the last year 2008 the electrical interconnection between Libya and Tunisia is expected to be synchronized on the 220 kV.

TABLE X
CONTRACTED ENERGY EXCHANGE (IN MWH)

	Libya – Egypt		Libya – Tunisia	
	Import	***Export***	***Import***	***Export***
2003	16708	–	25038	–
2004	8565	–	32184	–
2005	11164	504	32491	–
2006	30292	175	–	–
2007	6612	1875	–	–
Total	73341	2554	89713	0

It is expected to construct and operate the 500-400 kV link with Egypt by the year 2010 and with Tunisia 400 kV during the period 2010-2015. Under study is the electrical interconnection on 400 kV between Libya and Algeria and the submarine 400 kV cable between Libya and Italy.

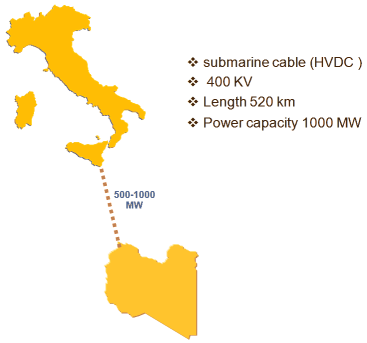


Fig. 5. Possible Interconnection with Europe: Libya – Italy (HVDC link)

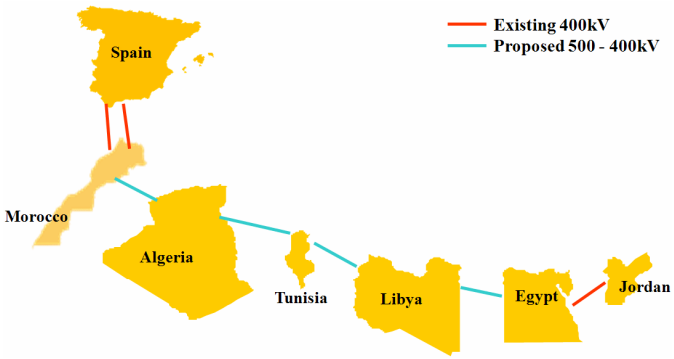


Fig. 6. Proposed 500-400 kV interconnection between ELTAM countries

VI. CONCLUSION

Completion of 400kV network permits power wheeling of about 500 MW in both directions. Availability of large quantities of natural gas will lead to development of new power station, by which we can export electricity to the neighboring countries.

ACKNOWLEDGMENT

I would like to thank prof. Ing. Michal Kolcun, PhD. for supervising my study.

REFERENCES

[1] The Euro-Mediterranean Energy Partnership. [online] [cited 12.3.2009] <http://www.auptde.org/NewSite/UploadFiles/Activityfile/153.ppt>
[2] General Electricity Company of Libya. [online] [cited 12.3.2009] <http://www.gecol.ly/gecol/index.php>
[3] Nasr M.: Electricity system of Libya and its future. In: SCYR 2009 - 9th Scientific Conference of Young Researchers : proceeding from conference: FEI TU, 2009. ISBN 978-80-553-0178-5.

State space controller for bidirectional DC/DC converter-buck mode

[1]*Matúš OCILKA*, [2]*Tomáš BÉREŠ*

[1]Dept. of Electrical, Mechatronic and Industrial Engineering FEI TU of Košice, Slovak Republic

[1]ocilka.matus@gmail.com, [2]tomas.beres@tuke.sk

Abstract—**This paper deals with designing of cascade state space controller for buck mode of bidirectional DC/DC converter. The control of converter is decomposed into outer voltage control and inner current control. First, the simple circuit of buck converter is created using the state space averaging method and simulated in Matlab/Simulink. The pole placement method is used to design the controller.**

***Keywords*—Buck converter, controller, state space model, pole placement**

I. INTRODUCTION

DC/DC converters are electronic systems which transfers input voltage to output load. There are many topologies of non-isolated DC/DC converters as buck, boost, Zeta, Ćuk. In this paper a buck mode of bidirectional cascade buck/boost converter is introduced which provides bidirectional flow of energy. Proposed control of converter is divided into three modes; buck mode, boost mode and buck/boost mode.

The aim of this paper is to design the controller for buck mode of converter. There are many control structures as voltage mode or current mode control, sliding mode control, delta sigma control, which provide output voltage independent of load or input voltage variations.

This paper deals with cascade state space controller for proposed mode of operation.

II. DC/DC BUCK CONVERTER

A. Modeling of buck converter

The simple circuit of converter is shown in Fig. 1. The converter consists of input voltage source, inductor, capacitor, switch and load resistor. For the sake of clarity the consideration of the following details shall be omitted: the influence of ESR (equivalent series resistance) of the output capacitor and the ohmic contribution of the inductive storage element on the control behavior; ideal switches are assumed. This can be done because the basic dynamic system quality is not affected [5].

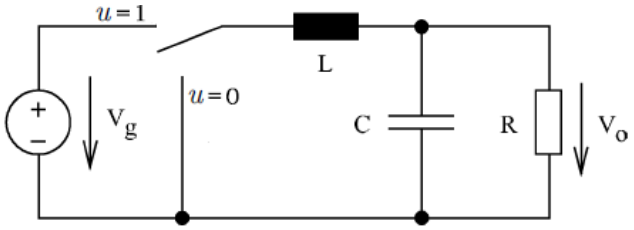


Fig. 1 The buck converter

First, let the switch position function to be *u=1*. Using Kirchhoff's laws we obtain set of equations

$$L\frac{di_L}{dt} = -u_c + V_g.u$$
$$C\frac{du_c}{dt} = i_L - \frac{u_c}{R} \quad (1)$$

When diode is in non-conduction mode *u=0* the equation results in

$$L\frac{di_L}{dt} = -u_c$$
$$C\frac{du_c}{dt} = i_L - \frac{u_c}{R} \quad (2)$$

By comparing the obtained particular dynamic systems descriptions, we immediately obtain the following *unified* dynamic system model. This result in

$$L\frac{di_L}{dt} = -u_c + V_g.u$$
$$C\frac{du_c}{dt} = i_L - \frac{u_c}{R} \quad (3)$$

We usually refer to this model as a switched model $u \in \{0,1\}$. The *average* converter model would be represented exactly by the same mathematical model, possibly by renaming the state variables with different symbols and by redefining the control variable *u* as a sufficiently smooth function taking values in the compact interval of the real line [0, 1]. In order to simplify the exposition, we shall refer to the model, with *u* replaced by u_{av}, as the *average model*. We shall however distinguish between the *average control input*, denoted by u_{av} and the *switched control input*, denoted by *u*. The average model of the Buck converter is then described by [3]

$$L\frac{di_L}{dt} = -u_c + V_g.u_{av}$$
$$C\frac{du_c}{dt} = i_L - \frac{u_c}{R} \quad (4)$$

B. State space model

Modeling using state space averaging is well known method since many years [4]. The state space averaged model of the converter can be expressed as

$$\dot{x} = Ax + bu$$
$$y = Cx \quad (6)$$

The state variables are the capacitor voltage and inductor current. Vector of state variables is then

$$x = [u_c, i_L]^T ; y = u_c \quad (7)$$

The matrices of the system can be written as follows

$$A = \begin{bmatrix} \frac{-1}{RC} & \frac{1}{C} \\ \frac{-1}{L} & 0 \end{bmatrix} ; b = \begin{bmatrix} \frac{V_g}{L} \\ 0 \end{bmatrix} ; c = [1 \quad 0] \quad (8)$$

III. Design Of Controller

A. Control structure

The general block diagram of cascade state space controller is shown in Fig. 2.

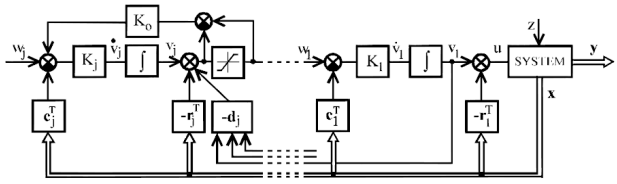


Fig.2 General block diagram of controller

Vector r^T realizes proportional gain of state vector and constants K_j are gain of the integrators. The control of buck mode is decomposed in outer voltage control loop and inner current control loop.

B. Current loop

First, consider the current equation of converter as inner loop of structure

$$\dot{x} = A.x + b.u + ez = Ax + b(v_2 - r_1^T x) + ez$$
$$\dot{v}_1 = K_2(u_2 - c_2^T x) \quad (9)$$

State space equation of current loop

$$\begin{bmatrix} \dot{x} \\ \dot{v}_2 \end{bmatrix} = \begin{bmatrix} A - br_2^T & b \\ -K_2 c_2^T & 0 \end{bmatrix} \begin{bmatrix} x \\ v_2 \end{bmatrix} + \begin{bmatrix} 0 \\ K_2 \end{bmatrix} u_2 + \begin{bmatrix} e \\ 0 \end{bmatrix} z \quad (10)$$

C. Voltage loop

According to Fig.2. the voltage control loop can be written as follows

$$u_1 = v_1 - r_1^T x - d_1 v_2 \quad ; \quad \dot{v}_1 = K_1(u_1 - c_1^T x) \quad (11)$$

State space equations of the new system result in

$$\begin{bmatrix} \dot{x} \\ \dot{v}_1 \\ \dot{v}_2 \end{bmatrix} = \begin{bmatrix} A - br_1^T & b & 0 \\ -K_2 c_2^T - K_2 r_1^T & K_2 d_1 & K_2 \\ -K_1 c_1^T & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ v_2 \\ v_1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ K_1 \end{bmatrix} u_1 + \begin{bmatrix} e \\ 0 \\ 0 \end{bmatrix} z \quad (12)$$

D. Pole placement

The main advantage of state space controllers is pole placement. This method allows placing poles of the system to obtain desired outputs. In this case we choose the damping factor ξ and time of regulation t_r. The natural frequency can be obtained from equation below

$$\omega_0 = \frac{1}{\xi.t_r}\left[3 - 0{,}5\ln(1-\xi^2)\right] \quad (13)$$

The poles of the system

$$s_{12} = -\xi.\omega_0 \pm j\omega\sqrt{1-\xi^2} \quad (14)$$

If system order is higher than two the rest of the poles are chosen as

$$s_i = -N\xi\omega_0 \; ; \; i=3.\ldots.n \quad (15)$$

The desired polynomial is then

$$P(s) = s^n + f_{n-1}s^{n-1} + \ldots + f_1 s + f_0 \quad (16)$$

Characteristic polynomial of the system can be obtained from state space model of system (10), (12)

$$P(\lambda) = \det(sI - A + br^T)$$
$$P(\lambda) = \lambda^n + f_{n-1}(r)\lambda^{n-1} + .. + f_1(r)\lambda + f_0(r) \quad (17)$$

By comparing between the characteristic polynomial (17) of system with the desired polynomial (16) gives the design equations of the feedback coefficients [1],[2].

E. Controller scheme

The designed controller was created in Matlab/Simulink according to equations (10), (12). V_{ref} is reference value of output voltage.

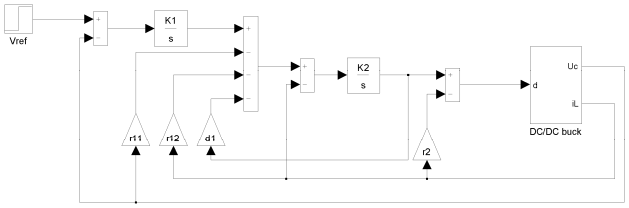


Fig.3 State space controller of buck converter

IV. Simulation Results

The controller was simulated in Matlab/Simulink to verify the properties of proposed controller and parameters of converter used in simulations are $L=47\ \mu H$,, $C=40\ \mu F$, $R=1.2\Omega$.

Damping factor and time of regulation for voltage loop $\xi=0.85$, $t_r=1ms$ and for current loop $\xi=0.85$, $t_r=0{,}5ms$. The desired poles for current loop are $s_{12} = -3.6410.10^4 \pm j2.2565.10^4$, $s_3 = -2.5487.10^5$ and poles of the voltage loop $s_{12} = -3.3367.10^3 \pm j3.4041.10^3$, $s3 = -1.0010.10^4$, $s4 = -1.3347.10^4$.

The coefficients of current loop are then $K_2 = 3.5932.103$, $r_2 = 0.1426$ and voltage loop coefficients $K_1 = 367.2795$, $d_1 = -12.9531$, $r_{12} = 0.9113$, $r_{11} = 0.6583$.

The performance of the converter with state space controller is shown in Fig.4.

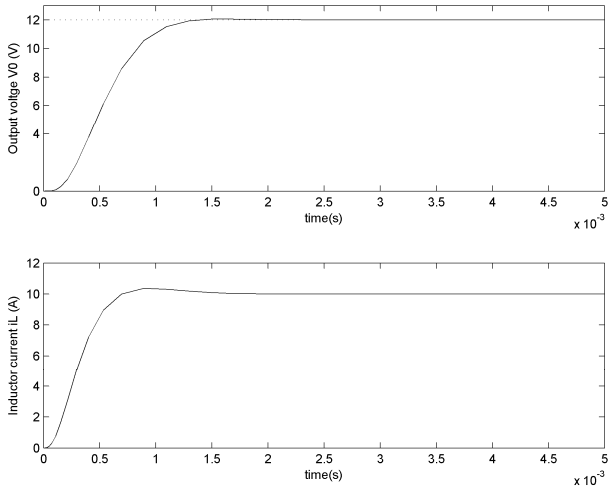


Fig.4 Performance of converter ($V_{ref}=12V$)

The reference of output voltage was set to 12V with nominal load of 10A. Output of converter reached the 5% of reference value in 1ms with minimal overshot. The inductor current settle down in 0.75ms again with minimal overshot caused by the damping factor which was set $\xi = 0.85$ in voltage and current loop. The simulation of the load step is shown in Fig.5. The load step causes a variation 1V in the output voltage and voltage settle down to its reference value in 1ms.

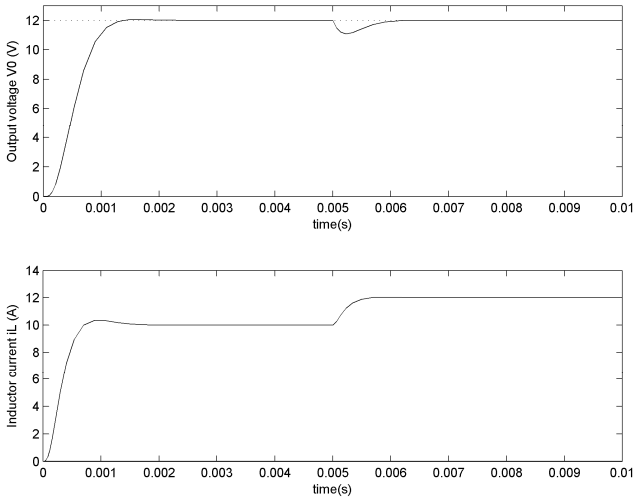


Fig.5 Performance of converter with load step ($I_z = 2A$)

V. Conclusion

In this paper the cascade state space controller for buck mode of bidirectional DC/DC converter was designed using state space averaging and pole placement method. The state space controllers advantages are pole placement and easy implementation by setting the time of regulation and damping factor. The proposed control structure was finally tested in Matlab/Simulink. Future work will be to design controllers for boost and buck/boost mode and compare all three modes with state space controllers to PI regulators.

Acknowledgment

This work was supported by Slovak Research and Development Agency under project APVV-0095-07 and by Scientific Grant Agency of the Ministry of Education of Slovak Republic under the contract VEGA No.1/0099/09.

References

[1] L. Zboray,F. Ďurovský, J.Tomko, *Regulované pohony*, Vienala, April 2000, ch.3

[2] L. Zboray,F. Ďurovský, *Stavové riadenie elektrických pohonov*, Vienala, 1995

[3] H. Sira-Ramirez, R. Silva-Ortigoza, *Control design techniques in power electronic devices*, Springer,2006, ch.2.

[4] G.Keller, D.Lascu, J.M.A.Myrzik, "*State-Space Control for Buck Converters with/without Input Filter*," Dresden, EPE 2005

[5] F. A. Himmelstoss, F. C. Zach, "*State space control for a step-up converter*,"INTELECT'91, Nov. 1991

[6] Z. Sütő, I.Nagy, "*Nonlinear dynamics and three-phase voltage source converters:review*", 16th Int. Conference on Electrical Drives and Power Electronics (EDPE 2007),,September 24-26,2007

[7] M.Olejár, V. Ruščin, M. Lacko, J. Dudrik. "*Bi-directional DC/DC converter for hybrid battery*," 16th Int. Conference on Electrical Drives and Power Electronics (EDPE 2007),September 24-26,2007

[8] J. Leuchter, P. Bauer, V. Řeřucha, P. Bojda, "*DC-DC converter with FPGA control for photovoltaic system*," 13th International Power Electronics and Motion Control Conference (EPE-PEMC 2008), 1-3 September 2008, Poznan, IEEE Catalog Number: CFP0834A-CDR, ISBN:978-1-4244-1742-1

[9] N. D. Trip, "*Analysis and experimental results of an active snubber for boost converters*," Proc.of the 6th Int.Conf. on Renewable sources and Enviromental Electro-Technologies – RSEE 06, ISSN 1454-9239, Oradea, pp. 125-128.

The type of chaotic sequences for signal transmission

Henrieta Palubová

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

henrieta.palubova@gmail.com

***Abstract*— The type of chaotic sequences for signal transmission is presented in this paper. The simulation results in Matlab are presented here. Chaotic sequences and another type of pseudorandom sequences (Walsh Hadamard, Golay, Gold) can be used in communications. Sequences derived from chaotic phenomena are actively being considered for spread-spectrum communications.**

***Keywords*—Logistic map, chaotic map, chaotic sequences, tent function.**

I. INTRODUCTION

Chaotic signals are non periodic, wide band, and more difficult to predict and to reconstruct. Chaos is a deterministic, random-like process found in non-linear dynamical system, which is non-converging and bounded. Moreover, it has a very sensitive dependence upon its initial condition and parameters. These properties make chaotic signals more difficult to intercept and to decode the information spreaded upon them [6]. In last decade there has been an increasing amount of interest in chaotic sequences [7], [1], [8]. Comparison of chaotic sequences and another type of pseudorandom sequences is described in [9]. The chaotic sequence can be used as spread-spectrum sequence in place of Pseudorandom sequence in CDMA communication systems. The three types of chaotic sequences for signal transmission are described in this paper. The using of these types of chaotic sequences is described in [5].

II. CHAOTIC SEQUENCES

The pseudo-noise sequences such as Gold sequences and Walsh Hadamard sequences are the most popular spreading sequences that have good correlation properties, limited security and are reconstructed by linear regression attack for their short linear complexity. Generator of pseudorandom sequences (Gold, Walsh Hadamard, Golay) can generate the finite number of states for several users. A chaotic sequence generator can get an infinite number of states in a deterministic manner and therefore produce a sequence which never repeats itself [2].

In this paper the three types of chaotic sequences are described. The first one is generated by improved logistic map, the second one is generated from chaotic map with different slopes and the third one is generated by tent function.

A chaotic map is a discrete-time dynamical system

$$x_{k+1} = f(x_k), \quad 0\langle x_k \langle 1, \quad k = 0,1,2,.... \tag{1}$$

running in chaotic state. The chaotic sequence

$$\{x_k : k = 0,1,2,....\} \tag{2}$$

can be used as spread-spectrum sequence in place of spreading code of conventional Direct-Sequence, Spread-Spectrum (DS/SS) CDMA communication systems. Chaotic sequences are uncorrelated when their initial values are different (Fig. 8)|, so in chaotic spread-spectrum systems, a user corresponds to an initial value.

A. Improved logistic map

Improved logistic-map is defined by:

$$x_{k+1} = f(x_k) = 1 - 2(x_k)^2, \quad x_k \in (-1,1) \tag{3}$$

The chaotic sequence:

$$\{x_k : k = 0,1,2,....\} = \{f^k(x_0) : k = 0,1,2,....\}, \tag{4}$$

generated by improved logistic-map (Fig. 1), is neither periodic nor converging (Fig. 2) , and sensitively dependent on initial value (Initial value is set to x_k = 0,2). The chaotic sequence is random-like, so probability and statistics can be used in discussing their characteristics [1].

The average of chaotic sequence is [1]

$$\bar{x} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} x_i = \int_0^1 x\rho(x)dx = 0 \tag{5}$$

$\rho(x)$ - density function does not depend on initial value

The auto-correlation function of chaotic sequence is

$$ac(m) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} (x_i - \bar{x})(x_{i+m} - \bar{x}) =$$

$$= \int_0^1 x f^m(x)\rho(x)dx - \bar{x}^2 = \begin{cases} 0.5 & (m = 0) \\ 0 & (m \neq 0) \end{cases} \tag{6}$$

If two chaotic sequences beginning with differential values x_{10} and x_{20} are not overlapping, the cross-correlation function of these two sequences is

$$cc_{12}(m) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} (x_{1i} - \bar{x})(x_{2(i+m)} - \bar{x}) =$$

$$= \int_0^1 \int_0^1 x_1 f^m(x_2)\rho(x_1, x_2)dx_1 dx_2 - \bar{x}^2 = 0 \qquad (7)$$

From these properties we know, that the chaotic sequences is identical with white noise whose average is zero [1].

B. *Chaotic map with different slopes*

To generate the chaotic sequence with biased values, chaotic sequence, when initial value x_k=0,2 shown in Fig.5 is used and it was made from the Bernoulli shift map (Fig. 3). This map is described by:

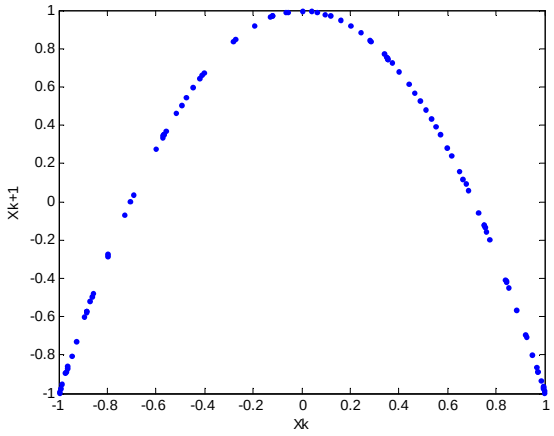


Fig. 1 Chaotic sequences (x_k in first step is set to 0.2)

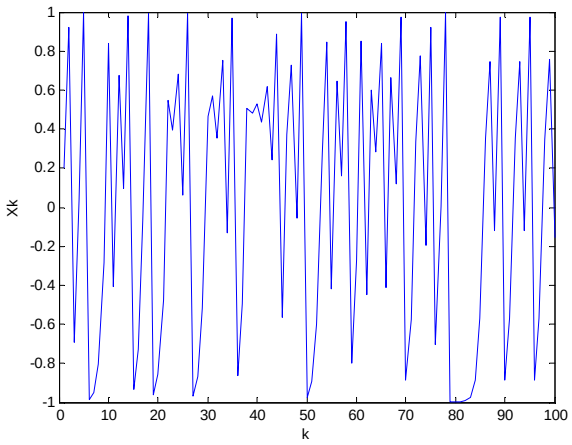


Fig. 2 Chaotic sequence generated from improved logistic map

$$x_{k+1} = \begin{cases} \dfrac{(1+r)x_k + q + r}{1-q} & (-1 \le x_k \le -q) \\ \dfrac{(1-r)x_k + q}{q} & (-q < x_k \le 0) \\ \dfrac{(1-r)x_k - q}{q} & (0 < x_k \le q) \\ \dfrac{(1+r)x_k - q - r}{1-q} & (q < x_k \le 1) \end{cases} \qquad (8)$$

The slopes of the map can be changed by deciding parameters q ($0.5 \le q \le 1.0$) and r ($0.0 \le r \le 1.0$).

As an example, chaotic sequence with biased values is shown in Fig.4. The parameters of chaotic sequences have been designed by author. From Fig. 3, it can be observed that the chaotic sequence includes many values near 1.0 (or -1.0) as both q and r approach 1.0, namely, it can be said that deviation was made to the values of the chaotic sequence [3].

C. *The tent function map*

The tent function is the next model of chaos. The map of tent function is described by:

$$x_{k+1} = f(x_k) \qquad (9)$$

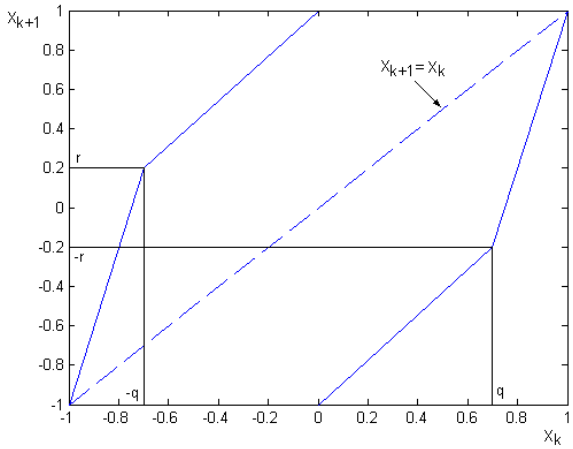


Fig. 3 Chaotic map with different slopes

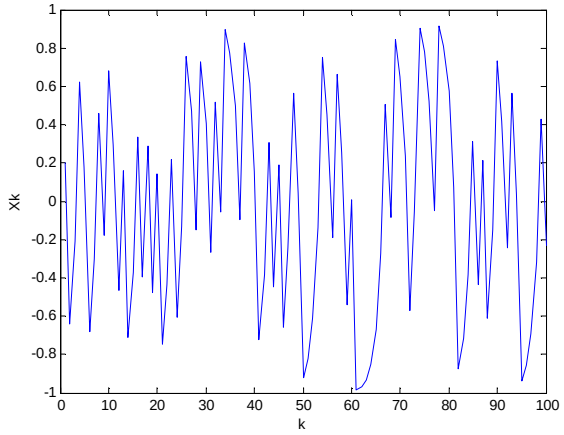


Fig. 4 Chaotic sequences with biased values (q,r) = (0.5, 0.1), x_k = 0.2

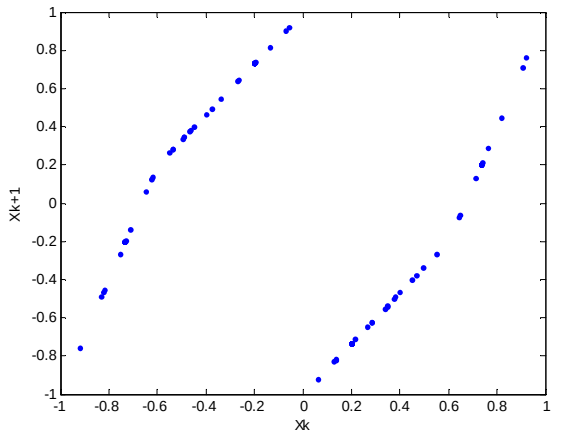


Fig. 5 Chaotic sequences (x_k in first step is set to 0.2)

The chaotic sequence generated by tent function is described by [4]:

$$x_{k+1} = f(x_k) = \begin{cases} \dfrac{2x_k + 1 - a}{a+1} & -1 \le x_k \le a \\ \dfrac{2x_k - 1 + a}{a-1} & a \le x_k \le 1 \end{cases} \qquad (10)$$

$$a \in [-1,1],\ x \in [-1,1]$$

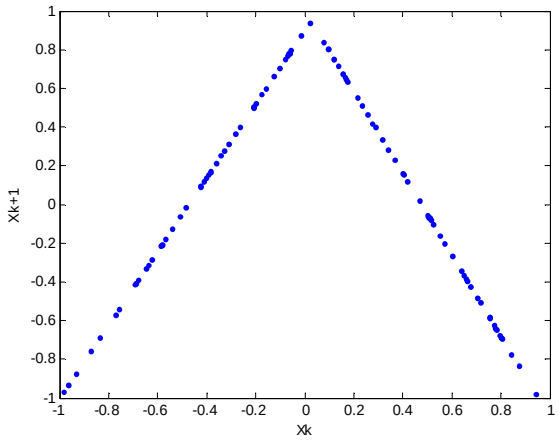


Fig. 6 The tent function

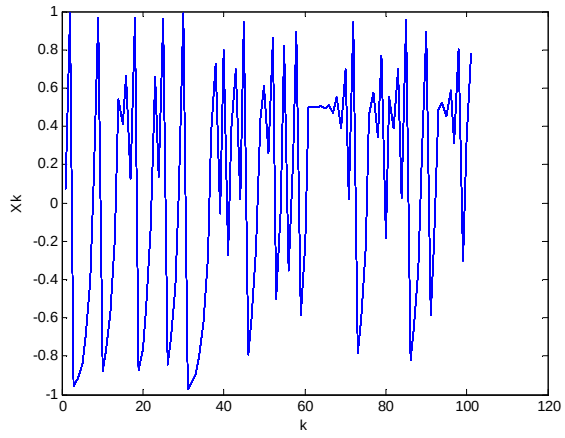


Fig. 7 Chaotic sequence generated by tent function (x_k = 0.075, a = 0.5)

As an example, Fig. 8 describes the crosscorelation function of chaotic sequence on Fig. 1 and Fig. 4. Fig. 9 describes the autocorrelation function of chaotic sequence on Fig. 1. The crosscorelation function and autocorrelation function are the very important parameters for signal transmission. If two signals are correlated they have worse transmission properties. Sets of non-correlated sequences with good autocorrelation and crosscorelation properties are required in order to provide low interference between users.

III. Conclusion

Three types of chaotic sequences for signal transmission are described in this paper. Chaotic sequences are generated from their chaotic or logistic maps. Figures of chaotic sequences, logistic maps, autocorrelation and croscorrelation functions in Matlab are presented here. The chaotic sequences presented in this paper can be used as spread – spectrum sequences in CDMA systems. A chaotic sequence generator can visit an infinite number of states in a deterministic manner and therefore produce a sequence which never repeats itself. Generator of pseudorandom sequences such as Golay, Zadof-Chu, Gold, Walsh Hadamard, can generate the finite number of states for several users. Comparison of chaotic sequences and another type of pseudorandom sequences in MC – CDMA Systems is described in [5].

Acknowledgment

This work was supported by the project VEGA 1/0045/10 – Nové metódy spracovania signálov pre rekonfigurovateľné bezdrôtové senzorové siete.

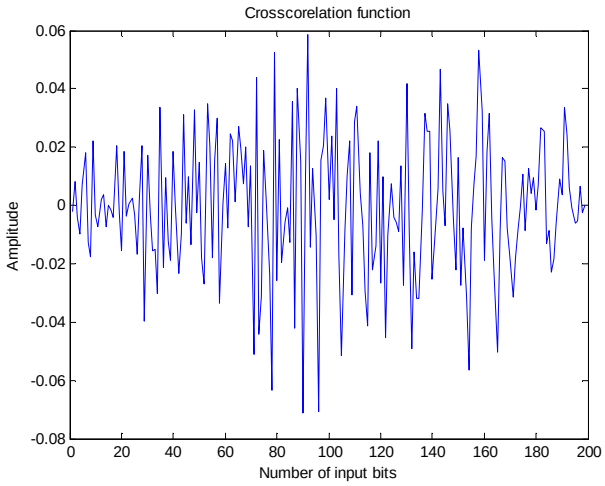


Fig. 8 Crosscorelation function (Chaotic sequence – Fig. 1 and Chaotic sequence – Fig. 4)

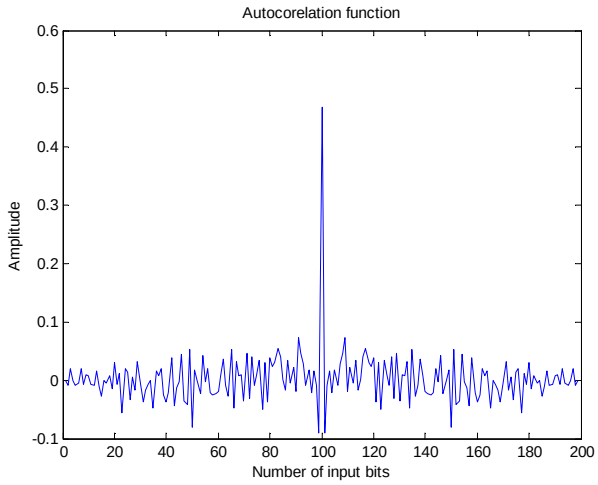


Fig. 9 Autocorelation function (Chaotic sequence - Fig. 1)

References

[1] W. Hai, H. Jiandong.: "Chaotic Spread – Spectrum Communication Using Discrete – Time Synchronization", The Journal of China Universities of Posts and Telecommunications, Vol. 4, No. 1, Jun 1997.

[2] V.Nagarajan, P.Dananjayan and L.Nithyanandhan.: "On The Performance Of Chaotic Spreading Sequence MIMO MC/DS – CDMA Systems Using NPGP", International Conference On

Control, Automation, Communication And Energy Conservation, jún 2009

[3] S. Arai, Y. Nishio, "Research on Differential Chaos Shift Keying Changing Deviation of Chaotic Sequence", 2007 RISP International Workshop on Nonlinear Circuits and Signal Processing, Shanghai Jiao Tong University, Shanghai, China, Mar. 2007.

[4] B. Smith M. Razím, J. Štecha,: „ Nelineární Systémy", ČVUT Praha, 1997

[5] H. Palubová,: „Chaotic sequences in MC-CDMA Systems", (unpublished work)

[6] G. Kaddoum, D. Roviras, P. Charge, D. Fournier-Prunaret,:" Performance of multi-user chaos-based DS-CDMA System over multipath channel", IEEE, 2009

[7] I. Vasilache, D. Grecu, C. Grozea, B. Cristea,:" Random Sequence Generator Based on Nonlinear Function", International conference, 2002, Bucharest

[8] P. H. Lee, S. Ch. Pei, Y. Y. Chen,:" Generating Chaotic stream chiphers using Chaotic systems", Chinese journal of Physics, vol. 41, No. 6, December 2003

[9] L. Cong, - L. Shaoqian,: "Chaotic Spreading Sequences with Multiple Access Performance Better Than Random Sequences", IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications, Vol. 47, No. 3, March 2000

Chaotic sequences in MC-CDMA Systems

Henrieta Palubová

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

henrieta.palubova@gmail.com

***Abstract*—The chaotic sequences in MC-CDMA Systems are presented in this paper. Performance evaluation and comparison of multi-carrier code division multiple access system model for different spreading sequences with chaotic sequences at the presence of Saleh and Rapp model of high power amplifier (HPA) is investigated. The simulation results in Matlab are presented here.**

***Keywords*—MC-CDMA, Saleh model, Rapp model, Chaotic sequences**

I. INTRODUCTION

Sequences derived from chaotic phenomena are actively being considered for spread-spectrum communications [10]. In recent years there has been an increasing amount of interest in chaotic sequences in CDMA systems. MC/DS CDMA is described in [11]. Comparison of chaotic sequences and another type of pseudorandom sequences in CDMA systems is described in [13], [14], [15]. This paper is deal with chaotic sequences in MC CDMA systems. Chaos is a deterministic, random-like process found in non-linear dynamical system, which is non-periodical, non-converging and bounded. Moreover, it has a very sensitive dependence upon its initial condition and parameters. The generation of orthogonal sequences is utmost importance in MC-CDMA systems, in order to increase the spectrum efficiency in multirate communications systems. In CDMA, sets of non-correlated sequences with good autocorrelation and crosscorelation properties are required in order to provide low interference between users [12].

Section II. described MC-CDMA system model, section III. deal with nonlinear models, section IV. described chaotic sequences versus another type of pseudorandom sequences in MC-CDMA systems.

II. MC-CDMA SYSTEM MODEL

In MC-CDMA, instead of applying spreading sequences, in the time domain, we can apply them in the frequency domain, mapping a different chip of a spreading sequence to an individual Orthogonal Frequency Division Multiple Access (OFDM) subcarrier. Hence each of OFDM subcarrier has a data rate identical to the original input data rate and the multicarrier system absorbs the increased rate due to spreading in wider frequency band.

In MC-CDMA transmitter, the information bits to be transmitted by a particular user, are firstly base-band modulated (QAM, PSK) into some modulation symbols and then are spreaded by using a specific spreaded sequence c_m. In the case of MC-CDMA, as the spreading codes Walsh codes, Gold codes, Zadoff-Chu codes, Golay codes and Chaotic sequence codes can be used. The spreaded symbols are modulated by multi-carrier modulation implemented by IFFT (Inverse Fast Fourier Transform) operation. The IFFT after parallel-to-serial conversion represents the input signal of a HPA (High Power Amplifier), (see Fig. 1). The receiver consists of serial-to parallel converter, FFT (Fast Fourier Transform), BMF (receiver-Bank of Matched Filters), block of linear or non-linear transformation (labelled as T) and a decision device. Here, the operation of a single-user receiver known as BMF consists of a set of simple matched filters (correlators). In order to extend BMF into a multi-user receiver, the T-transformation block is included in the receiver structure [3]. In this paper, the linear MMSE-MUD [4] as well as nonlinear MSF-MUD for MC-CDMA [5], [6] are considered. The T-transformation block in MMSE-MUD is represented by multi-channel linear Wiener filter. In the case of MSF-MUD, the T-transformation block is represented by a complex valued-multichannel nonlinear microstatistic filter (C-M-CMF). C-M-CMF is the minimum mean-square error estimator based on the estimation of desired signals by using a linear combination of vector elements obtained by threshold decomposition of filter input signals [5], [2].

The main benefit of combining OFDM with DS-spreading is that it is possible to prevent the obliteration of certain subcarriers by deep frequency domain fades [1].

A block diagram of the simplified baseband model of MC-CDMA transmitter is given in Fig. 1 [2].

The basic structure of receivers considered in this paper is sketched in Fig. 2.

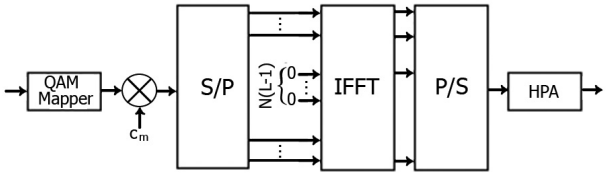


Fig. 1 MC CDMA transmitter

III. NONLINEAR MODELS

There are two major types of power amplifiers used in communications systems:

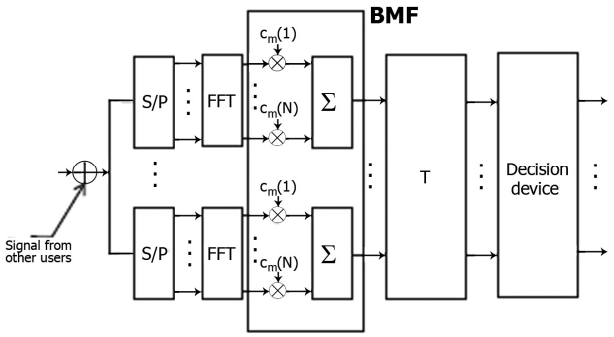


Fig. 2 MC CDMA receiver

- Traveling wave tube amplifiers (TWTA)
- Solid state power amplifiers (SSPA)

A common characteristic of both devices is that the signal at their output is a nonlinear function of the input signal at both the present and previous instants [7]. The output y(t) of the nonlinear amplifier is given by

$$y(t) = F(|x(t)|)\exp(j\Phi(|x(t)|) + jorg(x(t))) \quad (1)$$

where $|x(t)|$ is the amplitude (nonnegative voltage envelope) of $x(t)$, arg $(x(t))$ is the phase of $x(t)$, $F(\bullet)$ is the amplitude-to-amplitude (AM/AM) conversion and $\Phi(\bullet)$ is the amplitude-to-phase (AM/PM) conversion [9].

For Saleh Model is AM/AM given as

$$G(u_x) = \frac{\kappa_G . u_x}{1 + X_G . u_x^2} \quad (2)$$

and AM/PM as

$$\Phi(u_x) = \frac{\kappa_\Phi . u_x^2}{1 + X_\Phi . u_x^2} \quad (3)$$

The Saleh model is commonly used for TWTA modelling.

For Rapp Model is AM/AM given as:

$$G(u_x) = \frac{\kappa_G . u_x}{\left(1 + \left(\frac{u_x}{O_{sat}}\right)^{2s}\right)^{\frac{1}{2s}}} \quad (4)$$

and AM/PM as

$$\Phi(u_x) = 0 \quad (5)$$

The Rapp model is commonly used for SSPA modelling.

The AM/AM and AM/PM are nonlinear characteristics where nonlinearity depends on position of operating point.

The operating point of the amplifier is defined by input back-off (IBO) parameter, which is determined by the ratio between the saturation power of the amplifier and the average power of the signal. The HPA operation in the region of its nonlinear characteristic causes a nonlinear distortion of transmitted signal, that subsequently results in higher BER and outof-band energy radiation. IBO is given as

$$IBO = 10\log_{10}\left(\frac{P\max, in}{\overline{Px}}\right) \quad [dB] \quad (6)$$

The measure of effects due to nonlinear HPA could be decreased by the selection of relatively high value of IBO.

IV. Chaotic Sequences Versus Another Type Of Pseudorandom Sequences In MC-CDMA system

MC-CDMA performance analysis presented in this section is based on computer simulations. The basic scenario of the simulations is represented by the uplink MC-CDMA transmission system performing through AWGN transmission channel, at 16-QAM or 8-PSK baseband modulation, for K active users (K = 3 and K= 9).

As the spreading sequences, Walsh codes, Gold codes with period of 32 chips as well as complementary Golay codes, Zadoff-Chu codes and Chaotic sequences with period of 31 chips have been applied. The total number of sub-carriers has been set to N = 128. In order to avoid aliasing and the out-of-band radiation into the data bearing tones, the oversampling rate of 4 has been applied [2]. Then, Nu = 32 (Walsh codes, Gold codes) and Nu = 31 (complementary Golay codes, Zadoff-Chu codes and Chaotic sequences) carriers per transmitted block have been used for the spread 16-QAM and 8-PSK symbol transmission. The three types of systems are described here. The first one is the Linear MC-CDMA System, the second one is the nonlinear MC-CDMA System with Saleh Model and the third one is the nonlinear MC-CDMA System with Rapp Model.

A. Linear MC-CDMA System

The number of 100 000 input bits, the number of 3 users and the modulation type of 16-QAM and 8-PSK was used for simulations.

In the Fig. 3 the signal constellations at the outputs of 16-QAM mapper and BMF for Eb/No = 12 dB are given. For first user chaotic sequence is generated by chaotic map with different slopes with *q = 0.5, r = 0.1* and initial condition x_0 = *0.1*. For second user chaotic sequence is generated by logistic map with initial condition x_0 = *0.01*. For third user chaotic sequence is generated by tent function with initial condition x_0 = *0.012*, and parameter *a = 0.05*. The type of chaotic sequences is detailed described in [8].

In the Fig. 4, the BER vs. Eb/No for MC-CDMA transmission system for different spreading sequences and 16-QAM is given. The AWGN channel model and 9 users was used in this simulation. It can be seen from Fig. 4, that all the types of sequences and receivers have the similar performance, except chaotic sequences in combination with BMF. Chaotic sequences in combination with BMF have the worse performance.

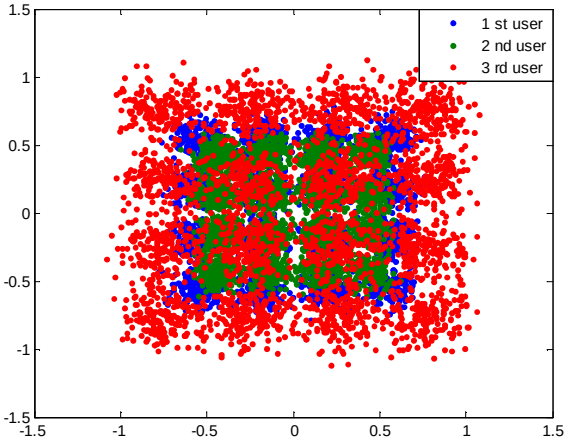


Fig. 3 Original symbol constellation at the output of the 16-QAM

B. *Nonlinear MC-CDMA System – Saleh Model*

For the specification of the Saleh model of HPA, the parameters $k_G = 2$, $\chi_G = \chi_\Phi = 1$ and $k_\Phi = \pi/3$ have been chosen.

The Saleh nonlinearity type has very destructive effect on QAM modulation (Fig. 5) [9]. The number of 100 000 input bits, the number of 3 users and the modulation type of 16-QAM or 8-PSK was used for simulation. In the Fig. 5, the signal constellations at the outputs of 16-QAM mapper and BMF for Eb/No = 12 dB are given.

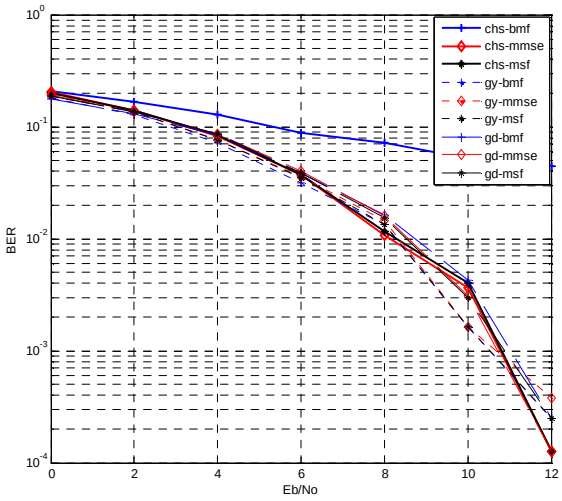


Fig. 4 BER vs. Eb/No for MC-CDMA transmission system for different spreading sequences and 16-QAM modulation

In the Fig. 6, the BER vs. Eb/No for MC-CDMA transmission system for different spreading sequences and 8-PSK is given. The AWGN channel model, 9 users and IBO = 2 dB was used in these simulations. It can be seen from Fig. 6, that all the types of sequences and receivers have the similar performance, except sequences in combination with BMF. All the types of sequences in combination with BMF have the bad performance.

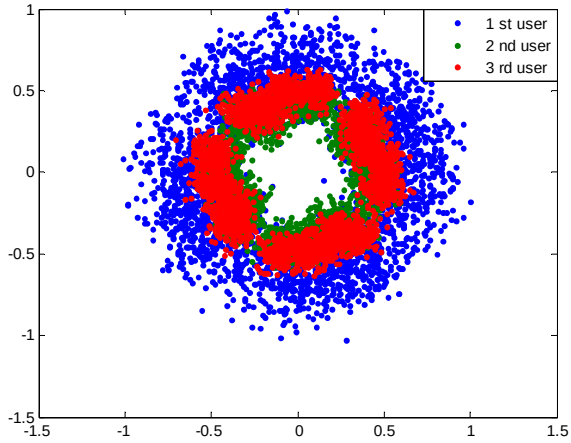


Fig. 5 Original symbol constellation at the output of the 16-QAM – Saleh model

C. *Nonlinear MC-CDMA System – Rapp Model*

For the specification of the Rapp model of HPA, its parameters have been set to $k_G = O_{sat} = 1$ and $s = 3$.

The number of 100 000 input bits, the number of 3 users and the modulation type of 16-QAM and 8-PSK was used in simulation. In the Fig. 7 the signal constellation at the outputs of 16-QAM mapper and BMF for Eb/No = 12 dB are given.

In the Fig. 8, the BER vs. Eb/No for MC-CDMA transmission system for different spreading sequences and 8-PSK is given. The AWGN channel model, 9 users and IBO = 2 dB was used in this simulation. It can be seen from Fig. 8, that the best performance can be provided when we used Golay sequences in combination with MSF-MUD, MMSE-MUD or BMF. When we used the chaotic sequences, MSF-MUD and MMSE-MUD have the same performance, receiver BMF has the bad performance. The worse performance has the Zadof-Chu sequences.

V. Conclusion

In this paper, the performance of MC-CDMA transmission system for two different models of HPA (Saleh and Rapp model), the different spreading sequences and receiver types is investigated. It has been found that Saleh model of HPA introduces much higher nonlinear distortion and causes more significant degradation of MC-CDMA transmission system performance than that of Rapp model. The best performance we can obtained when we used the Golay sequences in combination with MSF-MUD or MMSE-MUD. Chaotic sequences have similar performance like Golay sequences. MMSE-MUD and MSF-MUD have equivalent performance in linear and nonlinear MC-CDMA system. The worse performance we can obtain with using Gold or Zadof-Chu sequences. When we compare the modulation type, the best performance has 8-PSK. 16-QAM and 16-PSK have the similar performance. Presented results are performed by the author. It can be seen from simulation results, that chaotic sequences can be used for spread- spectrum communication. A chaotic sequence generator can get an infinite number of states in a deterministic manner and therefore produce a sequence which never repeats itself [11], whereas generator of

pseudorandom sequence (Golay, Zadof-Chu, Gold) can generate the finite number of states for several users.

ACKNOWLEDGMENT

This work was supported by the project VEGA 1/0045/10 – Nové metódy spracovania signálov pre rekonfigurovateľné bezdrôtové senzorové siete.

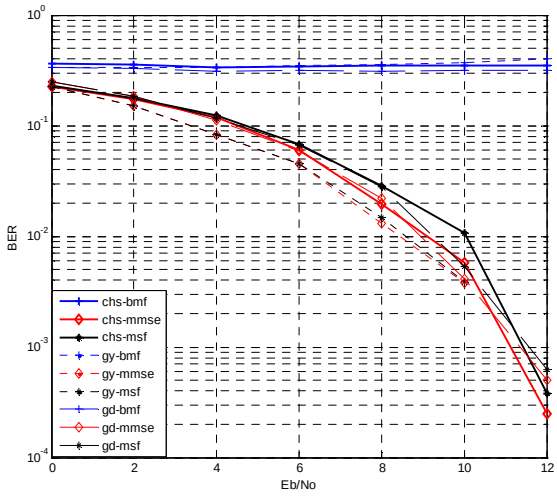


Fig. 6 BER vs. Eb/No for MC-CDMA transmission system for different spreading sequences (8-PSK modulation, Saleh model, IBO = 2 dB)

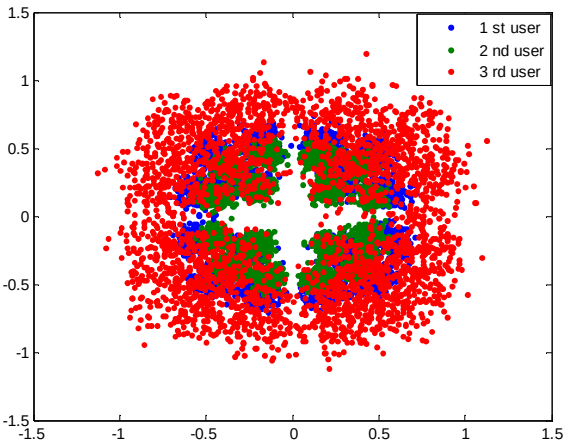


Fig. 7 Original symbol constellation at the output of the 16-QAM – Rapp model

REFERENCES

[1] http://wireless.per.nl/reference/chaptr05/cdma/rake. htm

[2] P. Drotár, – J. Gazda, – D. Kocur, – P. Galajda,: „Effects of spreading sequences on the performance of MC-CDMA System with nonlinear models of HPA“, Radioelektronika, 2008, 18th International Conference

[3] D. Kocur, - J. Čížová, - S. Marchevský, : “ Nonlinear microstatistic multi-user receiver”, Acta Electrotechnica et Informatica., 2003, vol.3, no.3, pp. 10-15

[4] L. Hanzo, – M. Munster, - B. J. Choi, – T. Keller,: OFDM and MC CDMA for Broadband Multi-User Communications, WLANs and Broadcasting, John Willey & Sons, Ltd, England, 2003

[5] J. Krajňák, - M. Deumal, - P. Pavelka, D. Kocur, P. Galajda, J. L. Pijoan,: “Multiuser detection of nonlinearly distorted MC-CDMA symbols by microstatistic filtering”, Wireless Personal Communications, Oct 2008, vol. 47, no. 1, pp. 147-160

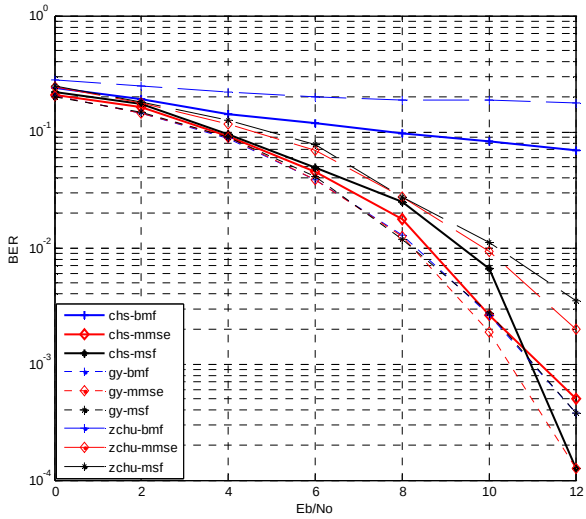


Fig. 8 BER vs. Eb/No for MC-CDMA transmission system for different spreading sequences (8-PSK modulation, Rapp model, IBO = 2 dB)

[6] J. Krajňák, - P. Pavelka, - P. Galajda, - D. Kocur,: “Efficient design procedure of microstatistic multi-user detector for nonlinearly distorted MC-CDMA”, In Proceedings of 17th International Conference Radioelektronika 2007, Brno (Czech Republic), 2007, pp. 147-152

[7] Dr. J. L. Pijoan Vidal – M. D. Herraiz.: „On OFDM Systems with low sensitivity to nonlinear amplification“, Presentation COST 289, 2002 - 2007

[8] H. Palubová,: „The type of chaotic sequences for signal transmission“, (unpublished work)

[9] L. Ch. Chang, - J. V. Krogmeier,: „Power optimization of nonlinear QAM system with data predistortion“, 14th IST Mobile and wireless Communications Summit, Dresden, Germany June 2005

[10] M. Suneel,: “Chaotic Sequences for Secure CDMA”, Ramanujan Institute for Advanced Study in Mathenatics, Chennai 600005, February 6-8, 2006

[11] V.Nagarajan, P.Dananjayan and L.Nithyanandhan.: “On The Performance Of Chaotic Spreading Sequence MIMO MC/DS – CDMA Systems Using NPGP”, International Conference On Control, Automation, Communication And Energy Conservation, jún 2009

[12] D. Sandoval – Morantes, – D. Munoz – Rodriguez,: „Chaotic sequence for multiple Access“, IEE Electronics Letters Online, No: 19980132, November 1997

[13] L. Cong, - L. Shaoqian,: ”Chaotic Spreading Sequences with Multiple Access Performance Better Than Random Sequences”, IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications, Vol. 47, No. 3, March 2000

[14] Z. B. Jemaa, - S. Marcos, - S. Belghith,:” Performance of the Super Stable Orbits based Spreading Sequences in a DS – CDMA System with a MMSE receiver”, 14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

[15] G. Mazzini, - G. Setti, - R. Rovatti,: “ Chaotic complex spreading Sequences for Asynchronous DS – CDMA – Part 1: System modeling and results”, IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications, Vol. 44, No. 10, October 1997

Cascade H-bridge Inverter for Photovoltaic System

[1]*Marek PÁSTOR*, [2]*Marcel BODOR*

[1, 2]Dept. of Electrical Engineering, Mechatronics and Industrial Engineering, FEI TU of Košice, Slovak Republic

[1]marek.pastor@student.tuke.sk, [2]marcel.bodor@tuke.sk

***Abstract*—The main objective of this paper is to describe a cascade H-bridge inverter with focus on photovoltaic applications. The cascade H-bridge inverter is compared to a single H-bridge inverter mainly from THDi and efficiency point of view.**

***Keywords*—cascade H-bridge inverter, current control voltage source inverter, multilevel converters, photovoltaics.**

I. INTRODUCTION

The solar energy and especially photovoltaics is one of the fastest growing industries in the world. There is a demand for high quality electrical energy and thus the use of photovoltaics is almost impossible without modern power electronics. If we omit the simplest PV battery charger there always have to be certain power conditioning unit (PCU) between the PV generator and the load whether to maximize the energy yield or to change certain qualities of the electrical energy. Whether it is a stand alone PV electrical generator or a grid connected system there is a demand to change the DC voltage to the AC voltage, to maximize the energy yield and to monitor the whole system. This is done by the mean of a PV inverter. The use of the PV inverter is to change the DC voltage to the AC voltage and to adapt the PV generator to the electrical load as well as to monitor the whole system. There are several types of PV inverters according to the topology. If there is a need for a galvanic isolation between the PV generator and the grid, the PV inverter with a transformer has to be use. The PV inverter can utilize a low frequency transformer with sufficient filter at the inverter's output or a high frequency transformer. The PV generator voltage does not always meet the required value and thus this voltage needs to be changed. This can be done by a DC/DC converter at the PV inverter's input. The DC/DC converter can utilize the high frequency transformer. This paper describes the cascade H-bridge inverter which can be used for photovoltaic applications.

II. CASCADE H-BRIDGE INVERTER

A. Basics

Cascade inverters belong to the multilevel power converters. Multilevel power converters are mainly used for medium and high power application due to utilization of several power semiconductor switches with separated DC sources connected in series. Multilevel power converters have several advantages over single level power converters [1]: staircase output voltage, low common mode voltage, low distortion input current, and lower switching frequency. The disadvantages are higher number of power semiconductor switches, more complex control technique and higher conduction losses.

B. Cascade H-bridge inverter

A single-phase structure of a 7-level cascade H-bridge inverter is shown in Fig.1. The nominal power of the proposed cascade H-bridge inverter is 3600W. Maximal power depends on the IGBTs and heat sinks.

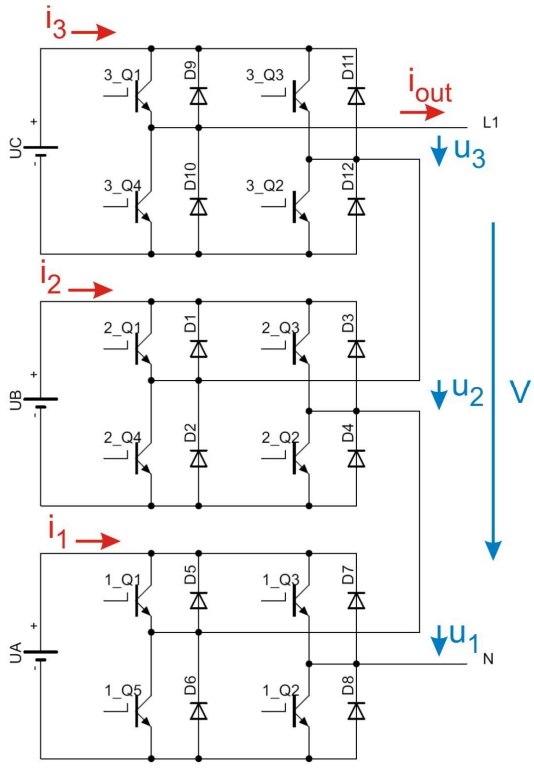


Fig. 1. Single-phase cascade H-bridge inverter with three separated DC sources (U_A = 240V, U_B = 120V, and U_C = 60V), capable to create 15 voltage levels at its output.

The number of output phase voltage levels n is defined by:

$$n = 2d + 1 \tag{1}$$

where:
d – is the number of separated DC sources.

However it is possible to create more voltage levels at the output of the cascade inverter. Each H-bridge converter can create positive, negative or zero voltage on its output with magnitude equal to the DC source. Thus there are 15 possible combinations for the cascade H-bridge inverter with 3 separated DC sources.

C. Output voltage control technique

There are several methods for a voltage control of the cascade inverter. One of them is the sinusoidal PWM from high switching frequency PWM modulation strategies [1].

The amplitude modulation index for the multilevel inverter is defined by:

$$m_a = \frac{A_m}{(n-1)A_C} \tag{2}$$

where:
A_m – is the modulation signal amplitude,
A_C – is the carrier signal amplitude,
n – is the output voltage level number.

The frequency modulation index is defined by:

$$m_f = \frac{f_C}{f_m} \tag{3}$$

where:
f_C – is the frequency of the carrier signal,
f_m – is the frequency of the modulation signal.

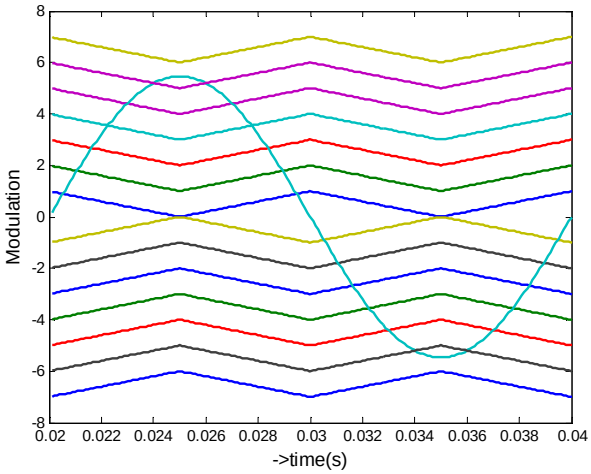


Fig. 2. Modulation and carrier signals for 15-level cascade H-bridge inverter (m_a = 0.78, m_f = 2).

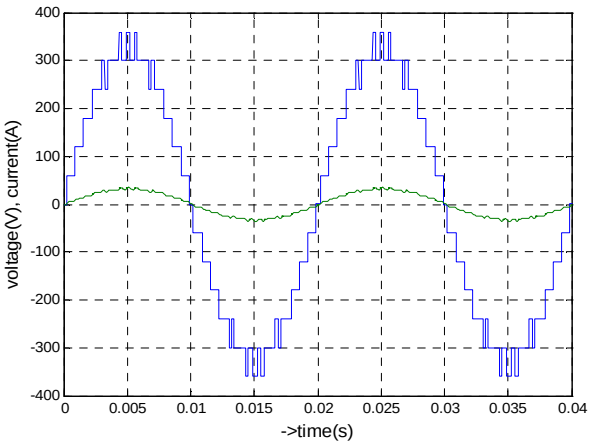


Fig. 3. Output voltage and current of 15-level (3 DC sources) cascade H-bridge inverter with voltage control (m_a=0.78, m_f=30, L=1mH, R=100Ω, U=232V, THDu=9,4%).

D. Current control technique

If we consider the DC/DC converter at the inverter's input this DC/DC converter acts as a voltage source. Thus the inverter must be a voltage source inverter (VSI). There are two main control strategies for VSI: thevoltage control (VCVSI) and the current control (CCVSI). They vary in the way they control the power flow. The VCVSI uses the control of the decoupling inductor voltage to control the power flow and the CCVSI uses the decoupling inductor current to control the power flow. The CCVSI is faster, can control active and reactive power flow independently but can not provide the voltage support to the load, can not operate without the grid. The CCVSI can be used for power factor correction due to the fact, that it can control the reactive power independently [2]. It also has a limited short circuit current compared to the VCVSI.

There are various techniques how to archive the current control in CCVSI. One of them is a predictive current control for voltage source inverters [3].

The easiest case is to use a simple RL filter to decouple the grid voltage *E* and the inverter's output voltage *V*. For circuit in Fig.4 it can be written:

$$\bar{V} = R\bar{I} + L\frac{d\bar{I}}{dt} + \bar{E} \tag{4}$$

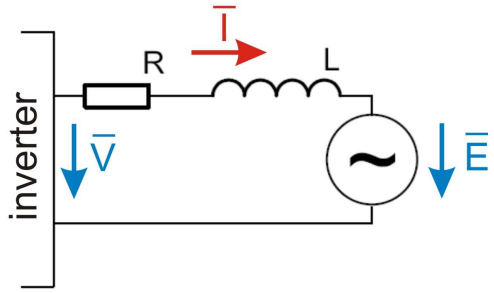


Fig. 4. The RL filter between the inverter's output and the grid used to decouple the output voltage and the grid and to filter higher harmonics.

If the consider the sampling period T to be sufficient small and the vectors $\bar{V}$ and $\bar{E}$ are constant between two sampling periods, current from (4) can be discretized as follows [3]:

$$\bar{I}_{(k+1)T} = e^{-\frac{R}{L}T}\bar{I}_{kT} + \frac{1}{R}\left(1 - e^{-\frac{R}{L}T}\right)\left(\bar{V}_{kT} - \bar{E}_{kT}\right) \tag{5}$$

The current value *I(k+1)T* is predicted by the Lagrange quadratic extrapolation.

E. Simulation results of the CCVSI

The same current control technique based on (5) was used for 3-level (one DC source: 400V) and 15-level (three DC sources: 240V, 120V and 60V) H-bridge inverter. The THDi of the injected grid current and estimated efficiency of the inverter disregarding the output filter were investigated.

It is more accurate to consider sampling frequency of the current controller rather than the switching frequency of the inverter. The controller chooses the best voltage vector at the inverter's output and thus each H-bridge switches with different frequency, as can be seen in Fig.5.

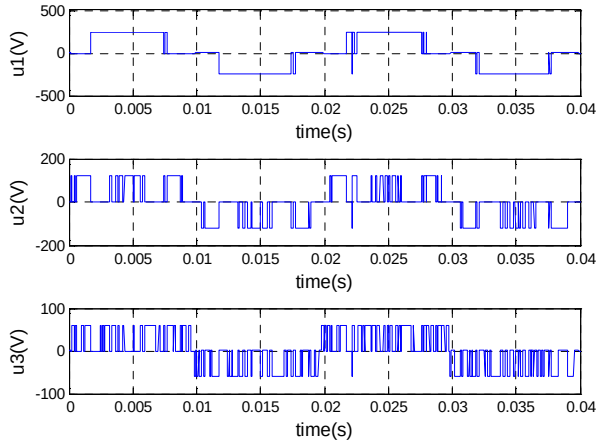


Fig. 5. Partial output voltages of the cascaded H-bridge inverter (amplitudes: 240, 120 and 60V), sampling frequency of the current controller f=10kHz.

The actual value of the output current of the inverter with respect to the desired value and its change is in Fig. 6. From Fig. 6 it can be seen that actual current tracks the desired current value and its changes with low THDi. There are no zero crossing spikes in the output current.

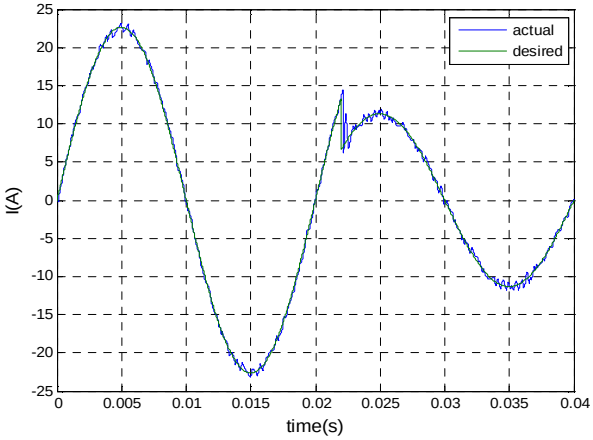


Fig. 6. Output current of the cascade H bridge inverter (desired and actual) created by combining partial output voltages shown in Fig. 5. The RL filter values: L=5mH, R=1Ω.

The same simulation of the output current that was done for the single H-bridge inverter is shown in Fig. 7. The output current THDi is significantly higher (THDi = 14%). Higher THDi also means higher uncontrolled reactive power. The response to the change in the desired value is similar as for the cascade H-bridge converter. This is caused by the same maximal voltage swing of the output voltage (±420V for cascade H-bridge, ±400V for single H-bridge).

THDi and efficiency of the inverters

The grid connected PV system is an electrical energy generator. There are two main points of view when considering such a system. It is important to ensure high energy yield and to meet standards for generator system (frequency, THD, EMI).The lower THD is achieved mainly by improved output filter. However lower THD requires more bulky and costly filter which has higher power losses. The cascade H-bridge inverter can achieve much lower THDi compared to the single H-bridge inverter and has lower EMI radiation due to the lower du/dt stresses. It can utilize lower switching frequency and decrease switching losses. It is less sensitive to the grid distortion. On overall, the cascade H-bridge inverter can produce higher quality electrical energy.

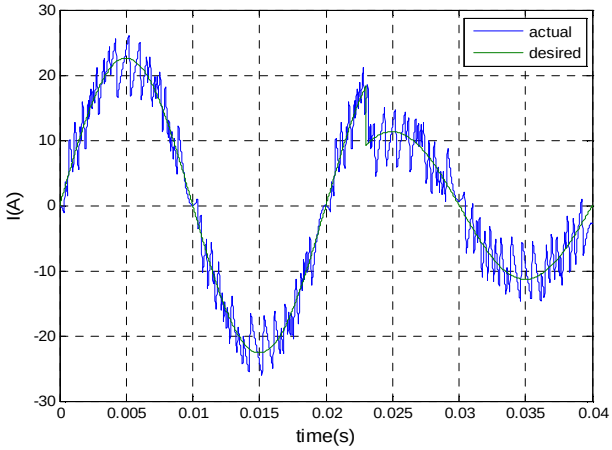


Fig. 7. Output current of the single H bridge inverter (desired and actual) supplied with DC voltage source 400V. The RL filter values: L=5mH, R=1Ω.

On the other hand there is a presumption for a lower efficiency compare to the single H-bridge inverter due to the increased number of power semiconductor switches.

The THDi and the efficiency of the single and the cascade H-bridge inverter were examined. The basic characteristics of both inverters are in Table I. Both inverters incorporate the same current control technique (5) and have the same output filter.

TABLE I
CHARACTERISTICS OF COMPARED INVERTERS

	single	cascade	
I_n	16	16	A
S_n	3680	3680	VA
U_{DC}	400	240, 120,60	V
IGBT	IRGB4056	IRGB4056	-
R_{filter}	1	1	Ω
L_{filter}	5	5	mH

The THDi of the grid current was simulated for various sampling frequencies of the current controller. The simplest RL filter was used at the inverters output. The results are shown in Fig. 8.

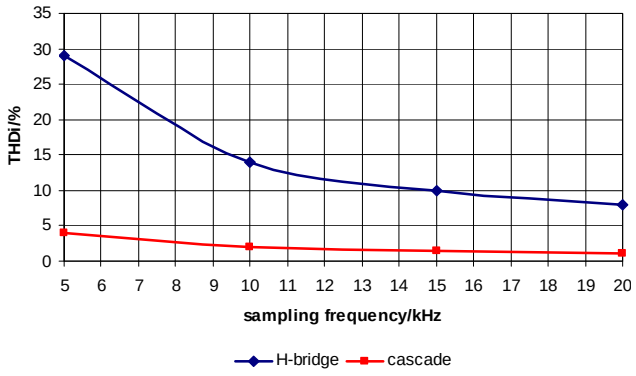


Fig. 8. The THDi of compared inverters versus current controller sampling frequency. The RL filter values: L=5mH, R=1Ω.

The desired value of THDi = 3% was never met with the single H-bridge inverter but THDi was lower than 3% for the cascade H-bridge inverter and frequencies above 10kHz.

The efficiency was examined for changing output power of the inverter. The sampling frequency of the current controller was set to 10kHz.

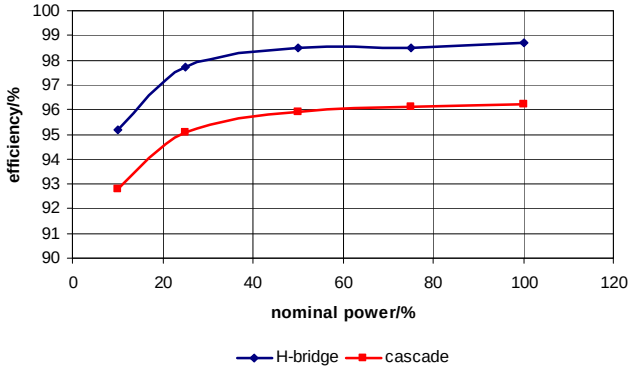


Fig. 9. The efficiency of compared inverters versus output power. The RL filter values: L=5mH, R=1Ω.

The efficiency of the inverter is not easy to estimate due to the changing current of the semiconductor switches. The calculation is based on the average value of the current.

TABLE II
COMPARISON OF SIMULATED LOSSES IN INVERTERS

Losses	single H-bridge	cascade H-bridge	
switching	5	5	W
conduction	36	103	W

Only the losses in the IGBTs are considerd disregarding diode losses (they are considered in total inverter's efficiency).

The switching losses are the same for the single H-bridge and the cascade H-bridge converter. Because there are three H-bridges in the cascade H-bridge inverter and only one H-bridge in the single H-bridge inverter, the switching losses per H-bridge are reduced in the cascade H-bridge inverter. On the other hand, the conduction losses are approximately three times higher. The $V_{CE(ON)}$ voltage of the IBGT is very important parameter when considering the efficiency of the cascade H-bridge inverter. The cascade H-bridge inverter gives the opportunity to choose the most suitable semiconductor switches for the each H-bridge an it gives a further possibility to increase the cascade H-bridge inverter's efficiency.

III. CONCLUSION

The cascade H-bridge inverter is an alternative to the single H-bridge inverter in photovoltaic systems. However cascade inverters are not popular in photovoltaic systems in nowadays. They are more expensive, have lower efficiency, require more complex control techniques. On the other hand they produce lower THD of the grid current and THD of the output voltage (Fig. 10), require smaller filters, can transfer more power and have smaller du/dt stresses.

There is a tradeoff between the number of output voltage levels and the switching frequency for the same number of DC sources at the input of the cascade H-bridge inverter. Less voltage levels mean lower switching frequency but it was shown that the high switching frequency can be transferred to the H-bridge with the lowest DC voltage (60V DC source in this case). Also the bridge with the highest DC voltage can operate at the fundamental frequency with a proper control technique (240V DC source in this case).

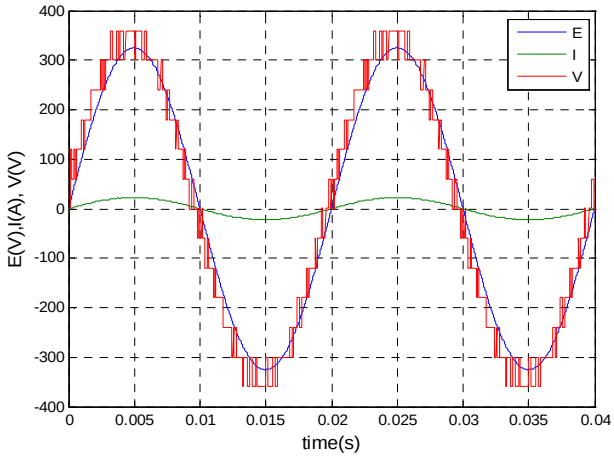


Fig. 10. The operation of the cascade H-bridge inverter with current control. The RL filter values: L=5mH, R=1Ω. The THDu of the output voltage V is 10%, THDi is 2%.

There is a need to increase the lifetime of photovoltaic inverters as well as their reliability. High voltage stresses decrease the lifetime of many electrical components [4]. Lower du/dt stresses of components in multilevel H-bridge inverter can help to meet these needs.

ACKNOWLEDGMENT

This work was supported by Slovak Research and Development Agency under project APVV-0095-07 and by Scientific Grant Agency of the Ministry of Education of Slovak Republic under the contract VEGA No. 1/0099/09.

REFERENCES

[1] S. Khomfoi, L. M. Tolbert, "Multilevel Power Converters" from M. H.Rashid, "Power Electronics Handbook", Second edition, Elsevier, 2007,ISBN10: 0-12-088479-8, pp. 451–482.

[2] S. H. Ko, S. R. Lee, H. Dehbonei, Ch. Nayar, "A Comparative Study of the Voltage Controlled and Current Controlled Voltage Source Inverter for the Distributed Generation System" [Online]. Available: http://www.itee.uq.edu.au/~aupec/aupec05/AUPEC2005/Volume1/S073.pdf

[3] G. S. Perantzakis, F. H. Xepaps,S. A. Papathanassiou, S. N. Manias, "A Predictive Current Control Technique for Three-Level NPC Voltage Source Inverter", [Online].

[4] "Current Demand of High Performance Inverters for Renewable Energy Systems" [Online]. Available: http://www.gpec.dee.ufc.br/publicacoes/Sdaher_Antunes.pdf

[5] C. V. Nayar, S. M. Islam, H. Dehbonei, K. Tan, H. Sharma, "Power Electronics for Renewable Energy Sources" from M. H.Rashid, "Power Electronics Handbook", Second edition, Elsevier, 2007,ISBN10: 0-12-088479-8, pp. 673–716.

[6] J. R. Espinoza, "Inverters" from M. H.Rashid, "Power Electronics Handbook", Second edition, Elsevier, 2007,ISBN10: 0-12-088479-8, pp. 353–404.

[7] L. M. Tolbert, F. Z. Peng, "Multilevel Converters as a Utility Interface for Renewable Energy Sources" [Online].

Novel Zero–Voltage and Zero–Current Switching Full-Bridge PWM Converter Using Simple Secondary Active Clamp Circuit

[1]*Ján PERDUĽAK*, [2]*Marcel BODOR*

[1]Dept. of Electrical Engineering, Mechatronics and Industrial Engineering, FEI TU of Košice, Slovak Republic

[1]jan.perdulak@student.tuke.sk, [2]marcel.bodor@tuke.sk

***Abstract* —A novel zero-voltage and zero-current switching (ZVZCS) full-bridge phase shifted pulse-width modulation converter is presented in this paper. A simple auxiliary secondary circuit is used on the secondary side consist of capacitor, inductance, two rectifier diodes and unipolar MOSFET transistor, provides conditions for ZVZCS – soft switching of IGBT transistor on the primary side of the DC/DC converter. The turning off the MOSFET transistor located on the secondary side provides reset both secondary and primary current and thus the conditions for ZVZCS is achieved. This paper presents detail theoretical analysis and experimental results. The appropriateness of using the proposed new topology diagram for high power application is confirmed.**

***Keywords*— phase-shifted PWM, power converter, insulated gate bipolar transistor, soft switching,.**

I. INTRODUCTION

Recently, the increasing demands for high performance load converters in power electronics are present. It also places great emphasis on increasing the switching frequency for reducing size and weight of the converters. However with increasing frequency the increasing switching losses occur mainly on the switching devices such as transistors. The power MOSFETs have many advantages such as very short switching times. It is one of the reasons why these switching devices are mainly used in ZCZVS FB PWM converters. However the MOSFETs are not suitable for high power applications. These days, IGBTs are replacing MOSFETs for high voltage, high power applications, since IGBTs have higher voltage rating, higher power density, and lower cost compared to MOSFETs [1]. On the other hand, the use of IGBTs is significantly reduced by their frequency switching, usually limited to 20 – 30 kHz because of their tail current characteristic [2]. To operate IGBTs at higher switching frequencies is required to significantly reduce turn – off switching losses. Many topologies have been developed to solve this problem [1] – [5]. This topology used auxiliary circuit on the secondary side to achieve ZCZVS and therefore to reduce the switching loss to zero. Generally the ZVS of the leading-leg switched is achieved by the similar manner as that of the ZVS FB PWM converters [3] – [4], [6] whereas ZCS of lagging-leg switches is achieved by resetting the primary current during the freewheeling period. The technical realization of auxiliary circuit which provides reset of primary current is realized in different ways. The converter proposed in [3] has simple auxiliary circuit which contains neither loss components nor active switches. Resetting of the primary current is achieved by using energy of leakage inductance and clamp capacitor placed on the secondary side. The converter [2] same as converter [3] contains neither loss components nor active switches. Resetting of the primary current is achieved using transformer auxiliary winding inserted into the secondary side what makes this auxiliary circuit more complex. The converter [7] contains active switch on the secondary side. This switch is used to control the clamping circuit. The clamp switch induces switching loss due to its hard switching, and the maximum output current is limited by the capacitance of holding capacitor [3]. The blocking capacitor on the primary side of the transformer winding is used in the converter [5]. The auxiliary circuit contains active switch and transformer auxiliary winding which make this circuit considerably complex and parameter design is complicated [2].

II. OPERATION PRINCIPLE

The detail description of proposed converter is in [6]. This operation principle description below is abbreviated form of description in [6].

The proposed converter (Fig.1) has nine operating modes within each operating half cycle. The equivalent circuits and the corresponding operation waveforms are show in Fig.2 and Fig.3, respectively.

Mode1- interval (t_0–t_1): The transistors T_1, T_2 are turned on with ZVS at t_0 because only magnetizing current flows through diodes D_1, D_2. The collector current of the transistor T_S, which is turned on at t_0 too, starts to flow and the capacitor C_C is discharged.

The rise of the collector current is in resonant way with the resonant frequency ω_{R1} different at no-load and short circuit in a range:

$$\sqrt{(L_0 + L_{CS}).\frac{C_0.C_C}{C_0 + C_C}} \le \omega_{R1} \le \sqrt{(L_0 + L_{CS}).C_C} \quad (1)$$

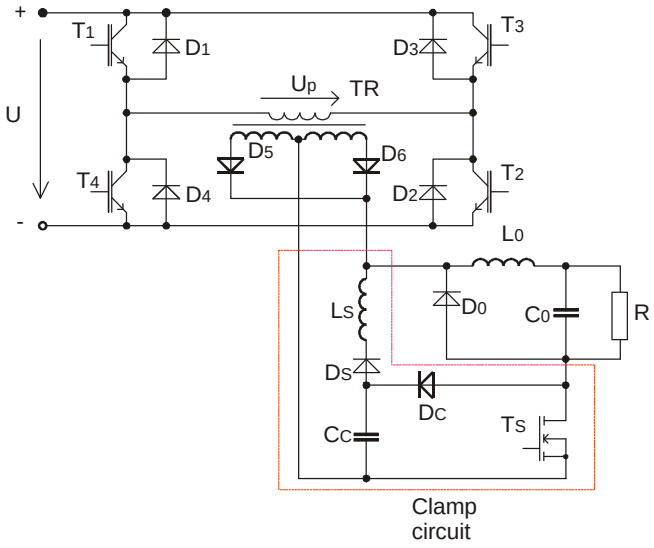


Fig.1. Circuit topology of the proposed converted

Mode2-interval (t_1–t_2): The transformer leakage inductance L_{LP} reflected to the primary side causes that primary current i_P is linearly increased with the slope U/L_{LP} while the secondary voltage u_S is zero as a result of commutation between output freewheeling diode D_0 and rectifier diode D_5.

Mode4-interval (t_3–t_4): Transistors T_1 and T_2 are conducting and the energy is delivered from the source to the load. The smoothing inductance current is a sum of the secondary current and inductance L_S current:

$$i_0 = i_S + i_{LS} \tag{2}$$

Mode5-interval (t_4–t_5): The primary current increases with the slope:

$$\frac{di_p}{dt} = \frac{U - n.U_0}{L_{LP} + n^2.L_0} + \frac{U}{L_m} \tag{3}$$

Where $n = \frac{N_P}{N_S}$ is power transformer turns ratio and L_m magnetizing inductance of the power transformer.

Mode6: interval (t_5–t_6): At t_5 the secondary transistor T_S turns off. At that time the commutation between transistor T_S and clamp diode D_C occurs and charging of the clamp capacitor C_C starts. Afterwards the commutation between D_C, D_5 and output freewheeling diode D_0 starts. In the mentioned commutation path the resonance occurs and rise of the current depends on the resonant frequency ω_{R2}:

$$\omega_{R2} = \sqrt{(L_0 + L_{LS}).\frac{C_0.C_C}{C_0 + C_C}}, \quad \text{for } R_0 = \infty. \tag{4}$$

$$\omega_{R2} = \sqrt{(L_0 + L_{LS}).C_C}, \quad \text{for } R_0 = 0. \tag{5}$$

During the commutation the energy stored in the leakage inductance is transferred to the clamp capacitor C_C and consequently an over-voltage ΔU_S appears on secondary voltage.

Mode7- interval (t_6–t_7): Only small magnetizing current i_m flows through primary winding of transformer. The output current flows trough output freewheeling diode D_0.

Mode8 - interval (t_7–t_8): In this interval the transistors T_1 and T_2 are turned off with ZCS. Only small magnetizing current i_m is switched off by transistors T_1 and T_2. The magnetizing current charges or discharges the internal output capacitances C_{OSS1} – C_{OSS4} of the IBGT transistors T_1 – T_4 respectively.

Mode9 - interval (t_8–t_9): At t_8 the freewheeling diodes D_3, D_4 starts to lead primary current and thus conditions for the ZVS for the transistors T_3 and T_4 are set up.

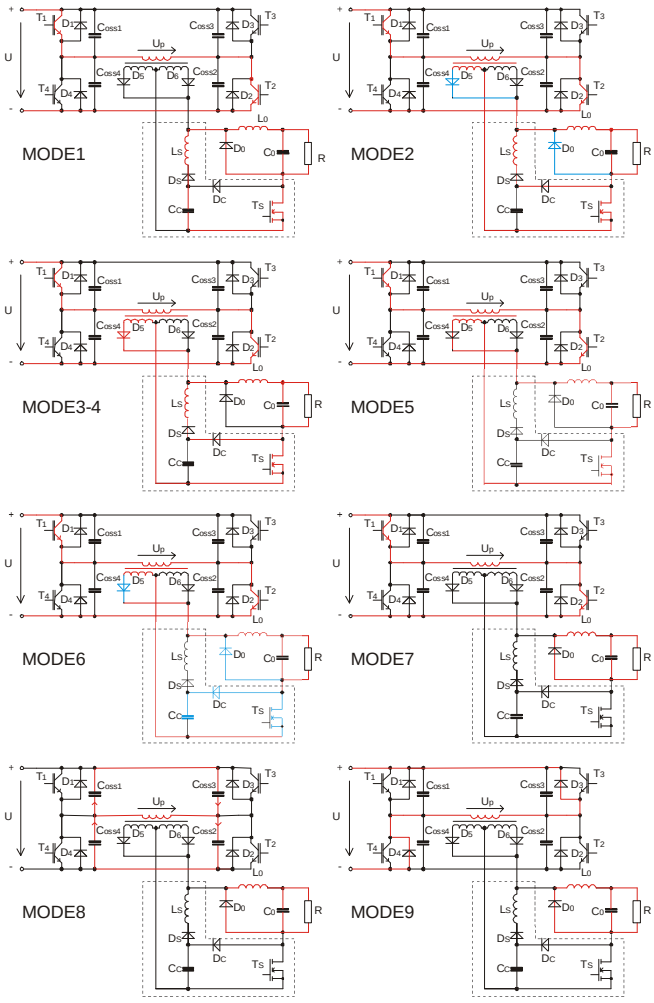


Fig.2. Equivalent circuit for each operation mode (**Note**–the blue color shows components among which commutations occur).

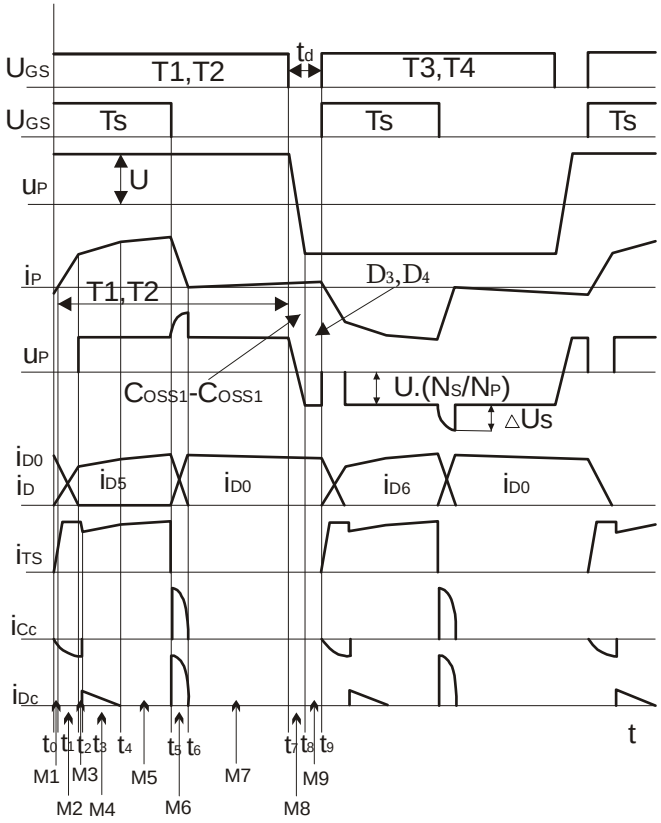


Fig.3. Operation waveforms of the proposed converter

III. Simulation Results

A simulation model in programme Orcad was created to verify the properties of the proposed converter. The simulations were made at input voltage U = 325V.

Parameters:

Transformer TR parameters:

Turns ratio n = 4,

Magnetizing inductance Lm = 1 mH,

Leakage inductance L_{LP} = 5 µH.

Clamp circuit parameters:

Clamp capacitor C_C = 400 nF,

Clamp inductance L_S = 5 µH.

Fig.4. shows the waveforms during turn-on and turn-off of the primary switch T_4. The influences of secondary active clamp circuit insure that all switching devices are switched softly. As we can see the leading-leg for transistor T_4 is switched softly and the switching loss is neglectable.

Fig.5 shows the secondary voltage u_{DS} and collector current i_D of transistor Ts (upper waveforms) in compare with switch waveforms of voltage u_{CE} and collector current i_C of transistor T_4 (bottom waveforms). It can be seen that influence of leakage inductance L_{LS} of transformer and clamp inductance L_S insure that turn-on of transistor Ts is under zero-current and just as in the previous case (Fig.4) the power losses can be neglectable, too.

Fig.6 and Fig.7, respectively show simulation waveforms of voltage u_{DS} and collector current i_D of transistor Ts (upper waveforms) in compare with primary voltage u_P and current i_P (bottom waveforms of Fig.6). After turned-off of secondary transistor Ts only small magnetization current i_m flows through the primary winding. This current charges and discharges the internal capacity of transistors T_1 – T_4, respectively and so the condition of ZVS is achieved. Fig.7. (bottom waveforms) shows the simulation waveforms of currents flows through clamp capacitor Cc, diode Ds and clamp diode Dc during cycle of operation Ts.

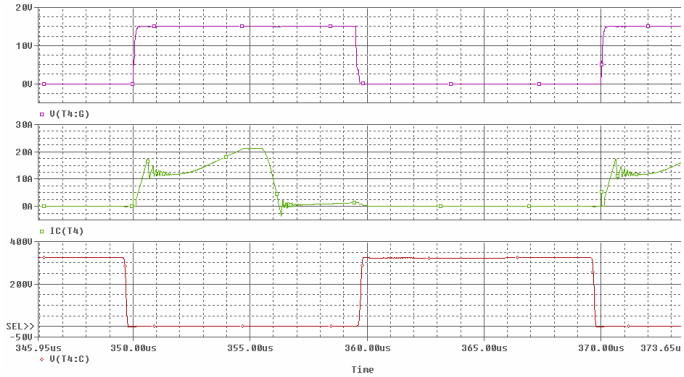


Fig.4. Turn-on and turn-off waveforms of the leading-leg and legging-leg switch T_4.

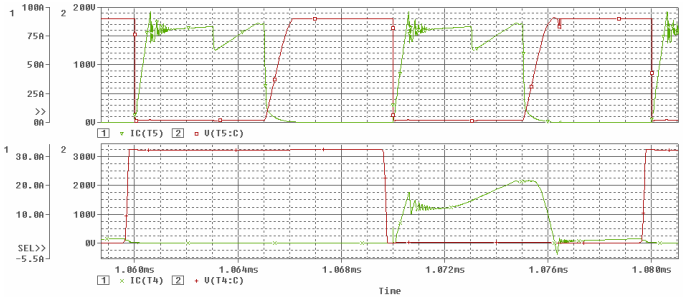


Fig.5. Waveforms of voltage u_{DS} and collector current i_D of transistor Ts (upper waveforms) and switch voltage u_{CE} and collector current i_C of transistor T_4 (bottom waveforms).

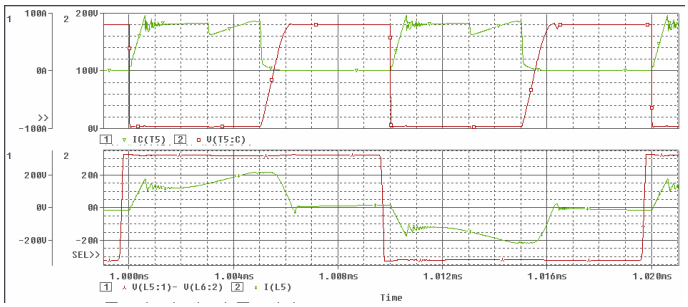


Fig.6. Waveforms of voltage u_{DS} and collector current i_D of transistor Ts (upper waveforms) and primary voltage u_P and current i_P (bottom waveforms).

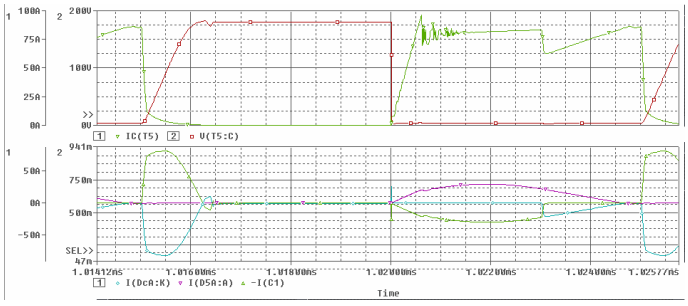


Fig.7. Waveforms of voltage u_{DS} and collector current i_D of transistor Ts (upper waveforms) and currents waveforms of Dc, Cc and Ds (bottom waveforms).

IV. CONCLUSION

A novel zero-voltage and zero-current switching (ZVZCS) full bridge phase-shifted PWM converter is presented in this paper. The properties of the proposed converter topology were analyzed. Theoretical analysis was verified by simulation. It is shown that using of secondary active clamp circuit the soft switching conditions for all switching devices T_1 - T_4 located on the primary side is achieved.

ACKNOWLEDGMENT

This work was supported by Slovak Research and Development Agency under project APVV-0095-07 and by Scientific Grant Agency of the Ministry of Education of Slovak Republic under the contract VEGA No.1/0099/09.

REFERENCES

[1] H. -S. Choi, J. –W. Kim, B. H. Cho, "Novel Zero-Voltage and Zero-Current-Switching (ZVZCS) Full-Bridge PWM Converter Using Coupled Output Inductor," *IEEE Trans. Power Electron*., vol. 17, No. 5, SEPTEMBER 2002, pp.641-648.

[2] J. G. Cho, J. W. Beak, D. W. Yoo, H. S. Lee and G. H. Rim, "Novel Zero-Voltage and Zero-Current-Switching (ZVZCS) Full-Bridge PWM Converter Using Transformer Auxiliary Winding," in *Conf. Rec. IEEE PESC`97*,pp.227-232.

[3] T. –F. Chen, S. Cheng, "A Novel Zero-Voltage Zero-Current-Switching Full-Bridge PWM Converter Using Improved Secondary Active Clamp," *IEEE ISIE, July 9-12, 2006*, Montreal, Québec, Canada, pp. 1683-1687

[4] J. –G. Cho, J. –W. Baek, Ch. –Y. Jeong, and G. –H. Rim, "Novel Zero-Voltage and Zero-Current-Switching Full-Bridge PWM Converter Using a Simple Auxiliary Circuit," *IEEE Trans. Industry Applications.*, vol. 35, no. 1, January/February 1999, pp.15-20.

[5] A. Jangwanitlert, K.J. Olejniczak, J.C. Bala, "An Improved Zero-Voltage and Zero-Current-Switching PWM Full-Bridge DC-DC Converter," *IEEE 2003*.227-232.

[6] J. Dudrík, V. Ruščin, "Voltage Fed Zero-Voltage and Zero-Current-Switching PWM DC-DC Converter," In: EPE-PEMC 2008: 13th International Power Electronics and Motion Control Conference : 1 - 3 September 2008, Poznan - Poland. S.l.: IEEE, 2008. p. 295-300. ISBN 978-1-4244-1742-1

[7] J. G. Cho, C. Y. Jeong, and F. C. Y. Lee. "Zero-Voltage Zero-Current-Switching Full-Bridge PWM Converter Using Secondary Active Clamp," IEEE Trans on Power Electronics, 1998, 13(4): 601-607

[8] J. Dudrík, V. Ruščin, "ZVZCS PWM DC-DC Converter with Controlled Output Rectifier," Acta Electrotechnica et Informatica *januar-march 2010. vol. 10, pp.12-17*

[9] Dudrik, J., Oetter, J.: Soft-Switching PWM DC-DC Converter for High Power Applications, EPE-PEMC 2006, Portorož, Slovenia, ISBN 1-4244-0121-6, CD, pp. 739-744

Multicriterion Decision and Products Competitiveness

[1]*Peter POÓR*, [2]*Juraj TIŽA*, [3]*Monika FEDORČÁKOVÁ*

[1]Fac. of Management and Economics, SjF TU of Košice, Slovak Republic
[2] Department of Metal Forming, Faculty of Metallurgy, HF TU of Košice, Slovak Republic
[2]Fac. of Management and Economics, SjF TU of Košice, Slovak Republic

[1]poorpeter@gmail.com, [2]juraj.tiza@tuke.sk, [3]monikafedorcakova@azet.sk

***Abstract*—Competitiveness can be defined as ability of company to offer equal or better conditions to customer. Not all the factors of competitiveness have an objective character and often not even measurable, but subjective perception by confronting customers with their requirements, values, or just moods. In a broader sense competitiveness is a superposition of various factors as quality, price, design, mark, utility and usefulness of product. This contribution practically described the method of multicriterion decision, where on sample of 31 cars with input parameters (price, equipment, consumption, performance, maximal speed, acceleration, design, emissions, volume of luggage compartment, front and side impact, pedestrian protection), is output defined. Since using the method chosen, the priority of parameters is defined by order in the entry table, differences in outputs are remitted when randomly organized table of inputs and outputs when classification by setting their priorities.**

***Keywords*— Multicriterion decision, product, competitiveness, car.**

I. PRODUCT COMPETITIVENESS VALUATION

The difficulty of commercial product success valuation depends on several factors, respectively. degree of novelty, as well as terms of their development, production and sales execution. There are different approaches and valuation methods. The most commonly used are various modifications of point valuation. They differ in number and usually by means of classification of evaluation criteria, as well by system of points. Evaluation of competitive ability of products can be subjected to economic analysis of selected indicators of competing subjects or consumer market research. When doing the economic analysis comparing the use of selected indicators we use synthetic indicators using:

Comparative - analytical methods – typical for the use of verbal indicators (quality, service, personnel structure ,...), level is expressed mainly by words (poor - average - excellent)

- SWOT analysis
- GE matrix
- Critical success factors;

Mathematics - Statistical methods - numerical indicators, mainly extensive (volume) or intensitive (proportional) are used here:

- mean values and the rate variability
- methods of statistical induction,
- depending on the method of analysis of qualitative characters,
- multi-criteria decision methods [1]

Evaluation of product suppliers can be done using the "O Meara" method, which administers these groups of factors:

- market fitness of product,
- market life and sales specification,
- manufacturability,
- incremental market potency [2].

II. MULTICRITERION DECISION METHOD

This method is based on the classification of product parameters qualifying using a variety of large definition indicators in the recommended composition according to the evaluation purpose. Best number of set and watched qualifying indicators is between 20-25. Principle of method is shown on Fig. 1.

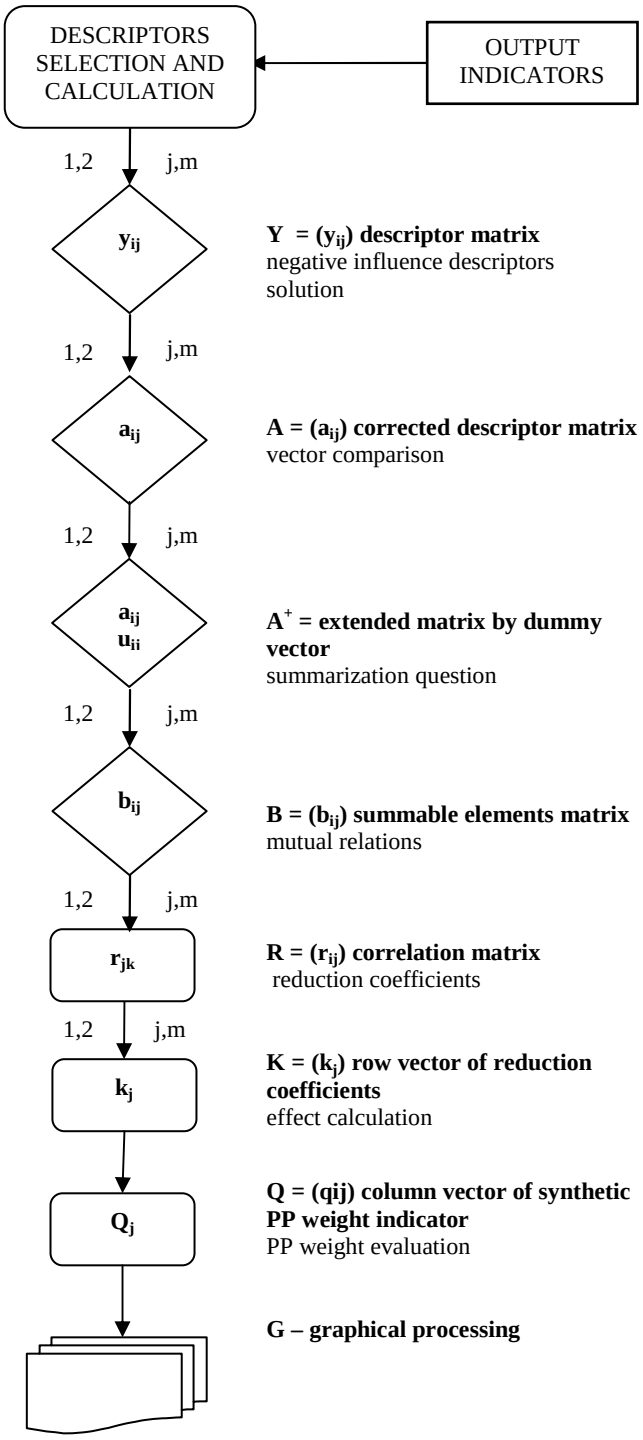


Figure 1 Multicriterion decision method [3]

Baseline data are represented as

$$Y = (Y_{ij}) \tag{1}$$

where the Y matrix elements form available qualified information about state (quality) of the product viewed as statistical features, other way descriptors. Each element Y_{ij} contains the quantitative variables, i.e. values that object $i = 1,2....m$ attained in descriptor $j = 1,2....n$. Those columns, which affect on the product is clearly positive, are reserved by multiplying with -1. This leads to the following matrix $A = (a_{ij})$.

Comparability of vectors is resolved in a manner that to the matrix A one more line, which represents a hypothetical object in the form of an artificial vector $U = (u_j)$, is added. Thus created matrix is known as A^+, to so it is true:

$$u_j \le a_{ij} \tag{2}$$

With this vector it is possible to compare vectors of real objects, since it is guaranteed to be smaller than any real vector object, which is

$$a_{ij} - u_j \prec 0 \tag{3}$$

Possibility of summation, by what we mean eliminating disparities and differences of rows are dealt on the principle of discrimination so that positive difference $(a_{ij} - u_j)$ is divided by standard deviation s_j, whose size is so, what are the values, from which it was calculated, and modifies row of numbers , where

$$s = \sqrt{\frac{\sum (a_{ij} - \bar{a}_i)^2}{n}} \tag{4}$$

Elements of *B* matrix are obtained by transforming the elements of *A* matrix by:

$$b_{ij} = \frac{a_{ij} - u_j}{s_j} \tag{5}$$

$B = (b_{ij})$ matrix elements are measurable dimensionless numbers, which can be summarized. Standard deviation also performs the role of importance– descriptor weight.

Mutual relations between descriptors are expressed by the correlation matrix $R = (r_{ij})$. This provides the basis for calculating reduction constants whose value is calculated when $(l \in j)$ as:

$$r_{ij} = \frac{\sum (a_{ij} - \bar{a}_j)(a_{il} - \bar{a}_l)}{\sqrt{\sum_i (a_{ij} - \bar{a}_j)^2 \sum_i (a_{ij} - \bar{a}_l)}} \tag{6}$$

which will then be

$$\begin{aligned} k_1 &= 1 \\ k_2 &= \left(1 - |r_{12}|\right) \\ k_3 &= \left(1 - |r_{13}|\right)\left(1 - |r_{23}|\right) \\ k_4 &= \left(1 - |r_{14}|\right)\left(1 - |r_{24}|\right)\left(1 - |r_{34}|\right) \\ k_n &= \left(1 - |r_{1n}|\right)\left(1 - |r_{2n}|\right)...\left(1 - |r_{n-1n}|\right) \end{aligned} \tag{7}$$

Based on those rapports we can come to term Q_j:

$$Q_j = \sum_j \frac{a_{ij} - u_j}{s_j} k_j \tag{8}$$

or

$$Q_i = \sum_j b_{ij}.k_j. \tag{9}$$

This relationship, writing down more in details, for example for the 1st object will be:

$$Q_l = \sum_j \frac{a_{1l} - u_2}{s_1} k_1 + \sum_j \frac{a_{2l} - u_2}{s_2} k_2 + ... + \sum_j \frac{a_{1n} - u_n}{s_n} k_n$$

(10)

Where:

Q_l - Value ....

a_{ij} - Value of j-th component of an artificial vector,

s_j - Standard deviation of the modified j-th descriptors

k_j - Reduction constants [3]

III. MULTICRITERION DECISION METHOD PRACTICAL EXAMPLE

In this case was solved the problem of determining the order of cars following these parameters: cost, equipment, consumption, maximal speed, design, the volume of luggage space, power, front and side impact, pedestrian protection, design, emissions.

As a set of reference objects, a sample of 31 cars with an equal engine volume 1.4 cm^3. The multi-criterion method was applied on 3 types of input data:

- table with input values arranged according to priorities,
- table layout without design input values,
- table with a random arrangement of input values.

Table 1 Input data

Car type	A	B	C	D	E	F
Seat Altea	18692	21	22	63	7,3	169
Seat Ibiza	11884	12	14	63	6,5	180
Seat Leon	18221	20	14	63	7	172
Kia Rio	11144	11	24	71	6,2	173
Kia Cee´d	14264	12	13	77	6,1	185
Renault Thalia	11582	15	11	72	7	186
Renault Mégane	15266	11	11	72	7,7	183
Renault Scenic	20876	12	11	72	7,2	174
Opel Corsa	14105	11	11	66	5,9	173
Opel Meriva	15432	11	19	66	6,4	168
Opel Astra	17258	12	13	66	6,1	180
Fiat Bravo	15366	17	13	66	6,7	179
Fiat 500	13261	15	16	74	6,3	160
Fiat Grande Punto	14105	11	14	70	6,1	165
Nissan Micra	13275	13	19	65	6,3	172
Honda Jazz	15930	15	12	61	5,6	170
Honda City	14241	12	19	61	5,6	175
Honda Civic	19884	18	19	61	5,9	170
Toyota Auris	16262	15	18	71	6,9	170
Citroen C2	13889	9	21	54	6	169
Citroen C3	13829	9	12	54	6,1	167
Peugeot 207	12747	16	11	54	6,3	170
Peugeot 206	11950	9	11	55	6,4	170
Peugeot 307	17195	11	11	65	6,5	172
Škoda Fabia	11947	13	10	63	6,5	174
Škoda Octavia Tour	15601	11	17	59	7,1	173
Škoda Roomster	13444	11	17	63	6,8	171
Ford Fiesta	12448	13	14	59	6,2	166
Ford Fusion	13577	13	14	58	6,5	163
VW Polo Comfortline	11840	14	11	59	6,4	175
Chevrolet Lacetti	13607	10	19	70	7,2	175

A - price (€)
B - equipment (points) – more points = better equipment
C – pedestrian protection (points) – more points = better pedestrian protection
D - performance (KW)
E - consumption (l/100km)
F - CO2 emissions (g/km)

Those parameters, which impact on the product is evidently negative, are reserved by multiplying (-1), as we can see in Tab. 2 In this case those are price, consumption and emissions.

Table 2 Entry data matrix for multicriterion decision with arrangement

Car type	A	B	C	D	E	F
1	-18692	21	-7,3	63	169	22
2	-11884	12	-6,5	63	180	14
3	-18221	20	-7	63	172	14
4	-11144	11	-6,2	71	173	24
5	-14264	12	-6,1	77	185	13
…						
…						
26	-15601	11	-7,1	59	173	17
27	-13444	11	-6,8	63	171	17
28	-12448	13	-6,2	59	166	14
29	-13577	13	-6,5	58	163	14
30	-11840	14	-6,4	59	175	11
31	-13607	10	-7,2	70	175	19

A - price (€)
B - equipment (points) – more points = better equipment
C – pedestrian protection (points) – more points = better pedestrian protection
D - performance (KW)
E - consumption (l/100km)
F - CO2 emissions (g/km)

After getting these values we calculated the correlation coefficient rij and artificial vector, mean value and standard deviation. The results of t-test are in table 3.

Table 3 t-test

T[x,y]	1	2	3	4	5	6
1	**-****	**+****	**-****	**-**	**-***	**+****
2	**+****	**-**	**-**	**-**	**+****	**-**

3	**-****	**+****	**-****	**-**	**-**	**-**
4	**+****	**-***	**+**	**+****	**+**	**+****
5	**+**	**-**	**+***	**+****	**+****	**-**
…						
…						
26	**-**	**-***	**-****	**-****	**+**	**+**
27	**+**	**-***	**-***	**-**	**-**	**+**
28	**+****	**+**	**+**	**-****	**-****	**-**
29	**+**	**+**	**-**	**-****	**-****	**-**
30	**+****	**+**	**+**	**-****	**+**	**-****
31	**+**	**-****	**-****	**+****	**+**	**+****

The correlation coefficient is a measure of linear dependence. It measures the strength of statistical dependence between two numerical variables. The correlation coefficient can acquire values from the interval <-1.1>. If the correlation coefficient equal to 0, it means that between the examined variables there is no relationship. If it is equal to one, the values themselves are directly dependent, the increase of one variable causes rise of another, when is equal to -1, an increase of one variable causes the other variable's decrease. Matrix of correlation coefficients is shown in table 4.

Table 4 Matrix of correlation coefficients

r[x, y]	x: 1	2	3	4	5	6	7	8	9	10
y: 1										
2	-0,406									
3	0,221	-0,176								
4	-0,137	0,046	-0,307							
5	0,038	-0,037	-0,271	0,384						
6	0,653	-0,321	0,055	0,370	0,243					
7	-0,116	0,258	0,340	0,005	-0,212	0,068				
8	0,203	-0,155	0,923	-0,267	-0,184	0,059	0,300			
9	-0,456	0,280	-0,249	0,081	0,363	-0,524	-0,200	-0,221		
10	-0,416	0,253	-0,238	0,453	0,101	-0,090	0,120	-0,229	0,168	
11	-0,081	0,080	0,035	0,046	-0,264	-0,148	0,289	0,046	0,047	-0,126

The reduction constants eliminate the multiplicity encounter of such an influence, which would occur when measuring objects quality in different changes when using different descriptors, see table 5.

Table 5 Reduction constants table

k1	1
k2	0,593667908
k3	0,64182828
k4	0,571034783
k5	0,416272997
k6	0,106182748
k7	0,316554112
k8	0,020471856
k9	0,051064411
k10	0,084007777
k11	0,2760510770

The ranking of dimensionless value of synthetic indicator (score) is set so that the greatest value marks the object with best properties, that are reflected in output table, see Table 6.

Table 6 Final table

Kia Cee´d	11,10330594
Kia Rio	10,95248835
Fiat 500	10,35243557
Renault Thalia	10,22159959
Honda City	9,720387903
Nissan Micra	9,708160493
Seat Ibiza	9,592689637
Fiat Bravo	9,463457922
Škoda Fabia	9,288454436
Opel Corsa	9,218325348
VW Polo	9,212285056
Toyota Auris	8,723808099
Fiat Grande Punto	8,614140072
Honda Jazz	8,590194289
Opel Astra	8,370171159
Peugeot 207	8,246988084
Chevrolet Lacetti	8,115960692
Škoda Roomster	8,08340407
Honda Civic	7,978672496
Citroen C2	7,844226863
Opel Meriva	7,782542214
Ford Fiesta	7,75506567
Seat Leon	7,611258278
Seat Altea	7,484036653
Peugeot 206	7,453412462
Ford Fusion	7,126617578
Renault Mégane	6,847664217
Peugeot 307	6,835556532
Citroen C3	6,429536476
Škoda Octavia Tour	6,082858216
Renault Scenic	4,475854964

IV. CONCLUSION

In this contribution ranking of cars on the basis of specific parameters was determined by the method of multicriterion decision. By the method was according to the tab. 6 best cars were evaluated: Kia Cee'd, Kia Rio, Fiat 500, Renault Thalia, Honda City. Multi criterion decision can be used in other situations, such as: optimization of production by comparison with the competition, where if the manufacturer knows how the customer prioritize the parameters of the product, he is able to adjust production to become more competitive. Similarly, the method can also be used to determine the current competitive market. This method can also be used to evaluate the characteristic parameters of products for the purpose of manufacturing, marketing, environmental burdens,

mutual comparison of own and competing products, finding hotspots features their products.

APPENDIX

This contribution was created within the solution of VEGA 1/0679/08 Integrates system for innovative projecting, planning, organizing and manufacturing planning.

REFERENCES

[1] Šutaj – Eštók, A.: Všeobecná ekonomická teória. EDÍCIA ŠTÚDIJNEJ LITERATÚRY, Košice 2006 – 212s. ISBN 80-3-741-X
[2] Mlákay, J.: Produkt a konkurencia. EKONÓM, Bratislava 2004 – 133 s. ISBN 80-225-1947-2
[3] Muránsky, J. – Badida, M.: Environmentálne aspekty navrhovania strojárskych objektov. Vienela, Košice 2003 – 367 s., ISBN 80-7099-741-9
[4] Horáková, I. – Kotas, P.: Marketing v současné světové praxi. Grada Publishing, Praha 1992 – 364 s. ISBN 80–85424-83-5
[5] Muránsky, J. – Badida, M.: Trvalo udržateľný rozvoj v strojárstve. Vienela, Košice 2003 – 251 s., ISBN 80-7099-519-X
[6] Fabianová, Katarína: Fiškálna decentralizácia a uplatňovanie funkcií verejných financií na obecnej a regionálnej úrovni 1 : zborník z vedeckej konferencie : 28. september 2009, Košice. - Košice : EkF TU, 2009. ISBN 978-80-553-0247-8. - S. 16-22.
[7] http://www.auto.sk

Multiple Target Tracking System for Through Wall Application

Jana ROVŇÁKOVÁ

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

jana.rovnakova@tuke.sk

***Abstract*—Through wall tracking of moving targets is of great interest for rescue, surveillance and security operations. In majority of such cases, the tracking of multiple moving targets is needed. Determination which measurements to associate with which targets being tracked and track maintenance belongs to the main tasks that the modern multiple target tracking system has to solve in addition to the target position smoothing. In this paper, a prospective implementation of such system is introduced and modified for through wall application. The experimental results obtained by the real radar signal processing confirm good performance properties of the proposed system.**

***Keywords*—Data association, MTT system, target tracking, radar signals, through wall radar.**

I. Introduction

Through wall tracking of moving targets can be very helpful in the situations where the entering of a room or a building is considered hazardous and it is desired to inspect its interior from outside through the walls. Examples include tracking of people in the dangerous environments, through rubble localization following an emergency or room monitoring for unauthorized intruders. Such tracking can be advantageously realized by ultra-wideband (UWB) radars which operate in a lower GHz-range base-band - approximately up to 3.5 GHz. Electromagnetic waves transmitted by mentioned devices show then reasonable penetration through most typical building materials including reinforced concrete, concrete block, sheet rock, brick, wood, plastic, tile or berglass [1].

The target tracking objective is to collect sensor data, i.e. UWB radar signals for the considered through wall aplication, from a field of view containing one or more potential targets of interest and then to partition these data into sets of observations, or tracks, that are produced by the same sources. Once tracks are formed and confirmed (so that background and other false targets are reduced), the number of targets can be estimated and quantities, such as target velocity, future predicted position, and target classification characteristics, can be computed for each track [2].

There is a fundamental distinction between single-target tracking (STT) systems and multiple-target tracking (MTT) systems. Because the STT systems are dedicated to a single target, there is no need to perform a complex data association function, such as that discussed later for an MTT system. However, consistency tests must be performed to ensure that the sensor is still pointing at the target. The tracking filter may be an analog device, but modern systems typically use Kalman filtering, such as that described in [3].

The extension of STT to MTT requires a complex data association logic in order to sort out the returning sensor data into the general categories of targets of interest, recurrent sources that are not of interest (such as background clutter), and false signals with little or no correlation over time. The gating, observation-to-track association and track maintenance functions are usually part of the overall data association function. First, gating is used as a screening mechanism to determine which observations are valid candidates to update existing tracks. Gating is performed primarily to reduce unnecessary computations by the association and maintenance functions that follow. The association function takes the observation-to-track pairings that satisfied gating and determines which observation-to-track assignments will actually be made. And finaly, track maintenance refers to the functions of track initiation, confirmation, and deletion. Modern MTT systems typically combine data association and multiple Kalman filter models that are running in parallel [4].

The intention of this paper is to introduce modern MTT system convenient for tracking of multiple targets moving behind walls. As a base, the MTT system used for driver assistance application has been utilized [5]. Its basic principle is outlined in Section II. The system modifications necessary for through wall application are described in Section III. The experimental results achieved by processing of real UWB radar signals obtained in through wall scenarios are given in Section IV. Finaly, a few concluding comments enclose the paper.

II. Basic Principle of Original MTT System

The functioning of the MTT system for driver assistance application introduced in [5] can be in short order explained as follows. Assuming recursive processing as shown by the loop in Figure II, tracks would have been formed on the previous radar scan. When new observations are received from the radar the processing loop is to be executed. Incoming observations are first considered by the Gate checker for updating of the existing tracks. Gating tests determine which possible observation-to-track pairings are reasonable, by attributing a cost to each pairing. The costs are calculated as the statistical distance between the predictions of the target states given by the filters and the observed state coordinates received from the radar. These costs are put together in a cost matrix which is then passed on to the assignment solver to determine the finalized pairings. The pairings are made in a way to ensure minimum total cost for all the pairings. The finalized observation-track pairings are passed on to the tracking filters

which use them for estimating the current states of targets and predicting the next states as well as the error covariance associated with these predictions.

The predicted states and predicted error covariance are used by the Gate compute function to define probability gates or windows around the predicted states. The dimensions of the gates being dictated by the prediction error covariance, these gates demarcate the probability boundaries for the next state coordinate measurements. The Gate Compute sub-function can be viewed as a first level of screening out the unlikely target-track associations in case of multiple observations falling close to a single prediction or vice versa. In the second level of screening, namely observation-to-track assignment, a strictly one-to-one coupling is established between observations and tracks.

The Track Maintenance sub-function consists of three blocks. The obs-less Gate Identifier identifies the gate where no observation falls. This indicates a probable disappearance of an already known target and hence the deletion of its track after confirmation. The New Target Identifier detects observations that fall outside all the gates. These observations are potential candidates for initiating new tracks after confirmation. The Track Init/Del block initiates new tracks or deletes existing ones when needed. The MTT is explained in detail in [4] and [6]. The tracking filters block in Figure II is particularly important. The linear Kalman filters are used for this block. The number of filters employed is equal to the maximum number of targets to be tracked.

III. System Modifications for Through Wall Application

The modifications of outlined MTT system have been done for the through wall radar device utilized in our measurements, i.e. for the M-sequence UWB radar system [7]. The modifications are detailed below according system parts in which they have been realized. The functions of Gate Computation, Gate Checker and Cost Matrix Generator have stayed without change [5].

A. Process Model

The process model mathematically projects the process current state to the future. This can be presented in a linear stochastic difference equation as

$$Y_k = AY_{k-1} + BU_k + W_{k-1} \tag{1}$$

In equation (1) Y_{k-1} and Y_k are n-dimensional state vectors that includes the quantities to be estimated. Matrix A is the assumed known state transition matrix which may be viewed as the coefficient of state transformation from instant $k-1$ to instant k, in absence of any driving signal and process noise. Matrix B is the assumed known control matrix, U_k is the optional control input and W_{k-1} represents zero-mean additive white Gaussian process noise (AWGN) with assumed known covariance Q.

The first modification relates only to adjusting of the numerical values of the matrix A, in which the radar pulse repetition time T occurs. In the original MTT system, the quantity T was equal to 0.02 seconds. For the M-sequence UWB radar, this quantity is set to 0.07 seconds. Then the matrix A has the following form

$$A = \begin{pmatrix} 1 & 0.07 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0.07 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

B. Measurement Model

To describe the relationship between the true state and the measurements (observations), a measurement model is required. It can be described as a linear expression

$$Z_k = HY_k + V_k \tag{2}$$

The observation matrix H in the measurement equation (2) relates the current state to the measurement (observation) vector Z_k and the terms V_k is a random variable representing the measurement noise. In the original MTT system, the measurement vector Z_k contains two elements - distance d and angle θ as shown below, and the state vector Y_k has the following form

$$Y_k = \begin{pmatrix} y_{11} \\ y_{21} \\ y_{31} \\ y_{41} \end{pmatrix} \qquad Z_k = \begin{pmatrix} d \\ \theta \end{pmatrix}$$

Here y_{11} is the target range, y_{21} is range rate, y_{31} is azimuth angle and y_{41} is angle rate.

Forasmuch as the M-sequence UWB radar measures only target ranges and not angles, it was not able to apply the described models to our measurements. Therefore, at first the target locations have been computed on the basis of trilateration process [8]. The obtained Cartesian coordinates have been then converted to polar coordinates and for such data the models have been utilized. This modification results in the change of observation-to-track to so-called plot-to-track association [9]. That means, the measured ranges cannot be more directly used to update the track information.

C. Kalman Filter

Given the process and the measurement models from (1) and (2), the Kalman filter equations used in the original MTT system are

$$\begin{array}{lll} (a) & \hat{Y}_{\bar{k}} & = A\hat{Y}_{k-1} + BU_k \\ (b) & P_{\bar{k}} & = AP_{k-1} + Q \\ (c) & K & = P_{\bar{k}}H^T(HP_{\bar{k}}H^T + R)^{-1} \\ (d) & \hat{Y}_k & = \hat{Y}_{\bar{k}} + K(Z_k - H\hat{Y}_{\bar{k}}) \\ (e) & P_k & = (I - KH)P_{\bar{k}} \end{array} \tag{3}$$

where $\hat{Y}_k$ is state estimation vector, $\hat{Y}_{\bar{k}}$ is state prediction vector, K is Kalman gain matrix, $P_{\bar{k}}$ is prediction error covariance matrix, P_k is estimation covariance matrix and I is an identity matrix of the same dimensions as P_k.

The covariance matrices Q and R occuring in equation set (3) have in the original MTT system the following numerical values

$$Q = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 330 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.3 \cdot 10^{-8} \end{pmatrix} \quad R = \begin{pmatrix} 10^6 & 0 \\ 0 & 2.9 \cdot 10^{-4} \end{pmatrix}$$

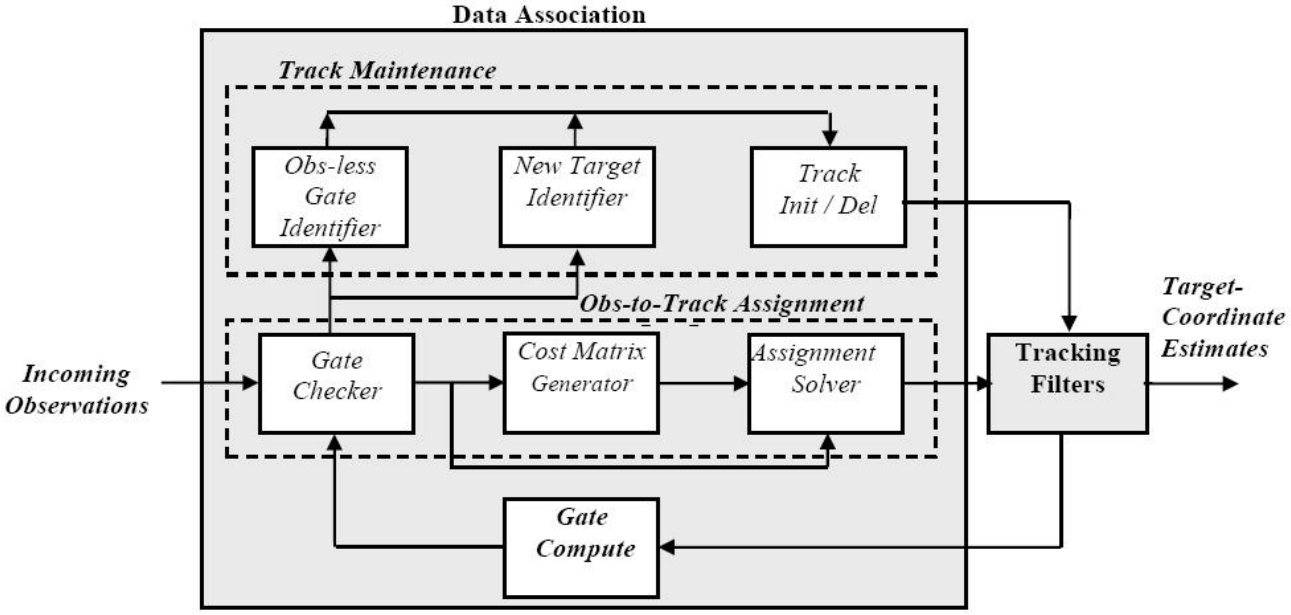


Fig. 1. Implementation scheme of the original MTT system used for driver assistance application

For the M-sequence UWB radar, new values of Q and R have been found based on processing of radar signals from through wall scenarios. The diagonal of matrix Q and the diagonal of matrix R have the following modified values

$$diag(Q) = (0 \quad 0.01 \quad 0 \quad 0.0001)$$
$$diag(R) = (0.1 \quad 0.01)$$

D. Assignment Solver

The assignment problem is stated as follows. Given a cost matrix of elements c_{ij}, find a matrix $X = \{x_{ij}\}$, such that

$$C = \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} x_{ij}$$

is minimized subject to

$$\sum_{i} x_{ij} = 1, \forall j$$
$$\sum_{j} x_{ij} = 1, \forall i$$

Here x_{ij} is a binary variable used for ensuring that an observation is associated with one and only one track and a track is associated with one and only one observation. This requires x_{ij} to be either 0 or 1. For finding matrix X, the Munkres algorithm is used [10].

The drawback of such aproach is that x_{ij} gets value 1 also in the case of zero raw or zero column of the cost matrix C. It results in misleading associations. The modifications, which corrects this drawback, rests in replacing of the matrix X by the matrix $X. * M$, where the operation $.*$ represents multiplication by elements with the "Gate Mask" matrix M.

E. Track Maintenance

A new target is identified when its observation fails all the already established gates, i.e. when all the elements of a row in the "Gate Mask" matrix M are zero. The case of "Observationless Gate" indicates the disappearance of a target from radar field of view. This is manifested when all the elements of a column in the "Gate Mask" matrix M are zero.

In the original MTT system, the "new target identifier" and the "Obs-less Gate Identifier" looks for 3 consecutive "misses" in 5 scans to confirm a new target or disappearance of a target, respectivelly. The counters are reset every five scans. The "Track Init/Del" initiates or deletes a track when needed.

Such conditions have not been satisfactory for the through wall target tracking. Here, the minimal number of successive observation-less gate identification for removal of correspondent track and the minimal number of successive new target identification for addition of new track have been set to value 10. In addition, the vicinity of new target identifications has to be fulfilled for initiation of new track.

IV. Experimantal Results

The performance of the modified MTT system is demonstrated by processing of the real radar signals acquired by the M-sequence UWB radar in two multitarget through wall scenarios. The radar system was equipped with one transmitting (Tx) and two receiving horn antennas (Rx_1, Rx_2), the positions of which are outlined in the bottom parts of Fig. 2. The distances between adjacent antennas were set to 1.3 m and there was no separation between the antennas and the wall. In both scenarios, two persons were moving in a gymnasium behind 24 cm thick wooden wall covered by tile.

The input data for the modified MTT system are depicted in Fig. 2(a) and 2(c). They represents the target locations computed by the trilateration process. Before the localization phase, the raw radar signals have been processed by methods of preprocessing, background subtraction, detection and trace estimation described in [11].

As can be seen from Fig. 2(b) and 2(d), the estimated tracks correspond very well with the true target trajectories indicated by blue squares. The tracks obtained for the first scenario are more accurate what results from the better input data. The loss of initial track positions for the second target (the black trajectory in Fig. 2(b)) was caused by the mutual shadowing of the targets. The complex input data for the second scenario effected the accuracy of the estimated tracks (Fig. 2(d)). However, the obtained results are still excellent.

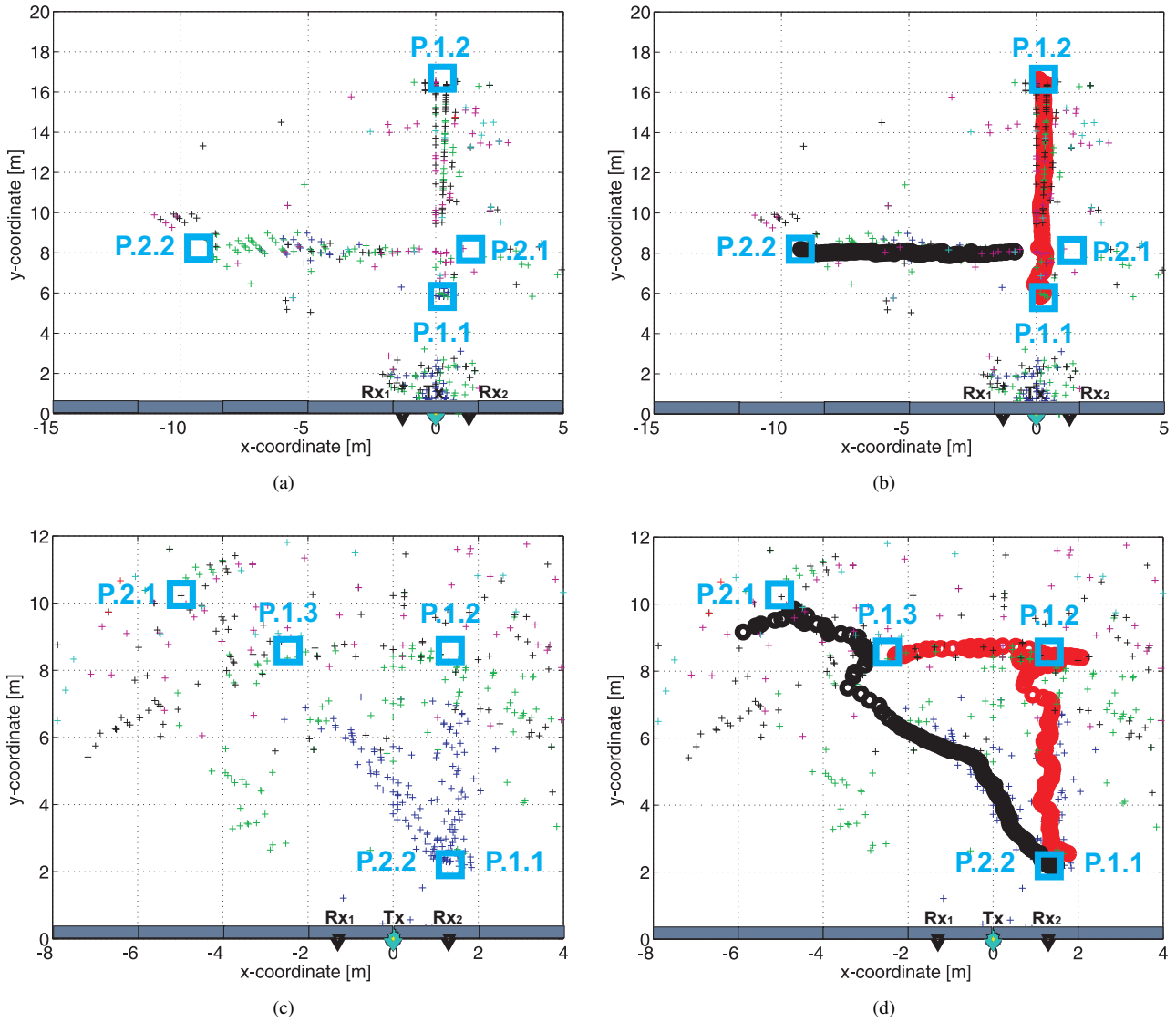


Fig. 2. Through wall scenarios with two moving targets, the true trajectories are indicated by blue squares. The first scenario: (a) estimated target positions, (b) estimated target tracks; The second scenario: (c) estimated target positions, (d) estimated target tracks.

V. Conclusion

The experimental results obtained by the processing of real UWB radar signals have confirmed good performance properties of the introduced modified MTT system for the through wall application. It has been shown that also in the case of inaccurate location estimations, the modern MTT system can greatly improved the estimations od final target positions. In spite of the demandingness of the solved tasks, the computational complexity of the whole tracking system is still suitable for a real time processing.

Acknowledgment

This work was supported by the Slovak Cultural and Educational Grant Agency (KEGA) under the contract No. 3/7523/09 and by the Slovak Research and Development Agency under the contract No. LPP-0080-09. This work is also the result of the project implementation Center of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research & Development Operational Program funded by the ERDF.

References

[1] S. Nag *et al.*, "An Ultra-Wideband Through-Wall Radar for Detecting the Motion of People in Real Time," *Proc. of SPIE - Radar Sensor Technology and Data Visualization*, vol. 4744, 2002.

[2] M. Kolawole, *Radar Systems, Peak Detection and Tracking*. Newnes, April 2003.

[3] M. S. Grewal and A. P. Andrews, *Kalman filtering: Theory and practice using MATLAB*, 3rd ed. Wiley-IEEE Press, September 2008.

[4] S. S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Artech House Publishers, August 1993.

[5] J. Khan, S. Niar, A. Rivenq-menhaj, and Y. E. Hillali, "Multiple target tracking system design for driver assistance application," *Design & Architectures for Signal and Image processing*, Nov. 2008.

[6] E. Brookner, *Tracking and Kalman Filtering Made Easy*. Wiley-Interscience, April 1998.

[7] D. Daniels, "M-sequence radar," in *Ground Penetrating Radar*. London, United Kingdom: The Institution of Electrical Engineers, 2004.

[8] E. Paolini, A. Giorgetti, M. Chiani, R. Minutolo, and M. Montanari, "Localization Capability of Cooperative Anti-Intruder Radar Systems," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, p. 14, March 2008.

[9] F. Folster, H. Rohling, and U. Lubbert, "An automotive radar network based on 77 GHz FMCW sensors," in *Radar Conference, 2005 IEEE International*, May 2005, pp. 871–876.

[10] Munkres' Assignment Algorithm, Modified for Rectangular Matrices, http://csclab.murraystate.edu/bob.pilgrim/445/munkres.html.

[11] J. Rovňaková, "Complete signal processing for through wall target tracking by M-sequence UWB radar system," Ph.D. dissertation, Technical University of Košice, Slovak Republic, August 2009.

Acquisition techniques to measure static characterization of High Resolution DAC

Martin SEKERÁK, Marian CHOVANEC

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

martin.sekerak@tuke.sk, marian.chovanec@tuke.sk

***Abstract*— A new method is presented for the dynamic characterization of high resolution Digital to Analog Converters (DAC). The methods are based on conversion of the DAC output voltage to time, that can be measured in the electronics much more accurate than amplitude. Method is based on the comparison of the DAC output voltage signal with a reference one. Proposed method is based on the time measurement loading using an oscilloscope. Information about time is acquired into the computer and then it is processed by PC. Time variation of output signal from DAC converter is used to evaluate the DAC transfer characteristic, the DNL and the INL. Measurement errors in such precision converters allow a better understanding of these errors for variable slope of the generated analog signal. Advantage for both methods is using simple instrumentation.**

***Keywords*— Analog to Digital converter, Differential Non Linearity, Integral Non Linearity.**

I. INTRODUCTION

The explosive growth in digital signal processing in recent years is due to its many beneficial properties. It also brought a sharp increase in production of electronic devices using digital signal processing in various areas of daily life such as telecommunications systems, computer systems, sensor technology, multimedia and many more. Digital processing has become very convenient but most of the signals in the real world are analogue. So there was a need to transfer analogue signal to digital before processing, as well as the processed digital signal to analogue, which is clearer and more used. For instance: sound, speech and video image and etc.

DAC conversion is the main way to convert a digital signal to an analogue real signal. Today electronic devices based on digital signal processing integrates more and more features, and each requires a higher requirement for further increases of the volume of processed data, speed but also accuracy and quality. Such a significant increase in performance and accuracy of today's electronic devices also requires the use of highly accurate, rapid and linear DAC converters. The implementation of such high quality DAC converters can be achieved by using suitable DAC architectures with their associated properties and technology used in production. Production of very accurate DAC converters, which resolution often exceeds the accuracy of the ADC converters for the world's major manufacturers of electronics is not already a problem even if each of them uses and enforces its own architecture.

The characterization of high resolution Digital-to- Analog converters is a challenge nowadays still open. The static characterization could be, in theory, fulfilled by means of high accuracy voltmeters even though the test duration makes it not practicable. But, still more challenging is high resolution DAC dynamic characterization. In this case, the use of a high speed/high resolution reference ADC (Analog to Digital Converter) able to gather the DAC variable voltage, should be required ([2], [3] - [10]).

These characteristic allows us to quickly and easily determine the actual properties, regardless of DAC architecture used to compare them with each other and further improve their properties more accurate evaluation of their mistakes.

II. THE PROPOSED METHOD FOR MEASURE DAC CONVERTERS

There are already several methods used for measuring the static and dynamic characteristic DAC converters [1].

The principle of the proposed method is based on the transfer amplitude of output signal from the DAC converter for time using fast comparators and measure time related with quantization level. This method was chosen because the measurement time using electronics available today is much more accurate than measuring amplitude. The obtained times for a high quality converters with high-resolution 14-bit or 16-bit, that corresponding only for voltage step LSB are very short. Their measure is not so easy and requires very fast electronic components and appropriate treatment and evaluation. The proposed ways to measure have been simulated in a simulation environment Lab View and has not been implemented.

III. METHOD OF HIGHLIGHTING A SMALL DIFFERENCE IN TIME

This method uses means of conversion output voltage from DAC converter for time corresponding to individual quantization levels and its subsequent measurements, which uses the properties of commonly available laboratory oscilloscopes, to achieve a sampling rate of 1 or 2GS/s. This high sampling rate allows very short time resolution of differences the proportion of errors at quantization level DAC

converter obtained from the comparator.

Block diagram of the proposed method called highlighting small differences in time is shown in Fig. 1.

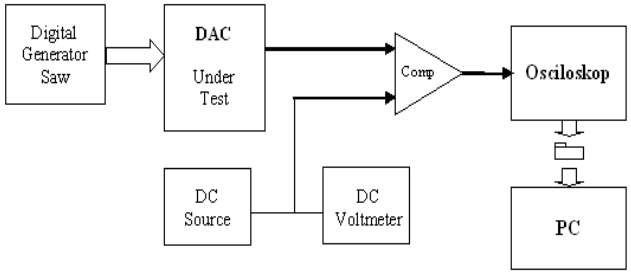


Fig. 1 Block scheme for method of highlighting a small difference in time

The Fig. 1 show that method requires the involvement of the DAC converter for digital output of code generator always addition +1 (digital generator saws). The output of the measured converter is connected to a fast comparator. On its second input connect U_{REF} from precision reference voltage source. The constant delay of comparator has been considered. The output of comparator is connected on the input of oscilloscope. Modern oscilloscope has possibility of storing samples in internal or external memory, what allows relatively easy to acquirement the samples and sends the file containing measured samples to PC where it will subsequently be processed and evaluated.

This method is relatively simple as it is mainly using the oscilloscope as an element in the measurement system. Oscilloscope serve as a quick sampling circuit and it loaded samples are stored in memory. Obtained record is not processed in real time so that does not require the use of fast components in measurement system. The disadvantage of proposed method is in limited memory space oscilloscopes and time required for the measurement converter. Memory can be expanded using an oscilloscope connected with PC and its use as an external memory, but also limited. Another disadvantage arises with high-resolution converters, which are time requiring to the individual quantization levels so short that not even sampling rate 1GS/s or 2GS/s is not sufficient. However, this shortcoming could be solved by using expensive high-speed oscilloscopes

IV. Behavioral Simulation for Method of Highlighting a Small Difference in Time

Method was not realized only verified through simulations, which the signal generator of saw (further sample increasing +1) represents the ideal output of the DAC converter which can be artificially added to a known failure. Simulation this method runs in cycles which in small steps (less than LSB) increase the value of precision reference voltage U_{REF} and compares them with the value of the output converter. Output of fast comparator produce a number of pulses related for individual quantization level of DAC converter. These pulses are evaluating and subsequent treatment as Differential Non Linearity (DNL) and Integral Non Linearity (INL). Because the measurement of such short time, or pulses of high frequency is susceptible to noise, simulation allows the addition of different types of noise, to assess their impact for accuracy when not only an ideal simulation environment.

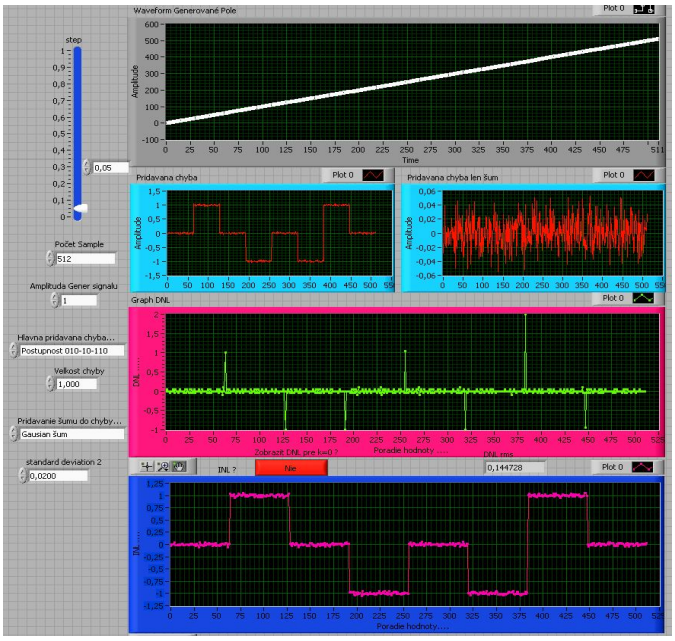


Fig. 2 Simulation for method of highlighting a small difference in time for DAC converter with 8 bit resolution in development environment

Obtain features DNL and INL in a development environment from simulation of method for highlighting a small time difference for the DAC converter with a resolution of 8bit and adding an error whose shape is shown on the graph with very small noise level can see in Fig. 2 and this same DAC converter with higher noise level in Fig. 3.

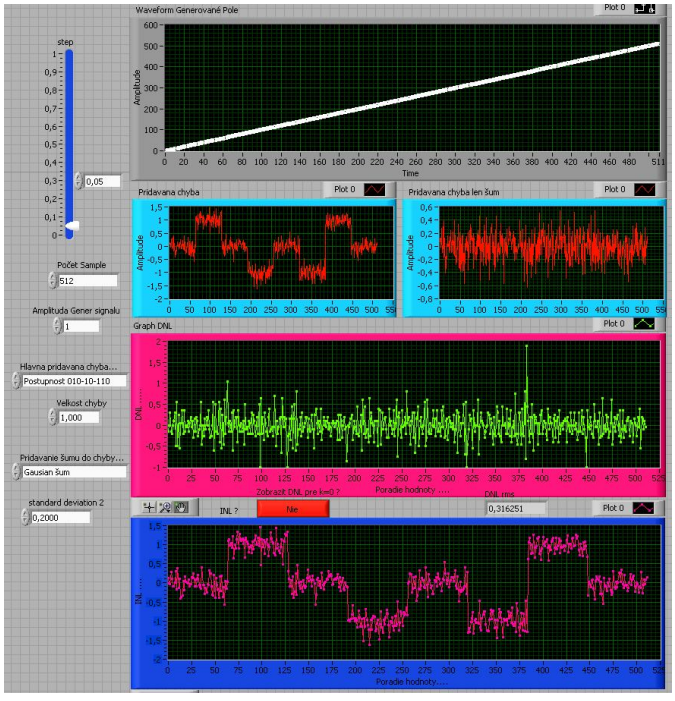


Fig. 3 Simulation for method of highlighting a small difference in time for DAC converter with 8 bit resolution in development environment

Method is simple and designed in preference for measuring defects larger than 1LSB with its big advantage is the lightweight devices used, for which we owe mainly use ordinary laboratory oscilloscope.

Weak point is the measurement of very short time, which

requires the use of very fast components and equipment, but on the other side is not required for high quality and linearity. The disadvantage is the setting for the DC voltage comparator, with a step less than quantization step itself, resulting time complexity methods.

V. CONCLUSION

Paper describes new method of measuring non-linearity DAC converters, which principle is based on the transfer of voltage to time and then measures it. Proposed method replaces implementation of the precise voltmeter with lower accuracy by the more precise time measurement. It allows achieve the sufficient accuracy for fast DAC with resolution 14 or 16 bit where are the voltages representing quantization level very small. This method described above provides a way that allows measure short time representing quantization level and obtain required characteristics.

Weak point this mean is in the comparator's own delay, that it's higher than the error that we want to detect, but here we assume that the comparator steal the same error in all measurements, and there can be corrected.

Behavioral methods under ideal conditions has been verified simulation, which showed that the method behaves as required, to implement problem may occur when you set up a very small voltage, which has to be less than quantization step (1LSB) of the tested converter. Voltage reference step size, relative to 1LSB determines the accuracy of measurement required characteristics of INL and DNL therefore this method is suitable for measure errors 1LSB and more.

Another very important fact in implementing the measure is also a very short time ($t_X << 1ns$), which will require a very fast electronic components and equipment and suppression of various interference.

REFERENCES

[1] D. L. Cari, D. Grimaldi, "Comparative analysis of different acquistion techniques applied to static and dynamic characterization of hight resolution DAC", *XIX IMEKO World Congress Fundamental and Aplied Metrology*, September 6-11, 2009, Lisabon, Portugal

[2] D.L. Cari, D. Grimaldi, "Static characterization of high resolution DAC based on over sampling and low resolution ADC", *Proc. Of IEEE Instrum. And Measur. Techn. Conf., IMTC 2007*, May 1-3, 2007, Warshaw Poland

[3] J. Le, H. Hosam, R. Geiger, C. Degang, "Testing of precision DACs using low-resolution ADCs with dithering", *Proc. Of IEEE Intern. Test Conf.*, Oct. 2006, pp.1-10.

[4] B.Vargha, J.Schoukens, Y.Rolain, "Static nonlinearity testing of digital-to-analogue converters," *IEEE Trans. on Instrum. and Meas.*, Vol.50, N°5, October 2001, pp.1283-1288

[5] A. Baccigalupi, Mauro D`Arco, A. Liccardo, M. Vardusi, "Implementation of high resolution DAC test station: A contribution to draft standard IEEE P1658", XIX IMEKO World Congress Fundamental and Aplied Metrology, September 6-11, 2009, Lisabon, PortugalB. Smith, "An approach to graphs of linear forms (Unpublished work style)," unpublished.

[6] Chun Wei Lin, Sheng Feng Lin, Shih Fen Luo, "A new approach for nonlinearity test of high speed DAC", *IEEE Int.Workshop on Mixed-Signals, Sensors, and Systems*, pp. 1-5, 2008.

[7] L.Angrisani, M.D'Apuzzo, M.D'Arco, "A new approach to linearity and intermodulation errors estimation in digital-to-analogue converters," in *Proc. of 9th Workshop on ADC modelling and testing*, 29 Sept. – 1 Oct. 2004, Athens, Greece, vol. 2, pp.859-864.

[8] J. Savoj, Ali-Azam Abbasfar, A. Amirkhany, B.W. Garlepp, M.A. Horowitz, "A new technique for characterization of Digital-to-Analog Converters in high-speed systems" *Design, Automation & Test in Europe Conf.& Exhibition,* 16-20 April 2007, pp.1-6.

[9] "Draft Standard for Terminology and Test Methods for Digital-to-Analog Converters", IEEE std P1658, September 2008.

[10]] Chun Wei Lin, Sheng Feng Lin, Shih Fen Luo, "A new approach for nonlinearity test of high speed DAC", *IEEE Int.Workshop on Mixed-Signals, Sensors, and Systems*, pp. 1-5, 2008.

Kalman filters for target tracking by UWB radar systems

Mária ŠVECOVÁ

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

maria.svecova@tuke.sk

***Abstract*—Kalman filters have become a standard technique used in a number of applications applied for target tracking due to their simplicity, optimality, tractability and robustness. This paper is focus on through wall target tracking by one or two independent UWB radar systems every equipped with one transmitting and two receiving antennas. For that purpose, linear Kalman filter, extended Kalman filter and unscented Kalman filter are used and compared based on real radar data processing. It is shown, that the linear Kalman filter preceded by suitable localization method provides the better target trajectory estimation than extended Kalman filter and unscented Kalman filter at systems with more antennas. It is identified if two UWB radar systems are used for target tracking. On the other hand, if a single UWB radar system is used, the target trajectories are comparable.**

***Keywords*—Kalman filtering, target tracking, UWB radar system.**

I. INTRODUCTION

The localization capability is becoming one of the most attractive features of wireless sensor network systems. Ultra wideband (UWB) radar systems as the special kind of wireless sensor network systems allow localize and track authorized or unauthorized targets with advantages in critical environments or under hindered conditions, e.g. through wall tracking of moving people during security operations, through rubble localization of trapped people after an earthquake or an explosion, through snow detection of trapped people after an avalanche, etc.

Moving target localization and tracking by UWB radar system, i.e. determining target coordinates as the continuous function of the time, is the complex process that includes such phases of radar signal processing as raw radar data pre-processing, background subtraction, detection, time of arrival (TOA) estimation, localization and tracking itself. The significance of these particular phases of radar signal processing can be found in [1] or [2].

In this paper we will focus on tracking as one of the phases of radar signal processing. We will assume that for moving target tracking UWB radar system or two independent UWB radar systems every equipped with one transmitting and two receiving antennas are used. It is expected that the antenna positions are known and TOA corresponding to the receiving antennas have been estimated by the particular phases of radar signal processing (raw radar data pre-processing, background subtraction, detection and TOA estimation).

Fot target tracking a number of methods have been proposed where the most important groups are reprezented by Kalman filters [3], [4] and particle filters [5]. There is a number of several variants of them in dependence on dynamic system. Here, three variants of Kalman filters for target tracking by UWB radar systems will be considered, namely linear Kalman filter, extended Kalman filter and unscented Kalman filter. Because internal equations of the Kalman filters are the same for various dynamic systems, we will therefore focus on detailed description of our dynamic systems and comparison of Kalman filters based on real UWB radar signal processing. The obtained results expressed by the target trajectory estimation show that the best accuracy of the target trajectory estimation is provided by linear Kalman filter.

II. KALMAN FILTERING

A. Problem statement

Let us consider a fundamental scenario of through wall tracking of a moving target by means of two independent UWB radar systems denoted as $\mathrm{RS_A}$ and $\mathrm{RS_B}$. Here, every radar system is equipped with one transmitting and two receiving antennas. It is assumed, that the antenna positions are known and their coordinates are given as follows:

- coordinates of the transmitting antenna of $\mathrm{RS_A}$ (Tx_A): $Tx_A = (x_{A,t}, y_{A,t})$,
- coordinates of the receiving antennas of $\mathrm{RS_A}$ ($Rx_{A,1}$, $Rx_{A,2}$): $Rx_{A,1} = (x_{A,1}, y_{A,1})$, $Rx_{A,2} = (x_{A,2}, y_{A,2})$,
- coordinates of the transmitting antenna of $\mathrm{RS_B}$ (Tx_B): $Tx_B = (x_{B,t}, y_{B,t})$,
- coordinates of the receiving antennas of $\mathrm{RS_B}$ ($Rx_{B,1}$, $Rx_{B,2}$): $Rx_{B,1} = (x_{B,1}, y_{B,1})$, $Rx_{B,2} = (x_{B,2}, y_{B,2})$.

Let $TOA_{R,i}(k)$ for $R = A, B$, $i = 1, 2$ represents TOA estimation of the electromagnetic wave observed at the time instant k transmitted by Tx_R, reflected by the target ($T(k) = (x(k), y(k))$) and received by $Rx_{R,i}$ for $R = A, B$, $i = 1, 2$. Here, we presume that $TOA_{R,i}(k)$ for $R = A, B$, $i = 1, 2$ at every time instant k have been estimated by the process described in [1]. In addition, TOA estimations have been observed at the discrete time instants $t_k = \Delta t.k + t_o$, $k \in N$, where Δt is an interval between two observation times. Then, the distance $d_{R,i}(k)$ for $R = A, B$, $i = 1, 2$ at every time instant k between Tx_R, $T(k)$ and $Rx_{R,i}$ can be expressed as $d_{R,i}(k) = c.TOA_{R,i}(k)$ where c is the propagation velocity of the electromagnetic wave. In our consideration, c is set to the electromagnetic wave propagation velocity in air. On the other hand, the distance $d_{R,i}(k)$ can be expressed also by means of antenna and target coordinates as follows:

$$\begin{aligned} & d_{R,i}(k) = r_{R,i}(k) + e_{R,i}(k) \\ & = \sqrt{(x(k) - x_{R,t})^2 + (y(k) - y_{R,t})^2} \\ & + \sqrt{(x(k) - x_{R,i})^2 + (y(k) - y_{R,i})^2} + e_{R,i}(k), \\ & R = A, B, \quad i = 1, 2 \end{aligned} \tag{1}$$

where $r_{R,i}(k)$ is the true Euclidean distance between Tx_R, $T(k)$ and $Rx_{R,i}$ and $e_{R,i}(k)$ is the component expressing the errors of TOA estimation.

Let $\mathbf{x}(k) = [x(k)\ y(k)\ v_x(k)\ v_y(k)]^T$ is the state of the system at the time instant k where $x(k)$, $y(k)$, $v_x(k)$, $v_y(k)$ are the cartesian coordinates of the target and their velocities at the time instant k, respective. The target motion can be described in matrix form as:

$$\mathbf{x}(k) = \mathbf{A}\mathbf{x}(k-1) + \mathbf{B}\mathbf{n}(k) \tag{2}$$

where the state transition matrix $\mathbf{A}$ and matrix $\mathbf{B}$ of the forms

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \ \mathbf{B} = \begin{bmatrix} \Delta t^2/2 & 0 \\ 0 & \Delta t^2/2 \\ \Delta t & 0 \\ 0 & \Delta t \end{bmatrix} \tag{3}$$

relate the state at time instant $k-1$ to the state at time instant k. It is assumed that the process noise $\mathbf{n}(k)$ is zero mean white Gaussian noise. The state model (2) projects the current state to the future.

A measurement model describing the relationship between the true state $\mathbf{x}(k)$ and the distance estimations $d_{R,i}(k)$ for $R = A, B$, $i = 1, 2$ at the time instant k can be expressed by using (1) as follows:

$$\mathbf{d}(k) = \mathbf{f}(\mathbf{x}(k)) + \mathbf{e}(k) \tag{4}$$

where

$$\begin{aligned} & \mathbf{d}(k) = \begin{bmatrix} d_{A,1}(k) \\ d_{A,2}(k) \\ d_{B,1}(k) \\ d_{B,2}(k) \end{bmatrix}, \ \mathbf{f}(\mathbf{x}(k)) = \begin{bmatrix} f_{A,1}(\mathbf{x}(k)) \\ f_{A,2}(\mathbf{x}(k)) \\ f_{B,1}(\mathbf{x}(k)) \\ f_{B,2}(\mathbf{x}(k)) \end{bmatrix}, \\ & f_{R,i}(\mathbf{x}(k)) = \sqrt{(x(k) - x_{R,t})^2 + (y(k) - y_{R,t})^2} \\ & + \sqrt{(x(k) - x_{R,i})^2 + (y(k) - y_{R,i})^2}, \ R = A, B, \ i = 1, 2, \\ & \mathbf{e}(k) = [e_{A,1}(k)\ e_{A,2}(k)\ e_{B,1}(k)\ e_{B,2}(k)]^T. \end{aligned} \tag{5}$$

Note that the state model for target tracking by one UWB radar system is the same as that one suggested for target tracking by two independent UWB radar systems (2). The measurement model (4) for that scenario is reduced to two equations belonging only to one radar system, e.g. $\mathrm{RS_A}$.

The problem of target tracking is to estimate target position and velocity trajectories from noise corrupted distances $d_{R,i}(k)$, $k \in N$. The linear, extended and unscented Kalman filters issued from the devised state model (2) and measurement model (4) can be used for that purpose.

B. Linear Kalman filter

The linear Kalman filter (LKF) supposes the linear state and measurement models. But in many cases, dynamic systems are not linear by nature as well as in our Scenario II-A. The localization of the target by suitable localization method can precede therefore target tracking by LKF.

The method of joining intersections of the ellipses (JIEM) [6] suggested for the target localization by two independent UWB radar systems can be used for that purpose. The target is localized here by using TOA estimates corresponding to the target to be tracked obtained for four receiving antennas under the conditions, that the positions of all antennas are known. The direct calculation method (DC) [1] can be used for the target localization by one UWB radar system. The target coordinates $\hat{x}(k)$, $\hat{y}(k)$ are the output of the localization estimated for every time instant k. For that scenario, the measurement model can be described as follows:

$$\mathbf{p}(k) = \mathbf{H}\mathbf{x}(k) + \mathbf{w}(k) \tag{6}$$

where $\mathbf{H}$ is measurement model matrix of the form

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \tag{7}$$

$\mathbf{p}(k) = [\hat{x}(k)\ \hat{y}(k)]^T$ is the vector containing the target coordinates estimated by localization and $\mathbf{w}(k)$ is the noise resulting from inaccuracy of $\hat{x}(k)$, $\hat{y}(k)$ estimates. It is assumed that $\mathbf{w}(k)$ is zero mean white Gaussian noise.

Then, the state model (2) and measurement model (6) are now linear. LKF can be used for the estimation of the target trajectory. The corresponding equations of LKF can be found in [3], [4], but their input variables are that given in this article.

In general, the Kalman filter is a recursive estimator that needs only the estimated state from the previous time instant and the current measurement to compute the estimate for the current state. It works in two steps: the prediction step, where the next state of the system is predicted given the previous measurements, and the update step, where the current state of the system is estimated given the measurement at that time instant. Typically, the both steps alternate, with the prediction advancing the state until the next scheduled observation, and the update incorporating the observation.

C. Extended Kalman filter

In the case of the extended Kalman filter (EKF) [3], [7], [8], the state transition or observation models need not be linear functions of the state, but they have to be instead differentiable functions. EKF linearizes the nonlinear state or measurement functions or both around the last estimated state by calculating the Jacobian matrix, by evaluating the partial derivatives of a vector with respect to the state variables. After that LKF equations [3], [4] can be applied.

D. Unscented Kalman filter

In the case of the unscented Kalman filter (UKF) [8], [9], [10], the state transition or observation models need not be linear functions of the state, too. UKF uses a deterministic sampling technique known as the unscented transform to pick a minimal set of sample points, called sigma points, around the mean. These sigma points are then propagated through the nonlinear functions, from which the mean and covariance of the estimate are then recovered. The points are chosen such that their mean, covariance, and possibly also higher order moments, match the Gaussian random variable. The estimated mean and covariance from the propagated sigma points are more accurate compared to ordinary function linearization.

Instead of detailed derivation of LKF, EKF and UKF equations we will focus in the next Section on their comparison on real data.

Fig. 1: M-sequence UWB radar system with horn antennas.

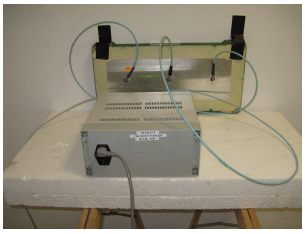


Fig. 2: M-sequence UWB radar system with spiral antennas.

Fig. 3: Scenario 1: room interior. **Fig. 4:** Scenario 2: room interior.

III. Experimental results

In order to illustrate the performances of LKF, EKF and UKF, two real scenarios of through wall tracking of a moving person have been analysed. The signals have been acquired by the measurements by one and two independent M-sequence UWB radar systems described in [11]. The radar devices have operated in the bandwidth of about DC-2.25 GHz with the total power about 1 mW. Impulse responses were acquired by a measurement rate of about 10 impulse responses per second what was sufficient to match the time variance caused by a walking person. Every radar system was equipped by one transmitting and two receiving double-ridged horn antennas (Fig. 1) or one transmitting and two receiving spiral antennas placed along line (Fig. 2). The transmitting antennas were located in the middle of receiving antennas.

The raw radar data acquired by the measurements have been processed by the procedure described in [1]. TOA estimations corresponding to all receiving antennas obtained by this procedure have been used for target tracking.

Scenario 1. Target tracking by UWB radar system

Scenario 1 was focused on the visual comparison of the tracking ability and accuracy of LKF, EKF and UKF.

The scheme of this scenario is outlined in Fig. 5. It is represented by moving person tracking through a brick wall covered by tiles with a total thickness of 24 cm by UWB radar system. The person was moving inside of the room (Fig. 3) from the position P(1), through the positions P(2), P(3), P(4), back to the position P(1) (Fig. 5). The distance between adjacent receiving antennas was set to 2.6 m. All antennas were placed 1.2 m above the floor. Other distances are schematically depicted in the Fig. 5.

Scenario 2. Target tracking by two UWB radar systems

Scenario 2 is represented by moving target tracking through bricky walls by two independent M-sequence UWB radar systems. The thickness of the first and the second wall has been 0.3 m and 0.43 m, respectively. A person to be tracked was walking inside of the room (Fig. 4) from the position P(1) through the positions P(2), P(3) to the position P(4) (Fig. 7). The first radar system denoted as RS_A has been equipped with spiral antennas (Fig. 2) and the second one denoted as RS_B with double-ridged horn antennas (Fig. 1). The distances between adjacent antennas was set to 0.175 m and 0.43 m, respectively.

The moving target trajectories for Scenario 1 are given in the Fig. 6 and for Scenario 2 in the Fig. 8. Here, by the localization methods estimated target positions for every observed time instants are depicted by the gray crosses. DC for Scenario 1 and JIEM for Scenario 2 have been used for that purpose. LKF applied to positions estimated by DC and JIEM computed the target trajectories depicted by red solid lines. By EKF and UKF estimated target trajectories are depicted by blue dotted lines and green solid lines with circles, respectively.

The alternative form of the visualization of the target position estimation accuracy is shown in the Fig. 8 for Scenario 2. For that purpose, the area bordered by the red dash-line is sketched. Inside this area, the true trajectory of the target is located. This area will be referred to as the region of the true positions of the target. The width of this region along the x- and y-coordinates is set to 0.5 m, which corresponds approximately to the width of a human body. Since the real width of the target is non-zero we can accept the estimated target position as the true one, if the target is located inside this region. This approach allows us to evaluate the target position estimation accuracy as the percentage of the "true" estimates of the target positions. This quantity can be evaluated as the ratio of the number of the target positions inside of the region of the true positions of the target to the total number of the estimated target positions. The percentages of the "true" target position estimates for LKF, EKF and UKF are given in Table I.

In Table I, the so-called average time of calculation is also brought out. This quantity represents the average time of the calculation of the target positions from estimated TOA by the tested Kalman filters at their implementation in MATLAB environment. Therefore, this quantity can be taken as an approximated measure of the filter complexity. For that application a standard PC can be used.

Now, following Fig. 6, 8 and Table I, we can evaluate and discuss the performance of Kalman filtering applied for target tracking. If we compare all target trajectories for Scenario 1 with one UWB radar system, the estimated trajectories are similar. The reason is too small number of receiving antennas. Another situation occured in Scenario 2 with two UWB radar systems, where the estimated target trajectories are quite different. By visual comparison of the target trajectory estimations and taken into account the percentages of the "true" estimates of the target positions it can be concluded that the best estimation of the target trajectory is provided by LKF. It results from the fact, that EKF linearizes the nonlinear system what can easily lead to divergence and great inaccuracies. On the other hand, UKF was very sensitive to input parameters what can be one of the reason of relatively bad performed estimation of the target trajectory. The best performance of the LKF is reached at the cost of its higher

Kalman filters	LKF	EKF	UKF
Percentage of the "true" estimates of the target positions [%]	73	57	68
Average time of calculation [ms]	12.85	1.09	3.85

TABLE I: Comparison of the Kalman filter performance.

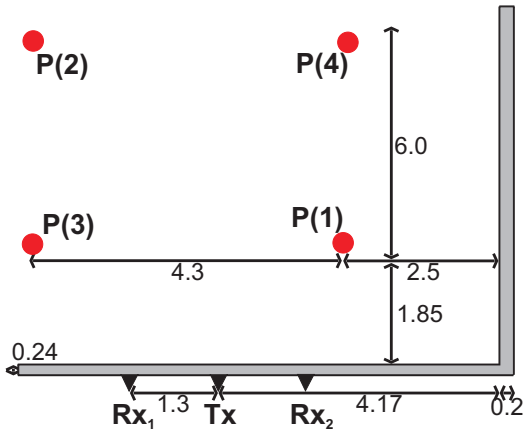


Fig. 5: The measurement scheme of Scenario 1.

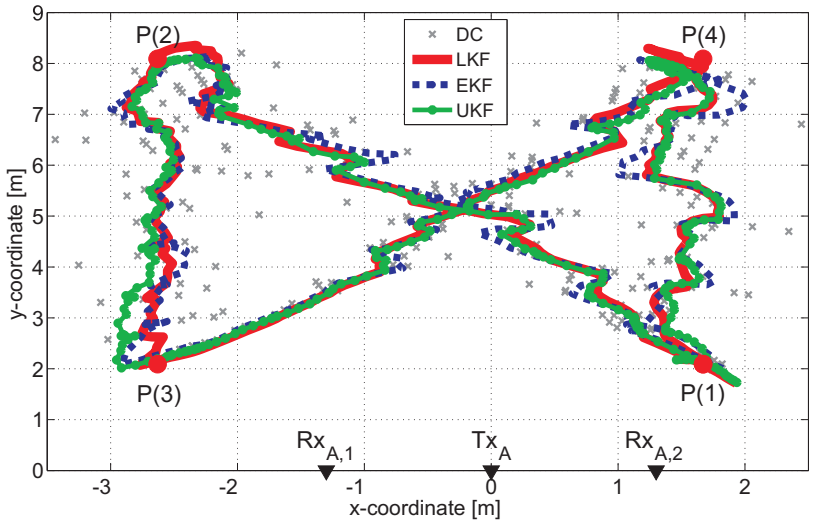


Fig. 6: The target trajectories for Scenario 1.

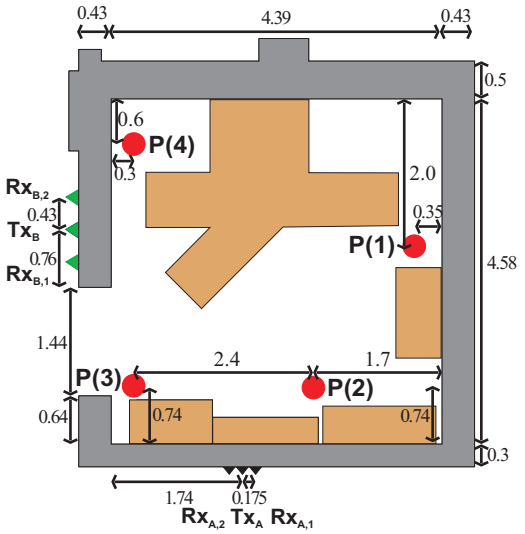


Fig. 7: The measurement scheme of Scenario 2.

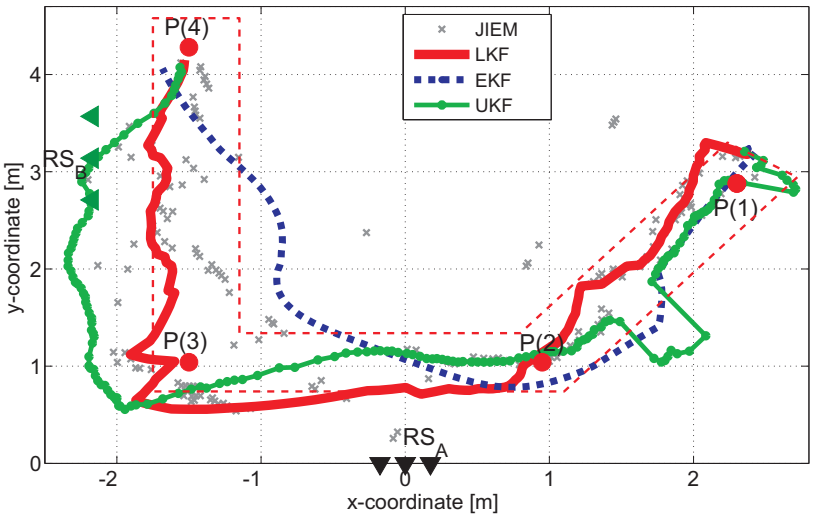


Fig. 8: The target trajectories for Scenario 2.

complexity in comparison with EKF and UKF.

IV. Conclusion

In this paper, we have compared LKF, EKF and UKF based on real radar data processing. We have estimated the moving target trajectories by using one or two independent UWB radar systems. We have shown, that the estimated target trajectories by LKF, EKF and UKF for one UWB radar system are comparable. LKF have achieved the better estimation of the target trajectory with comparison to other tested method for two UWB radar systems.

Acknowledgment

This work is the result of the project implementation Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] D. Kocur, J. Rovňáková, and M. Švecová, "Through Wall Tracking of Moving Targets by M-Sequence UWB Rradar," in *Towards Intelligent Engineering and Information Technology*, I. J. Rudas, J. C. Fodor, and J. Kacprzyk, Eds. Springer, 2009, pp. 349–364.

[2] J. Rovňáková, M. Švecová, D. Kocur, T. T. Nguyen, and J. Sachs, "Signal Processing for Through Wall Moving Target Tracking by M-sequence UWB Radar," in *Proc. 18th International Conference Radioelektronika*, Prague, Czech Republic, April 2008, pp. 65–68.

[3] E. Brookner, *Tracking and Kalman Filtering Made Easy*. Wiley-Interscience , 1998.

[4] M. S. Grewal and A. P. Andrews, *Kalman Filtering: Theory and Practice Using MATLAB*, 2nd ed. Wiley-Interscience, 2008.

[5] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House, 2004.

[6] M. Švecová and D. Kocur, "Target localization by intersection of ellipses," in *11th International Radar Symposium IRS-2010*, Vilnius, June 2010, accepted for publication.

[7] M. I. Ribeiro, "Kalman and Extended Kalman Filters: Concept, Derivation and Properties," *Institute for Systems and Robotics, IST*, 2004.

[8] S. J. Julier and J. K. Uhlmann, "A New Extension of the Kalman Filter to Nonlinear Systems," in *Proc. International Symposium on Aerospace/Defense Sensing, Simululation and Controls*, 2000.

[9] E. A. Wan and R. van der Merwe, "The Unscented Kalman Filter for Nonlinear Estimation," in *Proc. of Symposium 2000 on Adaptive Systems for Signal Processing, Communications, and Control*, Aug. 2000, pp. 153–158.

[10] H. Qasem and L. Reindl, "Unscented and Extended Kalman Estimators for non Linear Indoor Tracking Using Distance Measurements," *4th Workshop on Positioning, Navigation and Communication*, pp. 177–181, March 2007, Hannover.

[11] J. Sachs, "M-sequence radar," in *Ground Penetrating Radar*, 2nd ed., D. Daniels, Ed. London, United Kingdom: Institution of Electrical Engineers, 2004.

^{1}H and ^{13}C NMR study of polypropylene granulates

Magdaléna UHRÍNOVÁ, Oľga FRIČOVÁ, Viktor HRONSKÝ

Department of Physics, FEI TU of Košice, Slovak Republic

magdalena.uhrinova@tuke.sk, olga.fricova@tuke.sk, viktor.hronsky@tuke.sk

Abstract—Two samples of isotactic polypropylene prepared by classical Ziegler-Natta catalysis and metallocene-catalysed polymerization were studied by means of ^{1}H MAS NMR and ^{13}C CP MAS NMR techniques. The NMR spectra were measured in the temperature range of 20 – 100 °C. The strong narrowing of the ^{1}H MAS NMR spectra and the change of the shape of the ^{13}C CP MAS NMR resonance lines related to the individual groups of polypropylene were observed with the rise of the temperature. The changes of the spectra were related to the conformation changes in the amorphous region of the polymer. The ^{1}H and ^{13}C NMR spectra measured on both iPP polymers show that amorphous region of metallocene PP is more mobile than of that prepared by classical Ziegler-Natta catalysis.

***Keywords*— isotactic polypropylene, ^{1}H MAS NMR, ^{13}C CP MAS NMR**

I. INTRODUCTION

Isotactic polypropylene (iPP) is an important engineering plastic used in many different application areas. Although more than fifty years have passed from its birth, scientists still labour for improvement of its mechanical, optical, thermal, and environmental properties for new technology areas [1]. Isotactic PP is a stereoregular polymer with chains crystallized in helical form. The chains of iPP may be ordered into regions with different arrangement and different mobility and the microstructure of the iPP and its physical properties may be strongly affected by the preparation technique [2].

The nuclear magnetic resonance due to its sensitivity to morphology is a suitable tool for characterization of order and molecular motion, orientation, alignment and molecular dynamics of observed nuclei [3]. The influence of the preparation technology on the morphology of iPP is the interest of our study. The first stage of this study, presented in this paper, is to perform NMR experiments suitable for the evaluation of the effect of the sample preparation technique on the morphology of iPP and to find the basic relations between measured spectra and structure and processes going on in the studied materials. For this purpose the ^{1}H and ^{13}C NMR spectra of the iPP samples prepared by the classical polymerization using classical Ziegler–Natta catalyst and by the metallocene-catalysed polymerization were measured.

II. THEORETICAL BACKGROUND

A. Properties of examined materials

Isotactic polypropylene has a helical molecular chain conformation as the most stable conformation [4]. It is regarded as a three-phase system with amorphous, intermediate and crystalline phases, which differ in degree of alignment and mobility of chains [5].

In general, in iPPs prepared by metallocene catalysis the distribution of stereo defects is homogeneous and consequently the average length of isotactic sequences in this polypropylene (PP) is shorter than that in the PP prepared by classical Ziegler-Natta catalysis, where the formation of stereoblocks takes place. Because of the different configurational structure, the melting temperature of metallocene iPPs is lower than that of Ziegler-Natta iPPs. Metallocene polypropylenes are much more homogenous both in molecular weight, in tacticity and tacticity distribution, and chains resemble one another much more than when using Ziegler-Natta catalysis. In fact isotactic PPs obtained by classical catalysis can be considered as a mixture of very different types of chains. Short atactic chains are present even in the most isotactic commercial PPs [6].

B. Description of NMR techniques

The shape of the NMR spectrum depends on the interactions in which the examined nuclei participate. The Hamiltonian of a spin system with spin number 1/2, as is the case of ^{1}H and ^{13}C nuclei, in a static external field can be written as

$$H = H_Z + H_S + H_D + H_J, \quad (1)$$

where H_Z represents the direct Zeeman interactions of spins with the external magnetic field $\boldsymbol{B_0}$, H_S describes the indirect interaction of spins with $\boldsymbol{B_0}$ via electrons, H_D and H_J reflect the direct and indirect interaction between spins, respectively. In NMR experiments in solution the direct interaction H_D is averaged by the rapid reorientation of the internuclear vectors and high resolution spectra are detectable with detailed information about molecular structure or conformation of the investigated sample. However, in solids H_D broadens the main resonance line given by H_Z and completely masks all lines due to H_S and H_J. The so-called magic angle spinning (MAS) technique is often used to obtain a spectrum of higher resolution for solids. By this procedure the sample rapidly rotates around an axis inclined at an angle of 54°44’ to the static field $\boldsymbol{B_0}$ [4]. This technique eliminates not only the anisotropy of the interactions but by very high spinning rates also removes the effect of homonuclear dipolar interactions from NMR spectra. Slower spinning produces a set of spinning sidebands in addition to the line at the isotropic

chemical shift [7].

Molecular motion can average out some of the dipolar interactions and reduce the line width [8]. The half width $\Delta f_{1/2}$ of the peaks at their half height is the wider the less mobile are spins in the studied material.

To obtain highly resolved ^{13}C NMR spectrum dipolar decoupling (DD) and cross polarization (CP) techniques are used together with MAS. DD technique eliminates the effect of heteronuclear dipolar interactions by applying a resonant oscillating field $\boldsymbol{B}_1$ on ^{1}H spins, which rapidly changes the direction of the dipolar field between ^{13}C and ^{1}H. By CP the ^{13}C magnetization is enhanced by the equilibrium magnetization of abundant ^{1}H spins. This procedure is achieved by contacting ^{13}C spin system with ^{1}H system during the contact time, while Zeeman energies of both spins in the field $\boldsymbol{B}_1$ are equalized [4].

III. EXPERIMENTAL

A. Samples

Two samples of granulated iPP were studied. The first one denoted as TATREN HG 1007 was prepared by polymerization using classical Ziegler–Natta catalyst. It is predominantly isotactic PP with the crystallinity of approx. 55% (data from DSC), melting temperature $T_m = 163.6$ °C (DSC), glass transition temperature $T_g = 10$ °C (data from DMTA). The second sample denoted as PP METOCENE HM 562 N was prepared by metallocene-catalysed polymerization. It is predominantly isotactic PP with the crystallinity of approx. 52 % (DSC), melting temperature $T_m = 145.2$ °C (DSC), glass transition temperature $T_g = 12$ °C (DMTA).

B. Experimental conditions

The NMR measurements were performed on the Varian NMR spectrometer for solids installed at the Department of Physics, Faculty of Electrical Engineering and Informatics, Technical University of Košice. The spectrometer is equipped with an actively shielded superconducting magnet generating magnetic field of 9.4 T in the bore of the diameter of 89 mm (a wide bore magnet). ^{1}H and ^{13}C resonant frequencies corresponding to the above mentioned magnetic field are 400 and 100 MHz, respectively. All NMR experiments were carried out with a probe-head using the 4 mm rotor and under the spinning rate of 6 kHz. The cross polarization contact time was 1 ms and the proton-decoupling field of 85 kHz was applied during data acquisition.

The spectra were recorded in the temperature range of 20 – 100 °C.

IV. RESULTS

The ^{1}H MAS NMR spectra of TATREN recorded at room temperature and at 60°C are shown in Fig. 1. The spectra recorded at both temperatures consist of the broad and narrow lines reflecting chains within different morphological regions. In the spectrum detected at 60°C besides the line at 1.13 ppm also the spinning sidebands at the multiples of spinning frequency [7] can be seen. The line narrowing in this spectrum reflects the chain mobility increase [8].

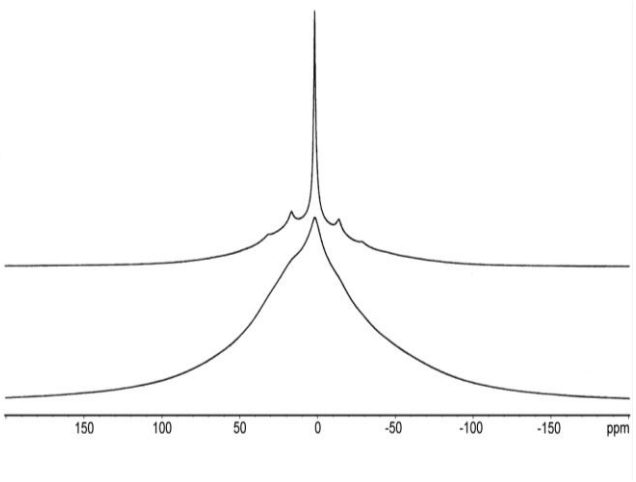


Fig. 1. ^{1}H MAS NMR spectra of TATREN measured at ambient temperature (bottom) and 60°C (top)

The ^{1}H MAS NMR spectra of METOCENE in the investigated temperature range are very similar to those of TATREN, however, above the room temperature the half widths $\Delta f_{1/2}$ of the central line of METOCENE are lower than those of TATREN. The values of the half widths $\Delta f_{1/2}$ of peaks in the spectra measured at temperatures 22 and 60 °C are presented in Tab. 1. The differences between these values indicate different mobility of the chains in the amorphous regions and therefore also structural differences in amorphous regions of investigated polymers.

TABLE I
THE HALF WIDTHS $\Delta f_{1/2}$ OF THE PEAKS OF ^{1}H MAS NMR SPECTRA

		TATREN	METOCENE
$\Delta f_{1/2}$ (kHz)	22°C	21,43	21,12
	60°C	0,84	0,61

The ^{13}C CP MAS NMR spectra measured on TATREN at room temperature and at the temperatures of 70 and 98 °C are shown in Fig. 2. The peaks related to the CH_2, CH and CH_3 groups within the chains of amorphous and crystalline regions are positioned at chemical shifts 44.14, 26.42 and 21.86 ppm, respectively. The observed changes in the spectra indicate that the temperature increase induces conformational changes of the PP chains that results in appearance of new peaks in the spectrum whose intensities increase with increasing temperature at the expense of those ones observed at room temperature. These peaks have chemical shifts 46.21 and 28.50 ppm and they are close to the original CH_2 and CH resonance lines, respectively. The change of the shape of the CH_3 resonance peak is also observed.

Similar changes in ^{13}C CP MAS NMR spectra of iPP with increasing temperature were observed by Kitamaru [4]. The additional peaks that appeared at higher temperatures are assigned to the amorphous component, which is at the room temperature in quasi-glassy state. The molecular

conformations in the amorphous phase are distributed over all permitted conformations stationary in time and randomly in space and the resonance line of the nuclei in this phase is distributed over a very wide chemical shift range centered to the same chemical shift as at higher temperature and so they are not clearly observed in the spectra measured at room temperature [4].

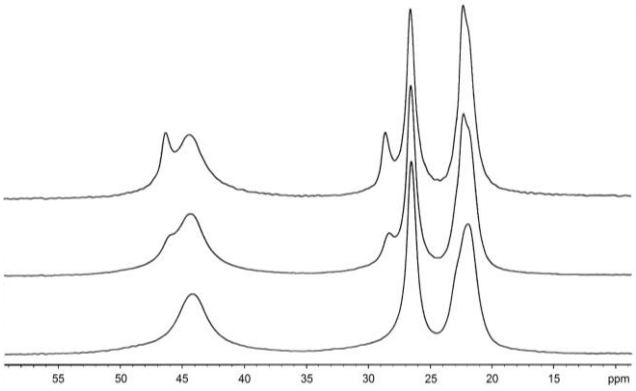


Fig. 2. ^{13}C CP MAS NMR spectra of TATREN measured at ambient temperature (bottom) and the temperatures of 70 (middle) and 98 °C (top)

Therefore, the changes of the spectra described above can be related to the amorphous regions of the investigated material and then under favourable conditions the NMR spectrum can give information on amorphous and crystalline regions separately.

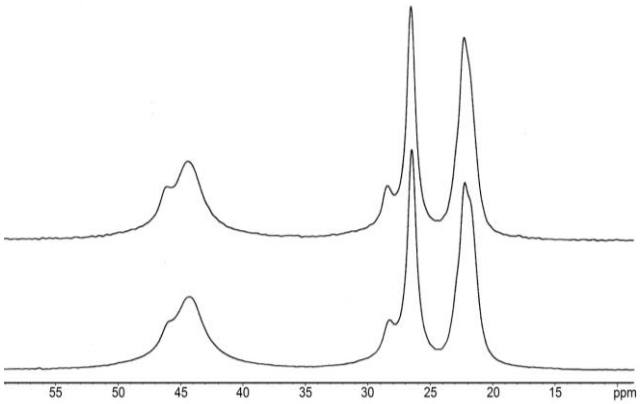


Fig. 3. ^{13}C CP MAS NMR spectra of TATREN (bottom) and METOCENE (top) measured at the temperature 70°C

In Fig. 3 ^{13}C CP MAS NMR spectra of both samples measured at 70°C can be compared. The basic features of the spectra are the same, but some fine differences between them can be observed. More distinct shoulder and peak of CH and CH_2 resonant lines, respectively, are observed in the spectrum measured on the METOCENE as compared with that measured on TATREN and so it is reasonable to assume that the amorphous phase in METOCENE is more mobile than in TATREN.

V. Conclusion

The NMR measurements on the iPP samples prepared by metallocene and Ziegler-Natta catalysis show the strong dependence of the ^{1}H and ^{13}C NMR spectra on the temperature. The strong line narrowing in the ^{1}H MAS NMR spectra and the change of the shape of the ^{13}C CP MAS NMR resonance lines related to the individual groups were observed with the rise of the temperature. The changes of the spectra are related to the conformation changes in the amorphous region of the polymer.

The ^{1}H and ^{13}C NMR spectra measured on both iPP polymers show that amorphous region of METOCENE is more mobile than that of TATRENE that was prepared by classical Ziegler-Natta catalysis.

It is reasonable to assume that the NMR experiments presented in this paper can be a good starting point for detailed study of the effect of the preparation procedure on the morphology of the investigated materials.

References

[1] M. Ratzsch, "*Special PP's for a Developing and Future Market.*", in *Journal of Macromolecular Science*, Part A: Pure Applied Chemistry, Vol. 36, 1999, pp. 1587 - 1611

[2] V. Busico, "*Microstructure of Polypropylene*", in *Progress in Polymer Science*, Vol. 26, Elsevier Science Ltd. 2001, pp. 443 - 533.

[3] F.A. Bovey, P.A. Mirau, *NMR of Polymers.* San Diego,California, USA: Academic Press 1996.

[4] R. Kitamaru, "*Phase Structure of Polyethylene and Other Crystalline Polymers by Solid-State 13C NMR.*", in *Advances in Polymers,* Vol. 137, Berlin Heidelberg: Springer - Verlag, 1998

[5] D. Olčák , "*Study of Motional Processes in Polymer Blends Composed of Isotactic Polypropylene and Ethylene – Propylene – Diene Monomer Rubber by Broad – Line ^{1}H – NMR.*", in *Journal of Applied Polymer Science*, Vol. 91, Wiley Periodicals 2003, pp. 247 - 252

[6] J. M. Gómez - Elvira, "*Relaxations and thermal stability of low molecular weight predominantly isotactic metallocene and Ziegler – Natta polypropylene.*", in *Polymer Degradation and Stability*, Vol. 85, Elsevier Ltd. 2004, pp. 873 - 882

[7] M. J. Duer, *Introduction to solid state NMR spectroscopy.* Oxford, UK: Blackwell, 2004, pp. 61 - 62

[8] L.Ševčovič, L. Mucha, "*Study of stretched polypropylene fribes by ^{1}H pulsed and CW NMR spectroscopy.*", in *Solid State Nuclear Magnetic Resonance*, Vol. 36, Elsevier Science 2009, pp. 151 – 157

Acknowledgement

We thank Prof. Ing. Ivan Chodák, DrSc. from the Polymer Institute of Slovak Academy of Sciences for providing the samples for NMR measurements and samples characteristics obtained by DSC and DMTA.

CFAR detectors for UWB radars: An Overview and Comparison

[1]*Daniel Urdzík*

[1]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]daniel.urdzik@tuke.sk

***Abstract*— In this paper different CFAR detectors for detection of multiple targets for UWB radar system will be described and their outputs compared. The cell averaging CFAR (CA-CFAR) cell averaging with greatest of (CAGO-CFAR), ordered statistics CFAR (OS-CFAR), clutter map CFAR (CM-CFAR) and cell averaging CFAR assumed for gamma distributed clutter will be represented. The properties of all detectors will be illustrated by real radar signal processing obtained by the measurement with the M-sequence UWB radar.**

***Keywords*— UWB radar, CFAR, order statistics, CM-CFAR**

I. INTRODUCTION

Ultra wideband (UWB) radars which operate in a lower GHz-range base-band (up to 5 GHz) produce results in their high spatial resolution, typically a few cm. This feature together with facts that UWB radar devices may be built small and light weight and that they employ low-power harmless electromagnetic waves which in the lower GHz range penetrate through most common building materials, is the reason, why UWB radars can be used for through wall tracking of moving targets [1]. These systems can be used in hazardous environments, where direct access is not possible or considered as hazardous, e.g. rescue operations or military operations. It has been shown in [2] that the trace estimation method can be used for through wall tracking of moving targets with an advantage.

This procedure consists of such phases of radar signal processing as: raw radar data pre-processing, background subtraction, detection, trace estimation, localization and tracking [2]. These phases of radar signal processing already have been analyzed e.g. in [3] - [8]. In this paper we will focus on the phase of target detection.

General problem of detection is to decide if a target is absent or present in examined radar signals. The most important groups of the detectors applied for radar signal processing are represented by sets of optimum or suboptimum detectors. Optimum detectors can be obtained as a result of solution of an optimization task formulated usually by means of probabilities or likelihood functions describing detection process. Here, Bayes criterion, maximum likelihood criterion or Neymann-Pearson criterion are often used as the bases for detector design. For the purpose of target detection by using UWB radars, detectors with fixed threshold [5], (N,k) detectors [5], IPCP detectors [5] and constant false alarm rate detectors (CFAR) [10] have been proposed. Between detectors capable to provide good and robust performance for through wall detection of moving targets by UWB radar, CFAR detectors can be especially assigned [2]. CFAR detectors provide adaptive estimation of an optimum threshold based on Neyman-Pearson criterion and under assumption, that the probability distribution function of the clutter is known. The problem of CFAR detector for UWB sensor networks have been analyzed for Gaussian distributed clutter (G-CFAR) in [10].

Based on the analyses of the clutter distribution for a great number of real radar signals obtained for through wall target tracking, we have found that the clutter distribution is do not follow a typical Gaussian distribution. In many situations where clutter distribution do not follow Gaussian distribution, a problem with target detection may occur when a value of computed threshold is too high or too low, which means that the target is either not detected or too many false alarms will be detected. In order to improve detection in UWB radar systems, we have focused on other CFAR detection methods that work under assumption of non-Gaussian clutter distribution, which in our case appears to be exponential distribution.

In this paper, cell averaging CFAR (CA-CFAR), cell averaging with greatest of (CAGO-CFAR) [9], cell averaging CFAR for gamma distribution (CAG-CFAR) [8], order statistics CFAR (OS-CFAR) [9] and clutter-map CFAR (CM-CFAR) can be given to improve the output of the detection phase of through wall tracking of moving targets by the trace estimation method. These CFAR detectors will be described presuming the exponential distribution and gamma distribution of the clutter probability distribution function. The performance of the proposed detection methods will be compared based on real radar signal processing obtained for through wall tracking of multiple targets. The obtained results will show that the described OS-CFAR is able to provide better detection of the targets to be tracked in comparison with that of G- CFAR, CA-CFAR, CAGO-CFAR, CAG-CFAR and CM-CFAR

II. CFAR DETECTION

The selection of particular CFAR detector is dependent on the clutter distribution and how the decision threshold is estimated in situations, where a target or multiple target echoes are present. Therefore in this section we will describe

basic principles of different CFAR detectors and how the final decision threshold is calculated.

The general scheme of the CA-CFAR detector is described in the Fig.1, where a sliding range window is used to analyze clutter power level in the vicinity of the test cell.

A. CA-CFAR and CAGO-CFAR

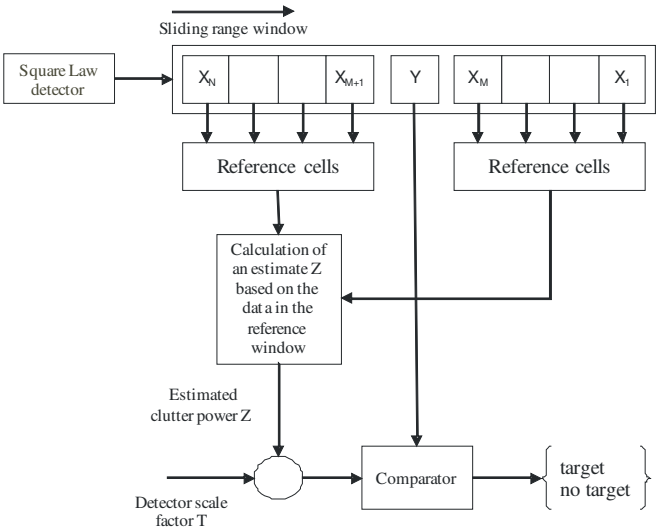


Fig.1 Scheme of CA-CFAR detector

Detection methods analyze the radar signal and reach the decision whether a signal scattered from target is absent (hypothesis H_0) or it is present (hypothesis H_1) in the examined radar signals. A detector discriminates between hypotheses H_0 and H_1 based on comparison testing (decision) statistics X and a threshold S. Then, the output of detector $D(Y)$ is given by:

$$D(Y)=\begin{cases}1, & X \geq S\\ 0, & X < S\end{cases} \tag{1}$$

In this paper it is assumed, that the background clutter can be described by a statistical model where range cells inside the sliding window, with length of N, contains statistically independent identically exponentially distributed random variables X_1, X_2...X_N. The probability density function of exponentially distributed clutter variables is given by the equation [9]:

$$p(x)=\begin{cases}(1/\mu)e^{-\frac{x}{\mu}}, & x \geq 0\\ 0, & otherwise\end{cases} \tag{2}$$

Then the probability of false alarm (P_{fa}) can be expressed:

$$P_{fa}=\int_{S}^{\infty} p(x)dx\,. \tag{3}$$

In the case of CA-CFAR [9], the sliding window is used to analyze local clutter power situations near the test cell. The mean power of clutter is estimated on the left and right side of the test cell:

$$Z=\frac{1}{N}\sum_{i=1}^{N} X_i\,. \tag{4}$$

When multiplied by a predetermined constant T, S provides an adaptive threshold to maintain constant false alarm rate:

$$S=T.Z \tag{5}$$

The scaling constant T of the threshold S is described for a given false alarm probability analytically by the expression :

$$T=\frac{1}{\ln(P_{fa})} \tag{6}$$

The CA-CFAR provides reliable adaptive threshold under assumption, that the clutter inside the sliding window is homogenous. This assumption is not true in many cases, so other modification of CA-CFAR was used to improve detection.

In the case of CAGO-CFAR [9], the average power of clutter Z is estimated differently where both sides of the sliding window are analyzed separately. The value of Z for CAGO-CFAR is estimated as:

$$Z=\max\left(\frac{2}{N}\left[\sum_{i=1}^{N/2} X_i\right];\frac{2}{N}\left[\sum_{i=(N/2)+1}^{N} X_i\right]\right). \tag{7}$$

B. CAG-CFAR

Another approach has been described by Mahafza et.al. [8]. In this case the structure of proposed CAG-CFAR is similar to that of in Fig.1. In this case the threshold is estimated differently where only a sum of all the cells of sliding window is taken into the account:

$$Z=\sum_{i=1}^{N} X_i\,, \tag{8}$$

where it is assumed that the variables follow the gamma distribution:

$$p(z)=\frac{z^{(N/2)-1}e^{(-z/2\psi^2)}}{2^{N/2}\psi^{M}\Gamma(N/2)},\quad z>0\,. \tag{9}$$

The adaptive threshold for CAG-CFAR is computed according (5), where the scaling factor T is calculated as follows:

$$T=\left(\frac{1}{P_{fa}}\right)^{\frac{1}{N}}-1\,. \tag{10}$$

The described CA-CFAR, CAGO-CFAR and CAG-CFAR give reliable results in scenarios, where only one target is present. In the case of multiple target scenarios, the threshold for the targets, that usually have smaller magnitude is not calculated correctly [10]. The other CFAR detectors had to be taken into consideration to provide reliable results in multiple target situations such as OS-CFAR.

C. OS-CFAR

The general scheme of the OS-CFAR detector could be described by sketch in Fig.1. The difference between CA-CFAR and OS-CFAR is in the estimation of clutter power from the cells of the sliding window. In case of OS-CFAR, the values of the signal samples in the sliding window (Fig.1) are sorted by the size of their magnitude [10]:

$$X_{(1)} \leq X_{(2)} \leq \ldots\ldots \leq X_{(N)} \tag{11}$$

The estimation of mean clutter power could be described as the selection of a single rank $X_{(k)}$ from the ordered statistics instead of arithmetic mean:

$$Z=X_{(k)}\,. \tag{12}$$

Assuming exponentially distributed clutter in the reference window for OS-CFAR, probability of false alarm can be

calculated as:

$$P_{fa} = k\binom{N}{k}\frac{(k-1)!(T+N-k)!}{(T+N)!}. \quad (13)$$

For practical applications only a few values of k are of interest. Experimental results showed, that the value $k=(3N/4)$ is a robust parameter. The different values of scaling factor T for given N and P_{fa} can be obtained by solving the equation, which can be derived from (13) by substituting $k=3N/4$:

$$\prod_{i=N}^{(3N/4)+1}(T+i) - \frac{N!}{P_{fa}(\frac{N}{4})!} = 0. \quad (14)$$

D. CM-CFAR

A clutter map divides the radar coverage area into cells on polar or rectangular grid. The clutter echo stored in each cell of the map can be used to establish threshold, it is then a form of CFAR. The size of each clutter map cell is equal or greater than radar resolution [12].

The structure of the CM-CFAR is shown in Fig.2, where clutter magnitude is estimated as:

$$\begin{aligned} Y(0) &= X(0) \\ Y(m) &= (1-a)Y(m-1) + aX(m) \end{aligned}, \quad (15)$$

where $Y(m)$ is estimated clutter value after m scans, a is a clutter map gain and $X(m)$ is clutter map update value.

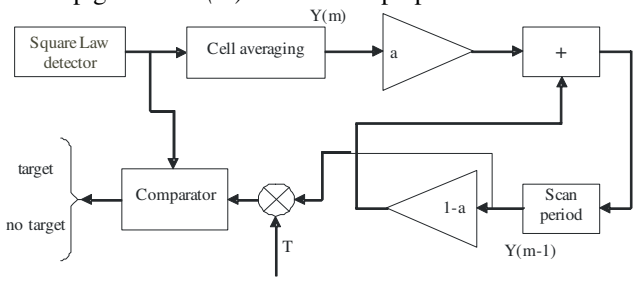


Fig.2 Scheme of CM-CFAR

Assuming that there are M resolution cells ($x(1,j)….x(M,j)$) in clutter map cell their average clutter level is calculated as:

$$X(m) = 1/M\sum_{i=1}^{M} x(j,m), \quad (16)$$

where value of j denotes order of the resolution cell in the clutter map cell. We can rewrite (15) into the general form as:

$$Y(m) = \sum_{n=1}^{m} h(n-m)X(n), \quad (17)$$

where

$$h(n) = \begin{cases} a(1-a)^n & 0 \le n \le m\text{-}1 \\ (1\text{-}a)^m & n = m \end{cases}. \quad (18)$$

For a given P_{fa} the scaling factor T can be obtained by solving the equation:

$$P_{fa} = \left(\frac{1}{(1+T(1-a)^m/M)}\right)^M \cdot \prod_{n=1}^{m}\left(\frac{1}{1+Ta(1-a)^{m-n}/M}\right)^M, \quad (19)$$

and the detection decision in case of CM-CFAR can be expressed as:

$$D(Y) = \begin{cases} 1, & x(m) \ge T \cdot X(m\text{-}1) \\ 0, & x(m) < T \cdot X(m\text{-}1) \end{cases}, \quad (20)$$

where value T is a factor for selection of desired constant false alarm rate and the value of $x(m)$ is tested radar echo similar to variable Y in Fig.1.

III. Experimental Results

The performance of the described detectors are demonstrated by processing of the real radar signals acquired by the experimental M-sequence UWB radar. The system clock frequency of the radar device is about 4.5 GHz, which results in the operational bandwidth of about DC – 2.25 GHz. The M-sequence order emitted by radar is 9, i.e. the impulse response covers 511 samples regularly spread over 114 ns. This corresponds to an observation window of 114 ns leading to an unambiguous range of about 16 m. Within the analyzed measurement scenario two persons were moving between tables in a classroom behind 24 cm thick concrete wall.

The analyzed scenario is represented by the radargram shown in Fig.3. In this scenario two targets are present one outlined by blue-dotted line (hereinafter “Second target”) and other outlined by red-dotted line (hereinafter “first targe”).

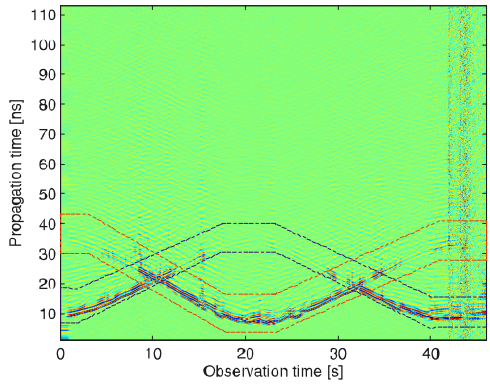


Fig. 3 Radargram with subtracted background

The output from G-CFAR detector is shown in Fig.4. Here, some parts of the second target are not visible, but also some parts of the first target have not been detected. The output of CA-CFAR is shown in Fig.5 and output of the CAGO-CFAR is shown in Fig. 6.

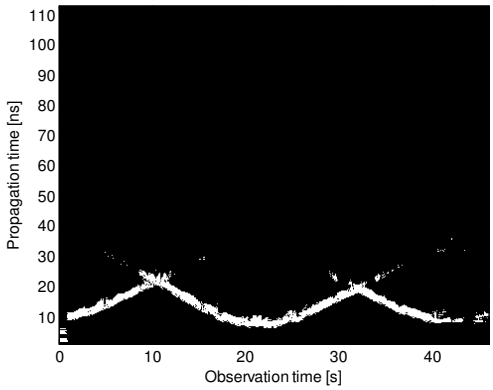


Fig.4 G-CFAR detector output

The OS-CFAR detector output is shown in Fig.7. The visual inspection shows, that in the case of OS-CFAR detector output, more parts of traces of second target traces are present than in G-CFAR, CA-CFAR and CAGO-CFAR. The outputs of CAG-CFAR and CM-CFAR are shown in Fig.8 and Fig.9. The comparison of all of described CFAR detectors show that the OS-CFAR appears to be superior, because in the areas outlined by red-dotted and blue-dotted line (Fig. 3), the

detector detected most of the parts of the targets.

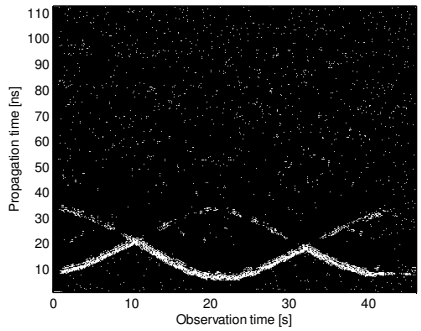


Fig. 5 CA-CFAR detector output

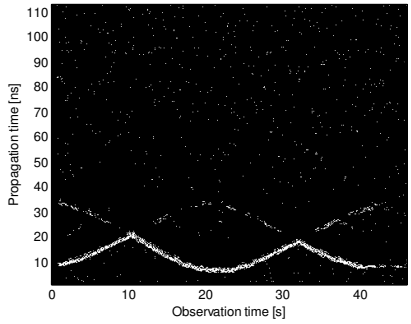


Fig. 6 CAGO-CFAR detector output

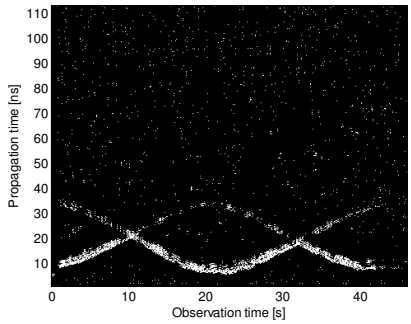


Fig. 7 OS-CFAR detector output

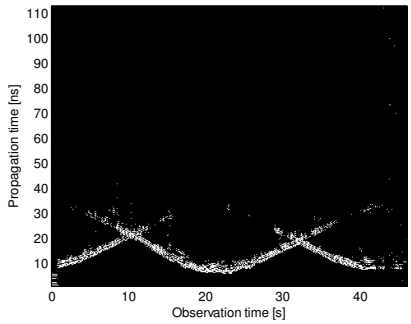


Fig. 8 CAG-CFAR detector output

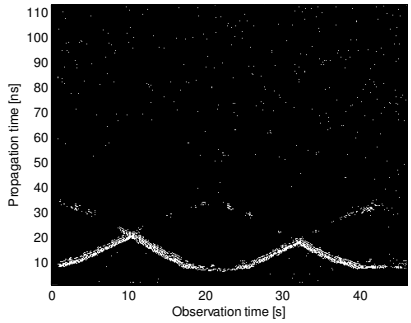


Fig. 9 CM-CFAR detector output

IV. Conclusion

This paper has been devoted to the problem of through wall detection of moving target by M-sequence UWB radar system. Here, some of the modifications of the different CFAR detectors have been described. The main advantage of proposed CFAR detectors is that they work under assumption, that clutter has different distributions and is not homogenous in scanned areas. The obtained experimental results approved the theoretical expectations of superiority of OS-CFAR in multiple target detection in UWB radar signal processing.

V. Acknowledgement

This work was supported by the Slovak Scientific Grant Agency (VEGA) under contract No. 1/0045/10. This work is also the result of the project implementation of the Center of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research & Development Operational Program funded by the ERDF.

References

[1] IMMOREEV, I. J Ultrawideband systems - features and ways of development. in *Ultrawideband and Ultrashort Impulse Signals, Second International Workshop*, Sevastopol, Ukraine, Sept. 2004, pp. 37–41.

[2] ROVNAKOVA, J: Complete signal processing for through wall target tracking by M-sequence UWB radar system, Dissertation Thesis, Kosice 2009.

[3] YELF, R., Where is true time zero?. in *Proceedings of the Tenth International Conference on Ground Penetrating Radar*, June 2004, pp.279–282.

[4] PICCARDI, M., Background subtraction techniques: a review. In *Proceedings of International Conference on Systems, Man and Cybernetics*. The Hague, The Netherlands, October 2004.

[5] TAYLOR, J. D. Ultrawideband Radar Technology. CRC Press, 2000.

[6] STANFORD, D., RAFTERY, A. E. *Principal Curve Clustering with Noise*. Tech. Rep. 317, Dep. of Statistics, University of Washington, February 1997.

[7] YU, K et.al.. *Ultra-Wideband Wireless Communications and Networks.* John Wiley & Sons, Ltd, Febr. 2006

[8] MAHAFZA, B. R. *Radar Systems Analysis and Design Using Matlab.* Chapman & Hall/CRC, 2000.

[9] ROHLING, H., Radar CFAR Thresholding in Clutter and Multiple Target Situations, *IEEE Transactions on Aerospace and Electronics Sysems,* vol. 19, No. 4, 1983, pp. 608-621.

[10] DUTTA, P. K., ARORA, A. K., BIBYK, S. B. Towards radar-enabled sensor networks. In *Proceedings of the Fifth International Conference on Information Processing in Sensor Networks. Special Track on Platform Tools and Design Methods for Network Embedded Sensors.* Nashville, Tennessee, USA, April 2006, pp. 467 - 474.

Fluctuations in electric circuits and the Brownian motion of particles

Gabriela VASZIOVÁ, Vladimír LISÝ

Dept. of Physics, FEI TU of Košice, Slovak Republic

gabriela.vasziova@tuke.sk, vladimir.lisy@tuke.sk

***Abstract*—In this contribution we explore the mathematical correspondence between the Langevin equation that describes the motion of a Brownian particle (or the noisy oscillator) and the equations for the time evolution of the charge in electric circuits, which are in contact with the thermal bath. The mean quadrate of the fluctuating electric charge and the mean square displacement of the Brownian particle are governed by the same equations that have been derived in the statistical mechanics for stochastic systems. We construct and solve these equations using an efficient approach that allows converting the stochastic equations to ordinary differential equations. From the obtained solutions, the autocorrelation function of the current and the spectral density of the current fluctuations are found.**

***Keywords*—Brownian motion, electric circuits, thermal fluctuations.**

I. INTRODUCTION

The mathematical correspondence between mechanical and electrical properties is often used to construct an electrical model of a given mechanical system [1]. This is a very useful way to predict the performance of a mechanical system, since the electrical elements are inexpensive and the measurements are usually very accurate. Such "analog computation" has been recently used also for stochastic systems in connection with the applicability of thermodynamic laws on nanoscales [2, 3] and with the so called fluctuation theorems that in the last decade attract a lot of attention not only in the statistical and condensed matter physics but also in the very different fields of science from nanotechnology to biology [4 - 6]. In electric circuits, the fluctuations have long been considered a nuisance - already the seminal works by Johnson and Nyquist on noise caused by thermal agitation of charge carriers were inspired by the problem of noise in telephone wires [7, 8]. On the other hand, when the studied system produces a frequency dependent (colored) noise, such noise contains information on the system. In electric circuits the information on the properties of the system is obtained from the measurements of the spectral density of the fluctuations, usually those of the current. The analogy with the noisy oscillator or the Brownian motion (BM) of particles can be very useful in the calculations of these fluctuations and interpretation of the measurements in circuits and *vice versa*.

In the present work we explore the analogy between the motion of a Brownian particle (BP) dragged by a moving harmonic potential and simple electric circuits in contact with the thermal bath, described by exactly the same equations. Using the methods of statistical physics we calculate the mean square displacement (MSD) of the BP. To do this, we use a method due to Vladimirsky [9] that allows one to convert the stochastic equations of the Langevin type to ordinary differential equations, which are much easier to solve. To our knowledge, the used method has not been applied to similar problems so far; the only exception is an old little known work [10] on the hydrodynamic BM. The efficiency of this method is immediately seen, especially in context of solving the generalized Langevin equations that are often used to describe various problems of anomalous BM [11]. Having found the MSD of the BP (which in electric circuits corresponds to the mean quadrate of the electric charge), it is then easy to evaluate the BP velocity autocorrelation function (VAF) (corresponding to the autocorrelation function for the electric current). From these functions we calculate the spectral density of the fluctuations, e.g., the spectrum of the colored noise produced by the circuits.

II. THE DRAGGED BROWNIAN PARTICLE AND SIMPLE ELECTRIC CIRCUITS

In the experiments [5] small (about 3 micrometers in radius) latex BP were dragged through water by a moving optical tweezer. This means that the BP was subject to an external harmonic potential with a time dependent position x_t^* of its minimum. For $t \le 0$ this minimum is at the origin, $x_t^* = 0$, whereas for $t > 0$ it moves with a constant velocity v^* (Fig. 1). Such a motion of the BP can be described by the Langevin equation

$$m\frac{d^2x_t}{dt^2} = -\alpha\frac{dx_t}{dt} + \xi - k\left(x_t - v^* t\right), \tag{1}$$

where x_t is the position of the BP at the time t, m is its mass, α is the Stokes friction coefficient, and k is the strength of the harmonic potential induced by the optical tweezer. The force ξ is the thermal white noise due to the kicks of the surrounding molecules. It has the zero mean and the property $\langle\xi(t)\xi(t')\rangle = 2k_BT\alpha\delta(t - t')$, where k_B is the Boltzmann's constant, T the temperature, δ is the Dirac delta function, and the brackets $\langle ... \rangle$ denote the statistical averaging. Such a model of a particle dragged by a spring through a thermal environment was

studied earlier, before the experiments [6], in the work [12]. Both in [12] and the later works [4, 13] the simplified equation (1) with m = 0 has been considered. That is, the overdamped motion of the BP was studied, assuming that the velocity relaxes quickly. This exactly solvable model was used to illustrate the fluctuation theorems and other predictions for systems evolving far from equilibrium.

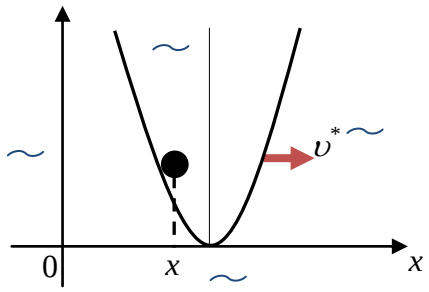


Fig. 1. Brownian particle dragged by a harmonic potential with a constant velocity.

Now, let us consider the electric circuit (Fig. 2), in which the resistor with the resistance R, the capacitor with the capacitance C, and the inductor with the inductance L are connected in series. They are subject to a voltage source $V(t)$. There is also a thermal noise generator next to the resistance, which reflects the fluctuations of the voltage drop across the resistor, $\delta V(t)$. The imposed voltage linearly increases with the time t, $V(t) = \kappa t$.

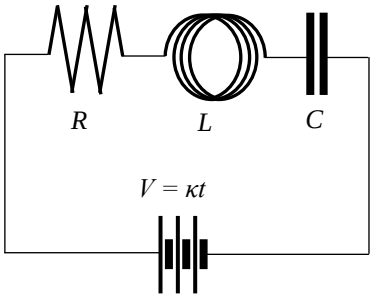


Fig. 2. Electric circuit with the inductor, resistor and capacitor in series. The voltage linearly increases with the time and the system is in contact with the thermal bath.

The equation for such a circuit

$$L\ddot{Q} + R\dot{Q} + \frac{1}{C}Q = \kappa t + \delta V(t)\,. \tag{1a}$$

exactly corresponds to (1) if the particle displacement x_t is replaced by the charge Q, L replaces the particle mass m, R and $1/C$ are for the friction coefficient and the elastic constant k, respectively, and the velocity of the harmonic well is replaced by the constant $C\kappa$. In [4, 6] another example of the circuit is given (Fig. 3), when the time evolution of the charge is described by essentially the same equation. In this electric circuit a resistor and an inductor are arranged with a capacitor in parallel and are subject to a constant, non fluctuating current source I. Energy is being dissipated in the resistor, which, according to the fluctuation-dissipation theorem means that there are fluctuations too. These fluctuations are described by a random noise term δV, which could be described by a voltage generator. The difference between (1a) and the equation describing the circuit in Fig. 3 is only in the term κt, which is now replaced by It/C. Thus the current I corresponds to the velocity υ^*. In what follows we continue to use the symbols for the BP as in (1), having in mind the above mentioned correspondence with the parameters of electric circuits.

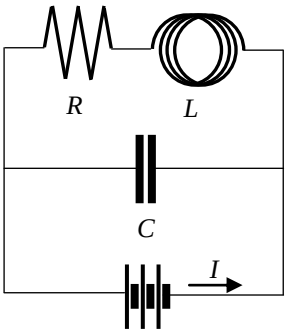


Fig. 3. An electric circuit with a serial inductor-resistor coupled to a capacitor in parallel.

Equation (1) can be solved by considering the motion of the particle with respect to the minimum of the harmonic potential. For our purposes (it will become clear below), more suitable is to separate the average motion of the BP (which results from the deterministic forces alone), from the stochastic motion. This can be interpreted as going to a comoving frame, which is what the motion of the particle would be if there would be no noise in the Langevin equations (1), (1a). The deterministic equation is (we consider a more general case than in [4, 12, 13]),

$$m\ddot{x}^* = -\alpha\dot{x}^* - k\left(x^* - \upsilon^* t\right). \tag{2}$$

Subtracting (1) and (2), we obtain for $x = x_t - x^*$

$$m\ddot{x} = -\alpha\dot{x} + \xi - kx\,. \tag{3}$$

Now we can apply a very efficient (but little known) rule due to Vladimirsky [9], which allows us to rewrite (3) for the MSD of the BP, $X(t) = \langle \Delta x^2(t) \rangle = \langle [x(t) - x(0)]^2 \rangle$, as follows:

$$m\ddot{X} + \alpha\dot{X} + kX = 2k_B T\,. \tag{4}$$

Thus we have an ordinary differential equation, which is much easier to solve than the original stochastic equation (1). According to the Vladimirsky's rule we have merely to substitute $x(t)$ in (3) by $\langle \Delta x^2(t) \rangle$ and replace the stochastic force driving the particle with the constant "force" $f = 2k_B T$. Of course, having the dimension of energy, f is not the true force; it only plays this role in the equation of motion for the particle "position" $X(t)$. This fictitious force begins to act on the particle at the time t = 0. Up to this moment the particle is nonmoving so that the equation of motion must be solved with the initial conditions $X(0) = V(0) = 0$, where we have introduced the "velocity" $V(t) = dX(t)/dt$. Using this rule, it is easy to obtain the MSD of the free BP, when $\upsilon^* = 0$, $k = 0$:

$$X(t) = \frac{2k_B T}{\alpha}\left\{t + \frac{m}{\alpha}\left[\exp\left(-\frac{\alpha}{m}t\right) - 1\right]\right\}. \tag{5}$$

However, the behavior of the dragged particle is very different. When $\upsilon^* \neq 0$ but $m = 0$ as in the cited works, the solution of (4) is

$$X(t) \approx \frac{2k_B T}{k}\left[1-\exp\left(-\frac{k}{\alpha}t\right)\right], \tag{6}$$

whereas the deterministic equation (2) has the solution

$$x^*(t) \approx \upsilon^*\left[t-\frac{\alpha}{k}+\frac{\alpha}{k}\exp\left(-\frac{k}{\alpha}t\right)\right] \tag{7}$$

if the harmonic potential is set in motion at $t = 0$ when the particle is at rest, $x^* = 0$. Note that one of the initial conditions for $X(t)$ now cannot be satisfied since the simplified equation (4) with $m = 0$ determines the value $\dot{X}(0) = 2k_B T/\alpha$ instead of 0. The more correct approach requires the inertial effects to take into account. This will be the subject of our further work. We believe that these effects, which are certainly important at short times, are important also in order to prove the validity of the fluctuation theorems for all times (in fact, the simplification $m = 0$ in [4, 12, 13] implies that only the long times are considered).

In the present work we shall not discuss the validity of the fluctuation theorems but, having in mind the applicability of the theory to the study of noise in electric circuits, we shall turn our attention to the fluctuation spectra in the circuits. To do this, we first calculate the total MSD of the BP, which in our case is

$$X_t(t) = \left\langle\left[x_t(t)-x_t(0)\right]^2\right\rangle = \left[x^*(t)\right]^2 + X(t), \tag{8}$$

with X and x^* from (5) or (see Fig. 4) (6) and (7).

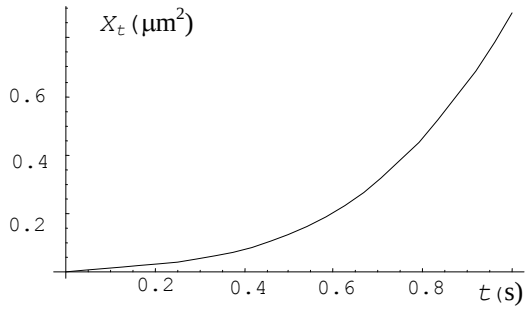


Fig. 4. Mean square displacement of the dragged Brownian particle in experiments [5] according to equations (6 - 8). The parameters are k = 0.1 pN/μm, $\alpha = 6\cdot10^{-8}$ kg/s, $\upsilon^* = 1.25\cdot10^{-6}$ m/s, and T = 290 K.

The spectral density of fluctuation spectrum is, according to the Wiener-Khinchin theorem, equal to the Fourier transform of the autocorrelation function of the quantity of interest [14]. So, for the fluctuation spectrum of the current I we have

$$S_I(\omega) = \frac{2}{\pi}\int_0^\infty dt\left\langle I(0)I(t)\right\rangle \cos\omega t\,, \tag{9}$$

which is the quantity usually measured [14]. In the BM the quantity corresponding to the current is the particle velocity, so that we have to calculate the spectrum of the VAF $\Phi(t) = \langle\upsilon_t(t)\upsilon_t(0)\rangle$. It contains only the part determined by the stochastic solution of (3), $\upsilon(t) = \dot{x}(t)$. It is seen by considering the autocorrelation function for $\dot{x}_t = \dot{x} + \dot{x}^*$ with respect to the laboratory frame, and taking into account the initial conditions for x^*. The correlator $\Phi(t)$ is determined through $X(t)$ as

$$\Phi(t) = \ddot{X}(t)/2 \tag{10}$$

(in the theory of the BM $\Phi(t)$ is called the time dependent diffusion coefficient). Thus, for the BP

$$S_\upsilon(\omega) = \frac{1}{\pi}\int_0^\infty \frac{d^2X}{dt^2}\cos(\omega t)\,dt\,. \tag{9a}$$

Now we can simply use the solution (5) and evaluate the spectrum from (9). If $k = 0$ but $m \neq 0$, we get (Fig. 5)

$$S_\upsilon(\omega) = \frac{2k_B T}{\pi}\alpha\left[\alpha^2 + (m\omega)^2\right]^{-1}, \tag{11}$$

which at $m \to 0$ has the limit $S_\upsilon(\omega) = 2k_B T/(\pi\alpha)$. If we assume from the beginning that $m = 0$, (6) yields a significantly different result

$$S_\upsilon(\omega) = -\frac{2k_B T}{\pi}\frac{k^2}{\alpha}\left[(k)^2 + (\alpha\omega)^2\right]^{-1}, \tag{12}$$

despite the fact that the limit of $X(t)$ (calculated at $m = 0$, $k \neq 0$) at $k \to 0$ is the same as the limit for $X(t)$ (calculated at $m \neq 0$, $k = 0$) at $m \to 0$, i.e. $X(t) = 2\,k_B T t/\alpha$.

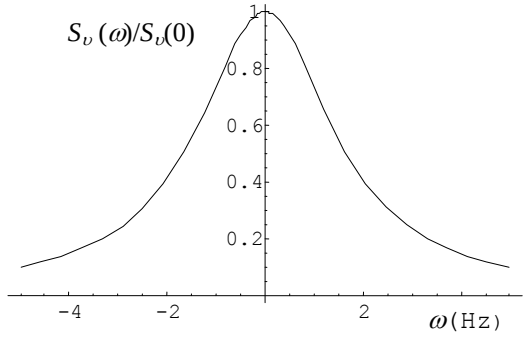


Fig. 5. Spectral density of fluctuations corresponding to Fig. 4.

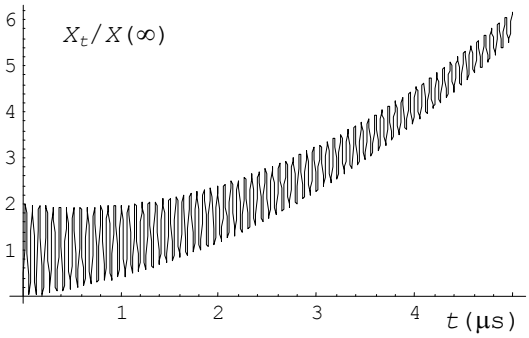


Fig. 6. Normalized mean square displacement for a dragged Brownian particle in a strong harmonic potential from (2), (4), and (8) with $m = 1.25\cdot10^{-13}$ kg. This and other parameters are taken from the experiment [5] except for a much larger elastic constant (k = 10 N/m).

The reason for the difference between (11) and (12) comes from a distinct behavior of $X(t)$ in the time. So, if $t \to 0$, $X(t)$ approaches 0 as $X(t) \approx 2k_B T t/\alpha$ ($m = 0$, $k \neq 0$), whereas $X(t) \approx k_B T t^2/m$ ($m \neq 0$, $k = 0$), which is the correct result for the BP. Due to the different time behavior the Fourier spectrum is different, too. This confirms our discussion on the fact that to obtain the correct solution we cannot assume from the beginning that $m = 0$. Assuming $m = 0$, one cannot obtain

correct results for the stochastic process $x(t)$ at the times $t \rightarrow 0$. An illustrative result for the MSD at $m \neq 0$ and large k, when the system exhibits oscillations, is shown in Fig. 6.

III. Conclusion

In conclusion, we have explored the correspondence between the motion of Brownian particles and the fluctuations in the electric circuits, which are in contact with the thermal bath. Using the established mathematical equivalence of the description of motion of the BP dragged by the optical tweezer and two simple electric circuits, we have calculated the time dependence of the mean quadrate of the charge in the circuit, the autocorrelation function of the fluctuating current, and the spectral density of the fluctuations. This was done using the method developed in the statistical physics and especially suitable for the calculation of the MSD of the BP, its VAF or the time dependent diffusion coefficient, which in the circuits exactly corresponds to the autocorrelation function of the current, $\langle I(t)I(0)\rangle$. Up to this moment our analytical calculations were limited to the simplest cases when the mass of the particle (the inductance in the circuits) or when the strength of the harmonic potential induced by the optical tweezer (the inverse capacitance in the circuits) are neglected. We have shown that the more general calculations, without these limitations, are necessary. For the future work a number of other questions arises. For example, when the BP moves in a liquid as in the experiments [5], for the correct description of the particle behavior (especially at short times) the inertial effects during the motion must be taken into account. This means that not only the particle mass should be nonzero, but also the memory effects play a role [15] (the state of the particle motion at the time t depends on the particle velocities and accelerations in the preceding moments of time). Mathematically it displays in the generalization of the Langevin equation, which becomes an integro-differential equation. A similar equation has been derived also for nanoscale electric circuits [2]. The statistical-mechanical approach used in the present contribution is applicable to such circuits as well, at least in the classical limit. Currently, the work in these directions is in progress.

Acknowledgment

This work was supported by the Agency for the Structural Funds of the EU within the project NFP 26220120021, and by the grant VEGA 1/0300/09.

References

[1] K. T. Tang, Vector Analysis, Ordinary Differential Equations and Laplace Transforms. Mathematical Methods for Engineers and Scientists, Part II. Berlin, Heidelberg: Springer-Verlag, 2007.

[2] Th. M. Nieuwenhuizen, A. E. Allahverdyan, "Statistical thermodynamics of quantum Brownian motion: Construction of perpetuum mobile of the second kind". *Phys. Rev. E* 66, 036102 (2002).

[3] A. E. Allahverdyan, Th. M. Nieuwenhuizen, "On testing the violation of the Clausius inequality in nanoscale electric circuits". *arXiv:cond-mat*/0205156v1, *Phys. Rev. B* 66, 115309 (2002) (see also comment on this article in: E. P. Gyftopoulos, M. R. von Spakovsky, "Comments on testing the violation of the Clausius inequality in nanoscale electric circuits". *arXiv* 0706.2842v1 [quant-ph]).

[4] R. van Zon, S. Ciliberto, and E. G. D. Cohen, "Power and heat fluctuation theorems for electric circuits". *arXiv:cond-mat*/0311629v2 (2004), *Phys. Rev. Lett*., 92, 130601 (2004).

[5] G. M. Wang, E. M. Sevick, E. Mittag, D. J. Searles, and D. J. Evans, "Experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales". *Phys. Rev. Lett*. 89, 050601 (2002).

[6] T. Taniguchi, E. G. D. Cohen, "Nonequilibrium steady state thermodynamics and fluctuations for stochastic systems". *J. Stat. Phys*. 130, 633-677 (2008).

[7] J. B. Johnson, "Thermal agitation of electricity in conductors". *Nature*, 119, 50–51 (1927); *Phys. Rev*. 32, 97-109 (1928).

[8] H. Nyquist, Phys. Rev. 29, 614 (1927); "Thermal agitation of electric charge in conductors". *Phys. Rev*. 32, 110-113 (1928).

[9] V. V. Vladimirsky, "To the question of the evaluation of mean products of two quantities, related to different moments of time, in statistical mechanics". *Zhur. Ekper. Teor. Fiz*. 12, 199-202 (1942) (in Russian).

[10] V. Vladimirsky, Ya. Terletzky, "Hydrodynamical theory of translational Brownian motion". *Zhur. Eksp. Teor. Fiz*. 15, 258-263 (1945) (in Russian).

[11] R. Klages, R. Günter, I. M. Sokolov. Eds., Anomalous Transport. Berlin: Wiley - VCH, 2008.

[12] O. Mazonka, C. Jarzynski, "Exactly solvable model illustrating far-from-equilibrium predictions". *arXiv:cond-mat*/991212v1.

[13] R. van Zon and E. G. D. Cohen, "Stationary and transient work-fluctuation theorems for a dragged Brownian particle". *Phys. Rev. E* 67, 046102 (2003).

[14] N. G. van Kampen, "Fluctuations in nonlinear systems". Chapter 5 of Fluctuation Phenomena in Solids. Edited by R. E. Burgess. New York: Academic Press, 1965, pp. 139–177.

[15] V. Lisy, J. Tothova, "On the (hydrodynamic) memory in the theory of Brownian motion". *arXiv:cond-mat*/0410222 (12 pp.).

Artificial Neural Network in Mechatronic System Control via Internet

Tibor VINCE

Dept. of Theoretical Electrical Engineering and Electrical Measurement, FEI TU of Košice, Slovak Republic

tibor.vince@tuke.sk

***Abstract*—The article presents regulation possibilities of mechatronic system via Internet and possible improvements of such control using Artificial Neural Network. Due to the development of Internet technique and speed increase of transmission, the inexpensive convenient communication approach is provided for the remote control system The paper also handles the advantages and disadvantages of Internet as a control and communication bus at different levels of the information hierarchy.**

***Keywords*—Remote control, Internet, Artificial Neural Network, Mechatronic system, Information architecture**

I. INTRODUCTION

There is huge effort to integrate different cooperating systems in one complex system. The basic problem is communication between these different modules of the system, especially when the modules are located in different locations. According to communication requirements, appropriate communication way has to be chosen.

Continual evolution of the Internet enables higher and higher communication requirements to be fulfilled. The Internet begins to play a very important role in industrial processes manipulation, not only in information retrieving. With the progress of the Internet it is possible to control and regulate remote system from anywhere around the world at any time. Distance remote via Internet, or in other words, Internet-based control.has attracted much attention in recent years

Such type of control bus allows remote monitoring or regulation of whole plants or single devices over the Internet. The design process for the Internet-based control systems includes requirement specification, architecture design, control algorithm, interface design and possibly safety analysis. Due to the low price and robustness resulting from its wide acceptance and deployment, Ethernet has become an attractive candidate for real-time control networks.

It is necessary to regulate mechatronic system in such remote regulation in some cases. The goal of the article is to explore existing possibilities for Internet based real-time regulation, eventual trends, review of advantages and disadvantages of distance remote via Internet at different levels of information hierarchy and possible solutions. The article presents regulation of mechatronic system as an example of such a real-time regulation and discusses possibilities of utilization Artificial Neural Network (as part of Artificial Intelligence).

II. ARTIFICIAL NEURAL NETWORK

An artificial neural network (ANN), often just called a "neural network" (NN), is a mathematical model or computational model based on biological neural networks. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase.

Since the early 1990's, there has been a growing interest in using artificial neural networks for control of nonlinear systems. Numerous applications have demonstrated that neural networks are indeed powerful tools for the design of controllers for complex nonlinear systems. Among different kinds of neural networks, the most widely used ones are multilayer neural networks and recurrent networks. In case of Internet-based Control is very important some kind of auto-adaptation.

By solving tasks in field of electric drives we meet following basic problems: system simulations, identification of system parameters, system state quantities monitoring, drive regulations and malfunction diagnostic. There is possible successful utilization of neural network in all these fields of problem. The most important neural network attributes in this field are: various nonlinear functions approximation, parameters settings based on experimental or learning data, data processing and robustness.

Two different models are used in identification models creation: mathematics and physics analysis and experimental identification. In case of complex subjects both methods are required. Neural network can be used as direct neural model connected parallelly or serial-parallelly in learning state. Neural network can be used also as inverse identification model in dynamic system. Today's computer science performance allows to replace classical methods of parameters estimation by automatic identification. Main advantages include complex test signals generation possibilities, sophisticated identification algorithms, on-line identification possibilities etc.

The condition of the Internet is a very varying parameter and the control system controlling via Internet has to compensate the variation. One of the solutions is to employ

the neural network. It is possible to teach neural network behaviour for different conditions of networks. The advantage of this solution is that neural network is a more universal tool and condition of the Internet can be used as a one of the many parameters that relate with controlling and regulation.

III. Information Architecture

As mentioned before, there is effort to integrate different subsystems in one complex system. Integration of information and control across the entire plant site becomes more and more significant. In the manufacturing industries this is often referred to as "Computer Integrated Manufacturing" (CIM). There is increasing use of microprocessor-based plant level devices such as programmable controllers, distributed digital control systems, smart analyzers etc. Most of these devices have "RS232" connectors, which enable connection to computers. If we began to hook all these RS232 ports together, there would soon be an unmanageable mess of wiring, custom software and little or no communication. This problem solution results in integration these devices into a meaningful "Information Architecture". This Information Architecture can be separated into 4 levels with the sensor/actuator level as shown in Fig. 1, which are distinguished from each other by“4Rs” principle criteria: [1]

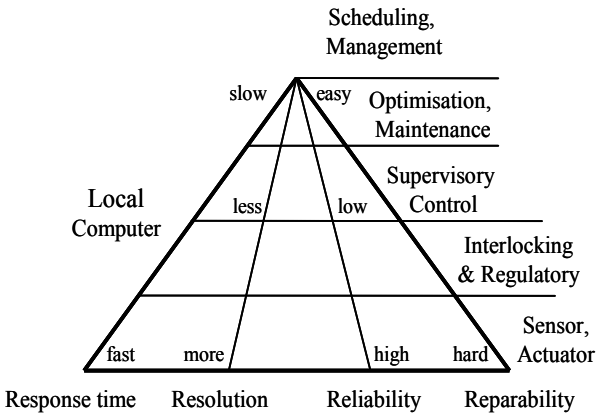


Fig 1. Information Architecture

The 4Rs criteria are: Response time, Resolution, Reliability and Reparability.

Response time: as one moves higher in the information architecture, the time delay, which can be tolerated in receiving the data, increases. Conversely, information used at the management & scheduling level can be several days old without impacting its usefulness.

Resolution: an Abstraction level for data varies among all the levels in the architecture. The higher the level is, the more abstract the data is.

Reliability: Just as communication response time must decrease as one descends through the levels of the information architecture, the required level of reliability increases. For instance, host computers at the management & scheduling level can safely be shut down for hours or even days, with relatively minor consequences. If the network, which connects controllers at the supervisory control level and/or the regulatory control level, fails for a few minutes, a plant shutdown may be necessary.

Reparability: The reparability considers the ease with which control and computing devices can be maintained.

Local computer on supervisory control level is able communicate with higher levels of information architecture via Internet, but there is also possibility to use the Internet also in lower levels of the Information architecture. The Internet can be linked with the local computer system at any level in the information architecture, or even at the sensor/actuator level. These links result in a range of 4Rs (response time, resolution, reliability, and reparability). For example, if a fast response time is required a link to the control loop level should be made. If only abstracted information is needed the Internet should be linked with a higher level in the information architecture such as the management level or the optimization level.

IV. Network Performance

There are more parameters in mutual relationship, which refer to network condition or network performance. One of performance parameters is *Latency*. Latency means a time required to transfer an empty message between relevant computers. Another parameter is *Data transfer rate*. Data transfer rate is the speed at which data can be transferred between sender and receiver in a network. The unit of this parameter is Bits/sec. For message transfer time calculating is equation 1. A third parameter of network performance is *Bandwidth*. Bandwidth is a total volume of traffic that can be transferred across the network. *Maximal data rate* formula is shown in equation 2. This maximum is only theoretical, not reachable in practice [5]

$$\text{Message transfer time} = \text{latency} + (\text{length of message}) / (\text{Data transfer rate}) \quad (1)$$

$$\text{Max. data rate (bps)} = \text{carrier Bandwidth} \cdot \log_2 (1 + (\text{signal/noise})) \quad (2)$$

The all parameters are pointing on the main disadvantage of controlling via the Internet – packets delivery delay. When packets are concurrently transported over an ordinary Ethernet, packets may experience a large delay due to contention with other packets in the local node where they originate and collision with other packets from the other nodes. By data transmission, four sources of delay spring up at each hop: nodal processing, queuing, transmission delay and propagation delay. The most significant part of total delay belongs to queuing. By queuing is considered the following equation 3:

$$TI = L * A/W \quad (3)$$

where TI is traffic intensity, L is packet length (bits), A is average packet arrival rate, and W is link bandwidth (bps).

If ratio L*A/W will be very small almost 0, average queuing delay is small. If ratio L*A/W rise up to 1, delays become large (exponentially) and if ratio L*A/W is bigger than 1 average delay is infinite, more “work” arriving than can be serviced.

V. Solution Approach

Adequate control software, appropriate computers on client

and server site and Internet with satisfactory connection speed are necessary for successful mechatronic system control controlling. Definition of "adequate" control software, computer and connection speed depends on concrete mechatronic system. In generally, regulation of mechatronic system may be considered as real-time regulation problem and time intervals in tens of milliseconds. The time intervals may vary significantly from every regulation system. For regulation system via Internet is very important if the regulation loop time interval must be under one millisecond, in milliseconds or may be over hundreds of milliseconds and more. In the architecture design, a remote regulation of mechatronic system via Internet generally includes three major parts: client, server and regulated mechatronic system. The general remote regulation system architecture is shown in Figure 2. The client part is the interface for the operations.

It includes computers, control software with user interface for operators or superior system. Client computer receives state information of mechatronic system, connection state and other information related to the system regulation via Internet. Received information will be processed and evaluated in remote computer.

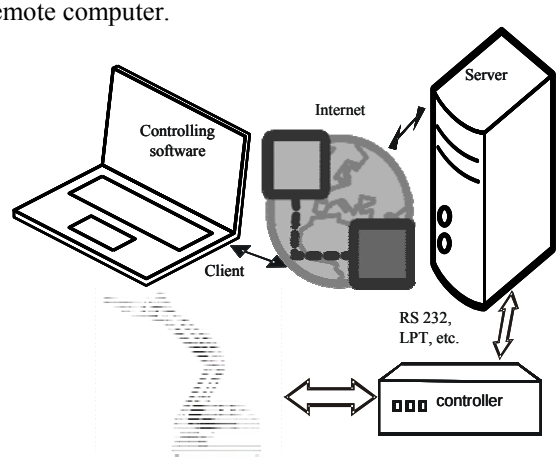


Fig 2. Remote system architecture

The server part contains a server computer, which is connected to the converter. Server contains all required drivers and devices for communication with the converter.

Communication of server with converter could be based on several ways (RS232, Profibus, CAN, USB, etc.). Sophisticated converters may be Ethernet enabled and may be connected directly to the Internet. But if the client computer is located in outside network – not in LAN network, where the converter is located, the server computer is recommended.

The third part of system architecture is the mechatronic system with the controller itself. Common way for distance regulation is, when remote client computer has limited functions – only start/stop of mechatronic system. The regulator itself (for instance PI regulator) is located on server site, or is implemented in converter.

But Internet speed progress open possibility for real-time control from client site, so there is possibility that Internet could be part of the regulation loop. Between client computer and server could be thousands of kilometers, or they could be in the same room. The difference is in the communication delay, but generally the system is the same. The communication service (the bus) can be achieved by wired connection, mostly Ethernet, or wireless – very popular WiFi. If regulated system is sufficiently slow, also GSM devices may be used for Internet communication.

VI. Conclusion

It is become to be a standard, that many control elements have been embedded with Internet-enabled functions, for example, PLC with TCP/IP stack, smart control valves with a built-in wireless communication based on TCP/IP protocol. There is possibility that some mechatronic system could be connected directly to the Internet. On the basis of done analysis it is evident that the existence of server as a gate to the Internet for mechatronic system is still highly recommended (because of capriciousness of Internet, computer crime and many other reasons). By utilizing of UDP Internet protocol it is possible to regulate real-time systems with tenths milliseconds of feedback. When compare Ethernet as a bus with other standard types of industrial bus, there are more advantages and disadvantages. The most powerful advantage is nearly unlimited size of bus, possible huge distance, open system of the internet protocols and accessibility of the Internet.

Acknowledgment

The paper has been prepared by the support of Slovak grant projects VEGA 1/0660/08, KEGA 3/6386/08 and KEGA 3/6388/08

References

[1] Yang, S.H., Tan, L.S., Chen X: Requirements Specification and Architecture Design for Internet-based Control Systems, proceedings of the 26th Annual International Computer Software and Applications Conference (COMPSAC'02), 2002.

[2] Hall E.: Internet Core Protocols: The Definitive Guide, O'Reilly & Associates (February, 2000) USA, ISBN: 1-56592-572-6.

[3] http://www.anybus.com/technologies/technologies.shtml, 10.2.2008.

[4] Kováč, D., Kováčová, I., Molnár, J.: Electromagnetic Compatibility - measurement, TU Košice publisher, 72 pages, ISBN 978-80-553-0151-8.

[5] Fonda C., Postogna F., "Computer networking basics ", ICTP WORKSHOP ON TELECOMMUNICATIONS: SCIENCE, TECHNOLOGY AND APPLICATIONS. Trieste, 15th Sep. - 3rd Oct. 1997

[6] Molnár, J., Kováčová, I.: Distance remote measurement of magnetic field. In: Acta Electrotechnica et Informatica, 2007, Vol.7, No.4, pp. 52-55, ISSN 1335-8243.

[7] Molnár, J.: Telemetric system based on internet. In: OWD 2009 : 11 International PhD Workshop : Wisa, 17-20 October 2009, p. 38-41. ISBN 83-922242-5-6.

[8] Kováčová, I., Kováč, D.: Non-harmonic power measuring. In: Acta Electrotechnica et Informatica. Vol. 8, No. 3 (2008), pp. 3-6. ISSN 1335-8243.

[9] Tomčík, J., Tomčíková, I.: Safety politics in the enviroment of the automatized and SCADA systems (In Slovak). In: EE Journal. Vol. 14, No. 1 (2008), pp. 46-47. ISSN 1335-2547

[10] Vince, T., Molnár, J., Tomčíková, I.: Remote DC motor speed regulation via Internet. In: OWD 2008 : 10th international PhD workshop : Wisla, 18-21 October 2008. p. 293-296. ISBN 83-922242-4-8.

2nd section: Informatics & Telecommunications

Architectural Knowledge and the Process of its Acquisition

Iveta ADAMUŠČÍNOVÁ, Attila N.KOVÁCS

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

iveta.adamuscinova@tuke.sk, attila.n.kovacs@tuke.sk

***Abstract*—Current orientation of software engineering to development of highly adaptable software systems presents the main force behind the efforts of finding the ways to efficiently improve the processes of software maintenance, mainly related to the detailed comprehension of system, its components, functionality and analysis of impacts of maintenance induced system's modifications.**

In this paper, we emphasize the importance of complex software understanding by presenting the concept of architectural knowledge and its principles. We also propose our new approach of dealing with the most challenging task related to its further usage – the problem of its automatic extraction and processing – by introducing the process of architectural knowledge acquisition.

***Keywords*—architectural knowledge, architectural knowledge acquisition process, software comprehension, software maintenance**

I. INTRODUCTION

Thorough comprehension of software system presents one of the most important preconditions for successful realization of its maintenance processes [1]. This fact is tightly related to always increasing demands after development of software systems that are easily and automatically adaptable to constantly changing requirements and environments which nowadays present a great competitive advantage [2]. This trend is the main cause of growing pressures on simplification, acceleration and effectiveness of individual processes of the whole software life cycle, in particular of those related to software maintenance. The efforts of integrating the approaches and concepts from various different disciplines, mainly from the area of artificial intelligence and usually related to knowledge engineering, present the natural results of this ongoing pressure [3].

In this paper, we present our approach of dealing with this challenge by analyzing the concept and principles of the so-called architectural knowledge, highlighting the importance of overall software understanding. The second part of this paper introduces our new proposal of overcoming the most cumbersome task related to its usage – the problem of automatic extraction and processing the architectural knowledge – by presenting the brief overview of the whole process of its acquisition.

II. CONCEPT OF ARCHITECTURAL KNOWLEDGE

The first part of this paper is aimed at introduction of the general concept of architectural knowledge. We emphasize our proposal of its definition, identify its components and afterwards we describe the main principles of our approach to the process of architectural knowledge acquisition.

A. Definition of Architectural Knowledge

Knowledge in general presents a key concept within the field of knowledge engineering. It can be defined as following [4]:

Knowledge presents understanding of a subject area. It includes concepts and facts about that subject area, as well as relations among them and mechanisms for how to combine them to solve problems in that area.

The specific type of knowledge which represents the current direction towards securing the efficiency of handling the system's modifications during the various phases of software life cycle is called *architectural knowledge*, i.e. knowledge about specific software system and its environment. This type of knowledge is usually hidden within the artifacts of software system and it is assumed that its acquisition, explicit representation and following use could significantly influence not only the quality of the whole software system's development but predominantly its stages of maintenance and evolution.

Although there doesn't exist any official definition of the architectural knowledge yet, based on [4], [5] and IEEE definition of software architecture [6], we conclude the main idea in our proposal of its definition:

Architectural knowledge presents a thorough understanding of specific software system. It is defined as the integrated representation of the software architecture, its structure, behavior, the architectural design decisions and the external context/environment.

The need of this type of knowledge originates from the common fact that the basic requirement for effective solving of problems during the challenging stages of software life cycle is its detailed *understanding* and that often the best source or, in some cases, the only source of the knowledge about existing software system is the software system itself [7], [8].

B. Components of Architectural Knowledge

Architectural knowledge presents the aggregation of all critical knowledge related to processes of software system's

maintenance, management and evolution. Among them, we distinguish three fundamental types [8], [9], [10]:

1. *Knowledge about mutual relations and dependencies between the artifacts of software system*. This type of knowledge helps to maintain the software system in the consistent state even after the implementation of system's modifications. Consistency is an inevitable condition for achieving the non-problematic maintenance/evolution of software systems.
2. *Knowledge about mutual relations and dependencies between the elements within the artifact of software system*. This type of knowledge is necessary for analyzing the impacts of processed modifications on other, often not obviously related, parts of the system.
3. *Knowledge about connections and relations between the software system and its environment* (e.g. other software systems). This type of knowledge is used to express the way of communication and dependencies between the cooperating systems and is also necessary for analyzing the impacts of performed internal modifications on external systems.

These types of critical knowledge which form the main components of system's architectural knowledge, are always very tightly related to the specific software system and they are contained in all its *artifacts* [5], [10] – in its models, source code, object code, diagrams, documentation, etc., which are stored collectively in system's project database.

The main problems of this concept, inhibiting it from its further use within software maintenance process, are that the architectural knowledge (i.e. knowledge about the system) is in principle stored separately of the system itself and the complexity of automatic extraction of architectural knowledge from its sources (i.e. system's artifacts). Because of these serious obstacles, presented approach leads inevitably to significant increase of the complexity of the whole process of accessing, searching, sorting and using relevant knowledge while implementing the required modifications with regards on preserving the consistency between the system's artifacts [10].

Therefore, the next chapter will be dedicated to introduction of the new approach to automatic acquisition of architectural knowledge from software system's artifacts.

III. Process of Architectural Knowledge Acquisition

In this chapter we offer our proposal dealing with the mentioned problems related to complexity of automatic acquisition process, as the architectural knowledge usually remains hidden among great deal of implementation or technological details [10], [11]. During the process, it's necessary to appropriately identify the sources of knowledge in order to separate redundant or to problem domain irrelevant information (e.g. implementation details) from the information that are supposed to form the useful parts within the system's complex architectural knowledge. Also, for an efficient use of the knowledge, it's inevitable to properly extract, process, store and present the resultant knowledge. So, within the following part of this paper, we're introducing our approach to gradual acquisition and processing of architectural knowledge. The whole process is depicted on Fig.1 and its individual phases are afterwards briefly described.

A. Identification of Architectural Knowledge Sources

The most important (and within our approach the only ones considered) sources of architectural knowledge of related software system are its *artifacts*. Models, source codes, object code, documents, etc. form the natural knowledge base, as they are not just the simple sources of data, but also contain the various procedures and principles of its processing and utilization, presented usually in a structured (or semi-structured) way [5], [8], [10], [11].

In this phase, it's necessary to determine the main aim of usage of system's architectural knowledge, as it presents the primary precondition identifying the direction of the following acquisition process. This specification influences the *selection of sources* of architectural knowledge which are relevant to identified aim. For example, if user wants to explore the proportion and types of dependencies between the particular components of a large system, it's probably unreasonable trying to extract this knowledge from available documents roughly describing their functionality. Instead, it'd be practical to extract it from source codes and/or models of suitable character.

B. Selection of Ontology Schema(s)

For effective usage of the knowledge, it's necessary to express it by some representation technique supplemented by a powerful interpretation language. Nowadays, ontologies backed up by *Web Ontology language* (OWL) present de-facto a standard in knowledge representation, supported by successful Semantic Web initiatives [3]. In our approach we use a subset of OWL called *OWL DL* (DL - Description Logic) to represent the architectural knowledge, mainly because of its ability to provide maximum expressiveness possible while retaining computational completeness, decidability, and the availability of practical reasoning algorithms [12].

In this phase, we select the *OWL ontology schemas* corresponding to sources identified in previous step. These schemas are pre-prepared for specific types of artifacts (e.g. schema for source code written in Java, behavioral diagram in UML notation etc.) and they're stored in *architectural knowledge repository*. The ontology schemas provide in general a specification of ontologies' elements and their semantics using the sets of classes, properties, restrictions and relationships and they represent a base for creating ontologies of artifacts' instances [12].

C. Extraction of Architectural Knowledge

Within this phase, the selected system's artifacts are parsed in order to extract the knowledge specified by artifacts-specific ontology schemas. The architectural knowledge is extracted and processed by the so-called *Processing modules*, they operate independently and their functionality or used procedures are strictly related to the nature of parsed artifact. Currently, we support three types of processing modules, extracting the knowledge from Java source code, bytecode and various UML diagrams. They traverse artifacts, searching for the elements specified in ontology schemas.

Found instances of elements are afterwards temporarily stored within *in-memory models* or *intermediate files* saving the relationships between classes and/or instances in form of *triples* (e.g. `TriedaA isOfType Class`, `TriedaA definesMethod getABC` etc.), which help to streamline

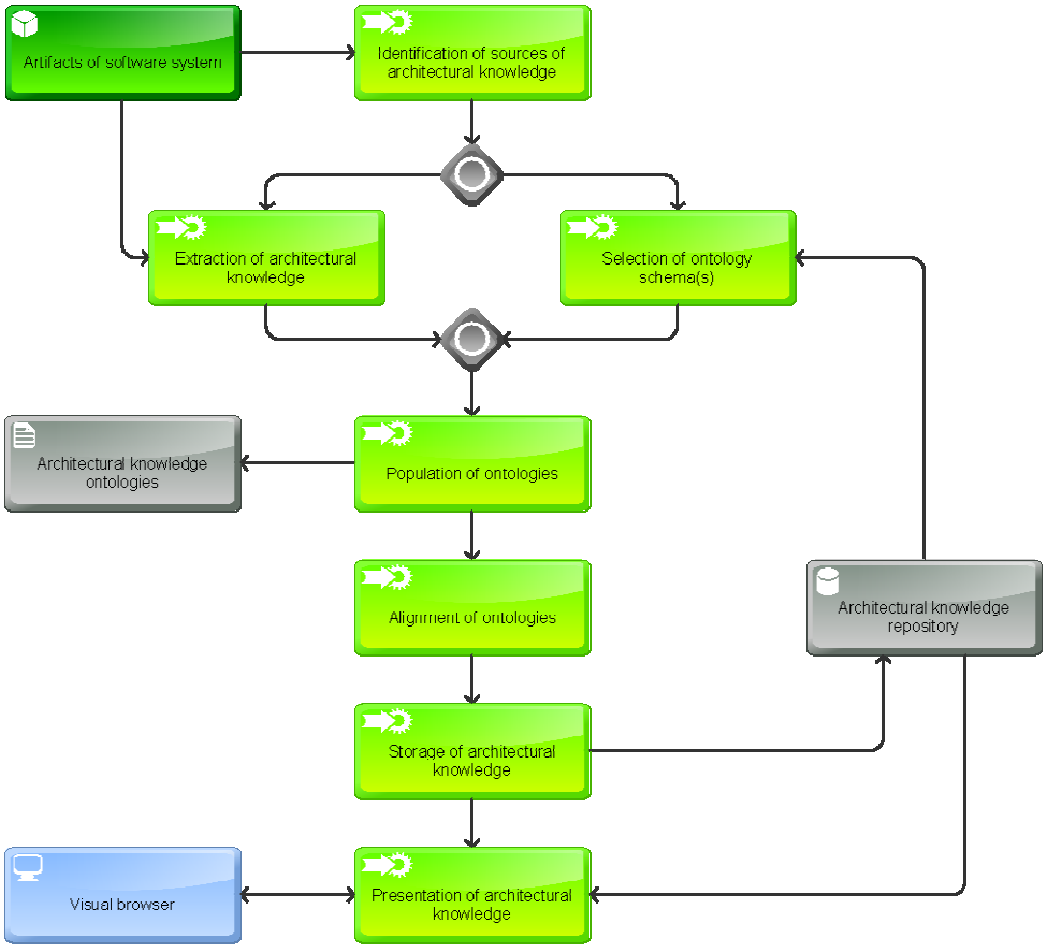


Fig. 1. Process of Architectural Knowledge Acquisition and its phases.

the subsequent phases of acquisition process by not using the original unprocessed artifacts, but files with pre-organized and structured *pseudo knowledge*. These temporal storages of ontologies' instances including the appropriate ontology schemas present the inputs for the next phase of the acquisition process.

D. Population of Ontologies

The main focus of this phase is the creation of artifacts' ontologies and subsequent population of these ontologies with extracted pseudo-knowledge stored in the in-memory models or intermediate files (created in previous phase).

The process of fully automatic population presents a nontrivial task, which is however significantly simplified by using our approach of temporal knowledge storage that is afterwards queried in order to fill the artifacts' ontology structures with concrete instances. After the process of population, we strongly recommend to *validate* the newly created ontologies by execution of the selected *reasoner*. Although this step is usually not necessary (as the underlaid ontology schemas are always validated during their creation), we suggest its inclusion, mainly for validating the highly populated ontologies (e.g. containing more than 2000 instances) [10].

E. Alignment of Ontologies

This phase serves as mediation for *interconnection of artifacts' ontologies* created within previous phases in order to obtain a complex architectural knowledge of the system. The process of establishing the relations between the concepts of existing ontologies presents nowadays a very cumbersome task. Actually, there is an active research into techniques to automate the process, but at this point, the task must be done by humans. While current tools can calculate class name and graph similarity metrics to try to give suggestions, they cannot yet consistently align ontologies automatically [13]. To achieve the *semi-automatic alignment* of ontologies' concepts of pre-prepared schemas, we created the set of *rules* which can be flexibly updated, reused (in other types of alignments) and shared. Currently, we use the support of *Jena framework* (i.e. Jena rules) which has its own rule engine [14]. However, in future, we plan to design the solution which would instead use the implementation of *SWRL* [15], a 'standard' rule language, to stabilize the automatization process.

F. Storage of Architectural Knowledge

In order to access and actually use acquired architectural knowledge, it's inevitable to store the individual artifacts' ontologies, their corresponding ontology schemas and alignment rules in common architectural knowledge repository. In our approach, it simply epitomizes a concept of *independent knowledge* base that forms an extension to existing software system to assist a user (during the system's maintenance process) but doesn't interfere with its original functionality.

G. Presentation of Architectural Knowledge

The acquired and processed architectural knowledge is afterwards presented by the *Visual browser*. The browser forms the semi-external part of extended software system

architecture. It cooperates with the architectural knowledge repository by using the wide set of pre-prepared *queries* on joint representation of architectural knowledge and provides the user friendly interface which is implemented as an interactive web page. Therefore, the user (e.g. maintainer) can easily navigate through the stored knowledge (represented by tree structure), study the system's structural/behavioral properties, inspect and analyze the potential impacts or side effects of particular modifications and observe its propagation into all system artifacts.

To ensure the continual recentness of the stored knowledge after modification of system's artifacts, it's necessary, in current state of our research, to repeat the phases of knowledge extraction and ontologies population. Afterwards, this new version of software system's architectural knowledge is stored in knowledge repository and is again easily accessible to the users by visual browser.

IV. Conclusion

In this paper, we presented the main principles of the essential element influencing the thorough comprehension of observed software system – the concept of architectural knowledge. We tried to propose our definition of concept and point out its main components. In the second part of paper, we introduced our approach to automatic acquisition of architectural knowledge with brief overview of process individual phases.

As a part of our future research, we consider the extension of the support for extracting and processing of some other types of available software system's artifacts.

Acknowledgment

This work is the result of the project implementation: Centre of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF.

The paper was prepared within the project "Life cycle and architectures of program systems based on the knowledge",

No. 1/0350/08, 2008-01-01 2010-12-31 with the support of VEGA.

References

[1] K G. Canfora, A. Cimitile, *Software Maintenance. Handbook of Software Engineering and Knowledge Engineering, volume 1*. World Scientific, 2001, ISBN: 981-02-4973-X.

[2] H. Yang, M. Ward, *Successful Evolution of Software Systems*, Artech House Publishers, 2003, 300 p., ISBN: 1580533493.

[3] H. Happel, S. Seedorf, „Applications of Ontologies in Software Engineering", *Proceedings from 2nd International Workshop on Semantic Web Enabled Software Engineering*, 2006.

[4] J. Durkin, *Expert Systems: Design and Development*, Macmillan: New York, 1994. ISBN: 0133486400.

[5] I. Gorton, A. Babar, „Architectural Knowledge Management: Concepts, Technologies, Challenges", *29th International Conference on Software Engineering - Companion*, 2007, pp. 170-171.

[6] I. Gorton, *Essential Software Architecture*, Springer-Verlag: Berlin Heidelberg, New York, 2006, 283 p., ISBN-10: 3-540-28713-2.

[7] A. Isazadeh, „Software Engineering: Integration", *Journal of Applied and Computational Mathematics*, Vol. 3, No. 1. 2004, pp. 56-66.

[8] I. Adamuščínová, *Znalosti, ich reprezentácia a využitie v životnom cykle a architektúrach softvérových systémov*, Písomná práca k dizertačnej skúške. KPI FEI TU v Košiciach, 2008, 112 p.

[9] M. G. B. Dias, N. Anquetil, M. K. de Oliveira, "Organizing the Knowledge Used in Software Maintenance", *Journal of Universal Computer Science*, vol. 9, no. 7, 2003, 641-658.

[10] I. Adamuščínová, M. Révés, Z. Havlice, "Using Architectural Knowledge in Process of Software Maintenance", *SAMI Proceedings*, pp. 83–88, 2010, ISBN: 978-1-4244-6423-4.

[11] J. Porubän, *Návrh a implementácia počítačových jazykov*, Habilitation thesis, Technical University of Košice, 2008.

[12] D. McGuiness, F. van Harmelen, "OWL Web Ontology Language Overview," [online], <http://www.w3.org/TR/owl-features/>, W3C Recommendation, 2009.

[13] S. Ponzetto, R. Navigli, "Large-Scale Taxonomy Mapping for Restructuring and Integrating Wikipedia", *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, Pasadena, California, pp. 2083-2088, 2009.

[14] Jena – A Semantic Web Framework for Java [online], <http://jena.sourceforge.net/>, 2009.

[15] I. Horrocks, P. Patel-Schneider, "SWRL: A Semantic Web Rule Language, Combining OWL and RuleML" [online], <http://www.w3.org/Submission/SWRL/>, 2004.

Abstract Adaptive Model for Intrusion Detection

Michal AUGUSTÍN

Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

michal.augustin@tuke.sk

***Abstract* — This paper presents abstract adaptive model for intrusion detection as part of computer security, it presents a method for automated creation of detection model for data mining. Abstract adaptive model creates detection on-the-fly, data is received by the sensors of IDS (Intrusion Detection System). This approach reduces the cost of deployment and implementation, because there is no need to create training sets or profiles.**

***Keywords* — Intrusion Detection, Attacks, Abstract Model, Statistical intrusion detection methods.**

I. Introduction

Many current approaches for Intrusion Detection and Prevention Systems (IDPS) apply data mining technologies. This approach creates detection models by applying data mining algorithms on a great amount of data gathered by the system. These models have been proven to by greatly effective [1] [2].

The disadvantage of data mining approach is that data, needed for the creation of training profiles is expensive to produce. Abstract adaptive model collects a creates detection models on the fly. Data mining IDPS collects data from sensors, which monitor the required aspects of the system. Sensor can monitor the network activity, system calls or access to the file system. Sensor extracts the raw data, which is monitored for further production of formatted data, that can be used for production. Data, collected by the sensors are evaluated by detectors by using the detection model. This model defines and determines whether the collected data is intrusive or not.

Algorithms used for the creation of detection model are generally divided into two categories: Misuse detection and anomaly detection. Characteristic for the training profiles is different for each category.

Misuse detection algorithms model the known behavior of the attack [3]. It compares the data from the sensors with known attack patterns from the training profiles. If data from the sensor matches with a known type of an attack, then these data is considered as intrusive. This model is acquired by training on large scale of data, in which attacks have been manually labeled, thus this data is very expensive to produce, because of the fact, that all the data has to be labeled as normal or as some form of an attack [4].

Anomaly detection algorithm models the standard (normal) behavior of the system [5], it compares data with normal patterns, which have been acquired from the training profiles. If the received data is in someway different from the normal behavior of the system, then anomaly detection model classifies the data as a potential attack. Anomaly detection model is very popular, because it gives the chance to detect new attacks [6] [7]. Most algorithms require the data from the training to be normal, it should not include any forms of attacks. Models from one environment do not have to be compatible and do not have to meet functions in different environments, which means that for the purpose of the best detection, data should be gathered for each environment in which IDPS will be implemented.

Abstract adaptive model can automate the process of gathering data from the sensors, creation of attack models and distribution of these models to detectors. System could use the advantages of new algorithm for the detection of anomaly, which effectively work also over noisy data [8]. This algorithm creates detection models and tolerates a small amount of intrusive data, which can be combined with "normal" data.

II. Abstract adaptive model

The proposed IDPS model architecture contains three main components, these components are: sensor, detector, adaptive model generator. The sensor receives the formatted data from the detector. The detector analysis and responds to potential intrusion. Sensor also sends its data to the adaptive model generator where it receives new detection models (by learning). After the adaptive model generator has learnt the new detection model, the model is then sent to the detector. Figure 1 shows this process

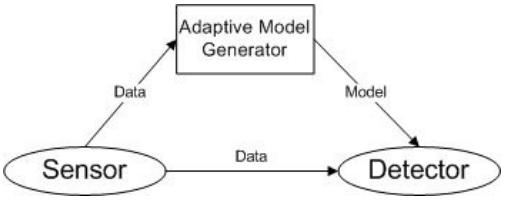


Fig. 1. Abstract Adaptive Model

The proposed adaptive model generator is composed of four components, these components are: data receiver, data warehouse, model generator and model distributor. Data receiver collects data from the sensor and converts them to such a form, so that they could be insert into the database or data warehouse. Components of data warehouse store the data to the database. Training sets or profiles can be generated from the model generator by using database queries. Components of the model generator create models by using training profiles, which have been acquired from the components of the data warehouse. The architecture of the adaptive model generator is shown in figure 2.

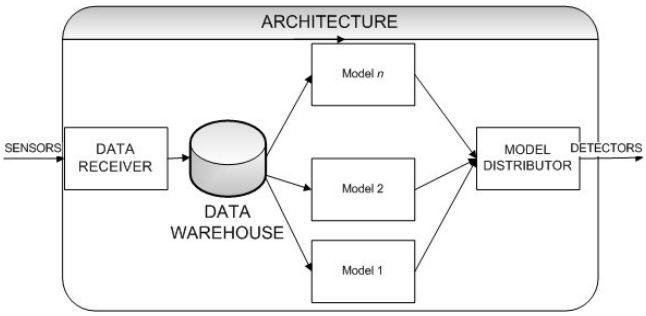


Fig. 2. Architecture of Adaptive Model Generator

III. Data representation

The data representation of the collected data (from the sensors) is generally in a raw type. Detection can be done by using an arbitrary evaluation engine and the model can be in the form of an neural network, set of rules or as an statistical model. Adaptive model generator has a robust mechanism, which deals with heterogeneous data and model representation, therefor it is considered useful to use XML technologies for data representation, as well as for models. XML can encode all the information needed for the input of information to the database.

IV. Data warehouse

The main advantage of storing data from the sensors in a data warehouse, is the ability to retrieve an arbitrary set of data just by querying. Database is very useful, because it is very ease to manipulate what data will be used for the training profile. When creating detection models base on system calls it is shown that, an effective model can be composed on the basis of modeling system calls for one program at a time [9]. If all the system calls are stored in the database, then it is possible to receive all the system calls for the given process just by one query.

V. Generation and Distribution Model

Adaptive model generation is designed to work with an arbitrary generation model algorithm and so the model generator can be viewed as a black box. Model generation algorithm has to be able to handle training data, which can contain some sorts of attacks. After the model generation has learnt, the distribution model sends the intrusion detection models to the detectors. The data representation of the collected data (from the sensors) is generally in a raw type.

VI. System efficiency

Efficiency is the key aspect in the design case. IDPS has to be sufficiently effective for a on-time attack detection and should minimize the work load on the system, which it protects. This feature is especially important in systems which are based on the monitoring of its hosts, because IDPS uses directly the resources of the system which it protects. It is suitable to implement the framework for the abstract adaptive model detection as a distributed system and so minimize the load, by minimizing the used resources on that particular system, which it protects.

VII. System response

System response in case of brute force attack can be solved by using a separate management network. In case of network overloading, system could send a response message via its management network. This architecture effectively isolates the management network from the production networks [11].

VIII. Principles of detecting anomalies

When using statistical methods of detecting anomalies a supposition is made that the behavior of the attach is significantly different from the normal behavior – statistical methods are used for modeling the user behavior and its differentiation from the behavior of the attacker. These techniques are used also for other subjects, such as user groups or programs.

To detect anomalies it is needed to determine which elements $\boldsymbol{x}$ were generated by the distribution of **A** and which elements were generated by the distribution of **N**, where elements generated by A are Anomalies and elements generated by N are Normal. For each element $\boldsymbol{x_t}$ its needed to determine whether it is anomaly or not. If it is anomaly move it to $\mathbf{A_{(t+1)}}$, in other case leave it in $\mathbf{N_{(t+1)}}$.

Likehood **L** for these distribution cases **D** in time $\boldsymbol{t}$ is:

$$L_t(\mathbf{D}) = \prod P_D(x_i) \quad (1)$$

A. NIDES/STAT

NIDES/STAT is a statistical method for real-time intrusion detection, it monitors the behavior of the subject in the computed system and adaptively learns what is considered as normal for individual subjects, such as users or groups [10].

If the observed behavior of the subject significantly differs from the expected behavior, then it is considered as potential intrusion (IDPS will give it a flag). The expected behavior is saved in the profile of the subject. Different measurements are

used for different aspects of the behavior.

The system periodically generates overall statistics T^2 reflects the abnormality of the subject. The value is the function of all abnormality values proceeding in the profile, and so, if there are n measurements $M_1, M_2, \ldots M_n$ model the behavior of the subject and if $S_1, S_2, \ldots S_n$ represent the measurement values of abnormalities M_1 *to* M_n the overall statistics T^2 given by:

$$T^2 = S_1^2 + S_2^2 + \ldots S_n^2 \quad (2)$$

(n independent measurements)

B. HAYSTACK METHOD

Haystack is a statistical detection algorithm for the identification of anomalies. Algorithm has the ability to analyze the users activity on the basis of four step process.

First step is generating session vector, which represents certain user activities. Vector $X = < x_1, x_2, \ldots x_n >$ represents the number of different attributes used for the representation of user activities of the given session.

Second step is to generate Bernoulli vector, which represents attributes, that are outside of the given session scope. Threshold vector T where $T = < t_1, t_2, \ldots t_n >$ and t_i is from $< t_i$ min, t_i max> is used to assist this step. The Bernoulli vector $B = < b_1, b_2, \ldots b_n >$ is generated so that b_i is set to 1 if x_i is not in the range of t_i ; b_i is set to 0 in other cases.

Third step is generating weight values for intrusion (for a particular intrusion type) from the Bernoulli vector and from the Weight intrusion vector $W = <w_1, w_2, \ldots w_n >$, where w_i describes the importance of the ith attribute in the Bernoulli vector for the detection of the given intrusion type. Weight intrusion value is the sum of all values (w_i), where ith attribute is not in the range of t_i and so: $\sum b_i \,.\, w_i$

Fourth step is generating the suspicion quotient, which represents how "suspicious" the session is with the intrusion type. The advantage of Haystack algorithm is better knowledge of attacks and better response to these attacks [10].

IX. TIME BASED INDUCTIVE MACHINE

Time-based inductive machine (TIM) is used to capture a user's behavior pattern. TIM discovers temporal sequential patterns in a sequence of events. These patterns represent high repetitive activities and have the functionality to provide predication. Temporal patterns are generated and modified from the input data using a logical inference. TIM can be applied to IDS and the rules are used to describe the behavior of patterns (user or group patterns) based on the audit history.

These rules are described as sequential event patterns that have the ability to predict next event from the given sequence of events. Simple rule produced is e.g.:

E1 - E2 - E3 -> (E4 = 95%; E5 = 5%) (3)

E1, E2, E3, E4, and E5 are security events. This rule says that if E1 is followed by E2, and E2 is followed by E3. Rules shows that event E4, is more likely to happen (95%), then to E5 (5%).

TIM has some limitations, it only takes into account the immediately following relationships between the observed events. The rule only represent the event pattern in which events are adjacent to each other. It can occur that a multiple task at the same time is needed to be executed, as a result the rules generated by TIM can not precisely capture the user's behaviour pattern [10].

X. CONCLUSION

Abstract adaptive model automatizes the process of gathering data from the sensors, creates models for detecting intrusions and distributes these models to detectors. This model gives the possibility of creation training profiles, that "learn" in the given environment (this requires time).

Adaption is done by the process of using statistical methods (NIDES/STAT, Haystack) for detecting intrusion or via time-based inductive machine which discovers sequential patterns in events. These methods can be automated and thus no user interaction has to be done.

There is also a possibility of immediate detection of intrusions by using detection methods that can be manually defined for the detector.

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0073-07 and VEGA grant project No. 1/0026/10.

REFERENCES

[1] W. Lee and S. J. Stolfo. Data mining approaches for intrusion detection. In Proceedings of the Seventh USENIX Security Symposium, 1998
[2] W. Lee, S. J. Stolfo, and K. Mok. Data mining in work flow environments, 1999
[3] Misuse Detection IDS Model, http://idstutorial.com/misuse-detection.php, 2009
[4] W. Lee, S. J. Stolfo, and K. Mok. Data mining in work flow environments: Experiences in intrusion detection, 1999
[5] Anomaly Detection, IDS, http://www.ciscopress.com/articles/article.asp?p=25334
[6] D.E. Denning. An intrusion detection model. IEEE Transactions on Software Engineering, 1987
[7] Stephanie Forrest, S. A. Hofmeyr, A. Somayaji, and T. A. Longstaff. A sense of self for unix processes, 1996
[8] Eleazar Eskin. Anomaly detection over noisy data using learned probability distributions. In Proceedings of the Seventeenth International Conference on Machine Learning (ICML-2000), 2000
[9] Stephanie Forrest, S. A. Hofmeyr, A. Somayaji, and T. A. Longstaff. A sense of self for unix processes. In Proceedings of the 1996 IEEE Symposium on Security and Privacy, 1996.
[10] Intrusion Detection Techniques, Peng Ning, Nort Caroline State University, Sushil Jajodia, George Mason University, http://discovery.csc.ncsu.edu/Courses/csc774-S03/IDTechniques.pdf, 2010
[11] Guide to Intrusion Detection and Prevention Systems, Karen Scarfone, Peter Mell, p. 3-1, Feb.2007

The Two-dimensional Map Analysis and Knowledge Representation for the Autonomous Navigation

[1]*František Baník,* [2]*L'ubomír Matis*

[1,2]Dept. of Electrotechnical, Mechatronic and Industrial Engineering, FEI TU of Košice, Slovak Republic

[1]frantisek@banik.sk, [2]lubomir.matis@tuke.sk

***Abstract*—The autonomous navigated vehicles are applied in present days more than ever before. Laser scanner is one of the interaction elements to create knowledge system. The article describes the data system of the two-dimensional laser scanner, results of analysis and knowledge representation used for the autonomous vehicle. Analysis of the result map was applied to the laboratory testing space. In the last chapter is a review of knowledge representations for the autonomous vehicles.**

***Keywords*—Autonomous navigation, autonomous vehicle, laser scanner, two-dimensional map, knowledge representation**

I. Introduction

The interaction instrument for the autonomous mode of the vehicle is needed. Spatial instruments can be constructed with distance sensors based on infrared, ultrasonic [1] or laser [2]. The two-dimensional (2D) laser scanner has been constructed as is described in [3]. Output data system can be modified for real vehicle purpose. Scanner data system is described in II. The measure data were transformed to 2D map of the environment. The basic environment for this testing was the space inside a laboratory. The map analysis is described in chapter III. Knowledge representation includes each measurement or scanning to hierarchical system, which is the base for the decision making in autonomous vehicle.

II. Scanner data system

The scanner was constructed as 2D system with 1D laser distance sensor and mechanical construction for rotation with this sensor. Stepping motor has been used for motion of mechanical construction, described in [3]. Control unit of the vehicle includes complete management of all processes in this scanner. Input informations before start scanning are start-angle, stop-angle and sampling rate. Start and stop angle can be set in range of $-180°...+180°$. Sampling rate is the number of stepping motor minimal steps which will be missed between scanning points, range is 1...3. The control unit loads the distance from the laser sensor and calculate angle from counter of the incremental sensor. This data are written down in the matrix 1 in the first row in form $[\delta, \gamma]$, where δ is the distance and γ is the angle. Then control unit sends impulses to the converter for the shift to the next scan point and this task is repeated. Result of the scanning is the matrix:

$$\begin{matrix} [\delta_1, \gamma_1] \\ [\delta_2, \gamma_2] \\ \vdots \\ [\delta_i, \gamma_i] \end{matrix} \quad (1)$$

The scan matrix is transfered to superior system and transformed to 2D map. Points of this 2D map are basic informations for the system of knowledge representation. The laser scanner has been experimentally tested inside the laboratory.

III. Analysis of the map data

Analysis of the real measure data brings the new ideas for modify of the map. The testing space is shown on figure 1.

Fig. 1. Testing space in laboratory

The testing parameters were set as $360°$ space and resolution 1. Scanning data were wrote down to matrix and then transformed to 2D map of the laboratory, figure 2.

The result map is a system of lines connecting each scanned point of the space. As we can see on figure 2, the result map includes real scanned objects and an imaginary lines between real objects. It is simple to find and filter an imaginary lines, because distance between two real points is much shorter than distance between the last point scanned on the first object and the first point scanned on the next object. Real object lines are black on the figure 2 and the imaginary lines are red. Imaginary lines present free space, that is interesting for the next scanning from another point of view. Free spaces are described as $S_1 - S_6$. Free space F (yellow color) is a range distance of the scanner and also present the interesting space. Blue lines C_1, C_2 are spaces between objects, that are narrow for vehicle motion and can be closed in the map as the fixed object.

IV. Knowledge representation

Sensing involves the collection of information, and also involves some preliminary treatment of the collected data, while control makes use of these data for immediate determination of control signals to bring the system configuration

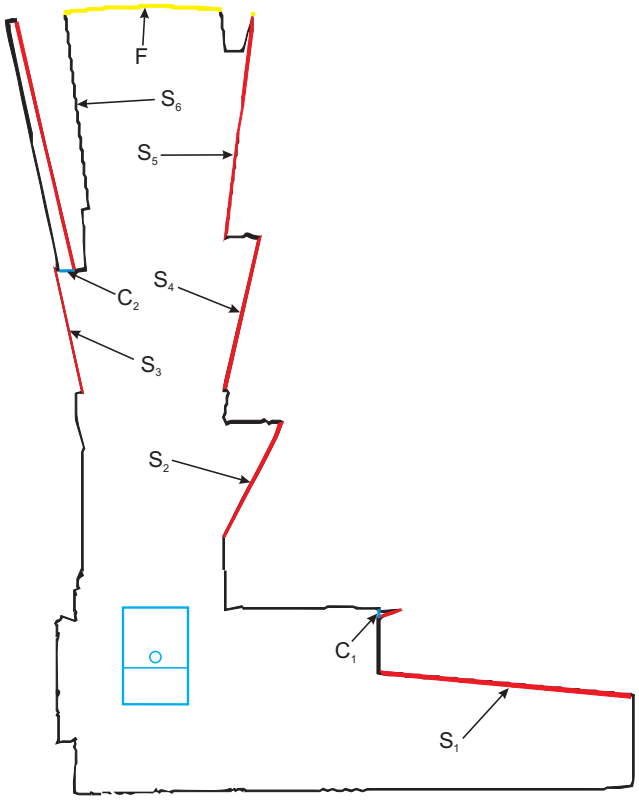


Fig. 2. The result map of the laboratory

to the desired one. Both these processes lack the sophisticated deliberation that makes a system intelligent. Such intelligent systems should be able to plan their own paths through an unknown environment, make decisions about their goals, and react to the decisions of other robots it senses.

Before the first levels of planning can actually take place, the information obtained from sensors has to be organized into suitable and useful forms. Structure of these forms creates system of the knowledge representation. Knowledge is central to a mobile robots ability to carry out its missions and adapt to changes in the environment. The knowledge subsystem must support acquisition of information from external sources, maintain prior knowledge, infer new knowledge from the knowledge that has been captured, and provide appropriate input to the planning subsystem. In order to carry out these responsibilities, there are different categories of knowledge required, such as task (also known as functional or procedural), and declarative, which includes spatial (or metrical). Representation schemes for the various types of knowledge must be chosen so as to provide the best performance and reliability. Many design decisions must be made, taking into account the real-time requirements of the robot control system, the resolution of the sensors, as well as the on-board processing and memory.

Representations useful for practical applications can be divided into a several groups:

A. *Spatial representations*

A large number of the mobile robot systems implemented have relied on spatial representations. Decomposing the space that the robot has to travel within into uniform or nonuniform regions (a geometric space) is one approach. Two commonly used geometric spaces are world space and configuration space (Cspace). World space is defined as the physical space that the robot, obstacles, and goals exist in. A particular location in world space can typically be represented by two to four parameters, where planar worlds with static environments require two parameters (x and y location) and 3D worlds with dynamic environments require four (x, y, z, and time). A configuration of an object may be defined as the independent set of parameters that completely specify the location of every point (or the pose) of the object [4]. The set of all possible configurations is known as the configuration space, and represents all of the possible poses of an object. The number of parameters necessary to specify the Cspace (or the dimensionality of the space) is also known as the degrees of freedom of the object. World space has the advantage of having objects in the world directly integrated into the space as opposed to having to compute potential object configuration interactions in Cspace. However, for nonholonomic robots, any path found in the Cspace is guaranteed to be collision free and realizable whereas a path found in world space may cause collisions with parts of the robot [5].

Grid-based structures [6] are a convenient means of capturing input from the robot's sensors, especially if multiple readings from one or more sensors are to be fused. They have the advantage of being easy to implement and maintain, due to their uniform, array-like structure. Figure 3 shows an example of a grid representation.

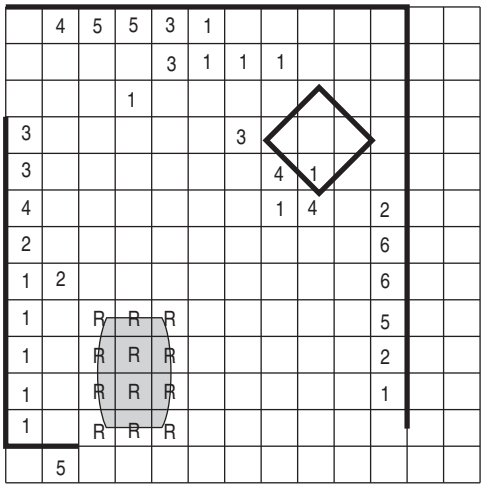


Fig. 3. Grid knowledge representation

Grid-based and other spatial representations vary in choice of coordinate systems and in the relationship to the robot itself. Some implementations use polar coordinates because the sensor data is returned in the form of distance (to object) and angle, reducing the number of calculations in constructing the map and in planning motion. The robot is always at the origin of the coordinate system in this case. However, it is more difficult to maintain a global map as the robot traverses the environment. The majority of implementations use a Cartesian coordinate system.

Other spatial representations are based on the geometric boundaries within the environment, such as planar surfaces [7]. These representations may augment the iconic or grid-based ones and often provide efficiencies by providing more compact descriptions of an environment, especially for indoor applications or highly structured environments.

B. Topological representations

Some systems represent the world via topological information [8], [9]. This enables them to reduce the amount of data stored and relate individual local maps together into a more global one. Topological maps provide qualitative information, noting significant entities in the environment, such as landmarks, and the connectivities and adjacencies amongst them, but do not provide exact coordinates or relative distances. Typically, topological information is implemented via graph structures, where the features are the nodes. The resulting maps are much more sparse and provide computational advantages in planning. Topological map of the figure 3 is shown on fig. 4.

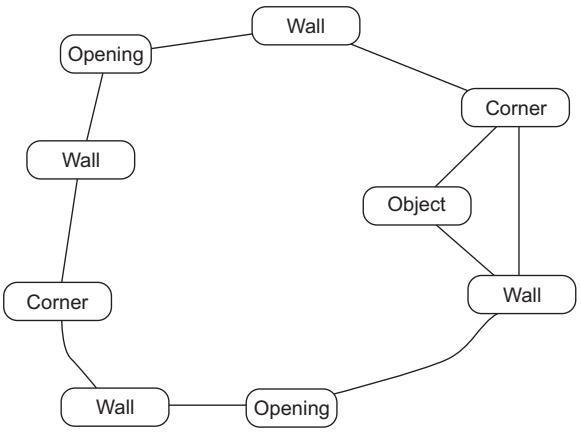


Fig. 4. Topological map

C. Symbolic representations

Symbolic representations provide ways of expressing knowledge and relationships, and of manipulating knowledge, including the ability to address objects by property. Much early work in robotics was carried out in the context of artificial intelligence research using symbolic representations [10].

D. No representation

Some have argued that representations such as those described earlier are too expensive to maintain and not valid due to the uncertainty inherent in trying to model the world [11]. This mindset paved the way for the robot architecture known as subsumption or behavior-based [12].

E. Multi-Representational Systems

Perhaps because of the overall complexity and difficulty of implementing a mobile robot, most implementations have relied entirely on a single representation approach. There are researchers who have chosen to expand the types of knowledge representations within their robotic systems to incorporate more than one type. The Spatial Semantic Hierarchy (SSH) is comprised of several distinct but interacting representations, each with its own ontology [13]. The SSH is based on properties of the human cognitive map and incorporates both quantitative and qualitative representations organized within a hierarchy.

The Polybot architecture [14] is designed to enable various modes of reasoning based on multiple types of data representations. Polybot is built upon a series of specialist modules that use any algorithm or data structure in order to perform inferences or actions. Since specialists may need to share knowledge, which they internally represent indifferent manners, a common propositional language for communicating information is part of the Polybot system.

A third example of a multi-representational architecture for mobile robots is 4D Real-Time Control System [15]. This architecture, with its hierarchical and heterogeneous world model, has been used in numerous types of implementations, ranging from underwater robots to autonomous scout vehicles. Several U.S. Department of Defense programs have selected 4D/RCS. These include the Army Research Laboratory's Demo III eXperimental Unmanned Vehicle.

V. Conclusion

The laser scanner data output is designed for wide range of opportunity. The basic output is matrix. The next step of processing the map is filtering the closed sections in the map. After this preprocessing, data can be inhered into knowledge representation. Spatial representation is the nearest to metrical data output. The analysis suggest the spaces in the map for the next exploration. Knowledge representation and this analysis technique is the base for decision making of the autonomous vehicle.

References

[1] S.-Y. Yi and B.-W. Choi, "Autonomous navigation of indoor mobile robots using a global ultrasonic system," *Robotica*, vol. 22, no. 4, pp. 369–374, 2004.

[2] P. Hoppen, T. Knieriemen, and E. von Puttkamer, "Laser-Radar based Mapping and Navigation for an Autonomous Mobile Robot," *IEEE International Conference on Robotics and Automation 1990*, pp. 948 – 953, May 1990.

[3] F. Baník, "2d laser scanner for the navigation of the autonomous vehicle," *9th Scientific Conference of Young Researchers of Faculty of Electrical Engineering and Informatics Technical University of Kosice*, no. 1, pp. 114–116, May 2009.

[4] Y. K. Hwang and N. Ahuja, "Gross motion planning - a survey," *ACM Computing Surveys*, no. 24, pp. 219–291, 1992.

[5] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. New York: Prentice-Hall, 1995.

[6] J. Borenstein and Y. Koren, "Real time obstacle avoidance for fast mobile robots in cluttered environments," in *IEEE ICRA*, Cincinnati, Ohio, 1990.

[7] D. C. W. B. Y. Liu, R. Emery and S. Thrun, "Using em to learn 3d models of indoor environments with mobile robots," in *Proceedings of the Eighteenth International Conference on Machine Learning (ICML)*, O. E. Brodley and A. P. Danyluk, Eds. San Francisco, California: Morgan Kaufmann Publishers, June - July 2001, pp. 329–336.

[8] E. Fabrizi and A. Saffiotti, "Augmenting topology-based maps with geometric information," *Robotics and Autonomous Systems*, vol. 40, pp. 91–97, 2002.

[9] D. Kortenkamp and T. Weymouth, "Topological mapping for mobile robots using a combination of sonar and vision sensing," in *Proceedings of the Twelfth National Conference on Artificial Intelligence*, AAAI. Menlo Park: AAAI Press/MIT Press, July 1994, pp. 979–984.

[10] D. Etherington, "What does knowledge representation have to say to artificial intelligence?" in *Proceedings of the Fourteenth National Conference on Artificial Intelligence*. Menlo Park, California: AAAI Press, 1997, p. 762.

[11] D. Kortenkamp, P. Bonasso, and R. Murphy, *Articial Intelligence and Mobile Robots: Case Studies of Successful Robot Systems*. Cambridge, MA: MIT Press, 1998.

[12] R. Brooks, "A robust layered control system for a mobile robot," *IEEE Journal of Robotics and Automation*, vol. 2, pp. 14–23, 1986.

[13] B. Kuipers, "The spatial semantic hierarchy," *Artificial Intelligence*, vol. 119, no. 1-2, pp. 191–233, 2000.

[14] N. Cassimatis, G. Trafton, M. Bugajska, and A. Schultz, "Integrating cognition, perception and action through mental simulation in robots," *Robotics and Autonomous Systems*, vol. 49, pp. 13–23, 2004.

[15] J. S. Albus, "4D/RCS Version 2.0: A Reference Model Architecture for Unmanned Vehicle Systems," National Institute of Standards and Technology, Gaithersburg, MD, NISTIR 6910, August 2002.

An automated headstone photo engraving

Mišél BATMEND

Dept. of Electrical Engineering, Mechatronics and Industrial Engineering, FEI TU of Košice, Slovak Republic

m.batmendijn@email.cz

***Abstract*—Nowadays, machine engraving of photos into solid materials such as marble or granite becomes very popular. Relatively cheap CNC (Computer Numerical Control) machines are available. The problem is that high quality photos are essential to obtain good results. This paper deals with principles of image processing applied on poor quality photos to get the best results. It also describes a model of CNC machine used for engraving.**

***Keywords*—photo engraving, image processing, CNC control, halftoning**

I. Introduction

Engraving photos into headstones is an old, commonly spread handicraft. It has been handmade for years (see Fig.1), but with birth of CNC machines there was a need to automate this action.

Fig. 1 An illustration of handmade portrait

Nowadays, a variety of engraving machines is available. Positioning system usually uses stepper motors in combination with acme screws or cogged belts. The main difference among machines is in engraving tool. Laser beam is used as a most expensive solution, rotating milling tools are also one of possible solutions. In our model we used simple but effective tool, which consists of vertically moving diamond "dotting" white dots into dark stone.

Concerning any of these tools, result of machine engraving is strongly dependent on quality of input image. Experience shows that skilled craftsman is able to get reasonably good results also from poor quality photos (like those on identification cards). Therefore some kind of image processing is essential. Variety of filters was applied on original images, trying to get the same result as handmade portraits.

With such an enhanced grayscale image, problem of converting to binary outstands. According to [1], there are many converting algorithms such as classical screening, direct binary search, error diffusion and other. Some of them are useless in this case, other provide only slight differences in a final portrait. The best one has to be chosen.

Finally, the binary image has to be coded and sent to a control system of the machine in order to be engraved. The most common way of controlling homemade CNC machines is use of commercial programs e.g. Mach3 or TurboCNC (see [2]). These programs are G-code interpreters. G-code is a list of instructions describing a path of a cutting tool. Interpreters work with CAD/CAM software, which generates G-code from CAD files. Connection between machine and PC is done by parallel port. Particular pins of port are directly controlling stepper motor drivers. Disadvantage of this solution is in use of outdated parallel port. Also precise timing is problematic. We decided to design control system of the machine using microcontroller.

To prove the theory, physical model of engraving machine was built. It was used to experiment with different halftoning and filtering methods, different kinds of stone and variety of mechanical setups.

II. Image Processing

A. Filtering

Usual size of engraved photo is 15x20 cm. Resolution of the machine is preset to 159 dpi. Therefore, original scanned image is converted to such a resolution, with grayscale color depth. Consequently different filters are applied. Among the great scale of available filters, 5 filters, Fig. 2, have been chosen as most useful.

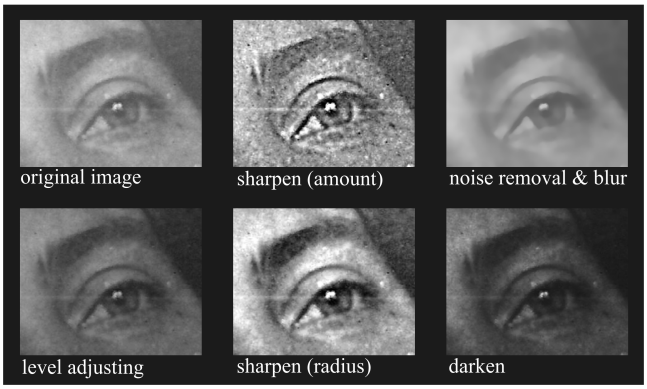


Fig. 2 Different types of filters applied on the same image

By applying a combination of these filters, virtually any picture can be enhanced. For different pictures, different combinations need to be used. The task of choosing filters is not deterministic, because criteria for "nice portrait" are more or less subjective. However, there are some hints that can help. *Sharpen (amount)* followed by *noise removal & blur* can be used at the beginning, to draw out contours of the image. Thereafter, *sharpen (radius)* will make a contrast between

light and dark areas of image. Finally, *darken* and *level adjusting* need to be applied in order to normalize brightness of resulting image. An example of adjusted image is shown in Fig. 3.

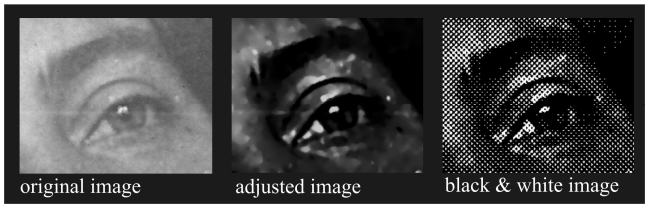


Fig. 3 Adjusted and B&W image after series of filters applied. Applied filters: *sharpen (amount)* 3x → *noise removal & blur* 1x → *sharpen (radius)* 1x → *darken* 2x → *level adjusting* 1x → *darken* 1x

B. Halftoning

Image halftoning is a process of converting high-resolution image to a low-resolution image, e.g. an 8-bit grayscale image to a binary image. Experiments show that cluster dot halftoning methods are more suitable for engraving than dispersed. Although, error diffusion methods produce higher quality halftones than classical screening, they do not look better when engraved. Thus, classical screen dithering algorithm with clustered dots seems to be the most suitable halftoning method for engraving into stone. MATLAB function *screen_19c()* using this method can be found in [1].

III. Data Coding

Once the image is in binary format, it must be sent from PC to control system. Generating of tool path code is done in PC program. Consider having an image as on Fig. 4.

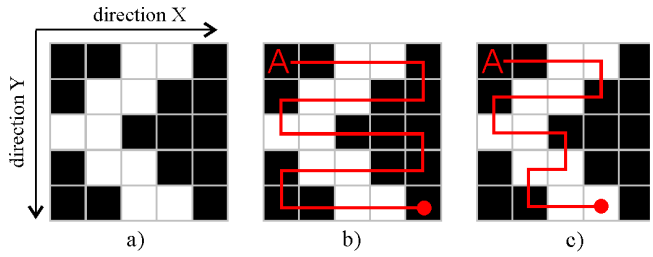


Fig. 4 Binary image and some of possible tool paths

Let's say that one step of stepping motor moves the tool by the distance of one pixel. Motor X operates in X direction, motor Y in Y direction. On each white pixel, the tool has to make a dot. On black pixel, tool takes no action. If the starting point of tool is in "A", then case c) on Fig.4 shows efficient tool path. Its computation is shown on Fig.5.

In our tool path scenario, there exist five possible actions, listed in Table I.

TABLE I
LIST OF POSSIBLE ACTIONS

Action	Code [decimal]	Code [binary]
motor X: step forward	1	0001
motor X :step reverse	2	0010
motor Y :step forward	3	0011
motor Y: step reverse	4	0100
tool :dot	5	0101

As can be noticed in flow diagram, particular actions do not occur simultaneously. Thus, the tool path can be coded as a series of actions. In each state, one action is taken. Actions are represented as 4 bit long word (Table I.)

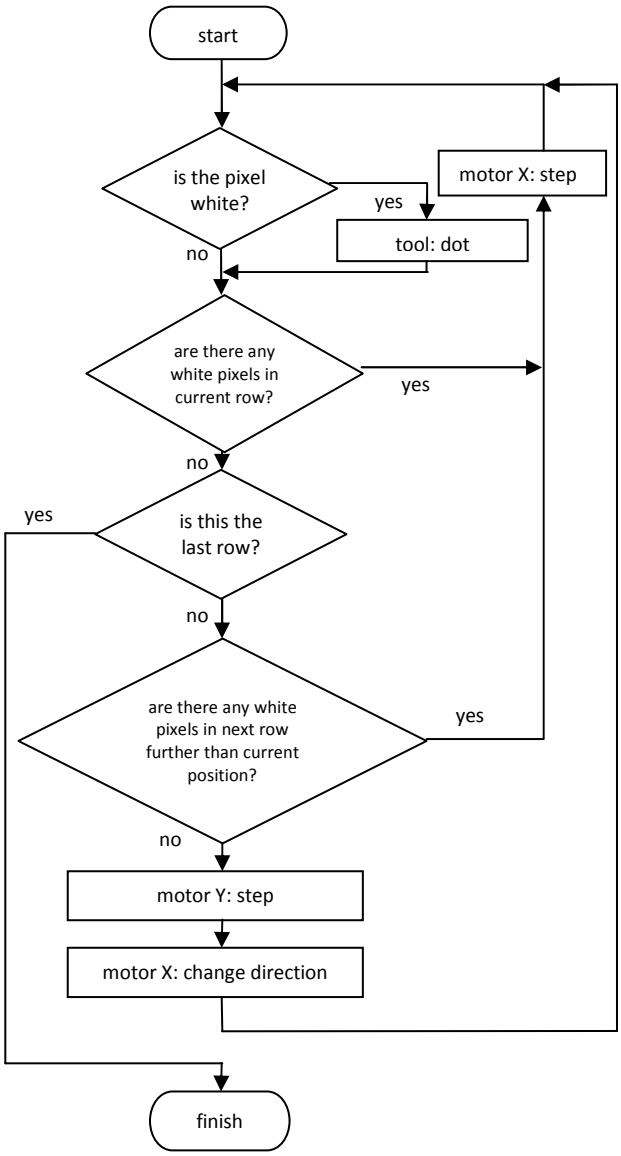


Fig. 5 Flow diagram for tool path

For coding of five actions only three bits are necessary. Fourth bit of word is added for future changes. The decimal code for the example image (Fig.4) would look as follows: 1 1 5 1 5 3 2 5 2 5 2 3 5 1 5 1 3 5 2 5 3 1 5 1 5.

IV. Control System

Once the data set representing the tool path is ready, it can be sent to a control system of the machine. The basis of control system is microcontroller ATmega162. It has 16MHz clock frequency, with 16K Bytes Flash, 512 Bytes EEPROM, 1K Byte SRAM memories. It is part of a control board where all necessary peripheral circuitry is implemented. Block diagram of control board with peripherals is on Fig. 6.

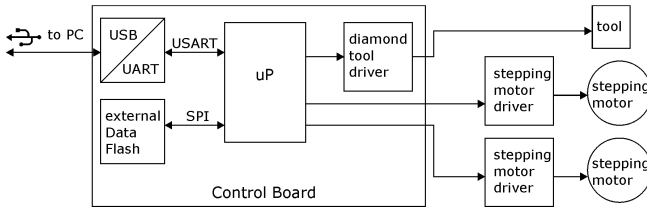


Fig. 6 Block diagram of control board and interfaces

Data or commands from highest level PC program come to control board through USB. Conversion to UART is performed by commercial integrated circuit FT232R. The data representing tool path are stored in 2M Bytes external flash memory AT45DB161D. It communicates with microcontroller through SPI (Serial Peripheral Interface). When all the data are stored, engraving routine can start. Operating of the machine is performed by PC program. Once the starting command is received, microcontroller reads the tool path data action by action and controls stepping motor drivers and a tool driver accordingly.

Two independent timers are used for timing of control pulses for stepping motors. Therefore exact timing for each motor can be achieved.

Tool driver is also integrated on control board. To comply with EMC standards it is decoupled by photocouplers. Vertical motion of diamond tool is achieved by electromagnet with moving core. The driver is a buck-boost power converter with MOSFET transistors driven by IRS2001 integrated circuit. It enables controlling of current flowing through electromagnet coil by PWM modulation of coil voltage. Thus, an electromagnetic force, which is proportional to current, can be controlled.

V. Physical Model

Physical model (Fig.5) was built to prove the theory described above. X and Y axes are moved by stepping motors combined with cogged belts. Belts turned out to be an ideal solution, because they are fast and provide enough precision. They also damp vibrations from tool. Common disadvantage of belts is elasticity, which leads to loose of precision while cutting. It should be pointed out that there is no cutting load, because the tool carriage stops moving every time it makes a dot. This constraint also leads to preferring stepping motors rather than servo motors.

Construction is mostly built out of Maytec profile system. A detailed documentation of mechanical construction and electrical parts can be found in [3].

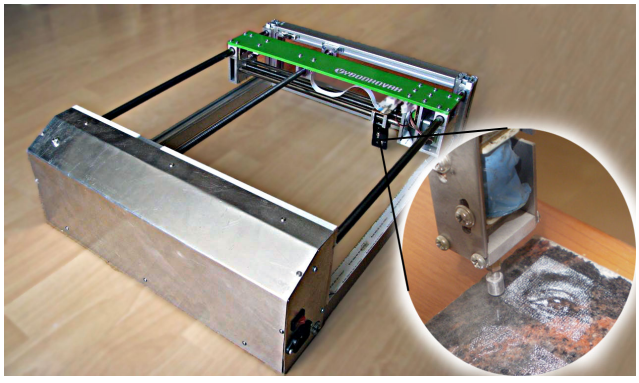


Fig. 5 Physical model of engraving CNC machine with detail of diamond tool and engraved image

An example of complete portrait is on Fig.6. As can be noticed, the binary image looks darker than engraved one. This has to be taken into account while adjusting image. Engraving of such a portrait takes two hours in average, depending on size of the portrait.

Fig. 6 Original grayscale photo (on the left), binary image (in the middle) and image engraved into black granite (on the right)

VI. Conclusion

Relatively simple and low-cost method was designed and also tested to engrave photos into headstones. It produces results that are comparable or even better than competitor ones (laser, milling tool). The machine has been used in real stonemasonry environment for two years. More than sixty portraits have been engraved. Lower quality photos and different types and colors of stone were used for engraving, providing satisfactory results.

In some cases, craftsman hand-enhanced some details to make the portrait brilliant. Especially face features (eyes, mouth) on very low quality images needed more contrast.

Further work can be focused on automation of image processing. Learning algorithms could be used for choosing a best set of filters for particular image. Also face detection [4] followed by facial feature extraction [5] can be used for independent facial feature enhancement (eyes, nose, mouth could be adjusted separately).

Different sizing of white dots can be achieved by applying different forces on the tool. Bigger and smaller dots might improve a contrast between dark and light areas of the portrait. Distribution of big and small dots could be based on grayscale image. Simple applying of threshold as well as sophisticated dithering methods could be used.

References

[1] V. Monga, N. Damera-Venkata, B. L. Evans, ”Halftoning toolbox for MATLAB” The University of Texas at Austin, [online]. Available: http://users.ece.utexas.edu/~bevans/projects/halftoning/toolbox/

[2] L. Davis,“ Hobby CNC Milling Machine“. In Servo 04/2005 pp.41-45 and Servo 05/2005 pp.51-54, USA

[3] M. Batmend, “Fyzikálny model polohovacieho stroja (Bachelor’s thesis)”, Dept. of Electrical Engineering, Mechatronics and Industrial Engineering, FEI TU of Košice, Slovak Republic, 2008

[4] S. A. Sirohey, “Human face segmentation and identification (Master’s thesis)“, [online], Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.48.810&rep=rep1&type=pdf

[5] K. Sobottka, I.Pitas, “A novel method for automatic face segmentation, facial feature extraction and tracking“, In Signal Processing: Image Communication, Vol. 12, No. 3, pp.263-281, 1998.

The proposal of beacon-based localization algorithm for Mobile Ad-Hoc Networks

Vladimír CIPOV

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

vladimir.cipov@tuke.sk

***Abstract*—The aim of the article is to bring a proposal of localization algorithm for MANETs that utilizes RSS (Received Signal Strength) measurement. The first part deals with the classification of localization methods used in Ad-hoc networks. The proposed algorithm was classified into one of the classes. For better understanding of the proposed algorithm, the second part describes the problems of negative influences of wireless communication channel. For simulation of algorithm in real operation was used the model that is derived from outdoor signal propagation models. The creation of the algorithm and its simulations are carried out in the MATLAB programming environment. The results of simulations are presented in clearly arranged charts and graphs.**

***Keywords*—localization algorithm, beacons, MANET, RSS**

I. INTRODUCTION

The MANET belongs to the group of wireless networks. They are robust, dynamic but simple with frequent changes of topology and without any central points. Some of the applications based on the ad hoc network existence require information about the arrangement of individual network terminals, i.e. their position.

Until now there have been several proposals of localization algorithm for wireless networks presented. It is generally known that there is no universal algorithm suitable for the localization in each environment and for all kinds of network topology. Therefore, the design of localization algorithms in mobile ad hoc networks is a very significant task of MANET networks.

The above mentioned reasons explain why it is desired that the number of designed and tested localization algorithms constantly grows. The main goal of this work is to simulate the localization method introduced in [1] in real environment using RSS (Received Signal Strength) measurement and make the method more effective.

II. CLASSIFICATION OF LOCALIZATION ALGORITHMS

Localization algorithms are classified into the following classes [2]:

- **Direct approach,**
- **Indirect approach**,
  - *Range-based*,
  - *Range-free*.

Classification of localization algorithms

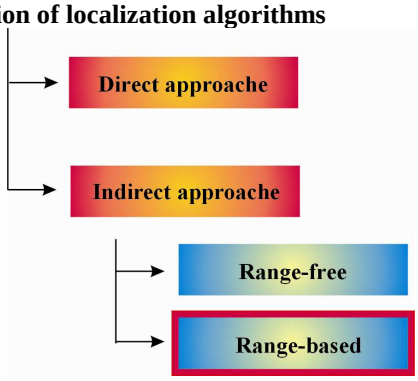


Fig. 1 Classification of localization algorithms

The **direct approach** is very cumbersome and not practical for large scale ad-hoc networks and networks with node mobility. In the GPS-based localization method, all nodes are equipped with a GPS receiver. This method adapts well for networks with node mobility, but it is not economically feasible to equip each node with a GPS receiver, because it is hardware-intensive. This method is not suitable for localization in indoor environment [2].

In the **Indirect approach** calculated unknown nodes position is relative to other neighboring nodes. Nodes with known position, called beacons, are equipped with GPS receiver or their position is manually configured. Other nodes compute their location with the assistance of these beacons [2].

Within the indirect approach, the localization process can be classified into the categories of **Range based** methods and **Range-free** methods [2]. In range-based localization methods, the location of a node is computed relative to other nodes in its vicinity. The absolute distance between the transmitter and receiver is estimated by the properties of received signal. The measured values of signal are used in some mathematical methods to compute the final location of unknown nodes. The accuracy of localization is subject to the transmission medium and surrounding environment. Range-free localization never tries to estimate the absolute point-to-point distance based on measurement of signal features. It is very appealing and a cost-effective alternative for localization in MANETs [2].

As is shown in Fig. 1 (red rectangle), the proposed algorithm belongs to the Range-based methods with indirect approach because it uses the measurement of RSS and method of circular trilateration.

III. Description of Mobile Radio Channel

The mobile radio channel defines the limitations on the performance of localization methods which use measurement the parameters of propagated signal to determine the distance between nodes.

Although such characterization does not exactly reflect reality, it has proved to be sufficiently useful and accurate to model mobile radio systems [3].

Authors in [3], [4] state that the degradation influences in mobile communication channel are a composite of few discrete effects as follows:

- **Large-Scale fading**,
- **Medium-Scale fading**,
- **Small-Scale fading**.

Large-Scale fading can be represented by few mechanisms. The simplest mechanism is the Free-Space loss where the communication channel takes place in ideal free space. The radio channel propagation characteristics are not specified. The space between transmitter and receiver is free of obstacles and the atmosphere behaves as a perfectly uniform and non-absorbing medium and the earth is treated as being infinitely far away from the propagating signal [4]. The models of path loss [3], [5] better represents the Large-Scale fading. The course of attenuation curve depends on the factors such as type of environment (indoor, outdoor, urban, rural, suburban) or high of antennas.

Medium-Scale fading is represented as a long-term fading or lognormal fading. Its variation is due to shadowing, terrain contour between the transmitter and receiver antenna [3].

Small-Scale fading is represented by rapid changes of signal power level in a very short time interval. This effect is represented by multipath fading of signal which is caused by reflection, diffraction, refraction and scattering of propagated signal [3], [5].

Both Medium-Scale and Small-Scale fading manifests itself in two mechanisms namely, time-spreading of the signal (signal dispersion) and time-variant behavior of the channel. These mechanisms will take place in two domains, time and frequency [4], [6].

IV. Description of Proposed Algorithm

Philosophy of this algorithm is similar to the method RSS [7] (Received Signal Strength) used as a localization method in cellular mobile networks. Each node situated in vicinity of minimum three beacons (in cellular networks base stations (BSSs)) is able to calculate its own position. The algorithm consists of several phases. In the first phase all mobile nodes which are in range at least 3 steady beacons are localized. The second and every other phase are like first one but difference is that all mobile nodes already localized in previous phase become virtual beacons. The algorithm is finished as [1]:

- Position of all mobile nodes was determined already, or
- It is not possible another mobile nodes position determine (undetermined node haven't sufficient number of beacons or virtual beacons).

The number of beacons with steady position and their deployment greatly affects the accuracy of proposed algorithm. On the other side, increasing number of beacons made localization algorithm less effective. The effort is made to use the smallest number of steady beacons.

Every phase of proposed algorithm consists of following four steps [1], Fig. 2:

1) Selection of a mobile node suitable for localization process,
2) The calculation of the distances between localizing mobile node and all beacons in its range,
3) Selection of appropriate beacons to calculate the position of the unknown node,
4) Use of trilateration to position calculation of localizing mobile node.

```
while(position of all nodes has not been
      detected or it is no longer possible
      to calculate the position of other
      nodes)
      for(detecting whether or not the node
          is a beacon)
          if(the node is not a beacon)
             detection of neighbour beacons
               to unknown node;
             calculation of distance to the
               beacons;
             selection the suitable trio of
               beacons [C];
                if(there is such a trio)
                   calculation the position
                     of unknown node;
                end
          end
      end
end
```

Fig. 2 Pseudocode of proposed algorithm

Each of the nodes includes the database of neighbors with information about RSS from them and also the information about the calculated distance to them by RSS. Selection of the appropriate beacons from this database yields to following criterion.

- All the trios of beacons which are considered are created.
- Only trios when all of the pairs of circles are intersected are considered.
- The trios which include the beacons with steady position are preferred.
- This nearest trio (according to received signal strength) of all is preferred. This trio creates the trio of beacons suitable for calculation the final position of unknown node.

If exists the suitable trio of beacons in range of unknown node, the localization process can be carried out. As mentioned to calculation the position of unknown node the process of trilateration is used.

V. Simulations and Results

Generated values of RSS are the result of interaction of all three types of negative channel influences which are described in section III.

For simplicity, **Large-Scale** was in general represented by simplified Frijs relationship [6] as a representation of the path loss in both the rural and also urban environment. The values generated by the Frijs relationship create the mean value of received signal strength for generating Gaussian distribution in step two – Medium-Scale fading and also the guide value for

Rayleigh and Rician distribution in Small-Scale fading.

As reported in literatures [3], [4], [6] **Medium-Scale** is simulated by normal Gaussian distribution. To simulate the Medium-Scale fading in MATLAB function `normrnd(μ,σ)` was used. The parameters μ and σ represents the mean value and variation of Gaussian distribution.

Small-Scale can be simulated by two mechanisms [3], [4]: ***Rayleigh distribution*** for urban environment with numerous obstacles and dominant Non Line-of-Sight communication was used. To simulate the Small-Scale fading due to Rayleigh distribution in MATLAB function `raylrnd(b)`was used. Mean value and also dispersion depends on parameter "b":

$$\mu = b\sqrt{\frac{\pi}{2}} \qquad (1)$$

$$\sigma = \frac{4-\pi}{2}b^2 \qquad (2)$$

Rici distribution for rural static environment without obstacles and dominant Line-of-Sight communication was used. To simulate the Small-Scale fading due to Rician distribution in MATLAB function `ricernd(v,s)`was used. Parameter "v" defines a lower limit of generated values and parameter "s" defines the variance of generated values.

In the simulations the impact of different variation of RSS due to type of environment was studied. Some of the features of proposed algorithm where studied as accuracy (the most important feature of algorithm), number of successfully localized nodes (success of algorithm), number of nodes localized in first phase (it is crucial because only the nodes localized in first phase estimate their position by all steady beacons which position is real and accurate) and number of needed phases (rate of algorithm). Several types of networks with different number of beacons and mobile nodes or different node sensitivity were created.

All types of simulations were hundredfold repeated for objectively evaluation. All simulations assume the area 1000x1000 meters. The sensitivity is the same for both the mobile nodes and also steady beacons.

Generating values for rural environment:

```
Free_space_loss=
  =10*logP_R(d)
Medium_Scale=
  =normrnd(Free_space_loss-3*σ, σ)
Small_Scale=
  =ricerdn(Free_space_loss,s)
The_resulting_attenuation=
  =(Medium_Scale+Small_Scale)/2
where:
o σ = [0, 0.25, 0.5, 0.75, 1]
o s = [0, 0.75, 1.5, 2.25, 3]
```

Generating values for urban environment:

```
Free_space_loss=
  =10*logP_R(d)
Medium_Scale=
  =normrnd(Free_space_loss-3*σ, σ)
Small_Scale=
  =(Free_space_loss)-raylrnd(b)
The_resulting_attenuation=
  =(Medium_Scale+Small_Scale)/2
where:
o σ = [1, 2, 3, 4]
o b = [3, 4.7, 6.4, 8]
```

A. Comparison the performance of proposed algorithm in rural and urban environment

The simulations assume the network with 40 nodes with node sensitivity -115dB and 3 steady beacons placed on [400,400] [600,400] [500,600]. In the following tables the comparison of performance of proposed algorithm is shown.

TABLE I
EVALUATION OF ALGORITHM IN RURAL ENVIRONMENT

RURAL ENVIRONMENT					
	Variance of RSS				
	Free-Space loss	σ=0,25 s=0,75	σ=0,50 s=1,50	σ=0,75 s=2,25	σ=1,00 s=3,00
Error of algorithm [m]	0	104,74	179,86	232,37	278,04
Successfully localized nodes [%]	99,6	98,8	98,6	98,45	98,2
Number of nodes localized in 1. phase	10,6	6,55	5,04	4,14	4,04
Number of phases	3,44	4,3	5,11	5,35	5,41

TABLE II
EVALUATION OF ALGORITHM IN URBAN ENVIRONMENT

URBAN ENVIRONMENT				
	Variance of RSS			
	σ=1 b=3	σ=2 b=4,7	σ=3 b=6,4	σ=4 b=8
Error of algorithm [m]	291,7	290,21	202,0	X
Successfully localized nodes [%]	86,5	52,5	7,5	X
Number of nodes localized in 1. phase	1,8	2	1	X
Number of phases	9,3	11	3	X

As seen from the tables the algorithm works better in rural environment. With increasing variance of RSS the accuracy decreases, number of nodes localized in the 1.phase decreases and number of phases increases. The information about parameters in urban environment is insufficient because the number of successfully localized nodes rapidly decreased.

As seen it was the problem when the negative influences of communication channel increased. The algorithm repeatedly did not work correct. In numerous cases of simulations the nodes in 1.phase were not localized. It was caused by large value of variance of RSS. The distances which were estimated

from measurement of RSS were not accurate calculated. The selection of suitable beacons for unknown nodes was not successful. Trilateration was not executed.

B. Influence to accuracy of proposed algorithm by different number of steady beacons.

The simulations assume the network with 40 nodes with sensitivity -115dB and rural environment where "σ" was changed from 0,25 to 1 and "s" was changed from 0,75 to 3.

The number and position of steady beacons were chosen as:

- One trio of beacons placed on [400,400] [600,400] [500,600],
- Five beacons, one common trio placed on [400,400] [600,400] [500,600] and two auxiliary beacons placed on [200,800] and [800,200],
- Two trios of beacons placed on [200,600] [300,800] [400,600] and [600,200] [700,400] [800,200].

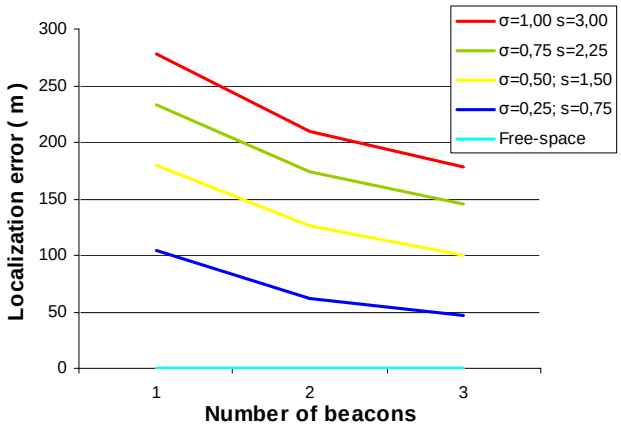


Fig. 3 Accuracy of algorithm for different number of steady beacons

It is clear that the accuracy of proposed algorithm significantly increased with using larger number of beacons. As was introduced the main goal of proposal was to use the minimum number of beacons, but using their correct number and correct deployment can bring better results.

The properties of proposed algorithm were also simulated in different types of environment with different number of mobile nodes and different node sensitivity. These results are described in words.

C. Change the number of mobile nodes

The simulation assume the network with sensitivity of nodes -115dB, rural environment where "σ" and "s" were changed as in section B., and 3 steady beacons placed on [400,400] [600,400] [500,600]. Number of nodes was chosen as 25, 30, 35, 40, 45 and 50. Different number of mobile nodes did not affect the properties of algorithm largely. Perhaps the number of successfully localized nodes in 1.phase and number of phases were affected. With increasing number of mobile nodes the number of nodes localized in 1.phase also increased and the number of phases decreased. However, to small number of beacons causes that it is not always possible localize nodes in the network. The 1.phase could not be realized because the nodes in vicinity of three steady beacons did not exist. Decreasing number of mobile nodes caused the problems with increasing influence of variation of RSS. The algorithm did not work correct.

D. Using different node sensitivity.

The simulation assume the network with 40 nodes rural environment where "σ" and "s" were changed as in section B. and 3 steady beacons placed on [400,400] [600,400] [500,600]. The sensitivity of nodes was chosen as -112dB, -113dB, -114dB, -115dB, -116dB, -117dB and -118dB. The values -112dB and -113dB where very small and caused problems in localizing. The success of method rapidly decreased. Number of neighbors to all mobile nodes was very small. With increasing sensitivity increased also the accuracy and decreased the number of phases of algorithm. The number of nodes localized in 1.phase increased.

VI. Conclusions

Localizing algorithm is better suitable for rural environment with extensive area and small number of obstacles. It is not suitable for urban environment. Great variance of RSS causes the problems in localizing process. The accuracy of algorithm with increasing number of phases decreases. It is caused by inaccurate calculation the final position of mobile nodes. These nodes become the new virtual beacons after each phase. Only the first phase uses the real position of steady beacons trio. The properties of algorithm such as number of nodes, their sensitivity and size of area must be concerted. Then it gives satisfactory results but only in rural environment. For future work the algorithm with method measurement the "time of arrival" will be studied. It is expected that the method "time of arrival" will give better results.

Acknowledgements

This work has been performed partially in the framework of the EU ICT Project INDECT (FP7-218086) and by the Ministry of Education of Slovak Republic under research VEGA 1/0065/10.

References

[1] Ľ. Doboš, V. Cipov, "Beacon based location algorithm for MANET terminals". *AEI Conference 2009*, Genoa, Italy, 7.-11. September 2009, pp. 28-35. ISBN 978-80-553-0280-5.

[2] A. Srinivasan – J. Wu, "A Survey on Secure Localization in Wireless Sensor Networks", *Encyclopedia of Wireless and Mobile Communications*, B. Furht (ed.), CRC Press, Taylor and Francis Group, 2008.

[3] P. Brída, J. Dúha, "Simulation of outdoor radio channel", *Proceedings of 6th International Conference Research in Telecommunication Technology RTT* 2005, 12.-14. 9.

[4] B. Sklar, "Rayleigh Fading Channels in Mobile Digital Communications Systems", Part I: Characterization: *IEEE Communication Magazine*, July 1997.

[5] L. Quing, "GIS Aided Radio Wave Propagation Modeling and Analysis", Blacksburg, Virginia, May 2005.

[6] Ľ. Doboš, F. Jakab, "Multimédiá v mobilných sieťach", Košice : TU-FEI, 2005. 196 s. ISBN 80-8086-000-9.

[7] P. Brída, P. Čepel and J. Dúha, "The Accuracy of RSS Based Positioning in GSM Networks", *In Proc. of 16th International Conference on Microwaves, Radar and Wireless Communications - MIKON 2006*, vol. 2, p. 541-544, ISBN 83-906662-7-8. Krakow, Poland, 22-26 May, 2006.

An Image filtration in distributed systems

[1]*Eva DANKOVÁ*, [2]*Peter JAKUBČO*, [3]*Marek DOMITER*

Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]eva.dankova@tuke.sk, [2]peter.jakubco@tuke.sk, [3]marek.domiter@tuke.sk

***Abstract*—Computer graphics is a section of informatic, which is expanding and still in progress. With growth of power are growing also requests on rendering quality mainly in computer graphics, which handle also with photorealistic renderng. In some cases is data visualization so difficult, that it is needed to realize calculations on distributed systems. To obtain image quality in project it is considering installing image filtration in ambient of distributed systems. Through projected system, the image quality will force by realization of photorealistic rendering. And also through application on distributed systems the filtration will not have a big affect on computing in term of time.**

***Keywords*— distributed system, filtration, photorealistic rendering, scene filtration**

I. INTRODUCTION

Three-dimensional photorealistic object rendering is a part of computer graphics, which in last year's attract more and more consideration. The main reason is rapid accumulation of computer power and new technology availability, thanks which we are able to execute image faster.

Goal of photorealistic rendering is to model entity of real or fictive world (film Avatar, 2009), so that sighting of the accrued image would afford experience of pursuing a real world. Photorealistic rendering is based on modeling physical facilities of light. The more it is trying to display entities of observing light more authentically, the faster is growing the processing time. The most used method for generating a high quality images working in image space is method of ray tracing. This method is based on tracing of light ray on the way from light source until the eye of the observer. Based on this information's collected in time of tracing its define colors of separate image pixels. Ray tracing is from the view of computing relatively difficult and so one of the possibility is to apply it on distributed systems. To make ray tracing more effective it is possible simultaneously with applying on distributed systems, to use also image filtration. The goal of the filtration is to smooth noise, to sharp other lines of image and simultaneously to don't increase the time needed for processing the image.

One of the import parts of the projected system by image processing is output part, which includes also filtration module. This module downloads parameters and color depth of image and then it built 3 matrixes needed for image filtration. Projected system, which is oriented on photorealistic rendering on distributed systems with fixation on image filtration, was realized at Department of Computers and Informatics FEI TUKE as a part of project KEGA 3/7110/09. Simultaneously it was supported by project APVV-0073-07 and KEGA 3/7110/09.

II. DISTRIBUTED SYSTEM

There are many different definitions in literature of distributed systems, but no one is adequate and no one is consistent with the other. The most screening is the next [5]:

„A distributed system is a collection of autonomous computers, which seems to user as a separate coherent system. "

This definition can be considered based on two aspects. The first one bear ship to hardware: computers are autonomous, and the second one is related with software, where from the view of access to system sources, distributed systems behave as a one computer. Based on this aspect can be defined following characteristics of distributed systems (DS) [5]:

- Differences between computers, communications details between them and inner organization of DS are hidden for user.
- Interaction between users and system is consistent and uniform aside from time and place of realization.
- Good DS are easy to extend.
- Good DS are high accessible (no ever).

Distributed system except listed characteristic should also fulfill some basic requirements [5]:

- *Equipment sharing* – system should allow to several applications to share system equipment (hardware, data).
- *Parallelism*
- *Openess* – system specification and all its interfaces are public.
- *Transparency* – users shouldn't know , if used equipment is local or remote.
- *Error resistant* – system should have ability to detect errors and to continue with processing after a part of distributed system occurs as unavailable.

In every distributed system hardware can be order in several ways, without reference to number of CPU. Layout can be different, especially in term of how they are with one another connected and how they communicate.

From the view of memory sharing, can be computers divided in two groups. The first group is built from computers, which have shared memory and are also called multiprocessors. The second group includes computers, which

don't share their memory, and which are also called multicomputer. The main difference between them is that by multiprocessors exists only one space for physical address and it is shared by all CPU. In case of multicomputer has every computer his oven memory [2].

There was projected several different topologies, where hypercube and grid are one of the most important (Figure 1). System selection is very important for effective system functioning, and therefore is in project used network environment, which built distributed system called also cluster [4].

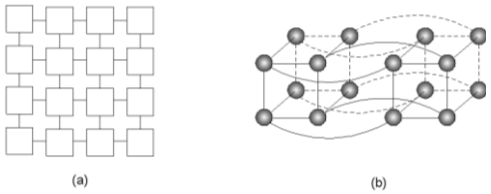


Figure 1. (a) Grid, (b) Hypercube

As hardware also software are very important for distributed systems. Software designates how a distributed system looks like and serves mainly as resources manager for basic hardware.

III. IMAGE FILTRATION

Image preprocessing is common name for operations with image at abstraction low level. Goal is to smooth noise, which can accrue by image digitalization and transmission. Also it can smooth or sharp image lines, which are important for next processing. Input and output of preprocessing methods are image data at low level of abstraction, namely matrixes, which are introducing digital image function [8].

Image preprocessing methods serve to better image from view of next processing, where are used information's about redundant data in image. Adjacent image elements have largely the same or similar luminance value. If it is possible to find image element out of drawing through accidental noise, than it is possible to repair his value based on average of luminance values in his ambient.

A. Image Noise

Image noise is casual, at most time undesirable variation of lightness or color image information. This type of noise can come from film crimping, electronically noise of sensors or input equipment circuit (scenery, digital camera). Image noise is best viewable in areas with low level of signal value, like shadowy areas, images with miserable lighting [9].

B. Impulse Noise

Image containing this type of noise have dark pixels in bright areas and bright pixels in dark areas. Impulse noise can be caused through "death pixels", errors in rendering from analog format to digital format and bits errors in transmission. Only some image points are devalued, which are substituted by constant value, which has following definition:

$$g(x,y) = \begin{cases} f(x,y), pre & 1-(p_0 + p_{255}) \\ 0, ak & p_0 \\ 255, ak & p_{255} \end{cases}$$

Value of new point is computed through random number generator. Summary of values p_0 a p_{255} are taking like "100%" a than follows computing with given probabilities. It takes number from interval <0, sum(p_0 , p_{255})).If the chosen number with value p_0 than pixel is replaced with minimal value, otherwise is replaced with maximal value [9].

C. Gauss Noise

Noise, which has probability graph of density in form of Gauss distribution [9]:

$$G(r,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{1}{\sigma\sqrt{2\pi}}} \quad (2)$$

Every point is specified through middle value μ and standard off-set σ.

IV. FILTRATION

Filtration is a file of image transformations, which transmit luminance values from input image on other luminance values from output image. Goal of this transmit is to underline, or suppress some its facilities. Very often we require suppressing of luminance different inside area, which includes noise. Choice of transmission depends on object amount [8]. It exist several filtrations types inter which belongs median filter, common average filter and edge detection.

A. Common Avarage

Common average filtrate image so, that new point luminance will be arithmetical average of point luminance's of its ambient. Areas which includes noise and are smaller than ambient size will be suppress .

Common average filtration is a special case of discrete convolution. For ambient with parameters 3x3 is convolution mask for this type of filtration. This method disability is that considerable measure spread edges. This problem can be solved with use of rotation mask. Around representative point is rotating a small mask. For every mask is computed luminance dispersion. New value is computed according to mask with the smallest dispersion. In this way edges will don't spread. [10].

B. Median Filter

Median filter belongs to nonlinear methods of image smoothing. The goal of this filtration is to eliminate big luminance differences in point ambient. Luminance values of point which falls into filtration mask are arrange according to their size. New luminance value will be median of this sequence. Median filter is suitable for suppressing of impulse noise. Disability of this filtration method is that thin lines and sharp angels could be damaged [10].

C. Edge Filter

Edge detection is a filtration method used to find object bounds, which are expressed as fast luminance changes. Tagged bound points can be connected in a form of lines or object outlines.

Is the sensibility by edge detection is very high, than it has trend to find also points in image, which can be adjudged as

an allowance of noise. If it is less sensible, it can conduce to loss of relevant edges. Parameters, which can be entering by edge detection, include size of the edge detection mask and sill value. Bigger mask is less noise sensible and lower sill value has addiction to reduce noise effect.

Edge are called places in image, where rapid change luminance value of image unction f(x,y). For process following of function f(x,y) is used gradient operator ∇. Gradient is a vector unit, which define the direction and size of the growth. Points with a big gradient value are regard as edge.

In analogue case, function gradient of two variables is calculated based on following equation:

$$f(x,y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) \quad (3)$$

Size of gradient can be defined according to following equation:

$$\nabla f(x,y) = \sqrt{\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2} \quad (4)$$

It is assumed that objects of image used in project will be relatively big, and noise wills occurs only in small areas, which differ from object with luminance. Then it is possible to remove noise in image through method based on common average without object defacing [10].

V. System proposition

To remove defects in image was projected a filter, which eliminates errors based on gauss or impulse noise. This filter smooth also sharp passing, which are built on places, where comes to expressive color changes.

Filter is projected so, that for his working are needed minimal 2 nodes. The master node is responsible for task distribution and slaves process these tasks and send them back to master node.

A. Master section

At the beginning master download from the file header size and color depth of the image. In consequence are built 3 matrixes, which are filled with data from the file.

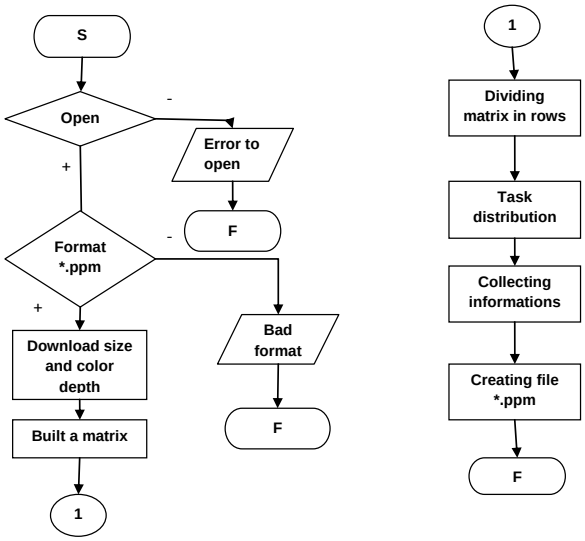


Figure 2. Flow diagram of projected filter, section master

This data represents color components of each image point. Master slave divide every matrix in parts (rows) and distribute them equally between nodes, where number of parts is equal to number of nodes. Number of rows, which are distributed to slaves differ at most of 1, in case, that number of rows is not dividable with number of slaves without residue.

Master collect data from slaves and save them in a new file, which will be in format *.ppm and program will corectlly finish. This new file is new filtered image.

B. Slave section

In first step every slave built a mask in a form of matrix of size 3x3 points. Slaves receive from master data.

Slave built a matrix for this data, which are initialized with one color component. Consecutive it's built bare output matrix. This matrix has identical size as matrix, in which are saved data received from the master.

It's passed step by step every point from matrix with color component and covers them with filtration mask. After that it is built a field, in which are saved all point values covered with filtration mask.

If no value from field of remembered points differ from other, than value of filtration point don't change in this point.

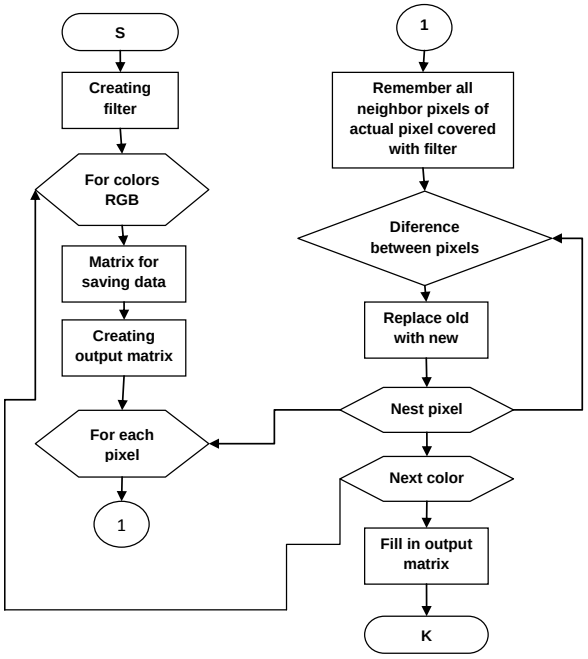


Figure 3. Flow diagram of projected filter, section slave

And in output matrix will be saved original value of this point. If difference in shade is big, than from the points in this area is computed arithmetical average. In output matrix is saved a new value. This is the way how slaves pass all matrixes with color components and points. Final matrix is forwarded back to master.

C. Contribution

Program, which was projected in MPI environmant with use oc C++, was aplicate on scene Mini_demo (Figure 4). This scene is orginaly Gilles Trans, which was publish at web page <www.oyonale.com> under icence Creative Commons By Attribution (<http://creative commons.org/licenses/by/3.0.

Scene contains different surfaces, tracing flats, reflections.

Projected filter was applied on rendered scene, which changed sharp passing's and places with a big difference between two color shades.

Figure 4. Scene sample mini_demo.pov©Gilles Trans, <www.oyonale.com>

On picture 5 it is possible to see that edge of the roof on the car are smoother (Obr. 5.a) than on original picture (Obr.5b).

a.)

b.)

Figure 5. Roof detail a.)Before filtration and b.)After filtration

TABLE I.
TYPE SIZES FOR CAMERA-READY PAPERS

Resolution 1024 x 768				
Number of nodes	2	5	10	15
Time [s]	1.93	1.74	1.49	1.11
Partial times [s]	1.93	1.72; 1.73; 1.73; 1.74	1.45; 1.45; 1.46; 1.48; 1.48; 1.47; 1.46; 1.49; 1.49	1.09; 1.10; 1.10; 1.10; 1.10; 1.10; 1.10; 1.10; 1.10; 1.10; 1.10; 1.10; 1.11; 1.11
Number of changed pixels	134708			
Resolution 2048 x 1536				
Number of nodes	2	5	10	15
Time [s]	8.26	6.73	6.43	5.78
Partial times [s]	8.26	6.63; 6.67; 6.68; 6.73	5.98; 6.04; 6.39; 6.39; 6.40; 6.41; 6.41; 6.43; 6.43	5.73; 5.74; 5.70; 5.74; 5.75; 5.75; 5.76; 5.76; 5.76; 5.77; 5.77; 5.77; 5.78; 5.78
Number of changed pixels	384020			

Filter was applied on picture with use of 2, 5, 10 a 15 nodes by to two different resolutions. It was monitored time needed for filtration and number of change points.

Measured values by realization of experiment are present in table 1. Dependence of filtration time and number of nodes is presented on Figure. 6.

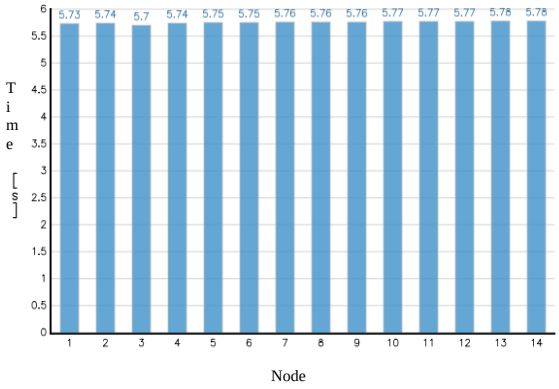


Figure 6. Graphical interpretation of partial times needed for filtration of mini_demo in resolution 2048x1536

Project was applied on distributed system, whereby was global time needed for filtration divided on separate nodes so the time needed for rendering wasn't influenced. Combination of image filtration and application on distributed systems was obtained more qualitative image without bigger time delay.

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under the contract No. APVV 0073-07, also the authors are pleased to acknowledge the financial support of the Cultural and Educational Grant Agency KEGA of the Slovak Republic under grant No. 3/7110/09

REFERENCES

[1] J. Kollár, "Metódy a prostriedky pre výkonné paralelné výpočty". 1. vydanie. *Košice*, Elfa, ISBN 80-89066-70-4, 2003.

[2] L.Vokorokos,"Digital Computers Principles", *Budapest*, pp. 232, Typotex 2004, ISBN 9639548 09.

[3] M. Jelšina, "Architecture of computer systems", *Elfa s.r.o., Košice*, pp. 467, 2002, ISBN 8089066402.

[4] B.Sobota, "Computer graphics and C language" KOPP, *České Budějovice*, 1996.

[5] S.A. Tanenbaum, M. Van Steen, "Distributed systems-Principles and paradigms", *Prentice Hall*, 2002 ISBN: 0132392275

[6] I. Wald, A. Dietrich, C. Benthin, "Applying Ray Tracing for Virtual Reality and Industrial Design", *Proceedings of the IEEE Symposium on Interactive Ray Tracing 2006*, pp. 177-185, Salt Lake City, USA, September 18-20, 2006.

[7] I. Foster, C. Kesselman, J. Nick, S. Tuecke, "The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration", Global Grid Forum, June 22, 2002. http://www.globus.org/alliance/publications/papers/ogsa.pdf

[8] R. Chellappa, "Digital image processing", IEEE Computer Society Press tutorial, 1992, ISBN 0818623616.

[9] B. Jähne, "Digital image processing", Springer, 2005,pp 607 ISBN 3540240357

[10] M. Nachtegael, D. Van Der Weken," Soft computing in image processing: recent advances", Springer, 2007, pp 500, ISBN3540382321.

Centralization of administration in academic computer network

[1]*Ing. Marek DOMITER,* [2]*Ing. Eva Danková,* [3]*Ing. Peter Jakubčo*

Dept. of Computer and Informatics, FEI TU of Košice, Slovak Republic

[1]marek.domiter@tuke.sk, [2]eva.dankova@tuke.sk, [3]peter.jakubco@tuke.sk

***Abstract*—Administration of computer network is for administrators difficult task. It is important that in large networks has been developed system for their administration, which makes administration of network services for administrators easier and will report network status. This paper deals with centralization, dynamic network devices management and unlike other computer networks periodical mass administration of users.**

The result is a system that was created in collaboration with the Faculty of Electrical Engineering and Informatics Department of Computers and Informatics and is used for administration of colleges computer network on the Technical university of Košice, which currently consists of about 150 network devices and 4000 computers.

***Keywords*—computer network, management, network devices.**

I. INTRODUCTION

Nowadays we encounter with information technologies daily. Every day there is a transfer of huge of information, which is made trough computer networks. This process is accompanied by many problems. If the described process should be controlled, a management system must be created.

Their main role is to ensure the smooth operation of network equipment and whole network too, which includes managing and monitoring network devices. Obtained information has to be transferred and processed. One of the solutions is implementing a centralized system. This approach includes centralized data storage and centralizing of services and tools.

There are large amounts of information systems, whose task is administration of computer network. But in these is absence of mass administration, which is required by model of academic student computer network. One of these systems was developed by Computer Network Laboratory – system Synets, which is designated for administration of computer network laboratory [4].

II. THEORY

A. Centralized system

Centralized systems use for their operating one logical node. In centralized system all programs are running on central location. [2] Centralized systems used nowadays and Legacy Systems work in centralized architecture with characteristic:

- Centralized system with multiuser processing – all intelligence is centered on central node
- Interaction via terminal or web browser – users communicate with central node trough computer network

Logical node consists of multiple physical computers which are redundant and creating cluster.

A **computer cluster** is a group of linked computers, working together closely so that in many respects they form a single computer. The components of a cluster are commonly, but not always, connected to each other through fast local area networks. Clusters are usually deployed to improve performance and/or availability over that of a single computer, while typically being much more cost-effective than single computers of comparable speed or availability. [5] There are two types of cluster: Load-balancing cluster and High-performance computing cluster [3].

By designing a centralized system is necessary to consider the ISO model for network management.

B. Model ISO for network management

System model for computer network management defined by The International Organization for Standardization – ISO/IEC 7498-4 consist of five areas [1] , which are the basis of the described system for administration of academic computer network:

- **Fault management** – encompasses fault detection, isolation and the correction of abnormal operation of the OSI Environment.
- **Accounting management** – enables charges to be established for the use of resources in the OSIE, and for costs to be identified for the use of those resources.
- **Performance management** – enables the behavior of resources and the effectiveness of communication activities to be evaluated.

- **Security management** – the purpose is to support the application of security policies by means of functions.
- **Configuration management** – identifies, exercises control over, collects data from an provides data to open systems.

III. DESIGN

A. System architecture

Proposed system for administration of computer network, which covers network devices maintanance and end devices - users computers mainstenance, consist of central data storage and network services. Network services maintain network and end devices or they are provided to users. Configuration data is stored in central data storage, what is the advantage

System design for administration of computer network is on Fig. 1.

Advantage of having configuration data for all services stored at central storage is independency of services and possibility of mass services setup by changing data at the central storage.

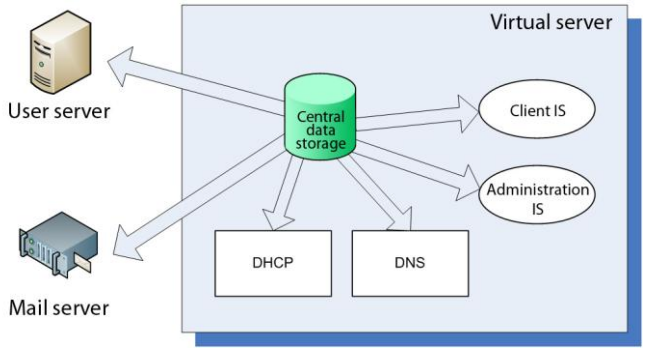


Fig. 1. All configuration data for administration of computer network are stored in central data storage. Data in data storage is maintained by information systems, network services are controlled by this data and user and mail servers use this data for authorization and authentication.

Unlike distributed systems, there is no needy of data synchronization, data recovery in case of communication failure and administration of multiple systems.

B. HW components

Since it is a centralized system, it is necessary to ensure high availability and sufficient computing power. The solution is to deploy more hardware devices that create a cluster.

In case of centralized systems is advisable to use combination of cluster solutions for acquiring high availability and high performance in one system.

Diagram of computer connections in HA cluster is on Fig. 2. It should be noted that described ways to ensure the high availability deal with the problem of failure of hardware components and software components, but not the operating system itself.

C. Data storage

Data storage in centralized system can by represented by any form. Information stored in centralized system is located on one place, so the chosen form has to be suitable for all data types collected and stored on data storage. Main data types are user's information and statistic information appertaining to users and computers.

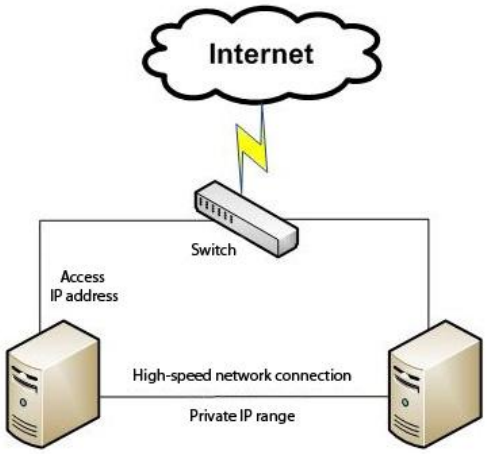


Fig. 2. Connection of computers in cluster, which provides high availability.

Given the wide range of data types, deployed in wide area network environments with huge amount of users and computers, it is preferable to use a relational database, over other solutions. As a particular type of database is used MySQL database.

The main advantages of a relation database in described system are:

- High reaction time to large amount of simultaneous requests – mainly insert operation (inserting of statistics information).
- Better opportunities of mass administration – multiple data editing and deleting.
- Easier deployment of communications environments IS – many of existing solutions of information systems offer only SQL database connection.
- Database consistency guaranteed by triggers and foreign keys (automatic statistic information deletion after deleting of particular user/device)
- Possibility to create view from multiple tables for needs of specific applications and services.

IV. SOLVING

A. Dynamic of system

As in academic sphere student computer networks are usually large and the resulting computer network consists of many sub networks that are managed by different administrators, it is necessary especially in case of centralized system to ensure its dynamism. The dynamical system consists of the possibility to do changes in network services directly from the IS interface without need to change any configuration or to restart provided network services. If there is no dynamic of centralized system, each administrator from sub networks have had central OS access for network services modifying and restart. This would be a problem for security and system continuity.

This problem is solved by DHCP, DNS and mail services modifying. Configuration of these services is dynamically loaded from central data storage and changes are stacked up immediately after actualization in data storage.

In case of DHCP IP addresses and configuration data are loaded directly from data storage. Recency of published DNS

information is ensured by database trigger which increment particular DNS zone serial number on information change in data storage. Next condition is automated network devices management.

B. Control of network devices

Conception of network devices control is going out directly from network structure and design. Network topology is displayed on Fig. 3. Automatically controlled switches, which are placed on network border, are directly connected to user's computers and other end devices. On these network devices is implemented dynamical setup configuration of each interface according to data placed in central data storage. Inter controlled setups belongs primarily interface turning on and off, configuration of belongings to correct community group - VLAN, configuration of maximal number of MAC addresses which can be connected to concrete interface, and editing or deleting of learned MAC addresses.

So that devices could be controlled and monitored, it is needed its initial configuration.

C. Configuring of network devices

For the needs of full use and the related monitoring initial configuration of network devices is needed. In large computer networks, manual configuration of each new device is time consuming and arduous. Also, subsequent monitoring of devices, which gather information about connected end devices appliances using the SNMP protocol and stored in a central data storage, provide overview of what is happening in the network. Device configuration can be divided into 2 pars:

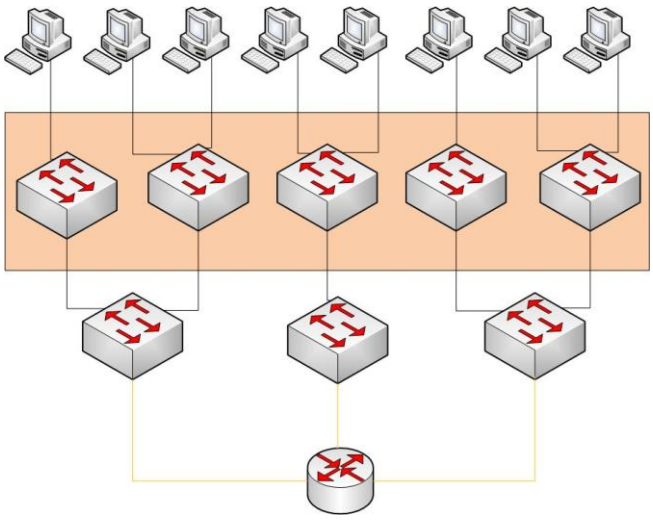


Fig. 3. Network topology with highlighted devices which are automated configured.

• **The basic configuration** - devices are accessible remotely

• **Automated configuration** - devices are configured by other applications

D. Basic configuration

Basic configuration of network devices consist of device connection to network and setting up device to remote connect. Then device has to be accessible and configurable from central system manually by administrator or automatically by applications.

This functionality is cover by:

- Setting up device IP address, network, virtual LAN id and default gateway.
- Setting up device virtual terminal and permit access from configuration server.
- Setting up SNMP access with rights for write from configuration server

E. Automatic configuration changes

As device is configured by described basic configuration, there are two possibilities to next configuration:

- **Partial configuration changes** – sequentially executed configuration commands, which gradually change device configuration
- **Change of whole configuration** - by upload of new configuration file

V. CONCLUSION

System for administration and monitoring of large academic computer network was designed regarding to environment in which the system was being created and is applied in. This environment consists of all colleges belonging to Technical University in Košice, which use this centralized system. The result is a centralized system that ensures continuous operation of computer network, monitoring and management of network devices and is implemented in the environment of information system for network administration associated with the central data storage based on SQL database MySQL. This system ensures a dynamic management of users - students and end devices – computers. Changes made via a web interface immediately reflect by network services in the computer network. The main approaches proposed for managing network devices is the use of automated management through the SNMP protocol directly from PHP environment of web information system. SNMP provides network management interfaces and message configurations for network devices.

By adding additional network services it is possible to dynamically expand an existing modular system and provide users additional accesses and services. The result can be applied in a large computer network, thus making network management easier for network administrators. System offers mass administration and life-cycle management of students and their computers in student computer networks.

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0073-07 and VEGA grant project No. 1/0026/10.

REFERENCES

[1] International standard ISO/IEC 7498-4, " Information processing systems – Open Systems – Interconnection – Basic Reference Model – Part 4: Management framework", 1989.

[2] US EPA, "Centralized vs. Decentralized Filing", http://www.epa.gov/, Nov 2009.

[3] Imrich, Jaroslav, "Failover cluster s protokolom CARP", http://www.linuxos.sk/, May 2007.

[4] Fecil'ak Peter, "Modular y based system for computer network management," Diploma work, KPI FEI TUKE, 2006.

[5] Bader, David, Robert Pennington, "Cluster Computing: Applications," The International Journal of High Performance Computing, 15(2):181-185, May 2001.

Automatic Set of Parameters of Energy Functional for Active Contours

Oľga DUĽOVÁ

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

olga.dulova@tuke.sk

***Abstract*— The active contours are nowadays a very popular segmentation technique. One of the ways how to find the optimal contour is a model which is based on the formulation of the minimization of the energy functional. The successful finding of solution depends on the choice of suitable weight parameters of the energy functional. As there are no exact rules how to set these parameters properly for the given picture, methods like "trial – mistake" are used. This article is focused on the automatic setting of the weight parameters through genetic algorithms, where the so called Greedy optimizing method is used.**

***Keywords*—Active contour, Genetic Algorithms, Energy Function, Greedy Algorithm**

I. INTRODUCTION

The task of the complete segmentation is the division of the picture on disjunctive areas which would –through their meaning- represent the objects in the picture. Different approaches are used for the segmentation: determination of threshold lines, growing of the areas, dividing and connecting of the areas, searching of borders and comparing with the sample. Many of these methods don't require any categorical information about the searched objects and they can offer very good results at the pictures without rustle or with a low level of rustle.

However, in the case when the quality of these pictures is low, these methods are not sufficiently exact. It means that the use of classical segmentation techniques either totally fails or it requires another step for the elimination of invalid borders. For the solution of such cases it is possible to use *active contours (snake*) which have been examined for such cases.

The active contours or Snakes have been presented for the first time by M. Kass, A. Witkin and D. Terzopoulos [1]. The active contour is a curve which aims under the influence of inner and outer forces to close the object which we want to obtain through segmentation. The inner forces influence the contour's smoothness during the deformation and the outer forces cause the movement of the contour in the direction towards the object's borders.

The active contours don't solve the whole problem of searching for contours in pictures. They depend on interactions with the user or on the interaction with a higher level of the picture's understanding. This interaction has to specify the approximate form and the start position for the active contour somewhere near the required contour[7].

According to the way of the curve's representation we can divide the active contours into two model types, parametric and geometric.

The classical parametric models were published in [1] for the first time. They are defined through the minimizing of energy functional which takes the minimum when the contour is smooth and stable on the object's border.

As an optimizing method for the optimizing of active contours based on the minimizing of the energy functional we can use the Greedy method. It is a method which is based on the searching around the point where the energy will be minimal and it moves to this position. This method is fast, flexible and stable. It is also effective, but it doesn't guarantee global optimizing.

The active contours don't solve the whole problem with the searching of contours in pictures. They depend on the interaction with the user or on the interaction with a higher level of the picture's understanding. This interaction has to specify the approximate form and the start position for the active contour somewhere near the required contour.

In this work is shown interpretations of simulation, which are acquired from the simulation program Matlab.

II. FORMULATION OF THE ENERGY FUNCTIONAL

The basis of Snakes based on energy minimizing [1] is to find a curve which will minimize the following energy functional:

$$E^*_{snake} = \int_0^1 E_{snake}(v(s))\,ds = \int_0^1 \{[E_{\text{int}}(v(s))] + [E_{image}(v(s))]\}ds \quad (1)$$

where E_{int} is the inner energy of the curve, E_{image} is the picture's energy (the potential energy).

The internal energy can be defined as

$$E_{\text{int}} = \alpha(s)\left|\frac{dv}{ds}\right|^2 + \beta(s)\left|\frac{d^2v}{ds^2}\right|^2 \quad (2)$$

The first term of the equation (1) receives large values in the case when the contour is not continuous. So it prevents the stretch of the contour. The bigger is the curving of the contour, the higher is the value of the second element of the equation.

The second element of the equation (2) is the functional of the potential energy and is calculated as the integral of the potential energy function alongside the curve V(s):

$$E_{image} = \int_0^1 E_{image}(v(s))ds \tag{3}$$

The function of the potential energy E_{image} is derived from the image data and it reaches smaller values on the object's border (in the case of following the gradient), or other attributes of interest.

If we use Greedy as the optimizing algorithm, we get the energy functional which is expressed like this:

$$E_{snake}(s) = \alpha(s)\left|\frac{dv_s}{ds}\right|^2 + \beta(s)\left|\frac{d^2v_s}{ds^2}\right|^2 + \gamma(s)E_{edge} \tag{4}$$

where the first and second derivation approximate every point searched in the surrounding of a chosen point on the contour. The weight parameters α, β, γ depend completely from the contour. That's why every point of the contour has interconnected values α, β and γ.

The first member of the equation (4) expresses continuity. At a common way of calculation it is only about the minimizing of the distance between the points which can cause clustering of points. This problem is in the Greedy algorithm [6] even worse as here the continuity is assessed only in a local way. There is a tendency to move the point to the previous point through which the distance from the following point gets larger. This causes a chain reaction, the movement of all points to their forerunners. It means that we will require from the algorithm that it keeps equal distances between the points and it shouldn't cause the clustering of the points.

We can express the energy of continuity through the following relation:

$$E_{cont} = \left(\overline{d} - |v_i - v_{i-1}|\right)^2 \tag{5}$$

Where *d* is the average distance of all points, $|v_i - v_{i-1}|$ is the distance between two actual points. Through this way we can achieve the situation that the distance between the points will be close to the average distance of points. The continuity value is finally normalized to the values from the range [0,1]. The new average distance of points *d* is calculated at every iteration.

The second member in the equation (4) is the curving. Its aim is to smooth the curve as much at it is possible. For the calculation the following relation is used:

$$E_{curv} = \left|v_{i-1} - 2v_i + v_{i+1}\right|^2 \tag{6}$$

The third member of the equation E_{edge} shows the gradient's size. The gradient's size can reach the values between 0 and 255. There is a significant difference between a point with a gradient's value of 240 and of for example 255. The values are normalized on the interval [0,1] according to the relation

$$(\min - mag)/(\max - \min) \tag{7}$$

Where *mag* is the gradient's value in the given point and *min* and *max* are the minimum and maximum from the surroundings of the given point. In the case that the difference max-min < 5 then for the minimum there is the maximal value reduced of 5. Through this we can prevent big differences in the gradient's value inside of big areas which are almost uniform. The Greedy algorithm will use the energies of the functionals so that it minimizes the compound functional [4]. This gives us the individual repetition in the development of the contour, where we can find all the points of the Snake. At first we assess the energy for each point and we remember it as the point's minimal energy. This will guarantee us that if a different point has been found whose energy is equally small, then the contour's point will stay on the same position. Afterwards, the 3x3 surroundings is being searched so that we can find out if there isn't another point with smaller energy in comparison with the determined point of the contour. If there is such a point, then it is integrated into the contour as a new point instead of the original point.

III. AUTOMATIC SETTING OF WEIGHT PARAMETERS

As we have already mentioned, the Snake is a curve which is being deformed through the influence of different forces. These forces must be balanced through the user who determines the segmentation object of the analyzed image. It means that he determines the weight, the parameters of the individual forces. In general, these parameters are determined through the method "trial – mistake".

In this work we will introduce the principle of the automatic setting of parameters through the use of genetic algorithms. Genetic operators such as crossbreeding, mutation and selection have been used for the contour's development [5].

A. Coding of chromosomes

The coding is a way in which the phenotype (the real representation of parameters) is transformed to the genotype (a chain of binary symbols). We have three real parameters which we need to code. At this method, it is necessary to know the field of the values for each parameter. We have $[\beta_{\min}, \beta_{\max}]$, $[\gamma_{\min}, \gamma_{\max}]$. If we want to assess the length, then it is necessary to define the maximal precision for each parameter, it means: $\Delta\alpha$, $\Delta\beta$, $\Delta\gamma$.

The chromosome's length will be given on the basis of the formula:

$$l = \left\lceil \log_2 \frac{\alpha_{\min} - \alpha_{\max}}{\Delta\alpha} \right\rceil + \left\lceil \log_2 \frac{\alpha_{\min} - \alpha_{\max}}{\Delta\alpha} \right\rceil + \left\lceil \log_2 \frac{\alpha_{\min} - \alpha_{\max}}{\Delta\alpha} \right\rceil \tag{8}$$

TABLE I
TEST PARAMETERS

	min	Δ	*Max*
α	0	0,01	1
β	0	0,01	1
γ	0	0,01	1

In our case the values have been encoded on chromosome lengths of 30 bit. We have used two-point crossbreeding. For the calculation of the mutations probabilities the following formula has been used:

$$\mu = \frac{1}{l} \tag{9}$$

It means that the mutation probability is 0,03 and this will guarantee that the chromosome will mutate in the least.

B. Quality assessment

At the measurement of the parameters' quality it is necessary to define the suitability of the chromosome. The suitability of the active contour can be given through the position of the contour's point or through the minimal global energy.

In our case we have used as the assessment of suitability the quality assessment according to the position of points in the Snake. It means, that we must have given the optimal contour, which can be obtained for example through methods of threshold setting, level-set or at complicated images through manual input of optimal points which we will compare with the solution that we get from the contour's deformation.

C. Setting of parameters

On the basis of the supervisional access of global solution we set the weight parameters for one image from the image group and we will use these parameters at the search of the optimal contour in the other images from the given group.

We place the points of the object's contour in the image manually and through this we define the optimal contour. We will look for the minimal area between the optimal contour and the contour obtained through the algorithm. The evaluative function of the genetic algorithm uses the Greedy algorithm for the determination of the parameters group's suitability. The algorithm will be calculated on 100 generations with 100 chromosomes. The parameters will converge during the individual generations. All parameters have the tendency to converge in the direction towards values with a higher or smaller precision [5].

IV. Algorithm for determination of global parameters

In the basic part we will create a genetic algorithm where as chromosomes there will be three parameters α,β and γ whose combinations will be encoded. We will calculate the suitability of the single individuals (of the parameter combinations) in two parts. At the beginning of the Greedy algorithm we will initialize the chromosomes and we will define the evolutional criterion of the final solution's quality. We will start the Greedy for every chromosome on the points of the initialization Snake. We will determine the single suitabilities through the quality criterion. Through the use of genetic operators selection, crossbreeding and mutation we generate another generation of individuals. We repeat this procedure for so long until the number of generations doesn't exceed the set limit or until we find the maximal suitability (the minimum summary of areas which are covered either only by the optimal contour or only by the Snake, which is zero pixels).

V. Experiments

The aim of the experiments which are introduced in this part was to show the development of the parameters during the evolution.

The evaluation of the solutions' quality was carried out on the basis of the following relation:

$$P(\alpha,\beta,\lambda) = \frac{S(v_{opt}) xor S(v_{snake})}{S(v_{opt}) + S(v_{snake})} \tag{10}$$

where $S(v_{opt})$ is the content of the area which is bounded through the optimal contour and $S(v_{snake})$ is the content of the area obtained from the Greedy algorithm for individual chromosomes.

In the evolutional algorithm was for the determination of individual solutions' suitability the simplified form sufficient:

$$P(\alpha,\beta,\lambda) = S(v_{opt}) xor S(v_{snake}) \tag{11}$$

as the algorithm is searching for the minimum of different areas in the same image.

In the fig. 1 there is the optimal contour which we want to achieve, this contour was obtained through the method of threshold setting. In the fig. 2 there are the points of the Snake at which we will start.

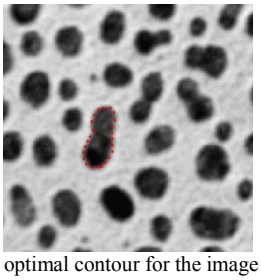


Fig. 1. The optimal contour for the image of a "cell"

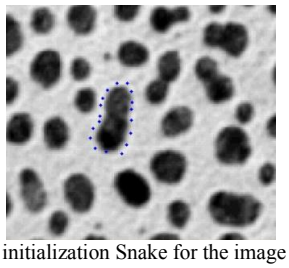


Fig. 2. The initialization Snake for the image of a "cell"

The evolution was activated on 50 randomly chosen individuals (chromosomes). As you can see in fig. 3, the suitability for the single individuals is in the range between 0 and 1000, which means that the difference between the optimal contour and the Snake is from 0 to 1000 pixels.

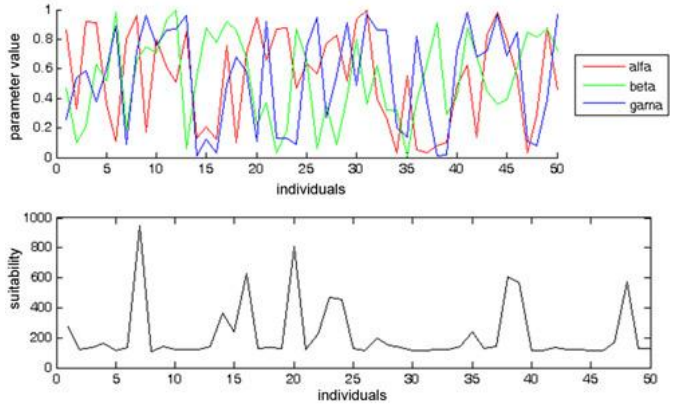


Fig. 3. Above: Initialization parameters, below: The suitability of initialization parametres

In the fig. 4 the development of the best individuals is shown during the evolution. The individual parametres change considerably and there is no linear releation between them. Fig. 5 shows how the suitability changes during the evolution.

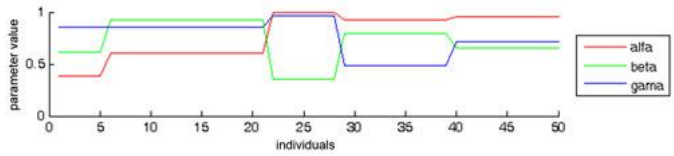


Fig. 4. Summary of the best parameter combinations in the generations

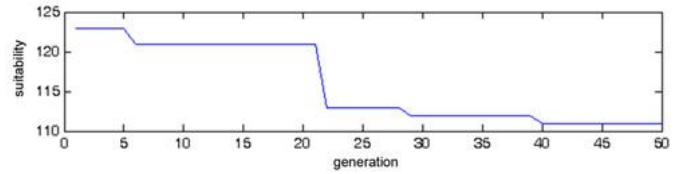


Fig. 5. The development of the best suitabilities in the individual generations

In the next picture a Snake after the activation of Greedy – throuhg the use of parameters obtained from the evolution - is shown.

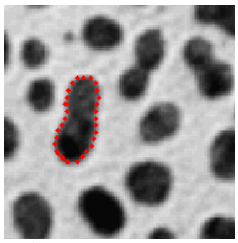


Fig. 6. The final contour of the evolutions, $\alpha = 0{,}960$, $\beta = 0{,}660$, $\gamma = 0{,}720$

At 50 generations the precision of over 97 % has been reached. In the table 2 the success of the best solutions during the evolution is shown.

TABLE I
SUCCESS OF THE BEST SOLUTIONS FROM THE INDIVIDUAL GENERATIONS

Image "cell"	Alfa	Beta	Gama	Success [%]
1.	0,39	0,62	0,86	92,44
2.	0,61	0,93	0,86	93,50
3.	1	0,36	0,97	94,92
4.	0,93	0,8	0,49	96,09
5.	0,96	0,66	0,72	97.54

VI. CONCLUSION

In this article we have shown the way of searching for the best parameters. The success of the solution was limited by the number of generations. This problem can be solved through restriction for the acquirement of boundary precision, however, this way can be often more time demanding. The use of the genetic algorithm at the automatic setting of parameters has its sense in the case if we want to use the calculated parameters on more similar images, as in the image on which the algorithm is learning the individual parameters, it is necessary to enter the optimal contour.

ACKNOWLEDGMENT

This work is supported by the VEGA project No 1/0386/08.

REFERENCES

[1] M. Kass, A. Witkin, D. Terzopoulos, Snakes: Active contour models, in *Proceedings of First International Conference on Computer Vision*, London, 1987, pp 259 – 269

[2] E. Demjenová, *Aktívne kontúry a ich uplatnenie pri segmentácii vláknitých objektov*, Písomná práca k dizertačnej skúške, Košice 2004

[3] M.S. Nixon, A. S. Aguado, *Feature Extraction and Image Processing*, An imprint of Butterworth-Heinemann, Linacre House, Jordan Hill, Oxford OX2 8DP, First edition 2002

[4] J. A. Tropp, *Greed is Good: Algorithmic Results for Sparse Approximation*, IEEE Transactions on Information Theory, Vol. 50, No. 10, October 2004

[5] J. J. Rousselle, N. Vincent, N. Verbeke, *Genetic Algorithm to Set Active Contour*, 10th International Conference Computer Analysis of Images and Patterns CAIP;2003, 25-27 aout 2003,Groningen, Hollande

[6] K.M.Lam, H.Yam, *Fast greedy algorithm for active contours*, Electronic Letters, 6th January 1994, Vol.30, No.1

[7] T. F. Chan, L. A. Vese, *Active Contours Without Edges*, IEEE Transactions On Image Processing, VOL. 10, NO. 2, February 2001

[8] P. Karch, *Graph cut segmentation*, In: SCYR 2009 : 9th Scientific Conference of Young Researchers : proceedings from conference. - Košice : FEI TU, 2009, ISBN 978-80-553-0178-5. pp. 162-165

[9] E. Ocelíková, D. Klimešová, *Preference ranking method for multi-criteria decision*, In: AEI '2009 : International Conference on Applied Electrical Engineering and Informatics : September 7-11, Italy, Genoa 2009. - Košice : TU, FEI, 2009 ISBN 978-80-553-0280-5, pp. 36-41.

Translation of Semantic Web Services Descriptions into a Planning Problem

Zoltán ĎURČÍK

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

zoltan.durcik@tuke.sk

Abstract— Semantic web services composition belong to relatively discussed theme in the area of semantic technologies and web services. Web services are programs accessible by network. In their descriptions (web service interface) is but exactly described how they work, how they are communicating together, and how we are able to access to this web services. Likewise would be necessary unify theirs interaction to certain standard communication protocol (e.g. SOAP). This allows interaction also between web services, which are programming in different languages. Best choices to automated web service composition appear to be artificial intelligence (AI) planners. Presented article show a possibility realizes automatic web service composition by using AI planners. In web service composition process is at first necessary transformations web service semantic descriptions into planning problem.

***Keywords*— composition, planner, OWL-S, PDDL**

I. INTRODUCTION

Web services are distributed programs located on networks, most frequently on internet. They are used by standard protocols, most frequently by HTTP (Hyper Text Transfer Protocol). Each of web service provides some function (for example translation from one language to another language). If there isn't possibility to achieve goal by one web service, we may try composition several web services together (e.g. in case, that we need translate word from one language to another, but don't exist for our request directly concrete service. Now we may try chain several web services.). For understanding web services composition problem, it is important understand web services standard and technologies. There is a briefly discussion of standard and technologies in following chapter.

II. WS STANDARD

Between most important standards, languages and protocols, which are related with web services, belong:

WSLD - Web Service Definition Language - is descriptive language, which serves to web service description with the aim of its correct usage and locates on networks. It enables abstract definitions of operations, which are given by service, and data, with which service works. Abstract definitions are them bonded to concrete protocols.

SOAP - Simple Object Access Protocol - is a protocol, which serves to communicating among web services over network (e.g. HTTP). This protocol was designed as platform and language (programming language) independent protocol, what means that it allows communicate between services programmed in various programming languages.

OWL-S - Semantic Markup for Web Services - goes out from OWL (Ontology Web Language), which is language for sharing and publishing ontologies. OWL-S description is ontology for web services description. It serves on more detail service description, its inputs, outputs, preconditions for web service execution, and effects after this execution. In more detail will be OWL-S described in chapter V.

III. AUTOMATED WEB SERVICE COMPOSITION

A draft of web service composition system is displayed on Fig. 1.

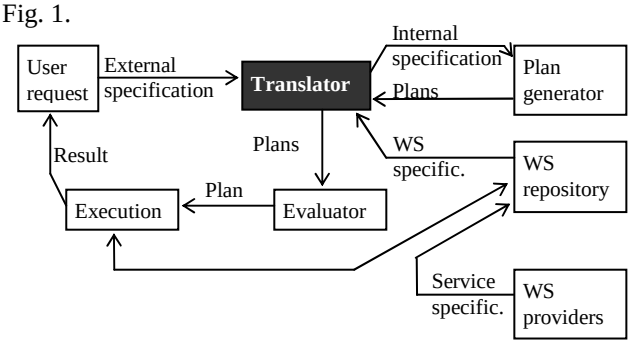


Fig 1. Automated Web Services Composition [4]

WS repository - serves to storing web services obtained from service providers.

Translator - serves to information processing obtaining from users and from web services repository. In many web services composition system is language by which user enter request, and a language which is used by algorithm to composition, different.

Process generator - is a set of algorithms, which choose atomic web services following the user request (query). These selected services next fulfill user request. At the end of this process is for us provided a set of atomic web services together with data and work flow among these web services.

Evaluation - occur in the case when are generated several different plans.

Execution engine - is a web services execution in order selected by planner. Result is provided to user.

It is important remember that the language, with which communicate users with system (i.e. external specification), and the language, which is used for composition–planning (i.e.

internal specification) are different. For external specification serves semantic standards as OWL, OWL-S, which describes web services, initial and goal state in really world. This information must be next transform into internal specification. Planner (i.e. plan generator) can work with this internal specification. Internal specification depends on type of plan generator. It may be plan generator based on Situation calculus, or Petri Nets, or planners based on AI planning methods as was mentioned above. At the last case we may use e.g. PDDL language for internal specification. The relation between OWL-S as external specification and PDDL as internal specification will be described in chapter VII.

IV. PLANNING WITH ARTIFICIAL INTELLINGENCE METHODS

As one of the more suitable choice for web service composition is use artificial intelligence planning methods [1, 7]. Planning problem can be represented as world model and it is possible write this model as pentad: **<S,S$_0$,G,A,Γ>. S** is representing a set of all possible states in given model, **S$_0$** is subset of S and marks initial state, **G** marks goal state of the planning problem. **A** is a set of available actions, from which each changes world state as passing from one state to another, and relation **Γ** is subset of SxAxS and define preconditions and effects for each action from A.

A relation between planning and web service composition is following: S_0 and G representing initial and goal states, which may be represented by ontologies, e.g. by OWL ontology. A is a set of actions and may be represented by available atomic web services. Γ is representing a state change function for each service. OWL-S service description presents itself as most suitable language for WS description and for directly connection with AI planning. By using OWL-S we can beside inputs and outputs directly describe also preconditions and effects.

Among most used planning methods belong [7]:
- state - space planning,
- graph oriented planning,
- hierarchical tasks networks planning,
- and planning by using logical programming.

V. OWL-S

Semantic Markup for Web Services (OWL-S) comes from OWL and it is ontology to web services description. OWL-S web service description (see Fig. 2.) consists from service profile, which describes what web services makes and what functionality provides, next from process model, which describes how web services communicates with clients, and from grounding, which specifies the ground properties of service as communication protocols, messages format, port type and likewise.

Operations are in OWL-S description represented as processes. There are tree kinds of processes [6]:
- atomic processes,
- composite processes,
- and simple processes.

Atomic process is process with one request message and returns one message as response. This process corresponds to action, for which is enough one interaction with web service. Atomic process is directly invocated. Atomic processes have no subprocesses.

Composite process is process, for which isn't enough one interaction with web service, and there is necessary more steps to execute this process. Composite process is decomposable into other (non-composite or composite) processes.

Simple process is used to provide abstraction view of some atomic process, or to simplified representation of some composite process.

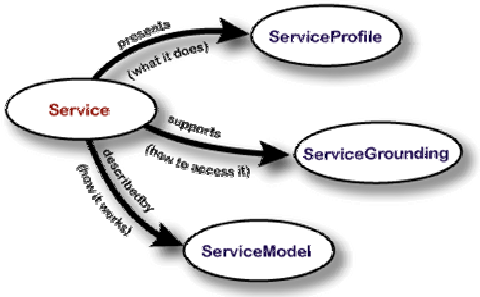


Fig. 2.Top level of service ontology [5]

Each process may include some properties, marked as IOPE – input, output, precondition and effect.

VI. PDDL

Planning Domain Definition Language (PDDL) was developed by Drew McDermott in 1998 for International Planning Competition[1]. Main goal of development this language was standardize language for planning domain. PDDL is a language for planning domain and problem description. It's inspired by STRIPS [9]. Likewise as STRIPS also PDDL use actions, with precondition, postconditions and effect. Precondition and postcondition serve to describe applicability this action in some state, and effect represent impact of execution this action on actual world, in which we accomplish this action. Planning task in PDDL language consist from two parts [2]:
- domain part
- and problem part.

These parts are located in separately files. World, in which we desire planning, is described in domain file. World knowledge is represented as predicates, and next this file includes also actions. Each action consists from parameters, preconditions and effects. Simple domain definition with one action you can see at Fig. 3.

```
(define (domain traveling)
  (:requirements :strips :equality)
  (:predicates (road ?from ?to)
    (at ?thing ?place)
    (person ?p)
    (vehicle ?v)
    (travel ?p ?v))

  (:action go
     :parameters (?person ?from ?to)
     :precondition (and (road ?from ?to)
        (at ?person ?from)
        (not (= ?from ?to)))
     :effect (and (at ?person ?to)
              (not (at  ?person ?from))))
```

Fig. 3. PDDL domain example

Now on fig. 3 we have five type of knowledge – predicates. "Road" is road (route) from one place (?from) to second place

[1] Drew McDermott - http://cs-www.cs.yale.edu/homes/dvm/

(?to). “At” is predicate, which inform us, that some person or thing (?thing) is located on place (?place). “Person” is predicate for person, “vehicle” for some vehicle (e.g. car, bus, and so forth), and predicate ”travel” inform us, that some person (?p) travel by some vehicle (?v). Now we have also one action in this domain definition. Action’s name is “go”, and means that we (“person”) need go from some place (“from”) to another place (“to”).

Actions in PDDL may be dividing in two classes:
- primitive,
- and composite actions.

Primitive actions are actions, which may be directly executed. Composite actions may be expansion into primitive actions or into other composite actions.

VII. Transformation OWL-S to PDDL

On fig. 3 we may see algorithm of web service composition based on PDDL.

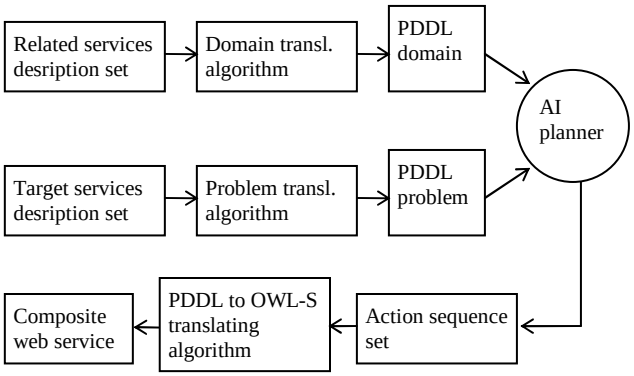


Fig. 3. Web service composition algorithm based on PDDL [3]

At the start of this process we have web services described by OWL-S language. From these descriptions we need obtain PDDL problem and domain definitions. After this transformation we input obtained definitions into some AI planner a resolve PDDL problem. Result from this subprocess is set of actions. Now we need translate this result into OWL-S composite process (again OWL-S). This process we can execute and return result to user.

As was mentioned above, OWL-S may contain three kinds of processes:
- atomic, composite and simple process.

For each process from this set we need transformation algorithm. Now we show a descriptions of some algorithm presented in [3]. There we have one function for each type of process. For example function translateAtomicProcess(k) translate process, which may be directly described by input, output, precondition and effect. There is an assumption, that each this process have or output, or effect, but not both [8]. Therefore this function is divided into two sub-functions, one for processes with effect, and one for processes with outputs. If we want translate composite process, we need invoke function according to control construct used in this process (e.g. if-then-else, sequence, choice and slit). For other algorithms see [3].

```
Input: OWL-S process model set K
Output: PDDL action sequence set AS
AS = 0;
For each atomic process k ∈ K do
  A = TranslateAtomicProcess(k);
  add A to AS;
End
For each atomic process k ∈ K do
  A = TranslateSimpleProcess(k);
  add A to AS;
End
For each atomic process k ∈ K do
  A = TranslateCompositeProcess(k);
  add A to AS;
End
```

Alg. 1. Translation OWL-S processes to PDDL actions

VIII. Conclusion

Presented article is allocated on possibility of web service composition by using AI planners. Beside this approach are also other approaches, e.g. situation calculus or composition by using Petri nets. But AI planners was shown as most suitable choice for web services composition, mainly for directly connections between planning problem and web services semantic description. A web service described by OWL-S is possible directly mapping on PDDL planning problem. Given transformation is of course only part of complete web service composition problem. This includes more additional problems, as e.g. interaction with users, web services management, knowledge management and so forth.

Acknowledgment

This work was supported by the Slovak Grant Agency of Ministry of Education and Academy of Science of the Slovak Republic under grant No. 1/0042/10 and is also a the result of the project implementation Development of Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120030) supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] Evren Sirin and Bijan Parsia. Planning for semantic web services. In Semantic Web Services Workshop at 3rd International Semantic Web Conference (ISWC2004)
[2] Ghallab, Howe, et al., PDDL – The Planning Domain Definition Language, version 1.2, AIPS-98 Planning Competition 1998
[3] Bo Yang, Zheng Qin, Composition Semantic Web Services with PDDL, Information Technology Journal 9(1), Asian Network for Scientific Information, 2010.
[4] Rao, J. – Su, X.: A Survey of Automated Web Service Composition Methods. In Proceedings of the First International Workshop on Semantic Web Services and Web Process Composition, SWSWPC 2004, San Diego, California, USA, 2004.
[5] OWL – Ontology Web Language - http://www.w3.org/TR/owl-features/ (2004)
[6] OWL-S – Semantic Markup for Web Services - http://www.w3.org/Submission/OWL-S/ (2004)
[7] Peer, Joachim: Web Service Composition as AI Planning - A Survey, Dissertation, University of St. Gallen, Switzerland, 2005.
[8] Evren Sirin, Bijan Parsia, Dan Wu, James Hendler, and Dana Nau. HTN Planning for Web Service Composition Using SHOP2, In Journal of Web Semantics, 2004
[9] NILSSON, Nils J. – FIKES, Richard E.: STRIPS: A New Approach to the Application of Theorem Proving to Problem Solving, Artificial Intelligence, 2(3):189-208, 1971.

Map building based on visual information from one camera

[1]*Juraj EPERJEŠI,* [2]*Miron KUZMA,* [3]*Jaroslav TUHÁRSKY*

[1,2,3,4,5,6]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]juraj.eperjesi@tuke.sk, [2]miron.kuzma@tuke.sk, [3]jaroslav.tuharsky@tuke.sk

***Abstract*—Map is the information in some form, which can be used to decide, if the robot can be positioned in particular space or not. Presented system provides the robot with the ability to create such map. Using visual information allows the system to do other operations with gained image and for example do some object recognition.**

***Keywords*—image processing, map building, mobile robot, vision**

I. Introduction

System uses movable camera for creation of map of environment for mobile robot. This map provides the robot with the information about its position in environment and possible locations to which it can go. Path planning is the basis for later more complex operations.

This system is implemented in MASS platform and therefore is not dependent on particular type or construction of robot. Nevertheless, several parameters must be set for the system to work with different robot.

II. Basic Construction of Robot

Basis of the system is IP camera movable in two different directions. Now used camera has range of horizontal turning from -171° to 171° and in vertical turning from -17° to 83°. Vertical and horizontal viewing angles are 40.5° and 54°. Height of focal point of camera above the floor is 148 mm.

For mobile platform is used the two-wheeled chassis assembled from LEGO Mindstorm kit.

Communication with camera is through wi-fi and with the chassis through Bluetooth.

III. Functional Scheme of Image Processing

- Image gained from camera
- Edge detection
- Determination of significant edge
- Filtration of unnecessary and erroneous points
- Correction of position of these points in image
- Conversion from image coordinates of point into map coordinates

This sequence is repeated for several images taken in different angles of camera and all points are then put into map.

A. Gaining of Image

The image is gained through CGI scripts from IP camera as often as is possible (as fast as system can process them). The system is not working with the video stream, but with still images. Now the system is working with images with resolution 320x240 pixels. The camera is capable to send images in resolution 640x480 pixels, but the system is then very slow.

B. Edge Detection

Gained image is processed by edge detector. Chosen was canny edge detector. This detector uses two masks moved through the image which count the gradient of brightness in horizontal and vertical direction. Based on these values, detector counts the direction of edge, which is then rounded into one of four basic directions.

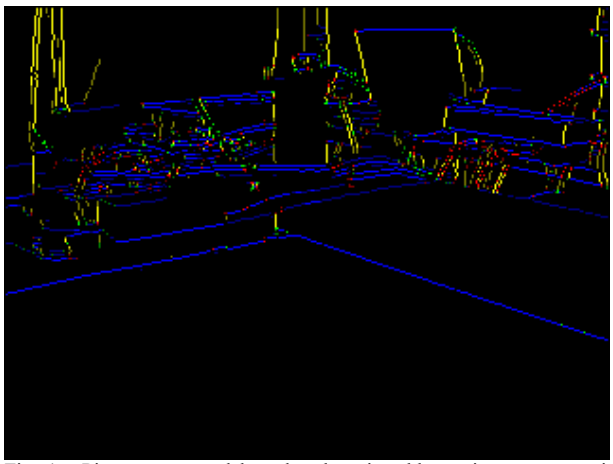


Fig. 1. Picture processed by edge detection, blue points are recognized as horizontal edges, yellow as vertical. Green and red pixels are detected as oblique in one, or another direction.

Several detectors were tested and as best was chosen Canny detector, because of its use of hysteresis. This practically means, that after first detection of edges by masks, masks are run through the image one more time, to find the points, where the gradient exceeds second threshold, which is lower then the firs one. These points must be connected to the already detected points. This allows for completion of edges,

where brightness may vary slightly while not detect points, which are unnecessary.

Edge detector in this system is slightly modified, because basic canny edge detection works with B/W images and this one works with color images. It counts the gradient for all colors, and if only one of them exceeds the threshold, the point is marked as edge.

C. Determination of Significant Edge

System works based on several prerequisites. That the floor is flat, with the plain regular color and that all obstacles are placed directly on the ground. These obstacles also have different color then the floor.

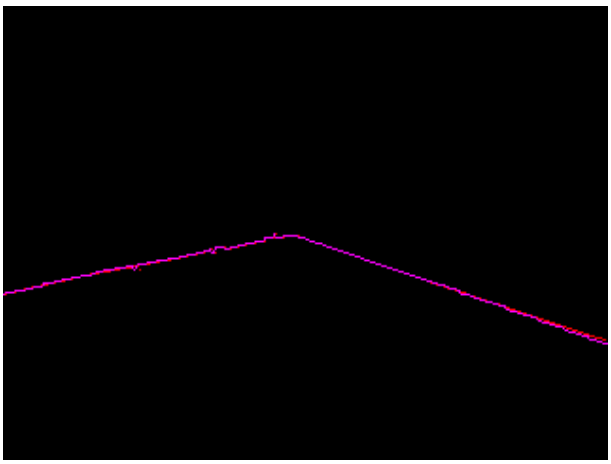
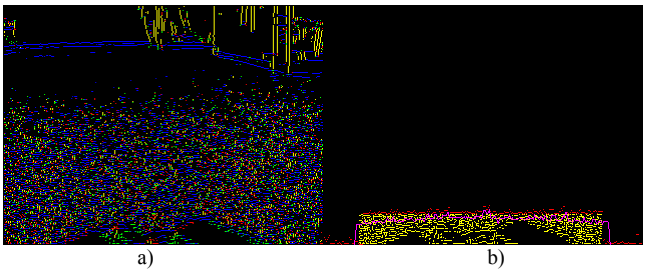


Fig. 2. Chosen significant edge. Red pixels represent the actual edge, and purple pixels are pixels after removing of radial distortion.

According to these prerequisites, closest obstacle creates the lowest edge on the image. Therefore, in the image with detected edges, all edges above the lowest one are erased. Based on construction of chassis and camera, closest detectable edge is about 20 cm from the center of the robot.

Another problem with this edge is that if the camera is rotated in particular angles, part of robot's own construction is visible on images in lower part. This leads to the detection of non-existent obstacle very close to the robot. Problem was solved by definition of ranges of angles in which the obstacles are detected from particular distance. These ranges were determined experimentally and are applicable only for this particular construction. Illustration of these ranges was done on highly textured floor, where almost all pixels around robot were detected as edge pixel with the exception of pixels located on robot's construction.



a) b)

Fig. 3. Definition of angles where construction of robot is visible.
a) edge detection on highly textured floor
b) chosen edge is higher than the construction

D. Filtration of Erroneous and Unnecessary Points

To determine the line segment, only the points on both ends are necessary. The chosen edge has still all of the detected points present. Basic condition is, that only points, where the difference of y coordination of point changes according to the foregoing difference.

Sometimes the edge is not detected correctly, for example, when there is shadow, or if the color of obstacle is close to the color of floor. Then errors, where one point of edge is not detected occur. The change in difference is significant but such point is clearly unusable by system.

Importance of this step is also in decreasing of number of points for which later computations must be performed which leads to faster run of the system.

E. Radial Distortion Correction

The objective of camera is round and the sensor is rectangular, which creates slight bend of image, which is most significant near the edges of the image. This effect can be easily removed by functions:

$$P'_X = P_X\left(1 - a_X\|P\|^2\right)$$
$$P'_Y = P_Y\left(1 - a_Y\|P\|^2\right) \quad (1)$$

Where P_X and P_Y are normalized coordinates of pixel in such way, that in the middle of the image are coordinates 0,0 and a_X and a_Y are coefficients determining the bend of the image in the direction of given axis.

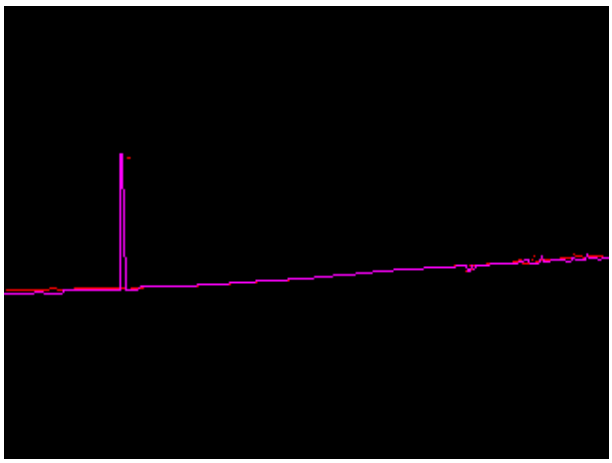


Fig. 4. Erroneous point detected. One pixel of the edge was not detected as edge and the second edge above the first was found.

F. Conversion of Coordinates from Image to Map

Based on the x coordinate of pixel in image and the horizontal angle of camera, angle between this point and the axis of the robot is calculated. Accordingly, from the y coordinate of pixel in image and the vertical angle of camera is calculated the distance from the center of the rotation of the camera.

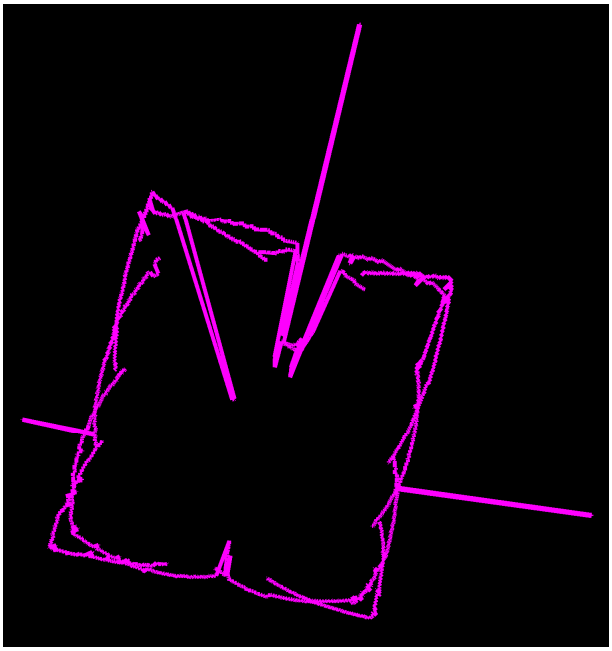


Fig. 5. Final map with visible problems. Three pixels on the edges were not detected correctly, left front part of robot's construction is visible and radial distortion is not solved yet.

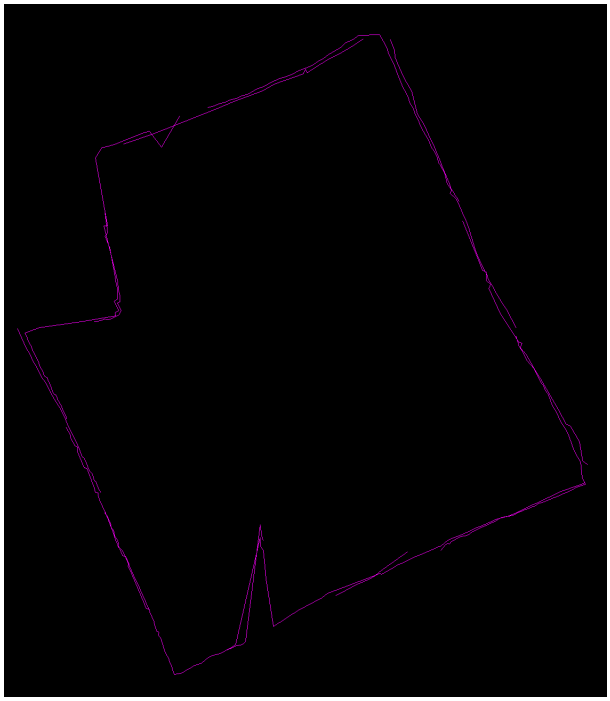


Fig. 6. Actual map with erroneous and unnecessary point filtration. The irregularity in lower part is electrical cable for the robot.

IV. Experiments

In the beginning, the experiments were focused on finding the precise coefficients for the used construction of mobile robot.

First determined constants were those concerning radial distortion. Then, the horizontal and vertical viewing angles were measured, followed by the size of steps in camera's movements. In this phase, only one image from camera was used. After that, the areas with visible robot's construction were defined as well as all previous parameters were tested on panoramic view created from several images. Current number of images is 12, which provides overlap approximately 50% between neighboring images.

One systematic error occurred during experiments. It was causing situation, where all simple views were good, but they did not fit together. As was later found out, the center of the image is slightly shifted against center of the camera by approximately 3°. After importing of this information into calculations, component parts of the map fit together.

After this point, filtering of unnecessary and erroneous points started.

V. Future Work

Next objective is to filter all points except the ending points of line segments from the map. This means, that also the points, which are found twice because of the overlap of images, should be processed and only one should remain.

Final step is to have two different maps, which overlap in some area and be able to connect them and create one bigger map from them. This will allow the map to be build incrementally.

VI. Conclusion

This article provides basic view of the system and how it is working. Also reveals steps leading to setting of several constants. Description of possible errors and their effect on final map is present too.

Acknowledgment

This publication is the result of the project implementation Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] Canny, J.: A Computational Approach to Edge Detection, IEEE Trans. Pattern Analysis and Machine Intelligence, 8:679-714, 1986
[2] http://en.wikipedia.org/wiki/Canny_edge_detector
[3] http://local.wasp.uwa.edu.au/~pbourke/miscellaneous/lenscorrection/
[4] http://www.pages.drexel.edu/~weg22/can_tut.html

IN. TRA. NET. - project for distance vocational education

[1]Daniel FÁBRI, [2]Martin SEKERÁK, [3]Marián CHOVANEC, [4]Chiara SANCIN, [5]Maria RICCIO

[1, 2, 3]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic
[4]Dida Network, Srl., Rome, Italy
[5]Faculty of Engineering, University of Sannio, Benevento, Italy

[1]dfabri@azet.sk, [2]martin.sekerak@tuke.sk, [3]marian.chovanec@tuke.sk, [4]csancin@gruppodida.it, [5]riccio@unisannio.it

***Abstract*—The paper presents a new project for distance vocational education and training for small and medium enterprises (SME) workers that integrates technology of remote laboratories, Learning Management Systems (LMS) and real technology control focused on current education and training. The method has been developed within international project IN.TRA.NET that is supported by European community. The idea of such a vocational education and training comes from experiences from previous experiences with remote laboratories for teaching university students some topics in area of measurement and electronics. The application of the remote laboratory technology, LMS structure and real technology is a new direction in SMEs workers vocation training that may decrease costs needed for their continual education.**

***Keywords* — Distance learning system, Remote laboratory, SME workers education,**

I. INTRODUCTION

Present workers in small and medium enterprises (SME) are required to be familiar with continuously developing modern and complex electronic apparatus as electronic measurement instrumentations (waveform generators, oscilloscopes, FFT analyzers) and control devices (PLC, numerical control machine, etc.). These apparatus are used not only in the automotive sector or in the telecommunication sector but also in all enterprises that have an automatic control of own production lines. On the other hand electric and electronic devices are evolved and renewed very quickly. For these reasons the wide diffusion of complex and last generation electronic apparatus includes the necessary updating for SME technicians because in this way the SMEs can acquire specific and innovative skills to improve own competitiveness. The upgrading activities unfortunately requires great commitment by the professionals involved who are forced to move away from their job and then to interrupt their productivity.

The objectives of the innovation transfer network (IN.TRA.NET) system are to create some specific learning services to educate professionals and workers and to give them a learning tools based on remote control of real industrial instrumentations and apparatus through e-learning technologies and methodologies. In this way IN.TRA.NET system can facilitate life-long learning activities of specialized technicians, especially in the field of process control, quality control and test engineering. By remote access to such operative equipments it is possible to:

- repeat and use the operative equipment more and more times and in real conditions;
- be trained on them even before they are available in the company (or they are available only in few sites or the Corporate);
- improve or update their technical skills related to the operating processes using those equipments;
- contribute to the some sector innovation processes.

II. ANALYSIS OF SMES' NEEDS

The first stage of the IN.TRA.NET. project covered research and analysis of SMEs needs. The User Need Analysis has been performed in all the European Countries involved in the Project Consortium. – Italy, Slovakia and Spain. The User Needs Analysis was required to identify the specific needs of SME workers and technicians concerning continuous updating activities for the acquisition of new and more skills in managing complex and latest generation instrumentations. The analysis was specifically finalized to the definition of what type of apparatus each SME has been interested to make it manageable in the IN.TRA.NET environment, what type of specific activity had to be realized and finally what type of specific theoretical contents had to be considered for the development of the learning contents.

The User Needs Analysis process was divided into three fundamental steps:

- detection of SMEs staff types convenient to be involved in the system;
- selection of some SMEs to be involved in the experimental activities related to INTRANET project ;
- individual interview with the each involved SME to detect their specific needs related to the IN.TRA.NET objectives.

The first step was conducted by the research group considering the main local industries and productions. This activity led to the identification of the following three main interest areas :

- Engineering industry;
- Agro-industrial sectors;
- Photovoltaic sector.

The second step of the User Needs Analysis selected four local industries, among theme participated to the first two steps, in accordance with the previous defined working areas and the results of the received questionnaire.

The last step was based on single meeting with each identified industry and the IN.TRA.NET. local partners to better know their specific needs related to the proposed project and to define their specific role and involvement in the future experimental activities.

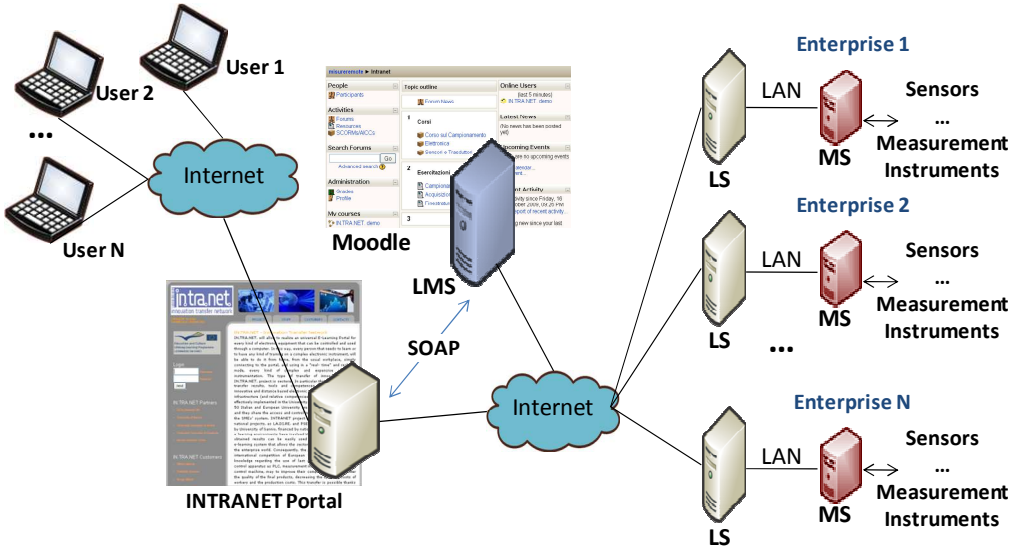


Fig. 1. General architecture of the IN.TRA.NET. environment

III. IN.TRA.NET TECHNOLOGY

The main design objectives, were defined by the research group in accordance with the general main objective of the IN.TRA.NET project that are the implementation and adaptation of the didactic distance learning services and methodology based on remote control of electrical apparatus and equipment to specific needs related to VET System as follows:

(i) portability: the visualization environment has to be portable on different hardware and operating systems;

(ii) usability and accessibility: the visualization and the management of an experiment have to be easy to understand and to perform, even for users that are not expert of information technologies, and the system features have to be accessed easily by students operating at University laboratories or at home;

(iii) maintenance: the maintenance costs should be reduced. This can be made possible through a client-server approach that eliminates the need for installing and upgrading application code and data on client computers;

(iv) client-side common technologies: students have to access to the system using their desktop computers based, with no need of powerful processors or high memory capacity, and connecting to Internet through low speed dialup connections;

(v) security: the remote access of the students through the Internet must preserve the integrity of recorded and transmitted data and of the system as a whole;

(vi) scalability: the system performance has not to be affected when connected users increase. Most of the proposed VLs cannot be considered an effective platform for delivering distance education.

The main functional requirements need to complete the access procedure to IN.TRA.NET system and to access to the specific services for remote control and remote visualization is composed by the following functionalities:

1. connection to the link;
2. authentication phase;
3. choice of the course;
4. choice of the didactic activity;
5. setting of some parameters concerning the operative system version and the video resolution;
6. visualization of the apparatus or equipment involved in the selected activity thanks to specific video capturing devices.
7. access to the specific activities for the remote control and monitoring of the involved apparatus.

The LMS is executed on a central server located at University of Sannio (Fig. 1). A Laboratory Server (LS) interfaces each experiment with the rest of the distributed architecture. There is a LS for each enterprise involved in the project. The Laboratory Server is the only machine in a measurement laboratory directly accessible through the Internet, while the other server machines constitute a private local network. For this reason the LS can also be used for security purposes in order to monitor the accesses to the measurement laboratory and to protect it against malicious accesses.

A Measurement Server (MS) is the server located in the enterprise that enables the interaction with one or more instruments or sensors depending on the specific experiment. A MS is physically connected to a set of different electronic measurement instruments or sensors by means of proper interfaces.

The preliminary results concerning the general design were achieved using architecture shown in Fig. 1.

The system is based on a web portal which interfaces with the LMS (Learning Management System), in order to take from it the contents and experiments to be provided to the users. The web portal allows making the system transparent to the user, so they have the support of the functionalities provided by the LMS, but without using it directly. The LMS is then connected to all the laboratories inside the university or in the company sites, where the experiments, instrumentations and industrial processes are located.

A. *Moodle as Learning Management System for the IN.TRA.NET. project*

The LMS has to be based on the client-server application model to deliver both administrative services and didactical facilities for the learning process. The administrative services allow to manage learners and to allocate learning resources such as registration, classroom, instructor availability, instructional material, fulfillment and on-line delivery. The LMS provides an infrastructure that can be used to rapidly create, modify, and manage content for a wide range of learning.

Moodle was chosen as the most suitable LMS platform to meet the requirements of IN.TRA.NET. project for many reasons:

- Moodle is a license free open-source software platform, users are free to download it, use it, modify it and even distribute it under the terms of the GNU license.
- It can be used on all servers that can use PHP such as Unix, Linux and Windows. It uses MySQL, PostgreSQL and Oracle databases, and others are also supported. Users can download and use it on any computer and can easily upgrade it from one version to the next, customize its options and settings.
- Moodle's infrastructure has a strong support for security and administration.
- It has multi-language support, this can encourage the cooperation between the different partners.
- It is modular in its design, supports authentication methods, activities, resource types, different data field type and can readily be extended by creating plug-ins for new functionalities.

This last feature is central to the design of the IN.TRA.NET. system since the platform, on one side has to deliver the typical functionalities of a common LMS, but on the other side it has to support innovative features related to the experiments delivery and remote control.

The greatest strength of Moodle is the community that has grown around it. Moodle's development has also been assisted by the work of open-source programmer community. This has contributed towards its rapid growth and continuous enhancement.

B. *Moodle Web Services and the IN.TRA.NET. Portal*

The INTRANET portal gives access to the system after an authentication procedure. Users access to the enterprise demos or experiments through standard Web browsers, without the necessity of specific software components on client-side.

The portal provides all the functionalities of the IN.TRA.NET. system, by means of a standardized interface to the LMS (Moodle).

In this phase, extending Moodle to use Web services allows the IN.TRA.NET. users to access their dedicated services and contents through the project portal in a friendly and easy way.

Interesting advantages can be planed for the spreading of the IN.TRA.NET. project when Moodle Web Services could give, for example, the possibility to share resources even though hosted on different LMSs, in different countries, by means of an unique account on the web portal of the project.

The software design of the system plans:

- to add to Moodle a module for the support of Web Service;
- to use the Java technology (servlets and JSPs) to realize the INTRANET web portal to access Moodle through SOAP over HTTP.

Simple Object Access Protocol, SOAP, is a lightweight protocol; its syntax is based on xml, for exchanging information in a distributed environment. The web services module resolves the lack of crucial classes in Moodle, for example User, Course, etc., their associated functions e.g. getMyCourses(), getResourses(), and processes SOAP requests from any clients or application that supports this protocol.

The IN.TRA.NET. portal will be realized by using Java technology to create dynamic web content. Java Servlets and JSP pages provide the dynamic extension capabilities for the portal. Apache Axis is the SOAP engine for the Moodle remote web services.

IV. 4. General identification of the IN.TRA.NET. services

An important task of the design phase was to identify the specific requirements of the experiments and demos that should be exported by means of the IN.TRA.NET. system. The companies involved in the IN.TRA.NET. experimental activities are SMEs of the Benevento province operating in different fields.

Coordination meetings with the target companies has been finalized (i) to better know their specific needs, (ii) to specify their specific role in the project and (iii) to define the application ambits and the specific service to be provided.

In general, the traced activities can be summarized in the five points below:

- Distance training of professionals on actual instrumentation, including support of lectures, experiments and tests for the verification of the acquired knowledge;
- Remote demo services, to be provided to companies which want to demonstrate their own products;
- Remote control of the production process quality;
- Remote control of the performance of a system or a process.

V. Conclusion

The knowledge that we gained during the realization of this project, is focused on the use by students too.

There are three main ways of getting the information, what we encountered while in distance learning or in an exploring process.

1. Deeper knowledge verification in this field.
2. Effective utilization of the exploring process regardless of time.
3. Real utilization, i.e. real instruments.

To ensure the proper functioning of real laboratories controlled remotely we need to put stress on the following criteria.

- remote control for this laboratory
- safe environment for people and laboratory equipment
- functionality of laboratory without controlling personal
- data transfer and its visualization

In this experiment we got experience where we put stress on system safety. This safety must be guarded against unpredictable manoeuvre that could damage the system, while leaving full control. Also it must be ensured for the person to be able to learn from his mistakes and give him opportunity to do so.

We got to a conclusion that using these steps students is motivated to self study. In this step students have courage to try new procedures without the risk of failing.

The listed knowledge are the basic properties where we emphasis independence, which opens new ways to research in our project.

At the moment the research group is working in extending Moodle to use Web services. Interesting advantages can be planed for the spreading of the IN.TRA.NET. project when Moodle Web Services could give, for example, the possibility to share resources even though hosted on different LMSs, in different countries, using an unique account on the web portal of the project. After the conclusion of the design phase the research group will work to the development of the chosen services, taking in account the results of the user needs and the specific requirements defined during the design activities.

ACKNOWLEDGMENT

The work is a part of project IN.TRA.NET. LDV/TOI/08/IT/493 supported by Leonardo Lifelong Learning Programme (50%).

This work is also the result of the project implementation Development of the Center of Information and Communication Technologies for Knowledge Systems (project number: 26220120030) supported by the Research & Development Operational Program funded by the ERDF (50%).

REFERENCES

[1] N. Ranaldo, S. Rapuano, M. Riccio, F. Zoino, “A Remote Laboratory for Electric Measurement Experiments: The Remote Displaying of Instruments”, Proc. of “The 19th Metrology Symposium”, Abbazia, Croazia, 26-28 Sept. 2005.

[2] N. Ranaldo, S. Rapuano, M. Riccio, F. Zoino “On the use of video-streaming technologies for remote monitoring of instrumentation”, Proc. of IMTC, Sorrento, Italy, 2006 pp.861-867.

[3] N.Ranaldo, S.Rapuano, M.Riccio, F. Zoino, “A Remote Laboratory for Electric Measurement Experiments: The Remote Displaying of Instruments”, negli atti della Conferenza Internazionale “The 19th Metrology Symposium”, Abbazia, Croazia, 26-28 Settembre 2005.

[4] S. Rapuano, F. Zoino, “A learning management system including laboratory experiments on measurement instrumentation”, IEEE Trans. On Instrumentation and Measurement, vol.55, N.5,2005, pp.1757-1766.

[5] P. Daponte, S. Rapuano, M. Riccio, F. Zoino “Remote Didactic Laboratory in Electronic Measurements: Quality of System Testing”, submission IMTC-2007, 1-3 May, 2007, Warsaw, Poland

[6] M. Drutarovský, J. Šaliga, I. Hroncová “Hardware infrastructure of remote laboratory for experimental testing of FPGA based complex reconfigurable systems”, Acta Electrotechnica et Informatica. - ISSN 1335-8243. - Vol. 9, No. 1 (2009), pp. 44-50

[7] M. Drutarovský, J. Šaliga, I. Hroncová, L. Michaeli “Remote laboratory for FPGA based reconfigurable systems testing”, IMEKO 19 World Congress : Fundamental and applied metrology : proceedings. – Lisboa, Portugal: SPMet, 2009. - P. 54-58

DWT based video watermarking

[1] *Peter GOČ-MATIS, Tomáš KANÓCZ, Radovan RIDZOŇ*

[1]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]peter.goc-matis@tuke.sk, tomas.kanocz@tuke.sk, radovan.ridzon@tuke.sk

Abstract— This paper describes a new method for video watermark embedding, using the knowledges already available from watermark embedding into digital static pictures. The watermark is embedded in transformed domain using Discrete Wavelet Transform (DWT). This paper also describes experiments conducted on the proposed watermark embedding method. The goal of these experiments was the test the robustness of the method presented in the paper against several watermarking attacks.

***Keywords*— Watermarks in video, attacks, transformation domain, DWT.**

I. INTRODUCTION

In recent years, there has been a rapid progress in the digital multimedia processing as well as in the internet technologies. Analog form of multimedia was practically replaced by the digital form of multimedia almost in the all the areas of the human life.

Digital multimedia has brought many advantages in comparison to the analog form of multimedia. The main advantages of digital multimedia form are easy processing and storage, compression and better noise resistance. Digital multimedia also has brought disadvantages. For example easy copying without quality degradation of copied multimedia. The copies are identical with the original multimedia and they can be transmitted over the worldwide internet. This illegal sharing is wrong for authors and distributors of the multimedia because they lose income.

Also with the progress of the peer-to-peer networks and fast internet the problems with the author's rights and copyright protection were established.

Approaches for multimedia content protection can be divided into two main groups [1], [2]:

- multimedia content protection during transmission,
- multimedia content protection after transmission.

Solution of multimedia content protection during transmission is the use of cryptographic methods which are based on content encryption of multimedia.

This paper deals with multimedia content protection after the transmission using digital watermarking.

II. DIGITAL WATERMARKING

Digital watermarking is a technique of embedding additional information into all kind of digital multimedia, whereby this data modification should be imperceptible [3], [4]. The embedded watermark carries information about author or distributor of multimedia and also provides data integrity check.

The form of embedded watermark may be symbol or number sequences, image information (logo) or segment of vocal signals. It depends on an application. The three basic parameters of digital watermarking are robustness, perceptual transparency and capacity [5].

A. Digital watermarking in video

The digital watermarking methods in video can be divided into three main groups [6]:

- methods based on watermarking in still images,
- methods based on video-time dimension,
- methods based on video compression standards.

B. Attacks on digital watermarks in video

Attacks on video watermarks represent all intended and unintended operations, which are executed by attacker with a goal to remove watermark from marked multimedia and get possession of unmarked content. Specific attacks applied to video are frame dropping, frame swapping, frame averaging, statistical analysis and unintended attacks for example compression of video sequences and affine transformations [7].

This paper focuses especially on the unintended attack of video compression and also on the intended attacks which are frame averaging, frame dropping and frame swapping.

III. THE PROPOSED WATERMARKING METHOD

The proposed method which performs watermark embedding into video content is based on Discrete Wavelet Transform (DWT). Reasons for the usage of this orthogonal transformation are its good results in applications which deal with image processing. The described method is based on watermarking in still images.

A. The watermark embedding block

Block of watermark embedding performs: video content loading, decomposition to still images, wavelet decomposition, watermark insertion, wavelet reconstruction and reconstruction the video content from still images. Inputs of this block are original video and embedded watermark. In Fig. 1 the watermark embedding block is shown.

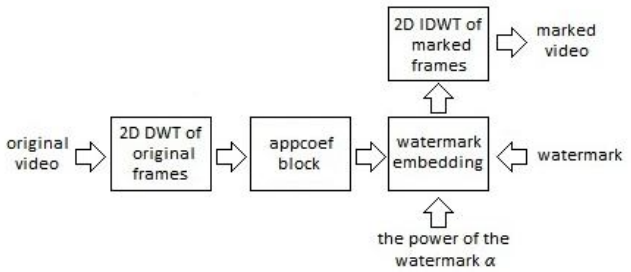


Fig. 1 **Watermark embedding block**

Six video sequences with different dynamic properties were used in experiments. These were (Dynamic – D, Mezzo Dynamic – MD and Lightly Dynamic – LD). Resolution of all video sequences is QCIF 352x288 pixels. Videos are cut into 10 seconds long sequences and the frame rate is 25 fps. Video sequences are uncompressed and they are in true color mode.

The embedded watermark is a binary image of a square-point star. This watermark is shown in the Fig. 2. The size of watermark depends on the resolution of original video and also depends on the level of wavelet decomposition. In our experiments size of the embedded watermark is 88 x 72 pixels.

Fig. 2 **Embedded watermark**

Proposed method performs watermark embedding into coefficients, which are obtained after two dimensional Discrete Wavelet Transform of second level. Watermark is embedded into approximation coefficients. The process of watermark embedding can be described by the following equation:

$$APK^{W}_{DWT}(i,j) = APK_{DWT}(i,j) + \alpha.W(i,j) \qquad (1)$$

where $APK^{W}_{DWT}(i,j)$ and $APK_{DWT}(i,j)$ represent block of marked and original approximation coefficients. $W(i,j)$ is embedded binary watermark and α (alpha) is the power of the embedded watermark. We can adjust the required robustness with this α factor. In proposed method the α factor is adjusts to values: 5, 7, 10 and 15.

In the proposed method watermark is embedded into all frames of original video sequences and also every color components (R, G, B) are marked. This approach increase robustness of embedded watermark against attacks like frame dropping, frame swapping and frame averaging.

B. The watermark extraction block

This block performs extraction of the watermark from marked approximation coefficients in the video. The process of the watermark extraction is shown in the Fig. 3.

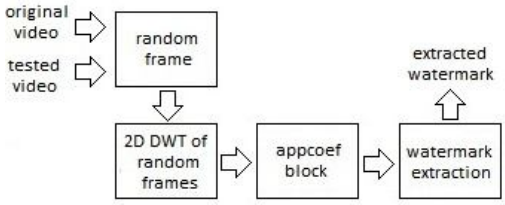


Fig. 3 **Watermark extraction block**

The inputs of the watermark extraction process are original video and marked video. The process of watermark extraction can be described by the following equation:

$$\alpha.W_{EXT}(i,j) = APK^{T}_{DWT}(i,j) - APK_{DWT}(i,j) \qquad (2)$$

where $APK^{T}_{DWT}(i,j)$ and $APK_{DWT}(i,j)$ represent block of marked approximation coefficients from the tested and original video. $W_{EXT}(i,j)$ is the extracted binary watermark and α (alpha) is the power of the embedded watermark. After extraction process elimination of factor α which gives information about robustness of watermark is necessary.

Watermark extraction is performed from a random frame of the marked video and also from every of color components (R, G, B).

IV. Experimental Results

The proposed method of watermarking was tested against unintended attacks MPEG standard compressions. Next attacks were frame averaging, frame dropping and frame swapping. These attacks are specific for a video and most used.

The quality of the extracted watermark after the attacks was judge by objective aspects, which were Mean Square Error (MSE), Peak Signal/Noise Ratio (PSNR) and also Bit Match (BM). The influence of the embedded watermark into approximation coefficients was measured by PSNR and MSE and values are presented in Table I. When the power of watermark α is increased the PSNR is decrease. The α factor is linked with robustness.

TABLE I
PSNR AND MSE COMPARISON

video		the power of the watermark α				
			α = 15	α = 10	α = 7	α = 5
D	1.	PSNR[dB]	36.09	41.17	42.11	48.13
		MSE	15.99	4.97	3.99	0.99
	2.	PSNR[dB]	36.09	41.16	42.11	48.13
		MSE	15.98	4.98	3.99	0.99
MD	1.	PSNR[dB]	36.09	41.17	42.11	48.13
		MSE	15.98	4.97	4.00	1.00
	2.	PSNR[dB]	36.14	41.22	42.13	48.18
		MSE	15.82	4.91	3.98	0.99
LD	1.	PSNR[dB]	36.11	41.17	42.12	48.13
		MSE	15.94	4.97	3.99	0.99
	2.	PSNR[dB]	36.09	41.18	42.11	48.13
		MSE	15.98	4.96	3.99	0.99

First attack was lossy compression. Six video sequences Dynamic (D), Mezzo dynamic (MD) and Lightly Dynamic (LD) were compressed by MPEG-1, MPEG-2, MPEG-4 and MJPEG. After compression watermark is extracted from a random frame and also from every color components (R, G, B). Finally watermark which is used for bit match calculation is averaged from the watermarks from components R, G and B.

The results of the watermarks extraction after lossy compression are shown in the TABLE II, TABLE III, TABLE IV and TABLE V. Bit Match is the quantity of identical pixels in the original and extracted watermarks. As can be seen from the tables the proposed method which performs watermark embedding into approximation coefficients is most robust against compression by MJPEG standard, next are MPEG-2, MPEG-1 and method is least robust against compression by MPEG-4.

TABLE II
THE QUALITY OF THE EXTRACTED WATERMARK AFTER MJPEG COMPRESSION.

Video		the power of the watermark α			
		$\alpha = 5$	$\alpha = 7$	$\alpha = 10$	$\alpha = 15$
		BM[%]	BM[%]	BM[%]	BM[%]
D	1.	87.42	95.34	98.22	99.72
	2.	87.03	96.61	98.82	99.92
MD	1.	88.38	96.21	98.99	99.78
	2.	86.19	95.17	97.13	98.99
LD	1.	87.78	96.56	98.60	99.70
	2.	89.17	97.57	99.04	99.84

TABLE III
THE QUALITY OF THE EXTRACTED WATERMARK AFTER MPEG-2 COMPRESSION.

Video		the power of the watermark α			
		$\alpha = 5$	$\alpha = 7$	$\alpha = 10$	$\alpha = 15$
		BM[%]	BM[%]	BM[%]	BM[%]
D	1.	82.75	90.66	97.38	99.45
	2.	82.77	91.89	97.84	99.64
MD	1.	82.53	90.69	97.73	99.62
	2.	85.13	93.07	97.87	99.57
LD	1.	82.88	90.34	97.01	99.68
	2.	84.01	91.92	97.90	99.68

TABLE IV
THE QUALITY OF THE EXTRACTED WATERMARK AFTER MPEG-1 COMPRESSION.

Video		the power of the watermark α			
		$\alpha = 5$	$\alpha = 7$	$\alpha = 10$	$\alpha = 15$
		BM[%]	BM[%]	BM[%]	BM[%]
D	1.	83.63	89.32	94.00	97.11
	2.	77.16	86.03	88.99	95.25
MD	1.	82.59	88.26	94.60	98.61
	2.	85.28	92.17	96.31	98.53
LD	1.	83.18	89.69	96.65	99.37
	2.	84.03	91.08	96.97	99.26

TABLE V
THE QUALITY OF THE EXTRACTED WATERMARK AFTER MPEG-4 COMPRESSION.

Video		the power of the watermark α			
		$\alpha = 5$	$\alpha = 7$	$\alpha = 10$	$\alpha = 15$
		BM[%]	BM[%]	BM[%]	BM[%]
D	1.	80.08	87.31	91.07	96.26
	2.	75.88	83.54	85.78	93.73
MD	1.	83.46	89.50	95.04	98.45
	2.	81.36	89.03	91.98	96.76
LD	1.	83.21	88.67	93.83	97.95
	2.	83.79	91.62	97.08	99.15

The results of the watermarks extraction after frame averaging are shown in the TABLE VI. The same watermark is embedded into all frames and the power of the watermark is $\alpha = 7$. As can be seen from the table quality of extracted watermarks depends on dynamic properties of video. Extraction is the best from lightly dynamic movie because there are small changes of approximation coefficients which give information about heavy features of image.

When the attacker averaged more than three frames, the quality of this frame is much degraded. Degradation of averaged frames is shown in the Fig. 4.

Fig. 4 **Five averaged frames**

TABLE VI
THE QUALITY OF THE EXTRACTED WATERMARK AFTER FRAME AVERAGING.

Video		the number of averaged frames			
		2	3	5	7
		BM[%]	BM[%]	BM[%]	BM[%]
D	1.	61.85	57.77	54.47	53.90
	2.	64.76	62.07	60.29	59.56
MD	1.	79.78	74.57	70.25	64.11
	2.	72.74	67.68	65.97	66.13
LD	1.	96.56	94.87	90.39	82.17
	2.	99.56	98.39	96.42	95.11

Last attacks were frame dropping and frame swapping. The method is robust against these attacks, because the watermark is embedded into all frames of original video. When the potential attacker dropped one or more frames, watermark from another frame can be extracted.

V. CONCLUSION

Experimental results shown, that the proposed watermarking method based on DWT is robust against the unintended attacks like lossy compression. Method is also robust against specific attacks on the video like frame swapping, frame dropping and frame averaging. The increasing of the robustness against attacks can be achieved by α factor increasing.

Experimental results highly depend on the dynamic properties of video. When the video content is less dynamic the extraction of watermark is better.

Primary disadvantages of the proposed methods are computing time and the need of the original video in the watermark extraction process.

ACKNOWLEDGEMENTS

The work presented in this paper was supported by Ministry of Education of Slovak republic VEGA Grant No. 1/0065/10, INDECT Grant (7th Research Frame Programme no. 218086) and Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF.

REFERENCES

[1] STALLINGS, W. Cryptography and Network Security: Principles and Practise. 3nd ed, Prentice Hall, 2002. 696 p. ISBN 978-0130914293.

[2] RIDZOŇ, R., - LEVICKÝ, D. Usage of different color models in robust digital watermarking. In *Radioelektronika 2009: 19h International Czech - Slovak scientific Conference,* April 2009, Bratislava, Slovak Republic.

[3] KATZENBEISSER, S., PETITCOLAS, F. Information Hiding techniques for Steganography and Digital Watermarking. Artech House, Boston, 2000.

[4] MILLER, M., COX, I., LINNARTZ, J.P., KALKER, T. A review of watermarking principles and practices, *In Digital Signal Processing in Multimedia Systems*, chapter 17, pp. 461-485, 1999.

[5] WU, M., LIU, B. Multimedia data hiding, Springer-verlag New York, Inc. 2003.

[6] DOERR, G. Security Issue and Collusion Attacks in Video Watermarking. PhD. thesis, Universite de Nice Sophia-Antipolis, 2005.

[7] CHAN, P. W. Digital video watermarking for Secure Multimedia creation and Delivery. PhD thesis, The Chinese University of Hong Kong, 2004.

Output Feedback Control Design

Daniel Gontkovič

Department of Cybernetics and Artificial Intelligence,
Technical University of Košice, Faculty of Electrical Engineering and Informatics, Slovak Republic
daniel.gontkovic@tuke.sk

***Abstract*—The linear matrix inequality based output memoryless controller design approach is presented in the paper. The design conditions are expressed in terms of matrix inequalities with the matrix rank constraints implying from an extended Lyapunov equation, which correspond to a feasible solution. Obtained formulation is the convex LMI problem for the output static controller design.**

***Keywords*—Linear matrix inequality, Lyapunov inequality, memory-free output feedback controller.**

I. Introduction

During its operation, a physical system is subject to external inputs, which may cause the undesirable effects on the system. The control system design task is attempts to stabilize the system to certain anticipated inputs and, additionally, to construct such new system abilities be able to follow desired inputs with a minimum value of track error.

Closed loop control has always been recognized to be of primary importance and has been object of intense research since the late fifties. It is recognized that most of the times feedback cannot be removed by another way for various reasons among which:

(i) it may be intrinsic in the physical mechanism generating the process variables,

(ii) process variables may come from an industrial plant where feedback loops cannot be open due to safety or production quality reasons,

(iii) the physical mechanism might be too complex and may not be easily manipulated in a open control structure.

There are many different methods to design closed loop control systems. Somebody prefers techniques based on the state-space system models (such linear quadratic control LQR, linear gaussian control LQG), others use techniques related on via pole assignment but a satisfactory general method has not been found yet [2].

The pole assignment (pole placement) is one well known design method, and although its practical usefulness has been continuously in dispute, it is the most intensively investigated in control system design. Subsequently, the state-feedback pole assignment in control system design can be noted as one from the preferred techniques in pole-placement techniques. While in the single-input single output case a solution to this problem, when it exists, is unique in a multi-input multi-output case various solutions may exist, and to determine a specific solution, additional conditions must be supplied in order to eliminate the extra degrees of freedom. This implies different suboptimal solutions, generally noted as the control system conservativeness.

In view of the optimization problems just formulated, at first it is necessary to interest in finding conditions for optimal solutions to exist. It is therefore natural to resort to a convex analysis which provides such conditions, where the main reason for studying convex functions is related to the absence of local minima. Those, a number of problems that arise in the state feedback control optimization, possibly formulated using Lyapunov function, bounded real lemma, positive real lemma etc. can be reduced to a handful of standard convex and quasi-convex problems that involve matrix inequalities (LMI). It is known that the optimal solution can be computed by using interior point methods [11] which converge in polynomial time with respect to the problem size and efficient interior point algorithms have recently been developed for and further development of algorithms for these standard problems is an area of active research. Some progres review in this research field one can find in [7], and various variants of statements of the design task, including the requirements of the controller parameter optimization, can be seen e.g. in [1], [13], and the references therein.

In the last years many significant results have spurred interest in problem of determining the control laws for the systems with constrained variables. One approach to the problem of finding the optimal results is the technique dealing with the constrained system transfer function. If this constrained problem is solvable, then it is possible to adapt the control law performance to given constrain. Therefore, special formulation can be given with the goal to optimize the output feedback controller parameters while the system state variables are constrained.

In this paper the design task of the stabilizing output memory-free controller for the constrained closed-loop system transfer function is translated into LMI framework and solved. The closed-loop system is characterized in the terms of convex LMIs, where the convex parameterization is based on the extended Lyapunov function. Problem formulation is straightforward adaptation of that one given in [3] and the used design method was directly inspired by the author's diploma work [6].

The paper was prepared in the frame of the author's doctoral study in the study branch Automatic Control on the Faculty of Electrical Engineering and Informatics of TU Košice, supervisor Prof. Dušan Krokavec, PhD.

II. Problem Formulation

The systems under consideration are linear dynamic multi-input/multi output (MIMO) systems, represented by the set of the state-space equations

$$\dot{\boldsymbol{q}}(t) = \boldsymbol{A}\boldsymbol{q}(t) + \boldsymbol{B}\boldsymbol{u}(t) \tag{1}$$

$$\boldsymbol{y}(t) = \boldsymbol{C}\boldsymbol{q}(t) \tag{2}$$

where $\boldsymbol{q}(t) \in I\!R^n$, $\boldsymbol{u}(t) \in I\!R^r$, $\boldsymbol{y}(t) \in I\!R^m$ are system state, input and output vectors, respectively, and $\boldsymbol{A} \in I\!R^{n\times n}$, $\boldsymbol{B} \in I\!R^{n\times r}$, $\boldsymbol{C} \in I\!R^{m\times n}$ are real matrices.

Problem of interest is to design the asymptotically stable as well as on the output-constrained closed loop system with the linear memory-less output feed-back controller of the form

$$\boldsymbol{u}(t) = -\boldsymbol{K}\boldsymbol{y}(t) = -\boldsymbol{K}\boldsymbol{C}\boldsymbol{q}(t) \tag{3}$$

There $\boldsymbol{K} \in I\!R^{r\times m}$ is the feedback controller gain matrix, and the output constrain is defined as

$$\int_0^t \gamma^{-1}\boldsymbol{y}^T(t)\boldsymbol{y}(t) \geq 0 \tag{4}$$

and $0 < \gamma \in I\!R$ is a positive constant.

Throughout the paper it is assumed the $(\boldsymbol{A}, \boldsymbol{B})$ is controllable, i.e.

$$\text{rank}\left[\begin{array}{cccc} \boldsymbol{B} & \boldsymbol{A}\boldsymbol{B} & \dots & \boldsymbol{A}^{-1}\boldsymbol{B} \end{array}\right] = n \tag{5}$$

III. Basic Preliminaries

A. Orthogonal Complement

Let $\boldsymbol{E}$, $\boldsymbol{E} \in I\!R^{h\times h}$, $\text{rank}(\boldsymbol{E}) = k < h$ be a real rank deficient matrix. The singular value decomposition (SVD) of $\boldsymbol{E}$ gives

$$\boldsymbol{U}^T\boldsymbol{X}\boldsymbol{V} = \left[\begin{array}{c} \boldsymbol{U}_1^T \\ \boldsymbol{U}_2^T \end{array}\right]\boldsymbol{X}\left[\begin{array}{cc} \boldsymbol{V}_1 & \boldsymbol{V}_2 \end{array}\right] = \left[\begin{array}{cc} \boldsymbol{\Sigma}_1 & \boldsymbol{0}_{12} \\ \boldsymbol{0}_{21} & \boldsymbol{0}_{22} \end{array}\right] \tag{6}$$

where $\boldsymbol{U}^T \in I\!R^{l\times l}$ is the orthogonal matrix of the left singular vectors, and $\boldsymbol{V} \in I\!R^{l\times l}$ is the orthogonal matrix of the right singular vectors of $\boldsymbol{X}$. The matrix $\boldsymbol{\Sigma}_1 \in I\!R^{k\times k}$ is a diagonal positive definite matrix of the form

$$\boldsymbol{\Sigma}_1 = \text{diag}\left[\begin{array}{ccc} \sigma_1 & \cdots & \sigma_k \end{array}\right], \quad \sigma_1 \geq \cdots \geq \sigma_k > 0 \tag{7}$$

which elements are the singular values of $\boldsymbol{E}$.

Using the orthogonal properties of $\boldsymbol{U}$ and $\boldsymbol{V}$, i.e. $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}_h$, as well as $\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}_h$ where $\boldsymbol{I}_{(\cdot)}$ is the identity matrix of appropriate dimension, gives

$$\left[\begin{array}{c} \boldsymbol{U}_1^T \\ \boldsymbol{U}_2^T \end{array}\right]\left[\begin{array}{cc} \boldsymbol{U}_1 & \boldsymbol{U}_2 \end{array}\right] = \left[\begin{array}{cc} \boldsymbol{I}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_2 \end{array}\right], \quad \boldsymbol{U}_2^T\boldsymbol{U}_1 = \boldsymbol{0} \tag{8}$$

respectively. Then

$$\boldsymbol{U}_2^T\boldsymbol{X} = \boldsymbol{U}_2^T\left[\begin{array}{cc} \boldsymbol{U}_1 & \boldsymbol{U}_2 \end{array}\right]\left[\begin{array}{c} \boldsymbol{S}_1 \\ \boldsymbol{0}_2 \end{array}\right] = \boldsymbol{0} \tag{9}$$

It is evident that for an arbitrary regular matrix $\boldsymbol{Y}$ yields

$$\boldsymbol{X}^\perp\boldsymbol{X} = \boldsymbol{Y}\boldsymbol{U}_2^T\boldsymbol{X} = \boldsymbol{0} \tag{10}$$

respectively, where a non-unique matrix $\boldsymbol{X}^\perp = \boldsymbol{Y}\boldsymbol{U}_2^T$ is the orthogonal complement to $\boldsymbol{X}$ (e.g. see [10]).

B. Matrix Inequality Equivalent Form

Let for $\boldsymbol{S} = \boldsymbol{S}^T > 0$ a matrix $\boldsymbol{K}$ has to satisfy the inequality

$$\boldsymbol{M}\boldsymbol{K}\boldsymbol{N} + \boldsymbol{N}^T\boldsymbol{K}^T\boldsymbol{M}^T - \boldsymbol{S} < 0 \tag{11}$$

Since there exists a matrix $\boldsymbol{R} > 0$ such that (see e.g. [13])

$$\boldsymbol{M}\boldsymbol{K}\boldsymbol{N} + \boldsymbol{N}^T\boldsymbol{K}^T\boldsymbol{M}^T - \boldsymbol{S} + \boldsymbol{N}^T\boldsymbol{K}^T\boldsymbol{R}\boldsymbol{K}\boldsymbol{N} < 0 \tag{12}$$

then completing square in (12) it can be obtained

$$(\boldsymbol{M}\boldsymbol{R}^{-1}+\boldsymbol{N}^T\boldsymbol{K}^T)\boldsymbol{R}(\boldsymbol{M}\boldsymbol{R}^{-1}+\boldsymbol{N}^T\boldsymbol{K}^T)^T - \boldsymbol{M}\boldsymbol{R}^{-1}\boldsymbol{M}^T - \boldsymbol{S} < 0 \tag{13}$$

$$\left[\begin{array}{cc} -\boldsymbol{M}\boldsymbol{R}^{-1}\boldsymbol{M}^T - \boldsymbol{S} & \boldsymbol{M}\boldsymbol{R}^{-1}+\boldsymbol{N}^T\boldsymbol{K}^T \\ * & -\boldsymbol{R}^{-1} \end{array}\right] < 0 \tag{14}$$

respectively. Inequality (14) can be used as an equivalent to (11) where the design parameter $\boldsymbol{R}$ is included.

Hereafter, $*$ denotes the symmetric item in a symmetric matrix.

IV. Feedback Controller Design

A. Lyapunov Inequality

Since there exists the output constrain, Lyapunov function can be chosen as follows

$$v(\boldsymbol{q}(t)) = \boldsymbol{q}^T(t)\boldsymbol{P}\boldsymbol{q}(t) + \int_0^t \gamma^{-1}\boldsymbol{y}^T(r)\boldsymbol{y}(r)dr > 0 \tag{15}$$

where $v(\boldsymbol{q}(t))$ is a quadratic positive definite function with the positive definite weighting matrix $\boldsymbol{P} = \boldsymbol{P}^T > 0$, $\boldsymbol{P} \in I\!R^{n\times n}$. Evaluating derivative of $v(\boldsymbol{q}(t)$ with respect to t it can be obtained

$$\begin{array}{c} \dot{v}(\boldsymbol{q}(t) = \dot{\boldsymbol{q}}^T(t)\boldsymbol{P}\boldsymbol{q}(t) + \boldsymbol{q}^T(t)\boldsymbol{P}\dot{\boldsymbol{q}}(t) + \\ +\gamma^{-1}\boldsymbol{y}^T(t)\boldsymbol{y}(t) - \gamma^{-1}\boldsymbol{y}^T(0)\boldsymbol{y}(0) < 0 \end{array} \tag{16}$$

and substituting (1), (2), and (3) into (16) gives the next result

$$\begin{array}{c} \dot{v}(\boldsymbol{q}(t)) = \gamma^{-1}\boldsymbol{q}^T(t)\boldsymbol{C}^T\boldsymbol{C}\boldsymbol{q}(t) - \gamma^{-1}\boldsymbol{q}^T(0)\boldsymbol{C}^T\boldsymbol{C}\boldsymbol{q}(0) + \\ +\boldsymbol{q}^T(t)(\boldsymbol{A}^T\boldsymbol{P}+\boldsymbol{P}\boldsymbol{A}-\boldsymbol{P}\boldsymbol{B}\boldsymbol{K}\boldsymbol{C}-\boldsymbol{C}^T\boldsymbol{K}^T\boldsymbol{B}^T\boldsymbol{P})\boldsymbol{q}(t) < 0 \end{array} \tag{17}$$

The design problem can be cast as a convex optimization problem. Therefore, (17) is guarantied to be fulfilled if the matrix inequality

$$\left[\begin{array}{cc} \boldsymbol{P}\boldsymbol{A}+\boldsymbol{A}^T\boldsymbol{P}-\boldsymbol{P}\boldsymbol{B}\boldsymbol{K}\boldsymbol{C} - \boldsymbol{C}^T\boldsymbol{K}^T\boldsymbol{B}^T\boldsymbol{P} & \boldsymbol{C}^T \\ * & -\gamma\boldsymbol{I}_m \end{array}\right] < 0 \tag{18}$$

is satisfied.

Inequality (18) is bilinear and has to be solved iteratively [9], [5]. Therefore, pre-multiplying left-hand side as well as right-hand side of (18) by the congruence matrix transform $\boldsymbol{T} = \text{diag}\left[\begin{array}{cc} \boldsymbol{P}^{-1} & \boldsymbol{I} \end{array}\right]$ results in

$$\left[\begin{array}{cc} \boldsymbol{Y}\boldsymbol{A}^T + \boldsymbol{A}\boldsymbol{Y} - \boldsymbol{B}\boldsymbol{X}\boldsymbol{C} - \boldsymbol{C}^T\boldsymbol{X}^T\boldsymbol{B}^T & \boldsymbol{Y}\boldsymbol{C}^T \\ * & -\gamma\boldsymbol{I}_m \end{array}\right] < 0 \tag{19}$$

where notation

$$\boldsymbol{Y} = \boldsymbol{P}^{-1} > 0, \quad \boldsymbol{X} = \boldsymbol{K}\boldsymbol{Y} \tag{20}$$

is introduced. Using by iterations obtained feasible solutions $\boldsymbol{Y} > 0$ and $\boldsymbol{X}$ satisfying (19) for given $\gamma > 0$ (if exists), the controller gain matrix can be found as

$$\boldsymbol{K} = \boldsymbol{X}\boldsymbol{Y}^{-1} \tag{21}$$

B. Unified Algebraic Approach

Another design method can be derived using unified algebraic approach. Now, inequalities (18), (19) can be written as

$$\begin{array}{c} \left[\begin{array}{cc} \boldsymbol{P}\boldsymbol{A}+\boldsymbol{A}^T\boldsymbol{P} & \boldsymbol{C}^T \\ * & -\gamma\boldsymbol{I}_m \end{array}\right] - \left[\begin{array}{c} \boldsymbol{P}\boldsymbol{B} \\ \boldsymbol{0} \end{array}\right]\boldsymbol{K}\left[\begin{array}{cc} \boldsymbol{C} & \boldsymbol{0} \end{array}\right] - \\ - \left[\begin{array}{c} \boldsymbol{C}^T \\ \boldsymbol{0} \end{array}\right]\boldsymbol{K}^T\left[\begin{array}{cc} (\boldsymbol{P}\boldsymbol{B})^T & \boldsymbol{0} \end{array}\right] < 0 \end{array} \tag{22}$$

$$\begin{array}{c} \left[\begin{array}{cc} \boldsymbol{Y}\boldsymbol{A}^T+\boldsymbol{A}\boldsymbol{Y} & \boldsymbol{Y}\boldsymbol{C}^T \\ * & -\gamma\boldsymbol{I}_m \end{array}\right] - \left[\begin{array}{c} \boldsymbol{B} \\ \boldsymbol{0} \end{array}\right]\boldsymbol{K}\left[\begin{array}{cc} \boldsymbol{C}\boldsymbol{Y} & \boldsymbol{0} \end{array}\right] - \\ - \left[\begin{array}{c} \boldsymbol{Y}\boldsymbol{C}^T \\ \boldsymbol{0} \end{array}\right]\boldsymbol{K}^T\left[\begin{array}{cc} \boldsymbol{B}^T & \boldsymbol{0} \end{array}\right] < 0 \end{array} \tag{23}$$

where $\boldsymbol{Y} = \boldsymbol{P}^{-1} > 0$.

Pre-multiplying left-hand side of (22) by the orthogonal complement

$$\boldsymbol{C}^{\circ T\perp} = \begin{bmatrix} \boldsymbol{C}^T \\ \boldsymbol{0} \end{bmatrix}^{\perp} = \begin{bmatrix} \boldsymbol{C}^{T\perp} & \\ & \boldsymbol{I}_m \end{bmatrix} \tag{24}$$

and right-hand side of (22) by the transposition of (24) gives

$$\begin{bmatrix} \boldsymbol{C}^{T\perp}(\boldsymbol{P}\boldsymbol{A}+\boldsymbol{A}^T\boldsymbol{P})\boldsymbol{C}^{T\perp T} & \boldsymbol{0} \\ * & -\gamma\boldsymbol{I}_m \end{bmatrix} < 0 \tag{25}$$

$$\boldsymbol{C}^{T\perp}(\boldsymbol{P}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{P})\boldsymbol{C}^{T\perp T} < 0 \tag{26}$$

respectively. Analogously, pre-multiplying left-hand side of (23) by the orthogonal complement

$$\boldsymbol{B}^{\circ\perp} = \begin{bmatrix} \boldsymbol{B} \\ \boldsymbol{0} \end{bmatrix}^{\perp} = \begin{bmatrix} \boldsymbol{B}^{\perp} & \\ & \boldsymbol{I}_m \end{bmatrix} \tag{27}$$

and right-hand side of (23) by the transposition of (27) gives

$$\begin{bmatrix} \boldsymbol{B}^{\perp}(\boldsymbol{Y}\boldsymbol{A}^T+\boldsymbol{A}\boldsymbol{Y})\boldsymbol{B}^{\perp T} & \boldsymbol{B}^{\perp}\boldsymbol{Y}\boldsymbol{C}^T \\ * & -\gamma\boldsymbol{I}_m \end{bmatrix} < 0 \tag{28}$$

Thus, sufficient conditions for the existence of $\boldsymbol{K}$ are given by (26).

Denoting, e.g.

$$\boldsymbol{M} = \begin{bmatrix} -\boldsymbol{P}\boldsymbol{B} \\ \boldsymbol{0} \end{bmatrix}, \qquad \boldsymbol{N} = \begin{bmatrix} \boldsymbol{C} & \boldsymbol{0} \end{bmatrix} \tag{29}$$

$$\boldsymbol{S} = -\begin{bmatrix} \boldsymbol{P}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{P} & \boldsymbol{C}^T \\ * & -\gamma\boldsymbol{I}_m \end{bmatrix} + \varepsilon\boldsymbol{I}_{n+m} > 0 \tag{30}$$

then $\boldsymbol{K}$ be a solution of (14) where $0 < \boldsymbol{R} \in I\!R^{r\times r}$, and $0 < \varepsilon \in I\!R$ are arbitrary design parameters. It is evident, that (26) have to be solved iteratively to obtain solutions for $\boldsymbol{P} > 0$ and $\boldsymbol{Y} > 0$ satisfying condition $\boldsymbol{Y} = \boldsymbol{P}^{-1} > 0$. Any conservative solution can be obtained using $\boldsymbol{P} > 0$ satisfying only (26).

V. Illustrative Example

To demonstrate the algorithm properties [6] it was assumed that system is given by (1), (2), where

$$\boldsymbol{A} = \begin{bmatrix} -2.500 & 1.000 & -0.500 \\ 8.125 & 8.250 & 6.375 \\ -11.250 & -16.500 & -10.750 \end{bmatrix}$$

$$\boldsymbol{B} = \begin{bmatrix} -0.500 & 1.000 \\ -0.125 & 0.500 \\ 2.250 & 0.000 \end{bmatrix}, \ \boldsymbol{C} = \begin{bmatrix} 7 & 6 & 5 \\ 2 & 4 & 2 \end{bmatrix}$$

The system is controlled into a stable state from non-zero initial conditions. The initial conditions of the considered MIMO system are $\boldsymbol{q}_0(t) = [-1\ 0.9\ 0]^T$. Using Self-Dual-Minimization (SeDuMi) package for Matlab [12] the conservative output-feedback gain matrix design problem given only by (26) was solved as feasible with

$$\boldsymbol{A}_c = \boldsymbol{A}-\boldsymbol{B}\boldsymbol{K}\boldsymbol{C} = \begin{bmatrix} -3.1311 & 0.4150 & -0.9618 \\ 7.8163 & 8.0029 & 6.1589 \\ -11.1263 & -15.6835 & -10.4840 \end{bmatrix}$$

which gives the stable closed-loop eigenvalues spectrum $\rho(\boldsymbol{A}_c) = \{-0.4578\ \ -1.0000\ \ -4.1544\}$.

A closed-loop autonomous system response is in the Fig. 1.

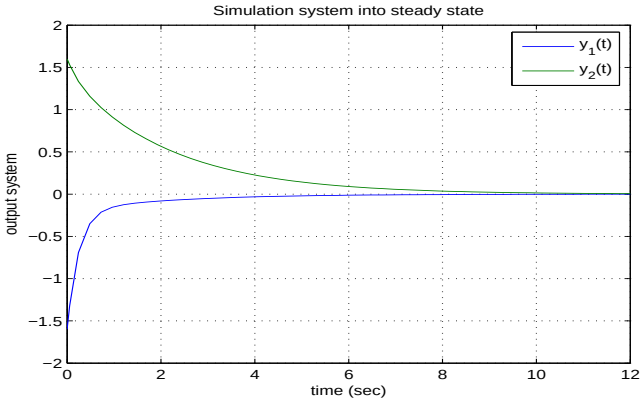


Fig. 1. Response of the closed-loop autonomous system

Acknowledgments

The work presented in this paper was supported by VEGA, Grant Agency of Ministry of Education and Academy of Science of Slovak Republic under Grant No. 1/0328/08. This support is very gratefully acknowledged.

VI. Conclusion

The output feedback memory-free gain matrix parameter design method is presented, as a modification of the unified algebraic control design method. The design principle use the standard LMI numerical optimization procedures to manipulate the gain matrix as a LMI variable. This way, the eigenvalues of the closed-loop system matrix $\mathbf{A}_c$ are placed in the left half s-plane. The presented algorithm extends Lyapunov inequality to take into account the prescribed constraint has been set on the output variables.

References

[1] D. Boyd, L. El Ghaoui, E. Peron and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, SIAM Society for Industrial and Applied Mathematics, Philadelphia, 1994.

[2] R.S Burns, *Advanced Control Engineering*, Butterworth-Heinemann, Oxford, 2001.

[3] A. Filasová, D. Gontkovič and D. Krokavec, Output feedback control design using unified algebtraic approach, *Proceedings of the 8^{th} International Symposium on Applied Machine Intelligence and Informatics SAMI 2010*, Herľany, Slovakia, pp. 259-262, 2010.

[4] A. Filasová and D. Krokavec, LMI-supported design of residual generators based on unknown-input estimator scheme, *Preprints of the 6th IFAC Symposium on Robust Control Design ROCOND '09*, Haifa, Israel, pp. 313-319, 2009.

[5] P. Gahinet, A. Nemirovski, A.J. Laub and M. Chilali, *LMI Control Toolbox User's Guide*, The MathWorks, Inc., Natick, 1995.

[6] D. Gontkovič, *Output feedback control of discrete-time linear dynamic systems*, Master Thesis, TU Košice, Faculty of Electrical Engineering and Informatics, Košice, Slovakia, 2009, (in Slovak).

[7] G. Herrmann, M.C. Turner and I. Postlethwaite, Linear matrix inequalities in control, *Mathematical Methods for Robust and Nonlinear Control*, Springer–Verlag, Berlin, pp. 123-142, 2007.

[8] I.C.S. Ipsen, *Numerical Matrix Analysis. Linear Systems and Least Squares*, SIAM Society for Industrial and Applied Mathematics, Philadelphia, 2009.

[9] D. Krokavec and A. Filasová, Decentralized control design using LMI, *Proceedings of the 8^{th} International Carpathian Control Conference ICCC 2007*, Štrbské Pleso, Slovakia, pp. 381-384, 2007.

[10] D. Krokavec and A. Filasová, *Discrete-Time Systems*, Elfa, Košice, 2008. (in Slovak)

[11] Y. Nesterov ans A. Nemirovsky, *Interior Point Polynomial Methods in Convex Programming. Theory and Applications*. SIAM Society for Industrial and Applied Mathematics, Philadelphia, 1994.

[12] D. Peaucelle, D. Henrion, Y. Labit and K. Taitz, *User's Guide for SeDuMi Interface 1.04*, LAAS-CNRS, Toulouse, 2002.

[13] R. E. Skelton, T. Ivasaki and K. Grigoriadis, *A Unified Algebraic Approach to Linear Control Design*, Taylor & Francis, London, 1998.

Control system for school robot manipulators

[1]*Ján ILKOVIČ*, [2]*Tomáš KAROĽ*, [3]*Rastislav HOŠÁK*, [4]*Juraj CHOVAŇÁK*

[1,2,3,4]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]jan.ilkovic@tuke.sk, [2]tomas.karol@tuke.sk, [3]rastislav.hosak@tuke.sk, [4]juraj.chovanak@tuke.sk

***Abstract*— The article discusses concept of technological process. It focuses on the way of controlling at all levels of distributed system. In the management of the standard communication protocols are used to link the individual application, which ensure that the operation of the system. The system is made up of PLC with program, kinematical model, visualization, database, which preserve process data and dates from the production plan. The article notices what is important to know by the creation of controlling system in the manufacturing corporation. Presented application is an example implementation of a distributed system controlling for a specific laboratory technology model that represents a link to the classification of products.**

***Keywords*—technological process, stepper motor, Programmable Logic Control (PLC), Dynamic Data Exchange (DDE), Open Database Connectivity (ODBC), data acquisition, database, communication, manipulator, kinematics, visualization, robotic workstation.**

I. INTRODUCTION

Solving problems [1, 3, 8, 10] of robotic workstations in development of all control's levels requires to take in account the various communications facilities and functional elements, integrate kinematic model, which is used to calculate direct (DKT) and inverse kinematic task (IKT) and data acquisition for the information level. Communication between levels of Distributed Control System (DCS) is implemented by using standard communication protocols, which are used to link those levels. Part II. Technological level of control describes hardware and PLC control, another part III. describes communication linkages and visualization, part IV. deals with data acquisition and part V. describes production plan on Manufacturing Execution System (MES) level. There are three selected technological processes as an illustration of the complex of model functionality. More detailed description of making the individual hardware and software parts are given in the next section.

II. TECHNOLOGICAL LEVEL OF CONTROL

Technological level of control is composed from manipulators hardware construction, actuators and control system. Stepper motors are model actuators that are interconnected to manipulator construction by using the cable gears. All used sensors in the model operating on optical principle. There are 15 stepper motors and 11 optical sensors in the model. Ten of sensors are infrared optical sensors and one is diffusive laser sensor. PLC is central control system, at which input/output cards are connected each of inputs and outputs used in the model. SLC500 is type name of control system. It consist of procesor SLC 5/03 CPU, two digital output cards using 24 voltage where each one has 32 digital transistor outputs and two digital input cards with 16 outputs at each one and also using 24 voltage.

Fig. 1. Workstation of robots

As the model has been given new concept of control, there has had to be changed also all other parts of technological level of control. So the origin manipulator control electronic part was replaced by new one that contains 24 galvanically isolated digital amplifiers for stepper motors and also one 8-block amplifier for sensors. All stepper motor coils were connected singly to SLC outputs, that is why we could control every stepper motor independently in any time.

All sensors and also digital joystick has been connected to the input cards. Connection between SLC processor and computer, where the other applications run and will be mentioned thereinafter, is realized via serial connection RS232.

Control application for SLC500 has been done in the development environment RSLogix500 by Rockwell Software. This application manages all signals from the model sensors and also drives actuators by signals from two digital output cards. Communication with other applications is created by the serial connection RS232 using the DDE communication protocol. This communication provides another Rockwell Software application - RSLinx.

The main task of SLC application is fulfill all motions that are received from supervisory level applications such as movements of the manipulators, conveyor belts and rotary conveyor. Visualization application and manipulator mathematic model are those mentioned applications by using which can user define model movements.

The SLC application also allows to control all actuators by using only development environment RSLogix500, but in that case user have to understand SLC source code.

There is also possibility to control first manipulator by

joystick that instantials only like a demonstration that model is functional.

By reason of model control simplification is SLC application design in such way that for activating an actuator is needed only to set up one bit and also the motor sense of rotation is given by another bit. At this concept every stepper motor has its own dibit, through the use of it is possible to control its motion. In this principle is possible to control whole model by using only 30 bits and this 30 bits are available for visualization application in Manual Mode via Direct Drive communication protocol provided by RSLinx.

Visualization application contains also Automatic Mode that allows to user send manipulator head containing jaws to the point in the space. Point in the space is defined by user by setting X-axis, Y-axis and Z-axis. For this mode has SLC application created variables for each stepper motor separately, to which are written values of steps, that have to be done by stepper motors to get manipulator head into the required position.

III. COMMUNICATION LINKAGES AND VIZUALIZATION

A. Communication software

Application of communication software was developed in Borland C + + Builder 6. It serves to link the PLC control, kinematics and visualization and an input (the server) for data from the camera of the current position of the wanted products in the area.

Its main task is to calculate the direct and inverse kinematics task [2, 3, 7], which is necessary for movement of manipulators in space, because it ensures the conversion of position in space to angles of rotation for stepping motors and vice versa.

The proposed communication software has scalable modular structure- shown in Fig. 2.

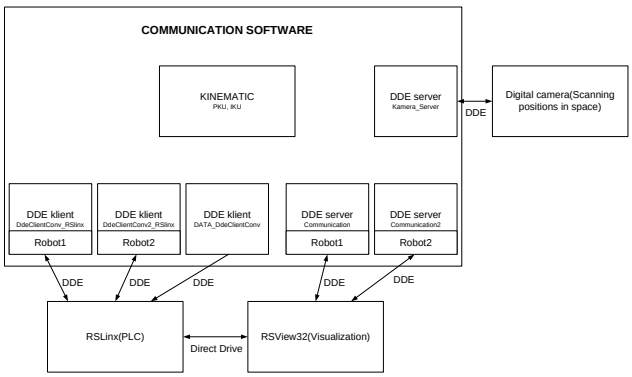


Fig. 2. Block diagram of the communication software

Description of the functionality of DDE servers:

- Comunnication – server, variables of which are used for purposes of kinematics. Designed for Robot 1.
- Comunnication2 – server with the same functionality as Comunnication but with kinematic linkage for Robot 2.
- Kamera_Server – an input interface for the camera system.

Description DDE Clients:

- DdeClientConv_RSlinx – its client's variables are linked to the control, auxiliary and check variables used in the PLC control for Robot 1.
- DdeClientConv2_RSlinx – the same functionality as DdeClientConv_RSlinx, but the control interface of Robot 2.
- DATA_DdeClientConv – connectioned himself are entered to the automat address (PLC) current position and orientation of the Robot 1 and 2, where are accessible for application RSSql.

B. Visualization

Solution of visualization [4,13] was designed in RSView32 version of 6:30:16, which is part of the software package of Rockwell Software [5, 6]. It serves to link the technology level with the operator. Its variables are directly related to bits and words (variables) used in the PLC control program. It has static nature, so that we have to monitor them visually on the technological workplace.

On this level desired positions and orientations are entered. They are sent through the DDE to the communication applications, where the appropriate kinematics calculation is realized (direct or inverse). After that action intervention is calculated and it is written through the DDE interface into the PLC. Then the interrupt is immediately performed or it waits for other values – this depends on the actual active mode. The function of visualization and its linkages with other applications used to control the robotic workstation are shown in the following figure (Fig. 3).

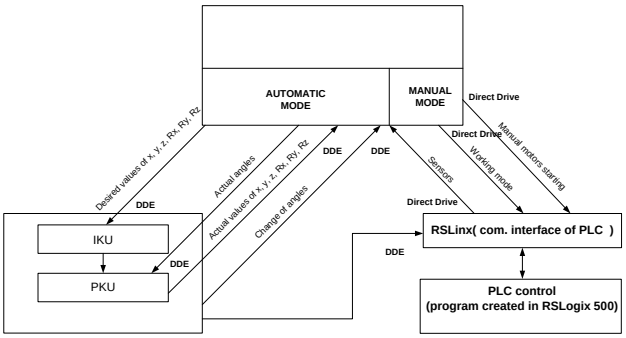


Fig. 3 Communication linkages of the visualization

Interfaces of the visualization are:

- Communications software – calculated kinematics from specific values and then entered as an action intervention into PLC automat via DDE.
- RSLinx – includes sensing elements, and direct starting engines in manual mode. This interface is made by communication protocol Direct Drive.

Robotics workstation may work in three modes (push the button in visualization by the operator):

1) Manual mode

Movement of the technological workstation in this working mode is fully controlled by the operator. He can move every moving part of the complex (if the motors perform movements of a certain part) or a particular stepping motor press the button with the arrow symbol, which represents the direction in which the movement is performed (Fig. 4).

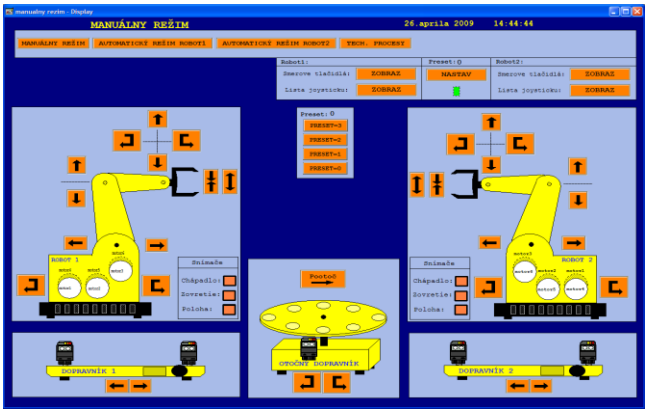


Fig. 4 Window of the manual mode robotics workstation

2) Automatic mode

In this mode (Fig. 5), unlike the manual mode, there is possibility of entering the desired position *x*, *y*, *z* and orientation R_x, R_y in the space. Orientation R_z is calculated from coordinates *x*, *y*, because the manipulator - its physical structure - has only 5 degrees of freedom [3]. Another functionality of this mode is, that we can teach the manipulator sequences of movements (point to point), we gradually enter the desired position and orientation and that are saved in the PLC 's memory.

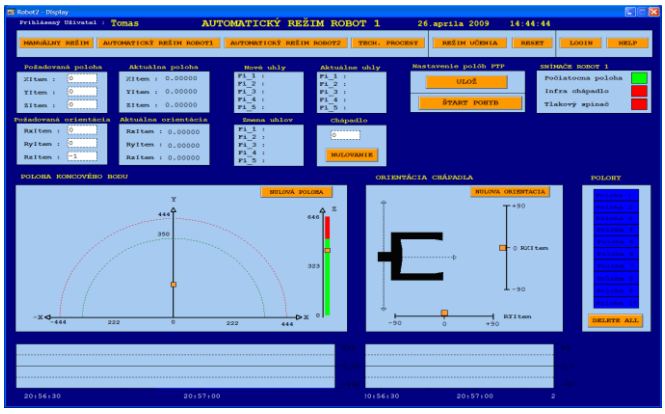


Fig. 5 Window of the automatic mode

3) Mode of technological processes

The last one mode - mode technological processes - can be activated only if desired sequences of positions are stored in memory of manipulators. If this condition is met, we can run technological process on the PLC controller by selecting (pressing) the button in visualization.

Visualization in this mode performs a supervisor role, which can abolish or restart the technological process – it depends on requirements and functions of operator.

IV. Data Acquisition from Technological Process

Data acquisition [8] is characterized of dataflow from technological process [12] through the individual processing steps to the data provision to the end-user.

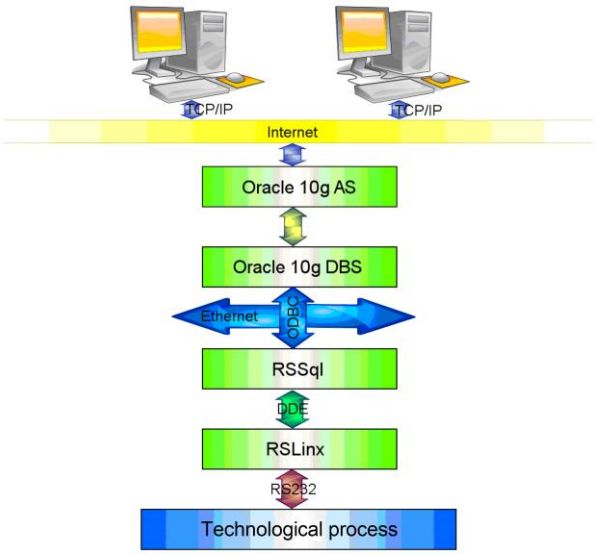


Fig. 6. Data acquisition from technological level of control.

Technological data was gain with RSLinx software from technological level of control. The computer, on which this application with application RSSql was situated, is connected with programmable logic automat SLC500. Application RSSql be used to connection communication protocols. RSSql intervene processing data to higher level of control. With DDE protocol was realized direct communication to the technological level of control and with ODBC protocol was realized direct communication created on platform Oracle 10g [9]. RSSql made links between data cells in PLC and columns in tables of data base server. RSSql was also used to start or stop the filling database with processing data.

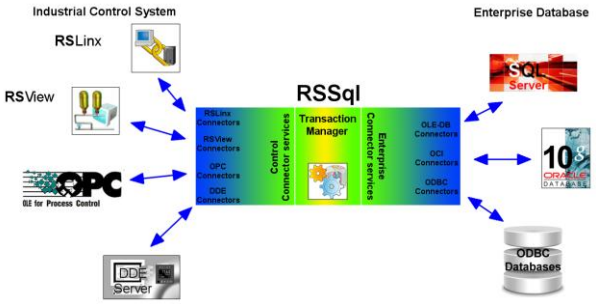


Fig. 7. Architecture of RSSql.

RSSql is transaction system, which can create full duplex connection between control system and database system. To the RSSql on the side of control is possible to connect different resources, for example RSLinx, RSView, DDE server or OPC (Object Linking and Embedding for Process Control) server. With ODBC protocol or OCI (Oracle Call Interface) can join in to the database servers, on the other side.

Data storage was relate to the time and date, information from optic gates situated on conveyor belts, information about activities of conveyor belts, actual positions of manipulator arms and manipulator tentacles in work area and regime in which all system works.

V. Manufacturing Execution System

Experience shows that the deployment of MES [11, 14] has

a propitious effect on the duration of the production cycle, largely eliminates work with paper documents, improves the quality of production (reduces the number of faulty pieces and defects) and reduces the time to enter data into the system. Nowadays companies are managed by production plan, which is completed on the basis of orders. Entering production plan in to the system performs authorized person.

On the database server [8, 9] is reserved table-space, which is administered by manager and by which production process develops on lower levels of control. The order with earliest completion date is chose from plan production during the technological process. This order is begin to execute.

Base of production plan are three tables: "ZOZNAMVYROBY", "PLANVYROBY" and table "UKONCENEZAKAZKY". Application "Plán výroby" works with them. This application is named Plan_vyroby.exe and claim for software Microsoft Visual Studio 2008. This application was developed in language C++ in remembered software. Necessary part is enlarging package for work with database system Oracle. Application accesses to the database with standard SQL orders. It consists with three compounds.

The first tab "Zoznam výrobkov" is assigned to entering of new products, which will be used in technological process with. User has a possibility to add bar code in to the database. The characteristic number of product type is assigned to each entered bar code automatically. User can take away product from the list in the case that product will be not part of production process in the future.

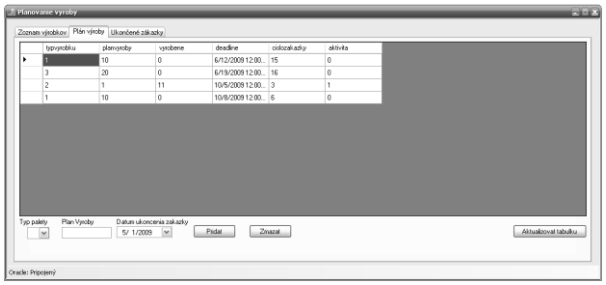


Fig. 8. Application – "Plán výroby".

The second tab "Plán výroby" intervene view to production plan. It makes possible to addition and deleting orders. Users have a possibility to definition properties of order. Be specific, which type of pallet should be made for offer – "Typ palety", how many of current pallet should be made – "Plán výroby" and completion date of offer – "Dátum ukončenia zákazky". Next possibility for users is deleting offers.

The third tab "Ukončené zákazky" has only information character and don't propose way of editing.

VI. Conclusion

The biggest benefit of this work is the practical implementation of the manufacture process for the structure of the DSR model on school system manipulators.
This model is intended mainly for demonstration purposes.
It was necessary to interconnect different types of equipment to a functional unit, so as to ensure the structure of the DSR.
The control system was divided into three parts

The first part [1] deals with the hardware equipment and the first line management.

Part Two [10] includes the SCADA/HMI control and kinematic model of the system manipulator.

The last part [8] is an information system operating at the level of database servers with the possibility of production planning.

Nowadays, the management companies seeks to automate the greatest amount of steps. After this manner they increase efficiency of production and less time. These factors directly affect the competitiveness at the trade. This thesis has been made in this spirit using basic parts of distributed system of controlling.

Acknowledgment

This work was supported by grant KEGA 037-011TUKE-4/2010 and VEGA - 1/0617/08.

References

[1] J. Ilkovič, "Návrh a realizácia technologickej úrovne riadenia modelu sústavy manipulátorov (Diplomová práca)," Košice, 2009.
[2] D. Krokavec, "Spracovanie údajov v robotike," Košice: Rektorát vysokej školy technickej v Košiciach, *85-628-85*,1985, pp. 260.
[3] J. Fedorčák, "Riadenie programového modulu pre riešenie priamej a inverznej kinematickej úlohy manipulátora (Diplomová práca)," Košice, 2007, pp. 90.
[4] D. Mudrončík, I. Zolotová, "Priemyselné programovateľné regulátory-Konfigurovanie, vizualizácia, kvalita softvéru," Košice: Elfa, *ISBN80-88964-45-8*, 2000, pp. 169.
[5] RSView32, "Technical Data. Rockwell Software," *VW32 - TD001A – EN – P*, 2000, pp. 8.
[6] RSView32, "USER'S GUIDE. Rockwell Software," *VW32-UM001D-EN-E*, 2007, pp. 776.
[7] "Teória priemyselných robotov," Ostrava, 2000, pp. 147, [online] [cit. 2009-12-2], available on the internet: <http://matescb.skvorsmalt.cz/robotika_kybernetika/teorie_prumyslovych_robotu.pdf> .
[8] R. Hošák, "Informačná úroveň riadenia modelu sústavy manipulátorov (Diplomová práca)," Košice, 2009.
[9] Oracle, "Oracle Database 10g Express Edition Tutorial," [online] [cit. 2009-12-2], available on the internet: <http://st-curriculum.oracle.com/tutorial/DBXETutorial/index.htm>.
[10] T. Karoľ, "Realizácia komunikačného a vizualizačného softvéru pre riadenie modelu sústavy manipulátorov (Diplomová práca)," Košice, 2009.
[11] I. Bélai, L. Jurišica, K. Kováč, J. Šturcel, "Control and diagnostics of the technological processes in manufacturing process based on application of ICT (Selected applications of ICT in enterprises, institutions and SMEs)," Bratislava: STU v Bratislave FEI, *ISBN 80-227-2303-7*, 2005, pp. 158.
[12] M. Franeková, K. Rástočný, "Modelling of disturbing effects within communication channel in area of safety related communication systems (Advances in Electrical Engineering)," *ISSN 1336-1376*, June 2007, pp. 63-68.
[13] I. Zolotová, S. Laciňák, E. Ocelíková, "New trends in supervisory monitoring and control," *International Conference on Applied Electrical Engineering and Informatics: September 8-11, Greece, Athens 2008*, Košice: FEI TU, *ISBN 978-80-553-0066-5*, 2008, pp. 102-105
[14] P. Božek, "Complex Control and Metrology Security of Automatized Trial System (Manufacturing Engineering, No. 2, volume IV) ," Prešov, *ISSN 1335-7972*, 2005, pp. 31-35.

Visualization of city agglomerations using 3D and panoramic pictures

František HROZEK

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

frantisek.hrozek@tuke.sk

***Abstract*— Visualization is nowadays very widespread, especially using 3D. It can be found everywhere, for example in entertainment, architecture, design or scientific research. In this paper is described a presentation part of 3D city agglomerations visualization. Problematic of presentation part is divided into four parts. Firs part deals about script creation which allows the smooth redraw of model given by the transition path. Second deals with rendering setting for photorealistic outputs. Third part is about using network rendering for speeding rendering calculations. Final part is about the panoramic images creation. The result of this work is the script for creating animations and panoramic images, serving as geographic information of individual positions on the map.**

***Keywords*—city information system, panoramic pictures, MAXScript, 3D Studio Max, V-Ray, visualization.**

I. INTRODUCTION

Today, more than ever, there is a great accent put on information about cities. Every city wants to simply and effectively mediate information about itself to citizens, tourists and investors. One way how to do this is city agglomerations visualization. This visualization can be created in 2D or 3D. Also, it is possible to combine these two approaches. This solution supports low hardware requirements and high information value.

The aim of this paper is 3D visualization of city agglomeration in the form of projection on n-angle around camera. Panoramic images were used for projection. These pictures were created from 3D model of visualized city agglomeration and serves as geographic information of individual positions on the map.

There is a plenty of specialized 3D software and number of modeling techniques with which 3D model of selected city agglomeration can be created [1][2]. It is necessary to choose right modeling technique and specialized software to get proper model. Second problem is a visualization of these models with additional information data displaying. Because the issue of 3D city agglomeration visualizations is wide-ranging, work described in this paper was divided into these three parts: modeling, presentation and visualization. Main goal of this paper is description of the presentation part.

Work on presentation part was divided into the following parts:

- Theoretically study basic issues of geographic systems, computer graphics, spatial modeling and visualization techniques,
- Propose a way of presenting the structure of a 3D model for 3D geographic information system with a focus on urban areas in the form of projections on the n-angle around camera,
- Propose a methodology for gathering, modeling, and subsequent visualization of models defined in paragraph 2 in form of photorealistic transformation,
- Develop a 3D model of a selected part of an urban agglomeration on the basis of points 2 and 3,
- Implement software using the scripting language which allows the smooth redraw of model given by the transition path.

Our choice of the city agglomeration area, which was visualized, was the Technical University of Kosice bounded by the adjacent streets: Letná, Komenského, Watsonová and Boženy Němcovej.

Panoramic images creation consists from following steps:

1. loading of the input data and coordinates in script,
2. parameters setting for script,
3. creation of animation using script,
4. parameters setting for photorealistic lighting,
5. rendering of all frames of animation,
6. creation of panoramic images .

Before rendering it was important to decide, how many computers we will use for rendering: single PC or multiple PC connected by network (network rendering).

II. SCRIPT DESCRIPTION

The most important part of our work was script which allows the smooth redraw of model given by the transition path. Since 3D Studio Max [3], which was used for rendering, contains its own scripting language, there was no need to seek other solutions. Name of this scripting language is MAXScript [4]. Created script consists from those parts:

- loading of input matrix,
- setting start and end coordinates,
- setting basic parameters,
- creation of animation,
- sorting out frames of animation.

MAXScript Overview

MAXScript is the built-in scripting language for 3D Studio Max and related products, such as Autodesk VIZ, character studio, Plasma and GMax.

MAXScript:

- language is specifically designed to complement 3D Studio Max,
- language syntax is simple enough for non-programmers,
- is rich enough to enable sophisticated programming tasks,
- integrates well into the 3D Studio Max user interface,
- supports formatted text as well as binary data input and output,
- can also be used as a high-level scene import utility.

The MAXScript language, as mentioned before, was developed to be used by artists as well as by Technical Directors and programmers. It provides relatively relaxed syntax rules and is more similar to BASIC than to C++ or Unix command line as is the case with Maya's MEL script. The internal structure of MAXScript has more similarities to LISP - MAXScript is an expression-based language.

Some notes on syntax specifics:

- **Semi-colons (;)** are not required for end of line termination but are allowed. They are only required to delimit multiple expressions in the same line,
- The language is completely **CASE-INsensitive**,
- **Variables** do not require explicit type declaration or value assignment. Uninitialized variables always return a special value 'undefined' which is equivalent to NULL,
- **Round brackets (parentheses)** are used to define code blocks and name spaces,
- **Arrays** are defined using #() and have no fixed element type - an array can contain any number of different elements with different classes, including other arrays,
- **Rectangular brackets []** are used for indexed or by-name access to sub-objects and array elements,
- Single-line **remarks** are inserted using double-dash --. Multi-line remarks are inserted using slash-star and star-slash pairs like /* some remark here */,
- **Properties** of objects are accessed either using a DOT notation (Object.PropertyName) or via GetProperty / SetProperty method calls.

Loading of Input Matrix

Input matrix determines which point of the scene is included to the animation and which not. Upon mutual agreement with the colleague who was responsible for the visualization part, input matrix got the shape of a square. Point in the lower left corner of the square matrix represents the coordinates [0,0]. Individual points of matrix are formed by numbers 0 or 1. These numbers notify script:

- 1 - point to be included in the animation
- 0 - the point is not to be included in the animation

For input matrix creation was used part of program, which serves for new project creation. This program was created in visualization part of our work. As input for this part of program was used picture of chosen city agglomeration, view from above. This picture was created in this part of the city agglomeration visualization. It is necessary, that the lower left corner of the picture must represent coordinate [0,0] in 3D Studio Max. In Fig. 1 square **A** shows the correct position of a segment designated to picture creation. Square B shows the incorrect position. Arrow shows the shift between correct and incorrect position that leads to collisions between the camera and objects in scene. Height and width of picture must by same.

Fig. 1 Correct (A) and incorrect (B) position of a segment designated to picture creation

Setting Start and End Coordinates

Setting start and final coordinates is used to determine the coordinates for start and end of animation. These coordinates do not need to be same as input matrix coordinates. This option is used to specify a smaller part of the input matrix, which can be used for dividing larger input matrix into smaller matrices and from these create small animations that allows processing on multiple computers.

Setting Basic Parameters

Basic parameters are:

- **step size** – serves to specify step size for camera, default value is 5 m
- **camera height** – serves to specify the height of camera, in which the camera will move around the scene, the default value is 2 m
- **number of renders** – serves to determine the angle for rotation of camera around its z axis, default value is 16 (360 / 16 = 22.5˚)

Parameter step size serves for computing grid density, which is one of the values needed for the input matrix creation. Value of grid density is computed by equation (1), where s is size of the square edge, k is step size and h_m is grid density.

$$\frac{s}{k} = h_m \tag{1}$$

Step size used in script must by same as step size used for grid density calculations. If different step size values are used, then collisions occur between camera and objects in scene.

Creation of Animation

Creation of animation consists of two parts:

- part in which is checked, if the coordinate is included in the animation or not,
- part of rotating the camera around its z axis.

The first part uses data from the input matrix, which are loaded into two-dimensional array of occurrence. This array represents the coordinate system in which the camera moves. If an array element contains 1, then the item is included into the animation. Then from the coordinates of this element and also entered step size is calculated camera position.

In the second part the camera begins to rotate around its z axis. The angle is calculated using the parameter number of renders. Initial rotation of the camera is 45˚ (fig. 2) in a clockwise direction, which was designed in tests with panoramic pictures. For each one turn of the camera the frame of animation is created. When the camera turns around z axis about 360°, then the next element of occurrence array is taken.

In the selected output directory, a copy of the input matrix and the starting and final coordinates of the animation are saved when **Vytvor animaciu** button is pressed.

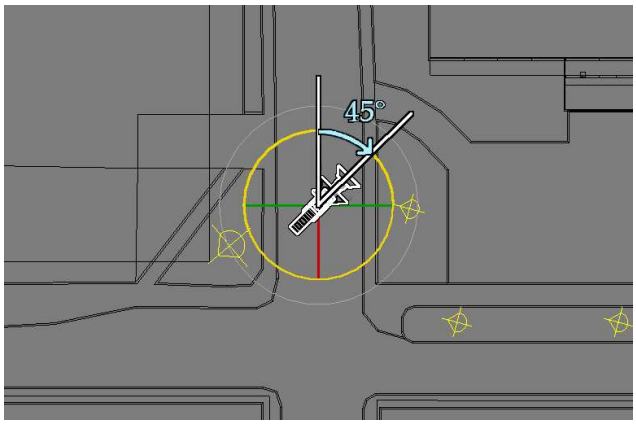


Fig. 2 Initial camera rotation, top view

Sorting out Frames of Animation

After rendering it was important to sort images into directories according to coordinates from which were created, because all images were rendered into the one output directory. Script uses copy of the input matrix, initial and final coordinate values for sorting of rendered images.

III. LIGHTING

Creation of realistic lighting was one of the key conditions for creating photorealistic images for our work. It was therefore necessary to choose the most realistic lighting, but also take into consideration rendering time. Also it was necessary to create a realistic sky for better impression of the real lighting [5].

Creation of lighting was divided into several points:

- selection of a suitable light source,
- sky creation,
- remove of color bleeding,
- lighting parameters setting.

Choosing the Light Source

When selecting a light source it was necessary to take into consideration the quality of lighting and rendering time. According to these criteria were then performed tests with different kinds of lights sources. The best results were achieved with lights sources Skylight [1] and VRayLight [6]. However, rendering time for Skylight was very high and therefore the only solution for photorealistic lighting remained light source VRayLight.

Sky Creation

Illusion of real sky was created by placing textured hemisphere into model. Advantage of this approach is that the sky change with the changing of camera position and rotation. Texture created for hemisphere, used in the model, was procedural.

Color Bleeding

Color Bleeding is transfer of color between nearby objects, caused by the colored reflection of indirect light. This is a visible effect that appears when a scene is rendered with Radiosity or full Global Illumination, or can otherwise be simulated by adding colored lights to a 3D scene. In our visualization was color bleeding unwanted effect and therefore we removed it. In Fig. 3 is shown visualized building with color bleeding effect and without it.

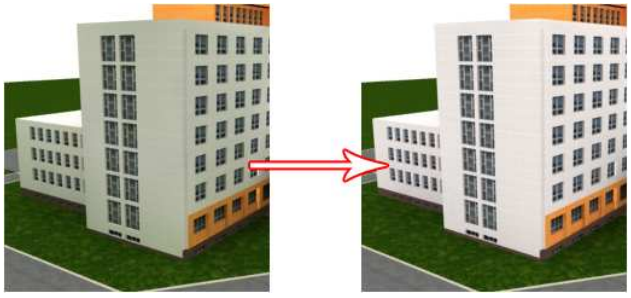


Fig. 3 Model with (left) and without (right) color bleeding

Lighting parameters setting

Another very important part of this work was to find compromise between the quality of rendered image and rendering time. In search for best results, many tests were done. In Fig. 4 is shown rendered image with the most appropriate setup for lighting parameters.

Fig. 4 Rendered image with most appropriate setup

IV. RENDERING

Because rendering with single PC would take long time, we used network rendering. Render network (render farm) was build from 20 PC. Configuration of used PC:

- CPU – AMD Athlon(tm) 64 Processor 3700+
- Memory – 1GB DDR
- HDD – Maxtor 160 GB
- GPU – 256 MB GeForce 7300 GT

Since 3D Studio Max [3] contains own network manager (Backburner Manager) for network rendering, it was no longer necessary to use any other software for network rendering. Rendering time for all images work was approximately 10,5 day.

V. PANORAMIC IMAGES

When were all images rendered and sorted, it was necessary to create panoramic images from them. For panoramic images creation was used the program Panorama Maker Pro 4 (trial version) from ArcSoft [7]. Each one panoramic image was composited from 16 rendered pictures. When was panoramic image created, its resolution was adjusted to 4096x1024 to speed-up their loading and processing on the graphic card.

VI. CONCLUSION

Output from presentation part of our work is script and panoramic images. Script was created with MAXScript scripting language and serves for animation creation and for sorting of pictures. Created panoramic images serves as geographic information of individual positions on the map.

A lot of details were leaved out in these panoramic pictures due to computing power of render farm. But with increasing power of computer hardware more and more details can be added to visualized scene of city agglomeration to improve photorealistic impression. These details can by higher polygon count of buildings, cars and other items from real life. Also higher and better values for lighting can be set.

ACKNOWLEDGMENT

This work is the result of the project implementation: Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF and is also supported by VEGA grant project No. 1/0646/09: Tasks solution for large graphical data processing in the environment of parallel, distributed and network computer systems.

REFERENCES

[1] B. Sobota, J. Perháč, I. Petz: *Surface modelling in 3D city information system, Journal of Computer Science and Control Systems*, 2, 2, 2009, pp. 53-56, ISSN 1844-6043

[2] B. Sobota, Cs. Szabó, J. Perháč, H. Myšková: *Three-dimensional interfaces of geographical information systems*. In: ICETA 2008 : 6th International Conference on Emerging eLearning Technologies and Applications : Information and communications technologies in learning : Conference proceedings : September 11-13, 2008, Stará Lesná, The High Tatras, Slovakia. Košice : Elfa, 2008. pp. 175-180. ISBN 978-80-8086-089-9.

[3] *3ds Max 9 User Reference*. URL: http://usa.autodesk.com/adsk/servlet/item?siteID=123112&id=10175188&linkID=9241177

[4] *MAXScript Reference 9.0*. URL: http://usa.autodesk.com/adsk/servlet/item?siteID=123112&id=10175420&linkID=9241175

[5] S. Kennedy: *3ds max 6: Animace a vizuální efekty*. Brno: Computer Press, August 2004. 554s. ISBN 80-251-0328-5

[6] *V-Ray help index*. URL: http://www.spot3d.com/vray/help/150SP1/

[7] *Panorama Maker 4 Pro*. URL: http://www.arcsoft.com/products/panoramamakerpro/

Advanced Temporal-spatial Error Concealment Algorithm for Video Coding in H.264/AVC

[1]*Branislavl HRUŠOVSKÝ*, [2]*Ján MOCHNÁČ*, [3]*Pavol KOCAN*

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]branislav.hrusovsky@tuke.sk, [2]jan.mochnac@tuke.sk, [3]pavol.kocan@tuke.sk

***Abstract*— Low bandwidth and high error rates are characteristic properties of mobile communication environments. Video transmitted over networks based on IP protocol is always subject to packet loss due to network congestion and channel noise. The major limitation of mobile networks is low transmission bit rate which demands the reduction of the used video resolution and high efficient video compression technique. Standard H.264/AVC, which is explained in this paper, is the newest video compression codec, which provides a distinct improvement of quality in comparison with previous video standards. Real-time transmission of video data in network environments, such as wireless network, is a challenging task, as it's impossible to retransmit the erroneous or lost macroblocks. Therefore there is a need for post-processing method, which try to restore the missing or corrupted video content by using the previously decoded video data. We used in our paper an advanced temporal-spatial error concealment technique for H.264/AVC coded video [3]. Simulations were done in computing environment Matlab using two standard video-sequences.**

***Keywords*—error concealment, frame, macroblock, slice**

I. INTRODUCTION

With the success of MPEG-2 in DVD and HDTV applications, the demand for higher coding efficiency in video-based services has grown significantly. In 2001, ITU-T Video Coding Experts Group (VCEG) together with the ISO/IEC Moving Picture Experts Group (MPEG) formed the Joint Video Team (JVT) to develop a new video coding standard with better coding efficiency. The committee completed the final draft of ITU H.264 also known as ISO MPEG-4 Part 10. The H.264/AVC (Advanced Video Coding) is thus a new video coding standard, which achieves much better compression than all other existing video coding standards. The H.264 supports video applications including low bit-rate wireless applications, standard-definition and high-definition broadcast television, video streaming over the Internet, delivery of high-definition DVD content, the highest quality video for digital cinema applications, etc.

II. H.264/AVC VIDEO STRUCTURE

A coded video sequence in H.264/AVC consists of a sequence of coded frames. Each frame is created by sampling the color values RGB (red, green and blue) of the captured image, according to the desired video resolution M x N pixels, where M is the number of pixel columns and N the number of pixel rows of the frame [1]. Each pixel is represented by a group of three color samples RGB. For the digital image processing, the RGB color samples (color space) are transformed to the YCbCr color space, where Y stands for the luminance (luma), Cb and Cr stand for chrominance (chroma) components. The color space transformation is given by (1) :

$$\begin{aligned} Y &= (0{,}229R + 0{,}587G + 0{,}144B) - 128 \\ Cb &= 0{,}433(B - Y) \\ Cr &= 0{,}877(R - Y) \end{aligned} \quad (1)$$

Because the human visual system is more sensitive to luminance than to chrominance, H.264/AVC uses a sampling structure in which the chroma component has one fourth of the number of samples than the luma component (half the number of samples in both the horizontal and vertical direction). This is called 4:2:0 sampling with 8 bits of precision per sample, as is shown in Fig.1. The undersampling of chroma samples reduces the amount of data per each frame without causing any degradation to the image quality [2].

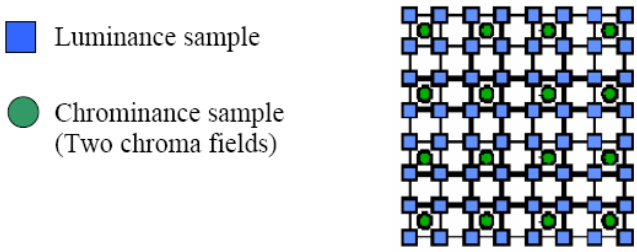


Fig. 1. Chroma format (4:2:0)

Each frame is partitioned into fixed-size macro blocks, where each macro block covers a rectangular frame area of 16x16 samples of the luma component and 8x8 samples of each of the two chroma components. Macro blocks are the basic building blocks of the standard for which the coding and decoding process is specified. The macro blocks are organized in slices, which generally represent subsets of a given picture that can be decoded independently. The slice is a group of macro blocks. Each picture may be split into one or several slices as shown Fig.2.

Each slice can be correctly decoded without the use of data from other slices provided in the same frame. Some information from other slices may be needed to apply the deblocking filter across slice boundaries [2].

The number of macro blocks in each slice can be set to constant value or it can be specified according to a fixed

number of bytes. Since each macro block is represented by variable number of bits, the encoder uses stuffing bits to fill the slice up to the desired bytes number. A slice can also be specified by using Flexible Macroblock Ordering - FMO.

Fig.2 also shown, that picture can be split into many macroblocks scanning patterns such as interleaved slices.

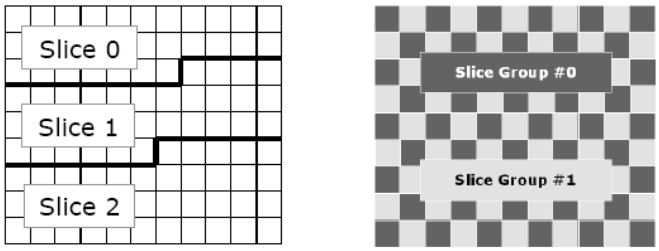


Fig. 2. Subdivision of video frames : a) basic slice mode b) FMO slice mode

III. Error concealment in H.264/AVC

The loss of transmitted data packets influences the quality of the received video. This problem is caused by the limited channel bandwidth used by the mobile communication networks. Since the real time transmission of video stream limits the channel delay, it is not possible to retransmit all erroneous or lost packets. Therefore there is a need for post-processing methods, which try to reduce the visual artifacts caused by bit stream error after locating the missing or defected parts of video data [4]. Error concealment methods which shall be implemented on the receiver side restore the missing and corrupted video content using the previously decoded video data. There are several error-resilience techniques. Forward, concealment, and interactive techniques.

Almost all forward techniques increase the bit rate since they add redundancy to data. Some of them may also require modifications of the encoder. Most interactive techniques need a feedback channel between the encoder and the decoder [5]. Interactive techniques will also introduce some delay and may therefore be unsuitable for real-time applications like mobile video communications. On the other hand, concealment techniques do not increase the bit rate, do not require any modifications of the encoder, and do not introduce any delay. This makes them a very attractive choice for mobile video communications. The Error concealment methods we can divided into two categories. Error concealment methods in time domain and in space domain. But the most effective methods are hybrid error concealment methods, which were created as a combination of temporal and spatial methods. One of the temporal-spatial algorithms is described in the next chapter.

IV. An advanced temporal-spatial error concealment algorithm for H.264/AVC video coding

Proposed modified algorithm for error concealment first explores the temporal correlation between successive frames. If similar blocks to that ones that are neighbors of the missing macroblock can not be found, then the spatial-based error concealment algorithm is used [3].

In order to find an estimate of the missing MB, 8x8 subblocks adjacent to it are used - U1, U2, R1, R2, B1, B2, L1 and L2, as is shown in Fig.3. First, for each of these subblocks, a matching subblock in the previous frame is determined. This matching subblock is found by searching a small area around the point corresponding to the center of each of the subblocks in the previous frame. The sum of absolute differences is used as the measure of similarity. Four of eight of these corresponding subblocks, namely U1', U2', R1', R2', B1', B2', L1' and L2' in the previous frame are shown in Fig.4. Then, eight blocks, namely X_U1, X_U2, X_R1, X_R2, X_B1, X_B2, X_L1, and, X_L2 which are connected to U1', U2', R1', R2', B1', B2', L1' and L2', respectively, are determined. The sum of squared border errors, between the estimated macroblock and its closest blocks, is computed for each of these eight blocks. One block from the above eight blocks, which value of sum is the smallest (is the most similar to lost macroblock) is chosen as a candidate to replace the lost macroblock. Thus, calculations are realized using (1) and (2) :

$$\hat{X} = \arg_{X\_U1, X\_U2, \ldots, X\_L2} \min \varepsilon^2 \tag{1}$$

$$\varepsilon^2 = \varepsilon^2_U + \varepsilon^2_R + \varepsilon^2_B + \varepsilon^2_L \tag{2}$$

Each of the border errors is defined in terms of adjacent pixels by (3), (4) :

$$\varepsilon^2{}_U = \left\| (\hat{x}_U - p_U) \right\|^2, \quad \varepsilon^2{}_R = \left\| (\hat{x}_R - p_R) \right\|^2 \tag{3}$$

$$\varepsilon^2{}_B = \left\| (\hat{x}_B - p_B) \right\|^2, \quad \varepsilon^2{}_L = \left\| (\hat{x}_L - p_L) \right\|^2 \tag{4}$$

The vectors $p_{U,}$ $p_{R,}$ $p_{B,}$ and p_L consist of the outside boundary pixels of the upper, right, bottom and left sides of the missing macroblock, respectively. The upper, right, bottom and left inner boundary pixels of the candidate macroblock are represented by the vectors $\hat{x}_U$, $\hat{x}_R$, $\hat{x}_B$ a $\hat{x}_L$.[3].

Once we have chosen a macroblock, we need to test its integrity as a suitable substitute to the lost macroblock. To do this, we compare the parameter ε2 with a local threshold. If ε2 is larger than this threshold, then we drop temporal concealment and use the spatial concealment. The local threshold is computed for each lost MB by calculating the sum of square distances of the inner boundary pixels of the chosen macroblock in the previous frame and the outer boundary pixels of that macroblock in the same frame , as is shown in Fig.5.

The choice of boundary pixels is made according to the condition of the neighboring macroblocks of the lost macroblock in the current frame. Fig.5 shows that if any of the neighboring macroblocks is corrupted, then the pixels that lie on that boundary of that macroblock, are disregarded in the calculation of the local threshold and the ε2 parameter.

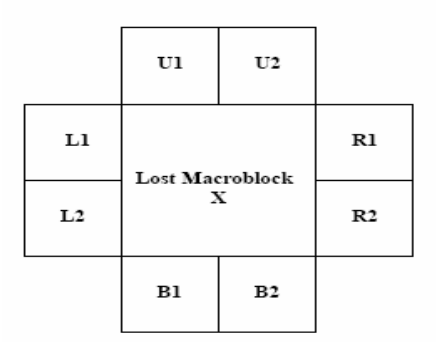


Fig. 3. 8 subblock adjacent to the missing macroblock

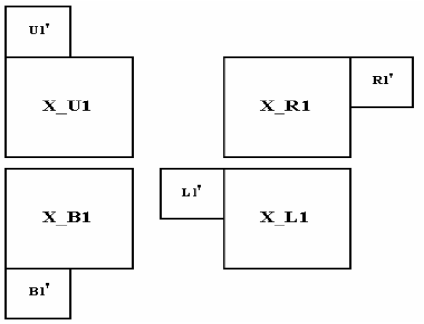


Fig. 4. 4 of 8 corresponding subblocks in the previous frame and candidate macroblock connected to them

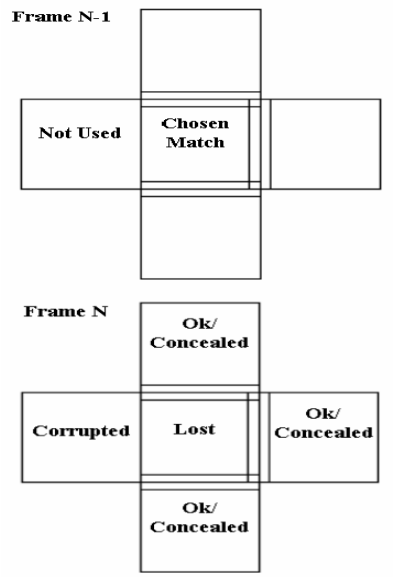


Fig. 5. Choice of boundary pixels used in calculating the local treshold and the boundary matching parameter ε2.

V. Simulation results

Presented algorithm was simulated in computing environment Matlab using two model video sequences : "Salesman" and "Mobile".

Video sequence 'Salesman' is video sequence containing little amount of movement. The background of the scene is static. Movement in this frame is represented by the movement of man's head and by his hands, which are still on the move. "Fig.6" shows where were lost macroblocks generated. Since the background is static, this part of the frame was concealed almost perfectly, as shown Fig.7. Some undesirable artifacts occured in the area of man's head, but during watching this sequence is almost impossible to recognize them.. The efficiency of our temporal-spatial algorithm is very good, as shown values in Table I.

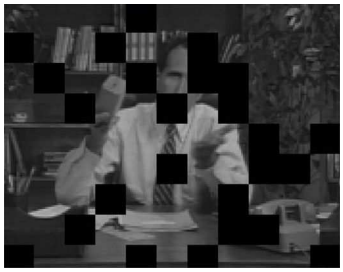


Fig. 6. Lost macroblocks in frame No.208 of the videosequence "Salesman"

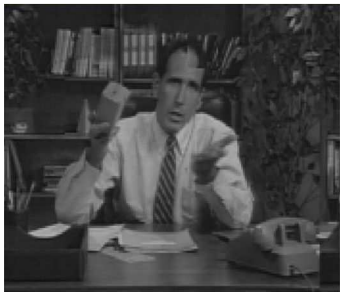


Fig. 7. Concealed frame No.208 of the videosequence "Salesman"

TABLE I
OBJECTIVE CRITERIA AND THEIR VALUES REPRESENTING THE QUALITY OF RECONSTRUCTED INFORMATION FOR FRAME NO.208 OF VIDEO SEQUENCE 'SALESMAN'.

MSE	MAE	NMSE	**SNR [dB]**
0.00023494	0.0034096	0.002794	**28.5377**

The second concealed frame presented in this paper is frame no.276 of the video sequence 'Mobile', as show Fig.8 and Fig.9. There are more types of movement on this frame. Movement of the train, movement of the calendar and movement of the pendulum, finally. All this types of movement make concealment of this frame more difficult. According to that, algorithm exploits particularly spatial concealment technique. Occurred artifacts are almost invisible, efficiency of used algorithm is good. Table II. shows objective criteria representing the quality of reconstructed information.

TABLE II
OBJECTIVE CRITERIA AND THEIR VALUES REPRESENTING THE QUALITY OF RECONSTRUCTED INFORMATION FOR FRAME NO.276 OF VIDEO SEQUENCE 'MOBILE'.

MSE	MAE	NMSE	**SNR [dB]**
0.0005597	0.004375	0.002646	**25.7737**

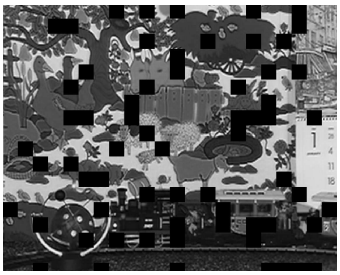


Fig. 8. Lost macroblocks in frame No.276 of the videosequence "Mobile"

Fig. 9. Concealed frame No.276 of the videosequence "Mobile"

VI. CONCLUSION

In this paper we present a low complexity but effective error concealment method. This algorithm first checks whether temporal concealment is feasible for frames. If not, then spatial concealment method implemented in the test model is used [6]. This algorithm does not require very complicated computations, hence is usable for various applications. We compared our simulation results with the simulation results obtained using only spatial error concealment algorithm, with weighted averaging method. We can declare our temporal-spatial algorithm as more effective in case of videosequence containing significant movement. Our simulation results have proved, that by exploiting temporal correlation between adjacent frames, error concealment can be improved significantly.

ACKNOWLEDGMENT

This publication is the result of the project implementation Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF and Grant of Ministry of Education and Academy of Science of Slovak Republic VEGA under Grand No.1/0045/10.

REFERENCES

[1] S. Kumar, L. Xu, K.M. Mandal, S. Panchanathan "Error resiliency schemes in H.264/AVC standard," *Elsevier J. of Visual Communication & Image Representation*,vol. 17(2), April 2006, 26 p.

[2] O. Nemethova, A. Al-Moghrabi, M. Rupp, "Error concealment methods for video transmission over wireless networks," *Diploma thesis*, Institutfur Nachrichtentechnik und Hochfrequenztechnik, Wien, May 2005, 86p.

[3] P. Nasiopoulos, L. Mendoza, H. Mansour, A. Golikeri, "An improved error concealment algorithm for intra-frames in H.264/AVC," *Circuits and System, 2005. IEEE International Symposium*, vol (1), pp.320-323.

[4] T. Stockhammer, M.M. Hannuksela, T. Wiegand, "H.264/AVC in wireless environments," *IEEE Transaction on Circuits and Systems for Video Technology*, vol.13, No.7, July 2003, pp.657-671.

[5] C. Y. Lee, Y. Altunbasak, R Mersereau, "A temporal error concealment method for MPEG coded video using a multiple-frame boundary matching algorithm," *Image Processing 2001. Proceedings. 2001International Conference*, vol.1, pp.990-993.

[6] Y. K. Wang, M. M. Hannuksela, V. Varsa, A. Hourunranta, M. Gabouj "The error concealment Feature in the H.26L Test Model," *in Proc. ICIP* , vol.2, September 2002, pp.729-732.

Extensible host language for DSL development

Sergej CHODAREV

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

sergej.chodarev@tuke.sk

***Abstract*—Area of domain specific languages is subject of active research. In this paper is presented proposal to create special extensible language, which can be used as a base for family of domain specific languages. This approach will allow reuse of common syntax, semantics and tools and will provide simple way to create new domain specific languages.**

***Keywords*—domain specific language, extensibility, modularity, software reuse.**

I. Introduction

Software engineering permanently faces growing complexity of solved problems and change of environment and requirements [1]. To deal with these forces we need to use higher levels of abstraction and reuse software components. Perspective way of raising level of abstraction is provided by domain specific languages.

II. Domain specific languages

Domain specific language (DSL) can be defined as a computer programming language of limited expressiveness focused on a particular domain [2]. Limited expressiveness there means limited features compared to general purpose languages (GPL), but DSL can be much more expressive and easy to use in its application domain by providing notations and constructs tailored toward this domain [3].

Main reason why DSLs are used is improvement of developers productivity. They provide higher level of abstraction and allow to express solutions in particular domain more clearly and concisely then general purpose languages. This helps developers to write code faster and simplifies maintenance. In addition DSLs can improve communication with domain experts, because language would operate directly with concepts from the domain.

There are three main styles of domain specific languages development: [2]

- *External DSLs*, which use a different language to the main language of the application that uses them.
- *Internal DSLs*, which use the same general purpose programming language that the wider application uses, but uses that language in a particular and limited style.
- *Language Workbenches*, which are IDEs designed for building DSLs.

A. External DSL

External DSL give language designer full control of appearance and behavior of the resulting language. It allows to choose syntax, which will be the most appropriate for expressing domain. But this advantage costs more expensive development, since it is needed to develop parser and translator/interpreter for new language.

Other option is to use existing syntax like XML or YAML, which allows using existing parser. But this approach naturally restricts syntax of DSL.

Development costs also include development of tools for new language such as editor or debugger. It is simple to define syntax highlighting in some existing editor. But development of more powerful development environment with features like code completion and incremental syntax checking require notable amount of work.

B. Internal DSL

Internal DSL can reuse parts of syntax and semantics of the host language, which simplifies development. In addition internal DSL can reuse existing infrastructure of the host language including development environment and documentation. It may also simplify learning DSL for its users, if they already know the host language.

Disadvantage of this approach is that internal DSL is restricted by the host language, which may add unnecessary noise to the DSL. Some languages, like Ruby or Lisp, have more flexible syntax which is more suitable for embedding internal DSL. But a lot of popular languages, like Java or C++, have less flexible syntax, so resulting DSL may be very limited and contain a lot of syntactic noise.

III. Extensible host language

As you can see, there is still area for improvement. Process of creating DSL and related tools should be simplified to allow developers to use potential of language oriented development.

As was mentioned, one way to achieve simplification is to use an existing syntax for DSL. This syntax should be simple and flexible, to allow express constructs of DSL without much noise. XML is not very suitable for this purpose, since it is in first place markup and data language and its syntax is too heavy. Possible candidates for the host syntax are s-expressions from Lisp or syntax derived from some programming language (like Ruby or Haskell).

Common syntax would also allow to create tools which could be shared by all DSLs based on this syntax.

Usage of an existing syntax will surely simplify language development, but this is not all we can do. A lot of languages may share parts of common constructs like arithmetic operators or conditional expressions. It is desirable to reuse these parts and free language developer from repetitive work. There should be possibility to share parts of languages. These parts should be therefore separated into reusable modules.

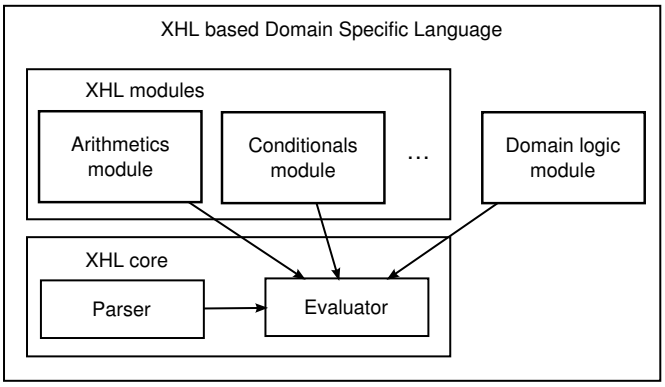


Fig. 1. Scheme of XHL based language

I will call this concept Extensible Host Language (XHL). This will not be a single language, but rather whole family of domain specific languages, which will share basic syntax and semantics, but will provide constructs specific for their domain.

With such host language, all we need to create new DSL is to select a set of predefined modules and provide our own modules for domain logic.

As you can see on Fig. 1, language created with XHL will consist of three main parts:

1) XHL core,
2) XHL modules,
3) user defined modules.

A. XHL core

XHL core will provide common parts for all new DSLs. Specifically it will define:

- Syntax for literals of basic data types and semantics for these types (integers, floating-point numbers, strings, booleans, lists, dictionaries).
- Syntax for identifiers (symbols).
- Syntax for expressions and statements.
- Evaluation mechanism, which defines the way how DSL code is evaluated. Evaluation functions for concrete operations and statements are provided by modules.

These are things needed in almost every programing language, so it is reasonable not to repeat them for every DSL, but rather share common core.

B. XHL modules

Everything besides the core will be provided by modules. Modules will define concrete statements and operators, which will be used in the language. Modules will be written in general purpose language, so DSL statements could be interconnected with other parts of the developed software system. It may be also possible for module to extend the syntax of the language.

Since a lot of statements and operators may be common in many DSLs, XHL will provide set of modules for this common functionality. They may provide:

- arithmetic operators,
- string or list operations,
- conditional expressions and loops,
- variables or constants,
- function definitions, etc.

C. XHL API

Important part of XHL will be its application programming interface (API) provided for development of modules and interaction with rest of the program. This API should provide simple way to create new DSL constructs in a similar way as normal functions or methods are defined. This includes checking type of parameters.

There should also be possibility to influence evaluation process. This is needed for example to create control structures, like branching and loops.

D. Advantages

XHL will share several advantages with internal DSLs [4]. Specifically it will allow to reuse a lot of ideas and artifacts including:

- *Syntax.* All DSLs built using XHL will share common basic syntax.
- *Semantics.* XHL will define not only syntax, but also semantics for common language elements.
- *Tools.* There can be common tools for all DSLs.
- *User knowledge.* User of DSL would benefit from the fact, that several DSLs will share common syntax, basic semantics and tools.

But in contrast to internal DSLs, XHL can be used even with languages with poor support for internal DSL development. Although it is also possible to embed general purpose extension language with more flexible syntax into programs.

Syntax of XHL can be specially designed for building DSLs, so it may be more appropriate for this purpose then most of general purpose languages.

In internal DSL, user of the language has access to whole host language. Sometimes this is not desirable. In contrast to this, XHL will provide only features selected by DSL designer. This can simplify operations like domain specific checking.

Compared to XML based external DSLs, XHL will provide cleaner and more concise syntax and API optimized for DSL development.

Mechanism of modules will also provide a way for combining different domain specific languages.

E. Disadvantages

XHL will not be suitable for all types of DSLs. Some would benefit from closer integration with general purpose language provided by internal DSL. For other languages XHL syntax could be too restrictive, so it would be more appropriate to use external DSL. Especially in cases, where predefined modules are not needed, it may be not much harder to develop new language with custom syntax.

IV. Prototype

As a demonstration of presented concepts I have developed prototype of XHL. Prototype uses syntax similar to Lisp. This syntax is very flexible, which is confirmed by the fact, that Lisp users are known for their long tradition of developing and using internal DSLs [5].

Similarly to Lisp, code consists of lists enclosed in parentheses. First item of the list should represent function name and rest of the list its arguments. Disadvantages of this syntax are inconvenient prefix form of operators and necessity to enclose all expressions in parentheses.

Prototype is implemented in Java. Modules are represented by Java classes. DSL functions are defined using normal Java methods marked with annotation `@Function`. In process of evaluation these methods are executed using Java reflection mechanisms. It is also possible to disable evaluation of arguments, so they can be processed as needed.

A. Example

On Fig. 2 you can see example DSL for simple 2D drawing. This language allows to draw simple 2D shapes using specified color. Domain logic module defines functions like `rgb`, `setcolor`, `draw`, and `rectangle`, but language also uses functions from predefined modules:

- `define` for defining constants and
- arithmetic operators (+, −, ⋆, /).

```
(define black (rgb 0 0 0))
(define gray (rgb 124 124 124))
(define base 50)
(define a 100)
(define a2 (* a 2))
(define pos (+ base (/ a 2)))
(define d 15)
(define square (rectangle a a))

(setcolor gray)
(draw (- pos d) (- pos d) square)
(draw (+ pos d) (+ pos d) square)
(setcolor black)
(draw base base (rectangle a2 a2))
```

Fig. 2. Example drawing DSL developed with XHL.

Implementation of this language is very simple providing that drawing functionality is already in place. In this case was used Java Graphics2D API with some helper classes. On Fig. 4 is presented part of the implementation of drawing module. You can notice that DSL functions are taking normal Java types as their parameters.

V. Related works

Area of domain specific languages is subject of active research. A lot of research is focused on improvement of parser specification and generation technologies, for example YAJCo parser generator [6]. This allows to simplify the process of language development. In contrast to these approaches, XHL has fixed syntax and parts of semantics.

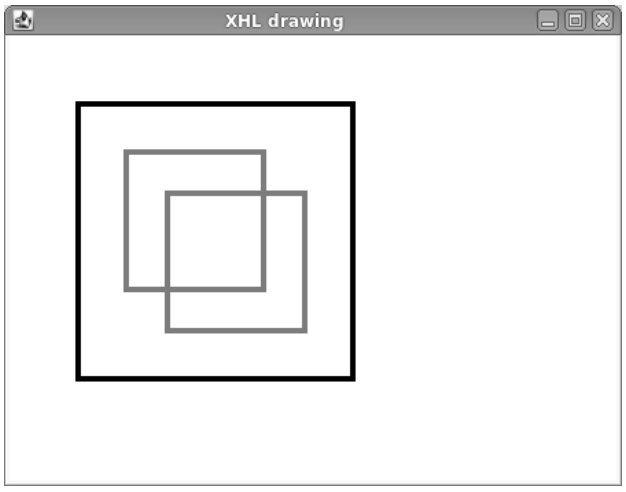


Fig. 3. Result produced by example drawing DSL script.

```
public class DrawingModule
        extends GenericModule {
    ...

    @Function
    public RectangularShape rectangle(
            double width, double height) {
        return new Rectangle2D.Double(
            0, 0, width, height);
    }

    @Function
    public void draw(double x, double y,
            RectangularShape shape) {
        ColoredShape cs = new ColoredShape(
            shape, color);
        cs.setLocation(x, y);
        canvas.add(cs);
    }
}
```

Fig. 4. Part of the implementation of XHL module for simple 2D drawing.

Similar approach to XHL is presented by Gel (Generic extensible language) [7]. This language is suitable as host syntax for DSL, but does not provide unified semantics and evaluation mechanisms as XHL. It should be possible to use Gel as syntax for XHL.

Other perspective area is presented by Language Workbenches, which provide IDE for language development [5]. They do not concentrate on textual form of the language, but use abstract representation as main form and allow language to have several editable projections (including graphical).

VI. Conclusion

Concept of Extensible host language for DSL development presented in this paper may allow simplified development of domain specific languages.

Possible area of future research is design of the host language syntax. It should be flexible, but at the same time simple and not much cluttered. Interesting question is a choice of typing system. Dynamic typing may provide better flexibility, but static typing allow better checking of DSL source codes.

References

[1] J. Greenfield, K. Short, S. Cook, and S. Kent, *Software Factories: Assembling Applications with Patterns, Models, Frameworks, and Tools*. Wiley, 2004.

[2] M. Fowler. (2009) Domain specific languages. [Online]. Available: http://martinfowler.com/dslwip/

[3] M. Mernik, J. Heering, and A. M. Sloane, "When and how to develop domain-specific languages," *ACM Comput. Surv.*, vol. 37, no. 4, pp. 316–344, 2005.

[4] P. Hudak, "Modular domain specific languages and tools," in *ICSR '98: Proceedings of the 5th International Conference on Software Reuse*. Washington, DC, USA: IEEE Computer Society, 1998, p. 134.

[5] M. Fowler. (2005) Language workbenches: The killer-app for domain specific languages? [Online]. Available: http://martinfowler.com/articles/languageWorkbench.html

[6] J. Porubän, M. Forgáč, and M. Sabo, "Annotation based parser generator," in *Proceedings of the International Multiconference on Computer Science and Information Technology*. Los Alamitos, USA: IEEE Computer Society Press, 2009, pp. 705–712.

[7] J. Falcon and W. R. Cook, "Gel: A generic extensible language," in *Proceedings of the IFIP TC 2 Working Conference on Domain-Specific Languages*. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 58–77.

Problem of the sequential algorithm for computer emulation

[1]*Peter JAKUBČO,* [2]*Milan VRÁBEĽ,* [3]*Eva DANKOVÁ*

[1,2,3]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]peter.jakubco@tuke.sk, [2]milan.vrabel@tuke.sk, [3]eva.dankova@tuke.sk

***Abstract*—In this paper, classical sequential algorithm for digital computers emulation is presented. The sequentiality of the algorithm results from the nature of control-flow processors, that are widely used in the present time. The standard algorithm works in a loop, and there it calls procedures for the CPU and the other computer components emulation. The sequentiality of the algorithm can have negative impact on the quality of the whole emulation. This problem is solved in the emuStudio platform in relatively effective way. The effectivity of the solution depends on some assumptions that the emulated components should accomplish.**

***Keywords*—CPU, computer emulation, control-flow processor, digital computer, sequential algorithm**

I. Introduction

The emuStudio emulation platform is a software developed at the Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice.

Design aim was to create a platform, that should be able to emulate various computer configurations, that are based on the architecture with von Neumann conception [1].

The main motivation was to create a support tool intended for the education process in the area of machine languages and computer architectures. There was taken into account the respect for simplicity, configurability and transparency, preserving the original functionality of emulated computer. In the present time the emuStudio platform is used in the education process at the Technical University of Košice for the third year by now.

II. emuStudio platform

The essential feature of the emuStudio platform allows to create and emulate various virtual architectures dynamically (it is an universal platform). Each virtual architecture is represented by interconnected and independent plug-ins, that represent individual components of the emulated computer. The software has been developed in the Java language, what has some advantages and disadvantages [2].

The user interactively chooses all the components needed for the designed computer architecture. Then the user interconnect some of the components in the abstract scheme editor and in such way designed computer is finally opened in the main module. The interactive creation of virtual computer configurations is possible thanks to universal communication model of plug-ins [3].

In the present time, there are only a few emulators created for the platform: a real MITS Altair8800 [4] in two variations, abstract RAM machine [5] (as the successful experiment of emulating other than von Neumann architecture), and own architecture called BrainDuck. More information can be found in [2], [6].

III. Basic structure of an emulator

The structure of any emulator depends mainly on the orientation of the emulated architecture. In the next, we will focus on instruction oriented architectures (ISA) [7].

The main activity of running emulation is repeated execution of the same sequence of steps, because the work of real running computer can be viewed as a repeated sequence of activities.

A CPU (processor) fetches, decodes and executes instructions in infinite loop. The other computer components are directly or indirectly controlled by the processor. Their activities are independent, parallel and can perform in synchronous or asynchronous way with the work of the CPU.

The implementation of parallel and asynchronous activity of emulated component is more difficult, because the starting time of the activity is not predictable (and it should not be), and depends on events that fire in the component. The nature of the most host processors does not allow the true parallel work – if we want to run the component in true parallel with CPU, it must not to be executed on the same host processor (if the processor has only one physical core). Mostly, the host processor does not have enough physical cores, where each component can be executed on it individually, each in separated thread (threads have to be supported in the host operating system).

Therefore an acceptable solution would be to use one thread per component, where the threads are executed on the same physical core of the host processor.

Generally, the nature of control-flow processors (here belong all computers of von Neumann type) defines sequential execution of programs. Seeing that an emulator is also a program executed on the computer, its activity (even if repeated), has to be sequential. More information can be found in [8].

The work of the emulator will be therefore described by a *sequential* algorithm. Inside the algorithm a scheduler needs to be implemented. The scheduler will allot the host processor time to individual emulated components, including emulated processor, in exactly defined order.

The basic algorithm of the emulator's work, using a simple scheduler, is described in the pseudo-code shown in Fig. 1.

In the algorithm, a specific time interval is allotted to every emulated component always in the same sequence. The

```
running = true;
(numberOfCycles, maxTime) = initTimeSyn();
while (running) {
    running = executeCPU(numberOfCycles);
    ints = generateInterrupts();
    emulateDevices(ints, maxTime);
    (numberOfCycles, maxTime) = timeSyn();
}
```

Fig. 1. Basic emulator algorithm (sequential)

component can perform an activity in that time interval . The time interval is estimated by the emulator, and the estimation depends on required emulation speed.

For the sake of clearness, lets imagine how a preemptive process scheduler works in an operating system. Each such scheduler allots fixed time interval to processes, allowing them to run in the time interval. When alloted time fires up, the operating system stops the running process and the next process is launched, with new allotted time interval.

Emulators work similarly, however when the time interval fires up, time synchronization must be performed, because it is not guaranteed that the components will be executed in the exact boundary of the time interval (sometimes the emulation can not be stopped right after fired time interval - it can result from the nature of the emulated component).

Usually the time interval is estimated for one emulation iteration for all components together, and finally the time synchronization is performed.

The time interval presented in Fig. 1 is divided into two parts:

- `numberOfCycles` - the maximal number of machine cycles of emulated processor that should be executed in single emulation step;
- `maxTime` - the maximal common time interval, in which all devices can perform activities in the single emulation step.

Besides the CPU and devices emulation, there are also external interrupts generated that come from devices. They are represented by the set of corresponding state variables in the CPU emulator. The variables are set only if certain conditions are accomplished. The service of the interrupts is realized outside of the CPU execution.

The time synchronization, if needed, ensures the balance of the emulation speed, so the emulation speed would be as close to required speed as possible. The final emulation speed (desired) *mostly* come out from real parameters of emulated components.

IV. The Algorithm Problem and its Solution in emuStudio Platform

The algorithm presented in the previous section (Fig. 1) deals with a problem, however it has not to be always exposed. As was said above, the algorithm is sequential. It means that steps of the algorithm are executed in an ordered sequence.

The `executeCPU` procedure interprets instruction flow and handles data flow. The input parameter of the procedure is the number of machine cycles of the CPU, during what this unit processes instructions and data. This number is usually big enough for an execution of hundreds instructions though (a single instruction consists of several machine cycles).

The problem is, that in a single call of the `executeCPU` procedure, there can be executed even several instructions working with the same device. As the device emulation procedure (`emulateDevices`) lies out of the `executeCPU` procedure, the evident become only the last such instruction. Therefore the device can not respond to the change immediately - the sequential algorithm forces the device to ”wait” for its time.

As an example, let’s present a situation where the processor communicates with a graphic card (Fig. 2). The processor in one emulation step may tell the graphic card to draw a point onto the screen, two times at the same position, but with different colors. In the real computer the user would see the change, but he would not in the emulator, because the graphic card will draw only the last point that appears in the performed emulation step of the processor.

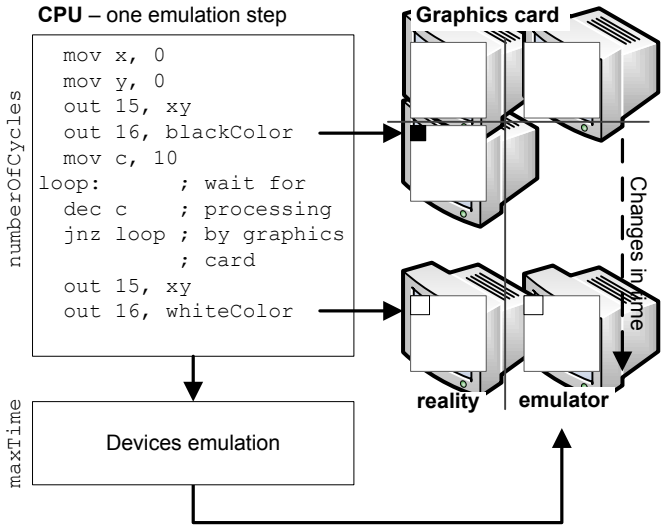


Fig. 2. Example of sequential emulation algorithm problem

There exist several possible solutions for the problem. The universal one is to use a queue that will store all changes that should be performed in a device, in the exact order. The queue would be internal part of the device. When the emulator allots the time to this device, the device, step by step, dequeues the changes and performs them. The device can own several queues though and their implementation can vary, depending on the way how the processor communicates with the device.

However, the disadvantage of such solution is a fact that the changes can ”pulse” from the user’s point of view. At first, the queue ”fills up” - in that time the user does not see any changes. Then the queue ”flushes” and the changes will appear gradually. If the filling/flushing time ratio is bigger than 1, long periods of inactivity will alternate with too quick changes.

In the emuStudio emulation platform another solution is used, depending on the ability of the host operating system (or at least Java) to work with threads.

The execution of each input-output instruction causes immediate call of corresponding input/output of the emulated component. The identification of the component is ensured by the communication model [3] that assumes quick response from communicating components.

If the response of the device will not be quick enough for the processor, the slow activity should be executed in an individual thread and then the execution should return back to

the processor. In repeated access to the device, another thread will be created for performing the input/output operations. However there exist cases when this approach is not possible to realize, however the cases exist only for the input operation (e.g. when processor needs the result immediately in order to complete current instruction execution). In any case, the device owns the responsibility for potential synchronization requirements.

Too slow responses of the devices can lead to system destabilization. There exists an empirical rule that says if the response will bound to one third of the time needed for `numberOfCycles` cycles execution, the latency influence to the system stability would be negligible.

The pseudo-code implementation shown in Fig. 1 does not depend on used emulation technique. The proposed solution affects the `executeCPU` procedure in a way that the `emulateDevices` procedure (in modified form) will be nested in a certain way to the `executeCPU` procedure.

V. CPU EMULATION

The core of an emulator is a processor, or CPU. However, ways by which the emulator executes demanded code can vary.

Even though the emuStudio platform does not limit the programmer to choose the processor emulation technique, in each implemented processor in the present time there is only interpretation technique used (the simplest one). The skeleton of the processor emulation algorithm (implementation of the `executeCPU` procedure), using interpretation as the emulation technique, is shown in Fig. 3.

```
cycles = 0;
while (cycles < numberOfCycles) {
    opcode = fetch_opcode(PC);
    PC = PC + 1;
    switch (opcode) {
        case MOV:
            ...
            cycles = cycles + MOV_CYCLES;
            break;
        ...
        case IN:
            port = fetch_operand(PC);
            PC = PC + 1;
            A  = readIO(port);
            cycles = cycles + IN_CYCLES;
            break;
        case OUT:
            port = fetch_operand(PC);
            PC = PC + 1;
            writeIO(port, A);
            cycles = cycles + OUT_CYCLES;
            break;
        ...
}
```

Fig. 3. CPU emulation using interpretation algorithm (*executeCPU* procedure)

In the algorithm the devices emulation uses two procedures: `readIO` and `writeIO`. The purpose of the procedures is to:

- identify a device connected to the port (the port is an operand of the input-output instruction),
- call the input/output of the device in the way that is defined by the communication model (transfer data).

A quick response is important for just these two procedures.

VI. CONCLUSION

In the present time the emulation thematics is actual enough, mainly its special case - virtualization.

The principle of the technique is to emulate as few processor instructions as possible, and let the rest of them to execute natively on the host machine.

However, the virtualization uses some emulation technique, too, and its similarity is indisputable with sequential emulation algorithm described in this paper.

The paper focuses on the description of basic computer emulation algorithm. We can see its modification in any emulator and therefore it plays an important role.

By the nature of control-flow processors, that are controlled by the instruction flow in an exact order, the algorithm is sequential.

Processor emulation in the standard algorithm is separated from the other computer components emulation, by what the individuality and independency of the components is emphasized. The components exchange the data to each other by certain communication interface.

However this standard approach can have negative impact on emulation quality, what the paper shows in an example.

The emuStudio emulation platform, that is used mainly in the education process, solves the problem effectively. It combines the processor emulation with the emulation of the other components in a way that the communication between the processor and the component is performed on demand, and quick response of the component is required.

REFERENCES

[1] J. von Neumann, "First Draft of a Report on the EDVAC," 1945. [Online]. Available: http://www.virtualtravelog.net/entries/2003-08-TheFirstDraft.pdf

[2] P. Jakubčo and S. Šimoňák, "EmuStudio a plugin based emulation platform," *Journal of Information, Control and Management Systems*, vol. 13, no. 1, pp. 33–46, 2009, ISSN 1336-1716.

[3] P. Jakubčo and M. Domiter, "Standardization of computer emulation," in *SAMI 2010 Proceedings, 8th International Symposium on Applied Machine Intelligence and Informatics*, 2010.

[4] MITS Inc., "Altair 8080 Operators Manual," 1975. [Online]. Available: http://www.classiccmp.org/dunfield/altair/d/88opman.pdf

[5] T. Kasai, "Computational Complexity of Multitape Turing Machines and Random Access Machines," *Publications of the Research Institute for Mathematical Sciences*, vol. 13, no. 1, pp. 469–496, 1977. [Online]. Available: http://projecteuclid.org/DPubS?service=UI&version=1.0&verb=Display&handle=euclid.prims/1195189815

[6] S. Šimoňák and P. Jakubčo, "Software based CPU emulation," *Acta Electrotechnica et Informatica*, vol. 8, no. 4, pp. 50–59, 2008, ISSN 1335-8243.

[7] M. Jelšina, *Architectures of computer systems: principles, structural organisation, function (in Slovak)*. Košice: Elfa, s.r.o., 2002, ISBN 80-89066-40-2.

[8] V. M. Barrio, "Study of the techniques for emulation programming," 2001, computer Science Engenieering, Facultat d'informatica de Barcelona , Universitat Politecnica de Catalunya. [Online]. Available: http://apple1.chez.com/Apple1project/Docs/pdf/Study_techniques_for_emulation_programming.pdf

Application of high performance computing in Markerless Augmented Reality systems

Radovan JANOŠO

Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

radovan.janoso@tuke.sk

***Abstract*—This paper focuses on potential areas of high performance computing utilization within one category of Augmented reality systems – Markerless systems. Initially classification of Augmented Reality (AR) systems is provided. Later on problem areas of marker less AR system are discussed. Towards the end proposal for potential solution is provided.**

***Keywords*—Augmented reality, markerless tracking, GPU computation.**

I. INTRODUCTION

These days augmented reality is making it from research labs to everyday life. Application of it to real conditions brings specific requirements to be addressed. This paper aims at providing description where high performance computing system can be utilized within AR system.

Well known taxonomy that addresses concept between real and virtual worlds is based on Paul Milgram [2] and Fumio Kishino [2]: Milgram's Reality-Virtuality Continuum. Continuum is visualized as line that is between reality and virtuality. Figure 1 shows continuum based on Milgram definition:

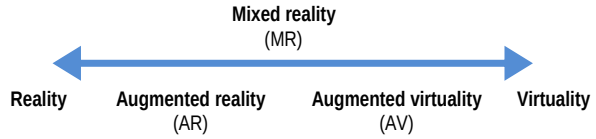


Fig.1 Milgram's definition of real to virtual world transition

Such taxonomy had been extended by Mann [9] primarily for reason of adding modification factor to either reality of virtuality. Augmented reality aims to complement real world and virtual reality to replace it. Mediality by definition brings concept of modification reality or virtuality respectively and therefore adding new categories as mediated reality and mediated virtuality. Such modification can be deliberate one or accidental.

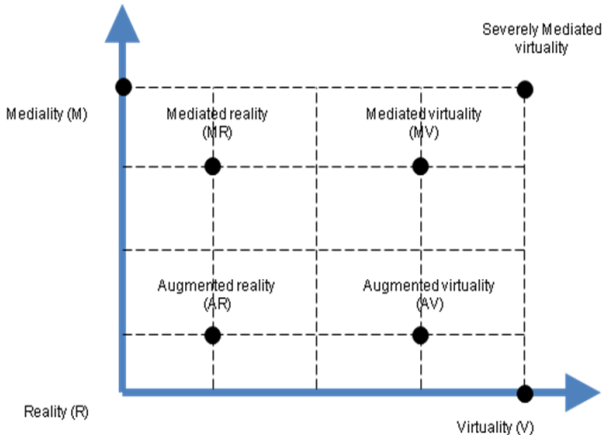


Fig.2 Manns's taxonomy

Augmented reality (AR) system is then system that supplements additional information into real world view. Division of AR systems based on how human observer views added information into real world is:

- Optical see through – semitransparent display is used to display added information to real world view.
- Video see through – real world image is displayed with AR added information. Common setup uses camera – display setup.

According to method how additional information is aligned with real world we recognize these two categories of augmented reality systems:

- Marker based systems – where special markers have to be added to objects and places in real world. AR system then replaces markers with related information in observers view.
- Markerless systems – where AR system needs to identify object and places in real world without using markers. Methods of object recognition with location, orientation and movement information aid are utilized.

Marker based AR system had been developed at DCI FEEI TU Košice in order to research, verify and demonstrate concepts of augmented reality [1]. Specific needs of AR systems for visualization methods were part of studies carried at DCI FEEI and are documented in [3] and [4].

II. MARKER BASED AND MARKERLESS TRACKING

As stated above there are two methods that perform tracking / alignment of real world with virtual objects. Advantage of first method resides in simpler implementation of AR system. Implementation of AR system integration with information system developed at DCI FEEI TU Košice as well as process for creation of marker based AR system is described in [11]. Examples of marker and marker replacement as it happens in AR system developed at DCI FEEI are presented in Figure 3 and Figure 4 respectively.

Fig.3 Example of marker that indentifies physical object (access card reader) located below marker

Fig.4 AR system view with marker replaced by virtual object created by AR system

However some shortcomings of marker based AR systems are obvious:

- Need of physical placement of marker on object to indentify it – not possible all the time
- Complex process of marker size and pattern selection for optimal recognition
- Complex process if marker location selection for fast and accurate recognition
- Optical conditions at place of marker can make it difficult for recognition

Markerless tracking represent more generic way compared to marker based systems. Absence of markers in such system increases complexity and that has direct impact on computational needs towards AR system. In relation to Figure 3 marker less AR system has to be able to recognize object in this case access card reader without aid of marker.

High level block diagram of AR system is presented in Figure 5.

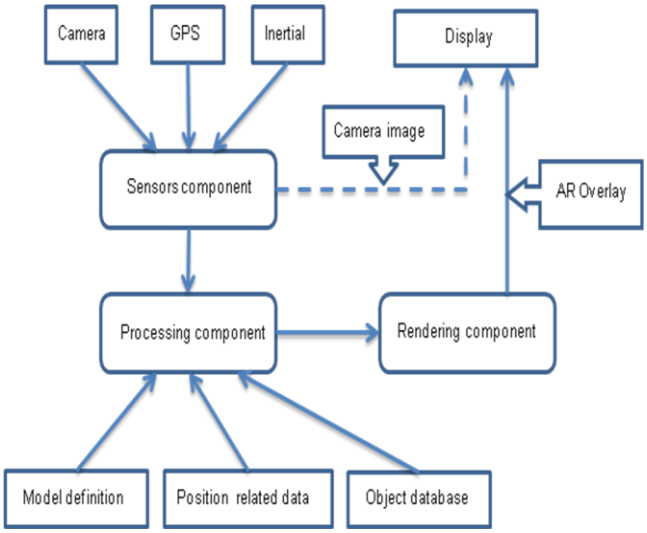


Fig.5 High level diagram of AR system

Sensors component collects information from sensors about environment where observer is located. *Camera* is used as primary data source providing visual information about environment. In case of video see through AR system camera's image is fed to display. *GPS* has quite important role in case of outdoor systems where it provides continuous feed of observer's position to AR system. *Inertial* sensors group includes for example accelerometers, gyroscopes and compass and work as aid to determine observer's position and orientation.

Processing component's role includes:

- process information provided by *sensors component* (perform object detection in images provided by camera, determine and classify observer's movement and orientation)
- compare recognized objects against object database and align position related data
- align processed information against defined model

Once alignment against model is done such data is provided to *rendering component* whose main role is to render overlay. Such overlay contains information created with AR system in relation to real world image.

High level diagram of AR system on Figure 5 can be extended with another dimension that defines whether system consists of *mobile* part (present with observer) or both *mobile* and *back end* (stationary available via network connection) parts. Having mobile part only has advantage in fast access to required data that is carried along. If there is need to have access to extensive data sets that cannot be placed on mobile part only or performance of mobile part is not satisfactory

combined solution is used.

With combined solution following split of components can occur:

- *sensors component* – mobile part only
- *processing component* – mobile and back end parts
- *rendering component* – mobile and back end parts

Model definition, position related data as well as object database can be split or rather mobile part contains subset of information that is stored on *back end* part. Emphasis is on performance of *back end* due to minimizing response time to mobile part. Large data sets and fast response times lead to high performance computing consideration for *back end* part.

III. Application of High Performance Computing in AR Systems

Typical set of tasks that is carried within AR system on top of data collected from sensors can be defined as follows:

- Image filtering and adjustment
- Edge detection
- Object / Text / Pattern / Face recognition
- Large data sets scanning / comparison
- 3D computation – 3D model alignment
- Rendering

Nature of tasks above makes then suitable for parallel processing. Approaches for high performance computing are these:

- computational clusters – many individual computers interconnected via high speed network to perform parallel computation
- GPGPU – general purpose computing on graphics processors utilizing many core setup of GPU
- supercomputers – massively parallel specialized computers

Out of abovementioned approaches GPGPU is one that stands out in case of AR systems. Advantage that GPGPU brings is that it can be used in AR systems with *mobile* part only or in both places for mixed systems where both *mobile* and *back end* parts are present. It is possible due to presence of GPU in *mobile* part that has GPGPU functionality.

Currently these frameworks are available for GPGPU computing:

- CUDA – supported by nVidia hardware [7]
- Brook+ - supported by AMD/ATI hardware [8]
- OpenCL – relatively new standard that is adopted by most of major GPU and embedded GPU manufactures [6]

OpenCL standard defines heterogeneous parallel programming on CPU and GPU. With embedded profile [10] it allows to have parallel computations on embedded type of hardware. One of possible setups for combined AR system is displayed on Figure 6. *Mobile* part of AR system operates on preprocessed data provided from back end part. *Back end* part has access to full set of data.

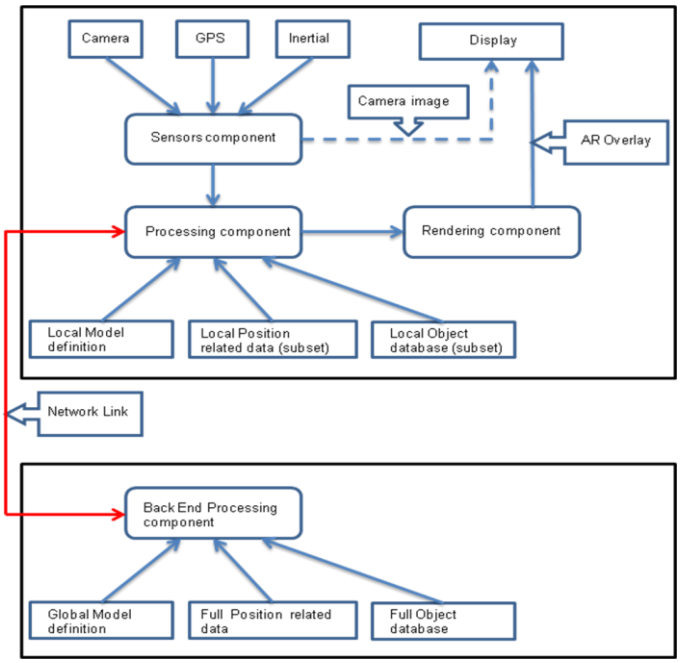


Fig.6 Possible setup of combined AR system

In this configuration Processing component combines information from sensors and provides it to *Back End Processing* component. According to given data *Back End Processing* component performs image processing and retrieves data related to position as well as executes search of full object database for matching entries. All found information is provided back to *processing component* of mobile part. That component updates local databases and sends information to *rendering* component. *Back End Processing* component determines what additional data shall be retrieved. It will do it based on location and inertial data as well as on knowledge of information that has been already sent to mobile part. These additional data are then provided to *processing component* of mobile part to update databases. It allows *mobile* part to have cached data and being able to perform faster. Described process is continuous one.

IV. Conclusion

This paper provides view on certain problems of markerless AR systems and points out direction to address them. More detailed work will be carried out to address individual problems – specifically markerless tracking algorithms and solution to implement combined AR system using GPGPU. Research in area of GPU acceleration for rendering and markerless tracking tasks will be carried out. As well as definition of interface between *mobile* and *back end* part of combined AR system will be established as part of future research.

Acknowledgment

This work is supported by VEGA grant project No. 1/0646/09: "Tasks solution for large graphical data processing in the environment of parallel, distributed and network computer systems"

REFERENCES

[1] Sobota, B., Perháč, J., Straka, M., Szabó, Cs.: Application of parallel, distributed and network computer systems in solving computational problems of large graphical data set processing; elfa Košice, 2009, ps. 180, ISBN 978-80-8086-103-2 (in Slovak)

[2] Milgram, P., Kishino, F.: A Taxonomy of Mixed Reality Visual Displays, IEICE Transactions on Information Systems, Vol E77-D, No.12, 1994, pp. 1321-1329

[3] Balun, M.: Visualization of site model using virtual reality technologies; MSc Thesis, DCI FEEI TU Košice, 2009 (in Slovak)

[4] Varga, M.: 3D model visualization using mixed reality technologies; BSc Thesis, DCI FEEI TU Košice, 2009 (in Slovak)

[5] Sobota, B., Straka, M., Perháč, J.: Some problems of virtual object modeling for virtual reality applications, Journal of Information, Control and Management Systems, Volume 6, No. 1, 2008, pp. 105-112 ISSN 1336-1716

[6] OpenCL specifications and presentations: Khronos group: www.khronos.org

[7] CUDA: reference materials from nVidia : http://www.nvidia.com/object/cuda_home_new.html

[8] Brook+: information gathered from AMD: http://ati.amd.com/technology/streamcomputing/AMD-Brookplus.pdf

[9] Steve Mann: Mediated Reality with implementations for everyday life; August 2002; http://wearcam.org/presence_connect/

[10] Pulli, K.: OpenCL in Handheld Devices ; http://www.khronos.org/developers/library/2009-hotchips/Nokia_OpenCL-in-Handheld-Devices.pdf

[11] Sobota, B., Janošo, R., Korečko, Š.: Using Augmented Reality Technologies in Information Systems.; MOSIS 2010: Hradec nad Moravicí, Czech republic, 27. – 29. April 2010 (in print)

Interactive off-line segmentation of moving objects in real traffic conditions

[1]*Vladimír Jeleň*

[1]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]vladimir.jelen@tuke.sk

***Abstract* – Exploitation of image processing methods in traffic control is becoming more popular in recent years. This article deals with segmentation in real traffic conditions. We used modern "graph-cut" segmentation method which is able to solve more complicated situations like e.g. occluding objects.**

Keywords – motion tracking, segmentation, graph-cuts, threshold, energy minimalization

I. INTRODUCTION

The number of people using personal motor vehicle to move from point A to point B is increasing and the traffic situation in big cities is near the collapse. As a result, more and more car accidents occur. One of possible solutions is to automate traffic control systems. Semiautomatic segmentation with motion prediction can help to analyze critical situations off-line.

In complicated situations like car accidents, objects or their parts are a occluded. It is very difficult to segment occluded objects using simple methods, like thresholding which assumes that brightness values of pixels belonging to objects is significantly different than pixels belonging to the background. Therefore, we have to use method which can incorporate apriori information about regions, borders, occluding and shape. One of such modern methods is Graph-cuts which is able to distinguish occluded objects with the same brightness levels as the background.

II. MATERIALS AND METHODS

A. Acquisition

We used computer with 2GHz Dual Core CPU, 3GB of RAM and 64MB VGA for our segmentation method. It is a common configuration. Several types of optical cameras and webcams for image acquisition was tested in experiment. (Vivotek IP camera, Logitech Web camera...) A common industrial camera placed above a crossroad was used too.

The situation on the crossroad was recorded and saved to AVI format for next processing and it was transferred into the computer, where it was subsequently analyzed by our software.

B. Software

Our application is a plug-in module for image processing software Ellipse and it was created in Microsoft Visual Studio 2008 using C++ compiler. This plug-in module combines segmentation with low level object tracking. Classes supporting Graph-cut calculations are available on web site of Vladimir Kolmogorov [8].

C. Labeling

The application of Min-cut/Max-flow algorithm in image processing first time was described in [1]. This algorithm is very effective for combinatory optimalization and for energy minimalization of many types of energy functions in computer vision.

In the next part basic information will be described about graphs and flows in the context of energy minimalization.

An oriented weighted graph $G = (\mathcal{V}, \mathcal{E})G$ consists of a set of nodes $\mathcal{V}$ and a set of oriented edges $\mathcal{E}$ that represents connections between the nodes. The nodes usually represent pixels or voxels.

A graph contains several additional special nodes called terminals. These nodes are usually marked *s* (source) where $s \in \mathcal{V}$ and *t* (sink) where $t \in \mathcal{V}$.

Terminal nodes, in computer vision, correspond with separated sets of labeled pixels, which can be divided into the *s-t* categories. You can see an example of a traditional *s-t* graph in the figure 1. Values of edges are represented by width of line.

Usually two types of edges occur in a graph: ***n-links*** and ***t-links***.

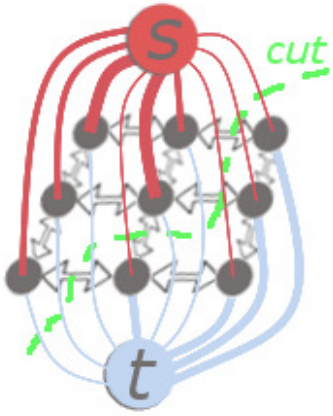


Fig. 1. : Visualization of s-t graph with Source and Sink terminals. Green line represent cut on position with minimal energy.

The first one connects pairs of neighboring pixels or voxels. It is a representation of neighboring system in the picture. Weights of n-links are responsible for penalization of difference in brightness values of neighboring pixels [2] [3].

The ***t-links*** connect pixels and terminals and their values represent the classification of pixels to the nodes *s* and *t*.

D. Segmentation

If $G = (\mathcal{V}, \mathcal{E})$ is undirected bipartite graph defined as set of nodes or vertices and edges ergo connections of neighboring pixels. Then it is possible to describe each pair of connected nodes in this graph by edge $e = \{p, q\} \in \mathcal{E}$.

The nodes in our case represent pixels and we have two terminals in the graph. ***Source*** terminal tagged as S is the representation of all pixels corresponding with an object in the picture and ***Sink*** terminal tagged as T which is the representation of the background. Each pixel is connected to both S and T terminals.

Non-negative weights w_e are assigned to each of the edges $e \in \mathcal{E}$ which contain n-links and t-links.

The *s-t cut* $C \in \mathcal{E}$ is a subset of edges in a graph. The cut divides nodes between terminals of a graph $\mathcal{G}(C) = \langle \mathcal{V}, \mathcal{E} \backslash C \rangle$. This divides the picture to object and background. The energy cost is given as a sum of weights placed on the cut.

$$|c| = \sum_{\varepsilon \in \mathrm{E}} w_\varepsilon$$

In our case *n-links* are placed on the boundary of the object to be segmented therefore their costs represent the cost of that boundary. On the other hand the separated *t-links* represent regional properties of the segment. Ergo cut with minimal cost is balance between border and region properties.

E. Interactive segmentation based on Graph-Cuts

The user marks regions in the picture that belongs to the object and regions that definitely are parts of the background [5] [4]. Number of marked regions depends on the user and the type of the task. Segmentation consists of three parts:

1. For each pixel inside the object is given a value that represents similarity of its intensity with the model. Low values represent better results.
2. For each pixel inside the background is given the value that represent similarity of its intensity with the model. Low values represent better results.
3. For each pair of neighboring pixels, where one is inside the object and the second is outside, is given a value that represents the similarity of intensities of both pixels. Low values represent their similarity.

Let it p be a pixel from the set of pixels P and A_p=*0* or *1* indicating that p is part of the background or the object.

Let it $R_p(A_p)$ be a function of similarity of pixels p (with values 0 and 1).

Let $B_{p,q}$ is the variance of the intensity changes of pixels p, q.

Then segmentation value is given by:

$$E = \mu \sum_{p \in P} R_p(A_p) + \sum_{(p,q) \in N : A_p \neq A_q} B_{p,q}$$

where N is the set of neighboring pixels. The first part of the equation represents the regional property and the second part represents the smoothness of the border and its continuity.

III. Image Analysis and Motion Tracking

A. Image analysis strategy

The first step is conversion of video signal into the stack of images. Each video frame is converted into 8-bit grayscale picture, because information about color is not important and this operation increases rapidly the efficiency of image processing in next step. We proceed with segmentation after this necessary preprocessing.

The user sets a seed line in the first two images from stack. That line must belong to objects. These two regions are used by the algorithm to automatically predict the motion and to calculate regions around the line that will be used for segmentation.

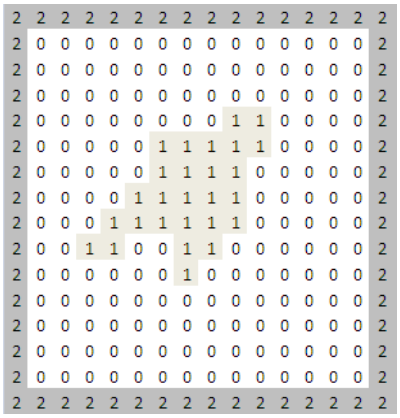


Fig. 2. : Pixels labeled by "1" seed region, pixels labeled by "2" correspond to background pixels. Label "0" have unlabeled pixels.

This setting defines three regions that represent:

- Part of the object (it is the expansion of the seed line and it is shown on Fig.2, it is the region labeled by "1")
- Part of the background (it is the frame of the scanning region shown on Fig.2, it is frame labeled by "2")
- Unallocated region (it is shown in Fig.2 and it is labeled by "0")

The algorithm finds the best border between regions representing the object and background in unallocated region labeled by zeros.

B. Motion tracking strategy

Prediction is the main feature typical for motion tracking. Extrapolation is the simplest form of trajectory prediction where the position of the object in the next frame is given by the position on previous frame shifted by vector (dx, dy) representing the differences of positions on last two frames. There are two important parameters in the model.

$$\bar{v} = (x, y)$$

where *x, y* defines the position of point B.

The difference vector

$$\overline{d} = (dx, dy)$$

this represents differences of individual parameters in last two frames. The predicted values determining the model position, rotation, shape for the present and for the next frame are given

$$\overline{v_{t+1}} = \overline{v_t} + \overline{d}$$

Interactive control of the program is a clear request of people who analyze images [6], [7] of traffic. Program should work as autonomously as possible but under the control of human operator. In critical situations user should be able to correct the algorithm by entering a new corrected seed region.

IV. Experimental results

We found that the graph-cuts based method is Nx slower than the method based on thresholding, but it is more precise. Their comparisons are shown in fig.4 (graph-cuts) and fig.3 (thresholding). The latter method is not useful to analyze of situation on crossroads.

Fig. 3. : Segmentation by threshold based method where object contour detection fails

Fig. 4. : Segmentation by graph-cut based method. Segmented object is occluded by car.

V. Conclusion

Segmentation of real traffic conditions is a very complex problem. It is necessary to consider many factors like natural constrains, quality of recording, objects size or their properties and many similar factors. In this contribution we developed a simple method which is able to trace moving objects off line. It can be useful e.g. for analysis of car accidents where the contour of studied object is detected. First pair of series requires manual labeling of seed region, then algorithm finds the contours on the subsequent frames automatically. Graph-Cuts method in the context of computer vision is robust to situations like occluding parts of objects where simple segmentation methods based on thresholding fail.

We plan to increase the efficiency of the algorithm and fully automate the process of segmentation in the future.

Acknowledgment

This work was supported by Research and Development Support Agency under project APVV-0682-07.

References

[1] GREIG, D. – PORTEOUS, B.SEHEULT, A.: Exact Maximum A Posteriori Estimation for Binary Images, J. Royal Statistical Soc. – Series B, vol. 51, no. 2, pp. 271-279, 1989

[2] ISHIKAWA, H. – GEIGER, D.: Segmentation by Grouping Junctions, Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 125-131, 1998

[3] BOYKOV , Y. – KOLMOGOROV, V.: Computing Geodesics and Minimal Surfaces via Graph Cuts, Proc. European Conf. Computer Vision, pp. 232-248, 1998

[4] LEGAULT, R. – SUEN, CH.: Optimal Local Weighted Averaging Methods in Contour Smoothing, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 19, no. 8, august 1997

[5] TOMORI, Z. – JELEŇ, V. – JANÁČEK, J.: Animal motion tracking using advanced image segmentation methods, 6th International Conference: Structure and Stability of Biomacromolecules, Košice, Slovakia, 2009, pp. 69-70.

[6] P. Das, *et al.*, "Semiautomatic segmentation with compact shape prior," *Image and Vision Computing,* vol. 27, pp. 206-219, 2009.

[7] C. Liu, *et al.*, "Interactive Image Segmentation Based on Hierarchical Graph-Cut Optimization with Generic Shape Prior," Image Analysis and Recognition, Proceedings, vol. 5627, pp. 201-210, 957, 2009.

[8] http://www.cs.ucl.ac.uk/staff/V.Kolmogorov/

Behaviour of software components parametrized by monads

[1]*Marián JENČIK*

[1]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]marian.jencik@student.tuke.sk

***Abstract*—This paper introduces *software components* modelled as concrete coalgebras by endofunctor of the category of Set. Endofunctor is parametrized by strong monad to captured a behaviour model and approach computing result.**

***Keywords*—Software components, coalgebra, monads**

I. Introduction

Expression *Software component* is very often use in computer world and can have a lot of meaning. In this contribution we under term *component* can introduce a number of services through a public interface which provides a limited access to it's internal space. Futhermore, it persists and evolves in time.

Components by [1] are, concrete coalgebras, over a state space supporting a set of observers (destructors) and actions which are explicitly specified. For each inhabitant of the state space, the corresponding behaviour arises by coinductive extension. In fact, *processes*, but not *components*, can be regarded as inhabitants of a final coalgebra. Coinductive types provide an abstraction for behaviours, just as inductive types characterise data.

II. Coalgebras

In computer science, coalgebra has emerged as a convenient and suitably general way of specifying the reactive behaviour of systems. Simple system specified in [2] by endofunctor $\mathcal{H}$: **Set** $\rightarrow$ **Set**, on the category of sets and functions is a pair $(\mathcal{Q}, s)$ consisting of a set $\mathcal{Q}$, set of states and a dynamics $s : \mathcal{Q} \rightarrow \mathcal{H}\mathcal{Q}$. Given coalgebras $(\mathcal{Q}, s)$, $(\mathcal{Q}', s')$ and *coalgebra homomorphism* is a function $f : \mathcal{Q} \rightarrow \mathcal{Q}'$ such that the square commute :

$$\begin{array}{ccc} Q & \xrightarrow{f} & Q' \\ {\scriptstyle s}\downarrow & & \downarrow{\scriptstyle s'} \\ H(Q) & \xrightarrow[H(f)]{} & H(Q') \end{array}$$

For example in the above case of $\mathcal{H} = (-)^{\mathcal{I}} \times \mathbf{bool}$, where $\mathcal{I}$ is a the set of inputs and a next-state function $\delta : \mathcal{Q} \times \mathcal{I} \rightarrow \mathcal{Q}$ is described the concept of functional simulation of deterministic automata.

This concept is suitable for description of simple system but if we want to descibe large systems we have to use *monads.*

III. Monads

The main idea using of *monads* is that computational effects are represented by a type constructor, an endofunctor in a suitable category $\mathcal{C}$ so that computations producing values type $\mathcal{O}$ are regarded as terms of type $\mathcal{B}\mathcal{O}$ representing the computation of values of type $\mathcal{O}$ from values of type $\mathcal{I}$ while producing some effect described by $\mathcal{B}$, in the diferent way, output values arise encapsulated in the effect specified $\mathcal{B}$. Such arrows will be referred to in the sequel as $\mathcal{B}$- computations.

A. Monads

A *monad* in a category consists of an endofunctor $\mathcal{B} : \mathcal{C} \rightarrow \mathcal{C}$ and the pair of natural transformations $\mu : \mathcal{B}^2 \Rightarrow \mathcal{B}$ and $\eta : Id_{\mathcal{C}} \Rightarrow \mathcal{B}$ understood as an associative multiplication and Its unit, such that the diagrams below commute :

$$\begin{array}{ccc} \mathcal{B}^3 & \xrightarrow{\mathcal{B}\mu} & \mathcal{B}^2 \\ {\scriptstyle \mu\mathcal{B}}\downarrow & & \downarrow{\scriptstyle \mu} \\ \mathcal{B}^2 & \xrightarrow{\mu} & \mathcal{B} \end{array}$$

$$\begin{array}{ccccc} Id_{\mathcal{C}}\mathcal{B} & \xrightarrow{\eta\mathcal{B}} & \mathcal{B}^2 & \xleftarrow{\eta\mathcal{B}} & \mathcal{B}Id_{\mathcal{C}} \\ & \searrow & \downarrow{\scriptstyle \mu} & \swarrow & \\ & & \mathcal{B} & & \end{array}$$

that is,

$$\mu \cdot \eta\mathcal{B} = \mathcal{B}\eta = id$$

$$\mu \cdot \mathcal{B}\eta = \eta \cdot \eta\mathcal{B}$$

If we are thinking of $\mathcal{B}$ as the encapsulation of a computational structure, its unit η represents the minimal such structure when a value $o \in O$ is embedded in BO.

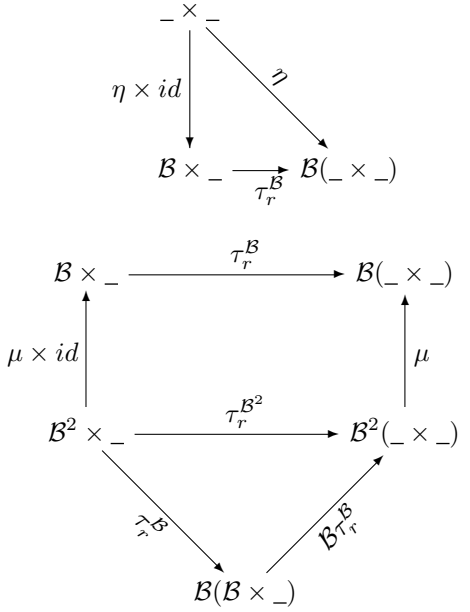
B. Strong Monads

A *strong monad* is simply a *monad* $\langle \mathcal{B}, \eta, \mu \rangle$ where $\mathcal{B}$ is a strong functor and both η and μ strong natural transformations. The characterisation law for strong natural transformations, entails the following additional axioms:

$$\tau_r^{\mathcal{B}} \cdot (\eta \times id) = \eta$$

$$\tau_r^{\mathcal{B}} \cdot (\mu \times id) = \mu \cdot \mathcal{B}\tau_r^{\mathcal{B}} \cdot \tau_r^{\mathcal{B}}$$

which express the commutativity of the following diagrams :



where $\tau_r^{\mathcal{B}^2} = \mathcal{B}\tau_r^{\mathcal{B}} \cdot \tau_r^{\mathcal{B}}$.

The main effect of *strong monads* is to distribute the free variable values in the context (_) along functor $\mathcal{B}$.

C. Powerset monads

Powerset monad $\langle \mathcal{P}, \eta, \mu \rangle$ is monad on the category **Set** with a natural transformations η and μ where for a set $\mathcal{C}$ let $\mathcal{PC}$ be the powerset of $\mathcal{C}$ and for a function $f : \mathcal{C} \to \mathcal{D}$ let $\mathcal{P}(f)$ be the function between the powersets induced by taking direct images under f.

For every set $\mathcal{C}$, we have a map $\eta_C : C \to P(C)$, which assigns to every element c of $\mathcal{C}$ the singleton c. A function $\mu_C \colon P(P(C)) \to P(C)$ can be given as : if $\mathcal{L}$ is a set whose elements are subsets of $\mathcal{C}$, then taking the union of these subsets gives a subset $\mu_{\mathcal{C}(\mathcal{L})}$ of $\mathcal{C}$. These data describe a *monad*.

IV. Coalgebraic models

As it was write above *components* are modelled as concrete coalgebras with specified initial conditions therefore we firstly must choose a family of functors to suitably components interfaces. Staring point is Set endofunctors as follows :

$$\mathcal{T}^{\mathcal{B}} = \mathcal{O}'^{\mathcal{I}'} \times \mathcal{B}(Id \times \mathcal{O})^I$$

where the sets $\mathcal{I}, \mathcal{I}'$ and $\mathcal{O}, \mathcal{O}'$ are, respectively, the input and output which ensure the flow of data and $\mathcal{B}$ is a *strong monad*. Monads play here an essential role as a way to encode in abstract terms different kinds of behavioural effects.

Each $\mathcal{T}$-coalgebra p over carrier $\mathcal{U}$ is written as split a $\langle \overline{o}_p, \overline{\alpha}_p \rangle$, where $o_p : \mathcal{U} \times \mathcal{I}' \to \mathcal{O}'$ is the observer, attribute or output function, and $\alpha_p : \mathcal{U} \times \mathcal{I} \to \mathcal{B}(\mathcal{U} \times \mathcal{O})$ stands for the coalgebra action, method or update function.

By instantiating $\mathcal{T}^{\mathcal{B}}$-interface we will get more specialised form of component. For example, by making $\mathcal{I}' = \mathcal{O}' = 1$, one obtains $\mathcal{T}^{\mathcal{B}} = \mathcal{B}(Id \times \mathcal{O})^I$, a shape for *functional components*.

A possible classification follows :

- *Functional components*
$$\mathcal{T}^{\mathcal{B}}_{\mathcal{I},\mathcal{O}} = \mathcal{B}(Id \times \mathcal{O})^I$$
- *Silent components*
$$\mathcal{T}^{\mathcal{B}}_{\mathcal{I},\mathcal{O}} = \mathcal{O}^I \times \mathcal{B}$$
- *Action components*
$$\mathcal{T}^{\mathcal{B}}_{\mathcal{O}} = \mathcal{B}(Id \times \mathcal{O})$$
- *Object components*
$$\mathcal{T}^{\mathcal{B}}_{\mathcal{I},\mathcal{O}} = \mathcal{O} \times \mathcal{B}^I$$

By [3], $\mathcal{T}$ is parametrized by a *strong monad*, $\mathcal{B}$, intended to capture a particular behaviour model associated to the temporal evolution of components. Such behaviour may be purely deterministic, in which case $\mathcal{B}$ is instantiated with the identity monad Id or rather more complex. By an appropriate choice for $\mathcal{B}$, different behavioural features might be considered. For example :

- `Partiality` - the possibility of deadlock, captured by the usual maybe monad $\mathcal{B} = Id + 1$
- `Nondeterminism` - itroduced by the finite *powerset monad*, $\mathcal{B} = \mathcal{P}(Id)$
- `Monoidal stamping` - parameter M should support a monoidal structure to be used in the definition of η and μ, $\mathcal{B} = Id + M$
- `Metric nondeterminism` - capturing situations in which, among the possible future evolution of component, some are more probable tha others. Isomorphism $\mathcal{P}(X) \cong \mathcal{X} \rightharpoonup 1$ suggests the extension of the powerset to a mapping expressing a richer notion of nondeterminism in which each possible state is assigned a confidence level, or probability, leading to $\mathcal{X} \rightharpoonup M$, for M a monoid. Its refinement to $\mathcal{P}(\mathcal{X} \times M)$ constitutes a *monad*.

V. Components As CoAlgebra

Software components by [4] are represented as dynamic systems. This system contains public interface and a private, encapsulated state. For each value of the state space, a corresponding behaviour, arises by computing its anamorphic image.

A typical example of software componet is LBuff: a component modelling a buffered channel which eventually looses messages. It is specife by pair of operations `put` and `pick` and we can illustrate it by *black box* diagram:

$$\begin{cases} \texttt{put} & : \mathbf{M} \to \mathbf{1} \\ \texttt{pick} & : \mathbf{1} \to \mathbf{M} \end{cases}$$

The `put` and `pick` operations are regarded as *buttons*, whose signatures are grouped together in the diagram :

- **M** stands for a message parameter type
- **1** for the nullary datatype
- **+** for datatype sum

The services are provided by component in terms of function $\alpha : \mathcal{U} \times \mathcal{I} \to \mathcal{P}(\mathcal{U} \times \mathcal{O})$, where $\mathcal{U}$ denotes the internal space state, $\mathcal{P}$ is a powerset monad and this function descibes how

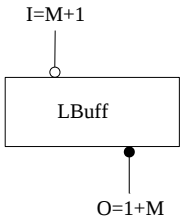


Fig. 1. Software component LBuff

component reacts to input stimuli, produces data and changes state.

Curried form of this notation can by written as $\alpha : \mathcal{U} \to \mathcal{P}(\mathcal{U} \times \mathcal{O})^{\mathcal{I}}$ that is involving transition functor :

$$\mathcal{T} = \mathcal{P}(Id \times \mathcal{O})^{\mathcal{I}}$$

If $\mathcal{P}$ is a finite powerset monad there is the posibility of non deterministic evolution. In fact, LBuff can be defined over $\mathcal{U} = M^*$, with `nil` as the initial state, and dynamics given by

$$\alpha_{LBuff}\langle u, \texttt{put } m\rangle = \{\langle u, \iota_1 *\rangle.\langle m : u, \iota_1 *\rangle\}$$
$$\alpha_{LBuff}\langle u, \texttt{pick}\rangle = \{\langle \texttt{tail } u, \iota_2(\texttt{head } u)\rangle\}$$

where `put` m and `pick` abbreviates $\iota_1\ m$ and $\iota_2\ *$, respective. In the general case, a *component* $p : \mathcal{I} \to \mathcal{O}$ is specified as a coalgebra in Set

$$\langle u_p \in \mathcal{U}_p, \overline{\alpha}_p : \mathcal{U}_p \to \mathcal{B}(\mathcal{U}_p \times \mathcal{O})^{\mathcal{I}}\rangle$$

Another example of a simple component is the specification of the reactive system underlying Ccs expression, for example $\mathcal{R} = \alpha.\beta.\mathcal{R} + \beta.0 + \beta.\mathcal{R}$. In this case the powerset monad is the appropriate choice and the attribute part may be considered trivial and, therefore, modeled by 1. Let Exp denote the set of Ccs expressions and $Act = \{\alpha, \beta, \ldots\}$ the set of actions. Our specification is

$$RSys = \langle R \in Exp, \langle o_{RSys}, \overline{\alpha_{RSys}}\rangle : Exp \to 1 \times \mathcal{P}(Exp)^{Act}\rangle$$

where $o_{RSys} = !_{Exp}$ and α_{RSys} is given by the following clauses,

$$\alpha_{RSys}(\mathcal{R}, \alpha) = \{\beta, \mathcal{R}\}$$
$$\alpha_{RSys}(\mathcal{R}, \beta) = \{0, \mathcal{R}\}$$
$$\alpha_{RSys}(\beta, \mathcal{R}, \alpha) = \{\mathcal{R}\}$$
$$\alpha_{RSys}(e, a) = \emptyset \text{ for all other } e \in Exp \text{ and } a \in Act$$

Components *RSys* can be represented in a diagrammatic form, making Its input and output interfaces explicit.

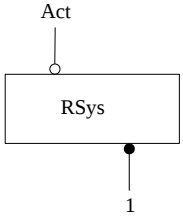


Fig. 2. Software component RSys

As a simple component we can use `Stack` too. `Stack` is a component with two observers (`top` and `isempty?`) and two actions (`push` and `pop`). We use the monad in the type of the corresponding coalgebra to force deadlock whenever an illegal `pop` is performed. Let *E* be a set and consider the state space is modeled by sequences of *E*. Then define

$$\texttt{Stack} = \langle <> \in E^*, \langle o_{St}, \overline{\alpha}_{St}\rangle : E^* \to ((E+1) \times 2) \times (E^* + 1)^{E+1}\rangle$$

where the operations have the expected definitions :

$$o_{St} = \langle \texttt{top}, \texttt{isempty?}\rangle$$

where

$\texttt{top} = \mathcal{B}s.\ |$ if s =<> then $\iota_2 *$ else $\iota_1\ (head\ s)$
$\texttt{isempty?} = \mathcal{B}s.\ |$ s =<>

$$\alpha_{St} = \overline{[\texttt{push}, \texttt{pop}] \cdot dl}$$

where

$\texttt{push} = \mathcal{B}(s, e).\ |\ \iota_1(< e > \frown s)$
$\texttt{pop} = \mathcal{B}(s, *).\ |$ if s =<> then $\iota_2 *$ else $\iota_1\ (tail\ s)$

Action `pop` has a dummy parameter (of type 1) which is made explicit and in the component interface represents the trigger for this action. We can illustrate component `Stack` by the diagram :

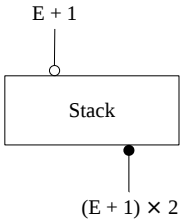


Fig. 3. Software component Stack

A. Components As Objects

Formulation *Components As Objects* is by [1] an alternative model for components in which there is no dependence between input stimuli and the ouputs produced. The relevant functor is

$$\mathcal{T}^{\mathcal{B}} = \mathcal{O} \times \mathcal{B}^{\mathcal{I}}$$

where $\mathcal{I}$ and $\mathcal{O}$ are interface types and $\mathcal{B}$ is a strong monad. A colagebra for this functor is given by the split od two functions

$$\langle o_p, \overline{\alpha}_p\rangle : \mathcal{U}_p \to \mathcal{O} \times (\mathcal{B}\mathcal{U}_p)^{\mathcal{I}}$$

where $o_p : \mathcal{U}_p \to \mathcal{O}$ is state observer (usually called the *attribute* in object-oriented programming paradigma) and $\alpha_p : \mathcal{U}_p \times \mathcal{I} \to \mathcal{B}\mathcal{U}_p$ is state update function (usually referred to as *method* or *action*). Note that often $\mathcal{O}$ instantiates to a Cartesian product $\prod_{x \in \mathcal{X}} \mathcal{O}_x$ of different, but simultaneously available, observers, whereas $\mathcal{I}$ takes the form of a sum $\sum_{y \in \mathcal{Y}} \mathcal{I}_Y$. of (state update, non interfering) operations. If $\mathcal{Y}$ is regarded as a set of action names, $\mathcal{I}_y$ denotes the type of the argument of operation y

The comorphism condition for functor is derived from the general case. $h : p \to q$ is comorphism from p to q if :

$$
\begin{aligned}
&(\mathcal{O} \times \mathcal{B}h^{\mathcal{I}}) \cdot p = q \cdot h \\
&\equiv \{ \text{ definition } \} \\
&o_p = o_q \cdot h \wedge \mathcal{B}h^{\mathcal{I}} \cdot \overline{\alpha}_p = \overline{\alpha}_q \cdot h \\
&\equiv \{ \text{ exponential fusion and absorption } \} \\
&o_p = o_q \cdot h \wedge \overline{\mathcal{B}h \cdot \alpha_p} = \overline{\alpha_q \cdot (h \times id)} \\
&\equiv \{ \text{ exponential universal } \} \\
&o_p = o_q \cdot h \wedge \mathcal{B}h \cdot \alpha_p = \alpha_q \cdot (h \times id)
\end{aligned}
$$

Components modeled in this way will be called, *object components*, after their similarity with notion of an object in *object-oriented programming*.

Behaviours of *object components*, which are animated in CHARITY by [5] are declared as inhabitants of the coinductive type.

$$data\ S \to obj(I, O) = ob : U \to O \mid ac : U \to I \Rightarrow BU$$

Simply example of *object component* is a `queue`. A `queue` similarly as `Stack` has two observers (`top` and `isempty?`) and two actions (`enq` and `deq`). Let E be a set and take sequences of E as the state space. We can define

$$Queue : E + 1 \to (E + 1) \times 2$$

as

$$\texttt{Queue} = \langle <> \in E^*, \langle o_{Qu}, \overline{\alpha}_{Qu} \rangle : E^* \to ((E + 1) \times 2) \times (E^* + 1)^{E+1} \rangle$$

and the operation in the usual way

$$o_{Qu} = \langle \texttt{top}, \texttt{isempty?} \rangle$$

where

$\texttt{top s} = (\ \texttt{s} = <> \to \iota_2\ *,\ \iota_1\ (last\ s))$
$\texttt{isempty? s} = \texttt{s} = <>$

$$\alpha_{Qu} = [\texttt{enq}, \texttt{deq}] \cdot dl$$

where

$\texttt{enq}(\texttt{s, e}) = \iota_1(< e > \frown s)$
$\texttt{deq}(\texttt{s, *}) = (\texttt{s} = <> \to \iota_2\ *,\ \iota_1\ (blast\ s))$

where, given a sequence s, $blast\ s$ return s without its last element.

VI. Conclusion

In this contribution we have defining software components modelled as concrete colagebras for Set endofunctors what is the first step to define behaviour of complex program systems. Software componets is system which encapsulates services and contains public interface and internal state. There is a limited access to services and their internal space because of we illustrated it like black box.

By instantiating interface parameters with particular sets, we got more specialized notions of component. We know two basic shapes of components, *Functional* and *Object - oriented*. *Functional* and *object components*, correspond to a monadic generalization of what is known as, respectively, Mealy and Moore machines in automata literature.

This contribution is focused into *object componets*. *Object component* is an alternative model for components in which there is no dependence between input stimuli and the ouputs produced.

Presented access served to experience obtaining about behaviour of complex program system and contribute to modification of the system in order to provide expected behaviour what is the main idea of my future work.

Acknowledgement

This work was supported by VEGA Grant No.1/0175/08: Behavioral categorical models for complex program systems.

References

[1] L. S. Barbosa, *Components As Coalgebra*. Minho, Portugal: PhD. Thesis, Universidade do Minho, 2001.

[2] J. Adamek, "Introduction To Coalgebra," *Theory and Application of Categories*, vol. 14, pp. 157–199, 2005.

[3] L. M. S. Barbosa, *Components As Processes:An Exercise In Coalgebraic Modeling*, Computer Science Department, University of Minho, Portugal, 2000.

[4] S. Meng and L. S. Barbosa, "On Refinement of Generic Software Components," *UNU/IIST Report*, vol. 281, pp. 2–3, May 2003.

[5] R. Cockett and T. Fukushima, "About Charity," *Yellow Series Report*, no. 92/480/18, June 1992.

DCT coefficients flipping as a method of image content protection

[1]*Tomáš KANÓCZ, Radovan RIDZOŇ, Peter GOČ-MATIS*

[1]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

tomas.kanocz@tuke.sk, radovan.ridzon@tuke.sk, peter.goc-matis@tuke.sk, dusan.levicky@tuke.sk

Abstract— The paper deals with information hiding in general and special focus is given to information content hiding in still images. A solution of information hiding of protected part of an image within the image itself is also presented. Three approaches are proposed – digital watermarking, DCT coefficients flipping and cryptography. The main focus of this paper is oriented to information hiding based on DCT coefficients flipping. In this case the information content protection is based on pixelisation and mosaic approach. Some experimental results are shown and advantages and drawbacks are also discussed.

***Keywords*— Information hiding, Discrete Cosine Transform, coefficients flipping, pixelisation, mosaic, digital watermarking, cryptography.**

I. INTRODUCTION

In the age of information it is necessary to have the right information at the right time. Information drives the world. In recent years, information security and security in general was not emphasized. They have become very important after attackers and terrorists started to use information and communication technologies for wrecking activities. But security in general is not only about information security. Security involves also security of people, objects and events.

Security is becoming a weak point of energy and communications infrastructures, bus and train stations, airports and crowded sites in general. Practically any crowded place is vulnerable, and the risks should be controlled and minimized as much as possible. Controlling access and being able to respond rapidly to potential dangers are facilities which every security system for such environments should have. One of the best ways to tackle such problems is through active observation. This explains the real need for intelligent surveillance and information systems in urban (and wider) areas. All of the mentioned techniques should be applied, whilst taking into consideration privacy protection.

II. THE INDECT PROJECT

The Integration Project INDECT (Intelligent Information System Supporting Observation, Searching and Detection for Security of Citizens in Urban Environment) contributes to the general vision of Topic SEC-2007-1.2-01 "Intelligent urban environment observation system" within FP7 concerning the future generation technologies which will neutralize threats in urban environments [1].

The **main objectives** of the INDECT project are:

- to develop a platform for: the registration and exchange of operational data, acquisition of multimedia content, intelligent processing of all information and automatic detection of threats and recognition of abnormal behavior or violence,
- to develop the prototype of an integrated, network-centric system supporting the operational activities of police officers, providing techniques and tools for observation of various mobile objects,
- to develop a new type of search engine combining direct search of images and video based on watermarked contents, and the storage of metadata in the form of digital watermarks.

One of the expected results is to develop a system, how to hide information in multimedia data, mainly in still images.

III. INFORMATION HIDING

Information technologies are used for the processing of valuable information more and more nowadays. Information processing, from information technologies point of view, consist of information storage, transmission, evaluation and interpretation of data. However, these data represent valuable information and they have to be protected in the following ways:

- only authorized persons should have access to these data,
- only authentic data should have been processed,
- there should be a method how to find out who have made, changed or removed these data,
- data should be confidential,
- data should not be denied, when they are needed.

Multimedia content protection approaches can be divided into two main groups [6][7]:

- multimedia content protection during transmission,
- multimedia content protection after transmission.

Multimedia content protection during transmission can be solved by using cryptographic methods, which realize multimedia content encryption. In this approach the multimedia content becomes unreadable after encryption by the sender. To access the multimedia content in the receiver,

decryption with proper key is necessary. Cryptographic methods can be divided into two main groups: symmetric cryptography methods and asymmetric cryptography methods.

The problem of multimedia content protection after decryption on the receiver side can be solved by using digital watermarking methods, which perform embedding of imperceptible information into the multimedia content and this information should not be easily removable by basic multimedia processing techniques.

In the INDECT project a solution of another problem is necessary. During crime investigation some parts of images have to be hidden from unauthorized people and other persons have to have access to the whole information content of the image. For example a car license number on a photo has to be hidden as confidential information. The same case is with a photo of a person. In some cases the face has to be hidden. Another request is to make the basic features of the protected part of the image recognizable. It should be clear that the protected part of the image is hiding a face, a car license number, a boat or other image content. On the other hand it should be impossible to accurately identify the true identity of a person or a car license number.

Information hiding can be achieved by decreasing the space resolution of the protected part of the image. This process is often referred to as **pixelisation**. In this case parts of the image which have to be hidden are undersampled and the rest of the image is preserved in full quality (Fig. 1). Two images have to exist, one censored and one uncensored.

Fig. 1 **Pixelisation example**

One solution of these problems is to keep two image databases, where one database contains images with hidden information and another one contains the uncensored images. But this solution has a big disadvantage because of two image databases.

Another solution is to hide information about the undersampled part of the image within the same image. In our approaches and experiments we exploit the features of the Discrete Cosine Transform (DCT).

IV. Discrete Cosine Transformation

DCT has been chosen for undersampling because of nearly ideal decorrelation. By zeroizing certain coefficients undersampling can be achieved. The formula used for direct 2D DCT calculation for an input luminosity matrix A and output coefficients matrix B is [1]:

$$B_{pq} = \alpha_p \alpha_q \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A_{mn} \cos \frac{\pi (2m+1) p}{2M} \cos \frac{\pi (2n+1) q}{2N} \quad (1)$$

for $0 \le p \le M-1$, $0 \le q \le N-1$,

where $\alpha_p = \begin{cases} 1/\sqrt{M}, p = 0 \\ \sqrt{2/M}, 1 \le p \le M-1 \end{cases}$ $\alpha_q = \begin{cases} 1/\sqrt{N}, q = 0 \\ \sqrt{2/N}, 1 \le q \le N-1 \end{cases}$,

and where M and N are the row and column size of A, respectively.

Information about average luminosity is obtained in the first transformation coefficient (called DC coefficient) and information about details are obtained in the rest of the coefficients (AC coefficients) (Fig. 2).

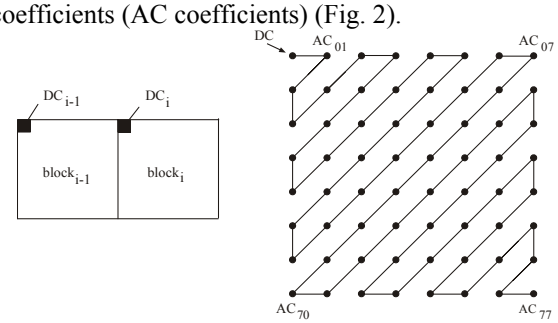


Fig. 2 DC and AC coefficient of the DCT transformation

V. Proposed Approaches

We have suggested three approaches how to resolve the problem of information hiding within the same image. The proposed approaches are:

- digital watermarking,
- DCT coefficients flipping,
- cryptography.

A. Digital Watermarking

Digital watermarks are primarily used for ownership and author rights protection and copy prohibition. Digital watermarking represents embedding of information within the content of the image. Embedded information should be undetectable by human visual system but has to be detectable by a detector, which is used in the watermark extraction or detection process [3][5].

In this case, AC coefficients values represent the embedded information of the undersampled part of the image. After DCT transformation DC coefficient remains at the same place and information about AC coefficients is spread within the image by using a digital watermarking technique. Unauthorized persons have access only to censored image and authorized persons can obtain image in full quality after the image reconstruction.

Small degradation of the whole image by digital watermark embedding is a drawback of this approach. The solution is to use Human Visibility System techniques in the process of watermark embedding [2].

B. DCT Coefficients Flipping

The second approach how to protect a chosen part of an image is to use DCT flipping. This method makes use of the properties of DCT. The DC coefficient carries the most information about a block. The other coefficients carry information about firmer details. If we keep the DC coefficient untouched, but shuffle the AC coefficients we get a distorted block, which is very similar to a pixelised block. A pixelised block has all of its AC coefficients zeroized. Only the DC coefficient is present. A permutation vector can be used as a key to unlock, or lock a protected image part. The permutation vector determines how the AC coefficients are shuffled.

No information loss is the biggest advantage of DCT coefficients flipping. The overall information contained within a block is not changed, just the order of its notation due to shuffling. This means that only the permutation vector is needed for image reconstruction and no additional information.

C. Cryptography

The last approach is to use cryptographic algorithms to hide information about AC coefficients. After the DCT transformation AC coefficients are encrypted by using symmetric or asymmetric cryptography algorithm. This approach is very similar to the previous approach with DCT coefficients flipping. In this case AC coefficients are not shuffled, as it was in the previous case, but encrypted.

VI. EXPERIMENTAL RESULTS

The DCT flipping method was implemented using a block size of 16x16. If a high resolution image is used, than a block size smaller than 16x16 makes the distorted face recognizable. As this method is intended to protect the identity of people, or other personal data, this is not an option.

Larger block sizes like 32x32 are better from the content protection point of view, but make it harder, or even impossible, to roughly guess what is the hidden part of the image, whether it is a part of a face or car license number.

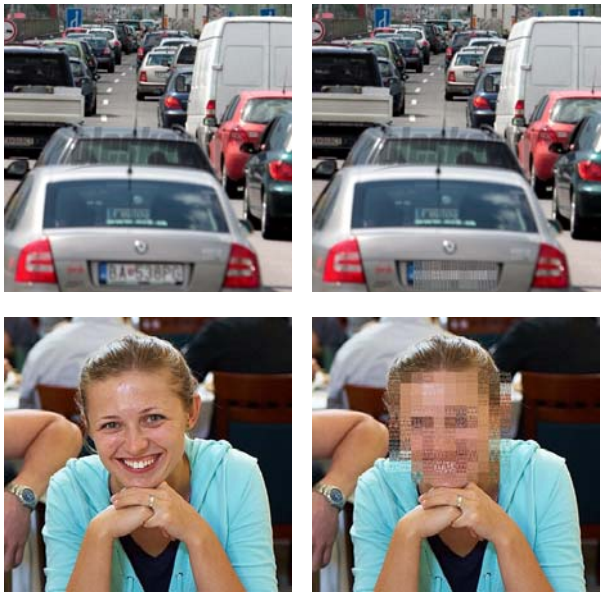


Fig. 3 Mosaic examples

Even though this method is theoretically lossless, during experiments some drawbacks where found. If we shuffle a block's AC coefficients, the output matrix after the inverse discrete cosine transform will also change. Before flipping the input luminance block consisted of whole positive numbers in the range from 0 to 255. These numbers represent the luminosity of a pixel, or the intensity of red, green or blue in an RGB image. After flipping the coefficients in the frequency domain and computing an inverse DCT on the block, we get real numbers, which in some cases can be even negative. If the block after inverse DCT contains non-whole numbers, just a small amount of distortion is inserted to the luminance block.

If the luminance block after inverse DCT contains negative numbers, this usually caused visible artifacts in the image. The reason of this is that if for example the output luminance block contains a value of 12.75, this number is rounded to 13, or 12. The rounding method varies from implementation to implementation. If we consider the worst case, the biggest error we get is the luminosity value of 1. Such a small variation in luminosity is undetectable for the human eye. On the other hand if get a value of -126.5 in a luminance block, this number will be rounded to 0. It means that if the protected image will be saved in a graphical format, the luminance value of -126.5 will be replaced with the value 0 and thus lost. If we load the saved picture and compute the DCT coefficients and than use the permutation vector to place the shuffled AC coefficients to their original position, a detectable image distortion will be introduced (Fig. 4). Saving the protected image as binary data can eliminate this error. On the other hand the introduced error is not too big, so it does not distort the image significantly. Other methods to eliminate this error are being researched.

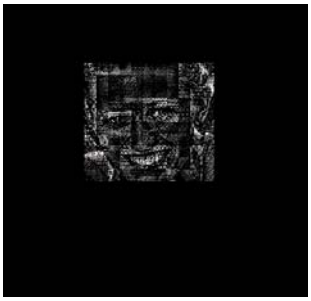


Fig. 4 Image difference (adjusted by gamma correction for better illustration)

DCT flipping can be implemented successfully to colored images, too. We used the YCbCr color model instead of plain RGB. Using YCbCr is computationally less demanding. To effectively distort a face, or a car license number, it is enough to modify the DCT spectrum of luminance Y. During our experiments several combination of YCbCr modifications were carried out. This color model allows different block sizes for the luminance and chrominance components of the image. Increasing the block size of the chrominance components to two times to that of the luminance components can introduce further errors to the image due to already mentioned rounding problems, but can make a face, car, or other object harder to identify. If for example there is a demand to hide the color of the distorted car it is even possible to swap the Cb and Cr components prior to shuffling (Fig. 5). Of course the inverse process must swap them back again to display the colors correctly after reconstruction. One of the disadvantages of this method is that it usually causes images padding, which will increase the size of the stored image slightly.

Fig. 5 Example of Cb and Cr components swapping

VII. Conclusion

Three solutions how to hide content of the image part were presented in this paper. One of them, based on DCT coefficients flipping, was closely introduced. Different experimental results in different color models were shown and advantages and drawbacks were discussed.

As was shown, big advantage of the DCT coefficients flipping method is no image degradation in the uncensored parts of the image in comparison with approaches based on digital watermarking. The main drawback of this method is limited number of possible permutations and it can be easy for an attacker to guess the proper permutation used for a DCT coefficients flipping. Another drawback of this method is the small distortion of the reconstructed part of the image. The elimination of this drawback is in the phase of research.

Acknowledgements

The work presented in this paper was supported by Ministry of Education of Slovak republic VEGA Grant No. 1/0065/10, INDECT Grant (7th Research Frame Programme no. 218086) and Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] Discrete Cosine Transformation, Matlab 2009b, Image Toolbox Tutorial.

[2] FORIŠ, P., LEVICKÝ, D.: Implementations of HVS models in digital image watermarking, Radioengineering. - ISSN 1210-2512. - Vol. 16, no. 1 (2007), p. 45-50.

[3] FURT, B., MUHAREMAGIC, E., SOCED, D.: Multimedia encryption and watermarking. *Springer*, 2005.

[4] INDECT Project, 7th Frame program, http://www.indect-project.eu.

[5] RIDZOŇ, R., LEVICKÝ, D.: DRM based on the robust digital watermarking, Science & Military. - ISSN 1336-8885. - Roč. 2, č. 2 (2007), s. 46-50.

[6] RIDZOŇ, R., LEVICKÝ, D.: Multimedia security and multimedia content protection, ELMAR-2009 : proceedings : 51st International Symposium. - Zadar : Croatian Society Electronics in Marine, 2009. - ISBN 978-953-7044-10-7. - P. 105-108.

[7] SENCAR, H.T., RAMKUMAR, M., AKANSU, A.N.: Data Hiding Fundamentals and Applications. Content Security in Digital Multimedia. *Elsevier*, 2004.

State of Art in Video Quality Measurement Standards

Martin KAPA

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

martin.kapa@tuke.sk

***Abstract*—With the increase of multimedia usage as a source of information and for a communication purposes, the measurement of video quality and corresponding user experience has become a crucial term. Today there are various companies and research groups that orient their research in this area and this paper provides overview of what has been done and what are the future trends.**

***Keywords*—Video quality evaluation, subjective quality measurements, objective quality measurements, VQEG, ITU-T, standardization.**

I. Introduction

Speech quality measurement has a very long history in the field of signal processing and telecommunication. Recently we have been able to see that quality assessment techniques were extended to audio and video as well. With the importance of user's mobility the corresponding multimedia services need to be highly reliable to earn the trust of its customers and industry has become more desperate for accurate and reliable objective video metrics. Quality measurement has a wide range of objectives, including codec evaluation, transmission planning, end-to-end quality assurance and client-oriented quality measurement.

Today we can find different approaches and standards that are trying to address numbers of issues related to video quality measurement, which include requirements, test plans, practices and many more. Within this paper I will focus on work that has been done in the field of video quality evaluation and its future trends. Therefore, it is important firstly to distinguish Quality of Service (QoS) and Quality of Experience (QoE). QoS is quite well understood and can be described as quality of service from the network performance and data transmission point of view. On the other hand QoE is still in the active state of research thus less defined. However, it can be described as perceived quality of the user.

In recent research we see that current approaches are only oriented to one specific video content type/application or to one scenario, which is not enough. Video quality metrics need to be more complex and cross-content to provide better correlation with subjective ratings that are really important for appropriate decisions on a suitable optimization method for video streaming.

The paper is structured as follows. We describe important terminology in Section II. In Section III we outline common standards for subjective quality assessment; Section IV introduces existing standards on objective quality assessment and Section V is orienting on current activities and future trends.

II. Foundation

In the following chapter, I outline the main terms needed to understand the main problematic of this paper.

User Experience is primarily understood as a personal occurrence that user achieve during the process of interaction with a product or service. However, we have to have in mind that user experience is highly influenced by different factors that are nearly unable to be measured like user expectations, experiences or state of mind. *Mean Opinion Score (MOS)* is literally a number defining QoE in range between 1 and 5 and it is used to express level of quality in multimedia (audio, VoIP or video)[Tab. I]. MOS for voice is standardized in ITU-T R. P.800[1] and recently it is frequently use for video quality evaluation as well. The value of MOS is commonly acquired by subjective assessment tests, where the attendants rate the audio or video quality of the test sequence.

TABLE I
Level of quality in MOS

MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible, but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very annoying

Video quality metrics can be classified into full-reference, no-reference and reduced-reference metrics based on the amount of reference information they require during the video quality tests:

- *Full-reference (FR) metric*: is a technique where we fully use reference video to compare its quality to test video. The whole process is done in two steps, the first step calculates the errors between original and distorted images and the second one has to pool the particular errors to a global quality assessment[2].
- *No-reference (NR) metric*: here we analyze only test video without the need to compare it with the reference video. However, this technique uses some prior information, like type of encoding to be able to look on codec-specifications.
- *Reduced-reference (RR) metric*: is a hybrid between FR and NR metric in terms of the reference information. This approach is suitable mainly cause of managing of the amount of reference information we use.

Another important factors that we need to fully consider while measuring of QoE are the values of video parameters like bit rate, video resolution, frame rate and codec; or network

parameters like bandwidth, delay, jitter and packet loss. *Bit rate* represents the number of bits processed during one time unit. *Video resolution* is a size of video image and is measured in pixels, where the numbers represent horizontal and vertical resolution. *Frame rate* specify a frequency at which the streaming device produce images that are called frames. *Video codec* is a software used for encoding and decoding of a digital data stream. *Bandwidth* is defined as the amount of data per one time unit that are delivered over physical network topology, from the source to the destination. *Delay* represent lapse of time while some action is awaited. *Jitter* is best represented as an end to end delay between one packet to the next, within the same stream. *Packet loss* defines percentage of lost packets.

III. Subjective Quality Assessment

Perceived quality is not a term that is easy to define or be exactly computed. When we consider video quality, each observer uses its own internal scale to constructing his judgment. This scale is most of the time not influenced just by the quality of video, but also its content and psychological state of particular observer. However, it still shows more accuracy in results of video quality measurements than objective metrics. It is caused by a fact that subjective measurements are more suitable for measuring of HVS (human visual systems) reaction on streamed video content. Standards for subjective assessment of video quality or speech and audio have been defined many years ago, visual quality assessment has been formalized in ITU-R Rec. BT.500-11[3] and ITU-T Rec. P.910[4]. These documents give recommendations about viewing distance, room illumination, normalized protocols, test duration, recruitment of observers, methods to detect incoherent observers (used to reject their votes) and methods to compute the precision of the mean opinion score.
For the subjective quality assessment there are several most common procedures that are used as *Subjective Assessment Methodology for Video Quality (SAMVIQ)* where subjects are able to replay videos to precise the judgments. It gives more proper results but requires a longer time to perform the assessments; *Absolute Category Rating (ACR)* is a single-stimulus method, which means that the subjects are rating each test video individually without comparison to a reference video; *Single Stimulus Continuous Quality Evaluation (SSCQE)* it is a type of assessment where subjects are dynamically rating video of duration typically 20-25 minutes using a slider with an associated quality scale; *Double Stimulus Continuous Quality Scale (DSCQS)* subjects can see pairs of video sequences (reference and impaired sequence) in a randomized order and *Degradation Category Rating (DCR)* here the subjects are rating test videos in comparing to known reference video. Mentioned testing procedures are used in different applications and variety of different rating scales. The results of subjective assessments are ratings of video quality by viewers, which are then transformed into MOS.

Subjective assessment is a proper approach for measuring of video quality, but it requires lot of experiments for appropriate results, which requires a longer time to perfom the assessments.

IV. Objective Quality Assessment

Objective video quality assessment techniques can be defined as mathematical models that are used to estimate results of subjective quality assessment. These techniques are based on metrics that can be automatically evaluated by a computer application. In compare to subjective assessment these metrics requires less experiments but usually they can not measure changes in video quality only seeing by HVS, which results in inaccurate measurements.

A. ITU-T Study Groups

ITU-T Study Group 9[1] is orienting its studies on issues of the usage of telecommunication systems for broadcasting of television and sound programs. The most recent work contains several ITU-T Recommendations on Voice, data and video IP applications over CATV networks (IPCablecom). In the field of video quality measurement they are trying to evaluate the amount of influence of interacting factors, such as source coding, compression, bit rate, delay, bandwidth, synchronization between the media and many others on perceptual audiovisual quality in multimedia services. The result of their work is to come with more accurate objective quality measuring technique that correlate the user opinion on the perceived audiovisual quality with the desirable preciseness.

ITU-T Study Group 12[2] is a leading group on performance, QoS and QoE. Recently this group is orienting on the development of software tools that will allow the modeling of potential network configuration and the prediction of the user impact of associated impairments. They have already developed a model for voice quality prediction, while the work on video quality prediction is still in progress. In the filed of QoS and QoE they provided standards like G.1000[5], G.1010[6], E.800[7], P.862[8] and Y.1541[9].

B. Video Quality Experts Group (VQEG)

VQEG[3] is a research group that contains professionals with various backgrounds in the filed of video quality assessment including members from international recognized companies. The group was founded in 1997 and the majority of the members are active experts in the International Telecommunication Union (ITU). Their work contains various projects like FRTV Phase I and Phase II, RRNR-TV, Multimedia Phase I and Phase II, HDTV and Hybrid Perceptual/Bitstream. *FRTV Phase I* was completed in year 2000; the objective of it was out-of-service testing. The test contains mainly video sequences with different profiles encoded in MPEG2. Results of subjective assessment were represented by the usage of DSQS. Unfortunately the results were indecisive, cause of statistical equivalent of most models. After the completion of first phase of testing, the second one followed-up (*FRTV Phase II*). The Phase II was targeting on full-reference metrics for SD TV application. This project was completed in year 2003. For better test results this project focused mainly on secondary distribution of the video sequence. Within the second phase there were produced 128 test sequences. The results of subjective measurements were collected using DSCQS(same as during the Phase I). Standards like ITU-T Rec. J.144[10] and ITU-R Rec. BT.1683[11] were based on the results of Phase II.

[1] See http://www.itu.int/ITU-T/studygroups/com09/
[2] See http://www.itu.int/ITU-T/studygroups/com12/
[3] See http://www.its.bldrdoc.gov/vqeg/.

The main target of the *HDTV* project was to analyze the performance of suitable models for digital video quality measurement in HDTV applications, secondary goal was developing of subjective datasets that in the future may be used as a foundation for the improvement of HDTV objective models. The test video sequences were encoded in MPEG2 and H.264 with transmission errors (packet loss, packet delay variation, jitter, overflow and underflow, bit errors, and over the air errors), display formats in the tests are 1080i at 50 and 60 Hz; and 1080p at 25 and 30fps.
Multimedia Phase I tried to evaluate metrics in multimedia scenario, with target in lower bitrates and smaller frame sizes (QCIF, CIF and VGA). The Phase I tests were done for a validation of full-reference, reduced-reference and no-reference models of objective quality measurements. The experiments contained 346 source and 5320 processed video sequences with a wide range of quality and transmission errors. These video clips were eveluated by 984 viewers. Standards like ITU-T Rec. J.247[12] and ITU-T Rec. J.246[13] were based on the results of this particular project. *Multimedia Phase II* is still in its early stages, not much happening yet, the results are expected to be released in September 2010.
RRNR-TV is one of the latest projects done by VQEG. The tests were done for SD Tv and each test contained 12 source clips and 34 test conditions with total of 156 test sequences. The sequences were encoded in MPEG2 and H.264. The result of this project is seven reduced-reference models, some of which may become future ITU recommendations.

C. ATIS IPTV Interoperability Forum (IIF)

ATIS IIF[4] is globally recognized as the leading developer of requirements and standards in the field of IPTV. One of the parts of ATIS IIF is QoS Metrics (QoSM) committee that targets on issues around QoS and QoE for IPTV. The most recent work of the QoSM committee is an implemention quide for QoS Metrics and an document of requirements for QoE in IPTV.

V. Future Trends

The latest statistics showed high increase of connected users, network elements and heterogeneity of physical connection (optical fiber, twisted pair and wireless). It is expected that, by 2011, about 3 billion hosts (devices that use the communications infrastructure including mobile and other type of handheld devices) will be connected to the internet from the 570 million of hosts in July 2008[14].In the last couple years, we are able to see dramatic surge in the field of different mobile devices for the users. Computers are getting smaller and mobiles are getting smarter, it would be necessary to turn our orientation in QoE to specific requirements of particular devices (size of the screen, hardware and software constraints, etc.). With this approach we would be able to define application QoE parameters and achieve the required level of QoE by the user, with decrease in usage of network resources. In this paper I compared different approaches for video quality measurement, we can see that if we want more accurate results we have to consider impact of video content and how it is seeing by HVS. For future video quality metrics and models it is essential to combine subjective and objective measurement techniques to provide more convenient results. Another issue that requires our attention is the increase of confusions among users because the variety of different algorithms. This resulted in usage of commercial products rather than free algorithms. In the future there is a need to define more flexible and on-demand validation process, suitable to assure comparability of video quality measurements across various platforms.

VI. Conclussion

The term of video quality measurement or estimation is very popular area among many different research groups and standardization organizations. This paper provides an overview of current status in the particular field. We can see that the research is very active. However, many different standards provide lot of misunderstandings among users and that is one of the main challenges that need to be solved. In this area we can see a gab within the video quality metrics caused by ill-considering of HVS factor and its importance in the accurancy of video quality measurements. This is an area that needs a lot of work in the future research.

References

[1] ITU-T P.800: Methods for Subjective Determination of Transmission Quality,International Telecommunication Union, Geneva, Switzerland, 1996
[2] M. Carnec, P. Le Callet, D. Barba, "Full reference and reduced reference metrics for image quality assessment," Signal Processing and Its Application, 2003.
[3] ITU-R Recommendation BT.500-11: Methodology for the subjective assessment of the quality of television pictures, International Telecommunication Union, Geneva, Switzerland, 2002.
[4] P.910, I.T.R.: Subjective video quality assessment methods for multimedia applications. Recommendations of the ITU, Telecommunications Sector
[5] ITU-T Recommendation G.1000: Communications quality of service: A framework and definitions, International Telecommunication Union, Geneva, Switzerland, 2001.
[6] ITU-T Recommendation G.1010: End-User Multimedia QoS Categories Series G: Transmission Systems and Media, Digital Systems and Networks Quality of Service and Performance, International Telecommunication Union, Geneva, Switzerland, 2001.
[7] ITU-T E.800: Definitions of terms related to quality of service, International Telecommunication Union, Geneva, Switzerland, 2008.
[8] ITU-T Recommendation P.862: Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, International Telecommunication Union, Geneva, Switzerland, 2001.
[9] ITU-T Recommendation Y.1541: Network performance objectives for IP-based services, International Telecommunication Union, Geneva, Switzerland, 2003.
[10] ITU-T Recommendation J.144: Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference, International Telecommunication Union, Geneva, Switzerland, 2004.
[11] ITU-T Recommendation BT.1683: Objective perceptual video quality measurement techniques for standard definition digital broadcast television in the presence of a full reference, International Telecommunication Union, Geneva, Switzerland, 2004.
[12] ITU-T Recommendation J.247: Objective perceptual multimedia video quality measurement in the presence of a full reference, International Telecommunication Union, Geneva, Switzerland, 2008.
[13] ITU-T Recommendation J.246: Perceptual visual quality measurement techniques for multimedia services over digital cable television networks in the presence of a reduced bandwidth reference, International Telecommunication Union, Geneva, Switzerland, 2008.
[14] F. Liberal, J.O. Fajardo, H. Koumaras, QoE and *-awareness in the Future Internet, Towards the Future Internet, p.293-302, IOS Press, 2009.

[4]See http://www.atis.org/IIF/.

Graph Cut Tracking of Ball in the Tube

Peter KARCH, Marián BAKOŠ

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

peter.karch@tuke.sk, marian.bakos@tuke.sk

***Abstract*—This paper deals with implementation of Graph cut segmentation for object tracking in dynamic images. The procedure proposed in this paper seeks to find a global optimal solution of Graph cut segmentation in local parts of input dynamic images.**

***Keywords*— Graph cut, image segmentation, tracking.**

I. INTRODUCTION

Tracking of moving objects is one of the tasks that can be performed using Graph cut segmentation. Several possibilities for tracking objects by using Graph cut methods are outlined in papers [12] and [13]. Graph cut segmentation is used to recognize a moving object where the results of segmentation in the previous steps are used to initialize Graph cut segmentation in the following steps. In this implementation the initial initialization is simplified so the user needs not to mark both the object of segmentation and the background of the tracked object [5], [6]. This implementation can be practically used to track a ball on the soccer field or in the laboratory conditions for tracking the ball in a tube or tracking the ball at an inclined plane.

The details of Graph cut segmentation method and object tracking are given in Section II. The descriptions of Model Ball&Tube are given in Section III A. Segmentation details are given in Section III B. Section III C gives the details of the tracking of the object. Section IV provides an application example of the object tracking using Graph cut segmentation implemented by Matlab for segmentation of a moving object.

II. GRAPH CUT SEGMENTATION

Graph cut segmentation [7] of an object from its background is interpreted as a binary labeling problem. Each pixel in the image is assigned to the binary label $A_k = \{O, B\}$, where O represents the object in the image and B represents the background in the image. These two labels are identified by terminal nodes S (source) and T (sink) and indicate hard constraints of segmentation. The labeling vector $A = \{A_1, \ldots, A_p, \ldots, A_{|P|}\}$ defines the final segmentation. The soft constraints that are imposed on boundary and region properties of L are described by the cost function:

$$E(A) = \lambda \cdot R(A) + B(A) \quad (1)$$

where

$$R(A) = \sum_{p \in P} R_p(A_p) \quad \text{(regional term)} \quad (2)$$

$$B(A) = \sum_{\{p,q\} \in N: A_p \neq A_q} B_{\{p,q\}} \quad \text{(boundary term)} \quad (3)$$

where

$$R_p(O) = -\ln \Pr(I_p \mid "obj") \quad (4)$$
$$R_p(B) = -\ln \Pr(I_p \mid "bkg")$$

and

$$B_{\{p,q\}} \propto \exp\left(-\frac{(I_p - I_q)^2}{2\sigma^2}\right) \cdot \frac{1}{dist(p,q)} \quad (5)$$

A. Object Tracking

The aim of an object tracker [10] is to generate the trajectory of an object over time by locating its position in every frame of the video. In its simplest form, tracking can be defined as the problem of estimating the trajectory of an object in the image plane as it moves around a scene. The tasks of detecting the object and establishing correspondence between the object instances across frames can either be performed separately or jointly. In the first case, possible object regions in every frame are obtained by means of an object detection algorithm, and then the tracker corresponds the objects across frames. In the latter case, the object region and correspondence is jointly estimated by iteratively updating object location and region information obtained from previous frames.

The main tracking categories:

- Point Tracking
- Kernel Tracking
- Silhouette Tracking

III. BALL TRACKING

A. Model Ball&Tube

The tube model [11] consists of a solid supporting part on which there is a glass tube fixed by a rotary mechanism through the shaft. The rotary mechanism allows the tube to tilt in both directions and thereby changes parameters of the system. This mechanism consists of a gearbox and a step motor. The feedback information on the actual tilt is obtained from a potentiometer which shaft is via gearwheel associated with the gearbox mechanism of the tube tilt. The infrared distance sensor is placed at the upper end of the tube, which allows measuring the height of the moving body in the tube. On the sensor there is located a potentiometer to its calibration and signaling LED diode which changes its brightness

depending on the distance of the body. The body is of spherical shape. The fan is located at the lower end of the tube, which affects the height of the ball by the speed of its rotation. The fan is located at the lower end of the tube, which affects the height of the ball by the speed of its rotation. The fan is powered by DC motor.

Model Tube (Fig. 1) consists of these basics parts:

1. Actuating device
   - DC motor
   - step motor
2. Sensing device
   - optical proximity sensor
   - step motor
3. Model
   - tube
   - body (ball)
   - fan
   - transmission

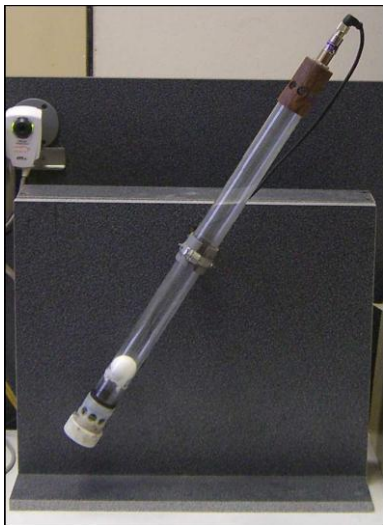


Fig. 1 Model Ball in the tube

Logical model of Tube

The tube is a controlled system. The basic objective of control [8] is to control the height of the ball at the desired value. The Schneider Electric system is used as the control system [9].

The functions are as follows:

- sensing real value of the ball height
- sensing real value of the tube tilt
- control height of the ball
- control tilt of the tube

Physical model of Tube

1. **Hardware layer:** tube, fan, ball, optical proximity sensor, DC motor, transmission, step motor, PLC Modicon TSX Premium with source and communication module for AS-i network, source for AS-i network, distribution module for AS-i network and input/output analog modules for AS-i network.
2. **Software layer:** programming tool for PLC (PL7 PRO).
3. **Communication layer:** The model communicates via input / output analog modules for AS-i networks. Data are transmitted to PLC automaton by AS-i network. The interface between the PLC and PC used RS 232 connection.

B. Segmentation

Graph cut segmentation is used for segmentation of the traced object. This method is initialized by determination of one or more points representing the object and of one or more points representing the background. The initialization in this proposal is modified for simplification so that the user has to label only the object being tracked. In this case it is the ball, which is labeled. This marked point is used as terminal S (source) that denotes the object for segmentation. This terminal is then used as the center of square that defines pixels of terminal T (sink) by identifying the background of segmentation. The user can determine the size of this square. The image segmentation is performed only locally in the labeled square, which accelerates the speed of segmentation.

C. Description of tracking

Segmentations of the object during the tracking are performed by individual frames. Initialization of Graph cut segmentation is performed in the first step on the first frame. After execution of this segmentation the center of the final object segmentation is used as the center of square that identifies pixels of terminal *T* (sink) in the next step.

IV. Experiments

For experimental verification of the proposal "BT.avi" video file was used as a source of the input image.

In the first step the user identifies the object that he wants to track by red label in Fig. 2 e.g. the ball in the tube. The black square is supplemented automatically and indicates the background of segmentation.

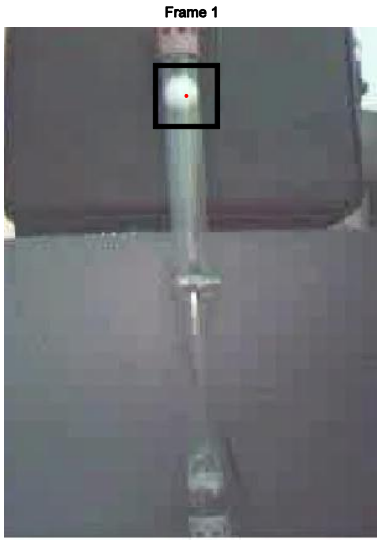


Fig. 2 Initialization

The following pictures show the ball tracking process for some frames. Successfulness of the object tracking depends on the lighting conditions, because in the low light conditions the object in the tube is not seen.

Fig. 3 Tracking - ball segmentation at frame 10

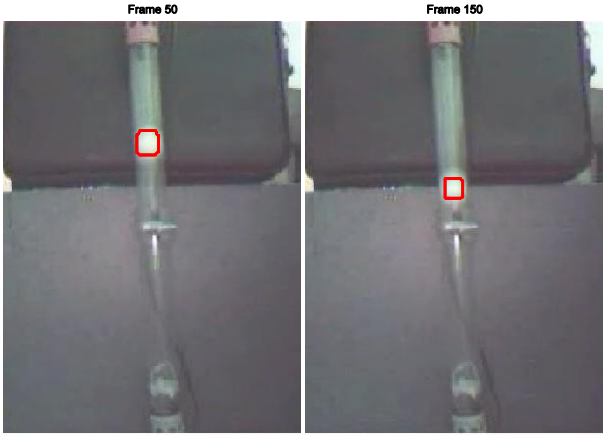


Fig. 4 Tracking - ball segmentation at frame 50 and 150

Fig. 5 Tracking – ball segmentation at frame 250

Matlab Wrapper for Graph cut [4] ([1], [2], [3]) is compiled and implemented with object tracking in Matlab R2007a. All the experiment results are given with CPU Core2Duo 2.13GHz, 3GB RAM and 256MB VGA in Windows XP Pro.

V. Conclusion

This algorithm was experimentally tested on the frames of the input video file. The next step in further experiments will be extension of the method of Graph cut segmentation using shape prior and a change in the source of input images to the IP webcam.

Acknowledgment

This work was supported by grant APVV-0682-07 and by the VEGA project No 1/0386/08. This publication is the result of the project implementation Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] Y. Boykov, V. Kolmogorov, “An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization In Vision”, *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no.9, pp 1124-1137, September 2004.

[2] Y. Boykov, O. Veksler, R. Zabith, “Efficient Approximate Energy Minimization via Graph Cuts”, *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no.12, pp 1222-1239, November 2001.

[3] V. Kolmogorov, R. Zabih, “What Energy Functions can by Minimized via Graph Cuts?”, *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no.2, pp 147-159, February 2004.

[4] S. Bagon, “Matlab Wrapper for Graph Cut”, December 2006. Online: http://www.wisdom.weizmann.ac.il/~bagon/matlab.html

[5] O. Duľová, I. Zolotová, “Using greedy and evolution algorithms for active contour”, *Process Control 2008: proceedings of the 8th international scientific - technical conference*, pp C085a-1-C085a-9, June 2008.

[6] E. Ocelíková, D. Klimešová, “Preference ranking method for multi-criteria decision”, *AEI '2009: International Conference on Applied Electrical Engineering and Informatics*, pp 36-41, September 2009.

[7] P. Karch, I. Zolotová, “An Experimental Comparison of Modern Methods of Segmentation”, *SAMI 2010: 8th IEEE International Symposium on Applied Machine Intelligence and Informatics*, pp 247-252, January 2010.

[8] M. Franeková, “Modelling of Communication Systems via SW Tools Matlab”, EDIS ŽU Žilina, 2003, ISBN 80-8070-027-3.

[9] I. Zolotová, S. Laciňák, E. Ocelíková, “New trends in supervisory monitoring and control”, *AEI '2008 : international Conference on Applied Electrical Engineering and Informatics*, pp 102-105, September 2008.

[10] A. Yilmaz, O. Javed, M Shah, “Object tracking: A survey”, *ACM Computing Surveys*, Vol. 38 , No. 4, 2006.

[11] I. Poruben, "Local and Remote Supervisory Control of Systems - Ball in Tube, Magnet” Diplomová práca. Košice: Technická univerzita v Košiciach, Fakulta elektrotechniky a informatiky, 2008.

[12] N. Xu, N. Ahuja, “Object contour tracking using graph cuts based active contours”, *Image Processing. 2002. Proceedings. 2002 International Conference on* , vol.3, pp III-277 - III-280, 2002.

[13] Z. A. Garrett, H. Saito, “Real-Time Online Video Object Silhouette Extraction Using Graph Cuts on the GPU”, *Image Analysis and Processing – ICIAP 2009*, pp 985-994 , 2009.

System optimalization based on relation ontology model comparison

Ján KAŽIMÍR

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

jan.kazimir@tuke.sk

***Abstract*—In this paper I will discuss the logic of my suggestion of an optimalization schema for combination of web content management system and organization where is that system deployed. This optimalization schema is based on comparison of relation ontological models.**

***Keywords*—Ontological modeling, ontology, optimalization, web content management system.**

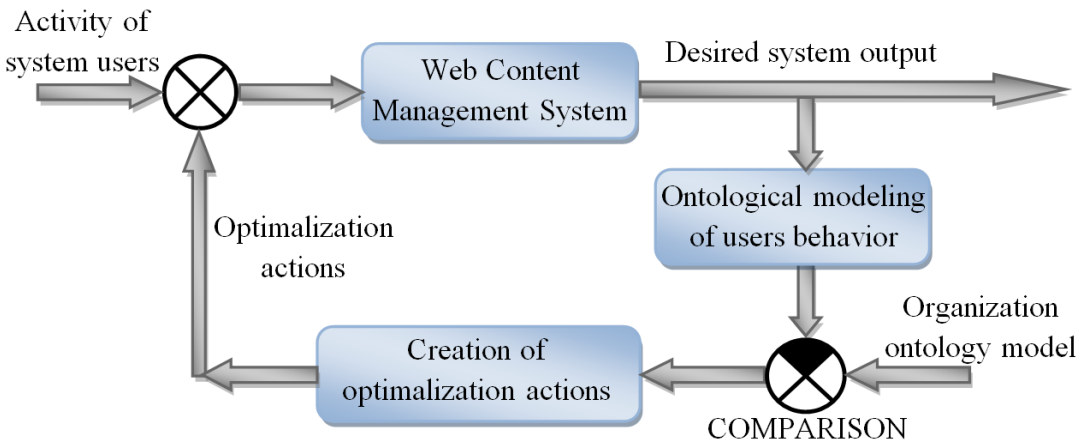
I. Introduction

In these hectic times, the efficiency is maybe the most important thing to maintain, or even enhance. One source of inefficiency is difference between desired output of an organization and actual state of output at same resources. To improve this kind of efficiency is important to minimize this difference. The biggest problem in solving this type of inefficiency is to accurately define this difference. Only if we are aware of this difference, we can effectively eliminate our problem with ineffectiveness within organization. Process of elimination of inefficiency is called optimization. In this paper I will describe logic of possible solution for this problem.

II. Organization And Web Content Management System With Optimalization As Regulated System

To better understanding of solution logic, I describe it on concrete possible deployment of my optimalization process. One of possible implementations of this optimalization process is application in web content management system. There is one important requirement for implementation and that is need of strong interconnection between deployment of web content management system (WCMS) and function of organization, where is this WCMS deployed. With this interconnection can mean that the output of WCMS is final product of organization. An example is journalistic web portal from inter-organizational perspective. In this case the concrete problem of inefficiency can be subscribed as difference of what users should do and what they are doing.

Now that we know the meanings of problem, we can proceed to solving part. The optimalization process consists of five major parts: WCMS, ontological model of organization, ontological modeling system of users' behavior, comparator and creator of optimization actions. Function of the optimalization system is displayed on Fig. 1. Further in this paper I will describe functionality of particular parts of the system.



I. Fig. 1. System optimalization schema

III. Optimalization System Blocks

A. Web Content Management System

In this system it is not needed to have special WCMS. All web content management systems that provide detailed logs about users actions in the system fit the bill.

B. Ontological Modeling of Users Behavior

This is maybe the most important part of whole optimalization system. This module provides all functionality needed for creation of an ontological model of user usual behavior. This is nonhierarchical model that can be represented as a relation $R \subseteq users \times actions$. Where *users* are concrete users of WCMS and *actions* represent actions that WCMS users can perform within WCMS.

C. Organization Ontology Model

This part of system is created manually. Specifically the creation of ontological model of organization is preformed fully manually. The model describes hierarchical layout of roles within organization. This model can be represented by relation $R \subseteq roles \times actions$, where *roles* are concrete roles within organization and *actions* are the same actions that are in first, relation.

D. Comparator

In this case the comparator is not just a simple negative summator of signals. Purpose of this comparator is to determine the ontology model difference as shown on Fig. 2. Every difference between those models is potential ineffectiveness. Each of those differences has to be properly examined by organizations management, because not all of those differences are strictly taken as our problem. There can be even some exceptions, which can be only mistakes in ontology modeling on each side of system. On side of users behavior, there can some of them been out of office during system log filling process. On the side of organization, there can be some roles, that are special and ontology engineer has disregarded those fact during modeling process.

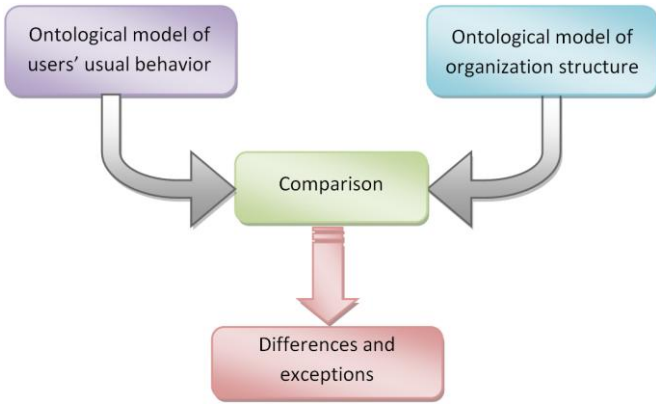


Fig. 2. Comparison logic

As the results of this comparison are list of differences and list of exceptions. Those lists pass to another step, which is the creation of the list of actions that may increase organization efficiency.

E. Creator of Optimalization Actions

There are can be three approaches to create optimalization actions depending on what we are willing to change in whole system:

1. Actions that will adapt organization structure.
2. Actions that will adapt WCMS.
3. Combination of both actions types.

First type of actions can be taken when we have users that are doing less or more than they are expected. In order to improve effectiveness we can descent user that isn't doing his job and we can promote user that is doing more work as is he expected to motivate him. My recommendation is to find cause of this difference first that we do changes to organization structure.

Second approach is to find reason of difference and try to modify web portal to ease users' access to parts and functions that they are use to use. Some of WCMS already have subsystem that dynamically alter web portal to user needs. But they are based on users' feedback, so the user must do more actions to customize his web part behavior. With using of optimalization process the feedback is provided automatically by using web parts and functions that they desire.

The third approach should be most common for all of uses, because it combine both advantages of using optimalization system.

IV. Why Ontology?

In this paper I often mentioned ontological modeling and ontology model, but I never explained why it is good to use ontologies. There are two major advantages that ontologies provide to this system.

First is that the ontology already have steady theoretical platform, so the theory of ontology should not change.

Second in order is that the ontology has its standardized modeling languages that have lot of existing software applications and tools. Because of that it is not needed to deal with it.

Last but not least advantage is that ontology was created to be tool for modeling the world and any type of knowledge. In my opinion there is no better way to model things than use of tool created specifically for that purpose.

V. Conclusion

As you could see, in this paper I have described just the very essence of optimalization process. In my further work I will provide closer look to all parts of optimalization schema.

References

[1] BOIKO, B.: „Content Management Bible“, Willey Publishing Inc., Indianapolis, USA, 2005
[2] CHIN, K.: „Evaluating and Selecting a Content Management Solution“, Gartner Symposium ITXPO, Orlando, USA, 2001
[3] MARTIN, N.: „DAM Right! Artesia Technologies Focuses on Digital Asset Management“, EContent, September 2001
[4] „Design and Development of an Ontology – Integrated Content Management System“, University of Leeds, Faculty of Engineering, 2005
[5] ANDREJKO, A. – BARLA, M. – BIELIKOVÁ, M.: „Ontology-based User Modeling for Web-based Information Systems“, In: Proc. of 15th Int. Conf. on Information Systems Development, Prague, 2005
[6] GUNTHER, G.: „Cybernetic Ontology and Transjunctional Operations“, In Self-Organizing Systems, February, 2004
[7] LAGOZE, C. – HUNTER, J.: „The ABC Ontology and Model“, Journal of Digital Information, Volume 2, Issue 2, November, 2001
[8] MIZOGUCHI, R.: „ Tutorial on ontological engineering: part 3: Advanced course of ontological engineering“, New Generation Computing, Vol. 22, No.2, p.198-220, January 2004
[9] POUCHARD, L. – IVEZIC, N. – SCHLENOFF, C.: „Ontology Engineering for Distributed Collaboration in Manufacturing“, AIS Conference, Arizona, March, 2000
[10] SMITH, B. – GRENON, P.: "The Cornucopia of Formal-Ontological Relations", Dialectica 58, No. 3, 2004, pp. 279-296
[11] SPYNS, P. – MEERSMAN, R. – JARRAR, M.: "Data Modeling versus Ontology Engineering", In SIGMOD Record, Vol. 31, Issue 4, New York, December, 2002, pp. 12 -17
[12] VILJANEN, K. – TUOMINEN, J. – HYVONEN, E. – MAKELA, E. – SUOMINEN, O.: "Extending Content Management Systems with Ontological Annotation Capabilities", Proceedings of the First Industrial Results of Semantic Technologies Workshop, ISWC2007, November 11, 2007

Automated Channel Changing in IPTV

[1]*Pavol KOCAN,* [2]*Ján MOCHNÁČ,* [3]*Branislav HRUŠOVSKÝ*

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]pavol.kocan@tuke.sk, [2]jan.mochnac@tuke.sk, [3]branislav.hrusovsky@tuke.sk

***Abstract*—Interest in digital video delivery increase rapidly. There are many of multicast streams delivering television broadcast today from which customer select one accordingly his interest. Delays in 'tune-in' acquisition of live streams emergent not only in the network and network equipment but even in the end user devices. This paper proposes technique of general applicability to minimize the channel changing delay perceived by the user in internet television. There are several techniques how to minimize particular delays. By using electronic program guide data could we predict time of the next channel change and with combination of the end user statistical record at receiver can we make channel changing smoother and more comfortable.**

***Keywords*— channel changing, EPG, FEC, IPTV, IGMP.**

I. Introduction

Watching television is still one of the most popular activities today. Tenths of providers, hours of entertainment each day. With growing interest in digital video delivery also grows the interest in providing of sufficient video quality and user friendly services that make watching of internet television (IPTV) more comfortable. There are several ways for satisfying quality and services demands, especially for IPTV providers where the receiver feedback is available.

This paper is organized as follows. Second section presents the theoretical background of IPTV channel changing delay, the third section presents our proposition for fluent channel changing if the user statistical data is used. We finish this paper with short conclusion and proposal for the next research.

II. IPTV Channel Changing Delay

Home user has available a lot of the television channels when using internet television service (IPTV). Channel changing delay can be described by several components:

- Real Time Streaming Protocol (RTSP) negotiation (if requesting streams or stream data),
- Internet Group Management Protocol (IGMP) delay,
- network latency,
- video and audio Random Access Points (RAP) acquisition,
- client stream buffering delay,
- synchronization between streams with Real-time Transport Control Protocol (RTCP) sender report,
- Forward Error Correction (FEC) block boundary acquisition
- and decoding delay in Set Top Box (STB).

A. RTSP Negotiation

RTSP is a protocol that can be used to initiate tune-in to a multicast stream or to request a unicast stream from a server. When it is used, it can take several packet round-trips of request/response before the RTSP server sends the first media packet. In some situations, this can be significant [3].

B. IGMP

Major part of channel changing time is IGMP latency. IGMP provides four basic functions for internet protocol (IP) multicast networks: join, leave, query and membership report. In an IPTV network, each broadcast television channel is an IP multicast group. The subscriber changes the channel by LEAVE-ing one group and JOINing a different group.

There are three versions of IGMP. IGMP version 1 is not used for IPTV because it does not include an explicit "LEAVE" capability. IGMP version 2 and version 3 can both be used for IPTV. Summarizes the major differences between IGMP v2 and v3 can be found in reference [2].

Multicast is an efficient distribution protocol for live streams as each network link needs carry only one copy of each stream. Routers in the network replicate streams from inputs onto the output links which have one or more clients.

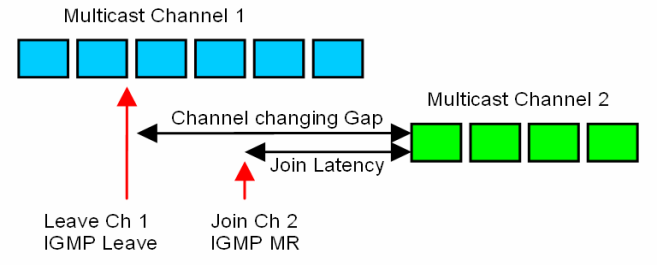


Fig. 1. Channel changing and IGMP delay

Clients indicate their interest in the multicast by sending a special control packet (IGMP), which is intercepted by the router. The routers in turn refer to each other to find, and forward, the packet stream [3].

In this example the subscriber is synchronized to channel 1 (azure frames). At a particular time, the subscriber issues a switch command to channel 2 (green frames), which triggers IGMP leave to the multicast stream group 1 and joins the multicast stream group 2. Then, the subscriber starts to receive multicast stream 2. Since "I" frames act as stream synchronization points for the subscriber decoder, the subscriber waits until an "I" frame is received. The received frames are discarded. The waiting time (channel changing gap) is determined by the number of frames being offered per second (Fig. 1). When the "I" frame is completely received and decoded, the subscriber is synchronized to the new stream, which is marked by STB [1].

C. RAP Acquisition

Video is classically 'differentially coded'. Very few coded frames contain all the information needed to create a complete set of decoded pixels. Instead, frames are coded as differences from one or more other frames.

In order to handle both recovery from loss and tune-in to live streams, 'independently decodable' frames or I-frames can be sent periodically. Decoder refresh frames or IDR-frames are I-frames that have the additional property that they mark a division of the sequence; frames displayed after the decoder refresh do not depend on frames before it. They are true Random Access Points.

Because these frames are so much larger (in bytes) than differentially coded frames, they are sent rarely in order to keep the bit-rate low. A delay occurs when tuning into a new stream, since the decoder must wait for a RAP before it can start displaying video. In addition, their size often means that traffic smoothing causes adjustment of their send time and that of adjacent packets [3].

D. Client Buffering

If there is deviation in the arrival time of packets from the relative timing that they would have if they were to arrive just-in-time to be played, then a de-jitter buffer is needed if we are to avoid under-run (starvation). This jitter has two causes: deliberately introduced jitter from the source, for traffic smoothing, and network-introduced jitter (e.g. caused by cross-traffic in network equipment). The source-introduced jitter tends to occur most, or even exclusively, in video, where the variation in coded frame size (in bytes) can be large (e.g. I-frames can be many times as large as B frames). A buffer is also needed if there is to be time to perform re-transmissions. A delay occurs during the time the client fills its buffer before starting to render [3].

E. FEC block boundary acquisition

Forward Error Correction (FEC), if used is usually applied to blocks of packets of the stream. Playout of the stream must be delayed at the client by a time equal to the largest such block in the stream, to allow time for blocks to be corrected if there is packet loss. Additionally, playout usually would not be started until the first packet of an FEC block, since lost packets before this point in the stream cannot be corrected by the FEC.

FEC codes may be systematic or non-systematic. In the case of systematic codes, the original data is sent followed by a number of "repair" packets which can be used at a received to recover original ("source") packets which were lost. In some cases, insufficient data may be received to perform the FEC recovery operation, in which case the received source packets may be passed to the A/V decoders for playout, potentially with loss concealment. There are still more research to improve this concealment techniques [3, 4].

F. Processing delay

Processing delays can occur in the terminal at a number of layers – network, RTP, codec, and so on. In general, this is a trade-off between terminal resources (memory, processor speed) and cost. However, there are recommendations that can be made to minimize the processing load on the end-system, and hence the delay.

III. Channel Changing Model

One of the simplest techniques for reducing channel switch times and to ensure viewer satisfaction that we present in this paper is to predict what channel the user will next ask for, and tune that in early. This could be done using the viewer statistical index (VSI) that is part of presented statistical model. Statistical model observe user interest and gathers statistical data about the viewed content, like genre type (movies, sport, news, story, etc.), watching time period, rate of channel changing. It's up to user whether on not to use this service in his STB receiver.

A. Genre Mark

An Electronic Program Guide (EPG) is an application used with digital set-top boxes and newer television sets to list current and scheduled programs that are or will be available on each channel and a short summary or commentary for each program. EPG is the electronic equivalent of a printed television program guide. An EPG is accessed using a remote control device. Menus are provided that allows the user to view a list of programs scheduled for the next few hours up to the next seven days. A typical EPG includes options to set parental controls, order pay-per-view programming, search for programs based on theme or category, and set up a VCR to record programs. Each digital television (DTV) provider offers its own user interface and content for its EPG.

If the time stamp from EPG is accurate enough it can be used to generate the genre mark that will inform receiver about the next genre in broadcasted television channel so the receiver can quick predict channel changing time. This genre mark can be obtained not only from EPG but also from broadcasted time code of server broadcasting to the customer. Proposed genre mark could contain program genre and precise time interval for each program.

B. *User Statistical Data*

Everyone is unique. The same can be said about television watching. Such uniqueness can be recorded in each receiver and offered to IPTV provider to better understand the customer demands. Statistical log implementation to the receiver is simple and consequently evaluated for next utilization. For mathematical statistic ideal is statistical mode, the most common value obtained in a set of observations [5]. Values could be obtained in one minute interval for better precision. This statistic is usefull to do for specific genre and/or for specific television programmes as for television channel whereas the channel could be general genre.

Output proposes statistical probability which channel is going to be switched after finishing current program. VSI of the most probably channel is generated and first two (three eventually) one could be ready to tune-in if necessary or displayed in Picture in Picture (PIP) form.

C. *Automated Channel Changing*

The process is illustrated on next two figures Fig. 2 and Fig. 3. There are three multicasted channels, blue one represent the first multicasted channel that is active currently. Customer watching sports (azure frames) after which the movie (blue frames) will follow in short time period. Sport program end time is obtained from EPG or provider time code and receiver is getting ready to leave multicast channel one. At the same time VSI recommend two (three) other channel that are displayed in PIP function on the screen. It's

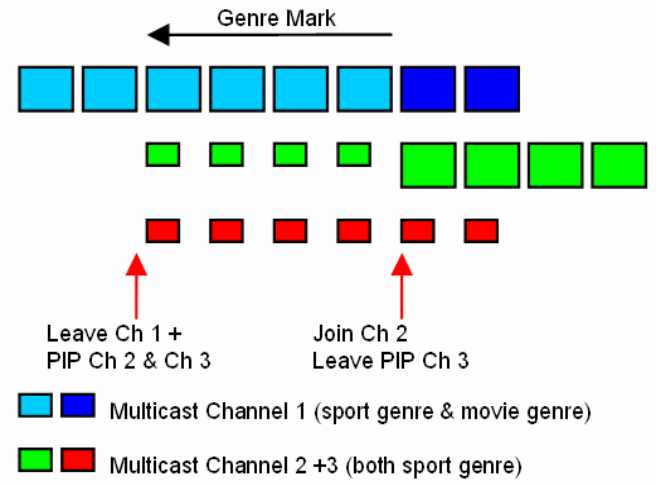


Fig. 2. Channel changing model

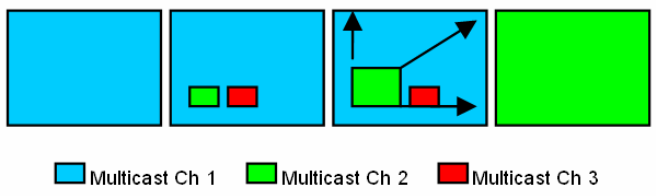


Fig. 3. Channel changing on screen

up the customer to select the next channel. Multicast channel two and three (green and red frames) are delivered at lower resolution as the channel one (quarter or just one eighth compared to channel one). If customer doesn't change the channel himself the channel will be changed automatically. Green channel is selected considering higher probability of customer interest in. After the channel change is done, multicast channel two is displayed on full screen and PIP function is finished.

IV. Conclusion

In this paper we focused on channel changing delay and service called automated channel changing. By using EPG data could we predict time of the next channel change and with combination of the end user statistical record at receiver could we make channel changing smoother and more comfortable.

Disadvantage of this technique is higher bandwidth consumption for multiple streams for each client, and this may use excessive bandwidth. For deployments based on Very High Bitrate Digital Subscriber Line (VDSL) and Fiber to the Home (FTTH) there may be enough bandwidth available.

In any case, predictive tuning is a technique that can be implemented at the client, most probably without the need for additional protocols in new type of integrated digital television receivers with direct connection to the internet protocol network.

Another research will include testing on simulation software NS-2. It is necessary to concentrate mainly on reducing packet drops by adjusting the stream's rates and selecting the right parameters for congestion control algorithm at the routers to ensure a satisfaction of using an automated channel channing in real service.

Acknowledgment

This publication is the result of the project implementation Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF.

This work is part of a research project supported by VEGA no. 1/0045/10.

References

[1] T.Janevski, Z. Vaneski, Statistical Analysis of Multicast versus Instant Channel Changing Unicast IPTV Provisionig, 16th Telecommunications forum TELFOR, 2008, p. 96-99

[2] Introduction to IGMP for IPTV Networks, 2007

[3] *Fast Channel Changing in RTP*, Internet Streamnig Media Alliance, December 2007

[4] J.Mochnáč, P.Kocan: *Performance evalutation of error resilience and error concealment in H.264*, 9th Scientific Conference of Young Researchers, Košice 2009, p. 64-65

[5] Weisstein, Eric W. "*Mode*." From MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/Mode.html

An Overview of Network Overlay used in Distributed Recording System

Michal KOHUT

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

michal.kohut@tuke.sk

***Abstract*—This paper is a short overview of the current state-of-the-art solutions based on the network overlay models widely used in popular file-sharing peer-to-peer systems. It describes and compares two basic types of network overlays; unstructured network overlays (Gnutella-like systems) and structured network overlays (DHT based systems). There is shortly introduced Golem - a distributed recording system for real-time processing multimedia data, while the attention is focused on the topology of its structured overlay. The paper shows benefits as well as drawbacks of currently used tree-based structure in Golem and suggests possible performance improvements by proposing a new unstructured network overlay which should be an efficient, simple, adaptive and fault-tolerant solution.**

***Keywords*—Network overlay, distributed system, peer-to-peer system, videoconference, multimedia recording**

I. Introduction

Distributed systems have emerged to play a serious role in industry and society only in last couple of years while the progress done in the computer networks over last 30 years is much more significant. Nowadays only raw computing power is not enough. Users require fast, reliable and secure systems with high level of availability. It is a distributed system which can fully meet all those requirements. However the engineering discipline of reliable distributed computing is still in its infancy. Only few distributed systems are truly reliable in the sense of automatic toleration of failures, guaranteeing availability and good performance even in a stress condition and offering sufficient level of security against various threats.

Large distributed environments, such as peer-to-peer systems, have brought up big interest in the complex distributed network systems without the centralized control. Compared with the traditional client-server architecture, a pure peer-to-peer system can be described as a distributed network system in which all participant computers (also known as peers or nodes) have symmetric duties and responsibilities. Key features of these systems are decentralized control, self-organization, dynamism, and fault-tolerance. All nodes act as both clients and servers to one another, leading to a large pool of information sources and computing power. There is no ”super” node which the function of other nodes depends on. There is no node whose failure can cause the global failure of the system. Any node can be substituted by any other node what makes the system truly fault-tolerant.

Constructing a distributed system, with completely decentralized control, which contains large numbers of common computers that randomly join and leave the network is non-trivial task. Significant impact on application properties such as performance, reliability and scalability has the topology of the overlay network.

An overlay network of a distributed system is a logical network built on top of an underlying physical network. Designing effective overlay models attract many academic researchers from the networking and the distributed systems communities. The ”beaty” lies in the simplicity of the solution and its ability to completely eliminate central authority. The ”ugliness” lies in the huge amount of traffic that makes the solution unscalable. It is very important to note that decentralized nature of the distributed system causes determining its global properties by local decisions. These local decisions are made by individual nodes and they are based only on local information. We are dealing with self-organized network of independent entities [1].

This paper is an overview of basic network overlay models which are used in peer-to-peer networks. It also introduces the overlay model used in our own distributed recording system Golem with the description of its benefits, drawbacks and possible improvements.

II. Distributed recording system

Golem[1] is a distributed system which performs automatic recording of videoconference meetings. This distributed recording system is being developed at Computer Networks Laboratory[2].

Primary task of Golem is to automatically make a record of a currently running videoconference. A videoconference, defined as a set of interactive telecommunication technologies, allows two or more locations to interact via two-way video and audio transmissions simultaneously. It is a live connection between people in separate locations for the purpose of communication, usually involving audio and video. Each videoconference meeting is completely unique as well as a multimedia stream coming from a videoconference. Data from a source of the stream are unrepeatable, there is only one chance to capture them and to make a real-time record of these data. If something in the system goes wrong, the recording procedure fails and there is no disaster recovery support (e.g. backup recording procedure), the record will be incompleted. And this situation is of course highly ineligible. Therefore the most critical part in the design of such kind of recording system is all above the *reliability*. The system processing real-time data must be :

- stable

[1]See http://atlantis.fw.sk/projects/golem/
[2]See http://www.cnl.sk

- fault-tolerant
- scalable
- reliable

It must stay online in any situation that may occur. The system must be able to recover from failures, be secured and fully functional as long as possible and moreover it should be simple and cheap to implement. Considering all these requirements, *Golem* has been designed as a *distributed system* consisting of common computers (assuming from tens to hundreds of them) with completely decentralized control.

The Golem system consists of two types of nodes :

- *manager node* - is responsible for the whole functionality of the distributed recording system. This node actually ”controls” the system. It builds the network overlay, it is a communication gate between the system and the outer world, it handles client's requests, it chooses nodes which will perform the recording of the events, it provides the monitoring of the other nodes etc.
- *storage node* - handles *only* the recording process. Its task is very simple. At given time the storage node starts recording and at the end of the event it stops recording. It is very important to have enough nodes of that type in the system to ensure a desired level of reliability.

Fig. 1 shows the architecture of the distributed recording system Golem.

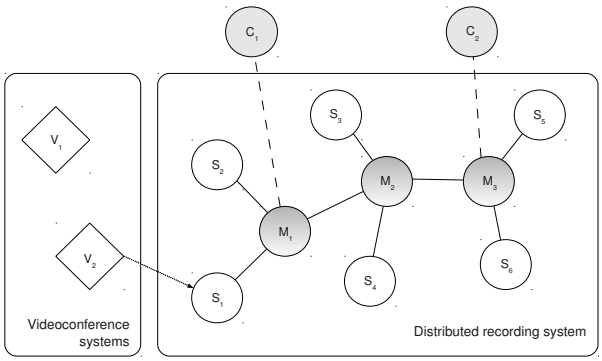


Fig. 1. Architecture of the distributed recording system Golem

Nodes marked as M_x represent the manager nodes, nodes marked as S_x are storage nodes. Manager nodes are interconnected to each other and they are also connected to the storage nodes. Storage nodes can connect only to the managers. The logical network topology of the distributed recording system is based on the layout of the the manager nodes. They are responsible for building the network overlay of the system. Layout of the storage nodes is balanced, every manager tries to connect approximately the same number of storages to itself. Storage nodes are being connected to the manager nodes equally.

Nodes marked as C_x are clients' stations or clients' applications which are connected to the system. They can connect to any manager node but only to that type of node. Nodes marked as V_x are videoconference systems - the source of video and audio data which is recorded.

The reliability of the distributed recording system is ensured by different approaches. Key feature of the system is *load-balancing*. Every time a new request for recording is entered, the most available storage node is chosen to make a record of it. The work load of the system is effectively balanced between all storages. Each storage node is also monitored by its parent manager node, its ”watcher”. If a storage node goes down, this watcher immediately finds a backup storage node to continue recording.

One of the most important thing which has a direct impact on the *performance and reliability of the distributed recording system is choosing* the topology of *the network overlay*. Efficient communication and optimal topology can ensure faster recovery procedures, shorter message delays, better usage of the underlying physical network and lower overhead traffic.

III. Network overlays

An overlay network is a virtual network of nodes and logical links that is built on top of an existing network with the purpose to implement a network service that is not available in the existing network. Overlays are used as a mechanism to deploy new distributed applications and protocols on top of the Internet. Participating nodes communicate with each other through *tunnels* which are virtual links between nodes. One tunnel always interconnects two nodes which may not be directly connected in the underlying network. These tunnels define the topology of the network overlay. A self-organizing overlay protocol ensures an efficient and connected topology even if the underlying network changes or nodes join and leave the network. Depending on the type of configuration of the overlay, there exists [2]:

- statically configured overlays
- dynamically configured overlays

Statically configured network overlays are the most currently deployed overlays. However this simple and straightforward technique has many disadvantages. Static configuration has a high management overhead when the overlay grows. It cannot adapt to the changes that occur in dynamic environment where nodes frequently join and leave the network. Improvement of the overlay performance due to changes in the underlying network is also impossible. These are only few reasons why the statically configured network overlays will be sooner or later completely displaced by the self-organizing overlays.

Dynamic or self-organizing network overlays are design to be highly adaptable and scalable. They can accept sudden changes in the network and they are able to adapt to them. Design and configuration of the dynamic network overlays is more difficult but further maintenance is incomparably lower than in the static overlays.

In the terms of network structure, dynamic network overlays are roughly classified into two categories:

- unstructured
- structured

Big progress in the development of the network overlay models has been done mainly in the last ten years when the peer-to-peer systems became extremely popular, especially by file-sharing systems like Napster. [3].

A. Unstructured network overlays

Unstructured overlay networks have been implemented in the first wave of the peer-to-peer systems. These overlays are characterized by random data placement and maintaining no global structure. The most known system from this category

is Gnutella [4]. To look for a file, unstructured systems use message flooding to propagate queries. In this process each computer connects to random members in a peer-to-peer network and queries its neighbors who act similarly until a query is resolved. Although such systems are fault-tolerant and resilient to users joining and leaving the network, their search mechanism does not scale. Queries for content that is not widely replicated must be sent to a large fraction of nodes.

Gnutella do not optimize the unstructured overlay and it does not collect routing information. The protocol is simple and there is no need to maintain routing tables. The drawback is overhead traffic for each message and the resulting inefficient network usage. The generated traffic is proportional to the number of tunnels. Another thing is the high network stress that results from the random overlay construction [5].

Other unstructured overlays like Narada [6] [7] and Scattercast [8] uses routing protocol to improve overlay performance. This protocol provides each node with information on all other participants in the system and on the optimal cost and path to each. This approach brings optimal routes between peers and also adaptability to the changes occurring in the underlying network. On the other hand there is overhead in maintaining the routing information. The size of routing tables which are being periodically exchanged between nodes is proportional to the number of nodes in the network.

B. Structured network overlays

Structured overlays, which implement a *Distributed Hash Table* (DHT) [9] data structure, were proposed to increase the scalability of unstructured systems. They assign keys to data items and organize the overlay nodes into a graph that maps each key to a responsible node. The graph is structured to support efficient data discovery by given keys but it does not support complex queries. Structured network overlays provide high level of scalability what means that DHT should work efficiently for overlay networks of arbitrary size. This implies handling node arrival and departures in a scalable fashion. However mapping the overlay to the underlying network is not so efficient as in the unstructured overlays. Also the handling the churn (movement of nodes - joining and leaving the network) requires more complex mechanisms. The most known DHTs are CAN [10], Chord [11], Pastry [12], Tapestry [13] [14], P-Grid [15] and Kademlia [16].

C. Evaluation of network overlay performance

Three basic metrics are used to evaluate performance of the network overlay [5]:

1) *Efficient usage of the underlying network* - each network overlay protocol should try to minimize relative delay penalty (RDP) and stress.
   a. *RDP* is the ratio of the latency experienced when sending data using the overlay to the latency experienced when sending data directly using the underlying network.
   b. *Stress* is defined as the number of overlay tunnels that send traffic over the same physical link, it means number of tunnels that are mapped over the same link.
2) *Scalability* - good overlay protocol tries always to minimize the overhead for maintaining the topology. Ideally the overhead should increase linearly with the increasing number of nodes.
3) *Adaptation* - network overlay should adapt to the changes occurring in the underlying network and also it should maintain efficient topology with the optimal routes.

IV. Golem network overlay model

Distributed recording system Golem is very similar to a peer-to-peer system. It consists of common computers (peers) which are connected to the Internet to create a large distributed environment for working with real-time multimedia. The main difference between the traditional peer-to-peer system and Golem system is that Golem is not used for sharing data. It is a distributed system where the nodes cooperate together to solve given task - to provide reliable recording service.

One of the most important rule in the design of the network overlay topology is [2]: *Design of self-organizing network overlay protocol must be tuned toward a particular application.*

Golem system currently uses a *tree-based* structured network overlay. The structure is *AVL tree* [17] - self-balancing binary search tree. Each node of the system must have its unique identification mark - *value*. Nodes in the AVL tree are ordered. For each node n applies that the value of any node in the left subtree is lower than the value of the node n and the value of any node in the right subtree is greater than the value of the node n. Any two nodes can be compared by comparing their values and so they can be added to the proper place in the tree structure.

A. Benefits

The main advantage of the usage of AVL trees is no need of explicit routing mechanism. Thanks to ordered tree structure each node knows where to forward message based on the value of a target node. Broadcasting messages is being done also very efficient, each node just forward message to all links except the one from the message was received. There is no traffic overhead and no duplicated messages.

B. Drawbacks

On the other hand, mapping the tree-based network overlay to the underlying network and its adaptation to the changes occuring in the underlying network is very poor and it requires global knowledge of the system. Another problem is an existence of a root node which may cause a bottleneck and also repair penalty for a node failure is generally higher in a tree. The maximum height of AVL tree is $1.44 * log_2 n$, where n is the number of nodes. In the real world it means that the maximum length between two endmost nodes in the system with 130 peers would be $2 * 1.44 * log_2 130 \cong 20$. So the message must pass 20 hops to reach the destination, it's simply too much.

C. Improvements

As mentioned before, the design of the network overlay should be tuned toward a particular application. Primary advantage of the structured overlay is efficient data searching in file-sharing peer-to-peer systems. However distributed recording system is not used for file-sharing. It is designed as

a reliable and stable recording system with high level of fault-tolerance. Therefore we assume that a possible improvement of Golem could be in usage of the unstructured network overlay, with emphasis on the effective mapping of the application overlay to the underlying network with the target to minimize values of RDP and stress. It is commonly believed that unstructured overlays have lower maintenance overhead than structured overlays, especially when there is a high churn rate.

Fig. 2 shows an example of efficient mapping of the application network overlay to the underlying network.

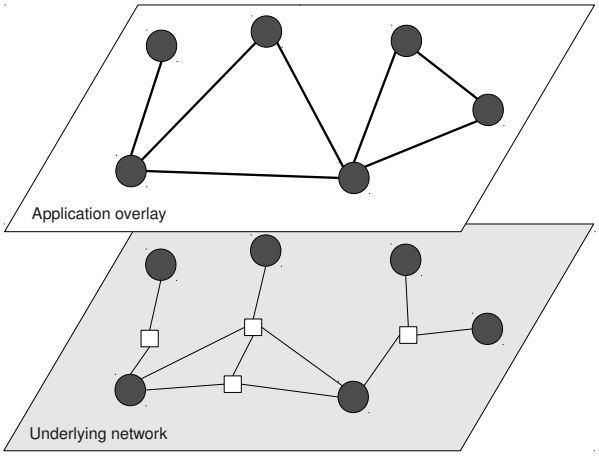


Fig. 2. Possible mapping of the application overlay to the underlying network

Serious problem of the most known unstructured network overlay Gnutella is its lower level of scalability. Lower scalability could be improved by employing available networking resources efficiently. Such kind of architecture should be completely adaptable to the changes occurring in the underlying network and it should also provide fast recovering mechanisms in case of nodes' failure. Another issue which is typical for the unstructured overlay is overhead traffic when broadcasting a message. This structure cannot ensure that one node will receive a broadcast message only once. One of the possible ways how to reduce broadcast traffic overhead is decreasing the number of neighbors for each node. However this method can lead to longer average path between two endmost nodes in the system.

Anyway, the deployment of the proposed unstructured network overlay in the distributed recording system Golem may bring at the first sight better results so this solution should be more explored.

V. Conclusion

In this paper we have made a short overview of the existing network overlay models which are widely used in the popular file-sharing peer-to-peer systems. Each model has its benefits as well as a drawbacks. Choosing the right model to use completely depends on the character of the specific application.

We have also introduced currently used network overlay model in Golem system, briefly described its features and tried to show possible improvement and future work. Usage of the unstructured network overlay in the distributed recording system Golem seems to be an efficient, simple and adaptive solution providing a good level of fault-tolerance. Assuming the number of participating nodes, from tens up to hundreds, the proposed overlay model could be also sufficiently scalable without causing the high traffic overhead.

References

[1] S. El-Ansary and S. Haridi, "An overview of structured overlay networks," *Theoretical and Algorithmic Aspects of Sensor, Ad HocWireless and Peer-to-Peer Networks*, 2005.

[2] S. Jain, R. Mahajan, D. Whetherall, and G. Borriello, "Scalable self-organizing overlays," *Technical report UW-CSE 02-02-02*, February 2002.

[3] Napster. (2000, Jun.) Open source napster server @ONLINE. [Online]. Available: http://opennap.sourceforge.net/

[4] Gnutella. (2000) The gnutella protocol specification @ONLINE. [Online]. Available: http://opennap.sourceforge.net/

[5] M. Ripeanu, I. Foster, A. Iamnitchi, and A. Rogers, "A dynamically adaptive, unstructured multicast overlay," in *Service Management and Self-Organization in IP-based Networks*, 2005.

[6] Y.-H. Chu, S. G. Rao, S. Seshan, and H. Zhang, "A Case for End System Multicast," in *Proceedings of ACM Sigmetrics*, 2002, pp. 1–12.

[7] Y.-H. Chu, S. Rao, S. Seshan, and H. Zhang, "Enabling Conferencing Applications on the Internet using an Overlay Multicast Architecture," in *Proceedings of ACM SIGCOMM*, 2001, pp. 55–67.

[8] Y. Chawathe, "Scattercast: An adaptable Broadcast Distribution Framework," *ACM Multimedia Systems Journal special issue on Multimedia Distribution*, 2002.

[9] S. Jain, R. Mahajan, and D. Wetherall, "A Study of the Performance Potential of DHT-based Overlays," in *Proceedings of the 4th Usenix Symposium on Internet Technologies and Systems (USITS)*, 2003.

[10] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Schenker, "A scalable content-addressable network," in *SIGCOMM '01: Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications*, 2001, pp. 161–172.

[11] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan, "Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications," in , 2001, pp. 149–160.

[12] A. Rowstron and P. Druschel, "Pastry: Scalable, Decentralized Object Location, and Routing for Large-Scale Peer-to-Peer Systems," *Middleware01:Proceedings of IFIP/ACM International Conference on Distributed Systems Platforms*, pp. 329–350, 2001.

[13] B. Y. Zhao, B. Y. Zhao, J. Kubiatowicz, J. Kubiatowicz, A. D. Joseph, and A. D. Joseph, "Tapestry: An Infrastructure for Fault-tolerant Wide-area Location and Routing," , Tech. Rep., 2001.

[14] B. Y. Zhao, L. Huang, J. Stribling, S. C. Rhea, A. D. Joseph, and J. D. Kubiatowicz, "Tapestry: A Resilient Global-scale Overlay for Service Deployment," *IEEE Journal on Selected Areas in Communications*, vol. 22, pp. 41–53, 2004.

[15] K. Aberer, "P-Grid: A self-organizing access structure for p2p information systems," *Proceedings of the Sixth International Conference on Cooperative Information Systems*, 2001.

[16] P. Maymounkov and D. Mazieres, "Kademlia: A Peer-to-peer Information System Based on the XOR Metric," *IPTPS02: Proceedings of the 1st International Workshop on Peer-to-Peer Systems*, 2002.

[17] G. Adelson-Velskii and E. M. Landis, "An algorithm for the organization of information," in *Proceedings of the USSR Academy of Sciences*. Springer, 1962.

Anomaly Detection Techniques for Adaptive Anomaly Driven Traffic Engineering

[1]*Jakub Kopka,* [2]*Martin Révés,* [3]*Juraj Giertl*

Dept. of Computers and Informatics, FEEI TU of Košice, Slovak Republic

[1]jakub.kopka@cnl.tuke.sk, [2]martin.reves@cnl.tuke.sk, [3]juraj.giertl@cnl.tuke.sk

***Abstract*—Traffic engineering (TE) has become mechanism for safe and efficient transportation of data in a computer network. TE uses statistical methods for prediction of a network traffic behavior. However, the traffic behavior will never match this predictions 100%. The aim of this paper is to describe detection techniques, which can be used for detection of the anomalous traffic in computer networks. We propose all known methods, which are suitable for the anomaly detection in different application domains and suggest the best techniques for the detection of anomalous traffic.**

***Keywords*—Computer networks, Computer networks management, Traffic control, Traffic engineering**

I. Introduction

Inadequate utilization of network resources is challenging problem for network traffic engineers. TE allows the optimization of the network resources usage through multiple mechanisms. This optimization is difficult due to dynamic nature of a network traffic. The network traffic can be characterized by several parameters. This parameters may come to be recorded and a model of traffic can be build. If one or more parameters will have a value that is different that one predicted by the traffic model, we call it an *anomaly*. This anomalies are monitored by a distributed monitoring system. The monitoring system can detect and localize the cause of anomalies and respond appropriately by reconfiguring the network. We call this approach *Adaptive Anomaly Driven Traffic Engineering*.

This paper is organized as follows. In the section II we describe and divide anomalies into categories, describe application domains, where anomaly detection techniques can apply and name the most known of them. Section III presents classification based anomaly detection techniques. In the section IV we present techniques for detecting contextual anomalies. Section V presents techniques for detecting collective anomalies. Section VI is conclusion of the previous sections and presents our suggestion, which of the detection techniques is the most suitable for our traffic engineering approach.

II. Classification of Anomalies and their Detection Techniques

A. Anomaly

An anomaly is a deviation from the common rule, type or form. The anomalies are patterns in data that do not conform to a well defined notion of a normal behavior [1]. In relation to the network traffic, the anomaly is any deviation of the expected traffic behavior. This anomaly significantly distincts from the normal traffic and influences one or more links in network. For example, Fig. 1 illustrates anomalies in a simple

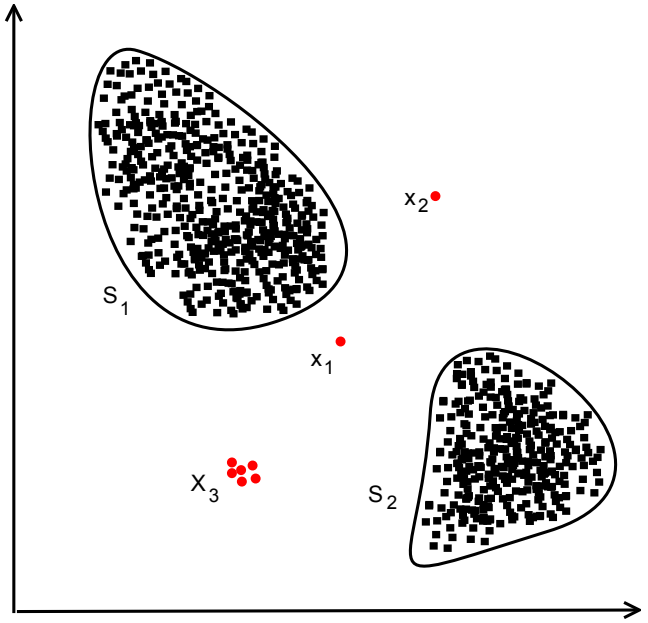


Fig. 1. A simple example of anomalies.

data set. This set of data has two areas of normal data S_1 and S_2. It is because the most of the data instances lie in these two areas. Points x_1 and x_2 are anomalies, because they are significantly far away from this areas. Also the set X_3 denotes anomalies, although it contains more than one data instance.

Detecting such patterns in the data, which are different from the normal expected behavior is called an anomaly detection. The anomaly detection is used in several application domains e.g. image processing, card fraud detection, pharmaceutical research, network intrusion detection systems and many other. All these application domains provides data which can be analysed for presence of anomalies.

Data are input for all anomaly detection techniques. These data are described by the set of attributes which can be binary, categorical or continuous. Data can be univariate or multivariate if it has one or multiple attributes repectively. Character of the data determines which of the techniques can be used.

The anomalies are divided into three categories:

- point anomalies,
- contextual anomalies,
- collective anomalies.

The point anomalies are single data instances, which are separated from the rest of the data. The contextual (conditional [2]) anomalies are such data instances, which are considered

as anomalies in a specific context only. In the other context can be considered as the normal data. This data have two types of attributes - behavioral and contextual attributes. The behavioral attributes denote noncontextual characteristics of the data, while the contextual attributes denote the relation between the data. The collective anomalies occur when whole data subset is divided from the normal data and considered as anomalous. Either point anomaly, or collective anomaly can be transformed to the contextual anomaly.

B. Anomaly detection

There exists more than one approach in the anomaly detection. A straightforward approach defines a region, which represents the normal data or behavior and declares any data, which do not fall into this normal data as an anomaly. However, there exist some factors which are challenging [1]:

- It is problem to say what normal is. A boundary between normal and anomalous behavior is often not so precise.
- When the anomalies are result of a malicious software, this software can try to mask these anomalies as a normal traffic pattern.
- The normal traffic pattern is dynamic. What is normal now, might not be normal tomorrow.
- It is very difficult to get the normal labeled data for building a prediction model.
- The data can contain a noise which is not considered as the anomaly, but it is not interesting for analyst and therefore is unnecessary, because can cause false detection of the anomalies.

The anomaly detection methods are different for different application domains and specific problems related to them.

Data labels denote data as normal or anomalous. Labeling a data is made manually, so aquiring of a fully labeled data is very difficult and expensive. Based on a type of the labeled data, the anomaly detection techniques can be divided into three categories:

- supervised anomaly detection techniques,
- semi-supervised anomaly detection techniques,
- unsupervised anomaly detection techniques.

The techniques working in the supervised mode need a fully labeled data. The typical approach in this mode is to build the predictive model of normal and anomalous data. All tested data are then compared to these models and denoted as normal or anomalous.

The techniques working in semi-supervised mode need for building of the predictive models the normal labeled data only. A specific type of this mode are techniques, which work with the anomalous data only and build the model of the anomalous behavior.

The unsupervised mode techniques do not need the training data to be labeled because they assume that there is much more normal data instances than anomalous ones.

III. Point anomalies detection techniques

A. Classification based techniques

These techniques work in two phases. During the first phase, which is called *learning*, a prediction model (*classificator*) is built using avalaible labeled data. The classificator can distinct between the normal and the anomalous data. During the second, *testing* phase, the tested data are classified into the normal or the anomalous classes. In the learning phase can model divide the normal data into several sets. When this occurs, such technique will be called a multi-class technique. When only one normal class exists, it is an one-class technique.

One group of classification techniques uses a classification agorithm based on neural networks. Such techniques can be used with the multi-class or the one-class data. Other techniques use the algorithms based on *Bayesian networks*, *Support vector machines* or *Rule based systems*.

The testing phase is generally very fast, because a predictive model has been built and testing instances are only compared to the model. These techniques also use algorithms that can distinguish between instances belonging to different normal classes. A disadvantage is that these techniques need training labeled data to build the predictive model.

B. Nearest neighbor based techniques

These techniques are based on the prediction that normal data instances forms neighborhoods, while the anomalous do not.

These techniques compute distances to the nearest neighbors or uses a relative density as the anomaly score. The first group of techniques use a distance to the k nearest neighbors as the anomalous score. The second group of techniques computes the relative density in a hypersphere with the radius d. Such anomalous score s can be computed as [3][4]:

$$s = \frac{n}{\pi d^2} \tag{1}$$

where n is the number of data instances. The advantage of these techniques is that they can work in the unsupervised mode, but if the semi-supervised mode is used, the number of the false anomaly detections is smaller [5]. The computational complexity of such techniques is relatively high, because between each pair of the data instances the distance is computed. Also the rate of the false anomaly detection is high, if a normal neighborhood consists only from few data instances.

C. Clustering based techniques

These techniques are very similar to techniques mentioned in the previous subsection. The problem of detecting anomalies which forms clusters can be transformed to the problem of nearest neighbor based techniques. Both, nearest neighbor based and the clustering based techniques are very similar. The clustering based techniques, however, evaluate each instance with a respect of the cluster it belongs to.

The first type of clustering based techniques assume that normal instances forms clusters (Fig. 2). Such techniques apply known clustering based algorithms and declare, whether any data belongs to the cluster or not. A disadvantage is that they are optimized to find clusters not anomalies.

The second type assumes that the normal data instances lie close enough to a closest cluster centroid (Fig. 3). Such techniques are not applicable, if the anomalies form clusters. Therefore there exists the third type of the clustering techniques, which assumes that the normal instances form large dense clusters, while anomalies form small clusters which are sparse (Fig. 4). Both previous types work in two phases. In the first phase, a clustering algorithm clusters data and in the second phase, it computes a distance as an anomalous score. The clustering based techniques can work in the unsupervised

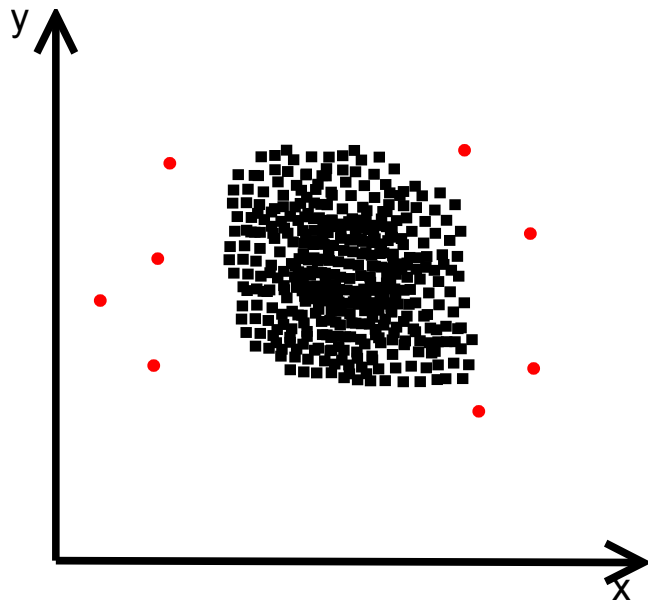


Fig. 2. Normal instances as a one big cluster.

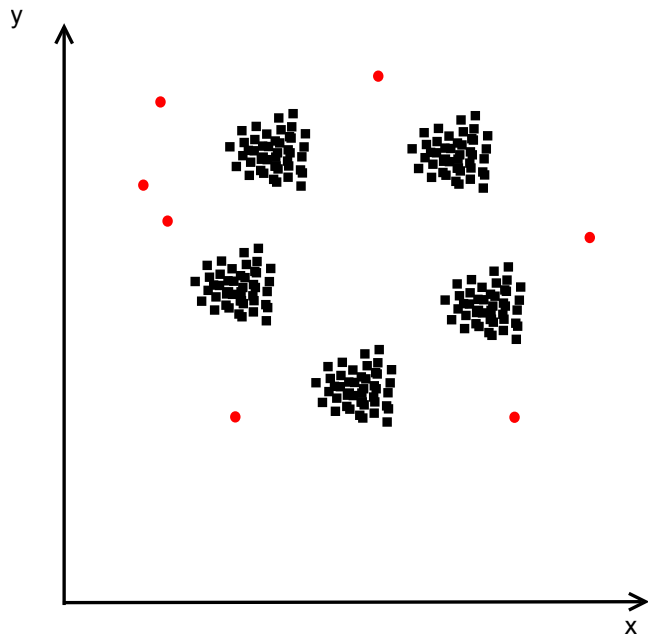


Fig. 3. Normal instances form more than one cluster.

mode, because clustering algorithms do not need labeled data. Once a model is built, the testing phase is very fast, because it just compares tested instances to the model. One main disadvantage is a high computational complexity and the fact, that these algorithms are not optimized to find anomalies.

D. Anomaly detection techniques based on statistical approach

The statistical methods are based on the following assumption [6]: *An anomaly is an observation which is suspected of being partially or wholly irrelevant because it is not generated by the stochastic model assumed*. It means that the normal data instances occur in the high probability regions, while the anomalies occur in the low probability regions of the stochastic model.

The statistical methods fit the statistical model to the normal data and then determines, if a tested data instance belongs to

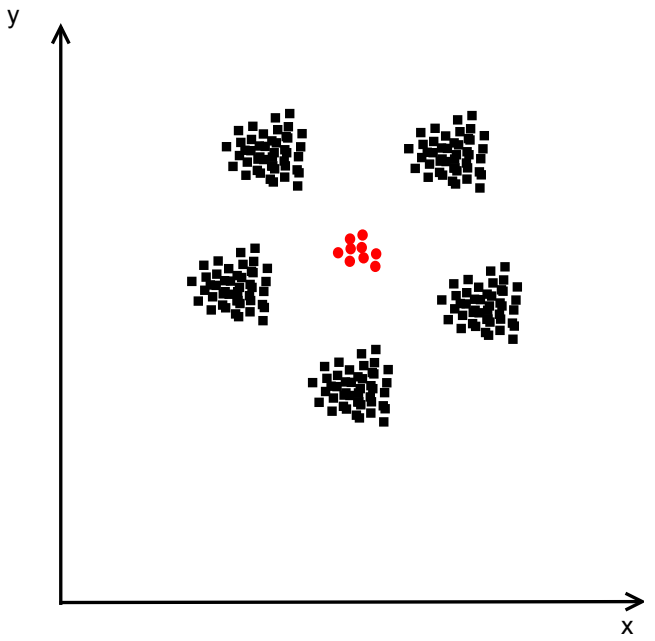


Fig. 4. Anomalous instances form a cluster.

the model or not. If technique assume the knowledge of the distribution it is called parametric [7], otherwise it is called nonparametric [8].

The nonparametric methods assume that model is determined from the given data. The most used techniques are the kernel function based and the histogram based techniques. The kernel function based techniques use *the Parzen windows estimation* [9]. The simplest techniques are the histogram based techniques. These techniques are widely used in intrusion detection systems. The first step in using of these techniques consists of building a histogram based on diferent values taken from the training data. In the second step is a tested data instance checked, whether it falls into one of the histogram bins.

The parametrical methods assume that data are generated by the parametric distribution Θ and the probability density function $f\,(o, \Theta)$, where o is an observation. The parametrical methods can be divided based on types of the distributions:

- Gaussian model based,
- regression model based,
- mixture of parametric distributions based.

The Gaussian model based methods use many known techniques like *box plot rule* or *Grubb's test*. In the Grubb's test for each test instance x, its z score computed as:

$$z = \frac{|x - \overline{x}|}{s} \tag{2}$$

where $\overline{x}$ is the mean and s is the standard deviation. The test instance is anomalous if

$$z > \frac{N-1}{\sqrt{N}} \sqrt{\frac{t^2_{\alpha/(2N),N-2}}{N-2+t^2_{\alpha/(2N),N-2}}} \tag{3}$$

where N is the data size and $t^2_{\alpha/(2N),N-2}$ is a treshold.

The regression based methods were used for time-series data. Techniques based on mixtures of the parametric distribution use different types of the distribution to model normal and anomalous data. If the normal data cannot be modeled by none of all known distributions, mixture of them is used.

The advantages of the statistical methods are that they are widely used and if a good model is designed, they are very effective. They can be used in the unsupervised mode with lack of the training data. The histogram based techniques are not suitable for detecting the contextual anomalies, because they cannot record an interaction between the data instances. Also choosing of the proper test method is nontrivial.

E. Other detection techniques

The aformentioned techniques are the most widely used. The other methods use the information theory techniques based on the *relative entropy* or the *Kolmogorov complexity*. These techniques can operate in the unsupervised mode and do not need a statistical assumption about data. Another techniques are the spectral anomaly detection techniques, which try to find an approximation of the data and determine the subspaces in which the anomalous instances can be easily identified. These techniques have high computational complexity.

IV. Contextual anomalies detection techniques

When detecting contextual anomalies, data instances have contextual and behavioral attributes. The context between data can be defined using sequences, space, graph or profile. The profiling is typicaly used for detecting of credit card frauds. For each of the credit card holders (each holder denotes separate context) is the behavioral profile built. Using the credit card for paying abroad can be labeled as the anomalous or the normal instance. It depends on the context and that is the card owner.

The problem of contextual anomalies detection can be transformed to the problem of point anomalies detection. It is necessary to identify the context and then compute an anomaly score. The other methods utilize a structure of data and use regression or divide and conquer approach. The advantage of these techniques is that they can identify the anomaly, which would be undetectable using the techniques in the previous section.

V. Collective anomalies detection techniques

These techniques can be divided into three categories:

- sequential anomaly detection techniques
- spatial anomaly detection techniques,
- graph anomaly detection techniques.

The sequential anomaly detection techniques work with sequential data and try to find the anomalous subsequencies. Such data can be system call data or biological data. The problem of detecting sequence anomalies can be reduced to the point anomaly detection problem [10][11].

Handling spatial anomalies includes finding subcomponents in the data. There exists few tehniques in this category. The image processing techniques using the *Markov fields* is one of them [12].

The graph anomaly detection techniques involve finding a anomaly subgraph in a large graph. The size of the subgraph is also taken into the consideration [13].

VI. Conclusion

In the previous sections we presented the most used anomaly detection techniques and defined what anomalies are. For our application domain, which is the traffic engineering, are suitable techniques which can work in the unsupervised mode, because it is hard to get fully labeled data which would cover all possible traffic in the computer network. Such techniques are using neural networks, statistical mathematical model or Bayesian networks.

Our future work should include design of a distributed system, which will collect data from the computer network and build a model of a normal traffic behavior. This distributed system will also detect anomalies using the unsupervised anomaly detection techniques. This system will react on the detected anomalies and reconfigure network.

Acknowledgment

The authors want to thank the entire staff of Computer Networks Laboratory at DCI FEEI at Technical University of Košice.

This publication is the result of the project implementation Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF & was partially prepared within the project "Methods of multimedia information effective transmission", No. 1/0525/08 with the support of VEGA agency.

References

[1] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," University of Minnesota, Tech. Rep., August 2007.

[2] X. Song, M. Wu, C. Jermaine, and S. Ranka, "Conditional anomaly detection," *IEEE Transactions of Knowledge and Data Engineering*, 2007.

[3] E. M. Knorr and R. T. Ng, "A unified approach for mining outliers," *Proceedings of the 1997 conference of the Centre for Advanced Studies on Collaborative research*, 1997.

[4] ——, "Algorithms for mining distance-based outliers in large datasets," *Proceedings of the 24th International Conference on Very Large Data Bases*, pp. 392–403, 1998.

[5] D. Pokrajac, A. Lazarevic, and L. J. Latecki, "Incremental local outlier detection for data streams," *Proceedings of IEEE Symposium on Computitional Intelligence and Data Mining*, 2007.

[6] F. J. Anscombe and I. Guttman, *Rejection of outliers*. Technometrics, 1960.

[7] E. Eskin, "Anomaly detection over noisy data using learned probability distributions," *Proceedings of the Seventeenth International Conference on Machine Learning*, pp. 255–262, 2000.

[8] M. Deforges, P. Jacob, and J. Cooper, "Applications of probability density estimation to the detection of abnormal conditions in engineering," *Proceedings of Institute of Mechanical Engineers*, vol. 212, pp. 687–703, 1998.

[9] E. Parzen, *On the estimation of a probability density function and mode*. Institute of Mathematical Statistics, 1962, no. 2.

[10] P. K. Chan and M. V. Mahoney, "Modeling multiple time series for anomaly detection," *Proceedings of the Fifth IEEE International Conference on Data Mining*, 2005.

[11] S. Budalakoti, A. Srivastava, A. Akella, and E. Turkov, "Anomaly detection in large sets of high-dimensional symbol sequences," NASA Ames Research Center, Tech. Rep., 2006.

[12] G. G. Hazel, "Multivariate gaussian mrf for multispectral scene segmentation and anomaly detection," *GeoRS*, vol. 3, pp. 1199–1211, 2000.

[13] C. C. Noble and D. J. Cook, "Graph-based anomaly detection," *Proceedings of the 9th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2003.

Clustering of Users Behaviour in IEC Font Design

[1]*Miron KUZMA,* [2]*Tomáš REIFF,* [3]*Zlatko FEDOR*

[1,2,3]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]miron.kuzma@tuke.sk, [2]tomas.reiff@tuke.sk, [3]zlatko.fedor@tuke.sk

***Abstract*—Clustering of users behaviour in IEC systems allows to build larger datasets, while not destroying users's individual characteristics. If we group similar users together, their collective behaviour is preserved. Within such a group of users we collect more data than from single individual user. Clustering algorithm is used to find, which users behaviuor is similar. In this paper we present application of users behaviour clustering in IEC font desing. We evaluated impact of clustering on behaviuor modelling using ten users data from computer modern latex font interactive design.**

I. Introduction

Clustering of users inputs in Interactive Evolutionary Computation (IEC) allows to build a collection of data sets with different users behaviors - preferences - approaches. In order to model users behavior in IEC we collect the pairs of visualized system data and corresponding user's evaluations. The record from single session of single user might be not big enough to allow to build sufficient model of user. But, if we try to collect data from several sessions and several users the problem of incompatibility of data will arise. The subjective character of IEC tasks leads to situations with two users having opposite opinions to the same presented information. Such conflicting data pose a problem for modeling tools based on function approximation techniques e.g. Neural Networks.

By clustering of multi-user data from multiple IEC sessions, we build groups of corresponding behaviors - preferences - approaches related to some particular task. These groups do not necessarily correspond to separate users, instead they might represent separate moods of users, or separate branches of solving given task. However, these groups should consist of less conflicting data then the whole collection of all data will be. The approach of clustering first and modeling later was first presented in NARA [14] and Mixture of Experts [9] or Counterpropagation [8] methods in the Neural Networks domain. Motivation was to divide the task into several simpler ones. Our motivation in IEC user modeling is to face conflicts in data resulting from inherent subjectivity of problem. Related to this scenario are the works on fuzzy rules approximation in KANSEI domain [10], [11].

We will study the clustering problem on the task of IEC design of fonts. Here, IEC interface is used to present alternative typesetting fonts to the user, and guided by his evaluations of these fonts, to iteratively search for the esthetically optimal font. This task was previously studied in [7], [12], [13].

According to [7] The reason to synthetize handwriting and to create fonts is that the synthetized handwrited characters enable us to personalize font for the user. It shows one's personal handwriting style.

The user is able to set the style of various documents according to his personal font, e.g. writing an e-mail, by using chat program, writing a blog and by other activities. From the recipient point of view, when he gets an e-mail or message written using personal font, he might feel closer to the sender and vice versa. We can say handwriting or "personal font" adds a feeling of personal touch [7].

One of the future directions of computational intelligence is humanized computational intelligence. One of such technologies is Interactive Evolutionary Computation (IEC). The term we explain in the following part of the paper. As we will see this research domain is famous with many of its successfull applications, the field of its potential application is wide.

The article published by Takagi in 2002 [1] gives a survey of the Interactive Evolutionary Computation (IEC). There exists a large variety of systems using IEC, eg. [2], [3], [4], [5], [6].

IEC is commonly used in artistic field, engineering field, and other fields. The research categories are: graphic art and computer generated animation, 3D computer generated lightning desing, music, editorila design, industrial design, face image generation, speech processing, hearing aids fitting, virtual reality, database retrieval, data mining, image processing, control and robotics, internet, food industry, geophysics, art education, writing education, games and therapy, social system. Another topic is the research of user interface. It focuses on human fatigue and tries to reduce its unwanted impact.

Interactive Evolutionary Computation (IEC) is a technique that involves evolutionary computation consisting of genetic algorithms (GA), evolutionary strategy (ES), evolutionary programming (EP), and genetic programming (GP). It aims to optimize the target system based on human subjective evaluation. Regular optimization methods can be used if the specifications or design goal of the target system is numerically given. However, there are many cases that the system performance is not measurable and only human can evaluate the system performance, for example, maximizing sound quality of a hearing aid for a user, generating computer graphics for my living room, generating Jazz-like music. Subjective evaluation includes both KANSEI scale such as preference and evaluation based on domain knowledge [1].

The Interactive Evolutionary Computation (IEC) as an optimization method involves Evolutionary Computation (EC). It is a method that uses subjective human evaluation. It is an EC technique thats fitness function is replaced by a human user, because we cannot provide the fitness function for the system.

Fig. 1 shows a general IEC system where the system output is shown to the user and user evaluates system outputs. The EC optimizes the target system to obtain the preferred output based on the user's subjective evaluation. The IEC technology embeds in the target system following: human preference, intuition, emotion, psychological aspects. We call these using

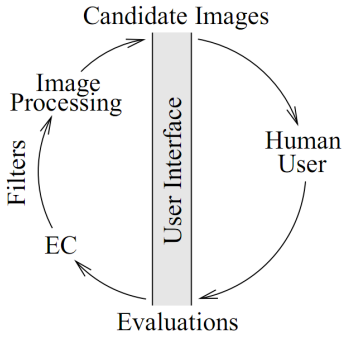


Fig. 1. General IEC System. [2]

a more general term KANSEI.

II. IEC Font Design

Our IEC approach focused on designing and implementing system that is able to help user to create a font, reduce the time needed for this process or give a basic idea of font to start with [13].

The Idea of the system complies to IEC basics. On Fig. 2. we have a user interface that is split into two main parts: the *"Control Panel"* and the *"Samples Panel"*.

The samples panel on the right side has 12 Font samples with marking buttons having marks from 1 - the worst - to 5 - the best - on them. The marks have influence on the evolution process. User has to evaluate the Font samples, if he does not evaluate the font samples, they are given the default mark of 0. User has to click on the mark that corresponds to his own preferences and intentions to evaluate the displayed samples. The next step of the evolution process is to consider the preferable action and click on the corresponding button.

We ran experiments with the system to justify its usability among users. Experiments were compared to the manual font design taking into account time, user-friendlyness, result - the designed font.

III. Data structure used in the system

The Font Evolving System uses as a base Donald Knuth's Computer Modern Font sized at $10pt$. The font configuration file, that is actually a metafont file, contains 62 variables/parameters. We chose from them 21 parameters that according to us, have major impact on the final font look [13].

We ran an experiment with 10 users to obtain data for further work. Users had to design a font they like. We recorded the whole session to have as much data as possible. For our further experiments the dimension number increased from 21 to 49. We collected 2112 font samples altogether during the testing sessions. The collected samples we will use in the eperimental part.

IV. Clustering of User Inputs

A. The Experiments Description

We will do two experiments. We have 2112 font samples collected from 10 users. We built two training sets. The first included all of the 49 parameters, and the second training set excluded the time variables from the first training set, to determine the influence of the time variables. The dimension of second training set was 46.

The first experiment is the clustering using SOM - Kohonen network. We use three kohonen networks with different mesh sizes. The smallest mesh, which is sized at $3x3$ units. The second mesh is sized at $15x15$ units and the largest mesh is sized at $60x60$ units. The largest mesh size has 3600 units altogether, so the number of all samples is smaller than the largest mesh size. Our implementation of kohonen network uses by default these parameters:

- h - adaptation height,
- R - radius of influence,
- γ - learning parameter,
- c - number of learning cycles.

The parameters had the same values for all meshes, excluding the parameter R, which was a function:

$$R = \frac{1}{3} mesh\_size\_x \tag{1}$$

We trained networks on both trainig sets and obtained a map of clusters for every pair of mesh and trainig set. We also obtained new sets of data where the number of cluster is included by every font sample.

The second experiment is based on feed-forward neural networks. We have training set of 2112 font samples obtained from the experiment one using the mesh size $3x3$ units. It is certain that we know to which cluster every font sample belongs. The topology of the neural network is:

- input units(48) - all variables from the training set excluding the cluster number and user evaluation(class),
- hidden units(3)
- output unit(1) - the user evaluation(class).

We created 10 neural networks. The first neural network used all of the samples in the training set. The remaining neural networks were created for every cluster, that way the training sample for every network was different according to the cluster distribution. The parameters of neural networks were:

- c - number of learning cycles, set to 5000
- $< -1, 1 >$ - interval for random weights initialization
- η - learning parameter, set to 0.2
- d_{max} - the maximum difference between a teaching value and an output unit which is tolerated, set to 0.05

The parameters had the same value for every neural network. The idea of the experiment is to determine the usability of the clustering.

B. The Experiments Results

We have done the clutering experiments with SOM maps. The results of the clustering are on the Fig. 3, 4, 5 and 6. On the Fig. 5 we can see the distribution of samples among the clusters. The cluster number was set to 3600, according to the $60x60$ units grid. The result is that users of the IEC System are displayed on the map as a compact set of points. We could say, every user when working with the system had his own intention. The observation for the $15x15$ units grid is similar to the previous observation - the samples from the same user are belong to the same group of cluster creating compact set of points on the kohonen map.

We have done also experiments to determine the presence of the time variable. We can say according to the comparative results on Fig. 3 and 4, 5 and 6 the presence or absence of the time variable does not have an impact on the clustering results.

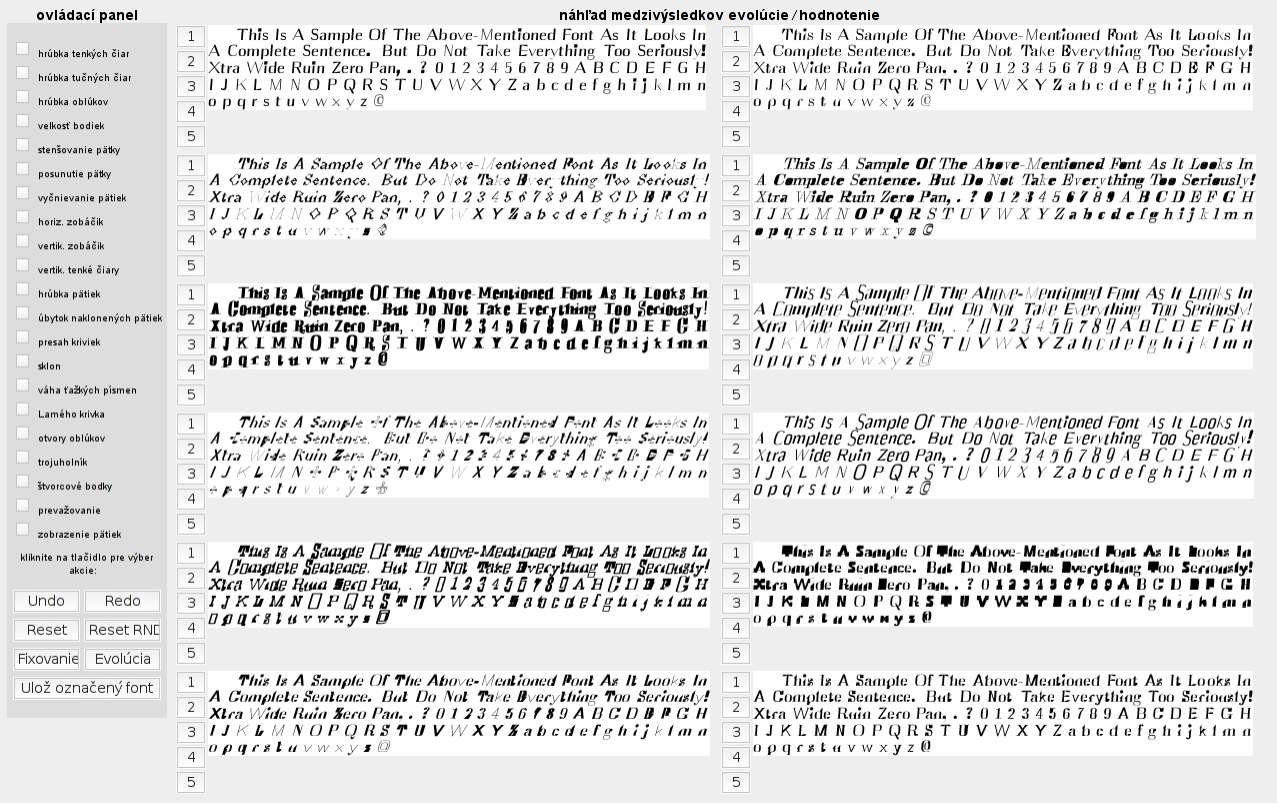


Fig. 2. Font Evolving System: User Interface. The *"Control Panel"* with actions on the left side and *"Samples Panel"* with Font Samples and their marking buttons on the right side. Evaluating samples on the right side and selecting the action from the control panel is the usual order of workflow by IEC programs.

We have done the next experiment after the clustering using the data obtained from the kohonen map. Here we wanted to observer the neural network leaning process using 10 pattern sets. We comapred the set with all patterns and the sets obtained from the clustering using $3x3$ units map and randomly generated 9 training sets from al the samples. We can see the results from this experiment on Fig. 7 and 8. We have observed the mean square error(MSE) by learning process of the neural network and we can say on Fig. 7 the training sets obtained from the clustering process had smaller MSE than the training set consisting of all the samples.

Fig. 7 and 8 illustrate the effect of clustering on the final user modelling. Fig. 7 shows training error on clustered data. Fig. 11 shows the results on randomly partitioned data in 9 disjunctive sets. Actually our results show no improvement by clustering yet. In future we will perform similar comparison but using more representative testing data(instead of training set). We also plan to incorporate user modelling output directly into user interface of font design application and directly subjectively evaluate impact of user modelling.

V. Conclusion

Clustering of user inputs in multi-user Interactive Evolutionary Computation shows potentila utility by visual evaluation of clustering results in our application. However, we did not obtain positive results yet with the application of clustering results into user modelling. In future we want further improve the evaluation of this user modeling impact, altohough there is also room for improvement with the clustering itself. In this paper we also want to emphasize the possibility of IEC application in font design and our application with LaTeX fonts. The downloadable package of the Font Evolving System development version resides on the main author's web page.

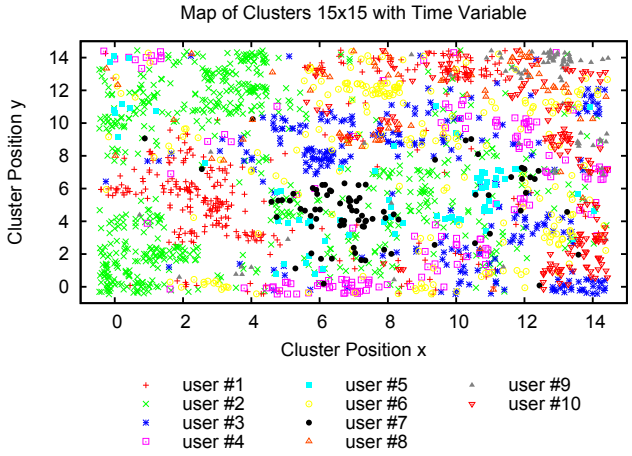


Fig. 3. Kohonen Map $15x15$ w/Time.

References

[1] H. Takagi, Interactive Evolutionary Computing: Fusion of the Capacities of EC Optimization and Human Evaluation *Proc. of 7th Workshop on Evaluation of Heart and Mind*, KitaKyushu, Fukuoka, (November 8-9, 2002)(in Japanese)

[2] JAKŠA, Rudolf; TAKAGI Hideyuki, Tuning of Image Parameters by Interactive Evolutionary Computation *Proc. of 2003 IEEE International Conference on Systems, Man & Cybernetics*, (SMC2003), Washington D.C., (October 5-8, 2003)

[3] JAKŠA, Rudolf; TAKAGI Hideyuki; NAKANO Shota, Image Filter Design with Interactive Evolutionary Computation *Proc. of the IEEE International Conference on Computational Cybernetics*, (ICCC2003), ISBN 963 7154 175, Siofok, Hungary

[4] NEUPAUER, Marek, Analysis of Medical Data using Interactive Evolutionary Computation *Master's Thesis*, Košice, Technical University of Košice, Faculty of Electrical Engineering and Informatics, Department of Cybernetics and Artificial Intelligence

[5] KOVÁČ, Július, Image Database Search Using Self-Organizing Maps and Multi-scale Representation *Master's Thesis*, Košice, Technical University

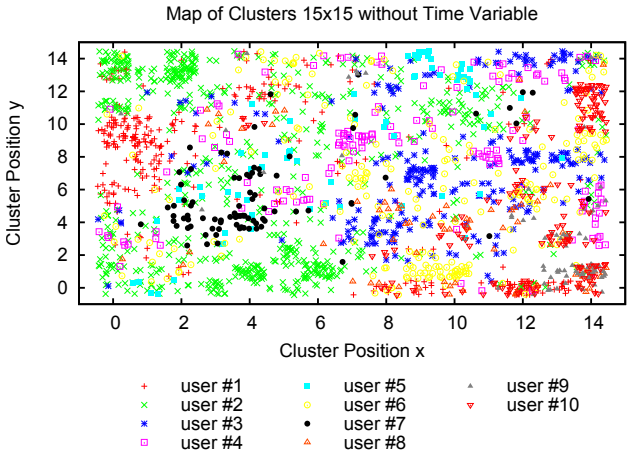


Fig. 4. Kohonen Map $15x15$ w/o Time.

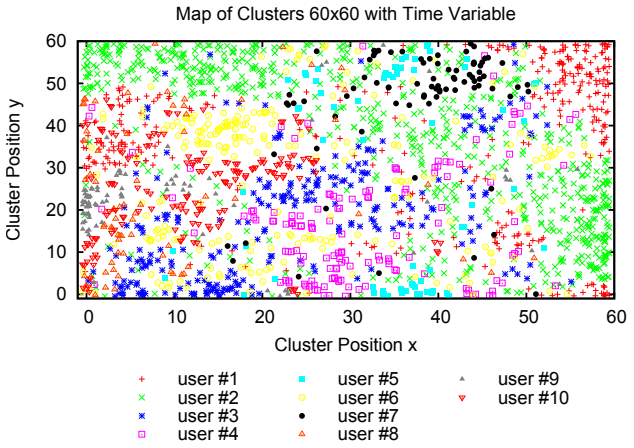


Fig. 5. Kohonen Map $60x60$ w/Time.

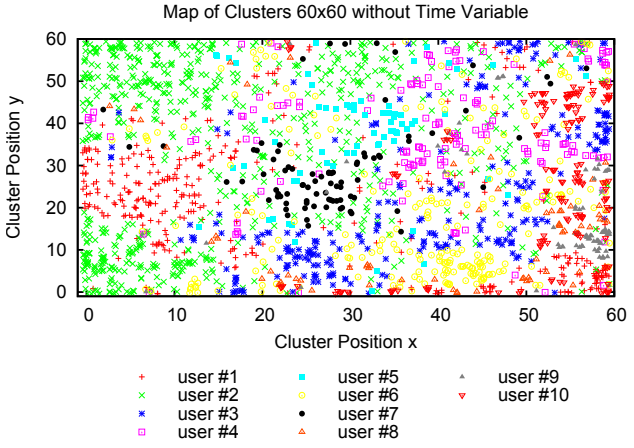


Fig. 6. Kohonen Map $60x60$ w/o Time.

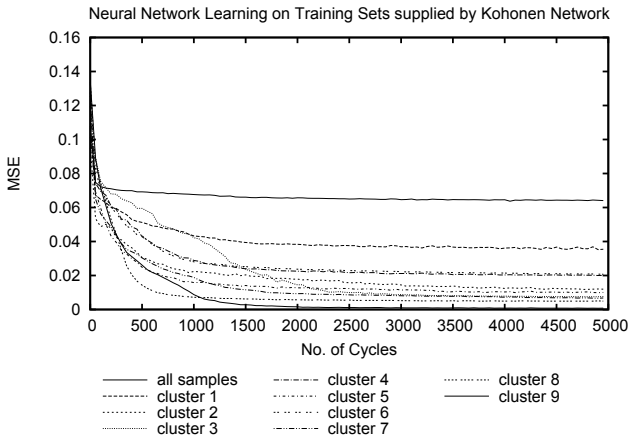


Fig. 7. Neural Network Learning Graph Using Training Set Supplied from Kohonen Network.

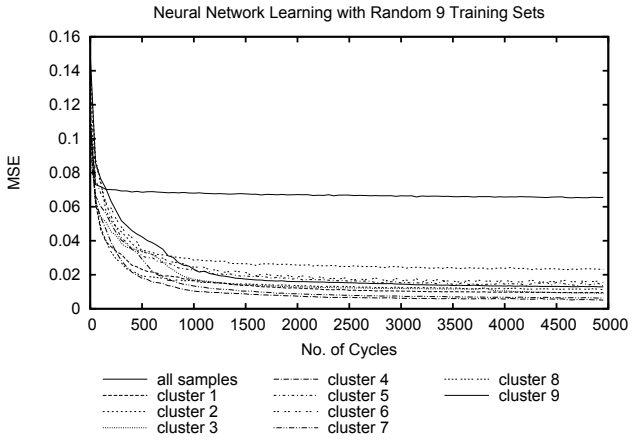


Fig. 8. Neural Network Learning Graph Using 9 Random Training Sets.

of Košice, Faculty of Electrical Engineering and Informatics, Department of Cybernetics and Artificial Intelligence

[6] PANGRÁC, Ľubomír, Interactive Evolutionary Computation for Satellite Image Processing *Master's Thesis*, Košice, Technical University of Košice, Faculty of Electrical Engineering and Informatics, Department of Cybernetics and Artificial Intelligence

[7] J. Dolinský, and H. Takagi, Synthesizing Handwritten Characters using Naturalness Learning *Proceedings of ICCC 2007*, 2007.

[8] R. Hecht-Nielsen, Counterpropagation networks, *Applied Optics*, Vol. 26, Issue 23, 1987, pp. 4979-4983.

[9] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, G. E. Hinton, Adaptive Mixtures of Local Experts, *Neural Computation*, No.3, 1991, pp. 79-87.

[10] Sz. Kovács, N. Kubota, K. Fujii and L.T. Kóczy, Behaviour based techniques in user adaptive Kansei technology, *Proceedings of the VSMM2000, 6th International Conference on Virtual Systems and Multimedia*, Ogaki, Japan, October 2000, pp.362-369.

[11] Sz. Kovács, Fuzzy Reasoning and Fuzzy Automata in User Adaptive Emotional and Information Retrieval Systems, *Proceedings of the 2002 IEEE International Conference on Systems, Man and Cybernetics*, Hammamet, Tunisia, October 2002.

[12] M. Kuzma, R. Jakša, and P. Sinčák, Computational Intelligence in Font Design, *Computational Intelligence and Informatics: Proceedings of the 9th International Symposium of Hungarian Researchers*, Budapest, November 2008, pp.193-203.

[13] M. Kuzma, Interactive Evolution of Fonts, *masters thesis*, Technical University of Košice, 2008.

[14] H. Takagi, N. Suzuki, T. Koda, Y. Kojima, Neural Networks Designed on Approximate Reasoning Architecture and Their Applications, *IEEE Transactions on Neural Networks*, Vol.3, No.5, September 1992, pp. 752-760.

Knowledge about Software Design Patterns in Software Architecture

M. Lakatoš*
* Technical University of Košice/Department of Computers and Informatics, Košice, Slovakia

Lakatos_matej@hotmail.com

Abstract— **During more and more growing up of complexity of software systems also raise sights on easy handle and understand-ability of whole system, management, maintenance and modification of software system. All efforts of this time are focused to connect knowledge about software system from analyst models, with implementation these models to a run code. For this reason is necessary to better understating between analysts and programmers point of views. Programmers are trying to get their written code to the more abstract level. One solution how to reach that is to use patterns methods techniques which at the last couple years have become more and more used. They are also representing by UML notations. On another side the UML is most of used language for keeping analyst ideas in models. This paper describes the application of software design patterns and illustrates on of these design patterns in UML notation. At the first there is describing of essential sort of software designs patterns. In the next part, there are mentions about most widely used software design patterns. In the last part there is a mention about MDA technology which can cover this knowledge about software architecture thought design patterns methods.**

Keywords—**design patterns, architectural patterns, knowledge, software architecture, Model Driven Architecture, software system.**

I. INTRODUCTION

An architecture design is most important in software engineering processes. Regardless of the variations in software engineering processes, software architecture provides the skeleton and constraints for software implementation. Software engineering is an iterative process that comprises multiple stages, including modeling, design, implementation, deployment, and maintenance. To finish each stages with good success rate depend on well understanding of clients requirements on a system and also well understand among analysts and programmers. Pfleeger[1] says that common reasons of break down software projects are on 13,1% uncompleted or no understandable requests, on 9,3% insufficient support from management side of suppliers, 8,7% change of requests and specifications, 8,1% no successful planning and so on. Software development in all stages requires knowledge about the system and also knowledge how to implement requests into the system. Well implementation is strongly depending on programmer's skills, knowledge of reading an analyst's requests and programmer's techniques. Design patterns methods are giving needed techniques to analytics and programmers how parts of developed system should be programmed. Design patterns are independent on program languages. They offer methods how to solve common problems, but no concrete implementation in specific programmer language. Design patterns is possible describes by UML notation, too and so analysts can during developing, or maintenance of a system to write how a programmer should implement their requests. Design patterns are something like a new language between analytics and programmers with formal notation. Model Driven Architecture (MDA) can help to better transforming analytics ideas written by design patterns to concrete platform. Is possible that programmers will need add next specification to a system so if a programmer use design patterns with good descriptions then MDA can automotive transform a programmer idea to model(s) and an analytic can better understand what a programmer made and what impact it will have on a developed system. This paper presents sort of design patterns and describes software problems which can these design patterns to solve and there is an example illustration of abstract pattern described by UML notation. The paper also points out on concept of MDA, exist specifications for MDA and tools which are based on UML for MDA concept.

II. SOFTWARE DESIGN PATTERNS

In software engineering, a design pattern[3] is a general reusable solution to a commonly occurring problem in software design. It is a description, or template for how to solve a problem that can be used in many different situations. Object-oriented design patterns typically show relationships and interactions between classes or objects, without specifying the final application classes or objects that are involved.

We can assume that ancestor of design patterns were objects. Patterns make them good with OOP community where experts-developers needed again and again apply same code steps. After time design patterns become to apply also in design, analyze and today we can to see patterns technique in different kind of and in more domain areas.

At a higher level there are Architectural patterns[2] that are larger in scope, usually describing an overall pattern followed by an entire system.

The software pattern is helping to create OOP design, because they do identify classes, instances, their relationships, responsibility to solve concrete technical problems so they help to speed up the development process by providing tested, proven development paradigms. Design patterns help to prevent small issues that can cause major problems, and it also

improves code readability for programmers and analytics that are familiar with the patterns. A team of Gang of Four[2] in their publication which is regards as bible of design patterns explains when, how and which concrete pattern to use. Design patterns have also problems. One of them is dispersal of a pattern in a developed software system. There isn't still an acceptable way to get pattern to graphic form after a programmer use him. Also there is problem to identify used pattern in source code in developed software.

The documentation for a design pattern describes the context in which the pattern is used, and the suggested solution. Each design patterns should to be describes by 13 sections (e.g. Pattern Name and Classification, Intent, Motivation (Forces), Structure, etc.)

A. *Design Patterns*

Design patterns were originally grouped into the categories Creational patterns, Structural patterns, and Behavioral patterns[2].

Creational patterns work with object creation mechanisms, trying to create objects in a manner suitable to the situation. The basic form of object creation could result in design problems or added complexity to the design. Creational design patterns solve this problem by somehow controlling this object creation.

Structural patterns ease the design by identifying a simple way to realize relationships between entities.

Behavioral patterns identify common communication patterns between objects and realize these patterns. By doing so, these patterns increase flexibility in carrying out this communication.

Some of Creational patterns are[2]:

- *Abstract factory pattern*, centralize decision of what factory to instantiate,
- *Factory method pattern*, centralize creation of an object of a specific type choosing one of several implementations,
- *Builder pattern*, separate the construction of a complex object from its representation so that the same construction process can create different representations,
- *Prototype pattern*, used when the type of objects to create is determined by a prototypical instance, which is cloned to produce new objects,
- *Singleton pattern*, restrict instantiation of a class to one object,
- Etc.

Some of Structural patterns are[2]:

- *Adapter pattern*, 'adapts' one interface for a class into one that a client expects,
- *Aggregate pattern*, a version of the Composite pattern with methods for aggregation of children,
- *Bridge pattern*, decouple an abstraction from its implementation so that the two can vary independently,
- *Composite pattern*, a tree structure of objects where every object has the same interface,
- *Decorator pattern* add additional functionality to a class at runtime where subclassing would result in an exponential rise of new classes,
- *Facade pattern*, create a simplified interface of an existing interface to ease usage for common tasks
- *Flyweight pattern*, a high quantity of objects share a common properties object to save space,
- Etc.

Some of Behavioral patterns are:

- *Chain of responsibility pattern*, Command objects are handled or passed on to other objects by logic-containing processing objects,
- *Command pattern*, Command objects encapsulate an action and its parameters,
- *Interpreter pattern*, Implement a specialized computer language to rapidly solve a specific set of problems,
- *Iterator pattern*, Iterators are used to access the elements of an aggregate object sequentially without exposing its underlying representation,
- *Observer pattern* Publish/Subscribe or Event Listener. Objects register to observe an event which may be raised by another object,
- *State pattern*, A clean way for an object to partially change its type at runtime,
- *Strategy pattern*, Algorithms can be selected on the fly,
- Etc.

An example illustration of graphics description of Abstract Factory design pattern.

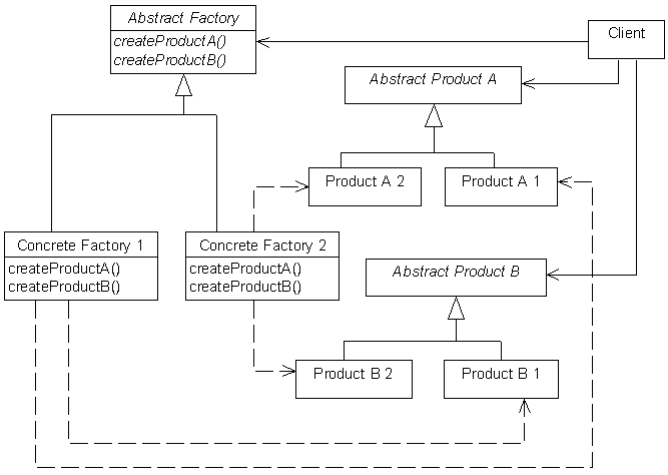


Figure 1. Abstract Factory pattern model in UML notation.

B. *Architectural patterns*[3]

Describe solutions for problems for architectural level. It gives description of the elements and relation type together with a set of constraints on how they may be used. In comparison to design patterns, architectural patterns are larger in scale. One of the most important aspects of architectural patterns is that they offer different quality attributes.

Some of architectural patterns are:

- *Layers*, a group of classes that have the same set of link-time module dependencies to other modules,
- *Model-View-Controller, Presentation-abstraction-control, Model View Presenter and Model View*

ViewModel, isolates "domain logic" (the application logic for the user) from input and presentation (GUI),
- *Multitier architecture (often three-tier)* - is a client-server architecture in which the presentation, the application processing, and the data management are logically separate processes,
- *Service-oriented architecture*, deployed SOA-based architecture will provide a loosely-integrated suite of services that can be used within multiple business domains.

III. MODEL DRIVEN ARCHITECTURE

Many organizations have begun to focus attention on Model Driven Architecture[4][5][6] (MDA) as an approach to application design and implementation. It provides a set of guidelines for the structuring of specifications, which are expressed as models and it supports reuse of best practices when creating families of systems.

The MDA[5] (Model Driven Architecture) technology is provided from OMG (Object Management Group). This group is focusing on making standards, which offers interoperability and portability of distributed OOP applications. Concept of MDA covers large part of exist specifications of OMG:
- UML (Unified Modeling Language),
- MOF (Meta-Object Facility),
- CWM(Common Warehouse Metamodel),
- XML (Extensible Markup Language),
- XMI (XML Metadata Interchange) a IDL (Interface Definition Language).

As defined by the Object Management Group (OMG), MDA is a way to organize and manage enterprise architectures supported by automated tools and services for both defining the models and facilitating transformations between different model types.

An idea of MDA is progressive specifying of models from higher layer of abstraction, which does include models of users without any relations to their implementations to lower layer which contains models directly mapped to source code.

Some authors call the UML in MDA as "UML as programming language", but just UML is not fully-fledged programming language.

Martin Fowler[6] defines three bases way of using UML:
- UML as sketch – UML is used only for catch of main ideas for developing software. This is main use of UML today,
- UML as blueprint – is trying to describe whole system more detailed.
- Automated transformations of models to source code. These models have to contain too much informations for successfully transformations from models to final source code,
- UML as programming language – system is completed described by models and these diagrams became of run code.

Next advantage of MDA is possibility to divide business logic from technology of platform. Result is that an application made with MDA concepts are simply implemented in large scale of platforms (CORBA, J2EE, NET…). MDA allows develop an application on higher level of abstraction and wherefore the MDA left more time to focus on business logic instead of spend a time for problems with implementation on a concrete platform.

OMG define four principles of MDA:
- Models expressed in a well-defined notation are important to understanding of systems
- The building of systems can be organized around a set of models by imposing a series of transformations between models, organized into an architectural framework of layers and transformations.
- A formal underpinning for describing models in a set of metamodels facilitates meaningful integration and transformation among models, and is the basis for automation through tools.
- Acceptance and broad adoption of this model-based approach requires industry standards to provide openness to consumers, and foster competition among vendors.

To support these principles, the OMG has defined a specific set of layers and transformations. OMG identifies four types of models: Computation Independent Model (CIM), Platform Independent Model (PIM), Platform Specific Model (PSM) described by a Platform Model (PM), and an Implementation Specific Model (ISM).

Computation Independent Model (CIM)[6]
A computation independent model is a view of a system from the computation independent viewpoint. A CIM does not show details of the structure of systems. A CIM is sometimes called a domain model and a vocabulary that is familiar to the practitioners of the domain in question is used in its specification

Platform Independent Model (PIM)[6]
A platform independent model is a view of a system from the platform independent viewpoint. A PIM exhibits a specified degree of platform independence so as to be suitable for use with a number of different platforms of similar type.

Platform Specific Model (PSM)[6]
A platform specific model is a view of a system from the platform specific viewpoint. A PSM combines the specifications in the PIM with the details that specify how that system uses a particular type of platform. In other words: the PSM is a more detailed version of a PIM. Platform specific elements are added. When defining a PSM a target Platform Model has to be available.

Platform Mode (PM)[6]
A platform model provides a set of technical concepts, representing the different kinds of parts that make up a platform and the services provided by that platform. It also provides, for use in a platform specific model, concepts representing the different kinds of elements to be used in specifying the use of the platform by an application.

On the other side the MDA technology has also opponents-critics. Martin Fowler[6] said that UML did arise from "sketch" notation (as tools for capture important ideas and communication with programmer), but for use in the MDA is not the UML ready to use, yet.

Also he says that design of sequence and activity diagrams is not better then to write code with modern programming language.
Steve Cook[7] compares the MDA with a DSM (Domain Specific Modeling). DSM is next of possible approach to a develop software, where main artifact is a model. DSM doesn't try about automatic transform of models, instate of is based on creation of a model for each part of a system (domain) especially and after that verify or approve these models mutually.

Tools working on UML base with MDA implementation[8]:

- AndroMDA –open-source product. Is using with another tools (ArgoUML, MagicDraw, Maven). Allows to write own transformational scripts also in JAVA, QVT, ALT,
- Enterprise Architect (Sparx Systems), commercial product. Complete tools for designing, selecting of requests, etc. Covers all 13th of UML2.1 models.
- OptimalJ - commercial product. Contains model of processes based on activity diagrams in UML2.0. Supports Eclipse platform,
- Borland Together 2006 - commercial product. Supports specification of (QVT) Query View Transformation, which allows run transfers among models. Supports also OCL2.0,
- PowerDesigner9 – commercial product. Supports 8th implementations of MDA techniques,
- Rational Rose – commercial product. Supports couple implementations of MDA techniques.

IV. Conclusion

In this paper, I presented software design patterns as cornerstone for design a software system. Software design patterns can describe almost all clients, analyst's requirements on a developing system with UML notation. This approach brings up some standard-unification steps how to design software system in same way, what gives a better understandability between architect ideas models of software system and code-developer teams. UML notation gives a possibility to use an activity diagram for tracking branches of software system. With better descriptions of used design patterns is better possibility of maintenance, design, analyze of software system. The paper also analyzed MDA which can represent knowledge about application domain and better to transform new requirements to analytic model(s), or to lower levels on MDA concept.

The future works will focus on possibilities of integrating the knowledge represented by software design patterns into software architecture what to improve some necessary changes concerning the maintenance process or reconfigure itself according to requirements.

Acknowledgment

This work was supported by VEGA Grant No. 1/0350/08 Knowledge-Based Software Life Cycle and Architectures.

This work is the result of the project implementation: Centre of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] S. L. Pfleeger, Design and analysis in software engineering: the language of case studies and formal experiments, in Plastics, New York, NY, USA, 1998.
[2] Gamma, E., a kol., Design Patterns: Elements of Reusable Object-Oriented Software, Addison-Wesley Professional, ISBN: 0201633612, 1994.
[3] Gregor Hohpe, Gregor Hohpe, Enterprise Integration Patterns, Melrose, MA 2003.
[4] Fowler, M., Model Driven Architecture, 2004.
[5] Fowler, M., UML Distilled: A Brief Guide to the Standard Object Modeling Language, Addison-Wesley Professional, ISBN: 0321193687, 2003
[6] Borland Software Corporation, Successful mplementation of Model Driven Architecture, 2007.
[7] Kleppe, Anneke G., a kol., MDA Explained: The Model Driven Architecture: Practice and Promise, Addison-Wesley, ISBN:032119442X, 2003.
[8] Kontio, M., Architectural manifesto: Choosing MDA tools, 2005.http://www.ibm.com/developerworks/webservices/library/wi-arch18.html

Towards Fast Construction of Static Speech Recognition Network

Martin LOJKA

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

martin.lojka@tuke.sk

***Abstract*—Static speech recognition networks are mainly used in automatic speech recognition systems based on weighted finite state transducers. With operations that are defined for weighted finite state transducers like composition, determinization, minimization and epsilon transitions removal the compact and optimized static recognition network can be created. Optimization process of static speech recognition network is computationally and memory intensive, especially determinization operation after composition of two transducers. In this paper we will face the problem and use an alternative way of construction of speech recognition network by using modified composition operation, which will allow us to overcome determinization operation after the composition.**

***Keywords*—Finite state transducers, Composition operation, Speech recognition, Speech recognition networks**

I. Introduction

Most of the stochastic speech recognition systems are based on statistical information in form of acoustic and language model. Each model is trained on large set of data. Acoustic model is often based on Hidden Markov Models (HMM) and contains information for classification of sounds to particular subword units, phones. For better speech recognition accuracy subword units with respect to neighbor units, the triphones are used. Language model is useful in concatenation of words into sequences. Usually stochastic n-gram language models are used, which can provide us a probability of a word based on its history. Bigrams (one word history is considered) or trigrams (two word history is considered) are usually used. For combination of those two models the pronunciation lexicon is used, which contains associations between words and subwords units.

In this manner the speech recognition is defined as follows[1]:

$$\hat{W} = \arg\max_{W}[P(W|O)] \tag{1}$$

and can be transformed to more acceptable form of

$$\hat{W} = \arg\max_{W}[P(O|W)P(W)]. \tag{2}$$

Where the probability $P(O|W)$ is provided by acoustic model and $P(W)$ by language model.

The speech recognition process can be explained as problem of finding the most probable path through *speech recognition network*, which is consisting of language model, acoustic model and pronunciation lexicon. Every path through this network is defining the output recognized sequence of words with assigned weights. This speech recognition network can be created as static before speech recognition process or dynamically through the process. In this paper we will focus our attention to static construction of speech recognition network.

As proposed by Mohri[2] the speech recognition network can be created using *weighted finite state transducer*(WFST) by using composition operation. Each of the models can be transformed into transducer, which is representing translation of one level of representation to another. Sequencing this translations speech recognition network can be created. This sequencing of transducers in order to recognize speech is called *speech recognition cascade*. Speech recognition network can be created by composition $L \circ G$, where L is transducer of pronunciation lexicon and translates input sequence of phones into words and G is assigning weights to valid sequences of words, which is helping the process to concatenate words into sequence. Next also the C WFST can be used for further expansion of speech recognition network for context-dependent phones, triphones[3]. The construction of speech recognition network R can be then formulated as follows:

$$R = C \circ L \circ G \tag{3}$$

In this paper we will consider only acoustic context inside words, and since the lexicon we have used has already associated words with context dependent units (triphones) we will not use the C WFST. All results in the end of this paper will be then presented only on composition L with G, thus the speech recognition network will have the form:

$$R = L \circ G \tag{4}$$

This static speech recognition network is highly redundant, so the optimization operations are used after composition, like determinization, minimization, and epsilon removal[4]. From these operations the determinization especially after composition is computationally and memory intensive. In this paper we will use an alternative way of constructing speech recognition network with modified composition operation.

II. Theory of WFSTs

A. Semiring

Speech recognition depends on a path through speech recognition network. We need to know how to handle weight on the path and how to combine weights from more than one path. This information depends on what semiring we are using. Semiring is defined as $(\mathbf{K}, \oplus, \otimes, \bar{0}, \bar{1})$, specifically by a set of values $\mathbf{K}$, two binary operations $\oplus$ and $\otimes$, and two designated values $\bar{0}$ and $\bar{1}$. The operations $\oplus$ is associative, cumulative, and has $\bar{0}$ as identity. The operation $\otimes$ is associative, has identity $\bar{1}$, distributes with respect to $\oplus$, and has $\bar{0}$ as annihilator. In the speech recognition

two weights semirings are particularly useful. First the *log probability* semiring $(\Re, \oplus_l, +, \infty, 0)$, where $\oplus_l$ is defined as $a \oplus_l b = -\log(e^{-a} + e^{-b})$ in counter part of probability semiring $([0,1], +, \times, 0, 1)$ in logarithmic domain. Second the *tropical semiring* $(\Re, min, +, \infty, 0)$, which is used as an approximation semiring to the log semiring.

B. Weighted finite state transducers

WFST can be specified as $T = (\Sigma, \Omega, Q, E, i, F, \lambda, \rho)$ over semiring $\mathbf{K}$ by a finite input alphabet Σ, a finite output alphabet Ω, a finite set of states Q, a finite set of transitions $E \subseteq Q \times \Sigma \times (\Omega^+ \cup \{\epsilon\}) \times \mathbf{K} \times Q$, an initial state $i \in Q$, a finite set of final states F, an initial state weight assignment λ and a final state weight assignment ρ [5].

The given transition $e = (p[e], l_i[e], l_o[e], w[e], n[e]) \in E$ is specified by a previous state or origin of the transition $p[e] \in Q$, a next or destination state $n[e] \in Q$, its weight of the transition $w[e]$, its input label $l_i[e]$ and its output label $l_i[e]$.

C. Operations with WFSTs

1) Weight Pushing (push): Theresulting transducer using this operation has "pushed" weights towards initial state. This operations has also his advantage in applying the weights as soon as possible for effective pruning during the speech recognition process.

2) Minimization (min): In this operation we can join *equivalent* states and obtain a transducer with less states and transitions. Two states of WFST are equivalent if the path to final state is labeled with the same symbols and weights of this path including weight of the final state are the same. Equivalent states can be joined without destroying the function of this WFST. In real case there aren't such equivalent states, but with weight pushing operation we can create them and so we can apply the minimization.

3) Determinization (det): The WFST is *deterministic* if there is one unique initial state and no two transitions leaving any state have the same input label. If the WFST is deterministic, the input sequence exactly determines the output sequence. This operation is a way to construct equivalent and deterministic WFST.

4) Epsilon transitions removal (ϵ-removal): This operation removes epsilon transitions from transducer. If an epsilon is used as output and input label, we can travel this transition without any input symbol and the transition produces no output symbol, so we can remove them without losing function of the transducer.

5) Composition ($\circ$): Consider two transducers T_a and T_b defined by (5). The T_a provides mapping from all sequences Σ_a^* to output sequences Ω_a^*, where Σ_a^* represents set of all input sequences that can be constructed from symbols in alphabet Σ_a. The same for output sequences represents notation Ω_a^*. The next WFST T_b in recognition cascade provides further mapping from Ω_a^* to Ω_b^*. This also means that the input alphabet Σ_b of T_b must be the same as output alphabet Ω_a of T_a, thus $\Sigma_b = \Omega_a$. This mapping can be done in one step by composition of transducers T_a and T_b. The resulting transducer is defined by (6)[6].

$$\begin{aligned} T_a &= (\Sigma_a, \Omega_a, Q_a, E_a, i_a, F_a, \lambda_a, \rho_a) \\ T_b &= (\Sigma_b, \Omega_b, Q_b, E_b, i_b, F_b, \lambda_b, \rho_b) \end{aligned} \tag{5}$$

$$T_c = (\Sigma_a, \Omega_b, Q, E, i, F, \lambda_b, \rho_b). \tag{6}$$

Let the writing $[T_a](\alpha \in \{\Sigma_a^*\}, \beta \in \{\Omega_a^*\})$ and $[T_b](\alpha \in \{\Sigma_a^*\}, \beta \in \{\Omega_a^*\})$ represent the mapping of transducers T_a and T_b, the function of composition of two transducers is defined by (7). The notation for this operation is $\circ$, thus $T_c = (T_a \circ T_b)$. The resulting weights on transitions are a $\otimes$-product of particular weights of original two transducers.

$$[T_c](\alpha, \beta) = [T_a \circ T_b](\alpha, \beta) = \bigoplus_{\gamma} [T_a](\alpha, \gamma) \otimes [T_b](\gamma, \beta)$$

The basic principle of composition can be summarized as follows if both transducers doesn't contain ϵ labeled transitions. [2]:

1) The initial state of the T_c is a pair of the initial states of the T_a and the T_b.
2) The final state of the T_c is a pair of final states of the T_a and the T_b.
3) In the resulting WFST T_c there is a transition from pair of states (q_1, q_2) to (r_1, r_2) for each transition in T_a from q_1 to r_1 and in T_b from q_2 to r_2, where the output symbol in T_a is the same as in T_b. The resulting transition is then labeled with input symbol from the transition in T_a and output symbol from the transition in T_b. The resulting weight is the $\otimes$-product of particular weights.

In real cases transducers often contains ϵ labeled transitions. The worst case is when T_a contains ϵ output labels and T_b ϵ input labels. In those cases we need to use composition filter to prevent of creations redundant ϵ paths in resulting transducer. In cases where only one transducer contains ϵ labels, they can be processed sequentially in composition process. The most important property of composition used in this paper is that output of composition of two deterministic transducers is deterministic transducer.

III. Speech Recognition Network

As stated in the introduction section in this paper only the pronunciation lexicon WFST and language model WFST according (4) will be used. This network is highly redundant, so we need to use optimization operations and so constructing network in form:

$$R = push\ min\ det\ \epsilon - removal(L \circ G) \tag{7}$$

From the composition of L and G the ϵ transitions are removed $(\epsilon - removal)$ and the result is determinized (det), minimized (min) and the weight are pushed $(push)$ towards initial state for better pruning during speech recognition process. In order to be able of performing the optimization operations after composition we need to preprocess input WFSTs in the following manner:

1) Language model WFST G is not deterministic because of ϵ transitions leading to back-off state. This will cause after the determinization operation high increase of states and transitions. Therefore to such transitions an auxiliary symbol $\#phi$ needs to be introduced.
2) Pronunciation lexicon WFST L is generally not determinizable, because of existence of homophones. Even without homophones this WFST may not be determinizable. The solution is to introduce here auxiliary symbols $\#0, \#1, ..., \#N$ to make the transducer determinizable.

An alternative way of construction of speech recognition network is:

$$R = push\ min\ \epsilon - removal(det(L) \circ (G)) \quad (8)$$

As we see the determinization operation is restricted only to lexicon WFST, which is much smaller than composition $L \circ G$ in (7), thus the construction of speech recognition network is faster. Using standard composition algorithm described here will result in the following problems. For standard composition the L transducer needs to have output labels on first transitions that are leaving the initial state for early matching with input labels of G transducer preventing creation of *useless* states with the composition process. Useless states are *non-coaccesible* states, which do not lie on a path between initial and final state. These states can be removed without affecting function of transducer. Because of late label matching the weights from transducer G are used later, which is ineffective for using pruning techniques during speech recognition. To overcome this problems a modified composition algorithm was developed, which is described in the following section.

IV. Modified Composition Algorithm

This composition operation, which will be presented here, is inspired by on-the-fly composition developed by Caseiro and Trancoso[7] for specialized composition of lexicon L transducer and language model G transducer. This method was later generalized by a Cheng[8] and Oonishi[9]. In on-the-fly composition a determinized L transducer is used in order to share paths through recognition network while decoding. For decoding process the token passing algorithm is used, where tokens are referencing to a position in transducer L and G.

The basic idea of fast on-the-fly composition algorithm is to disallow following ϵ transitions in L, which are not leading to words matching with G transducer and in this manner to prevent creation of useless states. This is done by labeling each transition with ϵ input label in transducer L with set of reachable output labels. If intersection of a set in L and current reachable input labels in G is non-empty set we can follow this transition and create new transition in resulting transducer. In this process we can also do the label pushing and weight pushing towards initial state. If the intersection is exactly one label, we can output this label in resulting transducer earlier with his weight in G. If the intersection is more than one label we will construct ϵ transition with minimal weight of matching labels in G (*look-ahead* technique).

The modified composition algorithm (on Fig.1) presented here is also using tokens for building output transducer, however here the tokens are referencing not only to state in L (q_l) and G (q_g) but also to a label ($pushed\_label$), for which the state was created to distinguish between states created (state definition (9)) by pushing various labels towards initial state. If no label was pushed then $pushed\_label = \epsilon$. Token also carries information about pushed weight ($pushed\_weight$)(10). The end state of both transducers is handled as transition with special input/output label (END), which will be converted back to end state after composition process. Reachable set for each transition is created as a list using deep-first search, each transition has assigned interval of labels in this list.

$$q_{out} = (q_l, q_g, pushed\_label) \quad (9)$$

$$t = (q_l, q_g, pushed\_label, pushed\_weight) \quad (10)$$

1) Create new token in position where $q_l = 0$, $q_g = 0$, $pushed\_label = \epsilon$ and $pushed\_weight = 0$. Push the token into stack S and repeat next steps until S is empty.
2) Get token from stack S
3) If state $(q_l, q_g, pushed\_label)$ is in output transducer then go to step 2
4) Push all transitions leaving from state q_l from L into stack M
5) Get transition e_l from stack M
6) Get anticipated label set of e_l and set of input labels from all transitions leaving state q_g of G. Find intersection between these two sets. If there is no intersection, no new transition in output transducer will be created, go to step 5 for next e_l.
7) Go through all matched transitions in G from state q_g and accumulate $semiring - sum(\oplus)$ of weights of matched transitions, which will be the $look\_ahead$.
8) If $pushed\_label \neq \epsilon$ and $pushed\_label$ is in the intersection
   a) Seek token through transition e_l in L.
   b) If $l_o[e_l] = pushed\_label$ then $pused\_label = \epsilon$.
   c) $pushed\_weight = 0$.
   d) Create new transition with input label $l_i[e_l]$ and ϵ output label (if $pushed\_label = END$ then output label is also END) from old token position to new token position with weight $w = 0$.
   e) If new token position is not in output transducer as start state of a transition then put token back into stack S.
9) If number of matched labels > 1 and $pushed\_label = \epsilon$
   a) Seek token through transition e in L.
   b) $pushed\_label = 0$
   c) Create new transition with input label $l_i[e_l]$ and ϵ output label from old token position to new token position with weight $w = w[e_l] \otimes look\_ahead \otimes pushed\_weight^{-1}$.
   d) $pushed\_weight = look\_ahead$.
   e) If new token position is not in output transducer as start state of a transition then put token back into stack S.
10) If number of matched labels $= 1$ and $pushed\_label = \epsilon$ then for all transition e_g leaving from q_g in G do
   a) If $l_i[e_g]$ is not the matched label go to step 3 for next transition e_g.
   b) Seek token position through transition e_l in L and e_g in G.
   c) If $l_i[e_g] = l_o[e_l]$ then $pushed\_label = \epsilon$ else $pushed\_label = l_i[e_g]$.
   d) Create new transition with input label $l_i[e_g]$ and output label $l_o[e_l]$ from old token position to new token position with weight $w = w[e_l] \otimes w[e_g] \otimes pushed\_weight^{-1}$.
   e) $pushed\_weight = 0$.
   f) If new token position is not in output transducer as start state of a transition then put token back into stack S.

Fig. 1. Modified composition algorithm

The following points are basic for this algorithm

Step 7 Computes look-ahead weight.

Step 8 Processing of token, referencing to state in output transducer, created by pushing a label.

Step 9 Processing of token referencing to state in output transducer, created with no pushed label and is passing through ϵ transition where number of intersecting labels is more than one.

Step 10 Processing of token referencing to state in output transducer, which was created with no pushed label and exactly one intersection was found or token is passing a transition with non-ϵ output label.

V. Experimental Resuls

In this section we will look at the testing of alternative construction of speech recognition network according (8) in comparison to the network constructed according (7). Specifically time and memory usage, which was needed for creation of the networks and word error rate (WER) as a function of real time factor (RTF) was tested.

The vocabulary for this task had 100k words, the language model was trigram (with 108,847 unigrams, 2,676,635

TABLE I
TIME AND MEMMORY USAGE

Network	Required Time	Memory Used
Standard Composition (Net. (7))	133 min	7.9GB
Modified Composition (Net. (8))	104 min	3.6GB

TABLE II
WER VS. RTF

	Standard Composition Net. (7)		Modified Composition Net. (8)	
Beam Width	WER	RTF	WER	RTF
130	19.34	0.61	21.71	0.57
150	16.01	0.92	16.35	0.87
200	14.26	1.65	14.11	1.68
250	13.92	2.11	13.7	2.12

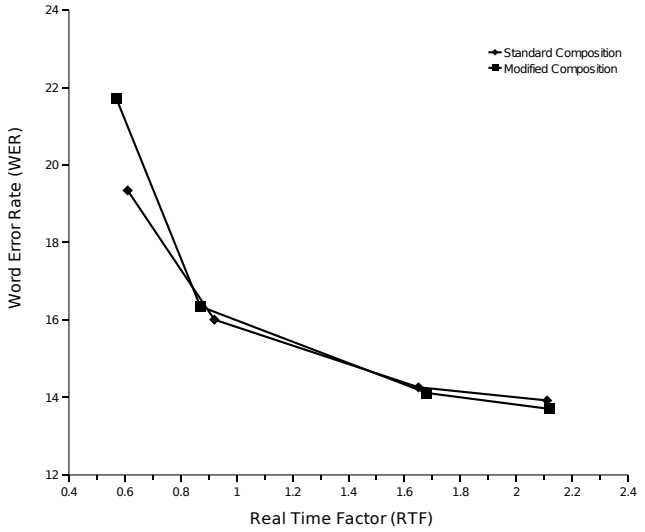


Fig. 2. Word Error Rate (WER) vs. Real Time Factor (RTF)

bigrams, 920,226 trigrams) and acoustic model was trained on 74.6 hours of parliament 16kHz sampled speech with 32 Gaussian mixtures. Testing was done on 74.85 minutes of parliament speech with 8,778 words.

For testing the WER Juicer: a weighted finite state transducers decoder was used[10]. Optimization operations and standard composition was used from AT&T FSM toolkit 4.0[11]. Modified composition was programmed in C++. Generation of L transducer from pronunciation lexicon and G transducer from language model in ARPA format perl scripts were used. Results of time and memory usage summarized in Table I and WER results in Table II.

As we see from Fig.2 alternative way of speech recognition network construction have similar results as the standard construction, the difference is in required time and memory usage, where alternative construction consume a fraction of them.

VI. CONCLUSION

In this paper an alternative way of constructing the speech recognition network based on WFST was shown and presented results in contrast to standard construction. Modified algorithm of composition was used to speed up the process, which had very close WER results to standard construction of network. In the future work more testing will be done and various network components will be used, like C, the context-dependency transducer in order to benefit also from cross-word triphones and various sequences of optimization operations. Next the algorithm will be more generalized in order to be able of composing transducer independently of existence ϵ transitions. The results will be presented in future papers.

ACKNOWLEDGMENT

The research presented in this paper was supported by the Slovak Research and Development Agency under research projects APVV-0369-07 and VMSP-P-0004-09 and is the result of the project implementation Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF.

REFERENCES

[1] X. Huang, A. Acero, and H.-W. Hon, *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2001.
[2] M. Mohri, F. C. N. Pereira, and M. Riley, "Speech recognition with weighted finite-state transducers," *Handbook on Speech Processing and Speech Communication*, 2008.
[3] M. Lojka and J. Juhár, "Finite-state transducers and speech recognition in slovak language," in *SPA 2009 : signal processing : algorithms, architectures arrangements, and applications*. Poznan : University of Technology, 2009, pp. 149–153.
[4] M. Mohri, "Weighted automata algorithms," pp. 213–254, 2009.
[5] M. Mohri, F. C. N. Pereira, and M. Riley, "Weighted finite-state transducers in speech recognition," *Computer Speech and Language*, 2002.
[6] A. Seward, "Efficient methods for automatic speech recognition," Ph.D. dissertation, Royal Institute of Technology Stockholm, 2003.
[7] D. Caseiro and I. Trancoso, "A specialized on-the-fly algorithm for lexicon and language model composition." *IEEE Transactions on Audio, Speech & Language Processing*, vol. 14, no. 4, pp. 1281–1291, 2006.
[8] O. Cheng, J. Dines, and M. Magimai-Doss, "A generalized dynamic composition algorithm of weighted finite state transducers for large vocabulary speech recognition," IDIAP, IDIAP-RR 62, 2006.
[9] T. Oonishi, P. Dixon, K. Iwano, and S. Furui, "Implementation and evaluation of fast on-the-fly wfst composition algorithms," in *Interspeech2008*, 2008.
[10] D. Moore, J. Dines, M. Magimai.-Doss, J. Vepa, O. Cheng, and T. Hain, "Juicer: A weighted finite-state transducer speech decoder," in *3rd Joint Workshop on Multimodal Interaction and Related Machine LEarning Algorithms MLMI'06*, 2006, iDIAP-RR 06-21.
[11] "Fsm 4.0." [Online]. Available: http://www.research.att.com/~fsmtools/fsm/

Dynamic Systems and Their Description: Calculus vs. Action Graphs

Martina L'al'ová

Department of Computers and Informatics,
FEI TU of Košice, Slovak Republic
http://www.tuke.sk

martina.lalova@tuke.sk

***Abstract*—In our paper we specify definition of dynamic system for the comparison of calculus and action graphs description. We look for the options of calculus in the case of dynamical systems. We want to use them to describe their structure and compare them with action graphs. We want to determine their strengths and weaknesses and identify range of opportunities for their use.**

***Keywords*—Dynamic systems, ρ-calculus, λ-calculus, matching power, π-calculus, μ-calculus, action graphs**

I. Introduction

In our research, we focus in the dynamic system, deterministic, discrete and closed in terms of surroundings. We use abstraction to simplify the structure of our system. We want to find the most suitable method for its description. We consider about characteristics of a structure that describes dynamic systems. We are looking for the best description of the characteristics of dynamic systems. By using the calculus we want to find ideal component for representation. This calculi we compare with action graphs. We also describe their advantages and disadvantages. In our research we want to deal with ρ-calculus, π-calculus and μ-calculus and their possibilities. We show the potention of action graphs and their relationship with other calculi.

II. Dynamic system

The notion system is used in many scientific fields. We are interested in systems used only in informatics. System is synonym for a set of independent but interrelated elements comprising a unified whole. We require that system has to be an abstraction for understanding basic properties and dynamics between the components. Values produced by the environment are independent, e.g. input values, controlling values. Values produced by the system are the dependent values (output values). System forms the ordered pairs of cause and consequence - the causality.

A dynamic system is a system in terms of behavior in relation to time, whose instantaneous state depends on previous states and external cues. A function of dynamic system depends on the exchange of information between the elements of the system and its environment by inputs and outputs.

A function of dynamic system depends on the exchange of information between the elements of the system and its environment by inputs and outputs.
We need to restrict the definition of dynamic system to deterministic, discrete one, closed regarding the surroundings, bounded by the number of values and continuous in with respect to values.

A dynamic system S is then defined as an eight-tuple of mappings:

$$S \equiv (T, X, U, \mathbf{U}, Y, \mathbf{Y}, \varphi, g),$$

where:

- T is a set of moments of time,
- X is a set of system states,
- U is a set of instantaneous values of input values,
- Y is a set of instantaneous values of output values,
- $\mathbf{U}$ is a set of admissible input functions

$$\mathbf{U} = u(t) : T \rightarrow U,$$

- $\mathbf{Y}$ is a set of admissible output functions

$$\mathbf{Y} = y(t) : T \rightarrow Y.$$

We define the orientation of time in the set of times T as an ordered subset of the set of real numbers. A set of input functions U is non-empty and it enables the unification and concatenation of admissible input actions.
States transition function φ is defined by values of states $x(t)$

$$x(t) = \varphi(t, t_0, x(t_0), u),$$

where $x \in X$ is a state and t_0 is an initial time moment. Its orientation in time is defined for all $t \geqslant t_0$ and it holds the identity

$$\varphi(t, t, x(t), u) = x(t),$$

where:

$$t \in T, x(t) \in X, u \in U.$$

If it holds that $u, \bar{u} \in U$ and $u(t) = \bar{u}(t)$
on interval $t_0 \leqslant t \leqslant t_n$, then we obtain

$$\varphi(t_0, t_n, x, u) = \varphi(t_0, t_n, x, \bar{u})$$

where $\bar{u}$ is the input value changed onto output and the function φ is uniquely determined by input actions. Output mapping g determines the output values

$$y(t) = g(x(t), u(t), t).$$

If we consider ordered pair $(t, x(t)), t \in T, x \in X$ to be an event of system from $T \times X$, then it holds that

$$y(t) = g(x(t), t).$$

III. DESCRIPTION OF DYNAMIC SYSTEM

On the base of observed properties and probabilities we try to formulate valid theses, which describe given system. We want to accomplish the description with the wide spectrum of expressibility which encapsulates the most important properties of a given system. A method for such nodes of systems are used pattern matchings. Recognition of the properties of a given system does not lead to recognition of fundamentals of the system. It can help to understand relations between basic entities. Observation of a system needs a high measure of abstraction (ρ-calculus). To recognize the fundamentals of a system means to understand the structure and its elements as a whole together with their interaction (functions and dependencies) as in a way of internally closed system. Subject of the observation defined in this way we can consider to be an object described by scientific facts. However described object has several properties and we are only interested in some of these. That is the reason why is a certain measure of abstraction and expliciteness requested.

Calculus is a scheme for construction of objects from elements. It is a set of rules for the schematic operations. There is finite number of atomic objects included into calculus from which we are able to build new onjects. There is also finite set of rules for the operations flow with objects. Derivation is the construction of objects according to that rules. Terms are strings of symbols from an alphabet, consisting of the signature and a countably infinite set of variables.

A term rewriting system consist of terms and rules for rewriting (reducing) these terms [8].

Formal: A signature Σ consists of a non-empty set of function symbols or operator symbol $G, H, ...$, each equipped with a fixed arity. The arity of a function symbol G is a natural number, indicating the number of arguments it is supposed to have.

The set of terms over Σ is indicate as $Ter(\Sigma)$ and is defined inductively:

- $x \in Ter(\Sigma)$ for every $x \in Var$
- if G si an n-ary function symbol $(n \geqslant 0)$ and $t_0, ..., t_n \in Ter(\Sigma)$, then $G(t_0, ..., t_n) \in Ter(\Sigma)$

The term t_i are called the arguments of the term $F(t_0, ..., t_n)$, and the symbol G the head symbol or root. Notation

$$G \equiv root(t).$$

Contex can be defined as a term containing zero, one or more occurrences of a special nstant symbol ϵ, denoting holes, i.e., a term over the extendet signature $\Sigma \cup \{\epsilon\}$. If C is a context containing exactly n holes an $t_0, .., t_n$ are terms, then

$$C[t_0, .., t_n]$$

denote the result of replacing the holes of C from left to right by $t_0, .., t_n$. If $t \in Ter(\Sigma)$ can be written as $t \equiv C[t_0, .., t_n]$, then context C is a prefix of t. If

$$t \equiv D[C[t_0, .., t_n]]$$

for some prefix D, then C is a subcontext of t. Klop's term rewriting graphs since his approach is nearest to ours needs. Klop denotes the node N_i by $x = N_i(y, z)$, whereas we use the notation $\langle y, z \rangle N_i(x)$. Klop uses $\langle x|...\rangle$ to denote the root of the graph, whereas we have the interface $()[...]\langle x \rangle$, which we will use in the Examples 1,2.

A. π-calculus

The π-calculus belongs to process calculi of the theoretical computer science and was originally developed by Robin Milner, Joachim Parrow and David Walker as a continuation of work on the process calculus CCS (Calculus of Communicating Systems) [10].

π-calculus is a formal method for describing communicating processes and analyzing their properties. It allows to describe the net of communicating processes and allows to model the connection changes between them. A basic element of π-calculus is the process interaction. Each process is taken as a black box which communicates with other processes through its ports. Mobility of system is dynamic system ability of changing its configuration of element and interaction between them.We are able to describe the behavior of communicating the concurrent processes by π-calculus. The π-calculus has an exact mathematical semantics which is very useful if we want to prove some useful statements. It is different from its predecessors CCS a CSP; and it supports the ability of changing the structure of a system which is requested mainly in dynamic systems.We can compare this kind of "mobility" with the mobility of objects in an object-oriented system, also in case with entities in some simulated world. π-calculus is specially useful for description of communication of the concurrents systems [11]. It enables the recursive definitions, defining truth values and branching; operations with lists, objects and λ-expressions. Like λ-calculus it does not contain any inbuilt data types. As it is possible to represent data by processes, expressivity of the calculus is not decreased but it brings ambiguity to the structure. For example, if our dynamic structure is based on the shared unit (e.g. memory), then encoding globally shared variable in terms of channels would be complicated.

Communication of a process is described using syntax

$$\pi ::= \quad x(y)|\bar{x}(y),$$

where $x(y)$ denotes that source port x sends a message to the target one y; $\bar{x}(y)$ denotes that a source port x become a target $\bar{x}$ and it receive a message from y.

For describing the communication of processes we use terms constructed according the following syntax:

$$S, N_a, \ldots, N_n ::= x|y|z|\bar{x}, \bar{y}, \bar{z}| \rightarrow |\circ$$

where $N_a, \ldots, N_n$ are nodes and S is the starting node. The nodes representing processes, $x, y, \bar{x}, \bar{y}$ are source and target ports, $\rightarrow$ is function and $\circ$ denotes the composition of connections. Dynamic change in the communication topology between the two processes is a mobility. During the communication at runtime the nodes obtain or lose the communication ports.

A system can by described by

- processes name,
- used ports,
- processes behavior description.

Processes are represented by nodes. Behavior of dynamic system change process N_a is represented by node with listening port x where it expects message e and on the sending port $\bar{y}$ where it sends message u, is describe as

$$N_a(x, y) = x(e), \bar{y}\langle u \rangle$$

We have to note: on one side of connection we have a sending port and on second side we have listening port. Seen of one node is end of connection sending, but from other node it can by a listening point.

Example 1:
We build the system with dynamic change of connections. For connection description we use input and output node interface. We describe a system P consisting of

- nodes $N_a, ..., N_n$
- starting node S
- connection l, k, r, p and m

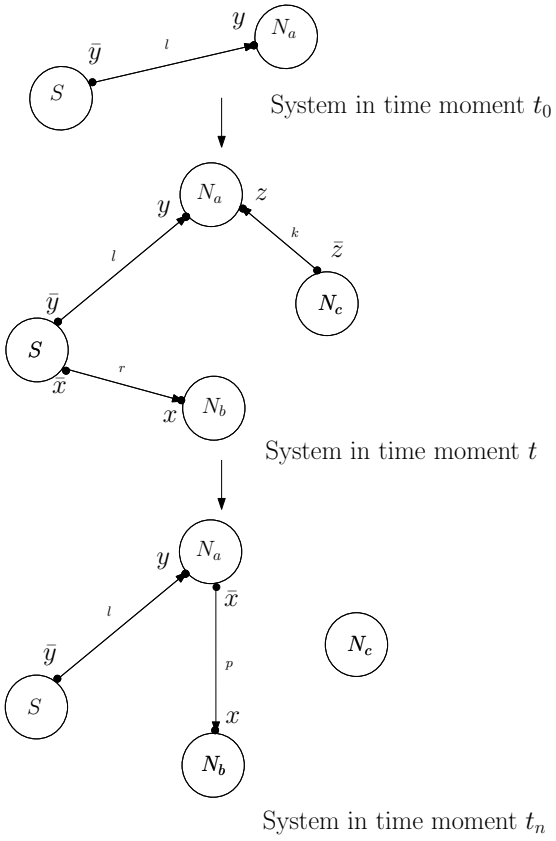


Fig. 1. Mobility of system P

Dynamic change is represented in $Fig.1$ as a port configuration change. For example mode N_a has changed listener port for sending. In the first time moment t_o the system has one connection l between start node S with sending port $\bar{y}$ and node N_a with listening port y. In the time moment t the start node S has two connections and two sending ports $\bar{y}$ and $\bar{x}$ as a result of configuration change of system P, where the connection r was created between nodes S with sending port $\bar{x}$ and node N_b with listening port x and the connection k between nodes N_a with listening port z and N_c with sending port $\bar{z}$. Configuration change of system in time moment t to system in time moment t_n depicted at Fig. 1 can by formulated by the term:

$$S|N_a|N_b|N_c = \bar{y}\langle x\rangle.S(y,x)|y(z).\bar{z}.N_a(y,z) \rightarrow \\ \rightarrow S(y)|\bar{x}.N_a(y,x)|x.N_b(x)$$

This entry has great power of content but the absence of his clear and illustrative. □
There may be two different reactions. The reaction always takes choice between the process that sends a message through the port and the process which through the same port, trying to commuicate. If this happens, it can take choice between the reaction. At one moment can have node to select a more of possible responses that could be made. This is the case with system P.

B. ρ-calculus

ρ-calculus was created to describe the basic elements of explicit description of objects, several terms, rules, abstraction, application and results [5], [4].
In ρ-calculus we use λ-abstraction of the form $\lambda X.N$. It is a generalization of the rule of abstraction

$$P \rightarrow N,$$

where P is actual arbitrary term and not necessary variable X. N is a consumed argument. Free variables from P are bounded in N by the application of the rule $P \rightarrow N$ on term M. We denote it by $(P \rightarrow N)M$. The evaluation of term we denote by $\delta(N)$ (by applying $\delta(N)$ we mean substitution to the term N) where $\delta(N)$ represents the solution of matching between P and M. For a more general description we use:

$$Terms : M, N, P ::= x|c|P \rightarrow M|MN|M \wr N|N[\delta]$$

If we take the structure from previous example and we place it to the term P then we are able to test matching between this structure and its mirror M which has properties of abstraction. We use notation

$$P \wr M$$

We can set the matching power of ρ-calculus for using arbitrary theory. In classical transcriptive terms it can lead to nondeterministic behavior of dynamic system [7].
The ability for parameterizing the ρ-calculus by a matching theory opens possibilities for describing the properties of the structure.

C. μ-calculus

μ-calculus is suitable for model checking. This algorithm is usable in deterministic system also in nondeterministic system. Necessity of using this calculus is indeed the creation of standard model; in our case model of system P and its mirrors. Therefore of model satisfying to standard, which we are able to describe [6], [2].
The most appropriate seems to be deterministic Kripke structures, which could satisfy for our defined dynamic system. As it requires model with final number of traces, it can work also with nondeterministic Kripke structures. However here is more difficult check in comparison with deterministic loop. Creating a model is not effective.
We have developed a special dynamic system. If we preserve marking input-output sets, than we create new input-output subset, because input for a one node is the output of the second node.

D. Action graphs

Theory which provides connection of the system structure description with its objects and relations and demonstrative graphical representation is encapsulated to the form of action graphs. Milner introduced action calculi as a framework for

describing models of interactive behaviour, where a graph corresponds to a process and the dynamics of a graph corresponds to interaction between processes. His motivation arose from analysing graphs and syntax of the π-calculus [10].
Action graphs are used for the description of several types of interactions including necessary functions. Computation of functions and interaction are described on the basis of the π-calculus. Simple action graphs are very similar to the graphs of terms.

Formally: action graph is a six-tuple

$$(N, E, S, T, L, C),$$

where

- N is a set of nodes,
- E is a set of edges,
- S is the initial node,
- T is a set of final nodes,
- L is the function label,
- C is a function.

Function label assigns a label to every edge and function assigns a node to a pseudocode. Every node of a graph represents a "region", every edge represents the function of interaction. We can create system P according by π-calculus [3]. In the system are defined interface ϵ, contexts and actions. Arrows C, D, G, H (edges from set E) are used as context [1], [9]. Context with domain ϵ consists of all used ports where a, b, d, e are actions of nodes S, N_a, N_b and N_c. Let pairs (l, r) and (m, n) be the reaction rule and R is set of reacts then reaction relations for start node S we write as:

$$\longrightarrow \stackrel{def}{=} \{(C \circ l, C' \circ r)(l, r) \in R\}$$

and relation $S \xrightarrow{C} N_a \stackrel{def}{\Longleftrightarrow} CS \longrightarrow N_a$, where we use action rule l of our system P. We write $S \xrightarrow{C} P_a$ for a member $(S; N_a)$ of this relation and l we call redex (Fig. 2).

Example 2:
We use structure from $Fig.$ 1. Here we builded system P with nodes S, N_a, N_b and N_c and connections l, k, r, p and m, which describe mobility of this system P.

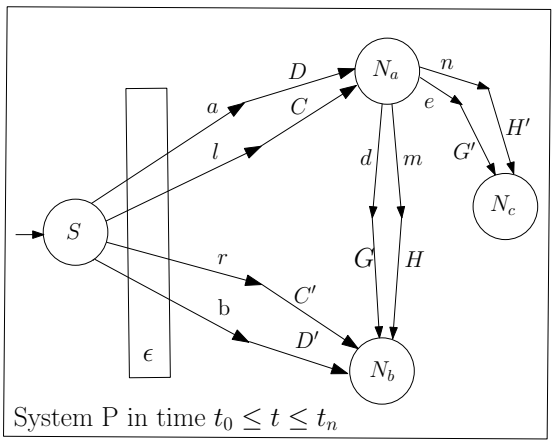


Fig. 2. Action graph of system P

In the system P are defined interface ϵ of node S, which is a "region" for action a, b. Start node S is a root of system P and we describe $S \equiv root(P)$. Reaction rules (l, r) are bounded in relation, they create context. Demarcation of border line around figure is symbolizing closed system in view of surroundings. □
Action graphs are suitable for using in category theory. They are mapping all objects and morfizms. They are more clear and illustrative. Action graphs are good graphical representation dynamic system properties.

Action graphs are similar to term graphs. Specific tool for action graphs exist and it allows the user to naturally switch between the syntactic and graphical presentations. The implementation includes a general matching algorithm for identifying redexes in a graph, and for reductions. [8]

IV. Conclusion

In this paper we described dynamical system with assistance of calculus and action graphs. We briefly compared their advantages for dynamic systems modeling and differences between them. There is close relationship between both ways, where is useful to modeling system by π-calculus, then to use abstraction of ρ-calculus for generalization of model properties and its power matching and finally to use μ-calculus for model checking.
We can use action graphs description of system for better illustrative of dynamic system structure, its objects and relations between them. Action graphs description of system can by created from any term of calculus used for dynamic system description.
Action graphs can be used for recursive notacion of formalism. Since recursive notacion we mean a case when a node can by subsystem described by another action graph. The graph inside the node is an abstraction $y(y)$, while the node itself determines the way in which that abstraction is used. The application of function to argument is represented by linking the node N_i argument and the node N_j to the N_k-node. This interesting properties of action graphs are the reason why their problematics will be a central domain of our research.

Acknowledgment

This work was supported by VEGA Grant No. 1/0175/08 Behavioural Categorical Models for Complex Programm Systems

References

[1] L. C. Andrew Phillips and G. Castagna, "A graphical representation for biological processes in the stochastic pi-calculus," *Transactions on Computational Systems Biology*, 2006, to appear.
[2] J. Bradfield and C. Stirling, "Modal mu-calculi." [Online]. Available: http://www.csc.liv.ac.uk/~frank/MLHandbook
[3] G. L. Cattani, J. J. Leifer, and R. Milner, "Contexts and embeddings for closed shallow action graphs," Tech. Rep., 2000.
[4] H. Cirstea, G. Faure, and C. Kirchner, "A rho-calculus of explicit constraint application," *To appear in the journal of Higher-Order and Symbolic Computation*, 2005.
[5] H. Cirstea, G. Faure, and C. Kirchner, "A Rho-Calculus of explicit constraint application," *Higher-Order and Symbolic Computation*, vol. 20, pp. 37–72, 2007. [Online]. Available: Papers/HOSC2007.pdf
[6] Y. Dong, B. Sarna-starosta, C. R. Ramakrishnan, and S. A. Smolka, "Vacuity checking in the modal mu-calculus," in *In Proceedings of AMAST'02, volume 2422 of LNCS*, 2002, pp. 147–162.
[7] M. Fernandez, I. Mackie, and F.-R. Sinot, "Interaction nets vs. the rho-calculus: Introducing bigraphical nets," in *Proceedings of EXPRESS'05, satellite workshop of Concur*, ser. ENTCS. Elsevier, 2005.
[8] J. W. Klop, "Term rewriting systems," 1992.
[9] B. König, "Description and verification of mobile processes with graph rewriting techniques."
[10] R. Milner, "The polyadic pi-calculus: a tutorial," Logic and Algebra of Specification, Tech. Rep., 1991.
[11] D. Sangiorgi and D. Walker, *PI-Calculus: A Theory of Mobile Processes*. New York, NY, USA: Cambridge University Press, 2001.

Determination of entry image data flows for scanning the environment in MATLAB/Simulink

[1]*Ľubomír MATIS*, [2]*František Baník*

[1,2]Dept. of Electronics, Mechatronics and Industrial Engineering, FEI TU of Košice, Slovak Republic

[1]lubomir.matis@tuke.sk, [2]frantisek.banik@tuke.sk

***Abstract*— In the following article, the model for the displaying of image in MATLAB /Simulink environment is solved. This is partial assignment for acknowledging object in the space.**

Article describes blocks for picture scanning, their setting and converting of colors for their right transformation to final picture.

Processed picture is then adjusted according the requirements of final model.

***Keywords*— colorful model, displaying, converting**

I. Introduction

One of the most important abilities of intelligent systems is their ability of perceiving surrounding environment as a group of objects placed in the area at certain place at certain time. In order to create a system able to react on its environment independently, it needs to be able to acknowledge this situation.

For the system to execute particular task, it needs to be able to acknowledge the important objects in its surrounding. In other words, taking in consideration these kinds of systems where we need to increase their intelligence, is the acknowledgment of objects the main aim which without this would not be possible.

Definition of objects emanates from methods of identifying patterns which these objects include. At present times, these methods are reaching quality level with high credibility even in the experiments reaching close to the real environment with their requirements. In particular, methods discussed are the ones of local characteristic types, which are trying to find beneficiary points at learned image models. From those points they gain the information to be used for their definition. The advantage is that these points will be appearing in higher numbers, which enhance the chances of finding sufficient number of points with close environments in the phase of identifying on similar model.

To obtain real image data from the environment, it is necessary to define the device, which will execute this process as well as adjust the color indications with enter the device.

II. Device to Obtain an Image

There is a video block from application library Simulink to be used for obtaining the image in the model. This enables to obtain image and image data flows from the device such as camera and other digital devices into the model.

Block retaining the image begins the process, initializes, does the settings and controls registration device. All this functions are taking place at the beginning of model realization. At the time of running, block with image data support one image picture for each simulating time step. According the needs, block can be assembled with one outcome port or three outcome ports corresponding with non-compressed color groups such as red, green, or blue, or Y, Cb, Cr. Individual configurations are depicted at (Chart.1) [1]

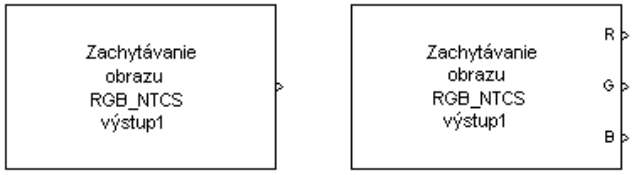


Chart. 1. Configuration of retaining of picture with one and with three ports.

III. Settings

For correct functioning it is necessary to set the parameters of block in dialog window (Chart. 2), where particular settings of displaying device are depicted. Certain areas which device is not dependant on are not depicted. If device does not need particular function, it will not be displayed in dialog window.

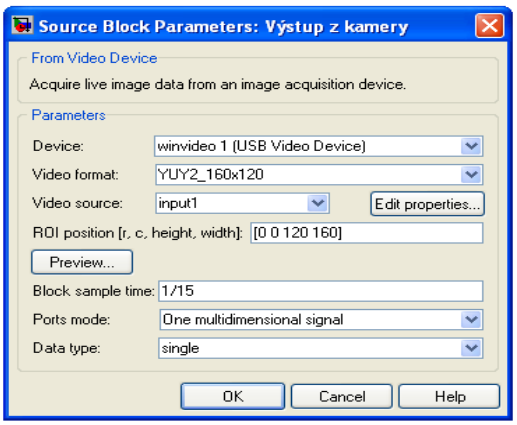


Chart.2. Dialog window with parameters

Device to obtain the image, camera we are connecting to, is displayed in the entry "Device". According to which devices is system connected to, the entries in the list change. After opening, all cameras supported by the block and connected to the model are displayed.

Next important step is setting of video format in entry "Video Format". It shows video formats supported by the device. This list changes with each device. If device supports cameras with particular resolution, these will be displayed on the list separately. By using particular camera and its characteristics, the resolution of pictures is set at figure160x120.

Entry "Video Source" serves for accessible entry sources of specified device. By using the "Edit Properties" key we adjust primary characteristics. By opening the window, the settings of primary specific characteristics of used camera, such as brightness and contrast are displayed (Chart. 4). Properties listed in the chart change according to the used device. Ones can be adjusted are marked by pencil icon or by opening list. It is not possible to adjust the items colored in grey. Changed data are automatically saved after closing the window.

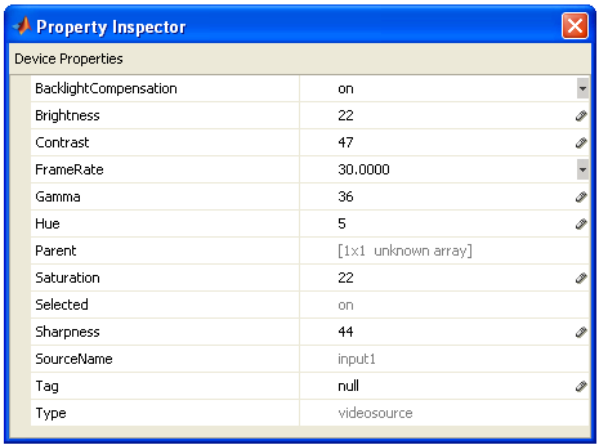


Chart.4. Dialog window of specific properties

Entry figures of row vector for closer determination of video image area are set in "ROI Position" entry. Under format we understand row, column, height, and width. Indicated by video resolution, basic figure for row and column is 0 if figures for height and width are set to the maximum option. If we want to display the whole sized image, we need to change the figures only at the pole of height and width.

After using the key "Preview Button", actual picture from camera is displayed. While reviewing the video, changes of settings are being executed, the picture is changed accordingly which enables to obtain needed picture during the command of model.

Window of timing patterning serves for determination of patterns during the simulation.

"Ports mode" are using on specify of one output port for all colors, or individual for every port (e.g. R, G, B). In the present case isn't necessarily separation color, is chosen one multidimensional signal of output signal that shall combined into one's graphics about information of signal for all colors.

"Data type" item displays data type of video building by the output of block. This data type indicates how image frames are output from the block to Simulink. It supports all MATLAB data types and single is the default. [2]

Following a settings individual parameters dialogue window display devices and running application oneself a screen will be displays environment in real time (Chart.3).

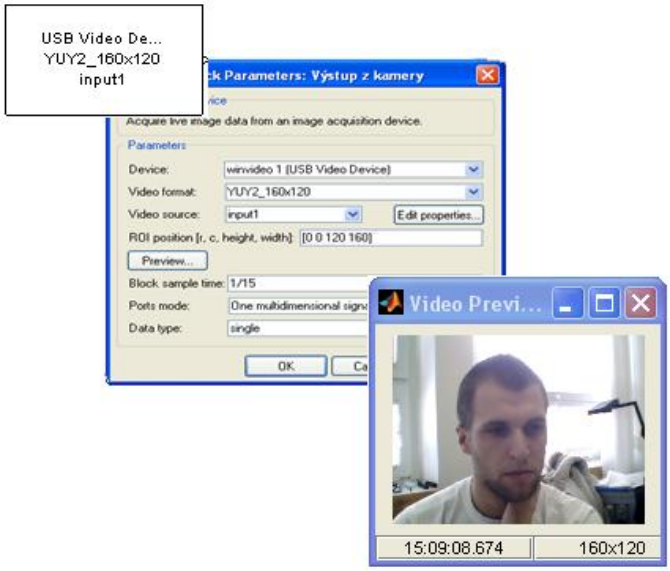


Chart.3 Scanning of environment in real time

IV. CONVERT COLOR INFORMATION BETWEEN COLOR SPACES

Output signal of block for image acquisition consists of color model with specific video separation.

For correct function of model is necessary this color model convert from Y'CbCr to RGB model.

Scanning of objects before converting is visible on (Chart.5)

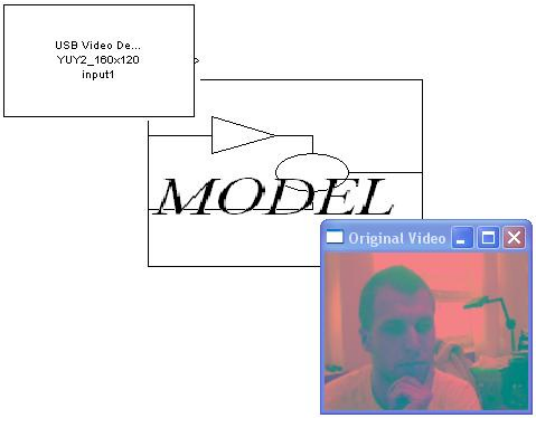


Chart.5 Scanning of object before converting of color model

The base of RGB model is additive folding color red, green and blue. Others colors are voice additive folder this color expressed by weighing sum of individual elements. Whereby have they colors bigger value, thereby is consequential color light. Model RGB is represented unit cube placing at the beginning unit coordinate system (Chart.6). Individual axis coordinate the system represent the size appurtenant - colored components in resultant color. Point in beginning coordinate system representing (0,0,0) sooty mould color and top (1,1,1) bleach color. In practice oneself unit cube separate on smaller

sections, most frequently 256 sections,(8 bit) i.e. intensity everyone's of image point can represent 24 bit. [3] [4]

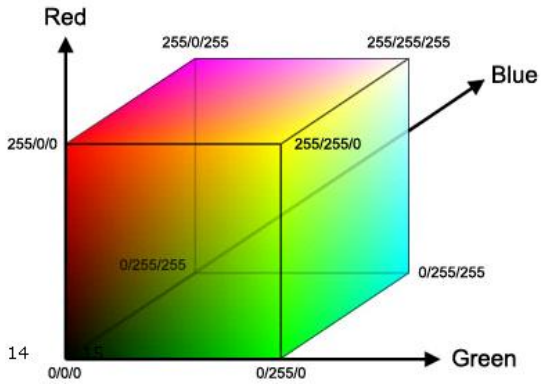


Chart.6 RGB description with gradual change colors one's walls of cube

Color model Y'CbCr is defined by standard CCIR 601. Y in picture introduces brightness, black and white or colorless a part of picture. Cb and Cr are in color differences blue and turn red colors.

Y, Cb and Cr are converted of RGB by definition CCIR recommended 601 but are standardized some that acquirement 256 value of 8- bit binary coding.

For value Y' by references 601 applies to:

$$Y' = 0.299\ \mathrm{R}' + 0.587\ \mathrm{G}' + 0.114\ \mathrm{B} \qquad (1)$$

The R'G'B' to Y'CbCr conversion and the Y'CbCr to R'G'B' conversion are defined by the following equations:

$$\begin{bmatrix} Y' \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + A \times \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} \qquad (2)$$

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = B \times \left(\begin{bmatrix} Y' \\ Cb \\ Cr \end{bmatrix} - \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \right) \qquad (3)$$

The values in the A and B matrices are based on choices defined by standard CCIR 601.
In (Tab.1) are illustration potential value of matrix A and B defined by standard 601.

Tab.1 Value in the A and B matrix defined by standard CCIR 601

Matrix	**Use conversion specified by = Rec. 601 (SDTV)**
A	$\begin{bmatrix} 0.25678824 & 0.50412941 & 0.09790588 \\ -0.1482229 & -0.29099279 & 0.43921569 \\ 0.43921569 & -0.36778831 & -0.07142737 \end{bmatrix}$
B	$\begin{bmatrix} 1.1643836 & 0 & 1.5960268 \\ 1.1643836 & -0.39176229 & -0.81296765 \\ 0.116438356 & 2.0172321 & 0 \end{bmatrix}$

Scanning objects after converting color models is illustrated on (Chart.7).

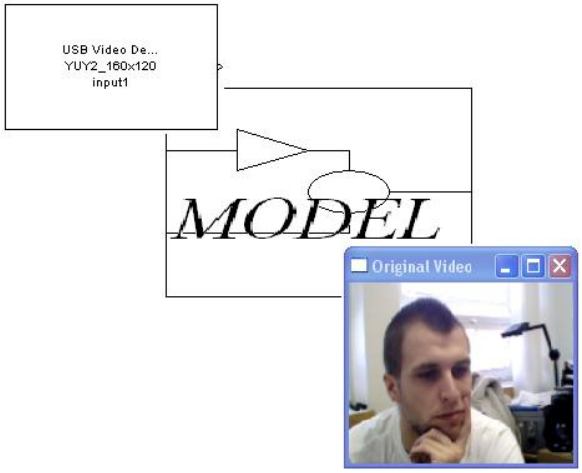


Chart.7 Scanning of object after converting of color model

Y′CbCr signals (prior to scaling and offsets to place the signals into digital form) are called Y'PbPr, and are created from the corresponding gamma-adjusted RGB (red, green and blue) source using two defined constants K_B and K_R as follows:

$$Y' = K_R \cdot R' + (1 - K_R - K_B) \cdot G' + K_B \cdot B' \qquad (4)$$

$$P_B = \frac{1}{2} \cdot \frac{B' - Y'}{1 - K_B} \qquad (5)$$

$$P_R = \frac{1}{2} \cdot \frac{R' - Y'}{1 - K_R} \qquad (6)$$

Where K_B and K_R are ordinarily derived from the definition of the corresponding RGB space.

Here, the prime ′ symbols mean gamma correction is being used; thus R′, G′ and B′ and to nominally range from 0 to 1, with 0 representing the minimum intensity (e.g., for display of the color black) and 1 the maximum (e.g., for display of the color white). The resulting luma (Y) value will then have a nominal range from 0 to 1, and the chroma (C_B and C_R) values will have a nominal range from -0.5 to +0.5. [3]

The form of Y′CbCr that was defined for standard definition use in CCIR 601 standard for use with digital component video is derived from the corresponding RGB space as follows:

$$K_B = 0.114$$
$$K_R = 0.299$$

From the above constants and formulas, the following can be derived terms for transfer color models that are description into terms (7) to (18).

Analog YPbPr from analog R'G'B' is derived as follows:

$$Y' = 0.299 \cdot R' + 0.587 \cdot G' + 0.114 \cdot B' \quad (7)$$

$$P_B = -0.168736 \cdot R' - 0.331264 \cdot G' + 0.5 \cdot B' \quad (8)$$

$$P_R = 0.5 \cdot R' - 0.418688 \cdot G' - 0.081312 \cdot B' \quad (9)$$

Digital Y′CbCr (8 bits per sample) is derived from analog R'G'B' as follows:

$$Y' = 16 + (65.481 \cdot R' + 128.553 \cdot G' + 24.996 \cdot B') \quad (10)$$

$$C_B = 128 + (-37.797 \cdot R' - 74.203 \cdot G' + 112.0 \cdot B') \quad (11)$$

$$C_R = 128 + (112.0 \cdot R' - 74.203 \cdot G' + 18.214 \cdot B') \quad (12)$$

Digital Y′CbCr is derived from digital R'dG'dB'd (8 bits per sample) according to the following equations:

$$Y' = 16 + \frac{65.738 \cdot R'_D}{256} + \frac{129.057 \cdot G'_D}{256} + \frac{25.064 \cdot B'_D}{256} \quad (13)$$

$$C_B = 128 + \frac{-37.945 \cdot R'_D}{256} - \frac{74.494 \cdot G'_D}{256} + \frac{112.439 \cdot B'_D}{256} \quad (14)$$

$$C_R = 128 + \frac{112.439 \cdot R'_D}{256} - \frac{94.154 \cdot G'_D}{256} - \frac{18.285 \cdot B'_D}{256} \quad (15)$$

The inverse transform is:

$$R'_D = \frac{298.082 \cdot Y'}{256} + \frac{408.583 \cdot C_R}{256} - 222.921 \quad (16)$$

$$G'_D = \frac{298.082 \cdot Y'}{256} - \frac{100.291 \cdot C_B}{256} - \frac{208.120 \cdot C_B}{256} + 135.576 \quad (17)$$

$$B'_D = \frac{298.082 \cdot Y'}{256} + \frac{516.412 \cdot C_B}{256} - 276.836 \quad (18)$$

V. Conclusion

Submitted article is centered on scanning and converting colour models that are a part of model for detection specific object in space based on colour distinction.

First a part of article is oriented on device for scanning video with setting for corrent function.

Second part is oriented on convert color information between color spaces of output signal of block for scanning image of color model with defined video resolution. For current function model is necessary this color model converted Y'CbCr to RGB. Individual passages between colour models are described relation.

Y, Cb a Cr are converted from RGB by definition CCIR recommendation 601.

References

[1] *Image Acquisition Toolbox™ User's Guide* © COPYRIGHT 2003–2010 by The MathWorks, Inc. pp. 406.
[2] *Image Acquisition Toolbox™ User's Guide* © COPYRIGHT 2003–2010 by The MathWorks, Inc. pp. 407-414.
[3] Charles A. Poynton (2003). Digital Video and HDTV: Algorithms and Interfaces. ISBN 1558607927.
[4] Nicholas Boughen (2003). Lightwave 3d 7.5 Lighting. Wordware Publishing, Inc. ISBN 1556223544

Analyzing application for network monitoring

Tomáš MIHOK, Miroslav ANTL, Martin RÉVÉS, Juraj GIERTL

Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

tomas.mihok@cnl.tuke.sk, miroslav.antl@gmail.com, martin.reves@cnl.tuke.sk, juraj.giertl@cnl.tuke.sk

***Abstract*—This paper deals with a monitoring of computer network traffic using BasicMeter tool based on the IPFIX architecture. Particular attention is given to analyzing application, modular design of the application web interface, and implementation of Java application for visualization of network traffic parameters.**

***Keywords*—Computer network, IPFIX, network monitoring, network traffic.**

I. Introduction

IPFIX (IP Flow Information Export) is a standard designed for obtaining and export of information about IP flows and thus provides information about the traffic in the monitored network [1]. Information is gathered through the metering process, which captures packets and generates flow records. Flow records are created according to defined templates, which consist of information elements defined by IPFIX information model [2]. Each information element has assigned a unique identifier, name and data type. Flow records are then exported in the form of IPFIX protocol messages [3] to the collecting process, which stores them in a database for future use, or sends them directly to the analyzing application for appropriate visualization for the user of monitoring system. In the Computer Networks Laboratory at the Technical University of Kosice we have developed the BasicMeter tool [4] in conformity with the mentioned IPFIX specifications (Fig. 1).

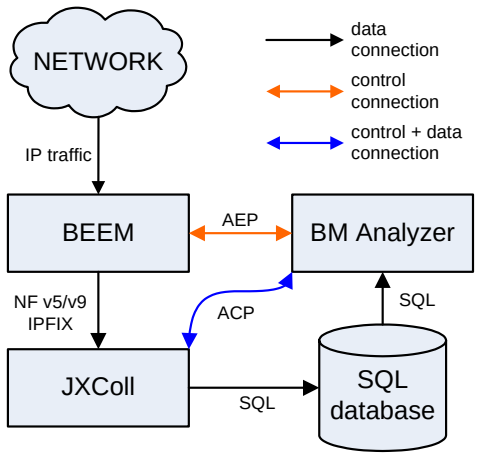


Fig. 1. Architecture of the BasicMeter tool.

Components of the BasicMeter tool:

- *BEEM* – BasicMeter Metering and Exporting Process,
- *JXColl* – Java XML Collector,
- *BMAnalyzer* – BasicMeter Analyzer,
- *ACP* – Analyzer Collector Protocol,
- *AEP* – Analyzer Exporter Protocol.

The BasicMeter as the IPFIX based tool has various applicabilities [5], such as the basic monitoring of network infrastructures, traffic optimization, QoS parameters monitoring, support for real-time and interactive applications, accounting for network services, security analysis.

In the following sections we describe modular concept of the analyzing application web interface, and implementation of Java application for visualization of network traffic parameters.

II. Modular concept of WebAnalyzer

WebAnalyzer sits on the upper level of BasicMeter architecture. Its main purpose is to allow both, display output and enable user to communicate with other BasicMeter parts via web browser. One of the requirements during the development of WebAnalyzer was its modularity.

Modularity is attribute of application which allows it to spread its functionality into separate modules. Reason for this is obvious - allows future programmers to easily add functionality without rewriting the original code of the application. Because of this requirement, it was important to choose proper approach and design simple concept that would fulfill this criterion [6].

Application itself was implemented with MVC (Model-View-Controller) design pattern in mind [7]. This pattern divides application into three main parts:

- Model – describes database or any other means of storing data.
- View – represents interface thus part that communicates with user.
- Controller – is a logic of application and it connects both model and view.

WebAnalyzer is written in Java programming language and a specific framework called Apache Wicket [8]. This framework is Java implementation of MVC design pattern aimed at web applications. Wicket is relatively new framework but it has proven itself as a powerful tool described by authors as component based.

Components are parts of web application that can represent link, button or anything else. For purpose of creating a modular web application, component named Panel has been chosen (Fig. 2). Panel itself provides place for other components. This way a programmer that wants to add a module only create a new panel and implement its functionality inside panel.

WebAnalyzer itself has functionality implemented only in its modules and can only display them. Although this concept is very simple, it does not require programmer to adjust module to any special requirements thus gives him full control and freedom.

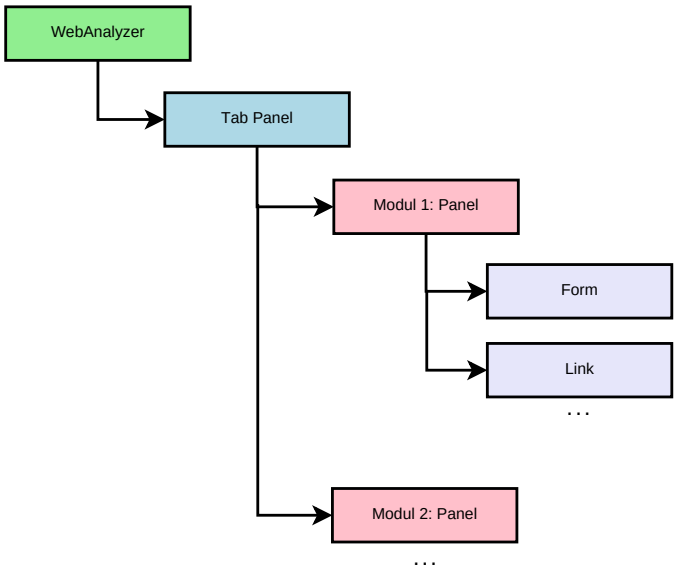


Fig. 2. WebAnalyzer architecture.

III. The BMAnalyzer application

Analyzer represents user front-end application of the BasicMeter tool. The role of the analyzer is to analyze the flow records collected by collecting process and to calculate and visualize the traffic parameters for the users. Traffic parameters can be measured for different purposes, such as traffic engineering, security analysis, usage based accounting, and many others. Therefore, several specialized analyzers exist. In this section, we describe the BMAnalyzer application (Fig. 3), which is designed for basic traffic analysis including visualization of observed network traffic in form of charts, tables, statistical data and others.

Description of individual parts:

- *INPUT* consists of modules for data reception from various sources.
- *BMA DB* is input module for communication with the database. It queries database and receives data for processing in the analyzer.
- *BMA ACP* is input module for communication with collecting process (JXColl) using the ACP (Analyzer Collector Protocol). It establishes connection with the JXColl and receives data for processing in the analyzer.
- *BMA ECAM* provides communication between *ECAM server* and *BMA Controller*. *ECAM* stands for Exporter Collector Analyzer Manager. It is module for management of all BasicMeter components.
- *BMAnalyzer Engine* forms the core of the application.
- *BMA filter* selects the necessary information elements and key attributes of processed flow records.
- *BMA data cache* collects selected information elements for *BMA* P_x modules.
- *BMA* P_x evaluates different traffic parameters P_x according to the required form of output.
- *OUTPUT* is the graphical interface, which serves for communication with the users.
- *BMA GUI* displays traffic parameters in desired way (chart, table, text, and others). It contains also interface for configuration of all BasicMeter components.
- *BMA Controller* receives settings from *BMA GUI*, configures all other BMAnalyzer components, and communicates with *ECAM*.

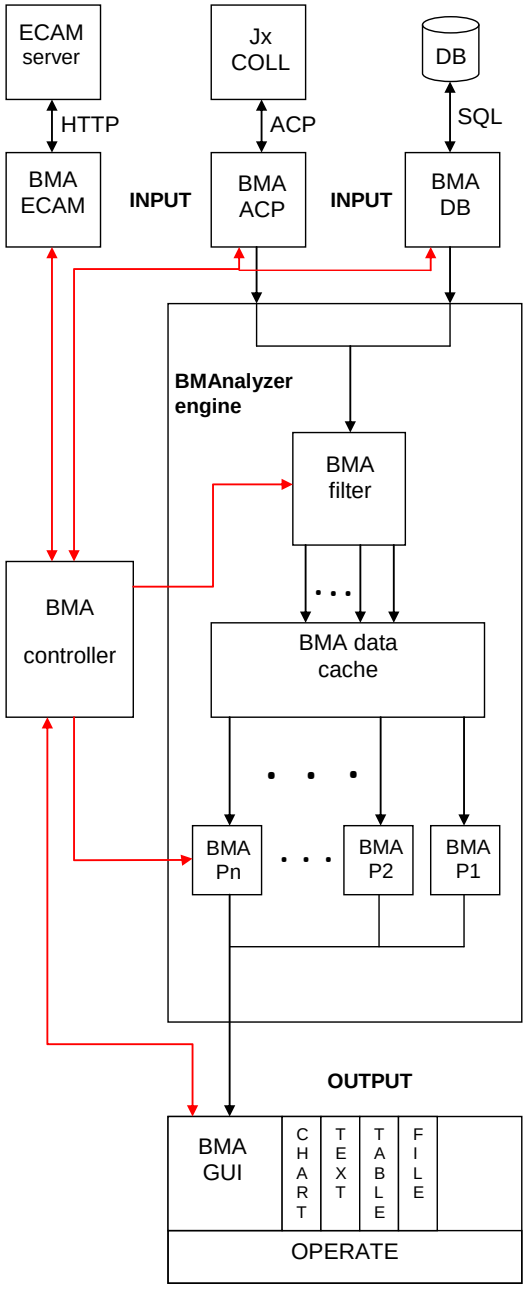


Fig. 3. Architecture of the BMAnalyzer.

The BMAnalyzer application is developed in Java programming language, so it becomes usable at various platforms. The program runs by opening the file `BMAnalyzer.jar` from folder `dist`. After starting the program displays its main window (Fig. 4).

The program allows to evaluate the amount of observed packets, octets and flows and visualize them using the charts. After pressing the *Create chart* button for creating a new chart, or the *Add plot* button for adding plots to the chart, filter window will be displayed (Fig. 5).

Filter allows to specify the type of data for evaluation by time, IP address and port. If time is selected, user has to enter start and end point of measured time range. The IP address can be chosen from the list of IP addresses captured from the network. It is also possible to enter the network address or the range of IP addresses. Selection of port numbers is similar to

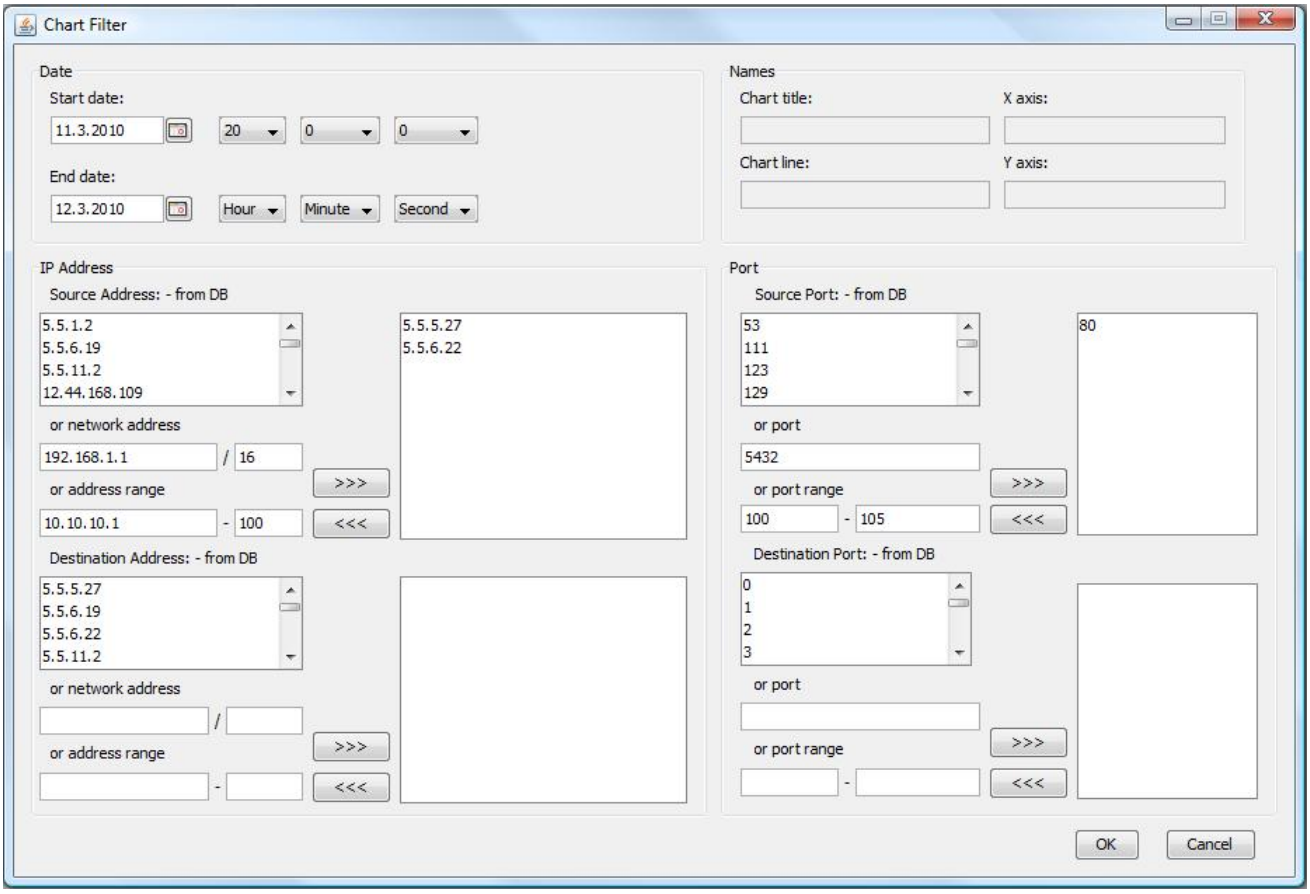


Fig. 5. Filter settings for adding plots to chart.

Fig. 4. The BMAnalyzer application.

selection of IP addresses. After the OK button is pressed, the choice is evaluated and request is sent to the database. Chart is created and shown in the analyzer window based on data received from database.

Current version of BMAnalyzer is capable to measure and visualize network traffic parameters such as byte rate, packet rate and flow count. Variation of selected traffic parameter in given time interval can be computed from user defined set of particular traffic flows. Byte rate plot is suitable for overview of utilization of network links and interfaces of devices. Packet rate plot can be useful to localize network invasive traffic sources. Using flow count monitoring it is possible to detect denial of service attacks, computer viruses and port scanning attempts.

IV. Conclusion

In this paper we introduced WebAnalyzer and BMAnalyzer which represent our approach to construction of analyzing application for network traffic monitoring. In the close future, WebAnalyzer will become a part of more complex system and will cooperate with more instances of BasicMeter over a network. We also plan to implement more features like a new interface and enhanced security.

We also plan to implement new features to the BMAnalyzer such as table output, statistical analysis and measuring of round trip time, which is important for performance analysis of network traffic interactive services. For more accurate selection of specific data from the database, filter will be extended. As a significant step we consider enabling real time monitoring of network traffic. To implement this part, ACP protocol will be used, which was created just for the purpose of direct communication between BMAnalyzer and JXColl. Finaly, BMAnalyzer will be integrated as a module to the WebAnalyzer, which will significantly increase its accessibility.

Acknowledgment

This publication is the result of the project implementation Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020)

supported by the Research & Development Operational Programme funded by the ERDF & was partially prepared within the project "Methods of multimedia information effective transmission", No. 1/0525/08 with the support of VEGA agency.

REFERENCES

[1] G. Sadasivan, N. Brownlee, B. Claise, and J. Quittek, "Architecture for IP Flow Information Export," RFC 5470 (Informational), Internet Engineering Task Force, Mar. 2009. [Online]. Available: http://www.ietf.org/rfc/rfc5470.txt

[2] J. Quittek, S. Bryant, B. Claise, P. Aitken, and J. Meyer, "Information Model for IP Flow Information Export," RFC 5102 (Proposed Standard), Internet Engineering Task Force, Jan. 2008. [Online]. Available: http://www.ietf.org/rfc/rfc5102.txt

[3] B. Claise, "Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information," RFC 5101 (Proposed Standard), Internet Engineering Task Force, Jan. 2008. [Online]. Available: http://www.ietf.org/rfc/rfc5101.txt

[4] F. Jakab, Ľ Košco, M. Potocký, and J. Giertl, "Contribution to QoS Parameters Measurement: The BasicMeter Project," *Conference Proceedings of the 4th International Conference on Emerging e-learning Technologies and Applications ICETA 2005*, vol. 4, pp. 371–377, 2005.

[5] T. Zseby, E. Boschi, N. Brownlee, and B. Claise, "IP Flow Information Export (IPFIX) Applicability," RFC 5472 (Informational), Internet Engineering Task Force, Mar. 2009. [Online]. Available: http://www.ietf.org/rfc/rfc5472.txt

[6] B. Stearns, M. Johnson, and I. Singh, *Designing Enterprise Applications with the J2EE(TM) Platform*. Reading, Massachusetts: Addison-Wesley, 2002.

[7] E. T. Freeman, E. Robson, B. Bates, and K. Sierra, *Head First Design Patterns*. Sebastopol, California: O'Reilly Media, 2004.

[8] M. Dashorst and E. Hillenius, *Wicket in Action*. Greenwich, Connecticut: Manning, 2008.

Free-space optical communication

Pavol Mišenčík

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

pavol.misencik@tuke.sk

***Abstract*— In today´s technological time, which is result of many inventions and discoveries of new technologies by users, but also service providers and network operators are increasingly imposed requirements for broadband Internet access. This paper presents the FSO system for transmission of large amounts of data.**

***Keywords*—Free space optics, modeling attenuation FSO, FSO channel modeling**

I. INTRODUCTION

Free Space Optics (FSO) is a fibreless, laser-driven technology that supports high bandwidth, with easy to install connections for the last mile access. Free Space Optics systems are starting to gain acceptance in the private market place as a solution to replace expensive fiber-optic based solutions. Optical wireless now allows services providers to cost-effectively provide optical bandwidth for networks, reducing Capex. Today the modern internet users are very much inclined towards the high bandwidth demanding applications, like video on demand, video conferencing, voice services, etc. FSO is a well-suited technology to make the high bandwidth of the backbone (Fiber network) available to the end user [2]. The main advantages of using FSO are that there is no licensing or tariffs for their utilization, there is no need to dig up roads, and they permit very high bit rates and thus a high bandwidth connection [4].

II. FSO TECHNOLOGY

FSO system, respectively optical communication free environment can be defined as a telecommunications technology that uses spread of light to transmit information between two points. It is a broadband telecommunications technology for line of sight, LOS technology which uses optical pulses modulated signals for wireless data transmission. Unlike fiber optics, where light pulses transmitted glass fiber the pulses of light transmitted through the atmosphere of a narrow beam, respectively free environment. FSO technology does this transfer using light beams instead of radio waves as the current wireless technologies. In other words, an FSO based optical communication lasers without optical fibers.

FSO system is based on optical wireless connectivity between units of the FSO. The basis of this FSO is FSO optical wireless unit, each unit consisting of the optical transceiver which provides full duplex communication, i. e. two-way communication. Each optical wireless unit uses an optical source with lens or telescope through which it transmits light free environment to the telescope or lenses other wireless units. These lenses, respectively telescopes are associated with highly-sensitive optical receiver via optical fiber.

FSO communication system is type point to point communication of two equal nicety optical transceiver mounted on a route with directly visibility. Usually these are transceiver mounted on roofs or windows of buildings and usually consist of the laser transmitter and optical detector as the receiver which is given full duplex communication. FSO systems operate in the order of tens of meters distance to several kilometers. FSO system operates in the infrared spectrum (IR) which adjoins of the visible spectrum. The human eye is invisible optical beams. It is therefore very important for the safety of using these systems. Operability of these systems in the frequency domain falls within the range of hundreds terahertz what analogy correspond to the operating wavelengths on the order of tenths of micrometers. Modern FSO systems operated in the infrared radiation wavelength range $0{,}75 - 0{,}85\ \mu m$ and $1{,}55\mu m$. In terms of safety and protection of the human eye is better choice for the application of laser working wavelength $1{,}55\mu m$. The wavelengt of $0{,}85\ \mu m$ optical beams to get across the cornea and lens on the retina which can occur in permanent damage to the eye. Conversely, using a system with a wavelength of $1{,}55\mu m$ optical rays are absorbed by the lens and cornea and thus there is no damage to the retina.

FSO systems provide transmisson speeds in the range of the order of hundreds of Mbps to several Gbps (100 Mbps, 155 Mbps and 622Mbps) There is commercially available transfer rate 2,7Gbps ,being tested speed 10Gbps. The aforementioned transfer rates giving rise to the transfer of large volumes of data at the time of friends, for example transmission of video, voice and various multimedia.

A. FSO principle

FSO principle is based on previous knowledge simply in the following five points:

1. required data coming from the network interface to the first FSO unit, where the digital signal will increase and converges to the optical signal,
2. broadcasting part of the first FSO unit by laser sends a signal through an optical lens or a telescope and then using the detector receives for optical signal transmission,
3. receiver part of other FSO unit accepts a beam of light through the lens or a telescope and then using detector the broadcast optical signal,
4. optical signal is obtained converges to the original electrical signal, amplifier is amplified and comes to the second interface of unit FSO,

5. inverse transfer, i.e. from the second unit to the first FSO unit is implemented in an identical manner – duplex transmission

To illustrate the points described above are shown in Fig. It should be noted that is it the simplified model of communication based on FSO technology, real communication take place in more difficult partial steps.

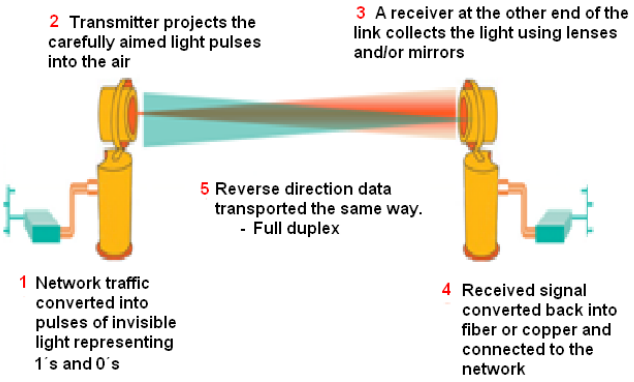


Fig.1 **FSO** principle [1]

B. *FSO advantages*

1) *Interference between devices*

There are not to interferences as on WI-FI card, because on the side of receiver is small point of the beam, so you can put around 7000 FSO facilities on 1kilometer square.

2) *Capturing data*

Capturing it is almost impossible for a small active range of the beam for the receiver. It would have to be used the same technology, moreover, the data going through this device can encode.

3) *Low latency*

FSO equipment has very low latency because it communicates the speed of light. It is also somewhat faster than the optical fiber because speed of fiber is limited by refraction of light.

4) *Improved speed*

Actual speed is 10 Mbps full-duplex. There are not delays (attenuation or waiting for waves) than with Wi-Fi. It has not speed control, so it can transmit data only 10 Mbps speed.

5) *The maximum distance range*

Distance is limited transmitter power and weather. The laser transmitter may reach several kilometers, but in severe fog loses the ability to transfer. Infrared wavelengths are better performance in fog, but not unlimited.

C. *Channel Modeling*

Power proposal for each communication system is heavily dependent on a precise understanding of the spread in the transmission channel. Selection of the appropriate modulation and coding is the most important for comprehensive model describing the channel attenuation and scattering characteristics of terrestrial FSO links. The aim of modeling channel is develop channel with similar properties will in fact be used to test the effectiveness of modulation. The proposal should take into account several system parameters. The effectiveness of land FSO links is depending mainly on climatic and physical characteristics of the selected areas. [1]

D. *Modeling attenuation*

Terrestrial FSO links have to deal with the atmosphere just above the Earth´s surface, where the maximum density due to gravitation forces. The atmospheric attenuation recognize attenuation molecular absorption, Rayleghiho scattering and aerosol scattering. Molecular absorption is a resonance effect of electrons and nuclei of atmospheric molecules. Aerosol scattering is caused by droplets and particles that are larger than the wavelength. Attenuation in the atmosphere FSO systems typically dominated by fog, but it is also affected by low clouds, rain, snow and the threshold of their various combinations. Molecular absorption can be minimized by appropriate choice of optical wavelength.

E. *Attenuation effects of rain*

Rain causes attenuation of optical signal. Rain drops are typically average about 0, 2 mm. Since the wavelengths are in the nanometer range, causing drops significantly less attenuation, much less then the fog. Scattering effects of rains drops is called non-selective because the size of the drops is much larger than the wavelength that causing effect independent of wavelength. [5]

Attenuation due to rain is given by:

$$\lambda_{rain} = 1{,}076.R^{\frac{2}{3}}\ dB/km \qquad (1)$$

where R is the rainfall rate in mm/hr. Figure 2. shows the simulation curves of rain attenuation.

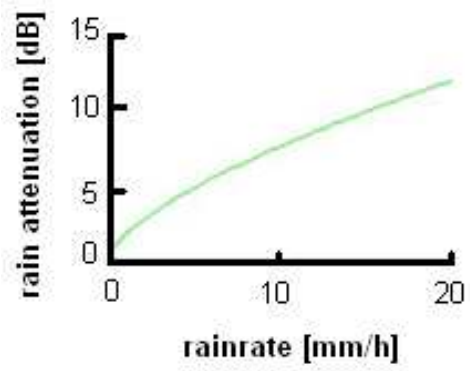


Fig. 2 Simulation curves of rain

F. *Attenuation effects of snow*

Snow causes attenuation depends on the wavelength optical signal. Attenuation due to snowfall was modeled as dry and wet snow and specific attenuation is given as follows:

$$\lambda_{snow} = aS^b\ dB/km \qquad (2)$$

, where S– snowfall rate in mm/hr
parameters a and b for the dray snow :
$a = 5{,}42.10^{-5}\lambda + 5{,}4959776;\ \ b = 1{,}38$
parameters a and b for wet snow :
$a = 1{,}023.10^{-4}\lambda + 3{,}7855466;\ \ b = 0{,}72$

L – introduced wavelength in nm

Figures 3 and 4 show the simulated attenuation for dry and wet snow. [3] In Figure 5 we see the difference between the attenuation characteristics of dry and wet snow.

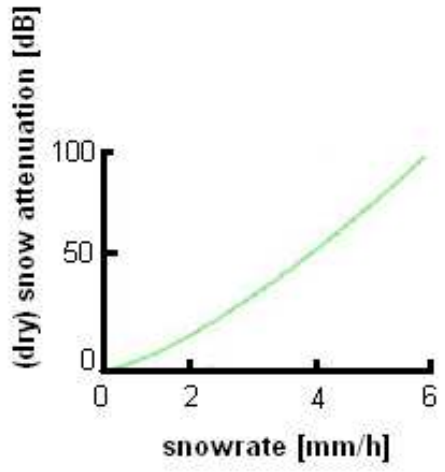


Fig.3 **Simulated** attenuation for dray snow

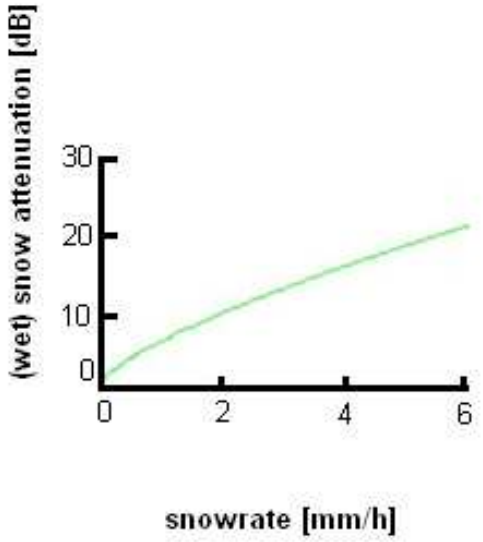


Fig.4 **Simulated** attenuation for wet snow

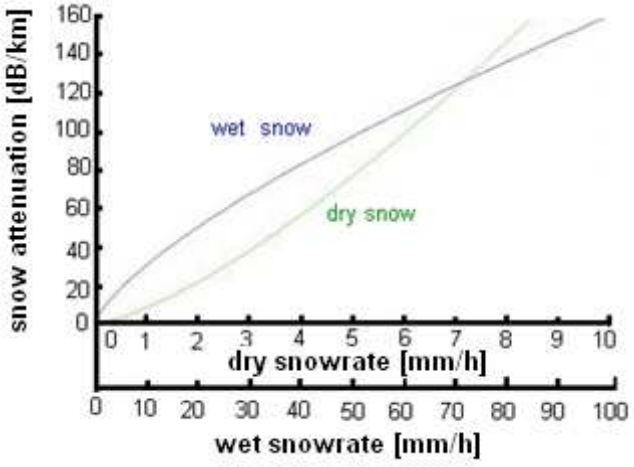


Fig.5 **Simulated** attenuation for wet and dry snow

III. Conclusion

Optical Wireless is an excellent nomadic broadband solution, supporting high bandwidth and services quality. This technology should be seen as supplement to conventional radio links and fiber optics. The use of low cost FSO-systems for private users. At the moment the main work in this field is to increase reliability and availability. Those two parameters of the FSO link are mainly determined by the local atmospheric conditions.

Acknowledgment

This work was partially supported from the grants VEGA No. 01/0045/10, project COST IC0802 and by Agency of the Ministry of Education of the Slovak Republic for the Structural Funds of the EU under the project Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020)

References

[1] S. Sheik Muhammad.: "Investigations in Modulation and Coding for Terrestrial Free Space Optical Links" (A dissertation submitted to the Faculty of Electrical Engineering and Information Technology): Graz University of Technology, Austria, March 2007

[2] S. Sheikh Muhammad - C. Chlestil – E. Leitgeb – M. Gebhart: "Reliable Terrestrial FSO Systems for Higher Bit Rates", In Proceeding at 8th International Conference on Telecommunications ConTEL, Zagreb, Croatia, 2005

[3] M. Mehra, M: " Free Space Optics : High bandwidth solution in network world" COIT 2007

[4] S. Sheikh Muhammad – P. Kohldorfer – E. Leitgeb: "Channel Modeling for Terrestrial Free Space Optical Links", IEEE ICTON 2005

[5] O. Koudelka – E. Leitgeb – S. Sheikh Muhammad: "Multilevel Modulation and Chnnel Codes for Terrestrial FSO links", IEEE 0-7803-9206-x/05 , 2005

Decentralized agent-based intrusion detection system with fully encrypted internal communication

[1]*Marián MIŽIK*

[1]Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]marian.mizik@tuke.sk

***Abstract*—Very important part of our daily lives are activities, requiring special level of private safety. Data security in any sort of eletronic communication is indispensable. Security, practicaly in any kind of software solution, is the more important, the more is system vulnerable to attacks of various kinds. Intrusion detection systems are one possible solutions for these problems. This kind of system is in most cases implemented as a primary security and monitoring center. Durability against failures and ability to regenerate from system errors is therefore problem worth research. This paper talks about basic decentralized architecture of intrusion detection system, focusing on minimalisation of single points of failure and supporting encryption of internal communication. It has ability to resize on the fly and ability to regenerate from error states without turning off.**

Keywords: **Network based intrusion detection system, distributed intrusion detection system, multicast, communication encryption, parallel computing**

I. INTRODUCTION

The research about intrusion detection systems using the agent based detection representative program approach can be separated into multiple levels. Some important and related researches are AAFID [4], DOS resistant IDS [5] mobile agent based IDS [6,7] and recoverable IDS approaches [8, 9]. Communication over multicast channel is occasionally used too [10, 11]

II. ARCHITECTURE

From a structural point of view, DIDS modules can be divided into two groups. Agents, which are monitoring the system, and logical modules, their work is detection of anomalies. In presented architecture, agent is an active part of system gathering information by monitoring selected environment. Stored information are analyzed and compared to default behavior tables. If there was some activity during monitoring, that was marked as suspicious, it is sent to other agents in the same network segment. Cooperation of agents on analyzing possible non default behavior of some activity has multiple pros. It prevents sending duplicate data to logical modules for deeper analyzing. If there is no success on agent level, there is no activity on logical module. Therefore, many unconfirmed possible attacks are removed before they are even deeply controlled by another layer of the system, so sharing of information between agents is decreasing amount of false alerts. If there is an agreement present in critical amount of detection rules, then these are moved to suspicious activity buffer prepare for deep analyses on logical module layer. Duty of this layer is to confirm, or deny the primary mark as a suspicious data, predetermination of importance, level of danger and appropriate reaction.

III. PROCEDURE OF ERROR STATE HANDLING

It is well known fact, that creating of false alert is normal behavior of every intrusion detection system. In case of distributed system, node that alarm is in false state and it needs to be corrected. In this architecture proposal, there is very important diversification in creation of false state into multiple levels. Result is then more optimized and resources usage split. Last but not least single points of failure are minimized. System regeneration is divided into two levels. Regeneration after wrong evaluation of monitoring data by agents (synchronization of agents) second level is recovery after wrong identification, or activity joining with some particular attack in logical IDS module (synchronization of logical modules) Following is important in case of agent synchronization. If the activity was finally evaluated as normal potential attack was cancelled. All agents that were reporting potential attack, will be synchronized with common knowledge. Of course possibility, that anomaly is located only in one segment of the network and can be detected only by small amount of agents is taken under consideration too.

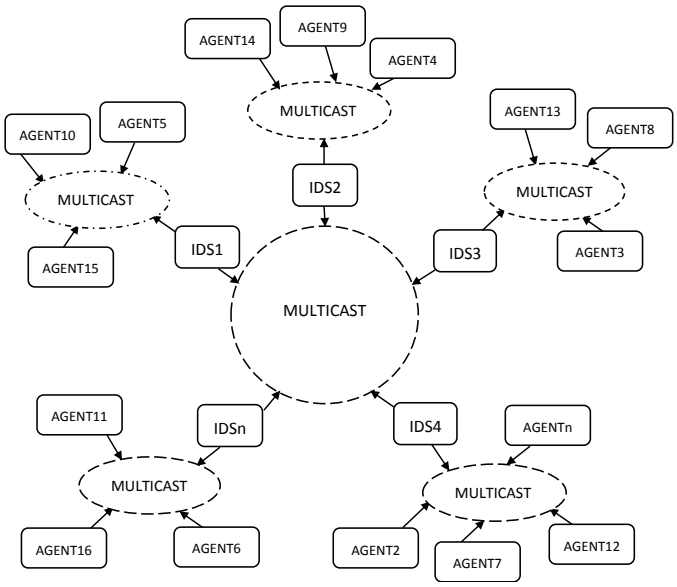


Fig. 1. System topology based on mentioned architecture

IV. INTERNAL COMMUNICATION

Internal communication is the key aspect in tendency to achieve milestones made at the beginning of his paper. Whole communication is based on hybrid encryption. Logical module is taking care of encryption key distribution for all agents that are signed to it. Communication between agents, will be implemented as a IP Multicast ring. Communication between logical modules will be implemented in the same way as IP Multicast ring(Fig 1). Initial encryption keys are built into the module during installation process. Every datagram inside of internal communication is carrying unique byte (or set of bytes) in its header. Thanks to this approach are all datagrams that belong to internal communication better recognizable from other communication inside the network. There is lower chance to attempt to decipher the invalid content, and the marked diagram is in most cases immediately discarded. Other uses are offered directly by the proposed DIDS, when a private communication between a group of agents, although it is available to them superior ITS junction, will not be taken into account by this junction as the unique byte is different for communication between the logical IDS nodes and data flow between sensors. As it has been said earlier, hybrid encryption is used for whole communication. After asynchronous key exchange, the symmetric encryption is used, and the communications may/can be monitored by potential attackers, the cipher used is changed in regular intervals. Another catalyst for the launch of a new redistribution cipher between the various elements of the proposed DIDS is the detection of a possible attack on one of the active nodes of the system. In this case before updating the communication is interrupted with the compromised element, whereas the already used one could be compromised.

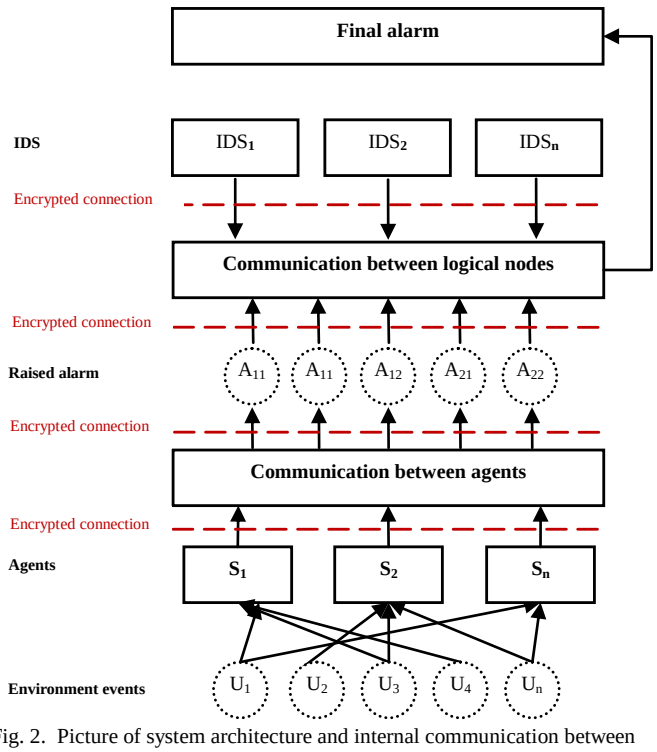


Fig. 2. Picture of system architecture and internal communication between single system modules.

V. INITIALIZATION OF NEW NODES

Initialization of new agent and its connection to internal communication works as follow. . Encryption key pair Pk and Vk is already present in basic installation package, as well as multicast channel identifier is present. First action is initialization request sent to default multicast channel. All logical modules can decrypt this message because keys are part of their installation. Node with smallest amount of nodes is chosen to continue the operation. Chosen node then send current cypher for communication between agents related to this logical node, bit sequence needed for communication with logical IDS nodes and bit sequence needed for communication between agents. These bit sequences are added to every datagram header. Last thing that agent receives is identifier of multicast channel for inter-agent communication. New agent responds to its superior logical IDS node confirming message encrypted with received encryption key and special bit sequence in every datagram. Superior node is sending list of all active agents as a response. After this action, initialization phase is successfully completed.

Connection of logical module to DIDS network, logon procedure and verification of incoming node is divided into multiple phases. Encryption key pair Pk and Vk is already present after basic installation, it also possess second pair of encryption keys for communication with its future agents and channel identifier of multicast. Node, that is performing logon action will send a request to whole multicast group. This request is encrypted with Vk key. Every member of multicast group will respond its own identifier. (encrypted with initialization public key Vk) Received list of IDs is sent back to multicast group. This procedure is necessary to confirm, if new node understand the communication, therefore is able to decrypt and encrypt data. If it is not, it is an attacker trying to use random packets, or data sniffed from previous logon attempts. After this security precaution will node with highest uptime send current cipher for synchronous communication inside the multicast group. Node will also receive new multicast channel identifier, where this communication is performed, because this channel was only channel for initialization. New node is then sending confirmation message to new channel encrypted with new key, informing others, that everything went as expected.

VI. CONCLUSION

Every layer of proposed DIDS is created from equal nodes. This information is important when taking to consideration, that every node in the architecture is redundant. Failure or absence of any node is not crucial for runtime. Agent nodes are able to move to replace other dead or attacked agent module module.

This paper presents architecture proposal of distributed intrusion detection system with focus on effective resource handling, security and encryption of internal communication and error state regeneration without need of system restart or

shut down. Architecture has minimized most of single points of failure, therefore it is showing better stability results than standard architectures. Part of responsibilities were moved to agents to lower system requirements as a result of faster detection of some false states.

ACKNOWLEDGMENT

This work was supported by the Slovak Research and Development Agency under the contract No. APVV 0073-07 and VEGA grant project NO. 1/0026/10.

REFERENCES

[1] Snort: *The open source network intrusion detection system*. http://www.snort.org.

[2] Ing. Martin Chovanec. *Distribuovaný systém detekcie prienikov využitím multisenzorovej fúzie dát*. PhD thesis, Košice, SR, 2008.

[3] Ramaprabhu Janakiraman, Marcel Waldvogel, Qi Zhang. Indra: *A peer-to-peer approach to network intrusion detection and prevention*. WETICE 2003: 226-231

[4] Jai Sundar Balasubramaniyan, Jose Omar Garcia-Fernandez, David Isacoff, Eugene Spafford and Diego Zamboni, *An Architecture for Intrusion Detection using Autonomous Agents*, COAST Technical Report 98/05, Purdue University, 1998.

[5] Perter Mell, Donald Marks and Mark McLarnon, *A denial-of-service resistant intrusion detection architecture*, Computer Network, Special Issue on Intrusion Detection, Elsevier Science BV, 2000.

[6] Kruegel et al., *Applying Mobile Agent technology to Intrusion Detection. Distributed Systems Group*, Technical University of Vienna, 2002

[7] P.C.Chan and V.K.Wei, "*Preemptive distributed intrusion detection using mobile agents*", in Proceedings of Eleventh IEEE International Workshops on Enable Technologies: Infrastructure for Collaborative Enterprises, Jun, 2002.

[8] E. Amoroso and R. Kwapniewski, *A selection criteria for intrusion detection systems*, in Proceedings of 14th Annual Computer Security Applications Conference, Phoenix, USA, Dec 1998.

[9] A. Birch, *Technical evaluation of rapid deployment and re-deployable intrusion detection systems*, in Proceedings of IEEE, 1992, International Carnahan Conference on Security Technology, Atlanta, USA, 1992.

[10] Sartid Vongpradhip, Ph.D. and Wichet Plaimart. *Survival Architecture for Distributed Intrusion Detection System (dIDS) using Mobile Agent.* Sixth IEEE International symposium on network computing and applications (NCA 2007)

[11] I-Hsuan Huang and Cheng-Zen Yang , *Design of an Active Intrusion Monitor System*, 0-7803-7882-2/03/$17.0002003 IEEE

Packet loss modeling

[1]*Ján MOCHNÁČ,* [1]*Pavol KOCAN,* [1]*Branislav HRUŠOVSKÝ*

[1]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]jan.mochnac@tuke.sk, pavol.kocan@tuke.sk, branislav.hrusovsky@tuke.sk

***Abstract*—Video transmitted over unreliable environment, like wireless channels or in generally any network with unreliable transport protocol, is subject to packet loss due to network congestion and channel noise. Therefore it is necessary to study impact of packet loss to obtain knowledge about effect on the video quality. Packet losses can be modeled through Markov chains based network models. In our paper we have analyzed in which way has occurred packet losses during transmission of video through wired network and we have computed parameters of the Gilbert model.**

***Keywords*—Gilbert loss model, ns-2, video streaming.**

I. Introduction

Most of the Internet traffic is controlled by the Transmission Control Protocol (TCP) protocol, which uses retransmission to control loss rates. However, TCP is not suitable for real time multimedia applications while it introduces some undesirable delay.

Another widely used transport protocol is the User Datagram Protocol (UDP). It is also unsuitable for multimedia transmission, while UDP does not guarantee ordered delivery of packets. Therefore the Real-time Transport Protocol (RTP) protocol based on UDP is widely used for multimedia streaming applications.

Packet loss is the main factor which degrades the visual quality of video content transported through networks based on IP protocol. Content of such real-time multimedia application should be delivered without any significant packet loss and with low delay.

Video transmitted over wireless environment, or in generally any network with unreliable transport protocol, is facing the losses of videopackets due to network congestion and noises of different kinds. By using highly efficient videocodecs problem is becoming more important. Visual quality degradation could propagate to the subsequent frames due to redundancy elimination in order to gain high compression ratio. Therefore it is necessary to know in which way the packets are lost and one of the possible ways to learn about losses is creation of networks model.

The impact of packet loss can be studied from recorded measurement traces of traffic and loss patterns. To generate error process with similar characteristics as observed in measurements, stochastic model can be modeled [1]. The most popular examples of such models are discrete-time Markov chain models. The use of discrete-time Markov chain models has been proposed in [2]. Discrete-time Markov chain models of increasing levels of complexity, including the 2-state Markov chain model have been described in [2], [3].

Obviously, Gilbert model is simple but its major drawback is inability to correctly model heavily tailed error runs. In such cases Hidden Markov models with up to five states are used to model the distribution of error and error-free burst lengths [4].

In our paper we have transmitted video through wired network and after that we have constructed Gilbert model for our network topology from recorded packet sequence numbers.

II. Packet Loss Modeling

In order to evaluate quality of transmission we have random variable *X*. If packet is not lost then *X=0*, otherwise *X=k* for *k* lost packets. After that we can build loss model with infinite number of states (m is infinite value. Such model gives us opportunity to model packet loss probabilities in dependence on burst lengths (several consecutively lost packets). For every additional lost packet which adds to the length of a loss burst a state transition takes place. If packet is correctly received, then the state returns to *X=0* [4].

State probability for system with $k > 0$ is $P(X \geq k)$.For finite number of received packets a, state probabilities of the system with $k > 0$ can be approximated with cumulative loss rate [4]:

$$p_{L,cum}(k) = \sum_{n=k}^{\infty} p_{L,n} \tag{1}$$

Cumulative loss rate for *k=0*, thus for no loss case, can be computed according the following equation:

$$p_{L,cum}(k=0) = 1 - \sum_{k=1}^{\infty} p_{L,cum}(k) = \tag{2}$$
$$1 - \sum_{k=1}^{\infty} \frac{\sum_{n=k}^{\infty} o_n}{a} = 1 - \sum_{k=1}^{\infty} \frac{k o_k}{a} = 1 - p_L$$

where o_k (o_n) is occurrence of loss with length k (n).

A. Gilbert model

Packet loss measurements on the Internet have shown that the probability of loss episodes of length *k* decreases approximately geometrically with increase of *k* [3]. Thus it is possible to use simpler packet loss model, e.g. Gilbert model.

Special case of *k*-th order Markov chain model is Gilbert model with *k*=2.

In this model, *0* represents state with no packet loss and on the other hand *1* represents the state of packet being lost.

The matrix for transition probabilities and for state probabilities can be expressed in form:

$$\begin{bmatrix} 1-p_{01} & p_{10} \\ p_{01} & 1-p_{10} \end{bmatrix} \begin{bmatrix} P(X=0) \\ P(X=1) \end{bmatrix} = \begin{bmatrix} P(X=0) \\ P(X=1) \end{bmatrix} \tag{3}$$

For unconditional probability P(X=1) holds the following equation:

$$P(X=1) = \frac{p_{01}}{p_{01}+p_{10}} \tag{4}$$

If previous packet is lost, then for conditional probability of having loss holds:

$$P(X=1|X=1) = 1 - p_{10} \tag{5}$$

Gilbert model memorizes only the previous state, thus the probability the next packet will be lost is dependent only on the previous state.
Transition probabilities p_{01} and p_{10} can be expressed with the following equations:

$$p_{01} = P(X=1|X=0) = \sum_{k=1}^{\infty} \frac{o_k}{a} \tag{6}$$

$$1 - p_{10} = P(X=1|X=1) = \frac{\sum_{k=1}^{\infty}(k-1)o_k}{d-1} \tag{7}$$

The probability of having a loss episode with length *k* [3]:

$$p_k = (1-p_{10})^{k-1} p_{10} \tag{8}$$

III. Experimental results

In experimental part of our work, we have transmitted video sequences through a fixed network using RTP protocol. Therefore we have modeled network topology with 12 computers and one video server in The Network Simulator–ns-2. The video server was source of video sequences, one PC was receiver of video traffic and other PCs have introduced some background traffic using FTP and CBR agents. The data rate used by this agents was 0.5Mbs and it was tried to transmit several flows in every node (1–6 flows). As a transport protocol RTP was used.

We have used Akiyo, Foreman, Mobile and Stefan video sequences with CIF resolution, and sequences created from previous sequences with length up to 1h40min.
The video sequences were coded using MPEG-4 codec, after that they were streamed with VLC using MPEG transport stream and captured with rtptools in order to obtain video in RTP packet format.
This converted files were transmitted through the network topology made in ns-2, each video sequence at least ten times. At the end of simulation we have obtained file with order numbers of transmitted packets and received packets. After that MATLAB has been used to compute transition probabilities and also unconditional probabilities as an average of partial results for every video sequence with standard length 300 frames and Gilbert model was made:

$$\begin{bmatrix} p_{00} & p_{10} \\ p_{01} & p_{11} \end{bmatrix} = \begin{bmatrix} 0.91537 & 0.79802 \\ 0.084629 & 0.20198 \end{bmatrix} \tag{9}$$

$$\begin{bmatrix} P(0) \\ P(1) \end{bmatrix} = \begin{bmatrix} 0.904179 \\ 0.095821 \end{bmatrix} \tag{10}$$

This process was repeated, but for background traffic were used more CBR streams (up to six). Again, transition probabilities and unconditional probabilities were computed as an average of partial results for every video sequence:

$$\begin{bmatrix} p_{00} & p_{10} \\ p_{01} & p_{11} \end{bmatrix} = \begin{bmatrix} 0.80151 & 0.77453 \\ 0.19849 & 0.22547 \end{bmatrix} \tag{11}$$

$$\begin{bmatrix} P(0) \\ P(1) \end{bmatrix} = \begin{bmatrix} 0.796 \\ 0.204 \end{bmatrix} \tag{12}$$

The conditional probabilities of loss $p_{11} = 0.20198$ in the first case and $p_{11} = 0.22547$ for second case are larger than the unconditional probabilities $P_1 = 0.095821$ and $P_1 = 0.204$ as it has been shown in [3]. These results also confirm that loss has appeared in bursts.
Interesting fact is also the length of loss episodes. In our network topology, the longest loss episode includes ten packets but the most frequent loss episode takes only two packets. Packet losses were concentrated mostly to loss episode with two packets. There were only several loss episodes with 10 packets.
After adding more background traffic the loss episodes have kept their allocation, the difference was in frequency of packet losses. As it was expected, more CBR streams led to bigger packet losses.

IV. Conclusion

In this paper, we have studied packet losses during transmission of video through a fixed network.
Even though the most used protocol for data transmission is TCP, RTP is more suitable for multimedia delivery, while it supports jitter compensation and out of sequence arrival detection, which have occurred during the transmissions on an IP network. Therefore we have used RTP as a transmission protocol in our ns-2 simulations.

To study impact of packet loss, packet loss models have been proposed in the literature. We have decided to use Gilbert model as a loss model, for which we have computed transition probabilities and unconditional probabilities. It is obvious from our results that losses have appeared in bursts as was expected. The longest packet loss episode takes ten packets, but the most frequent are episodes with two packets.

Acknowledgment

The work presented in this paper was supported by Grant of Ministry of Education and Academy of Science of Slovak Republic VEGA under Grand No. 1/0045/10.

This publication is the result of the project implementation Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] G. H. nad O. HOHLFELD, "The Gilbert-Elliott Model for Packet Loss in Real Time Services on the Internet." *Proceedings of the 14th GI/ITG Conference on Measurement, Modeling, and Evaluation of Computer and Communication Systems (MMB)*, pp. 269–283, 2008.

[2] M. YAJNIK, S. MOON, J. KUROSE, and J. TOWSLEY, "Measurement and modeling of the temporal dependence in packet loss," *Proceedings of Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM '99*, vol. 1, pp. 345–352, 1999.

[3] V. MARKOVSKI, F. XUE, and L. TRAJKOVIC, "Simulation and analysis of packet loss in video transfers using User Datagram Protocol," *The Journal of Supercomputing, Kluwer Academic Publishers*, vol. 20, pp. 175–196, 2001.

[4] H. SANNECK, G. CARLE, and R. KOODLI, "A framework model for packet loss metrics based on loss runlengths," *Proceedings of the SPIE/ACM SIGMM Multimedia Computing and Networking Conference 2000 (MMCN 2000)*, pp. 177–187, 2000.

Time Basic Nets: Time properties and the Reachability Problem

[1]*Attila N.Kovács*, [2]*Iveta Adamuščínová*

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]attila.n.kovacs@tuke.sk, [2]iveta.adamuscinova@tuke.sk

***Abstract*—Nowadays, lots of different kinds of Petri nets exist in which time aspect is incorporated. The most popular Petri net model for specification and verification of real-time systems are Time Petri nets. For analytical purposes of Petri nets the most suitable method is the reachability analysis. An important issue for time-critical systems is the end-to-end delay derivation in task execution. The CS-Class technique for reachability problem solution takes this issue into account for Time Petri nets. In this paper the CS-Class approach of Time Petri nets is applied to Time Basic Nets. Time Basic Nets represent the upper class of Time Petri Nets.**

***Keywords*—CS-Classes, Reachability problem, Time Basic Nets, Time Petri Nets.**

I. INTRODUCTION

Real-time systems, such as patient monitoring systems, aircraft control systems or traffic control systems, are very common in our everyday life. Even the smallest failure in such a system can cause enormous damages or loss of human lives. That's why these systems must be carefully and precisely verified.

Several extensions of Petri nets with incorporated time issue have been already proposed [1], [2], [3], [4], or [5]. The most used are Time Petri net, Stochastic Petri nets and Time Basic Nets. Time Petri nets are capable of modeling real-time systems but just a specific part of them. Time Basic Nets on the other hand have all the advantages of Time Petri nets plus they have a bigger modeling power.

II. TIME PETRI NETS

A. Basic Definitions

Time Petri Nets (TPNs) are Petri Nets where to each transition a static time interval (SI) is assigned [1] – [4]. The smallest time value of these time intervals is called Static Earliest Firing Time (SEFT) and the largest time value is called Static Latest Firing Time (SLFT). The Static Firing Interval of the transition will be the closed left bounded interval of times comprised between its SEFT and SLFT. For two time intervals $I_1 = [u_1, v_1]$ and $I_2 = [u_2, v_2]$ with $0 \leq u_i \leq v_i \leq +\infty$ we define $I_1 + I_2 = [u_1 + u_2, v_1 + v_2]$ and $I_1 - I_2 = [u_1 - u_2, v_1 - v_2]$.

A *state* in TPNs is a pair $S=(m, I)$ where m is a marking and I is a firing interval set (function) which associates with each enabled transition the time interval in which the transition is allowed to fire.

From the initial state a new state can be reached by a given sequence of firing times corresponding to a firing sequence. Since all time intervals assigned to transitions consist of real numbers the number of reachable states produced by the firing of a single transition is infinite. To handle this problem a *state class* is introduced.

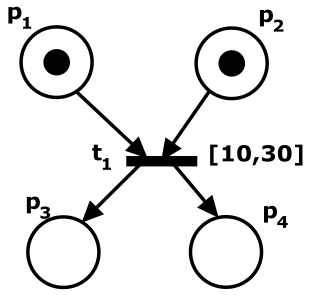


Fig. 1. A simple TPN

A state class represents all states reachable from the initial state by firing all feasible firing values corresponding to the same firing sequence. More formally, a state class is a pair $C = (m, D)$ in which m is the marking of the class and D is the firing domain of the class, which is defined as the union of the firing domain of all the states in the class. All states in the class have the same marking. A transition t is firable from class $C = (m, D)$ if t is enabled by marking M, and may fire before the minimum of all LFT's related to all enabled transitions. Firing rules in detail can be seen in [3] and [4].

B. Clock Stamped State Classes of Time Petri Nets

As we will see later, clock stamped state classes are very helpful in those cases, when we want to find the answer to the question, whether some process or action ends its execution until a specified time [3].

A clock stamped state class (CS-class) is a 3-tuple $C = (m, D, ST)$ where m is a marking; D is a firing domain, i.e., a set of constraints on the values of the time to fire for transitions enabled by current marking m. $D(t_i)$ represents the firing interval of an enabled transition t_i. The left bound of $D(t_i)$ is denoted as $EFT(t_i)$ (earliest firing time) and the right bound of $D(t_i)$ is denoted as $LFT(t_i)$ (latest firing time); ST represents the (global) time interval of the CS-class.

For an enabled transition t_i, $D(t_i)$ gives the global firing time interval of t_i. The word "global" means a relative counting of values to the beginning of the net's execution from the initial CS-class C_0. The initial CS-class is defined as $C_0 = (m_0, D_0, ST_0)$ where m_0 is the initial marking, D_0 contains all the static firing time intervals of the transitions enabled in

M_0, and ST_0 = [0, 0]. ST represents the global time delay interval in which the net runs from C_0 to current CS-class C.

The following firing rules guide the generation of all reachable CS-classes of a TPN. An enabled transition t_j is said to be firable at CS-class C_k if $EFT_k(t_j) \leq min\{LFT_k(t_i),\ t_i \in E(C_k)\}$, where $E(C_k)$ is the enabled set at C_k. Let $Fr(C_k)$ be the set of firable transition at CS-class C_k, and let

$$MLFT(C_k) = \min\{LFT_k(t_i), t_i \in Fr(C_k)\}. \quad (1)$$

where $MLFT(C_k)$ defines the minimum of latest firing times of all firable transitions in $Fr(C_k)$. The firable transitions in $Fr(C_k)$ can be divided into two groups: a) *inherited firable transitions* thaht were firable before C_k is reached and b) *new firable transitions* that begin firable at C_k. The firing of transition $t_f \in Fr(C_k)$ changes the CS-class to C_{k+1}. If CS-class $C_k = (m_k, D_k, ST_k)$ and $C_{k+1} = (m_{k+1}, D_{k+1}, ST_{k+1})$ then the following steps define transition firing rules:

1) Calculate $D_k(t_f)$, the feasible firing intervals of the firing transition t_f, by shifting right bound of $D(t_f)$ to MLFT(C_k) while keeping its left bound unchanged, i.e.,

$$D_k(t_f) = [EFT_k(t_f),\ MLFT(C_k)], ST_{k+1} = D_k(t_f). \quad (2)$$

2) The calculation of firing intervals of inherited firable transitions in CS-class C_{k+1} can be done following ways:
   a) Let $m'_{k+1} = m'_k - B(t_f)$ and collect (inherited) firable transitions at m'_{k+1}. Function $B(t_f)$ is responsible for removing tokens from input places of transition t_f.
   b) Let $D_{k+1} = D_k$ and delete from D_{k+1} all entries whose corresponding transitions are disabled by m'_{k+1}.
   c) For each inherited firable transition t_j ($t_j \neq t_f$) at m'_{k+1}, let

$$EFT_{k+1}(t_j) = \max(EFT_k(t_j), EFT_k(t_j)). \quad (3)$$

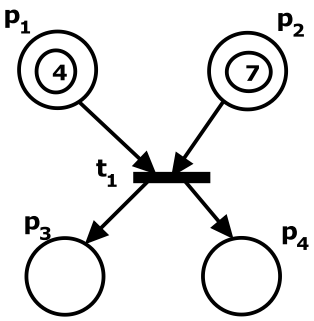
3) Calculate the firing intervals of new firable transitions after firing t_f:
   a) Let $m_{k+1} = m'_{k+1} + F(t_f)$ and collect new firable transitions. These transitions are firable in m_{k+1} but not in virtual marking m'_{k+1}. Function $F(t_f)$ is responsible for adding tokens to the output places of transition t_f.
   b) Add into D_{k+1} entries that corresponding new transitions at m_{k+1}: if t_j ($t_j \neq t_f$) is new firable transition at m_{k+1}, then

$$D_{k+1}(t_j) = SI(t_j) + ST_{k+1}. \quad (4)$$

   c) If t_f is still firable at m_{k+1} after its own firing, then

$$D_{k+1}(t_f) = SI(t_f) + ST_{k+1}. \quad (5)$$

Formal proofs and examples for the above mentioned approach can be found in [3]. In Section IV a modified version of this approach will be used to generate the reachable state classes of Time Basic Nets.



$$tf_{t_1}(p_1, p_2) = max(p_1, p_2) + 10 \leq max(p_1, p_2) + 30$$

Fig. 2. A simple TB net

III. TIME BASIC NETS

Time Basic nets (TB nets) are a particular case of Time Environment Relationship nets (TER nets) [8]. When we assume that the only types of tokens in TER nets are time values (chronos) then we get TB nets .TB nets have been introduced in [2].

A. Basic Definitions

A TB net can be characterized as a 6-tuple where P, T and F are, respectively, the sets of places, transitions, and arcs of nets. The preset of transition t, i.e., the set of places connected with t by an arc entering t, is denoted by ${}^\bullet t$. Symbol Θ (a numeric set) is the set of values (timestamps), associated with the tokens. A timestamp represents the time at which the token has been created. In the following, we assume Θ to be the set of non-negative real numbers, i.e., time is assumed to be continuous. Function tf associates a function tf_t (called time-function) with each transition t. Let *enab* be a tuple of tokens, one for each place in ${}^\bullet t$. Function tf_t associates with each tuple *enab* a set of value θ ($\theta \subseteq \Theta$), such that each value in θ is not less than the maximum of the timestamps associated with the tokens belonging to *enab*. At this moment we can define the enabling tuple, enabling time and the firing time.

Given a transition t and a marking m, let *enab* be a tuple of tokens, one for each input place of transition t. If $tf_t(enab)$ is not empty, *enab* is said to be an enabling tuple for transition t and the pair $x = <enab, t>$ is said to be an *enabling*. The triple $y = <enab, t, \tau>$ where $<enab, t>$ is an enabling and $\tau \in tf_t(enab)$, is said to be a firing. τ is said to be the firing time. The maximum among the timestamps associated with tuple *enab* is the enabling time of the *enabling* $<enab, t>$.

Firing occurrences, which ultimately produce firing sequences, define the dynamic evolution of the net (its semantics); markings represent the states and transitions represent events of the modeled system.

The following axioms must hold in TB nets: time never decreases; if the system does not stop, time eventually progresses. More axioms for TB nets can be found in [2].

In TB nets we can distinguish two time semantics: weak and strong time semantics. At this point we will describe the advantages and disadvantages both of them.

B. Weak and Strong Time Semantics

As it was mentioned above two time semantics can be considered in TB nest. These time semantics are weak time semantics (monotonic weak time semantics - MWTS) and strong time semantics. The following axioms hold for both time semantics.

Axiom 1: All the times of the firings of an MWTS firing sequence σ must be no less than any of the time stamps of the tokens of m_0.

Axiom 2: All the times of firings of an MWTS sequence σ are monotonically nondecreasing with respect to their occurrence in σ.

Axiom 3: For all $\sigma \in \Theta$, there exists *k*, $k \geq 0$, such that all firing sequences with at least *k* firings contain at least one firing whose time is greater than *τ*, i.e. the number of firings that can occur within a given time interval is bounded.

Axiom 1 requires that all firings must occur not earlier than the times associated with the tokens in the initial marking m0. Axiom 2 describes the monotonicity of the occurrences of firings in the sequence with respect to their firing times. Axioms 1 and 2 capture the fact that time never decreases. Axiom 3 states that if the system does not stop, time eventually progresses, or in other word, there exist no infinitely long firing sequences that take a finite amount of time. This property is often required in real-time system models [6].

For strong time semantics (STS) the following two axioms must hold in addition:

Axiom 4: No enabling tuple exists in the initial marking m_0 whose maximum firing time is less than maximum of the timestamps associated with the tokens in m_0. In this case the marking m_0 is called strong initial marking.

Axiom 5: Let *σ* be a monotonic weak firing sequence of a TB net with a strong initial marking. $\sigma = \langle y_1, y_2, \ldots, y_i, \ldots \rangle$ is a strong firing sequence if and only if for each transition *t* and for each reachable marking m_i, $1 \leq i$, there exists no tuple *enab* enabling transition *t* in m_i such that the time of firing y_{i+1} is greater than all the firing times of *t* under tuple *enab*.

C. Time Interval Semantics

Instead of time point semantic a more powerful time semantic can be used [7].

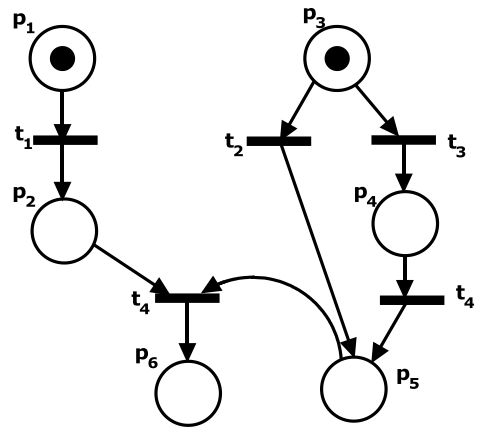
Interval semantics of TB nets give us the opportunity to assign a time interval (TI) to each token. This time interval specifies the time values in which the tokens can be created. Using TI instead of timestamps gives us a bigger modeling power. Any token (chronos) *τ* in TI is considered to be a TI $\tau = [\tau_i, \tau_a] \subseteq \mathbb{R}^+$, where $\mathbb{R}^+ = [0, +\infty]$. In TI semantics we replace any enabling tuple $enab = (m(p_1)...m(p_{|\bullet t|}))$ with a corresponding collection of TIs that is called *Time Interval Profile* (TIP). Besides the set operation $\cap$, $\cup$, $(\,)^c$ a new operation "+" is defined. For a given constant $c \in \mathbb{R}^+$ and TI $\tau = [\tau_i, \tau_a] \subseteq \mathbb{R}^+$ we have: $c + \tau = [\tau_i + c, \tau_a + c]$, $c \,.\, \tau = [\tau_i \,.\, c, \tau_a \,.\, c]$. For TIs τ' and τ'' we have:; $\tau' + \tau'' = \tau \Leftrightarrow \tau = [\tau_i, \tau_a]$, $\tau' = [\tau'_i, \tau'_a]$, $\tau'' = [\tau''_i, \tau''_a]$, $\tau_i = \tau'_i + \tau''_i$, $\tau_a = \tau'_a + \tau''_a$.

Given TB net $N_0 = (P, T, \Theta, pre, post, tf, q_0)$, then tft(enab) has for a given *enab* the unique representation

$$tf_t(enab) = \tau en + tf_t(0) \qquad (6)$$

where *τen* is a TI that depends on *enab*, $tf_t(0)$ is a TI that does not depend on *enab*. To put it another way, any t-generated TI τ_t can be represented as a sum of two TIs: τen-the determinate TI that depends on TIP *enab* in question and on *t* (or tf_t) and a constant TI $tf_t(0)$, which depends only on the structure of the TB nets in question.

According to the above mentioned unique representation of *enab* some interesting features can be found in [7] and [9].



$$m_0(p_1) = [0], m_0(p_1) = [0]$$
$$tf_{t_1}(p_1) = p_1 + 30 \leq p_1 + 50$$
$$tf_{t_2}(p_3) = p_3 + 10 \leq p_3 + 70$$
$$tf_{t_3}(p_3) = p_3 + 40 \leq p_3 + 90$$
$$tf_{t_4}(p_4) = p_4 + 20 \leq p_4 + 40$$
$$tf_{t_5}(p_2, p_5) = max(p_2, p_5) + 10 \leq max(p_2, p_5) + 30$$

Fig. 3. TB net with synchronization and concurrency

IV. CS-Class Approach in Time Basic Nets

As it was mentioned earlier, reachability problem for time-critical systems is quite different then for ordinary systems. Several researchers tried to solve this crucial problem [7] – [12]. Unfortunately, no general solution of this problem exists for Time Basic Nets.

For the TB net shown on Fig. 3 we will apply the generation rules introduced in section II. For this TB net a strong time semantic is used.

To use the generation rules we simply replace the names of the places in all time functions with concrete time values. The time values (time intervals) are written in square brackets, i.e. the marking m_0=([0],0,[0],0,0,0) shows, that the place p_1 and p_3 has one token with time value zero (marked as [0]) and place p_2, p_4, p_5 and p_6 has no tokens.

In TB nets the (global) time interval is not used in the same way as in the CS-Class of TPNs. In TB nets the (global) time interval represents the same time value as the time (time interval) of the newly created tokens.

For the TB net shown on Fig. 3 the initial CS-Class is C_0 = (m_0, D_0), where

$$ST_0 = [0, 0],$$
$$m_0 = ([0], 0, [0], 0, 0, 0),$$
$$D_0 = \{D_0(t_1) : [30, 50], D_0(t_2) : [10, 70], D_0(t_3) : [40, 90]\}. \qquad (7)$$

From the initial CS-Class C_0 three transitions are enabled; t_1, t_2 and t_3. After the firing of transition t1 the new CS-Class $C_1 = (m_1, D_1)$ can be computed in the following way

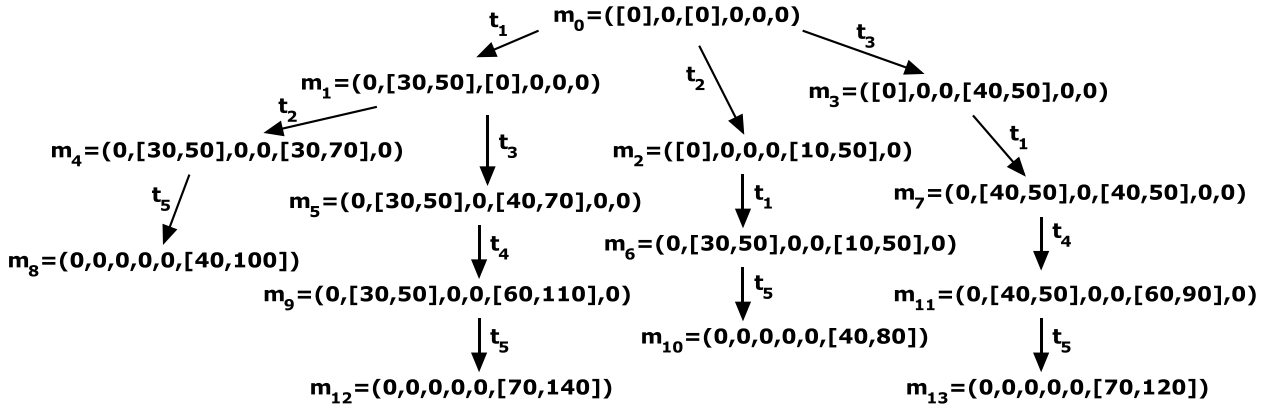


Fig. 4. Reachability tree of the TB net

$$
\begin{aligned}
&MLFT(C_0) = \min(LFT(t_1), LFT(t_2), LFT(t_3)) \\
&= \min(50, 70, 90) = 50, \\
&ST_1 = [EFT_0(t_1), MLFT(C_0)] = [30, 50], \\
&m_1 = (0, [30, 50], [0], 0, 0, 0), \\
&D_1 = \left\{ \begin{array}{l} D_1(t_2) = \max(EFT(t_1), EFT(t_2)), LFT(t_2) = \\ = [30, 70], \\ D_1(t_3) = \max(EFT(t_1), EFT(t_3)), LFT(t_3) = \\ = [40, 90] \end{array} \right\}.
\end{aligned} \tag{8}
$$

From the further computation of CS-classes we can create the reachability tree as described on Fig. 4. Questions like "Is the specified marking reachable until 20 time units?" or "Can this situation happen in the time interval [12, 24]?" can be easily answered by this reachabilty tree.

The complete reachability tree of CS-Classes is depicted on Fig. 4. Questions like "Is there any marking reachable until 20 time units?" or "Is this piece of metal ready to use in time interval [12, 24] or should I take another one?" can be easily answered by this reachabilty tree.

V. Conclusion

The most crucial problem in Petri nets is the reachability problem. It can be proved, that this problem is closely related to other problems like liveness, deadlock, boundedness or coverability problem.

For this reason we tried to find the solution for this problem. At first we proposed an approach which solves this problem for TPNs. Later on we introduced TB nets with their different time semantics, such as weak time semantics, strong time semantics and time interval semantics. For TB nets with STS the approach from Section II was applied because TB nets are far more general then TPNs.

Our future work will be focused on the further examination of unbounded TB nets. Currently we are working on a computer tool which will use TB nets to create and verify models of systems.

Acknowledgment

This work is the result of the project implementation: Centre of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF (50%); and also is supported by VEGA grant project No. 1/4073/07: "Formal specification of programming systems" (50%).

References

[1] M. A. Marsan, G. Balbo, G. Conte, S. Donatelli, G. Franceschinis, "Modeling with generalized stochastic petri nets", *J. Wiley Series in Parallel Computing*, 1995, 299 p.

[2] C. Ghezzi, D. Mandrioli, S. Morasca, M. Pezzé, "A unified high-level Petri net formalism for time critical systems", *IEEE Trans. on Software Engineering*, 2nd ed., vol. 17, 1991, pp. 160-172.

[3] J. Wang, Y. Deng, G. Xu, "Reachability analysis of real-time systems using time Petri nets", *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 30, 2000, pp. 725-736.

[4] J. Wang, "Timed petri nets: theory and application", *Kluwer Academic Publishers*, 1998, 282 pp.

[5] C. Ghezzi, S. Morasca, M. Pezzé, "Validating timing requirements for time basic net specifications", *Journal of Systems and Software*, vol. 27, 1994, pp. 97-117.

[6] T. A. Henzinger, Z. Manna, A. Pnueli, "Timed transition systems", *Lecture Notes In Computer Science*, vol. 600, Proceedings of the Real-Time: Theory in Practice, REX Workshop, 1991, pp. 226-251.

[7] Š. Hudák, "Time Interval Semantics of TB Nets", *Proceeding of international conference RSEE'96*, Romania, 1996, pp. 1-12.

[8] Š. Hudák, "Reachability Analysis of Systems Based on Petri Nets", 1st edition, Košice, elfa spol. s.r.o., 1999, 273 pp.

[9] J. Bača, Š. Hudák, "De/compositional Time Reachability Analysis", *Proceedings of the 6th International Conference*, Oradea - Felix Spa, Romania, 2001, pp. 60-65.

[10] E. W. Mayr, "An algorithm for the general Petri net reachability problem", *Proceedings of the thirteenth annual ACM symposium on Theory of computing*, Wisconsin, 1981, pp. 238-246.

[11] S. R. Kosaraju, "Decidability of reachability in vector addition systems", *Proceedings of the thirteenth annual ACM symposium on Theory of computing*, California, 1982, pp. 267-281.

[12] B. Berthomieu, D. Lime, O. H. Roux, F. Vernadat, "Reachability problems and abstract state spaces for time petri nets with stopwatches", *Discrete Event Dynamic Systems*, vol. 17, 2007, pp. 133-158.

Vector control of induction motor

Peter NGUYEN

Dept. of Electrotechnics, Mechatronic and Industrial Engineering, FEI TU of Košice, Slovak Republic

peter.nguyen@tuke.sk

***Abstract* — The article deals with possibilities of vector control induction motor. The article described two main methods of vector control, which are the direct and indirect vector control. In direct vector control are discussed two methods for determining the rotor magnetic flux, their advantages and disadvantages. In indirect vector control is also described method of determining the magnetic flux and moreover, this type of vector control is also verified by simulation.**

***Keywords*—induction motor, vector control, current model, voltage model, magnetic flux**

I. INTRODUCTION

At the present time induction motor drives are used in many industrial applications from low powers to high performance systems. This is due to the fact that induction machines have simpler construction in comparison with DC machines, and therefore they are cheaper, easier, more reliable, and have less servicing requirement.

In the past the induction motors were used mainly in applications, where speed control was not required. Main reason was lack of technical means and microprocessors, which would allow control an induction motor in full speed range. Today these problems are solved therefore the research is concentrate to control structures development, which can improve quality of control induction motor drives. The main disadvantage of induction motor is complexity of control. Compared with DC motor induction motor is more complex and non-linear system.

Where the drives are working mostly in steady-state, it is possible to used scalar control, also called volt-hertz control (V/f). The principal aim of the control is maintaining the constant ratio of voltage/frequency, thus the stator magnetic flux was constant. Because the scalar control does not give accurate results in transient state, and so is only used for simple applications, where is not required high accuracy and dynamics of control.

When the high accuracy and dynamics of control is required, it is necessary to use the vector control. A major revolution in the area of induction motor control was invention of field-oriented control (FOC) or vector control. The fundamentals of direct vector control were the first proposed in the early seventies by Blaschke [1] and indirect vector control by Hasse [2]. The vector control provides independent control of the torque and flux, similarly how it is in the DC machine control.

Direct torque control (DTC) was proposed by Depenbrock [3] and Takahashi [4]. The principle of this method is based on the control of stator flux vector position, so that the reference values of the torque and flux were obtained. These reference values can be obtained by selection of suitable switching combinations of inverter, thus by selection of the suitable voltage space vector. For the DTC is characteristic high dynamics of torque control. Disadvantage of the DTC is a larger oscillation torque and problems at a low speed.

The present research focuses on possibilities of the vector control without using speed sensor (called a sensorless vector control). The main reasons of elimination speed sensor are size and price reduction of drive, elimination of cable for a sensor and advanced reliability. The motor speed can be estimated by observers in the following ways:

- observers based on the mathematical model
- observers based on artificial intelligence
- observers utilizing the motor construction properties.

II. VECTOR CONTROL

A. Induction motor equations

Stator and rotor voltage equations of squirrel-cage induction motor are:

$$\boldsymbol{u_1} = R_1\boldsymbol{i_1} + d\boldsymbol{\Psi_1}/dt \quad \text{in stator reference frame} \qquad (1)$$

$$0 = R_2\boldsymbol{i_2} + d\boldsymbol{\Psi_2}/dt \quad \text{in rotor reference frame} \qquad (2)$$

Stator and rotor flux equations are expressed as follows:

$$\boldsymbol{\Psi_1} = L_1\boldsymbol{i_1} + L_h\boldsymbol{i'_2} \quad \text{in stator reference frame} \qquad (3)$$

$$\boldsymbol{\Psi_2} = L_2\boldsymbol{i_2} + L_h\boldsymbol{i'_1} \quad \text{in rotor reference frame} \qquad (4)$$

Where L_1 and L_2 are the stator and rotor inductance, L_h is the main inductance, i_1' is the stator current expressed in rotor reference frame, i_2' is the rotor current expressed in stator reference frame.

The motor torque can be expressed by the stator current and flux components as follows.

$$M_m = \frac{3}{2}p\boldsymbol{\Psi_1} \times \boldsymbol{i_1} = \frac{3}{2}p\left(\Psi_{1\alpha}i_{1\beta} - \Psi_{1\beta}i_{1\alpha}\right) \qquad (5)$$

Where p is number of pairs poles.

B. Vector control of induction motor

The vector control can be oriented on the rotor, stator or magnetic flux. Important part of vector control is determination of the magnetic flux vector position, since correct operation of the vector control depends on the determination of the flux position. The necessary angle for the

transformation of quantities from stationary reference frame α,β to synchronously rotating reference frame x,y and backward is computed from the flux vector position. The vector control can be classified into two groups, the direct and indirect vector control. The direct method obtains the position of flux vector either by measurement by Hall probe or the position is computed by using estimators, Kalman filter, Luenberger observer, etc. The indirect vector control calculates rotor flux from the model of rotor circuit and the measured value of stator currents or voltages and speed. It should be noted that all considered method are sensitive on parameter variation of the induction motor.

Direct vector control

In direct vector control the position of flux is usually obtained by estimation, since using Hall probes complicates manufacture of induction motor and increases its price. The flux vector can be estimated by using the voltage model or current model of the magnetic flux [5].

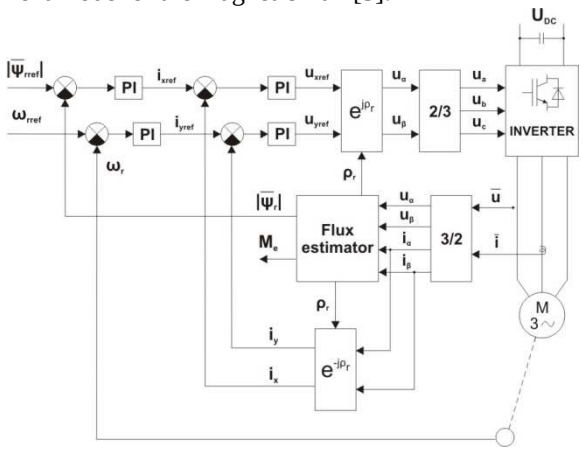


Fig. 1. Direct vector control block diagram

Voltage model

This method utilizes the stator voltage equation of induction motor. The machine terminal voltages and currents are sensed and the fluxes are computed in the stationary reference frame α,β.

$$\boldsymbol{\Psi_1} = \int (\boldsymbol{u_1} - R_1 \boldsymbol{i_1})\, dt \tag{6}$$

From the rotor flux equation (4) is expressed current i_2 and is substituted to the stator flux equation (3), and then is expressed rotor flux Ψ_2:

$$\boldsymbol{\Psi_2} = L_2/L_h\left(\boldsymbol{\Psi_1} - \sigma \acute{L_1} \boldsymbol{i_1}\right) \tag{7}$$

Equation (7) is further distributed to the real and imaginary components:

$$\Psi_{2\alpha} = L_2/L_h\left(\Psi_{1\alpha} - \sigma \acute{L_1} i_{1\alpha}\right) \tag{8}$$

$$\Psi_{2\beta} = L_2/L_h\left(\Psi_{1\beta} - \sigma \acute{L_1} i_{1\beta}\right) \tag{9}$$

Where $\acute{L_1} = \sigma L_1$ and $\sigma = 1 - L_h^2/L_1 L_2$ is coefficient of total leakage.

Angle of the rotor flux vector can be obtained from equation:

$$\cos \rho_2 = \Psi_{2\alpha}/|\boldsymbol{\Psi_2}| \quad \text{or} \quad \sin \rho_2 = \Psi_{2\alpha}/|\boldsymbol{\Psi_2}| \tag{10}$$

Where modulus of rotor flux can be computed as follows:

$$|\boldsymbol{\Psi_2}| = \sqrt{\Psi_{2\alpha}^2 + \Psi_{2\beta}^2} \tag{11}$$

The voltage model is not suitable for very low frequency (including zero speed), because at low frequency the voltage us is very low and therefore estimation of the magnetic flux is inaccurate. The estimation accuracy is affected by the variation of machine parameters R_1, L_1, L_2 and L_h, and particularly temperature variation of R_1 becomes more dominant [5].

Current model

In industry is often required, that the drive worked from zero speed. At low speed region is convenient using current model to estimate the rotor flux, which utilize current and speed signals. The rotor circuit equations expressed in the stationary reference frame are:

$$\mathrm{d}\Psi_{2\alpha}/dt + R_2 i_{2\alpha} + \omega_2 \Psi_{2\beta} = 0 \tag{10}$$

$$\mathrm{d}\Psi_{2\beta}/dt + R_2 i_{2\beta} - \omega_2 \Psi_{2\alpha} = 0 \tag{11}$$

Adding terms $(L_h R_2/L_2) i_{1\alpha}$ and $\left(L_h R_2/L_2\right) i_{1\beta}$ on both sides of the above equations, we get:

$$\frac{d\Psi_{2\alpha}}{dt} + \frac{R_2}{L_2}(L_h i_{1\alpha} + L_2 i_{2\alpha}) + \omega_2 \Psi_{2\beta} = \frac{L_h R_2}{L_2} i_{1\alpha} \tag{12}$$

$$\frac{d\Psi_{2\beta}}{dt} + \frac{R_2}{L_2}\left(L_h i_{1\beta} + L_2 i_{2\beta}\right) - \omega_2 \Psi_{2\alpha} = \frac{L_h R_2}{L_2} i_{1\beta} \tag{13}$$

Real and imaginary component of the rotor flux equation (4) is:

$$\Psi_{2\alpha} = L_h i_{1\alpha} + L_2 i_{2\alpha} \tag{14}$$

$$\Psi_{2\beta} = L_h i_{1\beta} + L_2 i_{2\beta} \tag{15}$$

Substituting equations (14) and (15) in equations (12) and (13) respectively, and simplifying, we get

$$\frac{d\Psi_{2\alpha}}{dt} = \frac{L_h}{T_2} i_{1\alpha} - \omega_2 \Psi_{2\beta} - \frac{1}{T_2} \Psi_{2\alpha} \tag{16}$$

$$\frac{d\Psi_{2\beta}}{dt} = \frac{L_h}{T_2} i_{1\beta} + \omega_2 \Psi_{2\alpha} - \frac{1}{T_2} \Psi_{2\beta} \tag{17}$$

Where $T_2 = L_2/R_2$ is rotor time constant.

Estimation accuracy of flux by current model is affected by the variation of machine parameter, and particularly the rotor resistance variation R_2 becomes dominant.

Since the voltage model flux estimation is better for higher speed and the current model estimation can be used at any speed, it is possible to have a hybrid model [5] (current model would be used for low speeds).

Indirect vector control

The only one difference between the direct and indirect vector control is manner of determination of flux vector position. The stator and rotor voltage equations and the stator and rotor flux equations are expressed in the synchronously rotating reference frame x,y.

$$\boldsymbol{u_{1k}} = R_1\boldsymbol{i_{1k}} + \frac{d\boldsymbol{\Psi_{1k}}}{dt} + j\omega_k\boldsymbol{\Psi_{1k}} \tag{18}$$

$$0 = R_2\boldsymbol{i_{2k}} + \frac{d\boldsymbol{\Psi_{2k}}}{dt} + j(\omega_k - \omega)\boldsymbol{\Psi_{2k}} \tag{19}$$

The magnetizing current i_{2m} is proportional to magnetic flux $\boldsymbol{\Psi_2}$ and is defined as follows:

$$\boldsymbol{\Psi_{2k}} = L_h\boldsymbol{i_{2m}} \tag{20}$$

The magnetizing current i_{2m} can be determined from the measured values of currents or voltages and speed.

$$T_2\frac{di_{2m}}{dt} + i_{2m} = i_{1x} \tag{21}$$

The magnetizing current is aligned with the x axis, and thus x component of stator current i_{1x} will respond rotor flux and will be called flux producing component. The second component of the stator current i_{1y} determines the motor torque and is called torque producing component. The immeasurable rotor current can be expressed by magnetizing current as follows:

$$\boldsymbol{i_{2k}} = \frac{L_h}{L_2}(\boldsymbol{i_{2m}} - \boldsymbol{i_{1k}}) \tag{22}$$

Angular speed $\omega_k = \omega_{2m}$ of synchronously rotating magnetic flux is given as:

$$\omega_{2m} = \omega + \frac{i_{1y}}{T_2 i_{2m}} \tag{23}$$

By integration of angular speed ω_{2m}, we get angle, which is necessary for the transformation between various reference frames. The speed control range in indirect vector control can be extended from zero speed to the field-weakening region [5]. Therefore the indirect vector control is very popular in industrial applications. However control accuracy is affected by parameter variations of machine.

III. SIMULATION OF INDIRECT VECTOR CONTROL

Before the proposed control is necessary to transform the three-phase system for the two-phase system, thus the number of equations describing the behaviour of the induction motor is reduced. Furthermore is required the transformation from stationary reference frame α,β to synchronously rotating reference frame x,y. This ensures that in the synchronously rotating reference frame the variables are not changed periodically. Due to non-linearity compensation can be designed conventional linear controllers.

Because the indirect vector control is very popular in industrial applications, such control was proposed. Detailed procedure for the proposal of vector control is described in [6]. As state variables were chosen following quantities:

$x_1 = i_{2m}$ $x_2 = i_{1x}$ $x_3 = \omega$ $x_4 = i_{1y}$

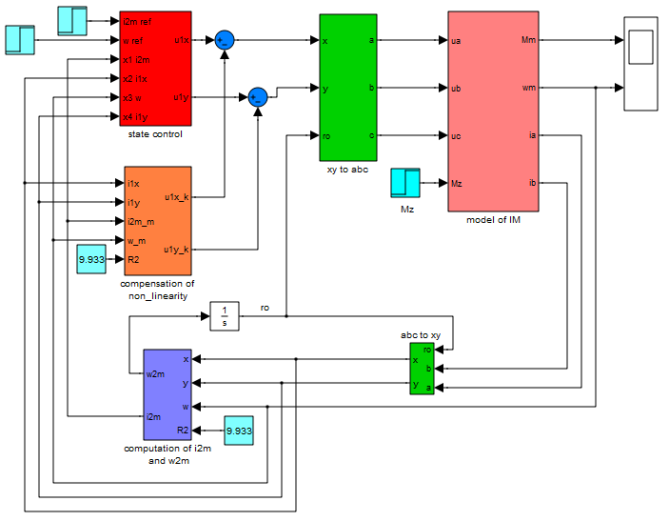


Fig. 2. Indirect vector control simulation block diagram

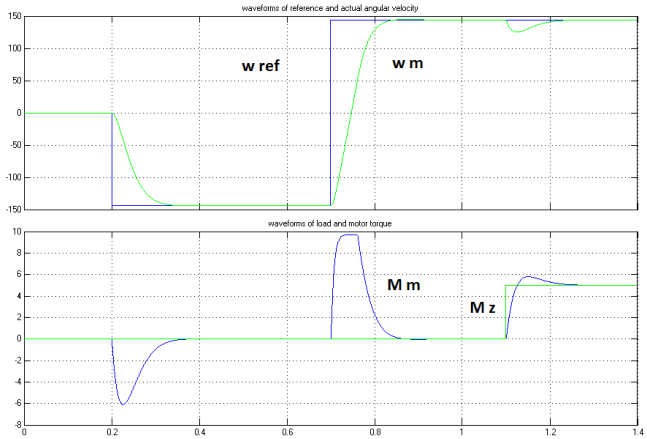


Fig. 3. Waveforms of reference and actual angular speed and torque

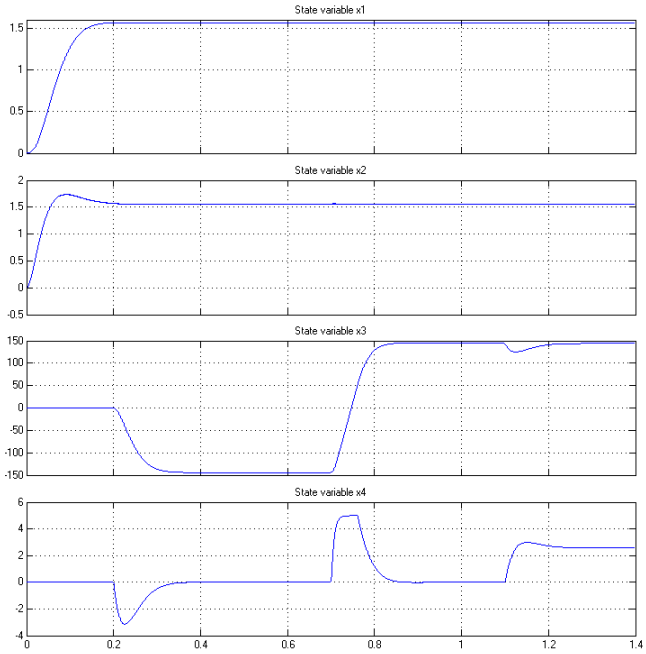


Fig. 4. Waveforms of state variables

IV. CONCLUSION

Where is the demand for high dynamics and accuracy of control, it is advisable to use vector control in combination with asynchronous motor. Currently induction motor is preferable to the DC motor despite a much more complex control. This is mainly due to its simple construction. Induction motor control problem was solved by the appearance of vector control and improvement of hardware and microprocessors. In the present research is focused on the vector control without encoder, i.e. sensorless vector control. Also, much attention is concentrated on the possibilities of using artificial intelligence in control and estimation of state variables.

ACKNOWLEDGMENT

The support provided by the grant VEGA 1/0006/10 is kindly acknowledgment.

REFERENCES

[1] F. Blaschke, “The principle of field orientation as applied to the new transvector closed-loop control system for rotating-field machines,” *Siemens Rev*., 1972.

[2] K. Hasse, “Zum Dynamischen Verhalten der Asynchronmachine bei Betriek Mit Variabler Standerfrequenz und Standerspannung,” ETZ-A , Bd. 9, p. 77, 1968.

[3] M. Depenbrock, “Direkte selbstregelung (DSR) fur hochdynamische Drehfeld-antriebe mit Stromrichterspeisung ,” ETZ Archiv Bd. 7 (1985), H.7, S.211-218.

[4] I. Takahashi – T.Noguchi, “A New Quick Response and High Efficiency Control Strategy of an Induction Motor,” IEEE Trans. on Industry Applications, 1986, vol. IA-22, no. 5, p. 23-25.

[5] B. K. Bose, “Modern Power Electronics and AC Drives,” Prentice Hall PTR, Upper Saddle River, 2002, p. 360-370.

[6] L. Zboray – F. Ďurovský – J. Tomko, “Regulované pohony,” Vienala, 2000, p. 194-197.

Easy Implementation of Domain Specific Language using XML

Marek NOVÁK

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

marek.novak@tuke.sk

***Abstract*—The paper presents a way how to create implementation of domain specific language (DSL) using XML technology relatively easily. The advantage of XML is better readability than general purpose programming language for non-programmers and existence of many tools for parsing XML tree structure. Therefore, development of new DSL is easier and less time consuming. The paper provides example of DSL built at Technical University in Košice as part of international project MonAMI.**

***Keywords*—domain specific language, XML, parser, MonAMI services.**

I. INTRODUCTION

There are some situations when software users, usually people without programming skills, have to inject some rules into the system or define conditions to achieve required functionality. To teach a user at least the basics of programming language in which the system is developed and allow to define rules directly in source code can be very inefficient. A configuration using domain specific language (DSL) seems to be a better solution. DSL is a computer language that is targeted to a particular kind of problem, rather than a general-purpose language which is aimed at any kind of software problem [1]. In comparison to general-purpose programming languages DSLs provide expression capabilities targeted directly to users domain and therefore they are easier to learn and use. Apparently graphical user interface (GUI) with drag-and-drop components provides the most intuitive way to define the rules for user who is non-programmer. Although GUI makes an average user work productively, an expert or a user with advanced skills in application domain can be slowed down [2]. Moreover, to develop robust GUI may be more complex task than building the system itself if the system is not expected to be very complex [3]. In general DSL development may be considered a difficult task because there are both domain knowledge and language development expertise necessary. Existing technologies as XML and parsers like Java Architecture for XML binding (JAXB) [4] with the assumption that rule set is not large and rules by themselves are not very complicated makes DSL development makes much easier.

The purpose of this paper is to show an example how to simply build DSL for a specific purpose. At Technical University of Košice (TUKE) we have developed our own simple DSL when working on project MonAMI to support research at Department of Biomedical Engineering, Automation and Measurements, Faculty of Mechanical Engineering.

II. DEVELOPING DSL WITH XML

As mentioned above, DSLs are designed to be useful for a specific task in a fixed problem domain. An advanced computer user uses daily DSL: configuration file, makefile, CSS etc. Expression capabilities of DSL are fairly limited. They are focused on a certain type of problem or domain, as its name implies, and on expressing narrow set of solutions within the context of that limited scope [2]. Simplicity is very important feature of DSL. A person familiar with domain must easily understand the domain language. Keywords must be very close to user's vocabulary. Syntax should be also simple and clear to facilitate user's work and focus on domain problems that user tries to solve. Depending on how DSL is implemented, we classify it on external or internal DSL [1].

External DSL – is designed to be independent on any particular language. An author of such language must decide about syntax, grammar and the way to parse the syntax. Any language and tools can be used to implement this DSL. For instance it can be Java and Groovy. When using external DSL, author is free to define syntax as he likes – to use symbols, operators, constructs and structures, which fit the best to the domain. On the other hand, it is necessary to define grammar for the language, to create a compiler to parse and process the syntax and map it to the semantics that is expected. Flexibility provided is an advantage, but it may be a really complex task to implement DSL well.

Internal DSL – is designed and implemented using a host language. The advantage is that author does not have to worry about grammar, parsers and tools. However, it brings disadvantages in form of constrains and limitations of the host language. Internal DSL provides easier implementation at the expense of flexibility.

The example of external DSL is also ANT build file [5], which uses XML representation. XML file is processed by the ant utility using the XML parser. Ant's vocabulary contains various terms, such as target and properties, which are valid in the domain and context of compiling and bundling code.

XML brings many advantages to DSL development, though it is not appropriate everywhere. XML-based DSL, grammar is described using DTD or XML schema [6] where *nonterminals* are analogous to elements and *terminals* to data

content. Element definitions determine grammar rules when the element name is the left-hand side and the content model is the right-hand side. XML documents form a tree structure that starts at the root, which corresponds to start symbol in grammar. DSL defined by Backus notation (BNF, EBNF) may be transformed easily to XML tree structure. There is no need to make a big effort to create a parser, since DOM parser or SAX (Simple API for XML) tool already provides this functionality. Since the parse tree can be encoded in XML as well, XSLT transformations can be used for code generation. Therefore, XML and XML tools can be used to implement a programming language compiler [7], [8].

III. MONAMI SERVICES

An intention to create our own DSL arose from a need of service configuration when cooperation on MonAMI project. The configuration should have been made primarily by postgraduate students from Department of Biomedical Engineering, Automation and Measurements without programming knowledge.

MonAMI - Mainstreaming on Ambient Intelligence project [9] (funded by the EU is 6th framework program. It was built on an assumption there will be a high percentage of population over the age over 65 who are still active, computer literate with the ambition to maintain their quality of life. Mission of the project is to improve daily activities and the quality of life of elderly and disabled people at home. It is based on mainstream systems and platforms which create one complex system comprising of different technologies. Services and applications developed with a "Design for All approach" are from following areas [10]:

- Home control, personalized communication interface, activity planning.
- Health control, medication.
- Safety and security at home, visitor validation, activity detection.
- Communication and information.

Department of Biomedical Engineering, Automation and Measurement took a part as centre for testing and validating technologies and services. Selected elderly people and people with disabilities will test services in a laboratory – Feasibility and Usability (FU) centre where the whole system is installed. There are different types of sensors for temperature, humidity, light level, motion and gas detection and different types of actuators as light, shutter and alarm actuator installed. People involved into project research are responsible for testing, satisfaction evaluation, measurement, services adjustment and configuration. Testing is divided into two phases. First phase comprises of evaluation, bug resolving, user insights incorporation and system adjustments to user's needs in laboratory. After this phase is finished, the system will be installed in real households. The role of people working on project is to prepare questionnaires for interviews with users, evaluate the answers, present elderly and disabled people offered MonAMI services in order to understand their purpose and functionally; and finally control, configure and create new services depending on user's ideas in the testing phase [11].

MonAMI service is an activity with strictly defined behavior. Implementation of services diverse:

- Collect data from sensors.
- Actuate devices when pre-defined conditions are fulfilled.
- Provide information to users in friendly way.
- Actuate devices by human input.

The services could be divided into two groups depending on an action taken when measured values exceed defined thresholds or when some values are detected e.g. smoke, gas. The first group represents actions taken by humans – carer responsible for disabled person is at the moment in a shop and is noticed by SMS about fall or heart attack of treated person. He can immediately call an ambulance or just check up the state of treated person depending on character of information. The second group represents actions executed automatically. In case smoke and high temperature is detected in a kitchen, fire department is automatically informed about this situation.

The whole system is implemented in Java programming language based on component oriented architecture OSGi. More detailed description of system architecture is not in the scope of this article but it is important to sketch service implementation. Particular sensors can be understood as services as well, which purpose is nothing more than reading values from sensors and providing them to other services. Each service is one component in the system that could be added or removed.

Behavior of majority of services can be simply defined by following formula: *if values from sensors exceed thresholds, take an appropriate action*. This condition has to be defined in particular service source code.

Here the problem has arisen, because thresholds could be made configurable by some simple interface, however, to add only one extra *AND/OR* condition could be a serious problem. To simplify the testing we have created a component, which parses XML configurable file where all services are defined by DSL. Structure of XML is defined by XSD and JAXB has been used as a parser.

IV. FLEXIAMI DSL

FlexiAMI is the name of component, which provides configuration capabilities for MonAMI services ("flexi" as flexibility in definition and AMI is taken from MonAMI) [12]. Keywords and service definition principle are close to user domain however users are tightened up by a relatively larger set of rules. They are straightforward, therefore easy to learn. Rules format is based on XML, which has been used because of many advantages, as mentioned also in previous part, the most important advantages are:

- hierarchical structure of final configuration file;
- self-describing and simple syntax;
- existence of validators;
- existence of APIs parsing the XML file;
- easy implementation in Java.

To provide the whole formal grammar is not necessary, however simple showcase can be very helpful. The configuration itself comprises of two parts – definition of

sensors and actuators represents first part and rules definition is the second.

First part could be following:

```
<sensor>
    <name>temperature</name>
    <type>TemperatureSensor</type>
    <location>Kitchen</location>
</sensor>
<actuator>
    <name>alarm</name>
    <type>AlarmActuator</type>
    <location>Kitchen</location>
</actuator>
```

Each sensor and actuator has to be declared by its unique identificator **<name>**, which is used consequently in rules, **<type>** specifies sensors and actuators according to MonAMI naming and **<location>** determines room where devices are installed.

Rules follow this diagram:

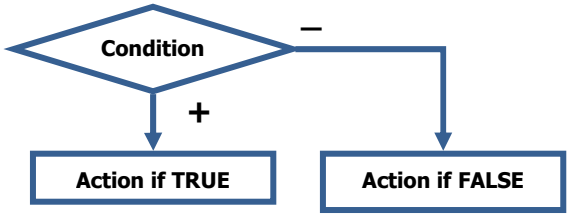


Fig. 1 Rule behavior diagram

Below is an example of rule, which uses temperature sensor, smoke sensor and alarm actuator. When smoke and high temperature is detected at once, alarm actuator is turned ON. Alarm actuator can represent beeper in house, alarm connected directly to firehouse station or some other notification method as SMS.

Definition is following:

```
<service>
    <name>FireService</name>
    <type>AmbientMonitoringAlarm</type>
<condition>
<and>
    <cond oper="eq" val="TRUE">smoke</cond>
    <cond oper="ht"
    val="40.0">temperature</cond>
</and>
</condition>
<action>
    <actIf val="ON">alarm</actIf>
    <actElse val="OFF">alarm</actElse>
</action>
<service>
```

Each service has its name and type according to MonAMI specification. Other bundles in OSGi system can find and use this service by defined type, which tells something about its behavior.

Condition comprises of three values: sensor identifier, operation and value to compare. The meaning of **<cond oper="eq" val="TRUE">smoke</cond>** is: *if smoke is detected in kitchen then TRUE*. The result of conditional expression can be *TRUE* or *FALSE*. The content of **<cond>** element refers to identifier in first part service definition, where is specified which sensor from which room should be used. Possible operations **oper="eq"** are: **eq, nq, ht, lt, he, le** denoted to: equal, not equal, higher than, less than, higher equal, less equal respectively. Compare value **val="TRUE"** depends on used sensor. Some sensors provide only two values – detected/not detected and some numeral values. Element **<and>** represents Boolean logical operator. Conditional expressions surrounded by this operator are translated as: *cond1 AND cond2 AND cond3*.

```
<and>
    <cond ... >...</cond>
    <cond ... >...</cond>
    <cond ... >...</cond>
</and>
```

Logical operators can be embedded to define condition evaluation priority. For example (1, 2, 3 here means conditional expression number, not sensor identifier):

```
<and>
    <cond ... >1</cond>
    <cond ... >2</cond>
    <or>
        <cond ... >3</cond>
        <cond ... >4</cond>
    </or>
</and>
```

It is evaluated as: *cond1 AND cond2 AND (cond3 OR cond4)*.

Condition embedding allows defining any type of logical expression. Elements **<and>** and **<or>** can surround arbitrary number of conditional expressions.

Execution of some services may depend on more aspects than only on sensor values. An example is time when sensor value is measured.

Elderly person is used to wake up at 8:00 am. When there is no motion detected in an hour after 8:00 am, carer is informed about this situation by SMS. This service is defined accordingly:

```
<service>
    <name>WakeUPService</name>
    <type>PersonPresenceDetector</type>
<condition>
<and>
    <cond oper="FALSE" val="TRUE"
    duration="3600">motion</cond>
    <cond oper="ht" val="08:00:00"
    >time</cond>
</and>
</condition>
<action>
    <actIf val="ON">sms</actIf>
</action>
<service>
```

In the first conditional expression there is one additional attribute **duration="3600"**. This attribute determines how long measured value has to be unchanged. In this case, no motion is detected during one hour. Duration value is set in seconds, because there could be some services when only

some seconds are needed. For example when person leaves the toilet and no movement is detected for 20 seconds, the light is turned off:

```
<cond oper="ht" val="08:00:00" >time</cond>
```

is translated as: *if actual time is higher than 8:00 am then TRUE.*

Service action comprises of two action expressions. One is executed when condition result is TRUE *(actIf)* and the other is executed when condition result is FALSE *(actElse)*. There can be some additional properties to service added by:

```
<properties>
    <prop name="name">value</prop>
</properties>
```

CONCLUSION

FlexiAMI component becomes very helpful in the phase of MonAMI services testing. DSL has enabled programmers to delegate configuration responsibility to people working on MonAMI project who are non-programmers. For that reason it was not necessary to change the application code after each new requirement of user. XML format and used elements have been descriptive enough and FlexiAMI users get used to them very quickly. JAXB parser significantly simplifies the DSL development process, and therefore XML technologies seem to be very effective for DSL creation for this purpose. There are also some other innovative approaches of DSL development as Annotation Based Parser Generator [13].

ACKNOWLEDGMENT

The authors would like to thank the European Commission for the support within the 6th Framework Program by the grant of the integrated project within the priority 2.3.2.10 e-Inclusion "Mainstreaming on Ambient Intelligence" - MonAMI IST-5-0535147, and Coordination action "Design for All for e-Inclusion" DfA@eInclusion IST 033838.

REFERENCES

[1] Fowler, M., *Domain Specific Language*. 2005, from Blog martingFowler.com: http://www.martinfowler.com/bliki/DomainSpecificLanguage.html"

[2] Subramaniam V. "*Creating DSLs in Java, Part 1: What is a domain-specific language?*", JavaWorld.com, March 2008, http://www.javaworld.com/javaworld/jw-06-2008/jw-06-dsls-in-java-1.html?page=2

[3] M. Fowler, "*Language Workbenches: The Killer-App for Domain Specific Languages?"* 2005. http://www.martinfowler.com/articles/languageWorkbench.

[4] Glassfish community, "*Java Architecture for XML Binding (JAXB)*", https://jaxb.dev.java.net.

[5] Apache Software Foundation, "*Apache Ant project*", http://ant.apache.org/

[6] W3C, "*XML Schema*", http://www.w3.org/XML/Schema.html

[7] R. Germon. "*Using XML as an Intermediate Form for Compiler Development*". 2001

[8] M. Mernik, J. Heering, A. M. Sloane, "*When and How to Develop Domain- Specific Languages*", ACM Computing Surveys, Vol. 37, No.4, December 2005, p. 316–344.

[9] Project Overview. (n.d.). Retrieved March 13, 2010, from MonAMI: *http://www.monami.info/*

[10] Technická univerzita v Košiciach, K. b.-t. (n.d.). *Inteligentné prostredie v službách hlavného prúdu.* Retrieved March 13, 2010, from Integrovaný projekt MonAMI: *http://web.tuke.sk/sjf-kbiaam/monami.pdf*

[11] D. Šimšík, J. Bujňák, "*Ambient technology and social services for seniors*". In: SAMI 2010 : The 8th International symposium on Applied Machine Inteligence and Informatics : Zborník príspevkov : 28.- 30.1. 2010, Herľany, Slovakia.

[12] Novák, M." *Ovládač pre zariadenia technológie 1-Wire pre platformu OSGi.*" Diploma thesis. May 2009. Košice: Technical University of Košice, Faculty of Electrical Engineering and Informatics.

[13] J.Poruban, M.Forgac, M.Sabo, "*Annotation Based Parser Generator*", Proceedings of the International Multiconference on Computer Science and Information Technology, 2009, pp. 707 – 714

Training improved acoustic models for IRKR system with extended training database

[1]*Marek PAPCO,* [2]*Stanislav ONDÁŠ*

[1]Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]marek.papco@tuke.sk, [2]stanislav.ondas@tuke.sk

***Abstract*—This paper describes training of acoustic models on extended Speechdat database with aim to reduce word error in ASR system as a part of IRKR. The origin database was extended to titles of towns and villages as they are recognized frequently due to services provided by IRKR system. Although the extending database is small in compare with original database, the results show that extension had postive effect to recognition accuracy of own names.**

***Keywords*—acoustic models training, speech corpus, voice dialog.**

I. Introduction

It is well known that good trained acoustic models are very important part of speech recognition system. For training of acoustic models the training database is very important. The best results of recognition are achieved with the acoustic models trained on similar data as the system will be onset [1]. IRKR system [2] that was developed in state project is the voice orientated dialog service that provide weather forecast, traffic guide and shedule of urban mass transportation in Košice in telecomunication networks in Slovak language and it is based on Galaxy architecture [3]. IRKR is a spoken language dialogue system that consist of Audio server (interface between telecomunication network and computer), Text-to-speech server (transforms text into acoustic form), Communication manager (control communication with user), Information server (find informations requested by user) and Automatic speech recognition server (ASR server). The acoustic models used for recognition are hidden Markov models of context dependent triphones. Acoustic models were trained on Speechdat database [4]. ASR server is specialized to recognition of isolated words and connected digits represented by date, time, town and bus stops names. Due to this fact the Speechdat database was extended to 580 utterances of titles of towns and villages in Slovakia. Each of the utterance duration is about 35 seconds. Utterances were recorded through VoIP channel with 8KHz sampling frequency to achieve real using conditions.

New acoustic models were trained with using scripts of Masper project [5]. For training of acoustic models the HMM toolkit HTK [6] was used.

II. Acoustic models training

Training of acoustic models was controlled by Masper scripts. Whole training process consist of some steps.

A. Features Extraction

First step in training is parametrisation of speech waveforms into sequence of feature vectors. Feature extraction is provided by tool *HCopy* that generate MFCC coefficients with zero, delta and acceleration coefficients. The cepstral mean normalization to MFCC coefficients was also used.

B. Monophones training

The flat start monophones training is started with tool *HCompV* that compute global means and variances to initialize the parameters of HMM model. Baum-Welch reestimation of acoustic model parameters is provided by tool *HERest*. After the prototype acoustic model is reestimated the next step is isolated word training. Using initialized acoustic model the Viterbi alignment to phones is performed with *HInit*. Using alignment to phones the acoustic model parameters with *HRest* are reestimated again.

C. Triphones training

Triphones training starts with creating triphone transcription from monophone transcription with tool *HLEd*. *HHEd* provide cloning monophone models and tying parameters. With *HDMan* the triphone lexicon and list of triphones is created. HMM parameters for triphone models are estimated by tool *HERest*. Increasing the number of Gaussian mixtures that model every emitting state of HMM model is provided by *HHEd*.

Using the training process described above there were trained 3 state left to right HMM models with up to 32 Gaussian mixtures. The block diagram of training process is illustrated in Fig. [1]. The first set of acoustic models was trained only on Speechdat database and second models set was trained on Speechdat database extended to utterances with titles of towns and villages.

III. Results

The acoustic models were tested on own names that contain names of the persons and titles of villages and towns. Results of speech recognition are illustrated in figure Fig. [2].

Due to using only tied models in IRKR system the table of results is omitted to phone models and contains results of tied-state triphone acoustic models in Table [I].

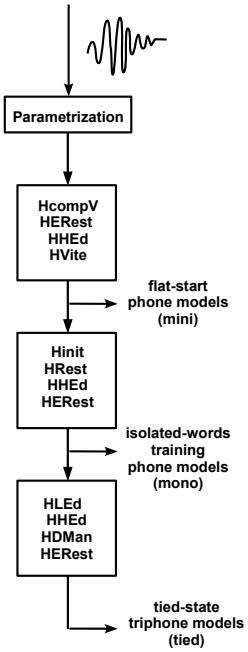


Fig. 1. Block diagram of training acoustic models.

Fig. 2. Results of recognition own names with acoustic models trained on Speechdat (SD) and Extended Speechdat database (ESD)

TABLE I
OWN NAMES RECOGNITION WITH TIED-STATE TRIPHONE MODELS

Acoustic Model	SD WER (%)	ESD WER (%)	Improvement (%)
tied_1_1	17.97	14.55	3.42
tied_1_2	18.37	14.96	3.41
tied_2_1	18.46	14.63	3.83
tied_2_2	17.56	13.82	3.74
tied_4_1	16.34	11.87	4.47
tied_4_2	15.2	10.65	4.55
tied_8_1	14.72	10.49	4.23
tied_8_2	14.07	10.09	3.98
tied_16_1	13.41	9.19	4.22
tied_16_2	12.6	8.86	3.74
tied_32_1	13.09	9.18	3.91
tied_32_2	13.82	9.76	4.06

IV. CONCLUSION

Results show that Speechdat database extension to towns and villages titles had influence to word error rate. In case of tied-state triphone models set the WER was 3.96% in average better in extended database. Tied-state triphone acoustic model with 32 Gaussians (*tied_32_2*) used in IRKR system had 4.06% worse WER than model trained on extended database. These results show that even if the origin training databse had about 80 hours af acoustical data the extending with small, but with aimed database to real onset conditions of the system had positive effect on WER of the ASR system.

ACKNOWLEDGMENT

The research presented in this paper was supported by the Slovak Research and Development Agency under research projects APVV-0369-07 and VMSP-P-0004-09 and is the result of the project implementation Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF.

REFERENCES

[1] M. Papco and M. Lojka, "Adaptation of acoustic models for robust speech recognition," in *SCYR 2009 - 9th Scientific Conference of Young Researchers*, FEI TU of Košice, 2009.

[2] J. Juhár, A. Čižmár, M. Rusko, M. Trnka, G. Rozinaj, and R. Jarina, "Voice operated information system in slovak," in *Computing and Informatics*, 2007, pp. 577–603.

[3] J. Juhár, S. Ondáš, A. Čižmár, M. Rusko, G. Rozinaj, and R. Jarina, "Development of slovak galaxy/voicexml based spoken language dialogue system to retrieve information from the internet," in *Interspeech 2006 - ICSLP : Proceedings of the Ninth International Conference on Spoken Language Processing*, Pittsburgh, Pensylvania, USA, September 17-21 2006, pp. 485–488.

[4] S. Lihan and J. Juhár, "Comparison of two slovak speech databases in speech recognition tests," in *ACOUSTICS High Tatras 06 : 33rd International Acoustical Conference - EAA Symposium*, Štrbské Pleso, Slovakia, October 4-6 2006, pp. 130–133.

[5] A. Žgank, Z. Kačič, F. Diehl, J. Juhár, S. Lihan, K. Vicsi, and G. Szaszak, "The cost 278 masper initiative-cross lingual speech recognition with large telephone databases," in *Proceedings of the 4th International Conference on Language Recourses and Evaluation*, Lisbon, Portugal, May 26-28 2004, pp. 2107–2110.

[6] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland, *The HTK Book (for HTK Version 3.4)*, revised for htk version 3.4 ed., December 2006, first published December 1995.

Visualization and Supervisory Control of Systems – Application Bells

Miloš PAVLÍK, Stanislav LACIŇÁK

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

milos.pavlik@tuke.sk, stanislav.lacinak@tuke.sk

***Abstract*— A lot of control systems are using computer workstations for process control, whether personal computers or workstations distributed over a network. Control interventions to control systems can be performed from multiple sites, using distributed workstations, possibly with a web browser. The point of article was to analyze in more detail the existing real model Bells in the laboratory. Visualization of model was created by using software RSView32 and supervisory control of the application using web browser is realized by using software RSView32 Active Display System.**

***Keywords*— visualization, supervisory control, model, control system**

I. INTRODUCTION

Currently, visualization has become part of the information and control systems of different technological processes. With the help of visualization we can design the graphical user interface of different machines, major technological equipment and processes.

Visualization systems have become standard part of automation. It's not only the part of large industrial plants, but has become part of the management systems in small workplaces. Current trends are becoming controlling and monitoring technological processes remotely, thus using a web browser [3].

II. VISUALIZATION OF PROCESSES AND SUPERVISORY CONTROL

A. Visualization

We often encounter with the acronym SCADA/HMI (Supervisory Control and Data Acquisition / Human-Machine Interface) in visualization, which states that it is a solution of operator interface. When we use naming visualization, is necessary to draw attention to possible distortion of the concept. As a SCADA system is called software product that is used universal as an instrument which can create a specialized product that visualize conditions and actions in a particular controlled object, enables its manual control, data entry of input parameters and executes several other functions, for example, long-term monitoring and documenting the development process. This is useful in showing the volume of production, its quality, disposal and raw materials, to determine the cause or offender for any problems, accidents and losses or to solving technical diagnostics. Visualization system serves as an interface between higher levels of corporate information systems and particular processes. Only in this context it is appropriate the term of visualization [3].

Visualization features can be characterized as:

- View information not only about the actual technological process, but also statistical information for monitoring and quality control process. This information may be appropriate for managers.
- Extending the possibilities of remote control by using various technology
- Imaging, processing and archiving of data coming from the process with alarms signalization, their validation, sorting by severity and the possibility of assigning an audio alarm
- Ability to communicate with subordinates or superiors stations from different manufacturers
- Diagnosis of various technological activities, as well as devices in periodic intervals, or according to special requirements [5]

The well-made interface HMI improves working conditions for users and also helps to regulate the error substantially and thereby reduce potential damage to the devices. Technical equipment may be formed by operator station and its communication tools.

B. Supervisory control

Supervisory control is a general term for control of many individual controllers or control loops, whether by a human or an automatic control system, although almost every real system is a combination of both.

Supervisory control often takes one of two forms. In one, the controlled machine or process continues autonomously. It is observed from time to time by a human who, when deeming it necessary, intervenes to modify the control algorithm in some way.

In the other, the process accepts an instruction, carries it out autonomously, reports the results and awaits further commands. With manual control, the operator interacts directly with a controlled process or task using switches, levers, screws, valves etc, to control actuators.

Supervisory control means that one or more human

operators are intermittently programming and continually receiving information from a computer that itself closes an autonomous control loop through artificial effectors to the controlled process or task environment.

III. CONTROLLED SYSTEM - MODEL BELLS

Function of bells is based on the bell oscillation where the clapper strikes the bell and issues the tone. It is necessary to ensure that the bells oscillate in defined path. This can be ensured by correct length and force action on the bell. It all depends on the bell consolidation and its axis of rotation. As the man is not used to pull his weight, so we need a different kind of traction force. The most common is the coil, where the iron core is pulled into it, which is fixed to the end of the rope.

Outer positions of bells are represented by sensors S1 and S2 and starting position by sensor S3.They give information about where the bell is located. As stated above, this technique is based on the crowding in the iron core into coil. The rope is fixed to the lever, which rotates the bell. The counterweight to iron core is located on the other side of the lever.

Bell is attached to vertical construction. Clapper is mounted on a flexible rope to ensure that the bell in the extreme positions of sensors struck to the clapper and also suppresses any overshoot of bells. Sensing elements are placed above the coil and the mechanical part is shown in Figure

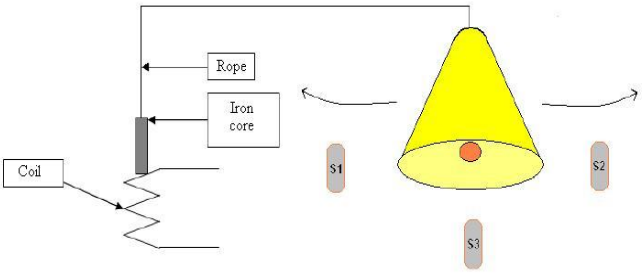


Fig. 1 The principal scheme of functionality of the bell technique

While cable is moving, linage is rotating and thereby it runs through inductive sensors, which senses the middle and outer positions of bell. Sensors used in this case are inductive. But for the mechanical design can also be used contact switches. But they are less reliable.

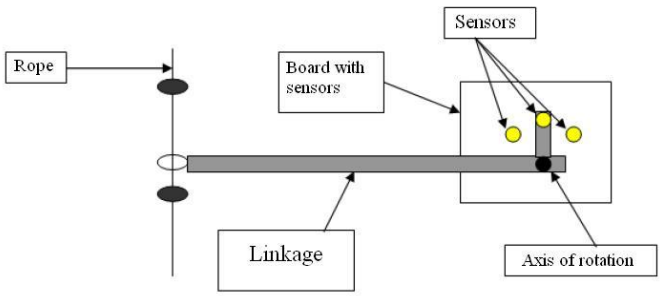


Fig. 2 The scheme of scanning drive

IV. PROCESS CONTROL LEVEL

A physical model of bells is composed of 3 parts:

- Hardware: sensors, DeviceNet technology network and its modules, programmable automaton PLC-5/20, Module FLEX I/O (1794-ADN), 1770-KFD interface (DeviceNet through RS232)
- Software: PLC programming tools
- Communication: DeviceNet industrial network technology that is used to connect sensors to the machine and technology DH + network that serves to connection machine and computer, where is RSLinx Gateway, which is attached application for supervisory management [3]

V. SUPERVISORY CONTROL AND VISUALIZATION

The application was designed in package of Rockwell Software called RSView32. Application includes the visualization of model Bells, trends and standard tools for diagnostics. It was designed for local control and remote control. Remote control via web browsers is implemented through a software package entitled RSView32 Active Display System.

RSView32 Active Display System is an extension of the product RSView32 and offers the same options like the option of displaying objects in real time, manage alarms, centrally manage files of graphic displays, automatically deploy client software through a network or automatically create a connection with client secondary server if the primary connection was interrupted.

Within the visualization the model Bells can be controlled, as well as we can monitor trends and elements of diagnosis, thus indications that the system is online, and if automatic control is not in conflict with the manual control of model Bells. Manual control is implemented using switches, which are installed directly on the case of model.

There are few differences between local and remote visualization of control system. Some commands are different in local and remote visualization. For example for opening graphic display in local visualization is used command “Display” but in remote visualization is used command “NavigateGFX”. To run a remote visualization is necessary to run an Active Display System server on startup of application. There are few more differences, but basically remote and local visualization are similar.

Variables that were necessary to the functional operation of visualization were type - I/O, and their values are received from control program, which was stored in programmable automaton SLC 5/04, which was connected with the module FLEX I/O using DeviceNet technology network.

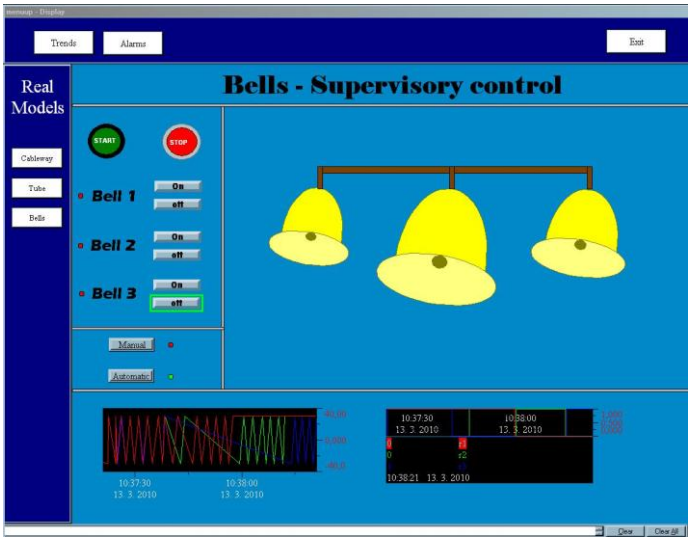


Fig. 3 Local visualization of model Bells

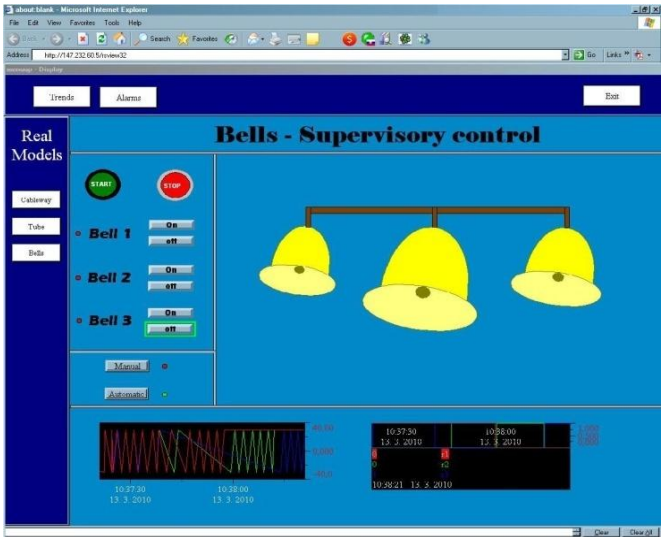


Fig. 4 Remote visualization of model Bells

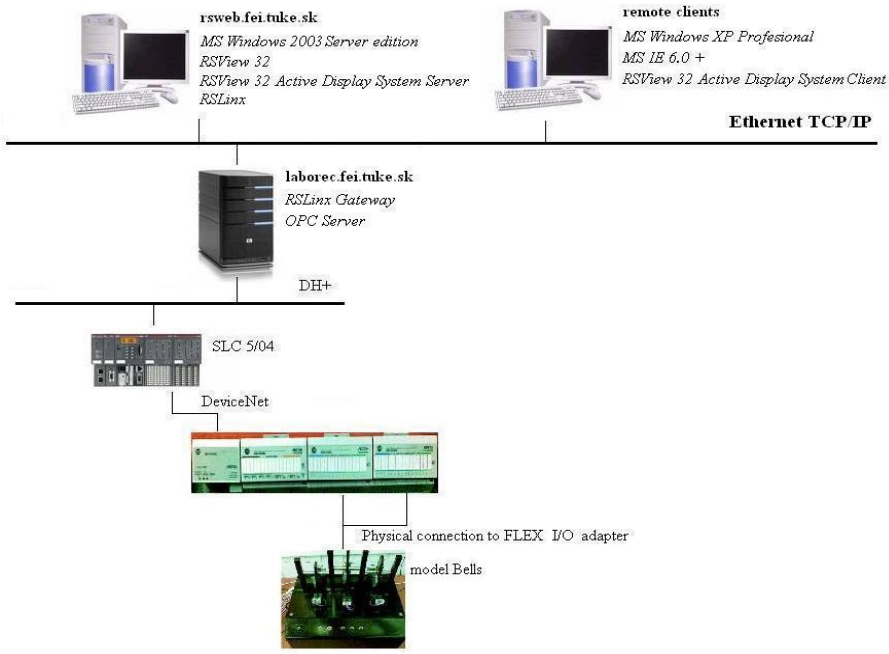


Fig. 5 Block diagram of connection of model Bells

VI. Conclusion

Visualization means the use of theoretical, technical, programming and communication funds for the visibility of defined objects in regard to technological or manufacturing processes and their automatic control system to support decision making and control in real time. New strategies help to advance the development of HMI and visualization, where the emphasis is on reducing the cost of training and development, as well as more convenient, simple and faster development of SCADA/ HMI applications.

Acknowledgment

This work was supported by grant KEGA 037-011TUKE-4/2010 and VEGA - 1/0617/08.

References

1. I. Bélai, L. Jurišica, K. Kováč, J. Šturcel, "Control and diagnostics of the technological processes in manufacturing process based on application of ICT: Selected applications of ICT in enterprises, institutions and SMEs." Bratislava: FEI STU in Bratislava, 158 p, 2005.
2. I. Zolotová, M. Bakoš, L. Landryová, "Possibilities of Communication in Information and Control Systems." *Annals of the University of Craiova: Series: Automation, computers, electronics and mechatronics*, Vol. 4, no. 2, pp. 163-168, 2007.
3. M. Pavlík, "Control of models Cableway, Bells and Tube and Integration to DSC on DCAI", Diploma thesis, Košice: FEI TU in Košice, 2009.
4. M. Novák, "Control of models Cableway and Bells within DSC of DCAI", Diploma thesis, Košice: FEI TU in Košice, 2009.
5. T. Andrek, "Multilayer Automation Architecture – Industrial Application Server – Lab Model Intelligent House", Diploma thesis, Košice: FEI TU in Košice, 2009.
6. J. Galdun, L. Takáč, J. Liguš, J.M. Thiriet, J. Sarnovský, "Distributed control systems reliability: consideration of multi-agent behavior." *SAMI 2008 : 6th international Symposium on Applied Machine Intelligence and Informatics*, pp.157-162, January 21-22, 2008.

On the Cartesian products with crossing number two

Jana PETRILLOVÁ

Dept. of Mathematics, FEI TU of Košice, Slovak Republic

jana.petrillova@tuke.sk

***Abstract*—There are several known exact results on the crossing numbers of Cartesian products of paths, cycles, and complete graphs. The aim of this paper is to characterize graphs G_1 and G_2 for which the crossing number of its Cartesian product $G_1 \times G_2$ equals two, if one of the graphs G_1 and G_2 is a cycle.**

***Keywords*—Cartesian product, crossing number, drawing, graph**

I. Introduction

Let G be a simple graph with vertex set V and edge set E. A drawing is a mapping of a graph into a surface. For simplicity, we assume that in a drawing (a) no edge passes through any vertex other than its end-points, (b) no two edges touch each other (i.e., if two edges have a common interior point, then at this point they properly cross each other), and (c) no three edges cross at the same point. The crossing number $cr(G)$ of a graph G is the minimum possible number of edge crossings in any drawing of G in the plane. It is easy to see that a drawing with minimum number of crossings (an optimal drawing) is always a good drawing, meaning that no edge crosses itself, no two edges cross more than once, and no two edges incident with the same vertex cross. The Cartesian product $G_1 \times G_2$ of graphs G_1 and G_2 has vertex set $V(G_1 \times G_2) = V(G_1) \times V(G_2)$ and any two vertices (u, u') and (v, v') are adjacent in $G_1 \times G_2$ if and only if either $u = v$ and u' is adjacent with v' in G_2, or $u' = v'$ and u is adjacent with v in G_1.

The investigation on the crossing numbers of graphs is a classical and however very difficult problem. The problem of reducing the number of crossings was therefore not only studied by the graph theory community, but also by VLSI communities and computer scientists. As a crossing of two edges of the communication graph requires unit area in VLSI-layout, the crossing number together with the number of vertices of the graph immediately provide a lower bound for the area of the VLSI-layout of the communication graph. The crossing numbers has been also studied to improve the readability of hierarchical structures.

As computing the exact value of crossing number of a given graph is in general an elusive problem, the crossing numbers of few families of graphs are known. Most of them are Cartesian products of special graphs. Let C_n and P_n be the cycle and the paths of length n, respectively, and let S_n denote the star $K_{1,n}$. Harary at al. [6] conjectured that the crossing number of $C_m \times C_n$ is $(m-2)n$, for all m, n satisfying $3 \leq m \leq n$. This has been proved only for m, n satisfying $n \geq m,\ m \leq 7$. It was recently proved by L. Y. Glebsky and G. Salazar [5] that the crossing number of $C_m \times C_n$ equals its long–conjectured value at least for $n \geq m(m+1)$. The crossing numbers of the Cartesian products of cycles and all graphs of order four are determined in [3], [7] and [11]. In [9], the crossing numbers of the Cartesian products of stars S_4 with paths and cycles were studied.

Kulli at al. in [10] started to study line graphs with crossing number one. For value two, the similar problem were solved in [1] and [8]. In this paper we extend the result obtained in MSc Thesis [12] by given the necessary and sufficient conditions for all pairs of graphs G_1 and G_2 for which the crossing number of the Cartesian product $G_1 \times G_2$ is one.

II. Preliminary results

Let us consider three graphs H, J, and K in Fig. 1. In the proof of the main result we need to known the crossing numbers of the graphs $P_3 \times H$, $P_2 \times J$, and $P_2 \times K$.

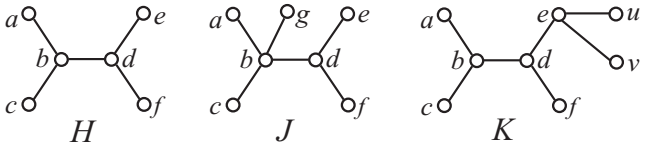


Fig. 1. The special graphs H, J, and K.

Lemma II.1 *$cr(H \times P_2) = 2$ and $cr(H \times P_3) \geq 3$.*

Proof. The graph $H \times P_2$ consists of three copies of H with the vertices $a_i, b_i, c_i, d_i, e_i, f_i,\ i = 0, 1, 2$ and of six paths $x_0 x_1 x_2$ for all $x = a, b, c, d, e, f$, see Fig. 2(a). Assume the subgraph $H_b \times P_2$ induced on the edges incident with the vertices b_i for all $i = 0, 1, 2$. The subgraph $H_b \times P_2$ is isomorphic to the graph $S_3 \times P_2$ and $cr(S_3 \times P_2) = 1$, see [7]. This implies that $cr(H \times P_2) \geq 1$ and if there is a drawing of the graph $H \times P_2$ with one crossing, none of the edges incident with the vertices e_i and $f_i,\ i = 0, 1, 2$, is crossed. But the planar drawing of the subgraph $H_{ef} \times P_2$ induced on the edges incident with the vertices e_i and $f_i,\ i = 0, 1, 2$, is unique shown in Fig. 2(b). In this drawing at most two of the vertices d_0, d_1, d_2 appear on a boundary of one region. Hence, in the considered drawing of the whole graph $H \times P_2$ with one crossing the vertex b_1 is placed in one region of the subdrawing of $H_{ef} \times P_2$ and the paths joining the vertex b_1 with the vertices d_0 and d_2 cross the edges of $H_{ef} \times P_2$. Thus, $cr(H \times P_2) \geq 2$. The drawing in Fig. 2(a) confirms that $cr(H \times P_2) \leq 2$ and therefore, $cr(H \times P_2) = 2$. In Fig. 2(c) there is the drawing of the graph $H \times P_3$ with four crossings. As $H \times P_2$ is a subgraph of $H \times P_3$, $cr(H \times P_3) \geq 2$. In Fig. 2(c) one can easy to verify that the removing of any edge from the graph $H \times P_3$ results in a graph which contain a subgraph homeomorphic with $H \times P_2$. Hence, if there is a drawing of the graph $H \times P_3$ with only two crossings, the removing of a crossed edge results in a drawing with at most one crossing. This contradicts the fact that every

graph homeomorphic to $H \times P_3$ has crossing number two and therefore, $cr(H \times P_3) \geq 3$. □

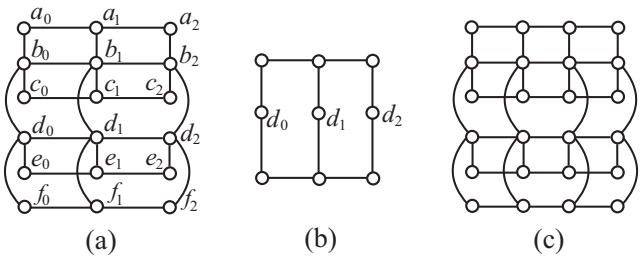


Fig. 2. The graphs $H \times P_2$, $H_{ef} \times P_2$ and $H \times P_3$.

Lemma II.2 $cr(J \times P_2) = 3$.

Proof. The graph J is obtained from the graph H by adding one new vertex g and the edge $\{b, g\}$. Thus, the subgraph $J_b \times P_2$ induced on the edges incident with the vertices b_i for all $i = 0, 1, 2$ is isomorphic to the graph $S_4 \times P_2$. In any drawing of the subgraph $J_b \times P_2$ there are at least two crossings, because $cr(S_4 \times P_2) = 2$, see [9]. Assume that there is a drawing of the graph $J \times P_2$ with only two crossings. Then on the edges of the subgraph $J_{ef} \times P_2$ induced on the edges incident with the vertices e_i and f_i, $i = 0, 1, 2$, there is no crossing and the subdrawing of $J_{ef} \times P_2$ is unique shown in Fig. 2(b). The similar consideration as in the proof of Lemma II.1 confirms that in the considered drawing also some edge of $J_{ef} \times P_2$ must be crossed. This contradiction, together with a suitable drawing of the graph $J \times P_2$ with three crossings, proofs that $cr(J \times P_2) = 3$. □

Lemma II.3 $cr(K \times P_2) = 3$.

Proof. The graph K consists of the graph H and two new vertices u and v and two new edges $\{e, u\}$ and $\{e, v\}$. So, the graph $K \times P_2$ contains $H \times P_2$ as a subgraph and therefore, $cr(K \times P_2) \geq 2$. On the other hand, a suitable drawing of the graph $K \times P_2$ with three crossings implies that $cr(K \times P_2) \leq 3$. Hence, if there is a drawing of the graph $K \times P_2$ with only two crossings, none of them appear on the subgraph $K_{uv} \times P_2$ induced on the edges incident with the vertices u_i and v_i, $i = 0, 1, 2$. The unique drawing of the subgraph $K_{uv} \times P_2$ without crossings contains at most two of the vertices e_o, e_1, e_2 on a boundary of one region and the same analysis as in the proof of Lemma II.1 confirms that in the considered drawing some edge of the subgraph $K_{uv} \times P_2$ must be crossed. This contradiction completes the proof. □

III. The main result

Tindira in MSc Thesis [12] proved the necessary and sufficient conditions for all pairs of graphs G_1 and G_2 for which the crossing number of the Cartesian product $G_1 \times G_2$ is one. More precisely, using special graphs in Fig. 3, he proved:

Theorem III.1 *Let G_1 and G_2 be connected graphs. Then $cr(G_1 \times G_2) = 1$ if and only if one of the following conditions holds:*
1) G_1 is F_1 or its subdivision and $G_2 = P_2$,
2) G_1 is homeomorphic with S_3 and $G_2 = P_2$ or $G_2 = C_3$,
3) G_1 is F_2 or F_{2+} and $G_2 = P_2$,
4) G_1 is F_3 or F_{3+} and $G_2 = P_2$.

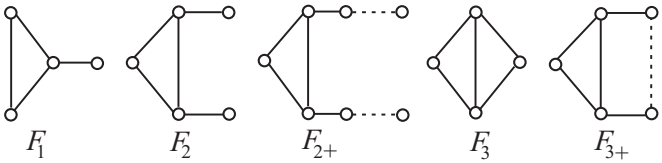


Fig. 3. The graphs used for characterization of Cartesian products with crossing number one.

The aim of the paper is to extend this result and characterize graphs G_1 and G_2 for which the crossing number of its Cartesian product $G_1 \times G_2$ equals two. Obviously, at least one of the graphs G_1 and G_2 does not contain a cycle. Otherwise the graph $G_1 \times G_2$ contains a subdivision of $C_3 \times C_3$ and it was proved in [11], that $cr(C_3 \times C_3) = 3$. If both graphs G_1 and G_2 contain a vertex of degree at least three, then the Cartesian product $G_1 \times G_2$ contains the graph $S_3 \times S_3$ as a subgraph. Asano in [2] proved that $cr(K_{1,3,n}) = 2\lfloor \frac{n}{2} \rfloor \lfloor \frac{n-1}{2} \rfloor + \lfloor \frac{n}{2} \rfloor$. The graph $S_3 \times S_3$ is isomorphic to the complete tripartite graph $K_{1,3,3}$ and therefore, $cr(S_3 \times S_3) = 3$. Hence, at most one of G_1 and G_2 contains a vertex with degree more than two. We solve the case if one of the graphs G_1 and G_2 is a cycle and we give the necessary and sufficient conditions for the case when the crossing number of their Cartesian product equals two. Let H be the graph in Fig. 1.

Theorem III.2 *Let G_1 is isomorphic to a cycle C_n, $n \geq 3$. Then $cr(G_1 \times G_2) = 2$ if and only if one of the following conditions holds:*
1) $G_1 = C_4$ and G_2 is homeomorphic with S_3,
2) $G_1 = C_3$ and G_2 is homeomorphic with S_4 or with the graph H.

Proof. It was proved in [7] that $cr(C_4 \times S_3) = 2$. This implies that for any subdivision S_3^s of the star S_3 the graph $S_3^s \times C_4$ has crossing number at least two. The drawing of the graph $S_3^s \times C_4$ in Fig. 4 shows that $cr(S_3^s \times C_4) \leq 2$. Hence, For any graph G_2 homeomorphic to the star S_3 we have that $cr(C_4 \times G_2) = 2$.

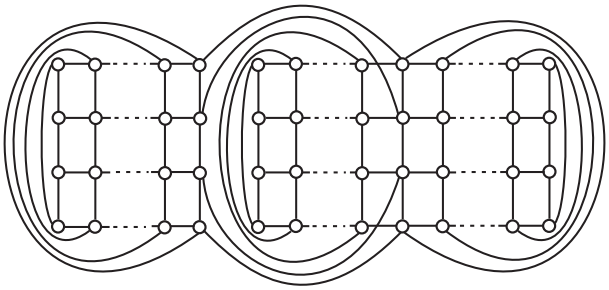


Fig. 4. The graphs $S_3^s \times C_4$ with two crossings.

The crossing number of the graph $C_3 \times S_4$ is two, see [9]. In [4] the value 2 for the crossing number of the graph $C_3 \times H$ is given. Hence, for any subdivision S_4^s of the graph S_4 and for any subdivision H^s of H, both graphs $C_3 \times S_4^s$ and $C_3 \times H^s$ have crossing number at least two. The reverse inequalities give the drawings in Fig. 5 and Fig. 6 of the graphs $C_3 \times S_4^s$ and $C_3 \times H^s$, respectively. This confirms that $cr(C_3 \times G_2) = 2$ for any graph G_2 homeomorphic to S_4 or H. It remains to prove that there are no other cycles C_n and no other graphs G_2 with $cr(C_n \times G_2) = 2$.

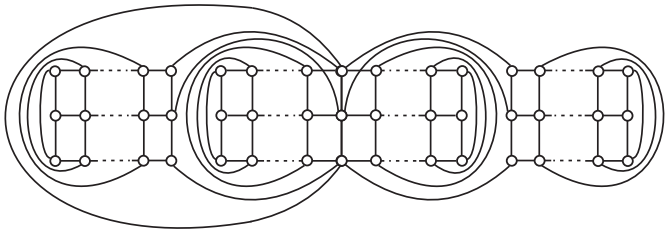


Fig. 5. The graphs $S_4^s \times C_3$ with two crossings.

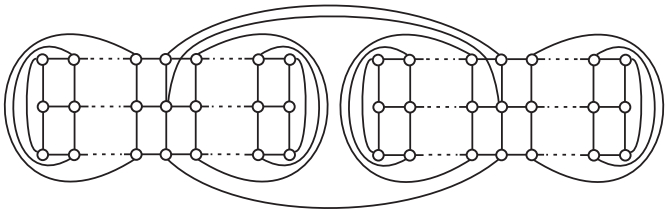


Fig. 6. The graphs $H^s \times C_3$ with two crossings.

Assume that $cr(G_1 \times G_2) = 2$. As $cr(C_n \times P_m) = 0$ for all $m \geq 1$, the condition $cr(C_n \times G_2) = 2$ enforces that the graph G_2 must contain a vertex of degree more than two, and the fact that $cr(C_n \times C_m) \geq 3$ for all $n, m \geq 3$ implies that G_2 does not contain a cycle. Hence, the graph G_2 must be a tree other than a path. Moreover, $G_1 = C_3$ or $G_1 = C_4$ because $cr(C_n \times S_3) \geq 3$ for $n \geq 5$, see [7]. The graph G_2 does not contain a vertex of degree more than four, otherwise the graph $G_1 \times G_2$ contains $C_3 \times S_5$ or $C_4 \times S_5$ as a subgraph. Both these graphs have crossing number more than two, see [3].

Consider first the graph $G_1 = C_4$. The graph G_2 does not contain a vertex of degree more than three, because $cr(C_4 \times S_4) = 4$, see [9]. The graph G_2 contains at most one vertex of degree three, otherwise $C_4 \times G_2$ contains a subgraph homeomorphic to the graph $H \times P_3$ with crossing number more than two, see Lemma II.1. Hence, the graph G_2 contains exactly one vertex of degree three. Every such graph is homeomorphic with the star S_3 and $cr(C_4 \times G_2) = 2$.

For $G_1 = C_3$, the graph G_2 has at most one vertex of degree four, otherwise $C_3 \times G_2$ contains a subgraph homeomorphic to the graph $J \times P_2$ with crossing number more than two, see Lemma II.2. The same fact implies that if G_2 contains one vertex of degree four, all other vertices are of degree at most two. So, G_2 is homeomorphic to the graph S_4 and $cr(C_3 \times G_2) = 2$. For the case when maximum degree of G_2 is three, lemma II.3 implies that G_2 has at most two vertices of degree three. Every connected graph with two vertices of degree three is homeomorphic to the graph H and in this case $cr(C_3 \times G_2) = 2$. As for the graph G_2 with less than two vertices of degree three $cr(C_3 \times G_2) < 2$, see [12], the proof is done. □

IV. Conclusion

Computing the exact value of crossing number of a given graph is in general an elusive problem. In this paper we solved only the case if one of the graphs G_1 and G_2 is a cycle and we proved the necessary and sufficient conditions for the case when the crossing number of their Cartesian product equals two. But there aren't all graphs for which the crossing number of the Cartesian product $G_1 \times G_2$ is two. It remains to prove the cases, if one of the graphs is a path.

Acknowledgment

The author thanks doc. RNDr. Marián Klešč, PhD. for help and precious advices.

References

[1] D. G. Akka, S. Jendroľ, M. Klešč, S. V. Panshetty, On line graphs with crossing number 2, *Univ. Beograd Publ. Elektrotehn Fak., Ser. Math.* **8** (1997), 3–8.
[2] K. Asano, The crossing number of $K_{1,3,n}$ and $K_{2,3,n}$, *J. Graph Theory* **10** (1986), 1–8.
[3] L. W. Beineke, R. D. Ringeisen, On the crossing numbers of products of cycles and graphs of order four, *J. Graph Theory* **4** (1980), 145–155.
[4] M. Draženská, M. M. Klešč, The crossing numbers of products of cycles with 6-vertex trees, *Tatra Mt. Math. Publ.* **36** (2007), 109–119.
[5] L. Y. Glebsky, G. Salazar, The crossing number of $C_m \times C_n$ is as conjectured for $n \geq m(m+1)$, *J. Graph Theory* **47** (2004), 53–72.
[6] F. Harary, P. C. Kainen, A. J. Schwenk, Toroidal graphs with arbitrarily high crossing numbers, *Nanta Math.* **6** (1973), 58–67.
[7] S. Jendroľ, M. Ščerbová, On the crossing numbers of $S_m \times P_n$ and $S_m \times C_n$, *Časopis pro pěstování matematiky* **107** (1982), 225–230.
[8] S. Jendroľ, M. Klešč, On graphs whose line graphs have crossing number one, *J. Graph Theory* **37** (2001), 181–188.
[9] M. Klešč, On the crossing numbers of Cartesian products of stars and paths or cycles, *Mathematica Slovaca* **41** (1991), 113–120.
[10] V. R. Kulli, D. G. Akka, L. W. Beineke, On line graphs with crossing number one, *J. Graph Theory* **3** (1979), 87–90.
[11] R. D. Ringeisen, L. W. Beineke, The crossing number of $C_3 \times C_n$, *J. Combinatorial Theory* **24** (1978), 134–136.
[12] R. Tindira, *Miery neplanárnosti grafov–priesečníkové čísla*, MSc Thesis, PF UPJŠ Košice, 2005.

Communication protocols in distributed simulation.

Igor Petz
Dept. of Informatics and Computer Science, FEI TU of Košice, Slovak Republic

igorp@centrum.sk

***Abstract*—This paper informs about communication protocols and architectures used in distributed simulation systems. Presented are TENA Architecture, DIS communication protocol and HLA communication architecture. TENA, DIS and HLA protocols are used for educational/testing simulations. TENA is used for live simulation purposes, for smaller simulations, DIS is standard communication protocol used in constructive simulations and for simulation federations and HLA is communication architecture prepared for larger simulations where different simulator types are used. HLA is most sophisticated and most variable way how to communicate between simulation nodes/federations.**

***Keywords*— Communication protocols, DIS, Educational simulation, HLA, TENA,**

I. Introduction

In the field of educational simulation different simulation protocol types are used. In present days TENA (Test and Training Enabling Architecture), DIS (Distributed Interactive Simulation) and HLA (High Level Architecture) simulation protocols are used. All these protocols/architectures are used for different purposes.

TENA is a communication protocol, which was defined as universal protocol, but in general it is used for simulation with simpler traffic. TENA was built on the concepts pioneered in JADS(Joint Advanced Distributed Simulation) and the HLA to support interoperability between the live testing/training range domain and the larger M&S community in this area.

DIS is standard communication protocol that is used for larger simulations. Larger simulations means that communicating cells are far from each other and larger means bigger data amount too.

HLA is one of the latest communication protocols. HLA is defined in IEEE standards.

II. Communication protocols

A. TENA

Communication in area of live testing/training range domain is based on TENA communication. TENA is defined as architecture and was defined in Foundation Initiative 2010, which was supported by Central Test and Evaluation Investment Program (CTEIP). This program is running under sponsorship of US MoD (United States Ministry of Defense). Under CTEIP project HLA architecture was defined for modeling and simulation too.

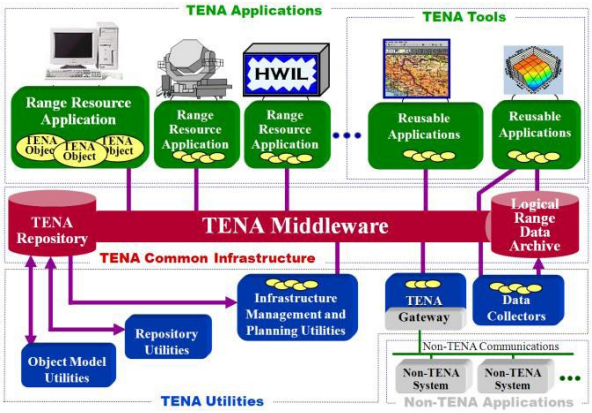


Fig. 1. TENA infrastructure[2]

TENA core component is TENA Common interface consisting from TENA Middleware, TENA Repository and TENA Logical Range Data Archive. TENA supports usage of more tools and applications.

As shown on Figure 1, TENA Middleware is core of TENA. TENA Middleware uses UML model oriented automatic code generator. TENA Middleware includes prepared API, that is less susceptible to generate errors than for example HLA/RTI API or other APIs using DIS. Integrated high level abstraction combined with infallible API allows quickly and precisely define concept of user's application. Reusable components are simplifying process of application development.

TENA Repository is storing all relevant data about TENA. It is big database, where unifying mechanisms for creating uniform communication environment are stored. TENA repository includes some standard TENA objects definition, implemented meta-data, TENA tools and utilities, software libraries for TENA Middleware

TENA Object model is representing models in TENA environment. It is defined as summary of abstract ideas described in UML or in own TENA text language marked as TDL (TENA Definition Language). These representations are transformed into C++ code and are represented in TENA Middleware.

TENA Logical Range Data Archive includes data that are tight to connected simulation ranges. Unlike TENA Repository, TENA Logical Range Data Archive can be divided between more computers.

TENA Middleware is high performance real time communication infrastructure with low loss rate. It is integrated in all connected applications together with object definition. TENA Middleware supports TENA meta-model

and communication between TENA Object models.

```
package Example
{
  local class Place
  { float64 orientation;
        float64 x;
        float64 y;
        optional float64 z;
  };
};
```

Fig. 2. TENA Object Model definition example in TDL

TENA brings common API for communication between objects, data transmitting and for connection into TENA Logical Range Data Archive. TENA is live architecture used in Slovakia too. TENA connects entities in MILES 2000 live simulation application (I-HITS implementation). This simulation is running on MTA (Military Training Area) Lešť.

B. DIS

DIS defines infrastructure for real-time simulation that can be distributed on different locations. This infrastructure supports connection between different simulation types. Live entities, constructive entities or virtual units can be connected. DIS is defined in IEEE Standard 1278.xx

DIS standard is used since 1992. Thanks to its simplicity it is used in many military and non-military simulations till now.

Basic DIS characteristics are:

Data packets are defined as PDU (Protocol Data Unit)

Set of rules for PDU traffic is defined

PDU enumeration is predefined

In DIS environment each connected node is working autonomously. Each simulator provides simulation of respective entities and is responsible for them. Standard DIS protocol in multicast or broadcast is used for data exchange. Communication is not controlled by one computer, no time synchronization is resolved. Each event and entity update has timestamp. Entire simulation is not depended on one node, it is possible to connect node to simulation after simulation starts.

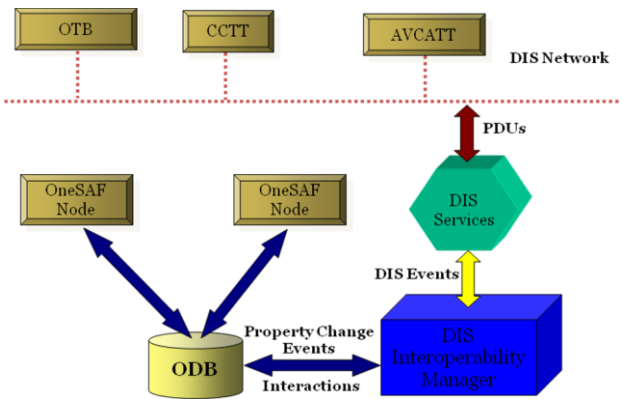


Fig. 3. Connecting OneSAF simulator into DIS network

DIS is an open architecture and can be implemented on different platforms. Enhancements can be made through experimental PDUs. On the market there are many DIS compatible simulators.[3][4]

In Slovakia is DIS used for virtual simulators connection, for OneSAF simulator connections and as general simulation network.

Field size (bits)	Fire PDU Fields	
96	PDU Header	Protocol Version—8-bit enumeration
		Exercise ID—8-bit unsigned integer
		PDU Type—8-bit enumeration
		Protocol Family—8-bit enumeration
		Timestamp—32-bit unsigned integer
		Length—16-bit unsigned integer
		Padding—16 bits unused
48	Firing Entity ID	Site—16-bit unsigned integer
		Application—16-bit unsigned integer
		Entity—16-bit unsigned integer
48	Target Entity ID	Site—16-bit unsigned integer
		Application—16-bit unsigned integer
		Entity—16-bit unsigned integer
48	Munition ID	Site—16-bit unsigned integer
		Application—16-bit unsigned integer
		Entity—16-bit unsigned integer
48	Event ID	Site—16-bit unsigned integer
		Application—16-bit unsigned integer
		Event Number—16-bit unsigned integer
32	Fire Mission Index	32-bit unsigned integer
192	Location in World Coordinates	*X*-component—64-bit floating point
		Y-component—64-bit floating point
		Z-component—64-bit floating point
128	Burst Descriptor	Munition—64-bit Entity Type record
		Warhead—16-bit enumeration
		Fuse—16-bit enumeration
		Quantity—16-bit unsigned integer
		Rate—16-bit unsigned integer
96	Velocity	*X*-component—32-bit floating point
		Y-component—32-bit floating point
		Z-component—32-bit floating point
32	Range	32-bit floating point
Total Fire PDU size = 768 bits		

Fig. 4. DIS Fire PDU structure

C. HLA

HLA for Modeling and Simulation (M&S), sponsored by the Defense Modeling and Simulation Office (DMSO), created a technical architecture for all military simulations to promote interoperability and reuse among simulation programs. HLA definition is tight to this intention and is representing new generation after DIS protocol. HLA allows communication between simulations without platform dependency. Communication must be managed by RTI (for example MAK RTI). RTIs for HLA are COTS (Commercial, off-the-shelf) products.

HLA is defined in IEEE 1516-2000 and is NATO standard (STANAG 4603).

HLA compatible simulations are connected into federation. Connected simulations are using common object template. Objects have defined data attributes and parameters.

Communication is running through events. Events are generated by RTI, where RTI decides who will receive generated event. Events have data parameters too.

Main HLA components are:

- Interface specification, where communication between RTI and external federates is defined
- Object Model Template
- HLA rules

Interface specification – RTI

RTI (RunTime Infrastructure) are applications managing HLA federation. RTIs are commercial APIs, where following subjects are defined:

Federation management

- Declaration management
- Object Management
- Ownership Management
- Time Management
- Data Distribution Management
- Support Services

Object Model Templates provides common framework for simulation communication in simulation federation. To create Object model templates it is needed to define public objects and attributes for whole federation – Federation Object Model (FOM) and for participating simulations (federates) – Simulation Object Model (SOM).

HLA rules define responsibility rules for federates and simulations. They are as follows [5]:

Federation must have FOM based on HLA Object template

All FOMs are represented on connected federates, not on RTI

All communication, all FOM changes must go through RTI

Communication with RTI is only based on interface specification

One attribute can be owned only by one federate

Simulations (federate) can have own SOM, created upon Object Model Templates

Single federates can manage own SOM, his attributes and rules when they will send information about attributes.

Each joined simulation can manage own local time so, that data are synchronized with other connected simulations.[1]

III. Conclusion

All described protocols are used in the field of educational simulations in Slovak military and in NATO countries. TENA is used for live simulation, where long time latency occurs (connected entities can be partially out of radio range). DIS is used for constructive simulation, where big entity number occurs. DIS is also used for communication through WAN. Virtual simulators are using DIS too. All new simulators are prepared to be used in HLA simulation federation, but HLA difficulty and problems with RTI configuration are still forcing use of DIS instead of HLA.

Although all these tree protocols/architectures are different, there are commercial solutions (gateways) to connect them. Good applications that are helping to interconnect TENA, DIS and HLA are for example VR Link or VR Exchange.

Acknowledgment

This work is supported by VEGA grant project No. 1/0646/09: "Tasks solution for large graphical data processing in the environment of parallel, distributed and network computer systems."

References

[1] U.S. Defense Modeling and Simulation Office (2001). RTI 1.3-Next Generation Programmer's Guide Version 4. U.S. Department of Defense.

[2] TENA, The Test and Training Enabling Architecture,Architecture Reference Document, Version 2002, Review Edition, November 2002, Foundation Initiative 2010 Project Office

[3] OTB Version 2.5 International User's Guide

[4] IEEE Std 1278.1a-1998

[5] IEEE 1516-2000.(2000). Standard for modeling and simulation (M&S) HLA high level architecture – framework and rules 2000. IEEE Standards

Modelling and control of systems with hybrid dynamics

Ľuboš POPOVIČ

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

lubos.popovic@tuke.sk

***Abstract*—This paper describes theoretical basis for modelling and control of hybrid systems, which involves continuous and discrete dynamics. Also describes nonlinear mathematical model of hybrid system represented by two tanks with liquid, which dynamics changes from system without interaction to system with interaction, linearization about two different working points for obtaining state space representations of this two modes of dynamics, design optimal controller for tracking reference trajectory, and presents simulation results for selected example.**

***Keywords*—Dynamical system, Hybrid system, Optimal controller with integrator, Reference trajectory, Two tanks with liquid.**

I. Introduction

Hybrid systems are currently the most discussed problems of control theory. Thus, hybrid systems are systems that involve continuous and discrete variables [5]. Existence of both types of variables, continuous and discrete gives the system hybrid character. Evolution of hybrid systems can be described by using equation, which contains a mixture of logic, discrete and continuous variables. The continuous dynamics of such systems may be continuous-time, discrete-time, or mixed, but is generally given by differential or difference equations. The main contribution in control theory are models describing the interaction between continuous dynamics described by differential or difference equations, and logical components described by finite state machines, if-then-else rules, propositional and temporal logic [1]. Design of controller for hybrid system is much complex, because controlled system switches between these various dynamics and therefore it is necessary design controller for each dynamics and include designed decision unit to control scheme for choosing and switching between controllers.

II. Two Tanks With Liquid

A. Nonlinear mathematical model

As an example of system with hybrid dynamics consider model of two tanks with liquid as shown on figure (Fig. 1). Dynamics properties of this system are changing over time and are described by system of differential equations. Transition between these dynamics occurs in certain switching time [2].

These two tanks on figure (Fig. 1) are in different height. Thus switch between dynamics occurs in moment when level of liquid in second tank exceeds height of bottom of the first tank. In this moment the mathematical model of whole system changes from system without interaction to system with interaction.

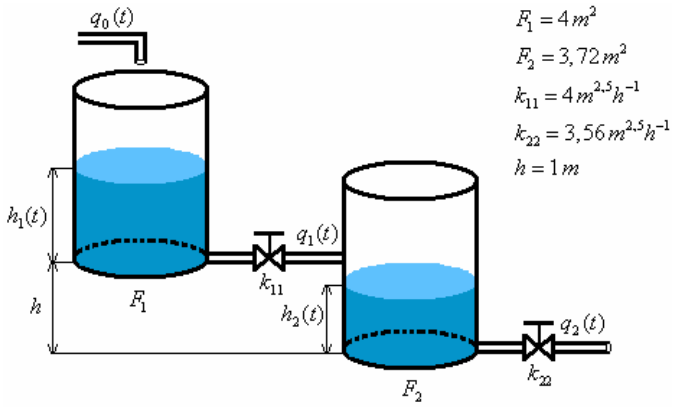


Fig. 1. Hybrid system of two tanks with liquid

For obtaining both models, it is necessary to formulate material balance for weight:
(Sum of mass flow on input) = (Sum of mass flow on output) + (Rate of accumulation of mass in whole system) [3]. Then for system of two tanks without interaction one can obtain:

$$q_0(t) = k_{11}\sqrt{h_1(t)} + F_1\frac{dh_1(t)}{dt}, \tag{1}$$

$$k_{11}\sqrt{h_1(t)} = k_{22}\sqrt{h_2(t)} + F_2\frac{dh_2(t)}{dt}, \tag{2}$$

and for system of two tanks with interaction:

$$q_0(t) = sign\left(h_1(t)-\left(h_2(t)-h\right)\right)k_{11}\sqrt{\left|h_1(t)-\left(h_2(t)-h\right)\right|} + F_1\frac{dh_1(t)}{dt}, \tag{3}$$

$$sign\left(h_1(t)-\left(h_2(t)-h\right)\right)k_{22}\sqrt{\left|h_1(t)-\left(h_2(t)-h\right)\right|} = F_2\frac{dh_2(t)}{dt} + k_{22}\sqrt{h_2(t)}, \tag{4}$$

Where F_1 is cross-section of the first tank, F_2 is cross-section of the second tank, k_{11} is valve constant of the first tank, k_{22} is valve constant of the second tank, h is height of bottom of the first tank (switching condition), $q_0(t)$ is input flow of liquid to the first tank (actuating variable), $q_1(t)$ is output flow of liquid from the first tank and simultaneously input flow of liquid to the second tank, $q_2(t)$ is output flow of liquid from the second tank. From equations (1) – (4) it is possible to create non-linear model shown on figure (Fig. 2).

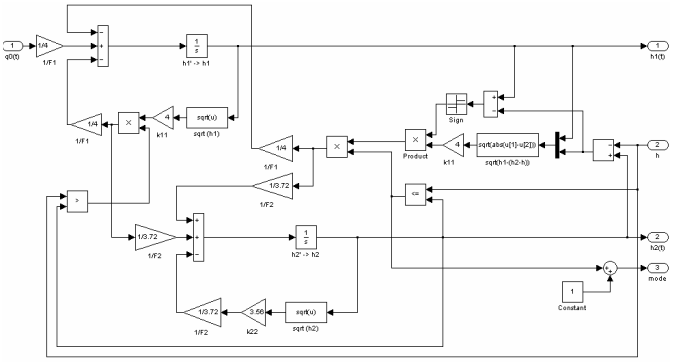


Fig. 2. Nonlinear model of two tanks with liquid designed in Simulink

Hybrid character of nonlinear model of two tanks with liquid is shown on figure (Fig. 3), where input flow of liquid $q_0(t)$ to the first tank was changed by step.

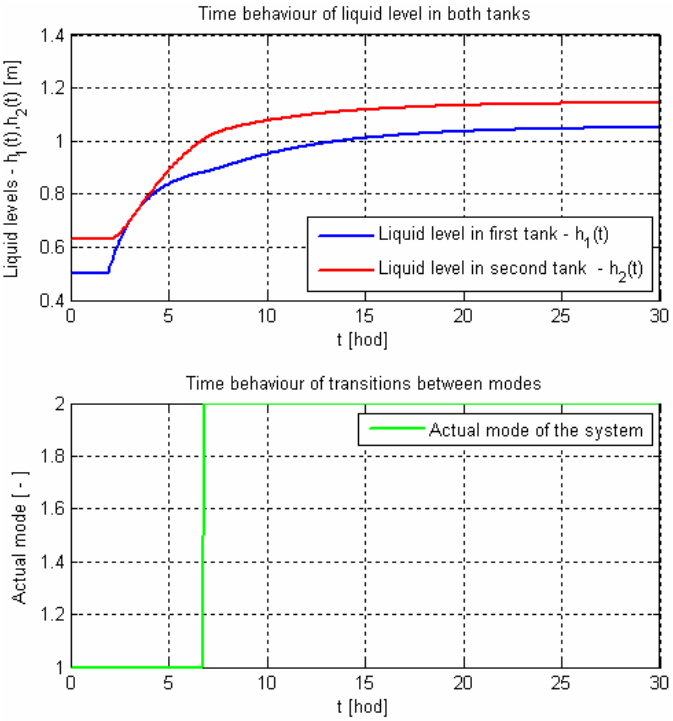


Fig.3. Time behaviour of liquid level and mode in both tanks

B. Linearized mathematical model

For successful design of controller, linearized mathematical model of controlled system should be available. In this case of hybrid system, it is necessary to obtain two linearized mathematical models, one for system without interaction and second for system with interaction.

As has been written, (1) and (2) represents behavior of system without interaction, but with nonlinear differential equations. For obtaining linearized mathematical model, nonlinear functions such as square roots must be linearized by using Taylor's series.

Due to the linearization, dynamical mathematical model for system without interaction was analyzed and mathematical model in steady-state was substracted from it. In steady-state (if input flow of liquid is constant in the long term), liquid level in both tanks have steady-state value and $dh_{10}/dt = 0$, $dh_{20}/dt = 0$, where h_{10} is liquid level of first tank and h_{20} is liquid level of second tank in steady-state. Liquid level of first tank was chosen and thus $h_{10} = 0.5\ m$. From steady-state it was possible to calculate all others variables such as steady-state value of input flow or liquid level of second tank. As input, inflow of liquid $q_0(t)$ was considered and as output, liquid level $h_2(t)$ in second tank was considered. These variables were used to derive linearized perturbation dynamical model of two tanks with liquid without interaction and after short mathematical modification, transfer function and equivalent state-space representation were obtained:

$$F_1(s) = \frac{\Delta H_2(s)}{\Delta Q_0(s)} = \frac{0.1901}{s^2 + 1.3094s + 0.4259}, \tag{5}$$

$$A = \begin{bmatrix} 0 & 1 \\ -0.4259 & -1.3094 \end{bmatrix},\ B = \begin{bmatrix} 0 \\ 0.1901 \end{bmatrix},\ C = [1 \quad 0]. \tag{6}$$

Equations (3) and (4) represent behavior of system with interaction, but still with nonlinear differential equations. It is therefore necessary to linearize these nonlinear differential equations in new steady-state, with assumption of omitting expression *sign*, and *absolute values*. Liquid level in second tank for system with interaction also depends on height difference between bottoms of both tanks – h. This difference should be included in the derivation of linearized perturbation dynamical model of two tanks with interaction. Liquid level of first tank was chosen and thus $h_{10} = 1.2625\ m$. From steady-state it was possible to calculate all others variables such as steady-state value of input flow or liquid level of second tank. These variables were used to derive linearized perturbation dynamical model of two tanks with liquid with interaction and after short mathematical modification, transfer function and equivalent state-space representation were obtained:

$$F_2(s) = \frac{\Delta H_2(s)}{\Delta Q_0(s)} = \frac{0.2688}{s^2 + 2.5011s + 0.4259}, \tag{7}$$

$$A = \begin{bmatrix} 0 & 1 \\ -0.4259 & -2.5011 \end{bmatrix},\ B = \begin{bmatrix} 0 \\ 0.2688 \end{bmatrix},\ C = [1 \quad 0]. \tag{8}$$

C. Comparison of nonlinear and linearized mathematical models

In this section comparison of nonlinear and linearized mathematical models of two tanks without and with interaction is presented. Simulation results were obtained in language for technical computing – Matlab/Simulink [6] and are shown on figures (Fig. 4, Fig. 5).

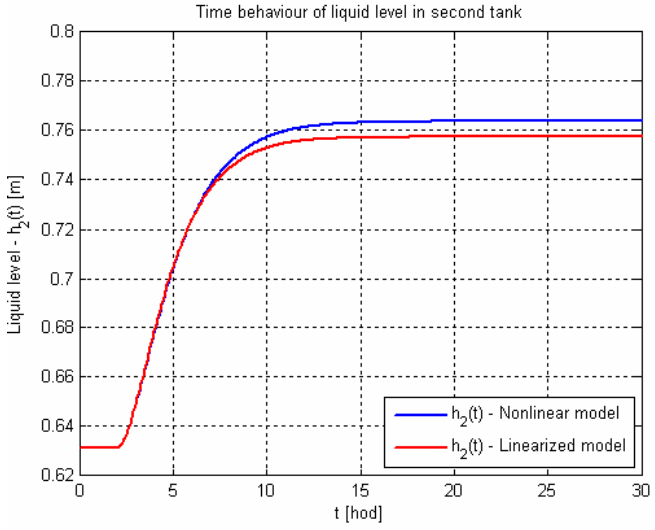


Fig. 4. Time behaviour of liquid level in second tank – two tanks with liquid without interaction

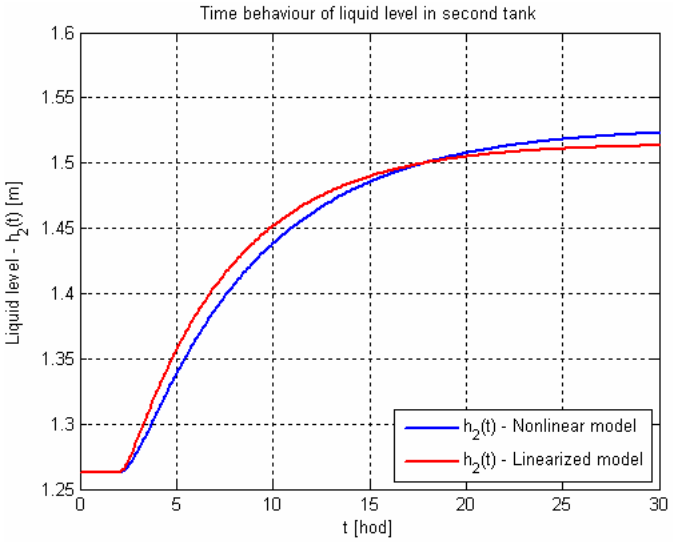


Fig. 5. Time behaviour of liquid level in second tank – two tanks with liquid with interaction

Both linearized mathematical models were used to design optimal state controller with integrator for tracking reference trajectory.

III. Design of Optimal Controller With Integrator For Tracking Reference Trajectory

Consider linearized dynamical system in state space representation:

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{9}$$

$$y(t) = Cx(t). \tag{10}$$

Problem of tracking reference trajectory is keep output variable $y(t)$ on reference trajectory $y_{ref}(t)$. For this reason control error was defined as:

$$e(t) = y_{ref}(t) - y(t). \tag{11}$$

Whereas control strategy using quadratic optimization doesn't contain integrator, it is necessary to integrate control error by integrator, describe by equation:

$$\dot{\upsilon}(t) = y_{ref} - Cx(t), \tag{12}$$

where υ is outputs of integrators. In this way a permanent control errors should be removed. Let's augment system (9), (10) expanding by equation (12):

$$\begin{bmatrix} \dot{x}(t) \\ \dot{\upsilon}(t) \end{bmatrix} = \begin{bmatrix} A & 0 \\ -C & 0 \end{bmatrix} \begin{bmatrix} x(t) \\ \upsilon(t) \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} u(t) + \begin{bmatrix} 0 \\ y_{ref}(t) \end{bmatrix}, \tag{13}$$

$$\begin{bmatrix} y(t) \\ \upsilon(t) \end{bmatrix} = \begin{bmatrix} C & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} x(t) \\ \upsilon(t) \end{bmatrix}. \tag{14}$$

Equation (13) and (14) can be rewritten in the form:

$$\dot{x}_1(t) = A_1 x_1(t) + B_1 u(t) + H y_{ref}(t), \tag{15}$$

$$y_1(t) = C_1 x_1(t), \tag{16}$$

where $x_1(t) = \begin{bmatrix} x(t) \\ \upsilon(t) \end{bmatrix}$, $A_1 = \begin{bmatrix} A & 0 \\ -C & 0 \end{bmatrix}$, $B_1 = \begin{bmatrix} B \\ 0 \end{bmatrix}$, $H = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

New augmented system has *(n+p)-th* order, where *n* is order of original system and *p* is a number of outputs. The main task is design a control algorithm, which will be able to compensate a vector Hy_{ref}, so the task is reduced to optimization problem with known disturbances. Criteria function is in form:

$$J = \frac{1}{2}\int_0^\infty \left[\langle e(t), Qe(t) \rangle + \langle u(t), Ru(t) \rangle \right] dt, \tag{17}$$

and control action is in form:

$$u(t) = -R^{-1}B_1^T \left[K_1 x_1(t) - h_1 \right]. \tag{18}$$

Constant matrix K_1 is steady-state solution of differential Riccati equation:

$$\dot{K}_1(t) = -K_1(t)A_1 - A_1^T K_1(t) + K_1(t)BR^{-1}B^T K_1(t) - C_1^T Q C_1, \tag{19}$$

and compensating vector h_1:

$$h_1 = \left\{ \left[A_1 - B_1 R^{-1} B_1^T K_1 \right]^T \right\}^{-1} \left[K_1 H - C_1^T Q H \right] y_{ref}. \tag{20}$$

If (20) is substituted to (18), the control action is:

$$u(t) = -R^{-1}B_1^T K_1 x_1(t) + R^{-1}B_1^T \left\{ \left[A_1 - B_1 R^{-1} B_1^T K_1 \right]^T \right\}^{-1} \cdot \left[K_1 H - C_1^T Q H \right] y_{ref}, \tag{21}$$

where:

$$L_1 = R^{-1}B_1^T K_1, \tag{22}$$

$$N_1 = R^{-1}B_1^T \left\{ \left[A_1 - B_1 R^{-1} B_1 K_1 \right]^T \right\}^{-1} \left[K_1 H - C_1^T Q H \right], \tag{23}$$

and then:

$$u(t) = -L_1 x_1(t) + N_1 y_{ref}. \tag{24}$$

If vector $x_1(t)$ spreads to the vectors $x(t)$ and $\upsilon(t)$, the control action is in the form:

$$u(t) = -Lx(t) - M\upsilon(t) + N_1 y_{ref}. \tag{25}$$

Matrix L and M are chosen from matrix L_1 and have appropriate dimensions, corresponding to dimensions of original system and numbers of integrator outputs [4].

According to equation (25) structural scheme of optimal controller with integrator can be designed (Fig. 6).

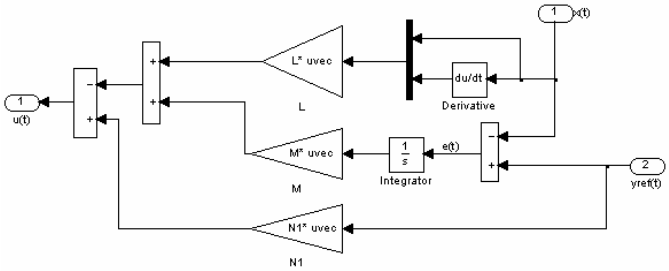


Fig. 6. Structural scheme of optimal controller with integrator in Simulink

This algorithm was used to design tracking controllers for both linearized dynamical models – system without interaction and system with interaction.

These controllers for tracking reference trajectory then were implemented in control structure for control of nonlinear hybrid system of two tanks with liquid. As has been written, switching between system without and with interaction represent hybrid dynamics.

IV. Results

In this section results of tracking reference trajectory are presented. Reference trajectory in case of two tanks with liquid means various required liquid level in second tank. Two tracking optimal controllers with integrators were designed as described in section III., and they are used to control nonlinear hybrid system of two tanks with liquid. Control structure for

tracking reference trajectory of system with hybrid dynamics is shown on figure (Fig. 7.)

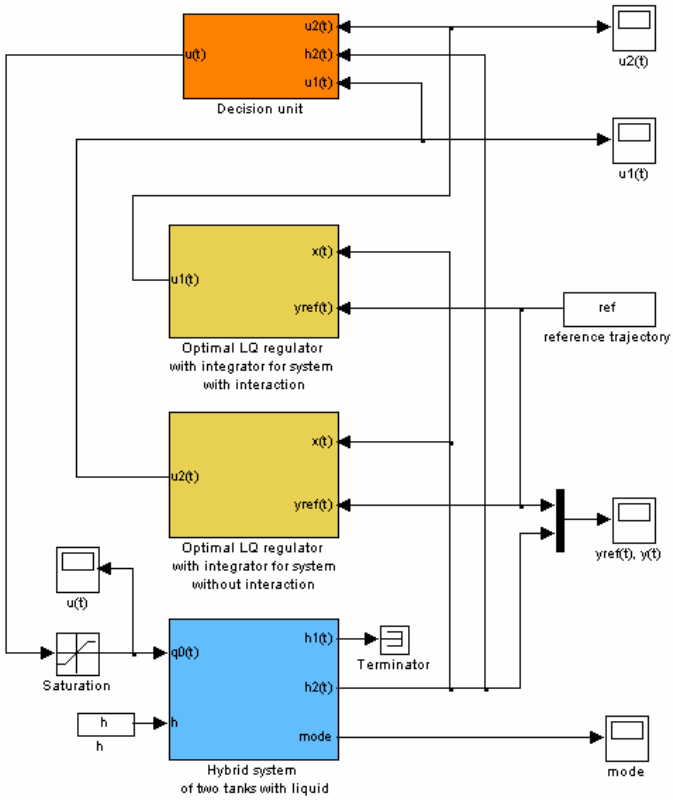


Fig. 7. Control structure for tracking reference trajectory of hybrid system – two tanks with liquid

As seen on figure (Fig. 7) decision unit must be added to the control structure to select and switch between controllers. Decision unit contains comparator, which compares actual liquid level in second tank with height of bottom of the first tank – *h*, which is also switching condition and selects appropriate control actions. Block "Saturation" was added just for constrain control actions. On figure (Fig. 8) are shown results of simulation for tracking reference trajectory and mode changes between system of two tanks with liquid without interaction and with interaction.

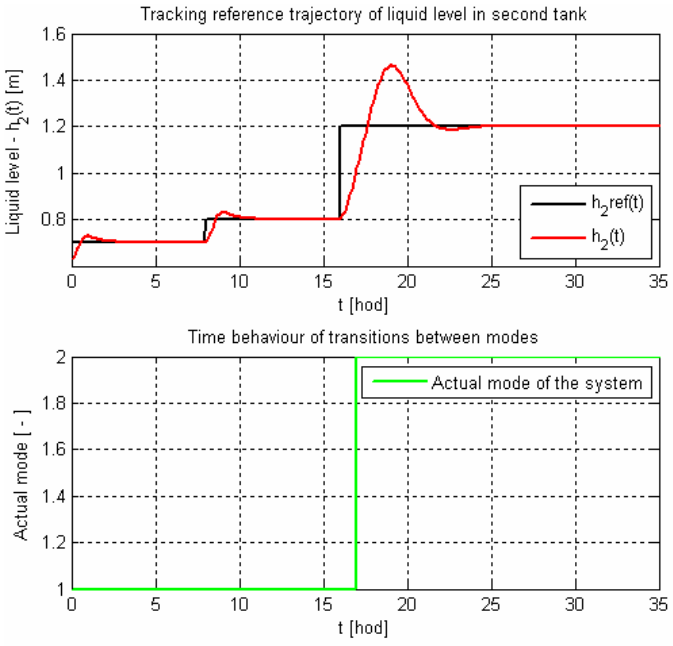


Fig. 8. Results of tracking reference trajectory of liquid level in second tank

V. Conclusion

Presented results and approaches to modeling and control of system with hybrid dynamics, shows new possibilities in designing of control systems. Presented control structure involves dynamical system on the lowest level, controllers on higher level and decision unit on the highest level of control system. This type of control structure is often called in literature as multi-level hierarchical control system. In the future this control structure could be extended into the discrete-time domain.

Acknowledgment

This publication is the result of the project implementation Centre of Information and Communication Technologies for Knowledge System (project number: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] Popovič, Ľ. – *Modelling of systems with hybrid dynamics.* The 9th Scientific Conference of Young Researchers at FEI TU of Košice, p. 79 – 82, ISBN: 978-80-553-0178-5

[2] T. Hirmajer – M. Fikar: *Optimal control of system with hybrid dynamics*. In: AT&P Journal. vol. 8, no. 12, 2005, pp. 81-84.

[3] M. Bakošová – M. Fikar – Ľ. Čirka: *Laboratórne cvičenia zo základov automatizácie* – dostupné na internete: http://www.kirp.chtf.stuba.sk/~cirka/vyuka/lcza/

[4] J. Sarnovský – A. Jadlovská – P. Kica: *Teória optimálnych a adaptívnych systémov*. Košice: Elfa, 2005, ISBN 80-8086-020-3, p. 173

[5] M. S. Branicky: *Studies in Hybrid Systems: Modelling, Analysis, and Control*, dissertation work. Massachusetts, 1995. pp. 198.

[6] Mathworks: *Full Product Family Help – Matlab Help*. Help for The Language of Technical Computing – Matlab/Simulink.

Multi-Agent Serving System: an open source middleware for artificial intelligence and robotics

Tomáš Reiff, Zlatko Fedor, Miron Kuzma

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

{tomas.reiff, zlatko.fedor, miron.kuzma}@tuke.sk

***Abstract*—This paper gives an overview of MASS, an open source middleware for artificial intelligence and robotics. MASS has a client-server architecture and can be used in a manner that clients work for a server or that server serves clients – and even many at the time thanks to multi-threaded system core. To separate middleware from actual functionality, MASS is using the plugin approach. Other properties of MASS can be stated as: simplicity, incrementality, user friendly interface and that it is based on .NET platform. Currently our middleware supports three robotic platforms: Sony AIBO, Lego NXT and partially Nao from Aldebaran Robotics.**

***Keywords*—middleware, artificial intelligence, robot, plugin, vision, speech**

I. Introduction

Artificial intelligence had always something to do with robots. Some plain explanation of this connection could be that we always wanted robots to be intelligent and their intelligence can only be called artificial. But at this time, everything is much more complex. There is a huge field of a research called robotics and another huge field called artificial intelligence. These fields should cooperate but the reality is often different. People from robotics are re-inventing methods from artificial intelligence and artificial intelligence people are working with commercial robots to demonstrate their methods.

There can be many reasons of this behavior. One can be that people from artificial intelligence are interested in the research of new methods so they are not interested in being an implementation support or consultants for a single robotic platform out of many. On the other hand people from robotics already started their research of various methods with the first robots they made, because they wanted to see them acting. Now they have their own branch of algorithms which are maybe reusable for the new platforms so they are not interested in study of artificial intelligence.

The proposed middleware is trying to work this out in some way. First of all, MASS is not trying to upload the functionality on the robot but it is using the robot as an input/output device while the processing is on a normal computer. With this approach a supported robot is only a plugin, i.e. small library with functions, which has a functionality to get video, audio, sensor data, etc. from the robot and the functionality to control a robot movement. This makes the things much simpler. Imagine that you have implemented a method for visual localization (as a plugin), which needs to react on a video stream by some movement. Then if a new robot plugin is made, you can just connect to its video stream and use its movement control to demonstrate your method on it. Nothing has to be re-made or changed.

II. Design Goals

We think that it is nearly impossible to develop a framework for artificial intelligence and robotics which would suit all the needs in these broad areas. MASS was designed to meet a specific set of challenges encountered during the years in our lab.

The biggest challenge had something to do with a fact that we have usually produced one purpose programs which were forgotten after we stopped using them and many times we have programmed the same functionality again. Another important challenge was to use different programming language because the usual C/C++ is not that popular for students anymore. Actually there were also many other challenges and together they formed the philosophical goals of MASS which can by summarized as:

- Plugin approach
- .NET based
- Client-Server architecture
- Simple
- Incremental
- User friendly
- Free and open source

In this section, we will elaborate these philosophies and show how they have influenced the design and implementation of MASS.

A. Plugin approach

Best way how to understand the reason why we have chosen the plugin approach and why it is suitable for such a middleware is an example. Imagine you want to create a face detection program which will work with the webcam and will use some kind of a neural network.

The worst approach, in way of reusability, would be to write a single executable program which will contain webcam frame capture and the neural network. When you will want to use another method to detect faces, you would have to implement it inside the same executable and change the "Do" function or make a new executable and copy paste at least the frame capture part to avoid re-programming it.

A better approach would be to create two libraries and one executable. First library would contain only the webcam frame

capture functions and the second one the method for face detection. Executable will just use appropriate functions from those libraries in appropriate order. When you will want to use another method, you will add it to the second library and modify the "Do" function in executable. It looks very similar to the first approach, but in this way it would be easier to reuse your code in future development by copying the library instead of copying the code.

But the plugin approach is a way better. Like in the second approach, you would create two libraries and one executable. First library would contain the webcam frame capture function and the second one the face detection method. Moreover these libraries would contain some information about their functionality. Now the executable would have an ability to search for such libraries in some folders and collect its information. The important part would be that you will choose which functions from which libraries will be executed in which order and in fact you will dynamically create the "Do" function without actually programming it. When you will want to use another method, you will create a new library with some information (let's call it now plugin) and you will just run the executable and choose that after the frame is captured, a function from this new library will be called to actually detect faces with the different method.

In an ideal case, plugin approach leads to a state that the only things you program are new algorithms.

B. .NET based

We have chosen .NET framework due to many reasons. This framework is a huge set of libraries with colorful functionality and everything is managed by one company – Microsoft. This is a warranty that it will go in some plotted direction and you won't need to rely on some untested code with unclear origin and license. Moreover, you can use .NET framework with three different languages C#, C++ and Visual Basic and all the languages are interpreted. This means that anywhere you have .NET framework installed, you will be able to run your functionality efficiently even without compilation and this is quite important for the plugin approach. One disadvantage of this framework is that it is officially available only on Windows platform but there is already its unofficial port to Linux and Mac called Mono. And yes, we would like to migrate to Mono but only if there will be an adequate replacement of Visual Studio IDE which is probably the biggest advantage of .NET framework.

It is true that Visual Studio is a commercial and expensive product (however not as expensive as Matlab) but on the other hand nearly every university is a member of MSDN Academic Alliance. This means students can have it for free and even with free Windows operating system. Once you are working with Visual Studio, your productivity will increase rapidly. Together with higher level language, the only one thing you will need to think about will be your algorithm.

C. Client-Server architecture

This architecture was chosen to make the system multi-agent. Imagine that you have a server with configured system for object recognition and you want it to recognize various objects. After you provide few images of nearby objects, you will find out that you are out of other objects. Here can be the multi-agent approach useful and even people from other continents can learn the system to recognize broad range of various object. They are even able to learn reasonable count of objects at a time because of multi-threaded core of MASS.

Here it is important that if client connects to a server, it is able to choose which ActualSystem, i.e. functional set of plugins, it will use, possibly choose other options and then receive all the necessary plugins with all needed files. Usual procedure of this initialization process is showed in Table 1.

Client	Server
Client type →	← List of running systems
Chosen system →	← List of Inputs available for chosen system
Chosen Input →	← List of Outputs available for chosen system
Chosen Output →	← List of necessary plugins and externals
List of needed plugins and externals →	← Needed plugins and externals, ActualSystem settings

Table 1 The usual client initialization process

After this initialization is further communication managed by Flow plugin used in chosen ActualSystem. Usually client sends commands with data and server sends back the results. The communication between clients and server is done by simple socket connections and MASS has built-in functions for sending and receiving of everything. You only need to take care that what is send is also received on the other side.

D. Simple

We have chosen to use a close set of plugin types, where plugin type has something to do with the general purpose of plugin. Then, every plugin type has a specific interface, i.e. list of defined methods and properties, so we know how to access functionality of every plugin of that plugin type. Because each plugin contains basic information like its name, description and type, we can cast it to its plugin type interface and then dynamically use it.

After some time of MASS development, we found out that this closed set of plugin types is sufficient with 7 members. Fig. 1 shows the MASS plugin types in the system scheme.

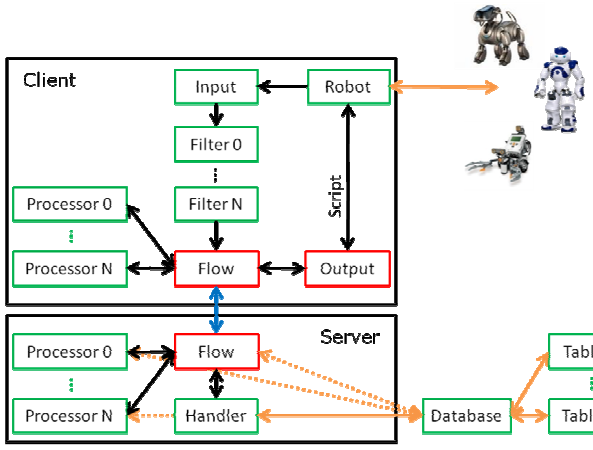


Fig. 1 MASS scheme – Red plugin types are required for every ActualSystem, Green are optional. Blue arrow represents socket connection

Flow

Flow plugin is on both sides of the system. On the client side it contains a form which is used to control the system runtime and sending the commands and its related data to the server part. Server part receives the commands and its data, makes the processing and sends back the results.

Output

Output plugin is showing the results received from the server part as a response to Flow commands and related data sent from the client part. Output can contain a script which reacts on current results. This script is being compiled dynamically and is usually used to represent a robot behavior.

Input

Input plugin is providing some kind of data to the system. Usually it has a thread which is collecting data and a thread which is taking the collected data, putting them into the system chain and waiting until the data are processed.

Filter

Filter plugin is receiving data from a previous plugin which could be Input or another Filter. After it gets data, it can transform them but it cannot change the form of data. That means that image data will change their look but they will be still image data.

Processor

Processor plugin is processing the data which are provided by a simple function call usually from the Flow plugin. In contrast with the Filter plugin, Processor can change also the form of the data, e.g. image data can be changed to a list of local features. This plugin type can be placed on both sides of the system.

Handler

Handler plugin is used for manipulation with database data or data in general. For example it can find the nearest neighbor for some descriptor from the database full of descriptors. Handler is therefore a suitable plugin where to implement clustering or classification functionality.

Robot

Robot plugin contains all the functionality needed for working with the robotic platform which it represents. Usually it contains robot specialized Input and Output plugins and functions for controlling the robot movement. All this robotic functions can be then accessed by the Output script.

E. Incremental

Incrementality is quite important for systems which are able to gain some kind of knowledge. Being incremental also means that the system is also able to store more and more knowledge. With this comes the need of database.

Because our middleware is programmed under .NET framework, we were able to develop our custom, tiny but handy C# database. With this database you can store any class and you only need to modify one row of it. It supports also linked tables and partially indexing. Because the Table class acts as a collection, you can simply use Language Integrated Query (LINQ) to perform the SQL-like queries.

F. User friendly

MASS has a simple and user friendly graphical interface with which you can create new ActualSystems by choosing, connecting and configuring the right plugins (Fig. 2).

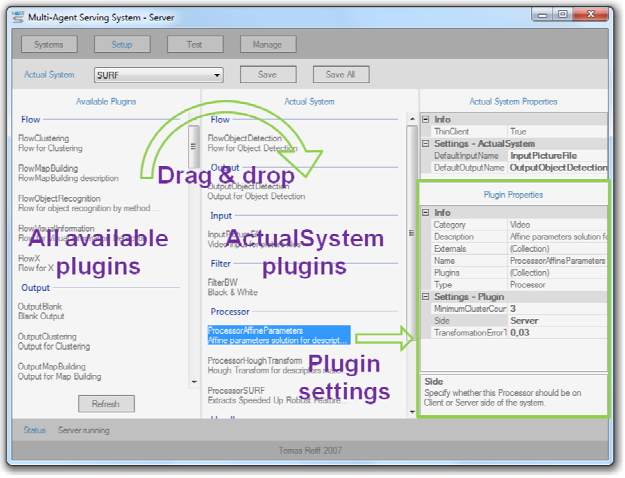


Fig. 2 Creating ActualSystem

Simple graphical interface is also used to control the runtime of ActualSystem (Fig. 3) or to view its output (Fig. 4). Moreover this interface is being generated based on what is specified in Flow and Output plugins of a current ActualSystem. If your ActualSystem has Input and Output plugins which can be used in a web interface, the web forms can be also generated.

A. Free and open source

We think it is very important for a middleware to be free and open source. Therefore anyone can try the web interface of the system, read the wiki pages, download the installation package or find the MASS repository address at http://brain.fei.tuke.sk. There are also links to some of our demonstration videos with MASS.

We would be happy to welcome new people who would like to contribute to the future development of this middleware or the development of plugins.

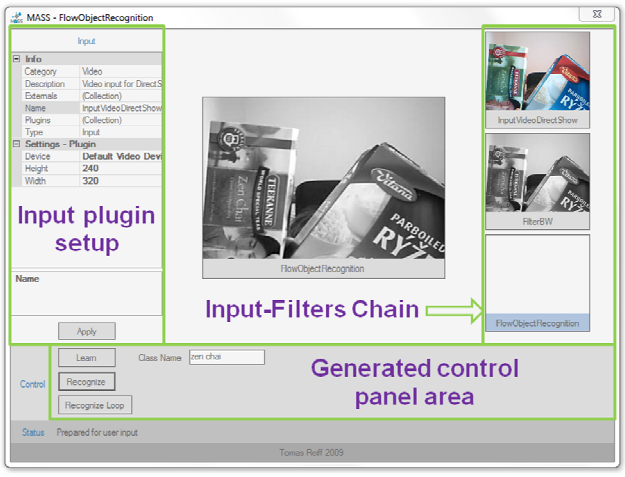


Fig. 3 Controlling ActualSystem runtime by Flow plugin

III. APPLICATIONS

Although there is still not a first release version of MASS, there are already some applications which are using it. We will

now briefly describe the two most important application areas and mention others.

Fig. 4 Output from ActualSystem

A. Object Recognition

The first application and also the guiding application of MASS was the object recognition [1], [10], [12]. Currently MASS installation package contains plugins necessary for extracting local feature descriptors of an object (Scale Invariant Feature Transform - *ProcessorSIFT*, Speeded Up Robust Features - *ProcessorSURF*), matching object descriptors (simple nearest neighbor with distance ratio - *HandlerNearestNeighbor*), clustering matched descriptors (Hough Transform - *ProcessorHoughTransform*) and drawing object parallelograms (solution of affine transform parameters – *ProcessorAffineTransform*).

With this plugins you can learn new object from a webcam (*InputVideoDirectShow*), picture file (*InputPictureFile*) and video file (*InputVideoFile*). Then you can recognize them like on Fig. 4.

B. Command Recognition

Command recognition was the second application done with MASS [2], [3], [13]. Latest publication is using MFCC coefficients together with DTW algorithm while the training set is being reduced by evolutionary algorithm. Currently MASS installation package contains of plugins for MFCC (*FilterHamming*, *FilterFFT*, *FilterABS*, *FilterOverlap*, *FilterMel*, *FilterLog*, *FilterDCT*, *FilterSpeechDetection*, *FilterFIR*), a library with DTW (*MASS.Audio*) and audio input plugins (*InputAudioDirectShow*, *InputSoundFile*).

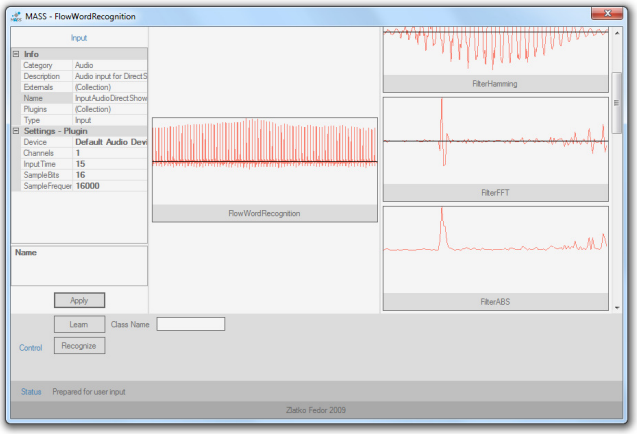


Fig. 5 Flow plugin for command recognition

C. Other

There were also some other applications which were hosted on MASS middleware. We can name few of them together with their references: foreground segmentation [6], recognition of visual information [7] and local image operators for improvement of object recognition [8].

IV. Conclusion

We have introduced a new open source middleware for artificial intelligence and robotics named Multi-Agent Serving System (MASS). Our middleware was developed to by simple, user friendly and based on the latest software technologies. It can be used to develop, test and demonstrate new algorithms on one of the supported robotic platforms.

Acknowledgment

This publication is the result of the project implementation Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] T. Reiff, Incremental system for object recognition, Written work for dissertation exam, Kosice, Slovakia, 2009.
[2] Z. Fedor, Command processing with use of artificial intelligence, Written work for dissertation exam, Kosice, Slovakia, 2009.
[3] Z. Fedor – T. Reiff, Classification of isolated words with Point-Border Artmap, In: SCYR 2009, Kosice, Slovakia, 2009.
[4] Z. Fedor – P. Sincak, AIBO talking procedure in multi-languages based on incremental learning approach, In: SAMI 2009, Herlany, Slovakia, 2009.
[5] T. Reiff – P. Sincak, Towards intelligent systems with incremental learning ability, In: Studies in Computational Intelligence, Vol. 243, p. 627-638, 2009.
[6] K. Chomjak, Detection of moving objects in image sequence, Master thesis, Kosice, Slovakia, 2009.
[7] M. Kica, Incremental system for recognition of visual information, Master thesis, Kosice, Slovakia, 2009.
[8] T. Sabol, Use of local operators for image quality improvement during recognition, Master thesis, Kosice, Slovakia, 2009.
[9] T. Reiff – P. Sincak, Multi-Agent sophisticated system for intelligent technologies, In: AEI 2008, p. 87-91, Athens, Greece, 2008.
[10] T. Reiff – P. Sincak, Multi-Agent sophisticated system for intelligent technologies, In: ICCC 2008, p 37-40, Stara Lesna, Slovaka, 2008.
[11] P. Sincak – T. Reiff, Incremental building of intelligent systems, In: Kybernetika a Informatika, p. 5, 2008.
[12] T. Reiff, Incremental system for image pattern recognition, Master thesis, Kosice, Slovakia, 2008.
[13] Z. Fedor, Incremental system for linguistic command recognition, Master thesis, Kosice, Slovakia, 2008.

Causal Model of Software Evolution using Domain-Specific Languages

Miroslav SABO

Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

miroslav.sabo@tuke.sk

***Abstract*—The paper presents the model of software evolution based on differentiating between two independent parts of the software system - domain-specific language reflecting application environment and implementation of solution reflecting the specification of the problem. The process of evolution is executed accordingly to the type of evolutionary change. The categorization of changes, made upon cause which induced the change, is also discussed in the paper.**

***Keywords*—cause of change, complexity of software system, domain-specific languages, language evolution**

I. Introduction

Laws of the software evolution were written by Lehman in early 1970s [1] but despite long time period, tools for effective solving of the problems identified by these laws have still not been developed. The model of software evolution proposed in this paper is targeted towards first two of the laws - *law of continuing change* and *law of increasing complexity*.

Law of continuing change states that effectivity of the system will be progressively deteriorated until it is continually adapted to changes in the application environment. Many solutions have been proposed to address this law [2], [3], [4], but success always came with the increase of complexity of the system. This negative side effect of the first law is also the main concern of the second Lehman's law. It states that as system evolves, its complexity will continually increase until progressive or anti-regressive effort is invested into maintaining or reducing it. What it means is as changes to the system are successively implemented upon each other, interactions and dependencies between system elements increase in an unstructured pattern and lead to an increase in system entropy. The best results in addressing this issue have been achieved with the generative methods of software development. In this approach changes are applied to the model of the system on higher level of abstraction and final implementation is generated from this model automatically.

Model of software evolution proposed in this paper introduces the differentiation of particular parts of the system representing the application environment and actual solution of the problem. Evolutionary changes can be applied that way directly to the subject it concers without increasing the overal complexity of the system. Categorization of evolutionary changes according to the cause of the change is also defined. Changes in each category are targeted towards specific part of the system, modification of which does not influence other parts therefore overal complexity of the system is well preserved during the whole process of evolution.

II. Inevitability of change

Change is the main characteristic of software evolution as software systems have to react on constantly evolving requirements and underlying platforms and other impulses from environment which they operate in. Changes are inevitable from different reasons:

- **New requirements on system** – requirements on system can change early in the process of software development but this phase may not always be convenient for their implementation from different reasons (e.g. firmly determined deadlines do not allow for unforeseen activities). On the other hand, it is the pressure from satisfied customers which are creating new requirements for functional extensions of the system.
- **Modeling of reality** – as the environment of the system dynamically evolves and changes, system must be continually adapted else it becomes progressively less satisfactory [1]. In extreme cases when systems are interconnected with application environment too tightly, environment is influenced by the system right after the deployment which in turn results in immediate need for adaptation of the system on these changes.
- **Bug fixing** – these requirements arise mainly in the testing phase.
- **Architectural changes** – significant changes in the structure of the system (e.g. system working with business processes evolves and increasing complexity requires integration of the rule engine which will interact with many modules within the system).
- **Enhancing the performance and reliability of the system**

III. Typology of software evolution

In the 1970s, Swanson proposed the typology of software maintenance [5] which was distinguishing between maintenance activities accordingly to the purpose which they were executed for:

1) **perfective** – any enhancements which increase the quality of the system (e.g. adding new features, increasing performance or system documentation).
2) **adaptive** – ensure the usability of the system after changes in environment or technical infrastructure of the system.
3) **corrective** – remove bugs from implementation, usually cover the solving of problems caused by discrepancies between requirements and actual implementation.

Some taxonomies [6] added another category:

4) **preventive** – this last category is often the subject of discussions, considered by some as part of perfective maintenance [7]. IEEE software engineering terminology standard [6] defines preventive maintenance as "maintenance executed with intention to prevent the problems before they even occur".

This typology had been rafined over time and based on the work experience, the classification of 12 types of software evolution and software maintenance [8] was defined later.

TABLE I
TYPOLOGY OF SOFTWARE EVOLUTION (E) AND SOFTWARE MAINTENANCE (M).

Object of change	Type of change	E/M
Business rules	Enhancive	E/M
	Corrective	
	Reductive	
Software properties	Adaptive	E/M
	Performance	
	Preventive	M
	Groomative	
Documentation	Updative	M
	Reformative	
Support interface	Evaluative	M
	Consultive	
	Training	

Complementary view on this topic presents Mens in his work [9] which is focused towards technical aspects of the software change. He proposes the taxonomy of software evolution based on characteristic mechanisms of change and factors which influence these mechanisms.

Even though precise fine-grained typology of software evolution is well documented, the model of software evolution proposed in this paper distinguishes only four fundamental types of evolution - perfective, adaptive, corrective and preventive.

IV. CAUSE-DRIVEN SOFTWARE EVOLUTION MODEL

Model of software evolution proposed in this paper is focused on elimination of the negative side effect of adaptation to the continually evolving environment - increased complexity of the system. The main idea is targeting the application of the changes strictly to those parts of the system implementation which represent the evolved objects in real world. For systems developed in general purpose languages this constitutes a complicated problem because development requires implementation of the concepts of application environment at first and just after that the actual solution may be implemented. Both implementations are tangled together and therefore adaptation of the system to enviromental changes requires identification of parts of the system to be adapted before adaptation can be executed. Even after that, change of the adapted code may be delegated further into system because of tangled code. The result is increased complexity of the structure of system. Proposed model of software evolution separates the implementation of application environment from the implementation of solution for the problem therefore evolution can be targeted directly towards the actual subject of change. Domain-specific languages, as technology following the principles of generative approach to software development, are utilized as tool for representation of application environment of the evolving system.

A. Software evolution in domain-specific languages

For the implementation of the change in software systems developed in general purpose languages (GPLs), its type is not relevant because all changes are applied on the same level - source code of the application. It does not matter whether changes relate directly to the change in specification of the application or are induced by the change in environment thus do not relate to the specification at all.

From the perspective of evolution, development of software systems in domain-specific lanaguages offers some benefits. Changes are always executed on the level which they directly relate to. Domain-specific languages (DSLs) are by definition [10] languages which directly reflect some specific domain. Therefore they can be considered as model of application domain [11]. The implication of this is that any changes arisen in the application domain should be reflected in appropriate DSL which models this domain. Contrary to software systems developed in GPLs, the impact of the implementation of such changes on models/specifications of systems developed in DSL is minimal or none [12]. Considering evolution induced by the change in environment and not by the change of definition of the problem, DSL approach follows these events precisely:

- changes in environment ⇒ changes in DSL and generator/interpreter
- (no) changes in definition of the problem ⇒ (no) changes in model/specification of the application

B. Causes of change

Changes occuring during software evolution can emerge from 3 different sources:

1) application environment
2) definition of the problem
3) specification of the solution

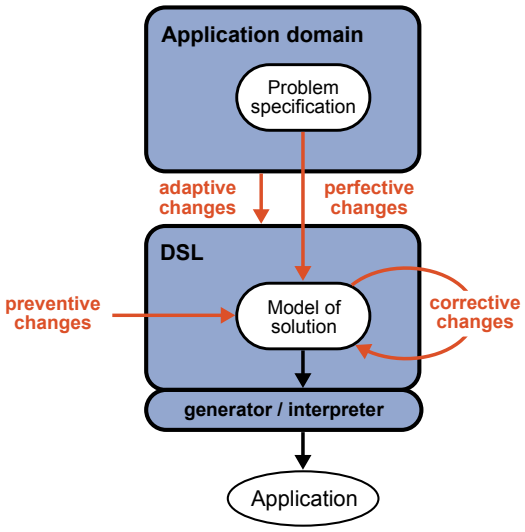


Fig. 1. Causal model of the software evolution.

Evolutionary changes of application environment are reflected in modification of DSL which models the application. The modifications include addition, removal or modification of the language constructs, modification of relations between the

constructs and modification of generator or interpreter. While implementing these changes, the specification of the solution can remain intact and new evolved version of the system will be created using modified generator or changes will be comprised in modified language interpreter [13], [14]. Changes caused by evolution of application environment belong among *adaptive changes*.

Changes in the definition of the problem are either caused by errors which occurred in previous versions (*corrective changes*) or result from new requirements on performance, usability, maintainability or other attributes of the solution (*perfective* and *preventive changes*). These changes are implemented directly in the model/specification of the solution or they can also induce change of the language.

V. Software evolution using causal model

Considering subject of the change implementation defined by the causal model, evolution of software systems can be divided into:

1) evolution of language
2) evolution of solution specification (program or model)

A. Evolution of language

Evolution of the language is closely connected to the manner in which the language has been designed. Domain-specific languages are generally designed in two ways - *internal* and *external*. Considering internal DSLs as higher level abstraction of GPLs, the evolution of such languages gets down to the regular evolution of code written in general purpose language.

On the other hand, external DSLs are usually designed using metamodeling languages in language workbenches specialized for language development [15]. After the model of language is created, the complete development environment for the new language, including tools such as editors, browsers, generators and interpreters, is generated automatically from the model of language. The evolution of such DSLs therefore includes only modifying the model of the language [10].

Similar approach to the evolution of external textual DSLs is provided by the language development tool YAJCo [16], [17] which is based on definiton of abstract syntax. Model of the language consists of the classes representing abstract syntax. Concrete syntax is defined upon abstract syntax classes through annotations. The change of the language, in the same manner as language workbenches, requires only modifications on the abstract and concrete syntax level and new generator for the language is created automatically.

B. Evolution of solution specification

Evolution of the solution specification, which is either in the form of program or model, is similar to the evolution of the program written in GPL. Use of DSL, however, brings some advantages specific for this approach such as specification directly in the concepts of the domain, domain-specific control checking and domain-specific optimalization.

The biggest advantage of using domain-specific languages, however, is that changes targeting solution specification can be implemented in a straightforward manner because implementation language (DSL) in which the changes will be applied is on the same level of abstraction as language in which the requirements are specified (language of domain experts). All in all, evolution of the software systems developed in domain-spedific languages is simple, easy to execute and less prone to errors.

VI. Conclusion

In this paper I have presented the causal model of software evolution. This model satisfies both of the first two Lehman's laws of software evolution - law of continuing change and law of increasing complexity. Preservation of complexity of software system during the process of evolution, as the main problem identified by these laws, is achieved by application of the evolutionary changes strictly to those elements of the system which represent the subject of change in real world. The part of the system to reflect the evolutionary change is determined with respect to the type of change. The categorization of types of changes, based on the cause which induced the evolutionary change, is also presented in the paper.

Acknowledgment

This work is the result of the project implementation: Centre of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] M. M. Lehman, "Laws of Software Evolution Revisited," in *EWSPT '96: Proceedings of the 5th European Workshop on Software Process Technology*. London, UK: Springer-Verlag, 1996, pp. 108–124.

[2] O. Greevy, S. Ducasse, and T. Gîrba, "Analyzing software evolution through feature views: Research Articles," *Journal of Software Maintenance and Evolution: Research and Practise*, vol. 18, no. 6, pp. 425–456, 2006.

[3] W. Cazzola, S. Pini, and M. Ancona, "Evolving Pointcut Definition to Get Software Evolution," in *ECOOP '04: Proceedings of the Workshop on Reflection, AOP and Meta-Data for Software Evolution*, 2004, pp. 83–88.

[4] S. P. Reiss, "Constraining software evolution," in *ICSM '02: Proceedings of the International Conference on Software Maintenance (ICSM'02)*. Washington, DC, USA: IEEE Computer Society, 2002, p. 162.

[5] B. P. Lientz and E. B. Swanson, *Software Maintenance Management: A Study of the Maintenance of Computer Application Software in 487 Data Processing Organizations*. Addison-Wesley Pub, 1980.

[6] I. O. Electrical and E. E. (ieee), *IEEE 90: IEEE Standard Glossary of Software Engineering Terminology*, 1990.

[7] N. Chapin, "Do We Know What Preventive Maintenance Is?" in *ICSM '00: Proceedings of the International Conference on Software Maintenance (ICSM'00)*. Washington, DC, USA: IEEE Computer Society, 2000, p. 15.

[8] N. Chapin, J. E. Hale, K. M. Kham, J. F. Ramil, and W.-G. Tan, "Types of software evolution and software maintenance," *Journal of Software Maintenance*, vol. 13, no. 1, pp. 3–30, 2001.

[9] T. Mens, J. Buckley, A. Rashid, and M. Zenger, "Towards a taxonomy of software evolution," in *ECOOP '02: Proceedings of the Workshop on Unanticipated Software Evolution*, 2002.

[10] J. Sprinkle, J. Gray, and M. Mernik, "Fundamental Limitations in Domain-Specific Language Evolution," *IEEE Transactions on Software Engineering*, vol. 35, no. 3, 2009.

[11] K. Czarnecki, "Overview of Generative Software Development," in *UPP*, 2004, pp. 326–341.

[12] J. Kollár, J. Porubän, P. Václavík, M. Forgáč, and J. Bandáková, "How to Adapt Programming Languages instead of Software Systems," *Computer Science and Technology Research Survey*, vol. 2, pp. 69–79, 2007.

[13] J. Kollár, J. Porubän, P. Václavík, J. Bandáková, and M. Forgáč, "Adaptive Compiler Infrastructure," in *Komunikačné a informačné technológie*, Tatranské Zruby, 2007, pp. 4–5.

[14] J. Kollár and J. Porubän, "Building Adaptive Language Systems," *INFOCOMP - Journal of Computer Science*, vol. 7, pp. 1–10, 2008.

[15] MetaCase, "MetaEdit+," http://www.metacase.com, 2009.

[16] M. Mernik and J. Porubän, "Language Design with Concrete/Abstract Syntax: LISA vs. YAJCo Compiler Generators Approaches," in *Informatics'09: Proceedings of the 10th International Conference on Informatics*, vol. 10. Koice: Elfa, 2009.

[17] J. Porubän, M. Forgáč, and M. Sabo, "Annotation Based Parser Generator," in *WAPL '09: Proceedings of the International Multiconference on Computer Science and Information Technology*, vol. 4, Mragowo, Poland, 2009, pp. 707–714.

Identifying Boundaries between API and DSL Approach to Language Development

Miroslav SABO

Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

miroslav.sabo@tuke.sk

***Abstract*—The paper presents different views on the notion of Domain-Specific Language. Since a lot of controversy exists around DSLs, various approaches on understanding the principles applied by DSLs are provided. Key characteristics of DSLs are analyzed and adjusting of the definition of DSL is proposed. The controversy concerning whether internal DSLs qualify for labeling a Domain-Specific Language is widely discussed and some propositions on this topic are concluded.**

***Keywords*—application programming interface, domain-specific language, fluent interface, language extensions**

I. Introduction

In software development, using the language of the domain of the problem which you intend to address, is becoming more popular and also considered as one excellent technique to producing software of high quality. The common name used for this kind of languages is *domain-specific languages* or shortly *DSLs*. Recently, however, the term DSL seems to have become overused, causing it to become vague. By understanding of many people omitting parentheses from function calls or using a block and including one or more well-named methods already qualifies a piece of code to be labeled a DSL. Since no precise definition of domain-specific lanuage has been widely accepted yet this happens quite often. This paper gives some insight on this problem and tries to provide the convenient explanation of what can what can not be classified as a domain-specific language.

II. Definition of DSL

The question what exactly is a domain-specific language is subject to debate but the idea of DSL is definitely not new.

The first time the concept of DSL has been used was in LISP which is a 40 years old programming language. Paul Graham states that the best way to write applications in LISP, is to extend LISP environment into a language to write that application [1]. For example, if you want to write a word processor, extend LISP to be a word processor language. According to Graham, a domain-specific language can be defined somehow like this:

Definition 1. *A domain-specific language is the way of treating the code as a language (where appropriate) by writing lower level code that eventually enables to write code and think in a higher level domain.*

This definition addresses the languages with syntax that can be twisted or bent to allow for more concise code [2], however, it does not rule out other approaches on how to achieve domain-specific code.

Arie van Deursen provides summary on an exhaustive survey on the literature available on the topic of domain-specific languages as used for the construction and maintenance of software systems [3] and proposes a definition for domain-specific language as follows:

Definition 2. *A domain-specific language is a programming language or executable specification language that offers, through appropriate notations and abstractions, expressive power focused on, and usually restricted to, a particular problem domain.*

The key characteristic of DSL according to this definition is its focussed expressive power, although one of the defining term - *problem domain* is rather vague and defined simply by the listing of domains.

Martin Fowler makes an important and useful distinction between *internal* and *external* DSLs. Internal DSLs are particular ways of using a host language to give the host language the feel of a particular language, while external DSLs have their own custom syntax and a full parser has to be written to process them. Considering that majority of the most common DSLs today are textual, he distinguishes also a third kind - *graphical* DSLs which require a tool along the lines of a language workbench [4]. The general definition of domain-specific language by his comprehension is:

Definition 3. *A domain-specific language is a computer language that's targeted to a particular kind of problem, rather than a general purpose language that's aimed at any kind of software problem.*

There are many more definitions of a domain-specific language out there but these are the ones which are being most referred to. They are all aiming for different aspects of the gist of a domain-specific language but some common characteristics stated by all definitions can be identified.

III. What makes a DSL a DSL?

A. Key characteristics

1) Focussed expressive power: Instead of aiming to be the best for solving any kind of computing problem, DSLs aim to be particularly good for solving a specific class of problems. A distinguishing feature of DSLs is their ability to precisely and concisely express the logic through code in a way that

somebody else reading the code can easily figure out what the original intention of author of the code was.

2) Narrow scope of applicability: A DSL cannot be too generic. The more it is generified, the less of a DSL it becomes.

B. Other characteristics

1) Small in size: DSLs usually offer a restricted suite of notations and abstractions as a result of constraining by the scope of the domain. In literature they are also called *micro-languages* and *little languages* [5].

2) Declarative in nature: As well as programming languages, DSLs can be viewed as specification or configuration languages. This is not a defining characteristic of DSLs but the majority of them appear usually in a declarative context.

3) Fluent readability: The concept of a DSL is generally used to describe a language that has a vocabulary suited for expressing concepts and/or syntax appropriate to a particular body of knowledge. The popular argument for creating DSLs is that they allow to write code that is easy to read and understand and does not hide its meaning behind language constructs such as loops or conditionals. The notation of DSLs is designed for use by a domain expert, as free of "code-like" syntax as possible. This characteristic is sometimes considered as a consequence of the high expressiveness of the language.

IV. Controversy over DSL

The question whether something qualifies for a domain-specific language or is just an application programming interface for the language has been around since the very beginning of DSLs due to absence of generally accepted definition. The greatest controversies stem mainly from ambiguity of two terms in the definition - *domain* and *language*.

A. Domain specificity

The main problem concerning the term *Domain* is that no definition prescribes what might and what might not be considered a domain. The constroversy over DSLs has arisen lately due to statements that *Ruby on Rails* is a domain-specific language for Web Application domain because it expresses a web applications behavior in web application native terms. First of all, Rails is not a programming language but web framework, however, it is not the main point of the statement to argue about. What is more doubtful is the area of specialization of Rails. The main objection of the opponents is that Web or Web Application is an area too broad and generic to qualify as a domain. This argument is based on valid reasoning because if Rails is awarded the DSL for the Web attribute, C might be as well called a Systems DSL or Assembler a Computer's DSL.

Following this thought, I would suggest to broaden the definition of a domain-specific language with a constraint for the domain:

Definition 4. *Along with the DSL being specific to the domain, the domain itself needs to be specific.*

B. The L in DSL

The single word, *Language*, creates the major discussion in the world of DSLs. Essentially, opinions on DSLs might be divided into two groups, depending on whether you interpret the term language strictly by the definition of a formal language [6] or loosely just as means of communication.

Definition 5. *A formal language is a (usually infinite) set of finite-length sequences of symbols defined (or generated) by a formal grammar.*

Martin Fowler wrote some thoughts of this subject [4] and he divided DSLs into two types: *external* and *internal*. External DSLs examples are SQL, Hibernate's HQL or Ant scripts. Internal DSLs would be certain aspects of Rails, or Javascript's Event.Behavior.

1) External DSLs: External DSLs are glaring example of DSLs. They are accepted as domain-specific languages even by the community understanding the term language accordingly to the strict Chomsky's definition and there is no discussion about them whatsoever.

They have their own custom syntax and full parser has to be written to process them. There is a very strong tradition of doing this in the Unix community [5]. Many XML configurations have ended up as external DSLs, although XML's syntax is badly suited to this purpose.

DSLs can be implemented either by interpretation or code generation. Interpretation (reading in the DSL script and executing it at run time) is usually easiest, but code-generation is sometimes essential. Usually the generated code is itself a high level language, such as Java or C#.

Domain concepts and relations between them are commonly designed by abstract grammar which can be implemented in many ways to provide the concrete syntax for DSL [7]. To name a few, concrete syntax implementations include own custom syntax, XML, diagrams to express statements, content management system, formatted spreadsheets to describe the intent or records read in from database tables. Having all of these as possible options, it is much more likely to use the right tool for the job depending on use case.

2) Internal DSLs: Internal DSLs are domain-specific languages using the syntax of an existing programming language, a host language. This approach has recently been popularized by the Ruby community although it's had a long heritage in other languages, in particular Lisp [1]. Although it's usually easier in low-ceremony languages like Lisp, effective internal DSLs can be created also in more mainstream languages like Java and C#.

They are very much constrained by the used host language. Since any expression must be a legal expression in used host language, a lot of thought in internal DSL usage is bound up in language features. A good bit of the recent impetus behind internal DSLs comes from the Ruby community, whose language has many features which encourage DSLs. However many Ruby techniques can be used in other languages too, although usually not as elegantly.

There are two approaches on how to implement internal DSLs, albeit the second quite controversial - via *language extensions* (embedded DSLs) or *APIs* (fluent interfaces). The

easiest way is to create a set of functions (or more likely these days classes with methods) where the method signatures define an API. The other way of implementing internal DSLs is via language extensions.

The question whether an API implementation should be considered as domain-specific language brings the major controversy over DSLs nowadays. What is it that distinguishes a DSL from a (really well designed) API? Is it method chaining? Is it the fact that API/DSL is readable by a domain expert? What does DSL enable you to do that API does not? So, is the term DSL in this case bad or misleading? How can a library written for a general purpose language be a language too? Isn't that just a well written API?

V. Distinguishing DSL and API

The impulse to create own little consistent API for expressing something when developing an application for specific domain is usually considered as good programming practice. Naming it a DSL, however, may somebody call an exaggeration. The preferred names for such APIs are *domain-specific dialects*, *domain-specific jargon*, or even *domain-specific slang*. The emphasis is put on "domain-specific" collocation and what it is actually referring to is *the code acting like a language* (suited for expressing concepts of a particular domain).

The same apply for language extensions. For example, the problem with calling the code written in Ruby a language is that it is still using the grammar of Ruby, just more expressive human interface has been written for the grammar.

The way of facilitating the use of domain-specific idioms may be classified as *using domain-specific language*. However, no new langauge has been constructed actually, therefore creating an API or adjusting the host language in any way can not be classified as *creation of a DSL*. The main idea to conceive is that when arguing about creation of a DSL, often more precisely might be saying using of DSL. Most of the time, no language has been created but use of a particular aspect of one has been made.

```
Engine e =
  new Engine(8, 3996, Fuel.DIESEL);
Transmission t =
  new Transmission(Transmission.MANUAL, 6);
Interior i =
  new Interior(AC.FULL, Audio.UNKNOWN,
  Seats.RACING, null);
return new Car(Model.Convertible, e, t, i);
```

```
car()
  .convertible()
  .engine()
    .cylinders(8)
    .capacity(3996)
    .diesel()
  .transmission()
    .manual()
    .gears(6)
  .interior()
    .airCondition()
    .racingSeats()
  .build();
```

Fig. 1. Comparison of the definition of an object in traditional approach and with fluent interface (Java programming language).

The synonym for internal DSLs looked at from the API direction if *fluent interfaces*. This term was coined by Eric Evans and Martin Fowler to describe more language-like APIs [4]. It tackles directly the difference between a DSL and an API - the language nature.

A. Fluent interfaces

Fluent interface represents a certain style of interface with the intent to do something along the lines of a DSL which it implements. This is why Evans and Fowler chose the term "fluent" to describe it.

The primary focus of API design is readability and flow. The key test of fluency is the domain-specific language quality. The more the use of the API has that language like flow, the more fluent it is. So far, fluent APIs have been mostly seen in declarative context (e.g. to create configurations of objects, often involving value objects).

VI. Spectrum of "DSL-ness"

Certainly there is a spectrum of conformance with DSL definition, accordingly to the expressiveness, for different approaches to DSL implementation. API and fluent interfaces fall at one end, closer to the GPL, because their expressiveness is considerably limited by the syntax of the host general-purpose language. Embedded DSLs (language extensions) are a little bit more expressive because host language offers some features to make syntax more natural. The real DSLs are represented by external DSLs which have no limitations.

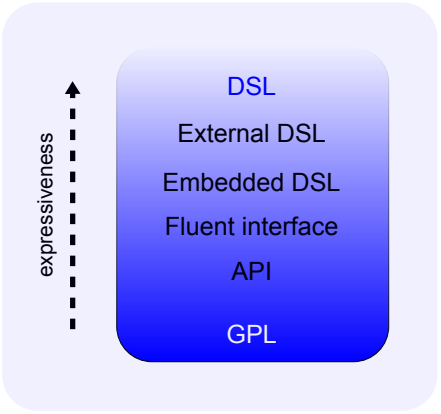


Fig. 2. Spectrum of DSLs accordingly to the level of focussed expressive power of the language.

VII. Conclusion

This paper presented different views on the topic of domain-specific languages. Since no widely accepted definition for a DSL exist, multiple definitions were provided, abstracting the key characteristics of a DSL from them in the following section.

The focus of the paper was on the resolution of controversy about whether API, which provides some domain logic, qualify for labeling a DSL or where is the imaginary boundary between an API and a DSL. According to the thoughts presented in the paper, it is obvious that problem stems from *ambiguities in terminology*.

Firstly, the term Domain does not have precise definition which causes disagreement on what qualifies and what does not for a domain. The extension to definition of a DSL concerning this issue has been proposed (*Definition 4.*).

Secondly, the term Language may be interpreted in many ways. The most strict view is understanding a DSL as a formal language. Chomsky's definition of formal language dictates that language is defined by formal grammar. By this definition, external DSLs are considered as the only languages qualifying for label DSLs. This view is quite rigid and I incline more towards the looser view on DSLs focusing on domain-specific expressiveness of the language rather than the language itself. However, this view brings the problem mentioned above:

Does API (or language extension) qualify for a DSL?

My answer is No, unless the following conditions are met:

1) *Domain-specific validation.* DSL must be able to provide error messages on the same level of abstraction as the solution - program or model specified in DSL.
2) *Domain-specific optimization.* Having domain knowledge captured in a DSL, it is possible to optimize the final implementation of the solution accordingly to this knowledge.

Meeting the second condition may be the subject of further research but I contend that language must necessarily provide domain-specific validation to qualify for a DSL.

ACKNOWLEDGMENT

This work is the result of the project implementation: Centre of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF.

REFERENCES

[1] P. Graham, *On LISP: Advanced Techniques for Common LISP*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.
[2] D. Flanagan and Y. Matsumoto, *The ruby programming language*. O'Reilly, 2008.
[3] A. van Deursen, P. Klint, and J. Visser, "Domain-specific languages: an annotated bibliography," *SIGPLAN Not.*, vol. 35, no. 6, pp. 26–36, 2000.
[4] M. Fowler, "Domain Specific Languages," http://martinfowler.com/dslwip, 2009.
[5] J. Bentley, "Programming pearls: little languages," *Communications of the ACM*, vol. 29, no. 8, pp. 711–721, 1986.
[6] N. Chomsky, *Syntactic structures*, ser. Ianua linguarum : Series minor ; 4. Mouton, 1962.
[7] M. Mernik and J. Porubän, "Language Design with Concrete/Abstract Syntax: LISA vs. YAJCo Compiler Generators Approaches," in *Informatics'09: Proceedings of the 10th International Conference on Informatics*, vol. 10. Koice: Elfa, 2009.

Using Template Matching Method to Compare the ECG Waves and Visualization of ECG Similarities

Peter SMOLÁR, Zlatko FEDOR, Juraj EPERJEŠI

Center for Intelligent Technologies/Department of Cybernetics and Artificial Intelligence, FEI TUKE Košice, Slovak Republic

{peter.smolar, zlatko.fedor, juraj.eperjesi}@tuke.sk

***Abstract* —This paper is concerned with processing and an analysis of ECG samples with using method Template matching. Mainly it deals with recognition of ECG samples of diagnoses of myocardial infarct and arrhythmia. Template matching method, which is used here, can find the best similarity between the test sample and ECG templates. According to the metrics it calculates their relative similarity too. For a visualization of relative similarities is used Kohonen network as processing. Input data were obtained from the project PhysioNet, gathered at the Institute of Cardiology at the University Clinic Benjamin Franklin in Berlin and digitalized in the National Metrology Institute, Germany under the name PTB ECG database. The outputs are the similarity coefficients of the twelve conventional ECG leads and the six basic parameters of waves. The results of our proposal with used methods for data preprocessing and implemented algorithm are comparable with the results obtained by systems based on neural networks classification. It has the potential to help physicians in the initial analysis and identification of the patient's condition.**

***Keywords*—template matching, diagnostic of heart disease, ECG**

I. INTRODUCTION

In compliance with the European Heart Health Charter [1] cardiovascular diseases represent the most frequent cause of death of women and men in Europe. They are annually responsible for more than 1.9 million deaths in the European Union. Cardiovascular diseases are also a main cause of disability and reduced quality of life. Speed and accuracy in determining the correct diagnosis often decide about life or death of the patient. Therefore, it is the area of continuous research.

As noted in [8], statistical methods are used in artificial intelligence and are capable to acquire new knowledge from test ECG samples and then accomplish the awareness for a doctor as decision support system. On the basis of comparison of two samples to determine their similarity and then to declare, that the sample belongs to the group of samples or not.

Using this method for the evaluation of ECG waves we can facilitate and accelerate the physician`s work. Important in this method is samples preprocessing. The main role has a computation of relative similarity of the ECG waves with diagnosis templates.

The main goal of this work is not to classify the samples with zero error, because this issue is very broad and it exceeded the scope of this paper. The intention is to try to facilitate and accelerate of the physician`s work and give him a important time to saving lives or reduce the consequences of the disease.

One of the simplest and the first approaches to pattern recognition is based on a Template matching [5] (comparison on the basis of the test samples and templates). Matching is the generic algorithm in pattern recognition, which is used to determine the similarity between two entities (points, curves, other services) of specified type. The template (or we can all it etalon) is the most important element of recognition in this method.

The test sample, which is an effort to recognize indications of diseases, is compared with template. The comparison is making with respect to the metrics and is calculated the similarity. It is necessary to make the normalizing changes of sample, in order to achieve the best similarity. In this paper is also described process of visualization of ECG similarities.

II. DESIGN OF SYSTEM FOR CLASSIFICATION OF ECG DATA USING THE TEMPLATE MATCHING METHOD

Our goal from as is mentioned in [8], is to create an application that would implement processing and analyze of ECG samples on the specific data. We try to classify the ECG sample to healthy class of data or to any class of selected cardiology diagnosis. Applications should be able to display the results numerically and graphically. Below is the block diagram of program (Fig. 1).

After loading the test ECG sample and a template of the fixed length from the database samples, it is calculated pulse of both data objects. According the size of pulses, the test sample or template is sent to horizontal scaling depending on, where smaller pulse was found. Test ECG wave is normalized with template, in order to normalize value of ECG pulse. Followed by a phase of vertical scaling and then it is finding the best similarity ECG template to test ECG sample using Template matching algorithm. Using the differential function is calculated relative mutual similarity. The outputs are the similarity coefficients of the twelve conventional ECG leads and the six basic parameters of waves.

The algorithm of ECG waves comparison is described here. To compare are used two vectors. One of them is a sample, which we find the most similar image of template and the template, which is equally large or smaller than the sample. We search on ECG sample the most similar image of template. The calculation of similarity is made on the basis of the specific metrics method.

The mathematics formula of calculation doesn`t depend on the used metrics, it depends on the size and dimension of the ECG sample and template.

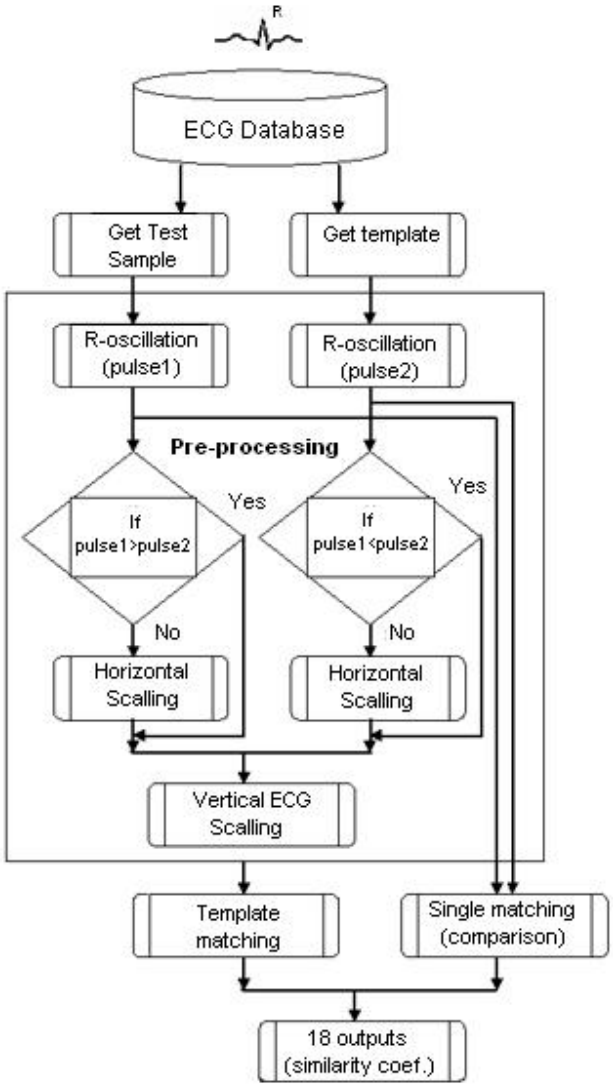


Fig. 1. Complete block diagram of our ECG comparative system, called CardioSys

During the comparison of samples are made different shifts on the ECG sample. The scaling adjustments are performed in order to find the most similar image on the sample.

The following comparison function is used for the comparison of one-dimension ECG curves:

$$r_i = \frac{\sum_{n=1}^{N} \left| (x_{n,i} - xref_{n,i}) \right|}{N} \quad (1)$$

$$r = \min_i (r_i) \quad (2)$$

$$i = 1, \ldots, k \quad (3)$$

Parameters:

$x_{n,i}$ - vector`s value of the sample at position n,

$xref_{n,i}$ - vector`s value of the template at position n,

n - position of value in the vector,

N - size of template vector,

r_i - similarity value in attached template to the ECG sample,

k - count of all different allowable positions of template to ECG sample,

i - identifier of the position.

III. Proposal of Visualization

In output data of the Template Matching method, we get twelve coefficients of similarity from twelve ECG leads and other six coefficients of similarity from base parameters of ECG wave. It means that in the end we have eighteen-dimensional space, which we want to visualize. As human receives maximally three dimensions (we not count a time dimension), we reduce dimensional space in our case to two dimensions. For this purpose we decided to use Kohonen neural network, which is unsupervised and it is possible to reduce feature space. About Kohonen network you see [2]. Kohonen network will have eighteen input neurons in input layer and two output neurons in Kohonen layer. It means that in the output we get x and y co-ordinates of one comparing point. See a Fig. 2.

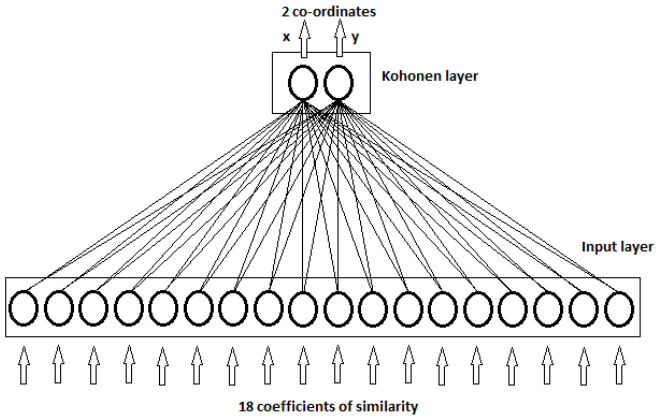


Fig. 2. Proposed Kohonen network for visualization of similarity coefficients

IV. The Group of Experiments 1 – Comparison of Test Samples to Templates According to Diagnosis

The goal is to determine, how the our CardioSys program classify (with a probability in percentage terms) diseased and

healthy ECG waves [8]. To detect, which lead is important for the diagnose (where is seen a pathological abnormality).

For first part of the experiment test ECG samples were from one type of diagnose (eg. arrhythmia) and for second part they were test ECG samples of healthy persons. Templates, which were compared to these samples, they were of two types: the healthy ECG waves and the type of diagnosis, which are test samples. In the other experiments test samples with other diagnoses were changed (and these are also replaced in the templates database).

To each test ECG sample were compared all templates in the database templates. The results were statistically evaluated according to the results of the comparison, which are eighteen. Coefficients of similarity of ECG-leads were twelve and another six values were the main coefficients of the ECG waves.

We calculated two types of statistical evaluations of final results. One type of summary computations was called types Min and second Avg.

The most similar template (Min): We find the most similar template from the set of all templates and according with winning template. The test sample is classified into the same category of diagnosis.

The most similar class of templates (Avg): After making the comparison is made the arithmetic average for each diagnosis and according with the best results of these averages is test sample classified to this diagnose.

Number of all the ECG samples, which are involved in the test process, was 337. The total number of matching combinations was 99 669. Many of the coefficients of similarity correctly classified to ECG waves with a probability of 100% or closely to this value. However, only in one direction. This means, if the parameter correctly classified test samples with some diagnose, he often incorrectly classified samples from healthy persons in the same database templates and reversely. We found out, that the classification of the voting form is not appropriate in determining the correct diagnose. It is preferable to use only similarity coefficients of those parameters, which in this case they know to classify sample the most exactly. Healthy ECG wave by using a statistical method Min classifies lead V6, and the method Avg classifies lead II, aVF and V5 as the best result.

In the experiment results described in [3], the neural network is able to classify acute myocardial infarction of 94.5% accuracy. Experiments in this work demonstrated, that the statistical method Min using a parameter for the classification of "Min. value (V), capable of correctly classified 100% of the samples with the diagnosis of myocardial infarction. In the statistical method Avg was classification result 98.28% using lead "v1". It should be noted, that both experiments use different classification methods, different sets of input test data and evaluation results.

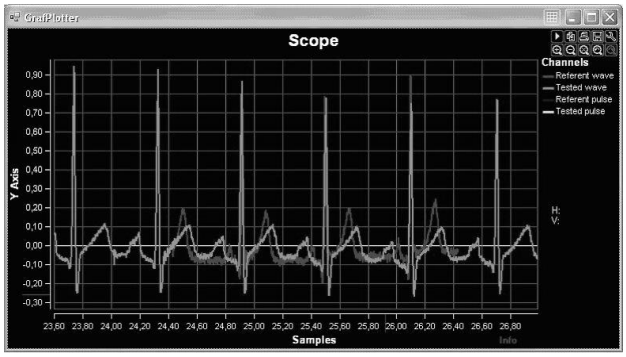


Fig 3. Example of comparison with the ECG sample wave and template

In the experiment results described in [4] neural network correctly identified 90.2% of patients with myocardial infarction of front wall and 93.3% of patients without myocardial infarction of front wall. Template matching algorithm can successfully identify 88.24% of patients with myocardial of front wall using method MIN and ECG lead "v3", 84.38% of patients without infarction of front wall. In the method AVG was classification accuracy 92.65% of patients with myocardial of front wall, using the ECG lead "i" and 90.63% of patients without a diagnosis of myocardial infarction. It should be noted that both experiments use different classification methods, different sets of input test data and evaluation results.

V. Experiment 2 – Visualisation of Output Similarities Using Kohonen Network

The basic intention of this experiment is try to visualize the results of the comparisons (i.e. similarity coefficients of twelve leads and another six base parameters of wave) and try to obtain new knowledge from this images about diagnosis after analyze.

In order to results would be able graphically visualize, we need to reduce a dimension of feature-based space from eighteen to two. For this purpose Kohonen network is proposed, which is described in section III.

First, the Kohonen network must learn across all data of comparisons of one test diagnosis. After learning is step of testing, where data are giving on the input and on the output we get x and y co-ordinates of one sample. Then these data are drawn as points.

A. Tests and experimental results

After several experiments with setting of Kohonen network is found out that the best configuration of parameters is:

- Cycles: 8
- Gamma coefficient: 0.9
- Neighborhood coefficient: 2
- Adaptive height: 0.8

We want visualize comparing infarction ECG samples with healthy ECG samples and conversely.

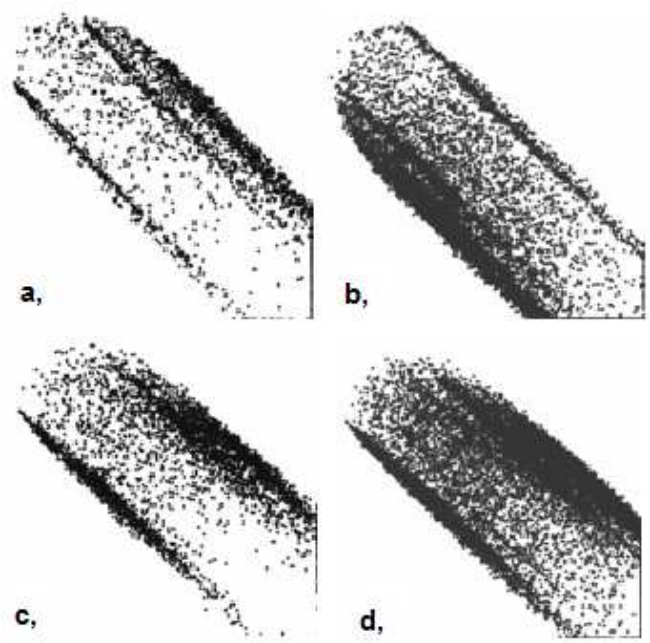


Fig 4. Images of graphically displayed similarity coefficients after reducing of dimension of featured space by Kohonen network from eighteen to two for myocardial infarction diagnose. Section A visualize comparing healthy ECG samples with healthy ECG templates, Section B visualize comparing healthy ECG samples with infarction ECG templates, section C visualize comparing infarction ECG samples with healthy ECG templates, section D visualize comparing infarction ECG samples with infarction ECG templates. Scale of images is 255 x 255 pixels.

B. Evaluation of experiment 2

Result are final images of graphically displayed similarity coefficients after reducing of dimension of featured space by Kohonen network from eighteen to two for myocardial infarct diagnose and healthy ECG samples. In Fig. 4 Section A visualize comparing healthy ECG samples with healthy ECG templates, Section B visualize comparing healthy ECG samples with infarction ECG templates, section C visualize comparing infarction ECG samples with healthy ECG templates, section D visualize comparing infarction ECG samples with infarction ECG templates. Coordinate of the image is 255 x 255.

From analyze of these images is deduced some conclusions. On the images can be found two types of models: clusters, where is a number of points very high and sparsely labeled areas, where the number of points is small per unit of area.

From number of clusters we can find out number of various types of one diagnose in templates or test waves.

If the unknown test ECG sample is in a cluster, where other samples belong with known diagnose, we can classify this unknown sample to same diagnose as known samples in this cluster.

Comparisons, where are situated outside of centers of large clusters, are not ordinary and matching this samples has uncommon results. It may means, that this samples or templates have a specific type of diagnose.

In a comparing of healthy ECG samples with templates of healthy ECG waves, we can deduce, that this images have similar pattern. However, in analysis of section D of Fig. 4, where are compared test samples with various diagnose, we can see, that each diagnose has its own pattern on the image, which is essentially different. This pattern can be characterized as a stamp of this diagnose (with this samples and templates), what is helpful in classifying of test ECG samples.

VI. Conclusion

The primary purpose of this paper was to implement a method of Template matching for the diagnosis and analysis of ECG samples, to perform basic experiments and to compare results with other scientific work in this area. As a ECG samples and the ECG templates were used data from the PTB Diagnostic ECG database. Specific diagnoses were selected: five types of myocardial infarction, arrhythmia, and samples from healthy persons. The experiments show that this method can classify the data, but the accuracy of the classification of each diagnose is different. There is a better to use for the final classification of the samples specific parameters of ECG curves (each ECG lead, the basic parameters of waves). Type of a selected parameter depends on the type of diagnosis. It was also confirmed, that important is the preprocessing of data (the correct pulse calculating, vertical and horizontal scaling of ECG samples, determining the good length of the sample).

Acknowledgment

This publication is the result of the project implementation Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] European Society of Cardiology, European Heart Network: European Heart Health Charter, [online], Available at the internet: <http://www.heartcharter.eu/download/Slovak.pdf>.

[2] P. Sinčák, G. Andrejková, G: *Neurónové siete, inžiniersky prístup (1. part)*. Kosice : Elfa, 1996. 107 s. ISBN 80-88786-42-8. Available at the internet: <http://neuron-ai.tuke.sk/cig/source /publications/books/NS1/html/>.

[3] V. G. Baxt, S. Shofer, F. D. Sites, J. E. Hollander: *A neural computational aid to the diagnosis of acute myocardial infarction*. Annals of Emergency Medicine [online], 2002, vol. 39, issue. 3, Available at the internet: <http://www.annemergmed.com/article/ PIIS0196064402725456/fulltext>.

[4] N. Ouyang, M. Ikeda, K. Yamauchi: *Use of an artificial neural network to analyse an ECG with QS complex in V1-2 leads*. Medical and Biological Engineering and Computing, 2002, vol. 35, num. 5, Available at the internet: <http://www. springerlink.com/content/j872u1gl11510438>.

[5] A. K. Jain, R. P. W. Duin, J. Mao: *Statistical pattern recognition*, A review, In IEEE Transactions on pattern analysis and machine intelligence, 2000. vol. 22, no. 1, [online] Available at the internet: <http://www.mts.jhu.edu/~priebe/COURSES/FALL2003/550.730/jdm0 0.pdf>.

[6] L. Cole, D. Austin, *Visual Object Recognition using Template matching*, [online], Available at the internet: <http://www.araa.asn.au/acra/acra2004/papers/cole.pdf>.

[7] R. M. Dufour,E. L. Miller, N. P. Galatsanos: *Template matching Based Object Recognition with unknown geometric Parameters*, In IEEE Transactions on image processing, 2002, vol. 11, no. 12, [online], Available at the internet: <http://ieeexplore.ieee.org/ stamp/stamp.jsp?arnumber=01176927>.

[8] P. Smolar, P. Sincak, R. Jaksa, *Application of AI in Cardiology*, In: SAMI 2010 : 8th International Symposium on Applied Machine Intelligence and Informatics : January 28-30, 2010, Herľany, Slovakia. - [s.l.] : IEEE, 2010. - ISBN 978-1-4244-6423-4. - S. 267-270.

Building efficient stochastic models of Slovak language for LVCSR

[1]*Ján STAŠ*, [2]*Daniel HLÁDEK*

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

[1]jan.stas@tuke.sk, [2]dhladek@gmail.com

***Abstract*—Language model is the one of the fundamental components in automatic speech recognition (ASR) system. For continuous speech recognition, building language models trained on huge text corpora with large vocabularies is needed. These systems are usually called large vocabulary continuous speech recognition (LVCSR) systems. The aim of this article is to present main model of building process for modeling the Slovak language. In this article the operations of data mining, text normalization, language models generation and additional steps for correct and efficient language modeling will be described. Experimental results for Slovak language models trained on news text corpus, retrieved from web pages will be evaluated and some main problems about modeling Slovak language in the end of this article will be discussed.**

***Keywords*—speech recognition, language model, *n*-grams, spelling check, text normalization, vocabulary**

I. INTRODUCTION

Automatic speech recognition (ASR) can be defined as independent, computer controlled transcription system of spoken language into readable text that works in real time. The aim of ASR system is to convert the speech signal correctly and efficiently in real time into the text form, independently of speaker, vocabulary, noise of surrounding environment and characteristics of microphone. The system of ASR, especially for LVCSR, is the complex area and consists of a number of components. One of these components is also building a language model.

Language model determines the probability of a sequence of words, as well as the word itself, this consequently helps to find the most probably sequence of words for ASR system, which corresponds to the acoustic information pronounced by the user. For ASR systems the stochastic language models (SLM) the most frequently are used. Those mainly consider the statistical dependency between individual words.

In general, the main aim of the SLM is to determine a priori probability $P(W)$ estimation of this sequence for an optional sequence of words $W=\{w_1w_2\ldots w_{n-1}\}$ and to extend the quickest and the most exact estimation of this sequence of words in the process of search strategy in ASR system. This a priori probability can be defined as follows [1]

$$\begin{aligned} P(W) &= P(w_1 w_2 \ldots w_n) = \\ &= P(w_1)P(w_2 \mid w_1)\ldots P(w_n \mid w_1 w_2 \ldots w_{n-1}) = \\ &= \prod_{i=1}^{n} P(w_i \mid w_1 w_2 \ldots w_{i-1}), \end{aligned} \tag{1}$$

where $P(w_1w_2\ldots w_{i-1})$ is the conditional probability of occurrence of word w_i conditioned by its history of words $w_1w_2\ldots w_{i-1}$. Such process of decomposition allow us to recognize for ASR system certain sequence of words during its pronouncing (in real time) and determine our probability $P(W)$ for decoding process purposes gradually.

However in practice, it is impossible to compute all of these probabilities for certain sequence of words, therefore the approximation is usually done by reducing the history of words to certain number. These SLM are called the *n*—gram models (for $n = 1$ we have the unigram model, for $n = 2$ we have bigram model etc.). Usually, in practice the trigram models for ASR systems are used. The trigram model can be defined as follows

$$P(W) = \prod_{i=1}^{n} P(w_i \mid w_{i-2} w_{i-1}). \tag{2}$$

In this case, the probability of word w_i is conditioned by history of two words, w_{i-2} and w_{i-1}. Main advantage of the *n*—gram models is simplicity of computation of their estimations of probabilities, which is based on computation of the relative occurrences of words, or sequences of words in the text set, called the *training set*, using *maximum likelihood estimation* (MLE) method [2]. For example, for computing the estimation of the bigram probability we can write

$$\bar{P}(w_i \mid w_{i-1}) = \frac{N(w_{i-1} w_i)}{N(w_i)}, \tag{3}$$

where $N(w_{i-1}w_i)$ is number of occurrences of bigram $w_{i-1}w_i$ and $N(w_i)$ is number of occurrences of unigram w_i.

Because every training set is finite and cannot include all word combinations, it might happens that such events can lead to the zero conditional probability, because a speaker can also pronounce a sentence that does not occur in training set. Moreover, zero probabilities lead to the errors in recognition. Therefore, every *smoothing technique* is based on this knowledge, which comprehensive summary can be found in [3], [4]. The problem of events that does not occur in training set, smoothing are resolved by more uniform redistribution of parts of probabilities of observed *n*—grams among *n*—grams that are not observed in training set. Using smoothing has not only better effect in recognition but also increases the accuracy of the SLM itself.

It is also necessary to limit the size of vocabulary that participates in the generating of the SLM. Moreover, with the size of vocabulary also increases the number of *n*—grams.

The storage of high-level *n*—grams (*n* = 3, 4, 5…) yields not only the problem with the memory requirements. Searching language models in decoding process is time consuming and requirements for computing are too high. Most of the high-level *n*—grams occur in training set just once. Therefore such SLMs need to be reduced, which can be done by using one of the *pruning techniques* that were published in [3], [5].

For efficient building of the SLMs, it is needed to create the optimal model just by using smoothing and pruning techniques. In addition, the right size of vocabulary should be chosen.

This article is organized in follows. First, the fundamental proposed model of building SLMs for Slovak LVCSR will be presented. Then, a short overview about every block of the proposed system will be described and some of problems in process of building correct vocabularies and SLMs will be mentioned. In following sections, the first experimental results of basic, smoothed and pruned SLMs and a short discussion about obtained results will be summarized. At the end of this paper the future intentions in language modeling of the Slovak language will be indicated.

II. Process of Building SLMs

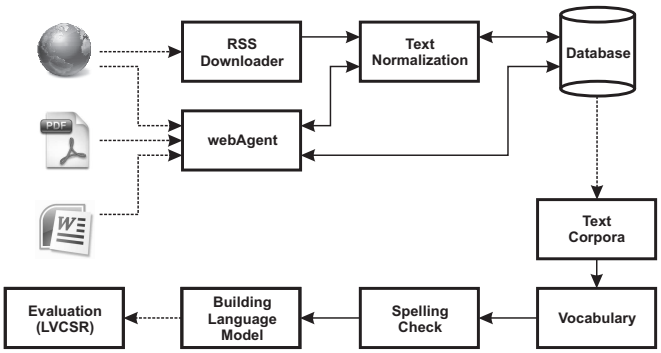


Fig.1. Demonstrative block scheme of the process of data mining, text normalization and building language model

Process of building SLMs for Slovak language consists of several parts, illustrated in Fig. 1. First, it is needed to collect a large amount of text data by automatic systems for text mining. These texts in uniform encoding (UTF-8) and raw text form are stored. Then all texts through the block of text normalization are passed. In this step, some additional text improvements are transacted. For example, into the verbal form are transcribed all numbers, abbreviations, symbols, etc. Then each of transcribed texts with description about their title, author, source etc. is stored in relational database. Selecting texts from database, the text corpora (general or domain-specific) are created and prepared for training SLMs. Training SLMs include operations of counting words, selection of vocabulary, spelling checking and generating various SLMs. Finally, consequential evaluations of SLMs are performed. In the following subsections, each of these steps for building SLMs in detail will be explained.

A. Text Mining

In the LVCSR system, it is needed to collect a large amount of text data in order to get the efficient SLM of given language. Texts are retrieved usually from electronic documents such as MS Word documents (DOC, RTF) or Portable Document Formats (PDF), etc. Another way is retrieving text data from web (HTML) pages. We use two automatic systems for text mining designed in our laboratory.

The first system, called **RSS Downloader**, retrieves the text data from HTML pages using RSS channels, which are manually predefined by user in configure manager. The system automatically expands every link in RSS channel and extracts text from every expanded HTML page, written in Slovak language, which in form of text file is stored [6], [7].

The second system, called **webAgent**, retrieves the text data from various Internet pages that are written in Slovak language. Besides text data, the system also collects links on the other web pages situated on a given web page. Moreover, the system is able to detect the encoding of given web page and also retrieves text data from PDF and DOC documents.

Both automatic systems have implemented tools for duplicity verification (at the level of the URI and content), spelling check and amount of various constraints for incorrect text exclusion, etc.

B. Text Normalization

The following steps include additional modifications of text data. All text data must fulfill following conditions:

- *Each of these sentences is in one line*. Sentence segmentation is not a simple task. Not every full-stop mark indicates the end of sentence (that can be indicated by some cardinal number or abbreviation). Usually, it is performed using appropriate regular expressions.
- *Text data are tokenized*. Token is a word, symbol, punctuation or hyphen. Cardinal numbers, dates, time stamp, abbreviations etc. are represented by.
- *Every word is mapped to lowercase*. In recognition step is not needed to especially recognize for example proper nouns. This part is component of a postprocessing step.
- *Numerals, dates, time stamps and other numbers are replaced by their pronunciation*. This step is one of the most difficult. Numbers must be transcribed in correct grammatical category, which is complicated in highly inflected language such as Slovak language. Correct grammatical form is usually obtained with the help of surrounding context by using hand-written rules (focus on the ending of the following word) or statistically.
- *Abbreviations and monetary units must be expanded*. Similar case like numbers.
- *All punctuation is deleted*, because is not needed in training process.
- *Sentences containing high count of grammatically incorrect words, words without diacritic marks and other ungrammatical events are filtered*.

C. Relational Database

Relational database is based on PostgreSQL and is closely associated with systems for text mining (see subsection A). In database both the raw texts obtained from web pages and normalized texts are stored. Duplicity verification by one of the system for data mining in process of insertion text data into the database is performed (both, on the level of the URI and content). Duplicity on the level of content by comparison

of hash codes of the entire blocks of text data is done. The database also contains a description about particular text data in form of metadata. Metadata involves URI of web page, article title, time stamp of received article and name of the source that where article were published.

D. Text Corpora

Text corpora are generated by selecting the set of articles from database in form of plain text. Usually, corpora are divided by the theme, for example corpus of news articles, corpus of blog pages, etc. Such corpora are known as domain-specific corpora. Then these corpora are further shifted to the training process.

E. Vocabulary

First step in the process of language modeling is building a vocabulary. Vocabulary includes a list of unique words that are performed by appropriate program for counting words. Because the number of unique words is usually too high (it can be contain grammatically incorrect words, etc.), the number of words is limited by the number of the most frequent words of a certain value (for LVCSR usually 100,000—350,000 of unique words).

F. Spelling check

Grammatically incorrect words or words without diacritics can occur in vocabulary, the spelling check is often needed. Spelling check is performed by separation all of the words that are not found in dictionary, which includes only grammatically correct words. In our case, the dictionary for spelling check is created by merging available Open Source dictionaries such as *aspell*, *hunspell*, and *ispell* [8] with dictionaries of proper nouns available on the Internet such as list of frequent names and surnames, cities, towns and villages, names of streets, geographic terms, names of companies, etc. Size of individual lists of unique words in dictionary for spelling check is illustrated on the Fig. 2. Remaining incorrect words are manually checked, corrected and added into the dictionary.

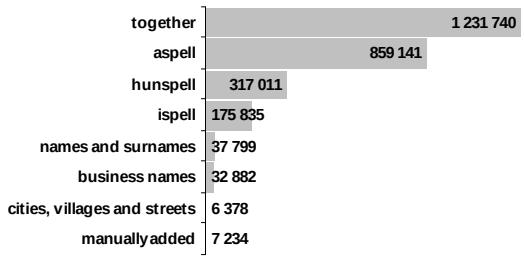


Fig.2. Size of particular dictionaries in dictionary for spelling check (number of unique words)

G. Building Language Models

For building language models toolkits are usually applied that by using various program tools can generate several types of SLMs such as smoothed models, pruned models, etc. One of the most universal toolkits is the SRI Language Modeling Toolkit [9]. Beside the necessary tools for building *n*—gram SLM in standard ARPA format, includes the set of tools for evaluation, pruning *n*—gram models, etc.

H. Evaluation

For evaluation SLM two standard measures are used, extrinsic evaluation using word error rate (WER) and intrinsic evaluation using perplexity.

WER is a common extrinsic measure of the performance of a speech recognition, which is defined by the formula

$$WER = \frac{N_S + N_D + N_I}{N}, \tag{4}$$

where N_S is the number of substituted words, N_D is the number of deleted words, N_I is the number of inserted words and N is the total number of words in reference. For evaluation of the WER the ASR system is needed to have itself.

Another measure is perplexity, which is defined as the reciprocal of the (geometric) average probability assigned by the language model to each word in the test set and is related to cross-entropy $H(W)$ by the equation [4]

$$PPL = 2^{H(W)} = \frac{1}{\sqrt[n]{P(w_1 w_2 ... w_n)}}, \tag{5}$$

where $P(w_1 w_2 \ldots w_n)$ is the probability of sequence of n words. Perplexity does not ensure necessary the increase in recognition itself, but usually highly correlates with such recognition improvement. Therefore, perplexity is often used if we do not have ASR system.

III. Experimental results

For speech recognition, we have used Julius high-performance LVCSR decoder (in version 4.1.3) adapted for the Slovak language. Recognition algorithm using Julius is based on a two-pass strategy. Therefore, it is necessary not to use only basic language models but also models trained on the reverse set of text data. In the case of two-pass strategy, the input data are processed in the first pass, and again the final search is performed again for the input using the result of the first pass to narrow the search space [10].

The acoustic model is made of context dependent phones [11], triphone HMMs with 32 Gaussian mixtures. It has been trained by using about 75 hours of annotated recordings of parliament speech.

The test data contains some manifestations recorded in parliament. The test data represent 75 minutes of speech and contain 8,778 words. The test vocabulary has 3,187 different Slovak words.

The corpus used for the training of word-based language models consists only of newspaper articles obtained from Internet pages. It contains about 2.5 million sentences with 30,820,506 Slovak words, 546,715 of them are in unique word forms. For building trigram language models three different sizes of vocabulary we used: 100k containing 100,275 unique words, 200k containing 200,711 words and 300k with 281,933 unique word forms. All language models (basic, smoothed and pruned) in standard ARPA format using SRI LM toolkit [9] were created.

We have tested four smoothing methods using SRI LM toolkit in training process: Katz model with Good-Turing discounts [4], absolute discount model [4], Witten-Bell model [4] and modified Kneser-Ney model [4] to the basic model and the effect of smoothing by evaluation of perplexity and WER were observed.

TABLE I
EVALUATION OF SMOOTHED MODELS

Evaluation	Size	Base	Katz-GT	Absolute discount	Witten-Bell	Modified Kneser-Ney
PPL bigrams	100k	1170.6	1170.6	1170.6	1170.1	1153.7
	200k	1289.1	1289.0	1289.1	1288.5	1269.6
	300k	1327.8	1327.8	1327.8	1327.0	1307.6
WER [%] bigrams	100k	23.97	23.99	24.03	23.99	24.07
	200k	22.90	22.88	22.85	22.85	23.07
	300k	22.86	22.86	22.86	22.92	23.13
PPL trigrams	100k	977.7	977.7	977.7	977.4	963.7
	200k	1075.8	1075.7	1075.8	1075.3	1059.5
	300k	1108.2	1108.2	1108.2	1107.5	1091.3
WER [%] trigrams	100k	18.76	18.77	18.77	18.77	18.85
	200k	17.57	17.59	17.56	17.51	17.65
	300k	17.08	17.04	17.08	17.07	17.21

Perplexity (PPL) and word error rate (WER) as a function of used smoothing method and language model sizes of different vocabularies.

Table I shows results of perplexity and WER for bigram and trigram language models of different sizes of vocabulary and various smoothing methods. As can we see, the modified Kneser-Ney algorithm significantly outperforms the others in the case of perplexity as it is published in [4], [7]. Unfortunately, perplexity is not always correlated well with result of recognition itself. For the condition of bigram model, the Witten-Bell algorithm usually performs better, whereas Kneser-Ney the worst, especially in the case of vocabulary size of 200k. Regarding to trigram language models, similar results we can be seen. Further we can observe that the increase of the order of *n*—gram language model requires larger vocabulary. The results show that the Witten-Bell and modified Kneser-Ney are the most interesting. It would be interesting to see how these smoothing techniques behave on several times larger training set, which might be subject promising of further research.

TABLE II
EVALUATION OF PRUNED MODELS

Threshold	PPL bigrams	WER [%] bigrams	PPL trigrams	WER [%] trigrams	Size [MB]
Base-KN	1288.5	23.30	1059.5	21.01	314
1e-8	1284.5	23.10	1070.4	20.46	243
25e-9	1302.4	23.31	1090.9	21.47	203
5e-8	1332.7	23.30	1122.2	21.20	160
75e-9	1358.7	23.31	1148.6	21.03	131
1e-7	1389.3	23.43	1175.3	20.92	110

Perplexity (PPL) and word error rate (WER) as a function of pruning threshold and language model sizes of a base 200k language model smoothed by Kneser-Ney method.

Table II shows language model perplexity and WER results evaluated on the test set for various pruning thresholds. We used relative entropy-based pruning [5] for the trigram model with Kneser-Ney smoothing, size of 200k. Only the one-pass strategy for recognition were used, because pruned language models trained on reverse texts were different in number of bigrams and trigrams. As is shown, pruning is highly effective in model size reduction. For the threshold equals 10^{-7} we have obtained a model that is 35% of the size of the original model with negligible degradation in recognition, both in the case of bigrams and trigrams. The best result, both perplexity and WER, was achieved when the threshold was 10^{-8}. Recognition results show, that the pruned model with this threshold is better in WER than the basic unpruned model for both cases, for bigrams and trigrams. It would be interesting to find the best pruning threshold (for the maximum reduction of size of the model) that would not cause a significant degradation in recognition, which will be subject for further research.

IV. CONCLUSIONS

In this article the complete building process of stochastic *n*—gram language models in case of the Slovak language have been presented. That includes process of text mining, text normalization and generation of language models. This is also the first attempt of such extent in research in Slovak language modeling. For efficient language modeling, both smoothing and pruning techniques were applied. This first experimental results lead to find the optimum ratio between the using one of the smoothing technique and appropriate pruning threshold. However, the fundamental problem is still the process of normalization of the text data into the shape most appropriate for correct training of models of inflective Slovak language.

In future work, we want to focus on the training of stochastic language models on several times larger training text set with optimal size of vocabulary. Further research should be also focused on the other types of language models suitable for inflective language such as adaptive models, morphologically motivated or morpheme-based language models.

ACKNOWLEDGMENTS

The research presented in this paper was supported by the Slovak Research and Development Agency under research projects APVV-0369-07 and VMSP-P-0004-09 and is the result of the project implementation Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research and Development Operational Programme funded by the ERDF.

REFERENCES

[1] J. Psutka, L. Müller, J. Matoušek, V. Radová, "*Mluvíme s počítačem česky*," Academia Praha, 2006, pp. 752, ISBN 80-200-1309-1.

[2] D. Jurafsky, J. H. Martin, "*Speech and language processing (2nd edition): An introduction to natural language processing, computational linguistics, and speech recognition*," Prentice Hall, Pearson Education, 2009, pp. 988, ISBN-13 978-0-13-504196-3.

[3] J. Staš, "*Morfologicky a morfematicky založené stochastické modely slovenského jazyka pre systémy ARR*," PhD thesis, 2010, pp. 117.

[4] S. F. Chen, J. Goodman, "*An empirical study of smoothing techniques for language modeling*," Technical Report TR-10-98, 1998, pp. 63.

[5] A. Stolcke, "*Entropy-based pruning of backoff language models*," In Proc. DARPA News Transcription and Understanding Workshop, 1998, pp. 270—274.

[6] M. Mirilovič, "*Stochastický jazykový model slovenského jazyka pre využitie v systémoch automatického rozpoznávania plynulej reči*," Dissertation thesis, 2009, pp. 142.

[7] M. Mirilovič, J. Juhár, A. Čižmár, "*Large vocabulary continuous speech recognition in Slovak*," In Proc. of the International Conference AEI'08 - Applied Electrical Engineering and Informatics, 2008, pp. 73—77, ISBN 978-80-553-0066-5.

[8] SK-Spell, "*Podpora slovenčiny v open source programoch*," [Online]: http://sk-spell.sk.cx/.

[9] A. Stolcke, "*SRILM – an extensible language modeling toolkit*," In Proc. of the International Conference on Spoken Language Processing, 2002, pp. 901—904.

[10] A. Lee, T. Kawahara, K. Shikano, "*Julius – an open source real-time large vocabulary recognition engine*," In Proc. European Conference on Speech Communications and Technology (EUROSPEECH), 2001, pp. 1691—1694.

[11] M. Mirilovič, J. Juhár, A. Čižmár, "*Comparison of grapheme and phoneme based acoustic modeling in LVCSR task in Slovak*," In A. Esposito et al. (Eds.), Multimodal Signals: Cognitive and Algorithmic Issues, LNAI 5398, Springer-Verlag, 2009, pp. 242—247, ISSN 0302-9743.

Visual Representation for Three-Dimensional Software Visualization

Kristián ŠESTÁK

Dept. of Computers and Informatics, FEI TU of Košice, Slovak Republic

kristian.sestak@gmail.com

***Abstract* — The paper presents a three-dimensional representation for visualizing software systems. The use of three-dimensional is almost as new as the more recent research directions and the focus is on the newer visualizations that make use of the extra dimension for the display of information. We present an overview of current research in the area, describing visual representations.**

***Keywords*— three-dimensional software visualization, 3D, software comprehension, visual representation, graphical representation.**

I. INTRODUCTION

The essence of software visualization consists of creating an image of software by means of visual objects. These visual objects may represent, for instance, systems or components or their runtime behavior.

To visually encode information, one can use text as well as two-dimensional (2D) or three-dimensional (3D) computer graphical representations. As software visualization mostly deals with software artifacts and their interrelations, graph-based representations play a major role.

Developing software systems is an arduous task, involving a set of related phases that spawn along the software lifecycle. During all these phases, software engineers need different ways to understand complex software elements.

Software visualization can be seen as a specialized subset of information visualization. This is because information visualization is the process of creating a graphical representation of abstract, generally non-numerical, data.

We focus on research in the areas of Knowledge-based software life cycle and architectures [1], [2], [7]

II. SOFTWARE VISUALIZATION

There are three primary types of visualization:

- Information visualization is defined as "the use of computer-supported, interactive, and visual representations of abstract data to amplify cognition" (Wakita & Matsumoto 2003).
- Scientific visualization is described by (Aref, Charles & Elvins 1994) as "when computer graphics is applied to scientific data for purposes of gaining insight, testing hypothesis, and general education"
- Software visualization is used to understand complex software systems and their lifecycles.

Software visualization is defined as "the use of the crafts of typography, graphic design, animation and cinematography with modern human-computer interaction and computer graphics technology to facility both the human understanding and effective use of computer software"

The goal is to provide a view on the software system on a higher level of abstraction which supports in understanding the software system.

The graphical representation of a software system is realized by displaying nodes for source code artifacts (e.g. classes, methods, attributes, etc.) and edges for relationships (e.g. inheritance, invocation, etc.). [3], [4], [6]

A. Software Visualization in 3D Space

For large, complex software systems, the comprehension of such diagrammatic depictions is restricted by the resolution limits of the visual medium (2D computer screen) and the limits of user's cognitive and perceptual capacities. One approach to overcome or reduce the limitations of the visual medium is to make use of a third dimension by mapping source code structures and program executions to a 3D space.

3D visualizations, the use of the third dimension is typically motivated by one or more of the following reasons:

Aesthetics: 3D graphics rendered with photorealistic rendering techniques are appealing to many people.

Evolution: Humans are used to 3D. It has been argued that the human visual system has been adapted to the real world – and this is a three-dimensional world.

Dimensionality: The third dimension can be used to add additional information to an originally two-dimensional representation. There are two notable instances of this:

- Multiple views: The same object can be shown in different ways by placing different, typically 2D, views of the object in the 3D space.
- History: The third dimension can be used as a time axis. Along this time axis, the states of an object at different points in time can be shown.

TABLE 1
3D STRENGTHS AND WEAKNESSES

Strengths	Weaknesses
+Greater information density. +Integration of local views with global views. +Composition of multiples 2D views in a single 3D view. +Facilitates perception of the human visual system. +Familiarity, realism and real world representations.	-Intensive computation. -More complex implementation. -User adaption to 3D metaphors and special devices. -More difficult for users to understand 3D spaces and perform actions in it. -Occlusion.

The work of Hubona, Shirah and Fout [Hubona et al. 1997] suggest that users' understanding of a 3D structure improves when they can manipulate the structure. Ware and Franck [Ware and Franck 1994] indicate that displaying data in three dimensions instead of two can make it easier for users to understand the data. In addition, the error rate in identifying routes in 3D graphs is much smaller than 2D [Ware et al.1993]. [5], [8]

III. VISUAL REPRESENTATIONS FOR 3D SOFTWARE VISUALIZATIONS

It is crucial not only to determine which information to visualize but also to define an effective representation to convey the target information to the user and support software engineering tasks.

Indeed, the design of software visualization must address a number of different issues, e.g., what information should be presented, how this should be done, what level of abstraction to support, etc.

Many representations for visualizing software have been proposed. For instance, some visual representations are based on abstract shapes such as graphs, trees and geometric shapes and others are based on real-world objects like 3D cities solar systems molecules video games metaballs 3D landscapes and social interactions among others.

Several 3D abstract visual representations based on graphs, trees, and geometrical shapes.

A. Abstract Visual Representations

Abstract visual representations based on graphs, trees, and geometrical shapes.

1) Graphs

A graph is a network of nodes and arcs, where the nodes represent entities such as procedures, objects, classes, or subsystems, while the arcs represent relationships between entities, such as inheritance or method calls. As seen in Fig. 2.

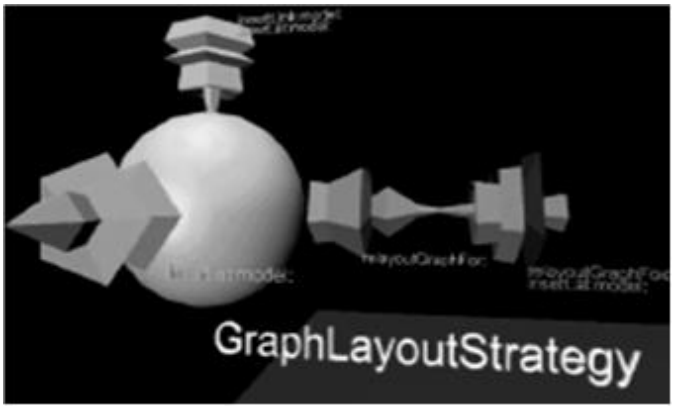


Fig. 1Node representations. Class and design patterns

2) Trees

A tree can represent many software entities such as subsystems, modules, or classes and the relationships between them such as inheritance or composition. Moreover, since trees have no cycles, unlike graphs, they are generally easy to layout and interpret. As seen in Fig. 3.

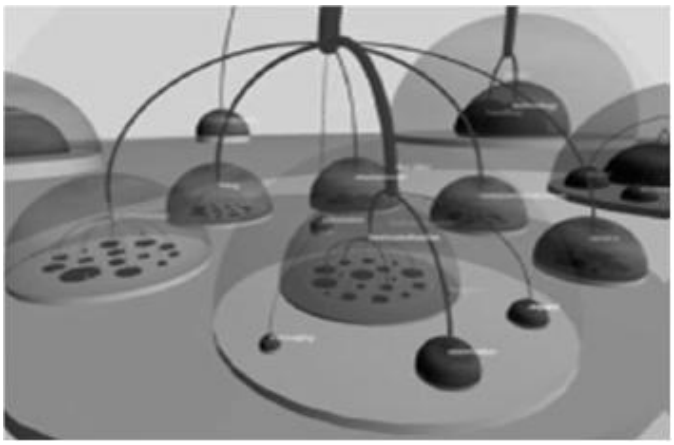


Fig. 2 Hierarchical Net 3D. Image courtesy of M.

3) Geometrical Shapes

Many software visualization tools use traditional node-link diagrams, but sometimes, they present scalability or layout problems. In an effort to explore new representations beyond graphs, several visualization techniques were proposed using abstract 3D geometrical shapes. As seen in Fig. 4.

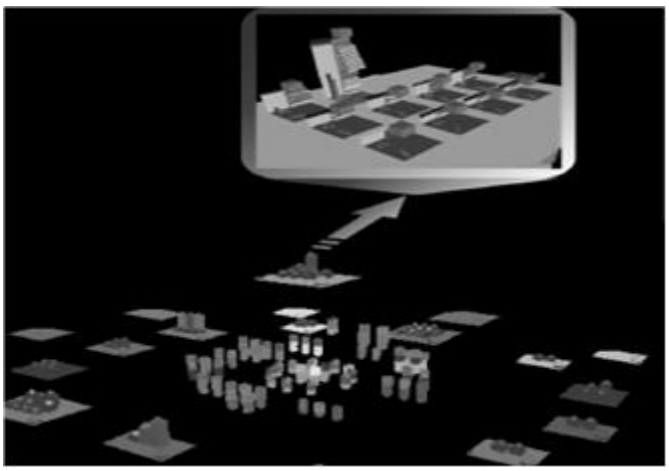


Fig. 3 Abstract software representations. Callstack and FileVis

B. Real World

Several researchers proposed using real-world metaphors. These techniques use well-understood elements of the world to provide insights about software.

Fig. 4 View over a software district showing many, possibly complex, methods

Further advancing the 3D space aspect of the visualizations, work by [5] moved to considering the use of virtual reality environments for software visualization. Software World was created to show that three-dimensions (in this case also showing the viability of real world metaphors) could be used to create automatable and scalable software visualizations. Buildings, cities, and also at the highest level atlas views, were used to represent Java source code. An example view of this visualization can be seen in Fig. 6. [2], [5]

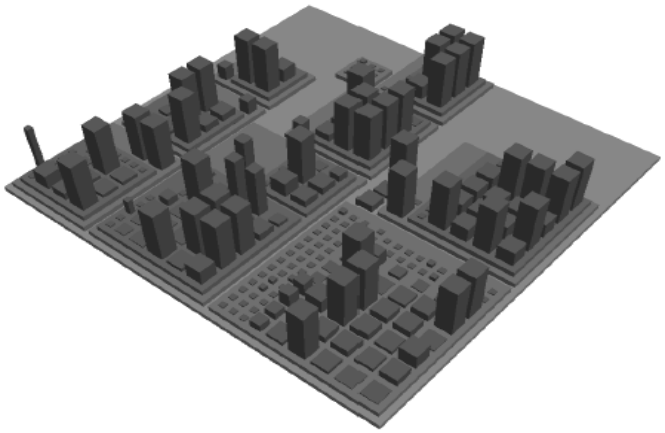


Fig. 5 Real-world metaphor. Visualizing Software Systems as Cities

IV. Conclusion

Visualizations developed for program and system comprehension should have two aims; to reduce the complexity of the perceived view of the software and to increase the user's understanding of the software.

Software visualization tools have to provide not only effective visual representations but also effective interaction styles to ease the exploration and help software engineers to achieve insight.

Using three dimensions for visualization adds an element of familiarity and realism into systems. The world is a three-dimensional experience and by making the visualization more like that world means there is less cognitive strain on the user.

One of the main problems for software visualization (and other forms of information visualization) is of trying to create a tangible representation o something that has no inherent form.

Given the complexity of software and the different problem solving characteristics of programmers, it is now well recognized that there is unlikely to be any one single visualization metaphor that can be considered most optimal for software visualization. [2], [7]

Acknowledgment

This work was supported by VEGA Grant No. 1/0350/08 Knowledge-Based Software Life Cycle and Architectures. This work is the result of the project implementation: Centre of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120020) supported by the Research & Development.

References

[1] Havlice, Zdenek et al. : Knowledge-based software life cycle and architectures. In: Computer Science and Technology Research Survey. Košice: TU, 2007. s. 47-68. ISBN 978-80-8086-071-4.

[2] Alfredo R. Teyseyre, Marcelo R. Campo, "An Overview of 3D Software Visualization," IEEE Transactions on Visualization and Computer Graphics, pp. 87-105, January/February, 2009.

[3] J. T. Stasko, J. Domingue, M. H. Brown, and B. A. Price, editors. Software Visualization— Programming as a Multimedia Experience. The MIT Press, 1998.

[4] Wysseier, Ch., "Interactive 3-D Visualization of Feature-traces", Master Thesis Philosophisch-naturwissenschaftlichen Fakultat der Universitat Bern, 2005.

[5] Knight, C., and Munro, M. "The Power of (Software) Visualization", Department of Computer Science Technical Report 01/00, January 2000.

[6] Holmberg, N., Wünsche, B., and Tempero, E. 2006. A framework for interactive web-based visualization. In Proceedings of the 7th Australasian User interface Conference - Volume 50 (Hobart, Australia, January 16 - 19, 2006). W. Piekarski, Ed. ACM International Conference Proceeding Series, vol. 169. Australian Computer Society, Darlinghurst, Australia, 137-144.

[7] Knight, C., System and Software Visualisation, Visualisation Research Group, Department of Computer Science, University of Durham, Durham, DH1 3LE, UK., 2000

[8] Šesták, K., "Using Virtual Reality Environment for Modeling Software System," 9th Scientific Conference of Young Researchers, Kosice 13th May 2009

Channel estimation error of comb-type pilot symbol arrangement in nonlinearly distorted OFDM system with iterative compensation algorithm

Ján ŠTERBA, Radovan BLICHA

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

sterba.jan@gmail.com, rblicha@gmail.com

***Abstract*—Utilization of high power amplifiers in their nonlinear part of characteristic at the OFDM transmitters lowers their energy consumption, but comes with the cost of nonlinear distortion inflicted into transmitted signal. When assuming perfect channel state information present at the receiver, only data symbols are nonlinearly distorted, what results in a severe performance decrease. However, in practice channel state information must be obtained through estimation process, which is usually based on insertion of so called pilot symbols into transmission and subsequent pilot symbols channel estimation. In this case, also transferred pilot symbols are subject to nonlinear amplification and distortion, which lowers the accuracy of channel estimation process, and further decreases the transmission performance. This article concerns with evaluation of channel estimation error for a comb-type pilot symbol arrangement in a situation of nonlinear amplification, and with an improvement of channel estimation error by the means of iterative algorithm for nonlinear distortion estimation, channel re-estimation and distortion cancellation.**

***Keywords*—OFDM, comb-type pilot symbol channel estimation error, nonlinear amplification, iterative algorithm.**

I. INTRODUCTION

Orthogonal Frequency Division Multiplex (OFDM) transmission scheme is one justified major candidate for Beyond 3G and 4G wireless communication systems [1], and has very promising potential to meet the demands for high data rate transmissions over multipath radio channels of future communication systems. OFDM based communication systems have large number of benefits in comparison with traditional schemes. On the other hand, high Peak-to-average power ratio (PAPR) of OFDM signal makes it very sensitive to the nonlinear amplification which results in the high Bit Error Rate (BER) penalty as well as to the enormous out-of-band radiation. These effects have harmful impact on OFDM system performance and therefore steps to mitigate these effects must be taken. Many different techniques have been introduced to mitigate large sensitivity of OFDM systems to nonlinear amplification. The conventional solution is to back-off the operating point of the nonlinear amplifier, but this approach results in the significant power efficiency penalty. Other alternative approaches are active constellation extension [2], tone reservation [3] or selected mapping [4], which are quite computational demanding. Another solution is to use nonlinear detector at the receiver side. The first contribution on this topic was proposed by Kim and Stuber in [5] for reduction of clipping noise of OFDM symbols by decision-aided reconstruction at the receiver. In [6] Declerq et al. proposed reducing the clipping noise in OFDM by introducing a Bayesian interference to the received signal. Finally Chen et al. [7] and Tellado et al. [8], proposed iterative techniques to estimate and eliminate the clipping noise in OFDM.

Besides the well-known and investigated BER degradation and out-of-band radiation, nonlinear amplification has in addition another important consequence. The pilot symbols, which are inserted into selected time and frequency positions of OFDM frame for the purpose of acquiring channel state information (CSI) are also nonlinearly distorted and therefore CSI acquired by traditional estimation techniques is also severely degraded.

In this paper, CSI error inflicted by nonlinear distortion on pilot symbols is measured for Saleh model of nonlinearity. Moreover, improvement of CSI error by the means of iterative algorithm for channel re-estimation and nonlinear cancellation [9] is measured and evaluated.

II. OFDM SYSTEM MODEL

The block scheme of the investigated OFDM system is presented in Fig. 1. Bits assigned for the transmission are first mapped into the complex-valued vector of 64-QAM constellation points. Then, every $m-th$ block of the mapped symbols is put into parallel streams using the serial to parallel converter. In the next step, pilot symbols are inserted into the OFDM frame according to comb-type pilot symbol arrangement, which is depicted in Fig. 2, with a uniform distribution of pilot symbols among all N sub-carriers. Then, obtained signal is send to the block of Inverse Fast Fourier Transform (IFFT) for OFDM modulation. If $x_f^m(n)$ is resultant T/N-spaced discrete-time vector, and m denotes the m-th block of the input

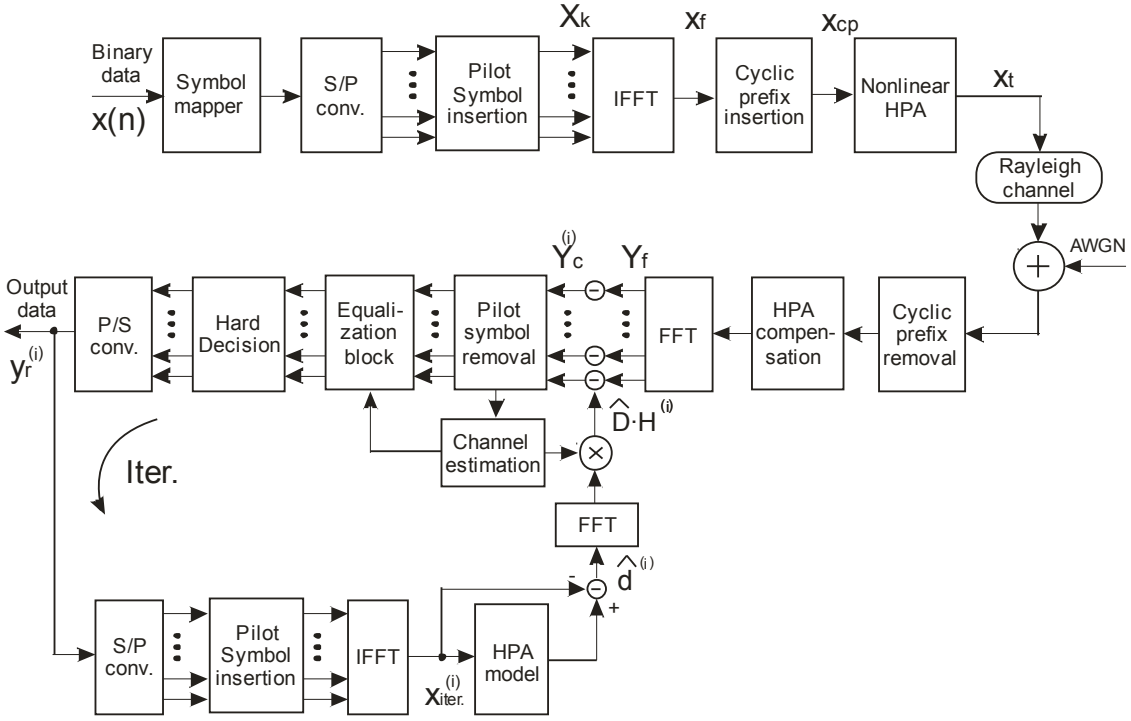


Fig. 1. The block scheme of the OFDM system with iterative receiver.

symbols, the IFFT can be described as follows:

$$x_f^m(n) = IFFT(X_k^m) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k^m \exp\left(\frac{j2\pi kn}{N}\right). \quad (1)$$

With the aim to mitigate the inter-symbols interference (ISI) caused by multipath propagation of the transmitted signal, carefully selected cyclic prefix (CP) is inserted. CP consists of cyclically extended part of the OFDM symbol over the time interval $[0, T_{CP}]$, and it's length is set longer than the delay spread assumed during the transmission. The resulting OFDM symbol after insertion of CP is given as:

$$x_{cp}(n) = \begin{cases} x_f(N+n), & n = -N_g, -N_g+1, ..., -1 \\ x_f(n), & n = 0,1,...,N-1 \end{cases} \quad (2)$$

where Ng is the length of a guard interval and N is the length of the OFDM symbol. Finally, the signal $x_{cp}(n)$ is amplified in a high-power amplifier (HPA), and sent to the antenna for transmission.

Nonlinear Amplification:

Both two most used high power amplifiers (TWTA), traveling wave tube amplifiers and solid state power amplifiers (SSPA) are not perfectly linear in their entire range of characteristic. If operating in it's linear part, their energy efficiency is decreased, but signal is undistorted. However, when lower energy consumption is required, characteristic must be widened to include for its nonlinear part. In this paper, Saleh model has been used to model the nonlinearity of HPA, which is typical model for TWTA amplifiers. The Saleh model can be described by the following amplitude-to-amplitude modulation (AM/AM) and amplitude-to-phase modulation (AM/PM) characteristics:

$$G(u) = \frac{\kappa_G . u}{1 + \chi_G . u^2}, \qquad \Phi(u) = \frac{\kappa_\Phi . u^2}{1 + \chi_\Phi u^2} \quad (3)$$

where $\kappa_G = 2, \chi_G = \chi_\Phi = 1$ and $\kappa_\Phi = \pi/3$ was chosen. The operation point of HPA is defined by so called input back-off (IBO), which is defined as:

$$IBO\ [dB] = 10 \log_{10}\left(\frac{P_{\max}}{P_{input}}\right) \quad (4)$$

The output signal x_t, as can be derived from Bussgang theorem [10], can be written as the sum of scaled version of the input signal x_{cp} plus uncorrelated distortion term $d(t)$:

$$x_t(n) = \alpha . x_{cp}(n) + d(n), \quad \text{where} \quad \alpha = \frac{R_{xy}(\tau_1)}{R_{xx}(\tau_1)} \quad (5)$$

where R_{xy} is the cross-correlation function of the signals $x_t(n)$ and $x_{cp}(n)$, and R_{xx} is the autocorrelation function of the signal $x_{cp}(n)$, and $\tau_1 = 0$. The complex scaling term α can be easily compensated by introducing correcting factor

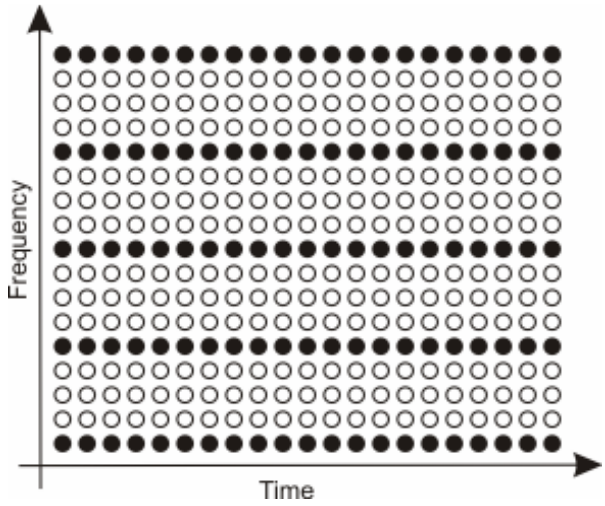


Fig. 2. An example of comb-type arrangement of pilot symbol in an OFDM frame.

$\alpha^*/|\alpha|^2$, however the distortion term $d(n)$ cannot be compensated by conventional receivers [11].

III. Iterative Compensation of Nonlinear Distortion

Investigated iterative receiver decreases the nonlinear distortion inflicted to data and pilot symbols by a nonlinear estimation and cancellation algorithm, firstly proposed in [12] for clipping noise mitigation, in an iterative manner.

In a backward loop of iterative receiver, received signal is send to the model of HPA with knowledge of nonlinear function $q(\cdot)$. Nonlinear distortion $d(n)$ can be estimated by subtracting nonlinearly distorted signal from its original version, thus obtaining estimate of distortion inflicted to data as well as pilot symbols:

$$\hat{d}^{(i)}(n) = q(x_{iter.}^{(i)}) - x_{iter.}^{(i)} \tag{6}$$

Although the estimate of the nonlinear distortion $\hat{d}(n)$ is degraded by the incorrect estimate of the transmitted signal $x(n)$, the overall performance will be improved if:

$$E[|d(n) - \hat{d}^{(i)}(n)|^2] < E[|d(n)|^2] \tag{7}$$

The estimate of nonlinear distortion is then transformed into frequency domain, multiplied with estimated CSI, and subtracted from the output signal $Y_f(k)$ of FFT, thus decreasing nonlinear distortion in an iterative manner:

$$Y_c^{(i+1)}(k) = Y_f^{(i)}(k) - \hat{D}^{(i)} \cdot H(k)^{(i)}. \tag{8}$$

Then, using the corrected pilot symbols, which are now less distorted by nonlinear distortion, the channel is re-estimated and used for the more reliable equalization and subsequent data detection. Then, the iterative process repeats, but now with less distorted data and pilot symbols by the nonlinear distortion. For more information regarding investigated iterative receiver and its mathematical description, see [9].

IV. Simulation Results

The simulations of OFDM communication system were obtained using the Monte Carlo computer simulations. The performance results were evaluated for 256 subcarriers with 64-QAM modulation for Saleh model of nonlinearity. Transmitted signals were generated by using 4-multiple oversampling, with cyclic prefix set to 3.33 µs, utilizing subcarrier spacing 18.74 kHz. The transmission channel was modeled as 4-tap COST 207 RA (Rural Area) Rayleigh fading channel, and zero forcing was utilized as equalization method. Channel estimation error was evaluated for comb-type pilot symbol arrangement, measured utilizing Mean absolute percentage error (MAPE), defined as:

$$MAPE = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| \frac{r_i - e_i}{r_i} \right| \cdot 100\% \tag{9}$$

where r_i is real value of CSI and e_i is estimated value of CSI obtained from pilot symbols at symbol positions.

In a results outlined in Fig. 3. is shown a development of CSI error in a communication system with changing noise conditions with IBO levels ranging from 1 to 12 dB. As can be observed from the figure, the worst case is obtained for IBO = 1 dB, with distortion of CSI obtained through estimation process with comb-type arrangement at approximately 40 %. The distortion then decrease gradually. An improvement of CSI error by means of iterative algorithm of nonlinear distortion cancellation for IBO = 4 dB, is shown in Fig. 4. As can be observed from the figure, for the noise value of Eb/No = 35 dB, the very first iteration of investigated algorithm improves the CSI error from 28,8 to 18,0 %, by a merely 10 %. The next iterations provided 4.5, 5, 4.2, 1.5 and 0.6 % improvement of CSI error.

Development of CSI error for different iterations of iterative receiver for and for conventional receiver and is

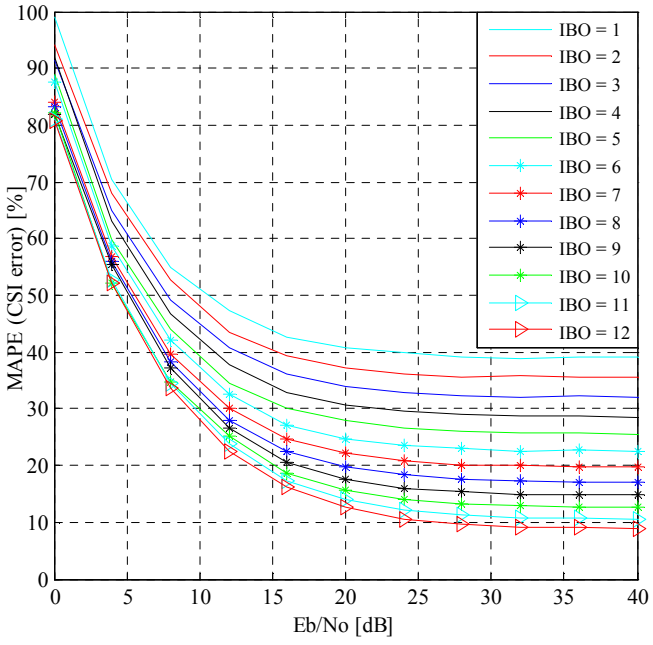


Fig. 3. Channel state information error for comb-type arrangement of pilot symbols and Saleh model of nonlinearity with different IBO values.

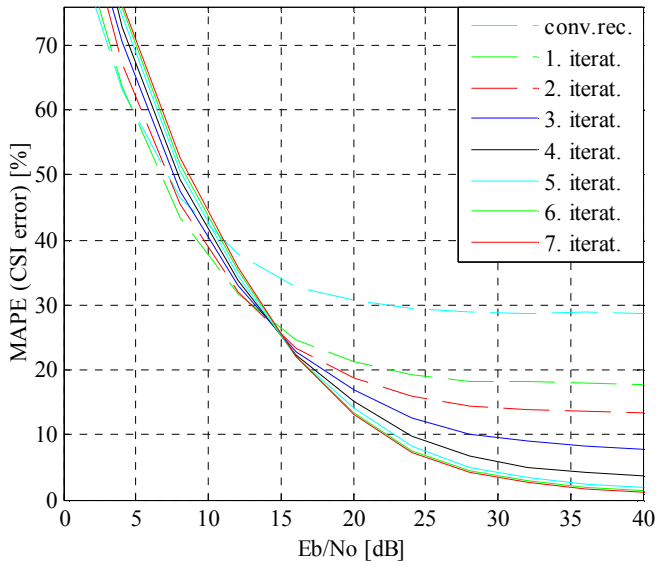


Fig. 4. Improvement of CSI error for different iterations of iterative receiver, Saleh model of nonlinearity, IBO = 4 dB.

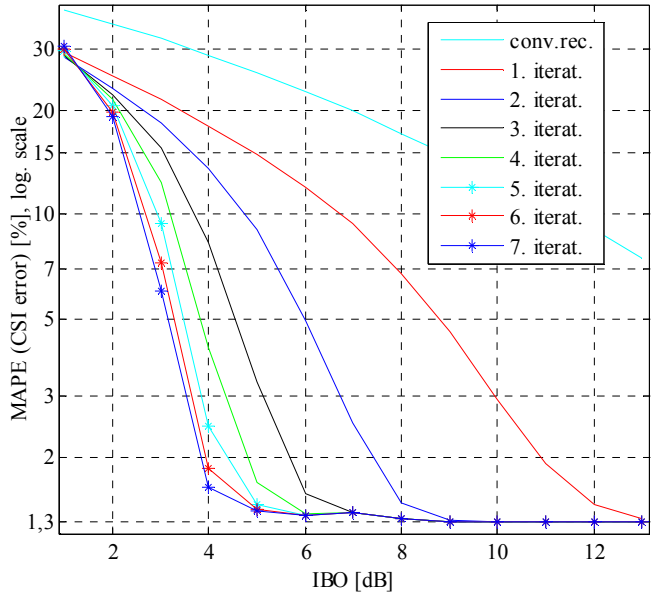


Fig. 5. Development of CSI error for OFDM transmission system with Saleh model of nonlinearity, IBO = 1-12 dB, and Eb/No = 35 dB.

depicted on Fig. 5. As can be seen from the figure, the first iteration of iterative receiver provided best improvement of CSI error, from 35 % to 25 % at IBO = 2 dB, from 25 % to 14 % at IBO = 5 dB and from 17 % to 7 % at IBO = 8 dB. For this level, the very next iteration provided improvement of CSI error to 1.5 %, very close to the lower bound of improvement. As can be seen from the results, the lower bound for improvement of CSI error for a comb-type pilot symbol arrangement by an iterative receiver is approximately 1.3 %.

The improvement of CSI error by iterative receiver is significantly slower, e.g. requires more iterations for lower IBO levels, requiring full 7 iterations to obtain 1.5 % CSI error at IBO = 4 dB, but faster at higher IBO levels, requiring just 3 iterations at IBO = 6 dB and 2 iterations at IBO = 8 dB to reach the same, 1.5 % level of CSI error.

V. Conclusion

In this article, channel state information error for OFDM communication error with comb-type pilot symbol arrangement is measured and evaluated for Saleh model of nonlinearity with IBO levels ranging from 1 to 12 dB. Additional improvement of channel estimation error is measured for 7 iterations of channel re-estimation and nonlinear distortion compensation algorithm. The results have shown that investigated iterative algorithm can be used efficiently to reduce the channel estimation error inflicted by nonlinear distortion in a case of comb-type pilot symbol channel estimation.

VI. Acknowledgement

This work is the result of the project implementation Center of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research & Development Operational Program funded by the ERDF. This work was supported also by the project COST Action IC0803: RF/Microwave Communication Subsystems for Emerging Wireless Technologies. This work has been also funded by Grant Agency SPP Hlavička.

References

[1] E. Dahlman, S. Parkvall, J. Skold, P. Beming, "3G Evolution. HSPA and LTE for Mobile Broadband", Academic Press, Elsevier, 2007.

[2] B.S. Krongold, D.L. Jones, PAR Reduction in OFDM via Active Constellation Extension. IEEE Transactions on Broadcasting, vol. 49, No.3, pp. 258-268, Sep. 2003.

[3] M. Deumal, A. Behravan, T. Eriksson, J. L. Pijoan, Evaluation of performance improvement capabilities of PAPR reducing methods. Wireless Personal Communications, vol. 47, no. 1, pp. 137-147, Oct. 2008.

[4] L.J. Cimini, N.R. Sollenberger, Peak-to-Average Power Ratio Reduction of an OFDM Signal Using Partial Transmit Sequences, IEEE Communications Letters, pp. 86-88, Mar. 2000.

[5] D. Kim, G.L. Stuber, Clipping noise mitigation for OFDM by decision aided reconstruction, IEEE Communication Letters, vol. 3, pp. 4-6, 1999.

[6] D. Declerq, G.B. Giannakis, Recovering clipped OFDM symbols with Bayesian inference, in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 157-160, Jun. 2000.

[7] H. Chen, A.M. Haimovich, Iterative Estimation and Cancellation of Clipping Noise for OFDM Signals, IEEE Communications Letters, vol. 7, pp. 305-307, Jul. 2003.

[8] J. Tellado, L. Hoo, J.M. Cio, Maximum-Likelihood Detection of Nonlinearly Distorted Multicarrier Symbols by Iterative Decoding, IEEE Transactions on Communications, vol. 51, pp. 218-228, Feb. 2003.

[9] J. Sterba, J. Gazda, M.H. Deumal, D. Kocur, Iterative algorithm for channel re-estimation and data recovery in nonlinearly distorted OFDM systems, Acta Hungarica, 2010. [Available Online: http://www.sterba.jan.szm.com/pdfs/tuke/acta_hun.pdf]

[10] J. Bussgang, Crosscorrelation function of amplitude-distorted Gaussian signals, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, Tech. Rep. no. 216, Mar. 1952.

[11] M.H. Deumal, Multicarrier communication systems with low sensitivity to nonlinear amplification, Enginyeria i Arquitectura La Salle, Universitat Ramon Llull, Barcelona, 2008.

[12] N.Y. Ermolova, N. Nefedoc, S.G. Haggman, An iterative method for non-linear channel equalization in OFDM systems, PIMRC 2004, Helsinki, Finland, pp. 484 -488, 2004.

Intelligent Tracking Trajectory Design of Mobile Robot

Peter ŠUSTER

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

peter.suster@tuke.sk

***Abstract*— This paper introduces a solution to the reference trajectory tracking problem done by a differential wheeled mobile robot Khepera II. The paper includes a mathematical model of mobile robot, which we use for the acquisition of a set training data for creating forward and inverse neural model. The purpose of the control structure was the reference trajectory tracking, which we verified using the Neural Network Toolbox of Matlab/Simulink.**

***Keywords*— mobile robot, MLP neural network, forward neural model, inverse neural model.**

I. INTRODUCTION

The primary task of every mobile robot in the industry is to track predefined trajectory form its initial to a final position. Track trajectory of mobile robot is possible by using neuro-fuzzy controller [9]. In our paper, we have used neuro approach for tracking trajectory. Training data, necessary for proposal nonparametric controller, we have obtained from simulation model of the robot, which was controlled the proposed control structure. Simulation model of the mobile robot was used for verify algorithms of tracking defined reference trajectory. Simulation model of the mobile robot is based on a real mobile robot Khepera II of K-team Corp. [8].

II. MATHEMATICAL – PHYSICAL MODEL OF MOBILE ROBOT

Created a model is based on several assumptions, namely that the robot moves on a perfect flat surface without sliding and also neglects the rolling resistance of the wheels. Position of the mobile robot is given by the coordinates x, y and angle θ, which represents the rotation of the mobile robot in relation to the chosen coordinate system. Mobile robot is controlled by the angular velocities of the wheels ω_L, ω_R. Between the angular velocities ω_L, ω_R and peripheral speeds v_L, v_R there are the following relations

$$v_L = r\omega_L \ , \ \ v_R = r\omega_R \tag{1}$$

where r is radius of the wheel. Position and rotation of the robot in the space can be based on (1) to express the following equations, which form a kinematic model of the mobile robot (Fig.1)

$$\begin{aligned} \dot{x}(t) &= v\cos\theta \\ \dot{y}(t) &= v\sin\theta \\ \dot{\theta}(t) &= \omega \end{aligned} \quad \Rightarrow \quad \begin{aligned} v &= \frac{v_L + v_R}{2} \\ \omega &= \frac{v_L - v_R}{b} \end{aligned} \tag{2}$$

where the inputs into the kinematic model are speeds wheels v_L resp. v_R and the outputs are x, y, θ. The kinematic model (Fig.1) allows us to determine the position and rotation of the robot under the condition that we know the initial state of the robot and we have updated information about the speed of the individual wheel [10].

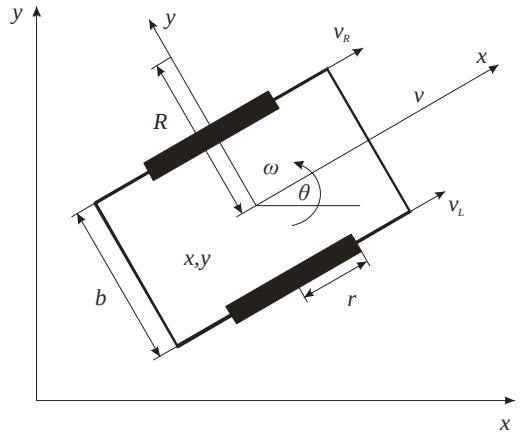


Fig. 1 Kinematic model of mobile robot

The kinematic model does not include friction forces acting on the wheel and the total mass of the mobile robot, so we have extended the mathematical – physical model about the dynamic part (Fig.2), which has the following shape:

$$\begin{aligned} ma_t &= F_L + F_R \\ J\varepsilon &= \frac{(F_L - F_R)b}{2} \end{aligned} \tag{3}$$

where tangent acceleration a_t is given by mass of the robot m and tangent forces F_L a F_R, which acting on the wheels due to change in the rotation speed. Angular acceleration ε is determined by the same forces, the moment of inertia of the robot J and distance between the wheels b [3]. Angular velocities ω_L and ω_R $(\omega_L = \dot{\theta}_L, \omega_R = \dot{\theta}_R)$ of the mobile

robot are driven by the voltage U_L and U_R. Differential equations expressing this fact have the following shape [1]

$$J\ddot{\theta}_L(t) + F_T\dot{\theta}_L(t) + F_L r = U_L$$
$$J\ddot{\theta}_R(t) + F_T\dot{\theta}_R(t) + F_R r = U_R \quad (4)$$

where F_T is friction force acting on the wheel. From equations (3) and (4), we have obtained dynamic model of the mobile robot in the state space :

$$\dot{x}(t) = Ax(t) + Bu(t)$$
$$y(t) = Cx(t) \quad (5)$$

where the state variables and their derivates have the following physical meaning:

$$x(t) = [x_1(t), x_2(t), x_3(t), x_4(t)] = [v(t), \omega(t), \omega_L(t), \omega_R(t)]$$
$$\dot{x}(t) = [\dot{x}_1(t), \dot{x}_2(t), \dot{x}_3(t), \dot{x}_4(t)] = [a_t(t), \varepsilon(t), \varepsilon_L(t), \varepsilon_R(t)]$$

and inputs into the system are:

$$u = [u_1(t), u_2(t), u_3(t), u_4(t)] = [F_L, F_R, U_L, U_R]$$

And outputs from the system are:

$$y(t) = [y_1(t), y_2(t)] = [x_3(t), x_4(t)].$$

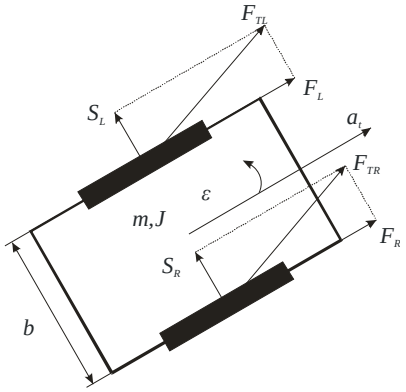


Fig. 2 Dynamic model of mobile robot

We programmed simulation scheme of the mobile robot (Fig.3)(Fig.4) in the Matlab/Simulink, based on the equations of the kinematic (2) and dynamic model (3) (4) :

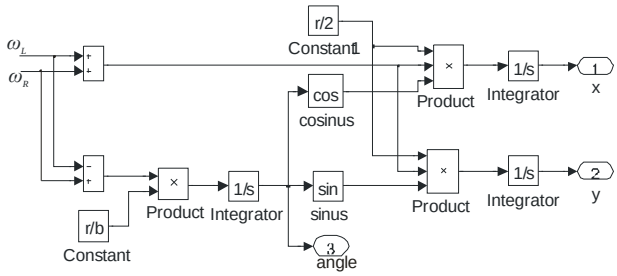


Fig. 3 Simulation scheme of mobile robot - kinematics

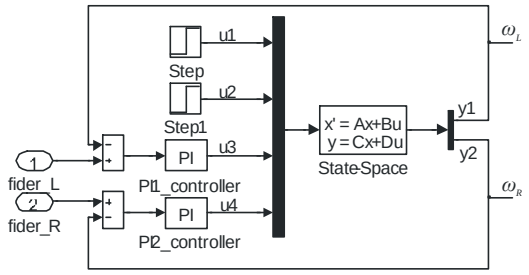


Fig. 4 Simulation scheme of model robot – dynamics

We proposed a control structure to ensure that the mobile robot can track one of the set reference trajectory [2]. The inputs into control structure of model robot are coordinates of current position of model robot *x*, *y* and coordinates of reference trajectory x_{ref}, y_{ref}. We have calculated Euclidean distance between current and desired position of the model robot by means of these coordinates. The outputs from control structure are angular velocities for left and right wheels. Subsystems control structure and model robot (Fig.5) we used in the simulation schemes for acquisition of training data necessary for design forward and inverse neural model. Simulations were carried out in the sample period $T_{vz} = 0{,}01s$.

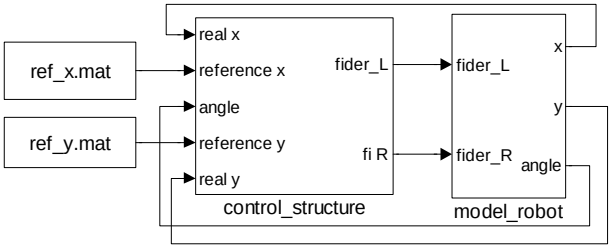


Fig. 5 Simulation scheme is designed for to simulate movements of the mobile robot

III. FORWARD NEURAL MODEL OF MOBILE ROBOT

Neural model, which approximates dynamic of the system is called forward model. Neural network is placed in parallel with identification system and error between output of the neural network $\hat{y}(k+1)$ and output of the dynamic system $y(k+1)$, the so-called prediction error, is used as training signal for neural network (Fig.6). Forward network of MLP type was used as neural network.

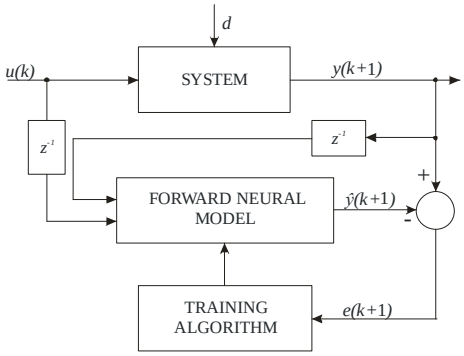


Fig. 6 Identification scheme based on output prediction error

If the output of the neural model is $\hat{y}(k+1)$ then we can express the equation by approximation

$$\hat{y}(k+1) = \hat{f}[y(k),...,(k-n+1),u(k),...,u(k-m+1)] \quad (6)$$

where $\hat{f}$ is represents the non-linear input-output representation by the neural model and $y(k)$ resp. $u(k)$ is n - output resp. m - input of the previous values [4].

Training data for proposal forward neural model, we obtained from simulation scheme to simulate the movement of the robot along defined reference trajectory. Reference trajectory was represented by vectors x and y coordinate. For training the forward neural model, we used a forward neural network of Multi Layer Perceptron (MLP) type with ten neurons in the input layer, with ten neurons in the hidden layer and with two neurons in the output layer (Fig.7). The training of forward neural model was carried out by the Levenberg-Marquardt algorithm using Neural Network Toolbox.

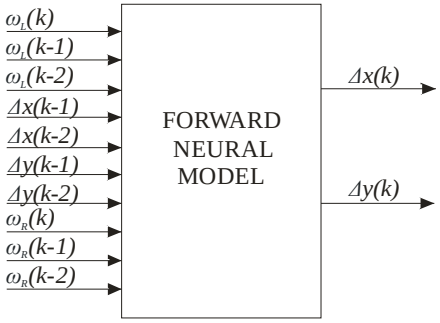


Fig. 7 Forward neural model of the mobile robot

The validation of the model is the next step after training of the neural model. The result of testing of trained forward neural model (Fig.7) is shown in the Fig.8. From picture (Fig.8) shows that forward neural model can approximate with accuracy, which meets for its further use at the tracking defined reference trajectory.

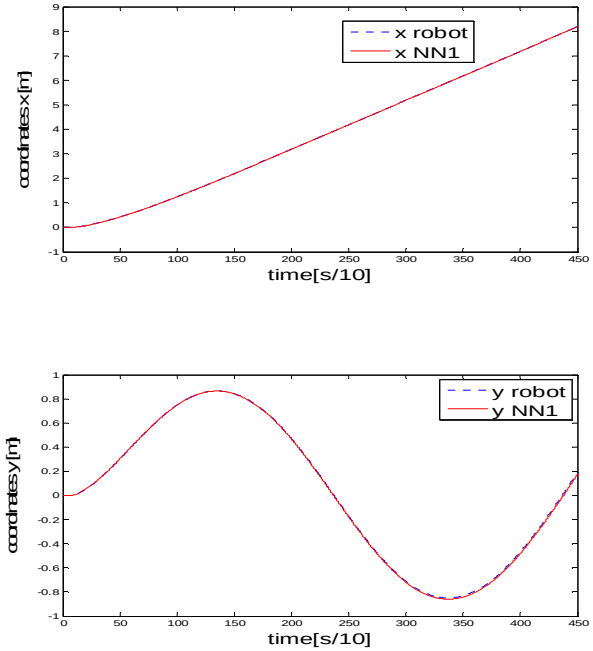


Fig. 8 Comparison outputs of the system and the forward neural model

IV. INVERSE NEURAL MODEL OF MOBILE ROBOT

Inverse neural model of the system is an important part of the theory of control. If the forward neural model was described by the equation (6), then the inverse model can be expressed in the form:

$$u(k) = f^{-1}\begin{bmatrix} r(k+1), y(k),.., y(k-n+1),... \\ ....u(k),....,u(k-m+1) \end{bmatrix} \quad (7)$$

where $y(k+1)$ is an unknown value, therefore it is substituted by the reference value of the control variable $r(k+1)$. To obtain inverse neural model, we have chosen General training architecture (Fig.9), which requires a known reference trajectory $r(k)$. Signal $u(k)$ is applied to the inputs of structure based on input predictive error with the aim of to obtain a corresponding system output $y(k)$, while the neural network is trained by the error $e_u(k)$, which is obtained as the difference of the neural model output $\hat{u}(k)$ and input signal $u(k)$ into the system [4].

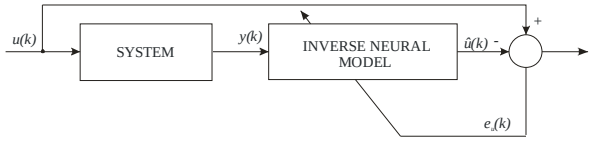


Fig. 9 General training structure

For training the forward neural model, we used forward neural network of Multi Layer Perceptron (MLP) type with fourteen neurons in the input layer, with five neurons in the hidden layer and with two neurons in the output layer (Fig.10). Training of forward neural model was carried out by the Levenberg-Marquardt algorithm.

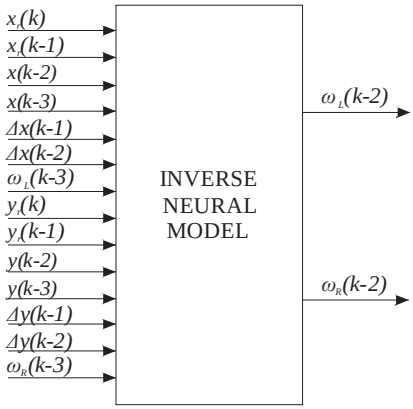


Fig. 10 Inverse neural model of mobile robot

We applied inverse neural model (Fig.10) together with forward neural model (Fig.7) into control structure IMC (Fig.12), which we have used for tracking defined reference trajectory. We proposed the IMC filter into control structure for better tracking trajectory. The goal of the tracking is to control the movement of the mobile robot from the point A to the point B along the chosen reference trajectory (Fig.11).

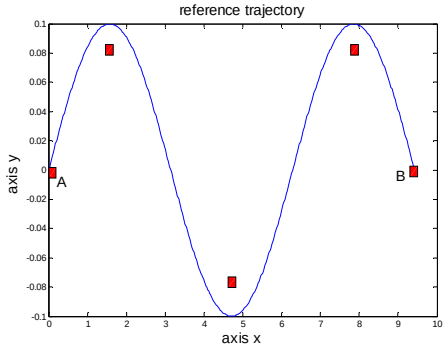


Fig. 11 Reference trajectory

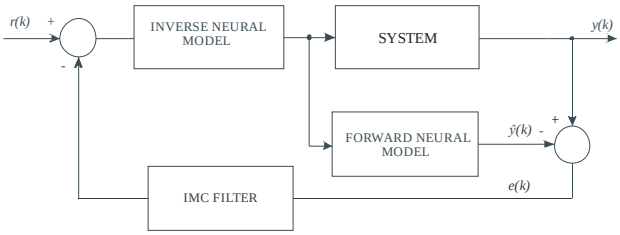


Fig. 12 Control structure Internal Model Control

The output from control structure IMC is current trajectory of simulation model of mobile robot (Fig.13), which is controlled nonparametric neural controller.

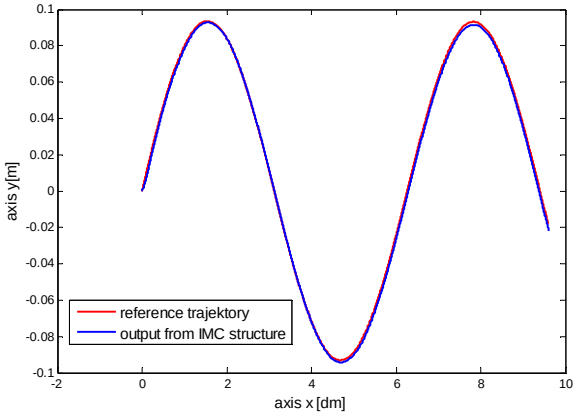


Fig. 13 Comparison the defined reference trajectory and output from IMC structure

From Fig.13, we can see that simulation model of mobile robot tracks the defined reference trajectory. We verified the functionality for other sinus trajectories with other amplitudes. When we have changed trajectory is necessary training the new inverse and forward neural model.

V. CONLUSION

We have analyzed the problem of tracking predefined reference trajectory of the mobile robot in the paper. As solution to the problem, we have proposed nonparametric neural controller, which we implemented into the control structure IMC together with forward neural model. Training data, necessary for proposal nonparametric controller, we have obtained from simulation model of the robot. The obtained knowledge in the field tracking reference trajectory of the mobile robot, we want to use for real mobile robot Khepera III, which are in our laboratory at the Department of Cybernetics and Artificial Intelligence.

ACKNOWLEDGMENT

This contribution is the results of the Vega project implementation: Multiagent Network Control System with Automatic Reconfiguration (No.1/0617/08), supported by the Scientific Grant Agency of Slovak Republic.

REFERENCES

[1] D. DOMINGUEZ, VRML and Simulink Interface for the Development of 3-D Simulator for Mobile Robots, *Proceedings of world academy of science, Engineering and Technology*, volume 25, November 2007. ISSN 1307 - 6884

[2] J. FIC, *Riadenie mobilného robota Khepera II s využitím metód umelej inteligencie.* Master Thesis, TU Košice, Faculty of Electrical Engineering and Informatics, Košice, Slovakia, 2009.

[3] M. GAJDUŠEK, F. ŠOLC, Generovaní časově optimální trajektorie pro mobilního robota s diferenciálním řízením. *AT&P Journal*, 2006, No. 2, pp 86-89.

[4] A. JADLOVSKÁ, J. SARNOVSKÝ, Aplikácia inverzného neurónového modelu nelineárneho procesu v štruktúre priameho inverzného riadenie. *AT&P Journal*, 2002, No.10, pp.75-77, No. 11, pp.84-86, No.12, pp. 46, 2002, ISSN 1335-2237

[5] J. KAJAN, Comparison of some neural control structures for nonlinear systéme. *Journal of Cybernetics and Informatics*, 2009, volume 8 2009. ISSN 1336-4774

[6] B. KIM, P. TSIOTRAS, Controllers for Unicycle-Type Wheeled Robots: Theoretical Results and Experimental Validation. *IEEE Transactions on robotics and automation*, 2002, No. 3, pp. 294-307.

[7] F. KÜHNE, W. F. LAGES, J. M. GOMES da SILVA, Mobile robot trajectory tracking using model predictive control. In.: *VII SBAI/II IEEE LARS*, São Luís, september 2005.

[8] K-TEAM. Dostupné na: <http://www.k-team.com >.

[9] I. MASÁR, Inteligentný regulátor na sledovanie trajektórie mobilným robotom. *Automatizace*, 2007, No 2, pp. 80-85.

[10] J. ŠEMBERA, F. ŠOLC, Modelovaní a řízení mobilního robotu s diferenciálním podvozkem. *AT&P Journal PLUS*, 2007, No. 1, pp 203-207.

Evolutionary approach for structural and parametric adaptation of NN for XOR problem

Jaroslav Tuhársky, Peter Smolár, Juraj Eperješi

Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

jaroslav.tuharsky@tuke.sk

***Abstract*— Paper describes experience with selected methods of structural and parametric adaptation of neural networks for real time learning and application in Reinforcement strategy. Non-linear function aproximation is tested and evaluated with these approaches with the aim of perspective application in Computer games and building an intelligence for Bots in the virtual reality. TWEANN and NEAT methods are tested and experimental and theoretical study was accomplished on these methods. They are based on evolutionary computation and optimization of neural networks structure and synaptic weights. Application potential for supporting the intelligence of NAO humanoid robots is mentioned in the conclusion of the paper.**

Keywords— neural networks, evolutionary algorithms, Neuroevolution, TWEANN, NEAT, genetic algorithms.

I. INTRODUCTION

Evolutionary algorithm (EA) is used for solving optimization problems and one of these tasks could be a search for optimal Neural Network (NN) and its topology.

Finding the optimal neural networks by using EA may consist of NN topologies optimization - searching NN topology able to solve the problem and of NN synaptic weights (SW) optimization - search for suitable values of SW. As it is described in [6], neuroevolution (NE), the artificial evolution of neural networks using genetic algorithms (GA), has shown great promise in complex learning tasks.

II. NEAT

The method NeuroEvolution of Augmenting Topologies (NEAT) was created by K. O. Stanley and R. Miikkulainen, from the Texas University in Austin, described in [6]. From the same publication is the following description.

A. Genetic Encoding

NEAT's genetic encoding scheme is designed to allow corresponding genes to be easily lined up when two genomes cross over during matting. Genomes are linear representations of network connectivity.

B. Historical Markings of genes

Whenever a new gene appears through structural mutation, a global innovation number is incremented and assigned to that gene. The innovation number thus represents a chronology of the appearance of every gene in the system. When crossing over, the matching genes in both genomes with the same innovation numbers are lined up, see [6]

C. Protecting Innovations through Speciation

Speciating the population allows organisms to compete primarily within their own niches instead of with the population at large. This way, topological innovations are protected in a new niche where they have time to optimize their structure through competition within the niche. In NEAT is the measure of the compatibility distance of a different structures a simple linear combination of the number of excess E and disjoint D genes, as well as the average weight differences of matching genes W, including disabled genes, see (1) [6].

$$\delta = c_1 \cdot \frac{E}{N} + c_2 \cdot \frac{D}{N} + c_3 \cdot \overline{W} \qquad (1)$$

$\overline{W}$ – average of SW differences of matching genes
E – number of excess genes
D – number of disjoint genes
N – number of genes in larger genome (for normalization because of its size)
c_1, c_2, c_3 – coefficients
δ – compatibility (gene's) distance

III. IMPLEMENTATION OF NEAT APPROACH

For the implementation of experiments we have proposed and implemented the software in the creation of which we was inspired by the method NEAT, see chapter II. We've been using evolutionary calculations, namely the GA to find the simplest topologies with optimized SW for XOR problem.

A. Representing individuals

The population is made up of individuals - NN. Particular individual, which we call the genome, contains a list of "genes of links" [6].

In the initialization of the population are individuals with a minimum NN topology, see Fig.1., whose structure is made up of only input and output layer, i.e. without hidden layers, and with the philosophy that their structure is rising only when it is appropriate for a given solution, see Fig.2. Input layer consists of two input nodes, output layer of one output node. Node "bias" was incorporated into the topology so that we can introduce the entry of the external world to all of the neurons.

For each neuron in the NN, we used the same sigmoid activation function.

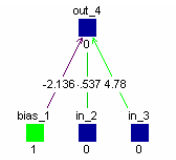


Fig. 1. Initial NN topology – 4-Node NN

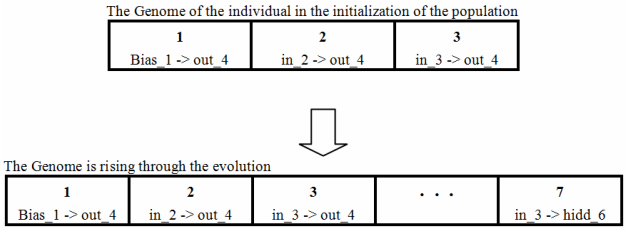


Fig. 2. The Genome is rising through the evolution

IV. EXPERIMENTS

A. *Experiment example*

This experiment shows that in 500 generations, the program NEAT created 4-Node, 5-Node, 6-Node and 7-Node NN. Figure 3 shows the number of individuals pertaining to the topology in the certain generation, as well as the emergence and disappearance of species in the population.

In this case, the 7-node NN was created at the end of the experiment, in the 450. gen. there were only 5 such individuals in the whole population, see Fig.3. Therefore, the program had sufficient time to search SS (State Space) of SW of 6-Node NN and founds its optimization in the 350. gen., i.e. founds the optimal topology and values of NN's SW able to solve XOR task. The 5-Node NN was optimized in the 130.gen. and the search for the 4-Node NN, see Fig.1 (which we know that is not able to solve XOR task) the program stopped in 260.gen., so that entire 4-Node type was thrown away from the population, see Fig.3 and charts of SSE (Sum of Squared Error) during the evolution which are shown in figures Fig.4-Fig.8

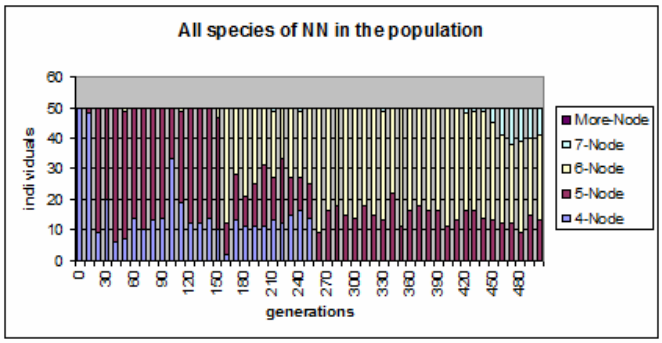


Fig. 3. Number of individuals in the population.

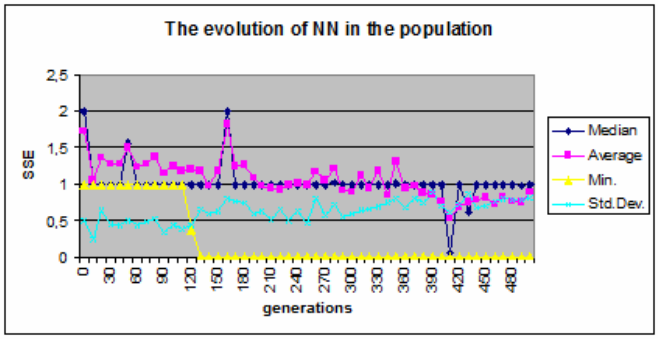


Fig. 4. SSE of NN through evolution in all population.

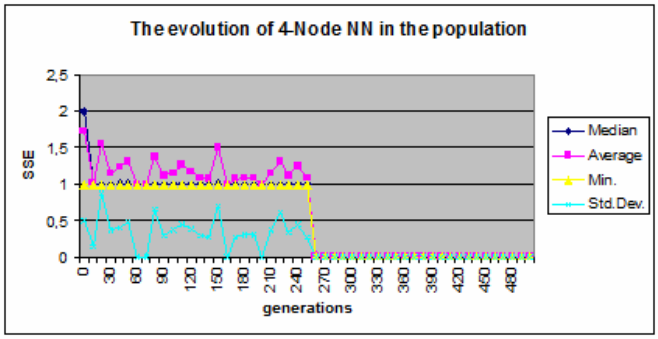


Fig. 5. SSE of the 4-Node NN through the evolution.

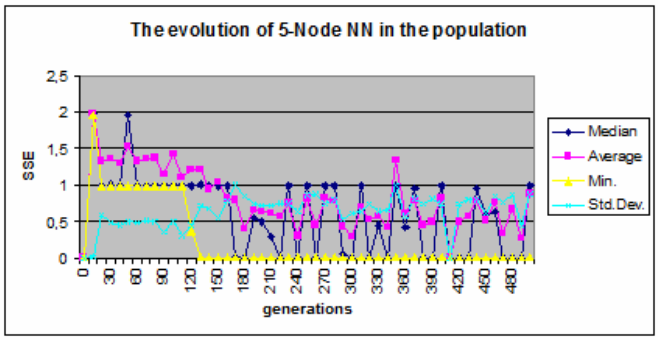


Fig. 6. SSE of the 5-Node NN through the evolution.

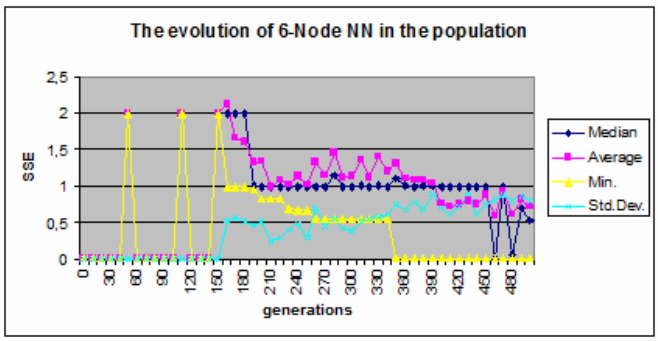


Fig. 7. SSE of the 6-Node NN through the evolution.

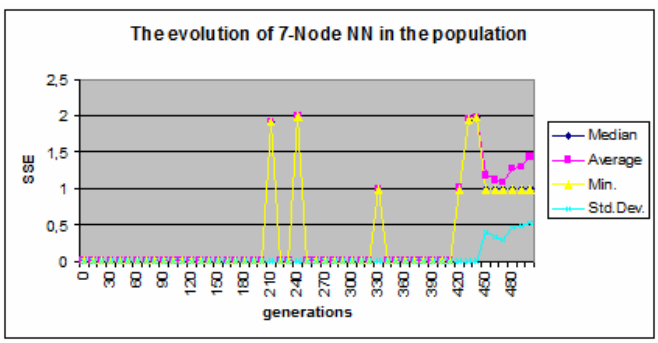


Fig. 8. SSE of the 7-Node NN through the evolution.

The reason why the 7-Node NN was not optimized, is due to lack of time (generations) needed for sufficient scan of NN's SW state space.

In most cases to find the simplest (5-Node) topology of NN able to solve XOR task (where SSE = 0) only 100.gen. were needed, but for NN with more complex structure we need more generations for its optimization.

Figures Fig.9 - Fig.12 shows individuals of particular species which were evolved on the end of the evolution process.

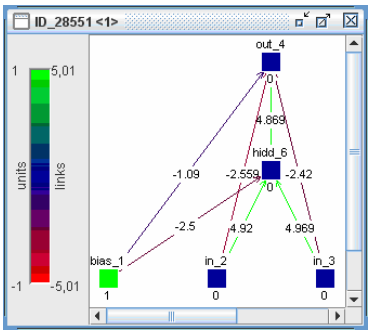


Fig. 9. The individual from species No1

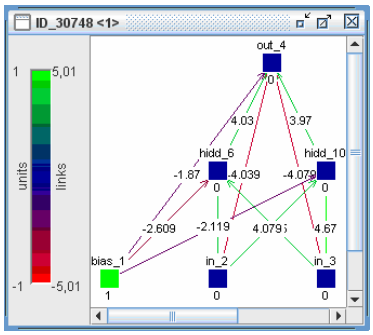


Fig. 10. The individual from species No2

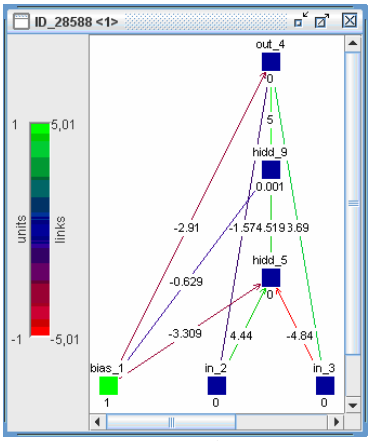


Fig. 11. The individual from species No3

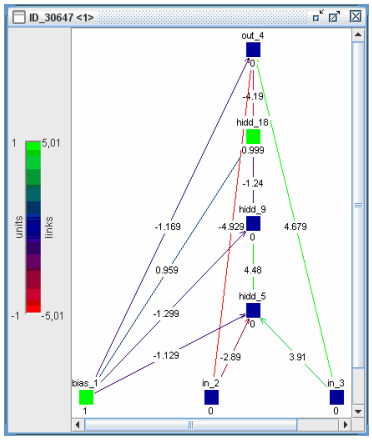


Fig. 12. The individual from species No4

Tab. 1. SSE of the best individual of each species at the end of evolution

INDIV. ID	SSE	INDIV. ID	SSE
ID_28551	3.09e-05	ID_30748	1.19e-08
ID_28588	0.006266	ID_30647	1

V. Conclusion

The functionality of the NEAT method was tested for its ability to evolve various topologies of NN able to deal with XOR problem.

This work has shown strong and weak points of the system TWEANN, as well as the NEAT system, and outlined possible pitfalls, which can be given when using these systems. Tested approach is very effective for the problems of TWEANN which were easily solved.

Acknowledgment

This publication is the result of the project implementation Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] Peter Sinčák, Gabriela Andrejková (1996) Neuron Networks (engineering approach) part 1 and part 2 (in Slovak), Elfa, Kosice.

[2] Marian Mach (2009) Evolutionary algorithms – elements and principles, Elfa, Kosice. ISBN 978-80-8086-123-0.

[3] Vladimír Kvasnička, Jiří Pospíchal, Peter Tiňo (2000) Evolutionary algorithms (in Slovak), STU, Bratislava. ISBN 80-227-1377-5

[4] Vladimír Mařík, Olga Štěpánková, Jiří Lažanský et al. (2001) Artificial Intelligence 3 (in Czech), Academia, Praha. ISBN 80-200-0472-6

[5] Vladimír Mařík, Olga Štěpánková, Jiří Lažanský et al. (2003) Artificial Intelligence 4 (in Czech), Academia, Praha. ISBN 80-200-1044-0

[6] Kenneth O. Stanley, Risto Miikkulainen (2002) Evolving Neural Networks through Augmenting Topologies, The MIT Press Journals.

[7] Kenneth O. Stanley, Risto Miikkulainen(2002) Efficient Evolution of Neural Network Topogies, Proceedings of 2002 Congerss on Evolutionary Coputation. Piscataway, NJ : IEEE.

[8] Kenneth O. Stanley, Boddy D. Bryant, Risto Miikkulainen (2005) Real-Time Neuroevolution in the NERO Video Game, IEEE Transactions on Evolutionary Computation.

[9] Jaroslav Tuhársky, Tomáš Reiff, Peter Sinčák (2009) Evolutionary Approach for Structural and Parametric Adaptation of BP-like Multilayer Neural Networks. Proceedings of the 10th International Symposium of Hungarian Researchers on Computational Intelligence and Informatics, November 12-14, 2009, Budapest, pp. 41-52 ISBN 978-963-7154-96-6

Introduction to Social Networks and Exploitation of Network Data

[1]*Gabriel TUTOKY*, [1]*Ján PARALIČ*

[1]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]{gabriel.tutoky, jan.paralic}@tuke.sk

***Abstract*—This paper presents interaction between knowledge discovery and social networks, and possible exploitation of network data. After brief introduction to knowledge discovery we present Social Networks. We deal with definition of Social Network, and with their representations by graphs and matrices. In second part of this paper we discuss special type of Social Network – Affiliation network and also possible representation of these kind of networks. In the last section we propose approaches for exploitation of data from Social Networks and we present our future work.**

***Keywords*—Social Network, Knowledge Discovery, Data Mining, Affiliation Network, Graphs, Representation of networks.**

I. INTRODUCTION

Social network (SN) is a concept very well known in nowadays, but SNs were introduced and defined many years ago, exactly at the start of previous century [1]. The huge amounts of scientific articles were published with this theme, but many of them were faced with deficient of source data. The rapid growth of SN analysis was facilitated with big popularity of Internet. Generally, we can say that it is almost infinite data source, especially for network data.

Many web portals provide international SNs of unimaginable dimensions, e.g. by [2] in July 2009 the five most visited SN portals were Facebook, MySpace, Blogger, Twitter and WordPres with more than ten millions accesses per month.

Except these gigantic (worldwide) SNs, there are also available middle and small SNs for specific communities. One of such examples is the Slovak portal birds.sk oriented on students and young people who want to express and spread their opinions and reflections [3]. Another example is portal esvety.sk, which wants to invite people of real word communities to build their own social network. [4].

II. DATA MINING

Knowledge discovery (KD)[1] is a process of (semi–) automatic knowledge extraction consists of several steps: *Business understanding*; *Data understanding*; *Data preparation*; *Modeling*; *Evaluation* and *Deployment* [5].

There are many definitions of KD [6], [7], [8] and [9], but the most suitable definition is by [10] and [11] which states: *KD is nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data*, where term pattern goes beyond its traditional sense to include models or structure in data.

A. Data for Knowledge discovery

Traditional data for KD may exist in several forms, e.g. in computer files written by humans, business information in SQL databases or in other standardized database formats, automatically recorded information by machines (logs of devices, binary data streams, etc.). All of these data forms describe identifiable *objects* (usually from real world) and *relations* between them [7].

Each of the examined objects is described by a set of values corresponding to their measurable properties. In KD the set of values assigned to an object are called attributes and usually are recorded as one row or one instance in the table. Thus data set is set of all reachable information usually stored in the table (see Table I.). Each record (row) in the table is one object, in our case one person, which attributes are stored in columns whereas in this particular example the last column "class" has specified significance and serves for classifying into predefined categories [11].

TABLE I.
DATA SET OF TRADITIONAL DATA

NAME	SEX	AGE	NUM. OF FIENDS	CLASS
Robert	man	32	2	social
Joseph	man	30	1	non social
Catherine	woman	26	1	non social
Mary	woman	27	3	social
Thomas	man	29	3	social
Alice	woman	28	3	non social

B. Knowledge discovery Tasks

In [7], there are many forms of knowledge discovery defined, such as *Data Warehousing and OLAP*; Mining *Frequent Patterns, Associations and Correlations*; *Classification and Prediction*; *Cluster Analysis, Mining Stream, Time-Series and Sequence Data*; *Graph Mining, Social Network Analysis and Multirelational Data Mining*; *Mining Object, Spatial, Multimedia, Text and Web Data*.

III. SOCIAL NETWORKS

In the rest of the paper we focus on knowledge discovery in social networks data. "*What is a Social Network?*" One of the "traditional" answers is that a social network consist from a set of *nodes* (or network actors) connected to each other by one or more types of *ties* [12], or by [13]: social network is a set of

[1] Sometimes referred as data mining, although data mining is in fact one particular step in the knowledge discovery process.

socially-relevant nodes connected by one or more *relations*. Nodes, or network members, are the units that are connected by the relations whose patterns we study. These units are most commonly persons or organizations, but in principle any units that can be connected to other units can be studied as nodes. From the point of view of KD, the most appropriate definition is by [7]: social network is a *heterogeneous* and *multirelational* data set represented by a *graph*. The graph is typically very large, with nodes corresponding to *objects* and edges corresponding to *links* representing relationships or inter-actions between objects. Both nodes and links have *attributes*. Objects may have class labels. Links can be one-directional and are not required to be binary.

A. Network Data

Network data are different from traditional data. They consist from representation of one (or more) relation(s) between actors [12]. Usually are stored in tables with the same number of rows and columns. First row and first column of the table are representing actors, whereas other cells of the table are representing relations between them [14] (see Table II.).

TABLE II.
DATA SET OF NETWORK DATA

Actor	*Robert*	*Jozeph*	*Catherine*	*Mary*	*Thomas*	*Alice*
Robert	–	0	0	0	1	1
Joseph	0	–	1	0	0	0
Catherine	0	1	–	1	0	0
Mary	0	0	1	–	1	1
Thomas	1	0	0	1	–	1
Alice	1	0	0	1	1	–

Network data consist of two types of variables: *structural* and *composition*. Structural variables are measured on pair of actors and are the cornerstone of social network data sets. Structural variables measure ties of a specific kind between pairs of actors, e.g. friendships between people, or trade between nations. This kind of data is represented by 0 and 1 in the Table II., where 0 means absence of the tie and 1 means presence of the tie between actors[2] [12].

Composition variables are measurements of actors' attributes. There are standard social and behavioral attributes, and are defined at the level of individual actors, e.g. we might record gender, race, or ethnicity for people, or geographical location, act. [12]. In Table II., composition variables are names of particular actors which should be expanded by data from Table I.

B. Types of Social Networks

Many different types of SNs exist in the real world, and they are not always coming from social context. Examples of them are technologic, business, economic, or biological SNs. We can distinguish SNs by distinct set of entities on which the structural variables are measured to: *one-mode*, *two-mode* and *higher-mode* SNs.

1) One-mode networks

One-mode networks are dominant type of SNs with just a single set of actors, e.g. people, organizations, nation act. The actors themselves can be of a variety of types: subgroups, organizations, or communities. Relations between them that can be studied are: *Individual evaluations*; *Transactions of transfer of material resources*; *Transfer of non-material resources*; *Interactions*; *Movement*; *Formal roles*; or *Kinship*.

2) Two–mode networks

A two-mode network involves measurements on two sets of actors, or on a set of actors and a set of events.

Two Sets of Actors. These networks describe for example companies and its employees, or authors and their articles. Typical analysis of such networks is between actors of one type and actors of second type, because it is not possible to create ties among actors of the same type. Usually in this case, just one type of actors should create tie (sender), and second type of actors should accept tie (receiver) [13].

One Set of Actors and One Set of Events. It is special type of two-mode network, commonly referred as *affiliation* network. It arises when one set of actors is measured with respect to attendance at, or affiliation with, a set of events or activities. Actors (the first mode) are related to each other through their joint affiliation with events or activities (the second mode). The events are often defined on the basis of membership in clubs or voluntary organizations, attendance at social events, sitting on a board of directors, or socializing in a small group [12].

IV. REPRESENTATIONS OF NETWORK DATA

Based on [12], there are three forms of network data representation: *Graph theoretic* – is most useful for centrality and prestige methods, cohesive subgroups ideas, as well as dyadic and triadic methods; *Sociometric* – is often used for the study of structural equivalence and blockmodels; and *Algebraic notation* – is most appropriate for role and positional analyses and relational algebras[3].

A. Simple Graphs

A graph is a model for a SN with an undirected dichotomous relation; that is, a tie is either present or absent between each pair of actors. In a graph, *nodes* represent actors and *lines* represent ties between actors (see Fig. 1 a)).

Graph G is ordered couple (V, E), where V is non-empty set of vertices and E is set of subsets, each one consisting of two elements from set V. Elements of set V are called *vertices* of the graph G and elements from set E are named as *edges* of the graph.

We will use only graphs with finite set of vertices (there exist graphs with infinite set of vertices). A graph G with vertices V and edges E is noted as $G = (V, E)$, the set of vertices in a concrete known graph G is labeled as $V(G)$, accordingly the set of edges is $E(G)$.

Graph is combinatory object that giving the elements of two sets into relationships. Graphs are visualized as projection into the plane, the vertices (nodes) are the points in the plane and edges are expressed as a straight line or spline (connection or link) between the points. This visualization of the graph is also called *diagram of the graph* [15] (see Fig. 1).

Subgraph. Graph G' is a subgraph or factor of graph G if set of vertices $V(G')$ are subset of vertices $V(G)$ and set of edges $E(G')$ is subset of edges $E(G)$, so $V(G') \subset V(G)$ and $E(G') \subset E(G)$, we write $G' \subset G$ (see Fig. 1 b)).

[2] If we assume that is not possible to create cyclic ties of one actor to him/herself.

[3] For more details see [12], section 3.

Complete graph at $n \geq 1$ vertices is graph $C_n = (V, E)$, where $|V| = n$ and E includes all possible two element subsets of vertices.

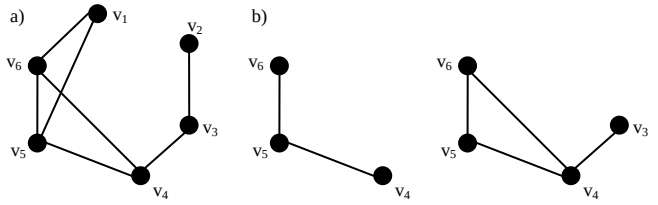


Fig. 1. a) Diagram of the simple graph; b) Diagrams of subgraphs.

B. Directional and Valued Graphs

Many relations are *directional* in SNs. A relation is directional if the ties are oriented from one actor to another. The import/export of goods between nations is an example of a directional relation. A directional relation can be represented by a *directed graph* $\vec{G}$, or *digraph* for short. A digraph consists of a set of nodes representing the actors in a network, and a set of arcs directed between pairs of nodes representing directed ties between actors. The difference between a graph and a directed graph is that in a directed graph the direction of the lines is specified (see Fig. 2) [12].

Often SN data consist of valued relations in which the strength or intensity of each tie is recorded. Examples of valued relations include the frequency of interaction among pairs of people, or rating of friendship between people in a group. Thus, next step in the generalization of graphs and digraphs is to add a *value* or *magnitude* to each line or arc (see Fig. 2). Valued graphs are the appropriate graph theoretic representation for valued relations [12].

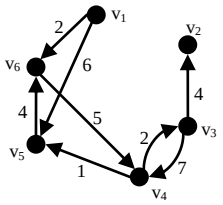


Fig. 2. Diagram of the valued directional graph

C. Matrices

The information in a graph G may also be expressed in a variety of ways in matrix form. There are two such matrices that are especially useful. The first is the *sociomatrix* (discussed below), and the second is *incidence* matrix [16].

A sociomatrix is the primary matrix used in SN analysis, also called as *adjacency* matrix [12]. This matrix indicates whether two nodes are adjacent or not. For one-mode networks is sociomatrix of size $g \times g$ (g rows and g columns), and there is a row and column for each node, and the rows and columns are labeled 1, 2, …, g. The entries in the sociomatrix, x_{ij}, record which pair of nodes are adjacent. There is a 1 in the $(i,j)^{\text{th}}$ cell (row i, column j) if there is a line between n_i and n_j, and a 0 in the cell otherwise (see Table II).

More formally, sociomatrix of graph $G = (V, E)$ or digraph $\vec{G} = (V, E)$ with vertex set $V = \{v_1, v_2, \ldots, v_n\}$ is a square matrix $B = (b_{ij})$ of order n, and its elements are equal [15]:

$$b_{ij} = \begin{cases} 1, & \text{if } (v_i, v_j) \in E \\ 0, & \text{otherwise.} \end{cases}$$

V. Affiliation Networks

Affiliation network (AN) differ in several ways from the types of SN [12]. First, ANs are two-mode networks, consisting of a set of actors and a set of events. Second, ANs describe collections of actors rather than ties between pairs of actors.

A. Properties of Affiliation networks

Most importantly, since ANs are two-mode networks, we need to be clear about both of the modes. As usual, we have a set of actors $N = \{n_1, n_2, \ldots, n_g\}$, as the first of two-modes. In SNs we also have a second mode, the *events*, which we denote by $M = \{m_1, m_2, \ldots, m_h\}$. The event in an AN can be a wide range of specific kinds of social occasions; e.g. social clubs in a community, treaty organizations for countries, and so on.

Another important property of ANs is the *duality* in the relationship between the actors and the events. However, the duality in ANs refers specifically to the alternative, and equally important, perspectives by which actors are linked to one another by their affiliation with events, and at the same time events are linked by the actors who are their members.

Duality of an AN means that we can study the ties between the actors or the ties between the events, or both. Focusing on events, two events have a pair-wise tie if one or more actors are affiliated with both events, this we will refer as *overlapping* events. When we focus on ties between actors, we will refer to the relation between actors as one of *co-membership* [12].

B. Representing Affiliation networks

ANs are in the graph theoretic representation represented by *bipartite graph* (see Fig. 3). A bipartite graph is a graph in which the nodes can be partitioned into two subsets, and all lines are between pairs of nodes belonging to different subsets. Thus, each mode of the network constitutes a separate node set in the bipartite graph. Since there are g actors and h events, there are $g + h$ nodes in the bipartite graph.

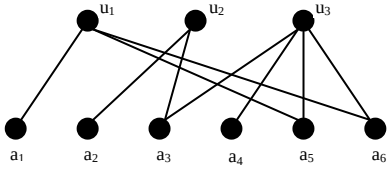


Fig. 3. Diagram of bipartite graph of an Affiliation network

Formally, c*omplete bipartite graph* $C_{m,n} = (V, E)$, where $m, n \geq 1$, is graph, in which $V = \{n_1, \ldots, n_g\} \cup \{m_1, \ldots, m_h\}$; $E = \{n_i, m_j\} : i = 1, 2, \ldots, g; j = 1, 2, \ldots, h$ [15].

In the Sociometris, ANs are represented by matrix that records the affiliation of each actor with each event. This matrix, which we will call an *affiliation matrix*, $\mathbf{A} = \{a_{ij}\}$, codes for each actor, the events with which the actor is affiliated. Equivalently, it records for each event, the actors affiliated with it. The matrix **A**, is a two-mode sociomatrix in which rows index actors and columns index events. Since there are g actors and h events, **A** is a $g \times h$ matrix, where $(i,j)^{\text{th}}$ cell is equal:

$$a_{ij} = \begin{cases} 1, & \text{if actor } i \text{ is affiliated with event } j \\ 0, & \text{otherwise} \end{cases}$$

Actor	*Event 1*	*Event 2*	*Event 3*
Robert	1	0	1
Joseph	0	1	0
Catherine	0	1	1
Mary	0	0	1
Thomas	1	1	1
Alice	1	1	0

Fig. 4. Example of an Affiliation matrix

VI. Proposal of Social Networks Exploitation

A. Social networks of small communities

Small real communities and their networks can have several targets. One of these targets is grouping of people and creating new relationships between them. The relations depend on real world activities, but correct representation of these relationships is crucial in analysis of these, small communities networks.

SN of small communities has several advantages in contrast of gigantic networks such as Facebook, LinkedIn, or others. We can analyze small networks in whole and also in parts, because there are usually still sufficient counts of members. Other, very important advantage of small networks is that they can be analyzed by visualization tools. This analyzing technique is of course very sensitive to the number of network members.

B. Network Data exploitation

Extracted knowledge from the network data can be used in many ways. Some of them can be used for business targets, increasing of working process effectiveness, or for customizing and forming of social network by itself.

In case of small community, the extracted knowledge should be very useful for achieving of community targets. For example, small community target is to group people who are not previously known each other in a group during some action or activity. Thus it is possible to make new friendships between group participants.

Representation of relationships by SNs is very useful for future partition of people into groups. Except data of past affiliation in the groups we can store an additional data in SN. We can customize ties between people by their intensity of communication or by other relations (not exactly) added by humans, like grouping in their own (virtual) groups or their discussions in forums.

VII. Conclusion

In section II. we briefly presented Knowledge discovery, and its relation to Social Networks which were presented in all other parts of this paper. Section III. dealt with definitions and basic principles of Social Networks. Finally, in section IV. we presented some possible representation of networks. Affiliation networks as special type of Social Networks were presented in section V., and in the following, section VI., we made a proposal for exploitation of data from Social Networks.

Next future work will be oriented to data preprocessing of small community network, their analysis and visualization. After this, we will start experiments with modeling of network and analysis of variety modeling methods with reflection to its expression power.

Finally we want to use data from SN for specific tasks, e.g. for partitioning of people into groups during actions and compare it with classical methods (methods without reflection of past data), or for creation of new friendships automatically by providing possibilities for communication and interaction between people.

VIII. Acknowledgements

This work was supported by the Slovak Grant Agency of Ministry of Education and Academy of Science of the Slovak Republic under grant No. 1/0042/10 and is also a the result of the project implementation Development of Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120030) supported by the Research & Development Operational Programme funded by the ERDF..

References

[1] Wikipedia. *Social Network*. [Online] Wikipedia Foundation, Inc., Social Network. http://en.wikipedia.org/wiki/Social_network.

[2] Nielsen Company. 2009. *Social Media Stats: Myspace Music Growing, Twitter's Big Move*. Nielsenwire. [Online] Nielsen Company, 17. júl 2009. http://blog.nielsen.com/nielsenwire/online_mobile/social-media-stats-myspace-music-growing-twitters-big-move/.

[3] Birdz. *Birdz.sk*. [Online] Birdz. http://www.birdz.sk/.

[4] Jarošová Gabriela. 2009. *Sociálne siete po slovensky*. FWD. [Online] RFD.sk, June 4, 2009. http://fwd.etrend.sk/vsetko/socialne-siete-po-slovensky.html.

[5] PARALIČ Ján. 2003. Objavovanie znalostí v databázach. Košice : Elfa, 2003. http://people.tuke.sk/jan.paralic/prezentacie/OZ/ObjavovanieZnalostivDB.pdf. ISBN 80-89066-60-7.

[6] Data-Mining Concepts. *Data-Mining Concepts*. http://media.wiley.com/product_data/excerpt/24/04712285/0471228524-1.pdf.

[7] Han Jiawei, Kamber Micheline. 2006. *Data Mining: Concepts and Techniques*. Second Edition. San Francisco : Morgan Kaufmann Publishers, 2006. ISBN 978-1-55860-901-3.

[8] Larose Daniel. 2006. *Data Mining: Methods and Models*. New Jersey : John Wiley & Sons, Inc., 2006. ISBN 978-0-471-66656-1.

[9] Two Crows Corporation. 2005. *Introduction to Data Mining and Knowledge Discovery*. Third Edition. U.S.A : Two Crows, 2005. ISBN 1-892095-02-5.

[10] Fayyad Usama, Piatetsky-Shapiro Georgy, Smyth Padhraic. 1996. *The KDD Process for Extracting Useful Knowledge from Volumes of Data*. ACM, 1996. p. 27-34. http://portal.acm.org/citation.cfm?id=240464.

[11] Bramer Max. 2007. *Principles of Data Mining*. London : Springer, 2007. ISBN 978-1-84628-765-7.

[12] Wasserman Stanly, Faust Katherine. 1994. *Social Network Analysis*. Cambridge : Cambridge University Press, 1994. ISBN 978-0-521-38707-1

[13] Marin Alexandra, Wellman Barry. 2009. *Social Network Analysis: An Introduction*. London : Forthcoming in Handbook of Social Network Analysis, 2009. http://www.chass.utoronto.ca/~wellman/publications/newbies/newbies.pdf.

[14] Hanneman Robert, Riddle Mark. 2005. *Introduction to Social Network Methods*. Riverside : University of California, 2005. http://www.faculty.ucr.edu/~hanneman/nettext/.

[15] Klešč Marián. 2006. *Diskrétna matematika*. Košice : Technická univerzita v Košiciach, 2006. ISBN 80-8073-698-7.

[16] Borgatti Stephen. 2004. *Introduction to Graph Theory*. 2004. http://www.steveborgatti.com/papers/graphtheory.doc.

Controlling of drug targeting by magnetic field

[1]*L. Vaľová*

[1]Dept. Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1]lucia.jancurova@cern.ch

***Abstract*—A special focused magnet, designed for the use in the magnetic targeted drug delivery system, was constructed. From the theoretical calculation of the adhesion condition for a magnetic fluid drop in magnetic field with obtained design showed, that the constructed focused magnet generates a sufficient magnetic force for the capture of a magnetic drop on the vessel wall and can be used 2.5 - 3 cm deeper in organism comparing with prism permanent magnet what could enable to the non-invasivity of the magnetic drug targeting procedure. The maximal values for magnetic field and gradient of magnetic field are 0.38 T and 101 T/m.**

***Keywords*—magnetic drug targeting, magnetic fluid, focused magnet**

I.INTRODUCTION

The standard treatments of cancer is the surgical removal of tumor and subsequent chemotherapy, irradiation or other methods. The choice of therapy depends on the location and size of tumor and stage of disease.

Oncology doctors seduce their patients to chemotherapy by convincing them it is a great success and the hope for cure is high. Science, however, says something quite different. Results of the 14 year research revealed shocking findings, indicating the overall benefit of the chemotherapy for 5 year survival rate of adults with cancer is 2.3% in Australia and 2.1% in the USA.

One way to improve the methodologies such as chemotherapy, is the targeted chemotherapy and the targeted transportation.
Magnetic fluids, thanks to their properties, are widely used in biomedicine and biotechnology. They can be easily controlled by external magnetic field, transported to the required location in the body and localised where needed . These features are used in the targeted transport of drugs.
The particles are maintained in the desired position by magnetic force arising from the fact that the particle, having a magnetic moment m in an inhomogeneous magnetic field of induction *B*, is driven by force:

$$\vec{F} = (\vec{m} . \nabla)\vec{B} \qquad (1)$$

Relatively simple calculations show, that the maintaining the position of the magnetic particle requires the speed of blood to be kept below 100 mm/s in arteries and 0,5-1 mm/s in capillaries. The magnetic field must be such that the product of its gradient and flux is maximal [1].

A local therapy achieved through the targeted transport may increase the efficacy of a medicament by its concentration in the tumor site. The advantage of this method is smaller toxic effect on healthy cells, an application of lower doses of drugs, because they reach the required destination with much greater efficiency [2].

Current technologies of the targeted transport allow to locate more than 70% of the dose in the target tissue with a minimal interaction and toxicity to normal cells and to 8 times increased concentration of the drug in the tumor when only 1/3 of normal chemotherapeutic dose is administered.

Current research on methods to target chemotherapy drugs in the human body includes the investigation of biocompatible magnetic nano-carrier systems, e.g., magnetic liquids such as ferrofluids [3]. The use of biocompatible magnetic fluid as potential drug carrier appears to be a promising technique. Due to their superparamagnetic properties the magnetic fluid drops can be precisely transported, positioned and controlled in desirable parts of blood vessels or hollow organs with the help of an external magnetic field [4]. The motion of magnetic drop within the body is controlled by the combination of magnetic force and a hemodynamic drag force due to blood flow. The models which investigate the interaction of an external magnetic field with blood flow containing a magnetic carrier substance are based on the Maxwell and Navier-Stokes equations, where a static magnetic field is coupled to fluid flow. This is achieved by adding a magnetic volume force to the Navier Stokes equations, which stems from the solution of magnetic field problem [5]. In order to effectively overcome the influence of blood flow the magnetic force must be larger than the drag force. The conditions for holding a magnetic fluid drop on a blood vessel wall were investigated by Voltairas et al.[6]. In this work the non-uniformity of considered magnetic field as higher only close to the magnetic pole, what was regarded as a major technical problem that has to be resolved in order for the drug targeting to remain essentially non-invasive. The aim of our work was to construct a focused magnet, which enables to achieve maximal magnetic force in deeper position, to map its magnetic field and to find the adhesion condition for a magnetic fluid drop in magnetic field with obtained design.

II.RESULTS

All the achievements required a time-consuming mathematical and physical calculations lasting several weeks when exexutedon a current computer. Therefore, we decided to speed up the calculations significantly by using local PC clusters, which mostly consist of dozens of computers or by using Grid involving several hundred computers. These solutions allowed to reduce several days calculations down to

several hours. Specifically, we have built a PC cluster consisting of 4 PCs (12 cores) with the PROOF software.
The PC cluster together, with prepared software environment, has been used for numerical solution of physical tasks formulated in the objectives of the thesis.
A magnetic field of classical rod magnets, when applied by a magnet into deeper positions in the body, does not supply a geometry of magnetic field and its gradient sufficiently, so we used the concept of a focusing magnetic field in order to achieve better results in a marking of drugs magnetically.

Following the above mentioned considerations and taking into account the technological simplicity, we proposed the construction of compound focused magnet. Its cross section is schematically drawn in Figure 1

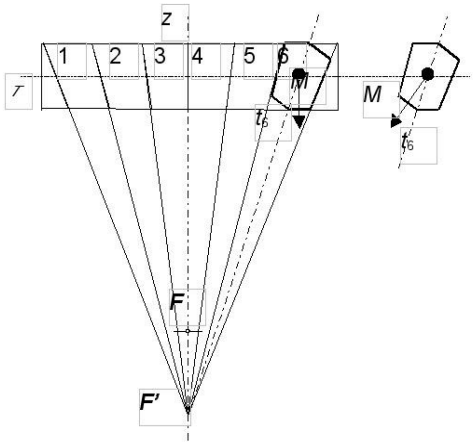


Figure 1: The crossed section of the focused magnet

An idea of focused magnetic field was exploited for the calculation of magnetic field maps using GRID.
An outcome is illustrated in Figure 2.

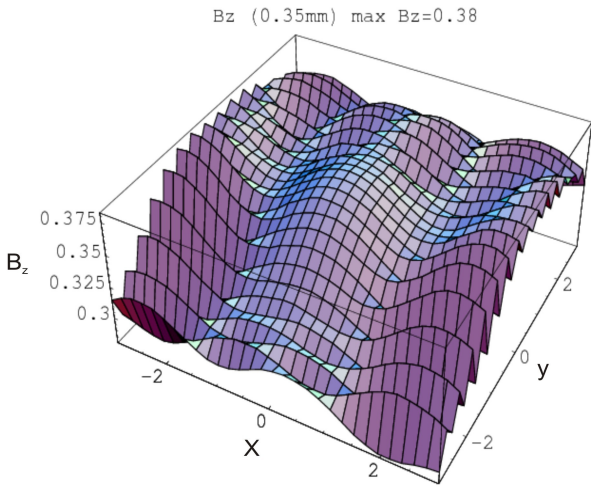


Figure 2: The profile of magnetic field generated by focused magnet

Magnetic field generated by focused magnet was measured using 3D Hall probes. The maximum value of the magnetic field near the surface of the magnet (0:35 mm) was estimated to 0:38 T.

To simulate the movement of drug particles in the blood circulatory system under the influence of magnetic field, produced by focused magnet, we need to describe a value of a vector magnetic field in any point.
An interpolation function of magnetic field of focused magnet, which was measured using the same point probes, was interpolated by a 10-degrees polynomial with a shape given in [7].

To find the coefficients of the interpolation polynomial for each component B_x *(x, y, z)* <-3, 3>, B_y *(x, y, z)* <-3, 3>, B_z *(x, y, z)* <-0, 35; 4.85> on the GRID with step h_x= 0,2, h_y=0,2 and h_z=0,5 a ROOT program was written. The program starts counting from the 3-degrees polynomial and adds one degree up to 10 degrees to the magnetic field. The components B_x, B_y, B_z at y = 0:0, are shown.

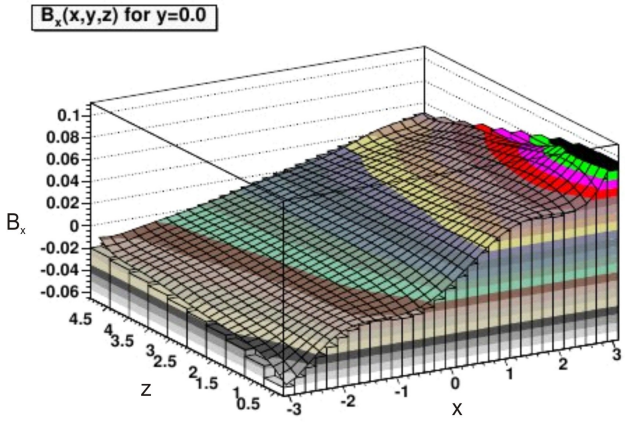


Figure 3: The x-component of magnetic field

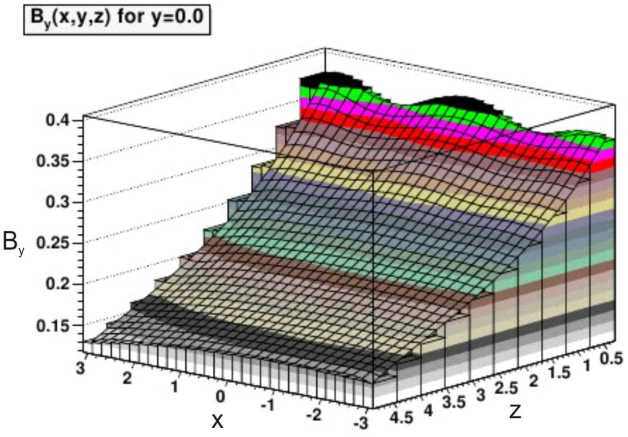


Figure 4: The y-component of magnetic field

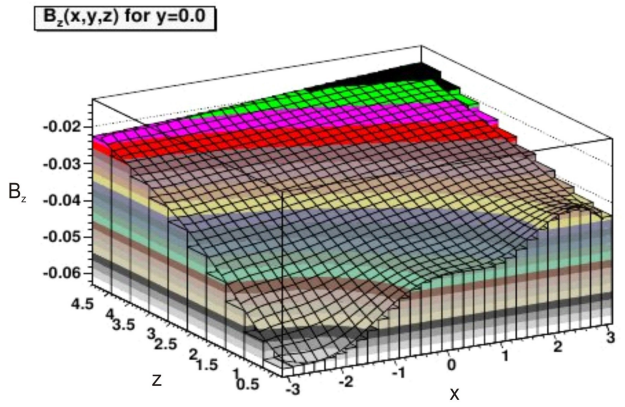


Figure 5: The z-component of magnetic field

III. CONCLUSION

In summary, we have mapped the magnetic field and used its profile for numerical calculations determining the upper limit of the average speed of the blood flow, in which the magnetic field is able to capture the drug bound to the magnetic beads on the walls of blood vessels.

	Femoral artery	Carotid artery
B_0 [T]	0.195	0.234
u_0 [m/s]	0.462	0.841
u_{exp} [m/s]	0.05-0.35	0.1-0.6
F_m [kN/m^3]	5.992	105.025
dB/dx [T/m]	7.747	107.824
M [mT]	0.975	1.170

Table 1: The model – comparison with experiment.

Thus, the results obtained (Table 1) allow to develop a greater holding force at a greater distance than conventional magnets i.e. they proved an ability of the magnet to generate a sufficient magnetic field inside the body (2 to 3 cm, deeper than conventional magnets), thereby contributing to the non-invasive drug transport.

B_0 [T] the magnetic field at the point of capture,
u_0 [m/s] speed value calculated using simulations,
u_{exp} [m/s] the experimental value,
F_m [kN/m^3] the volumetric magnetic force required to maintain the magnetic drug,
dB/dx [T/m] the gradient magnetic field in a place of capture,
M [mT] the magnetization of used magnetic fluid.

ACKNOWLEDGMENT

This publication are the results of the project implementation: Centre of Information and Communication Technologies for Knowledge Systems (Project number:26220120020) suported by the Researcher & Development Operational Programme Funded by the ERDF, FEI TUKE Grant VEGA 1/0617/08 and JINR Dubna protocol No: 3920-6-09/10

REFERENCES

[1] T. Neuberger, B. Schöpf, H. Hofmann, M. Hofmann, B. Rechenberg, J.Magn.Magn.Mater.293 (2005) 483.
[2] A. A. Kuznetsov, V.I. Filipov, O.A. Kuznetsov, V.G. Gerlivanov, E.K. Dobrinsky and S.I Malashin, J.Magn.Magn.Mater.194 (1999) 22.
[3] E.K. Ruuge, A.N. Rusetski, J.Magn.Magn.Mater.122 (1993) 335.
[4] U.O. Häfeli, J.G. Pauer, J.Magn.Magn.Mater.194 (1999) 76.
[5] R. Ganguly, A.P. Gaind, S. Sen, I.K. Puri, J.Magn.Magn.Mater.289 (2005) 331.
[6] P.A. Voltairas, D.I. Fotiadis and L.K. Michalis, J.Biomech. 35 (2002) 813.. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.
[7] L. Jancurova, J. Jadlovský, J. Chovanak, P. Kopcansky et al., The Concept of Focused Magnet for Targeted Drug Delivery, Preprint (2009) E11-2009-174

Linear feature transformations in speech processing

Peter VISZLAY

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

peter.viszlay@tuke.sk

***Abstract*—The paper describes two linear transformations used in speech preprocessing in automatic speech recognition systems. These transformations are applied in feature vector extraction of speech signal with the purpose to increase the recognition accuracy. The first described method is Linear discriminant analysis (LDA). It provides the classes separability with maximalization of the between-class scatter and the within-class scatter ratio. The second method is called Principal component analysis (PCA). It represents the data along their largest variance by computed principal components. These two dimension reduction and decorrelation techniques result in higher performance of the recognition system. For both methods is given a mathematical description and their comparison.**

***Keywords*—class, classificator, dimension, discriminant analysis, feature extraction, feature vector, linear transformation, principal component, scatter matrix**

I. Introduction

The human - computer interaction by speech is in the present a modern form of human - machine communication. One of the crucial processes required for such voice based communication is automatic speech recognition (ASR), which converts the speech signal into a text form. This process consists of two phases - feature extraction and classification.

The *feature extraction* is a selection process of acoustic vectors. Various methods of acoustic analysis on the signal samples at the obtaining of the feature vectors are applied. An ideal feature vector should contain information enabling the unambiguous classification into individual phonetic classes considering the dimension of the vector. Therefore, to achieve the best recognition results the appropriate selection of parameters to be included in the final acoustic vectors is very important [1]. The obtained acoustic vectors contain only the relevant, nonredundant information. If the parameters of the features are not properly extracted, the succes of the recognition is limited [2]. The feature extraction methods provide a new set of parameters, which is a linear combination of the original set of parameters. These methods can be also described as *linear feature transformations*.

The *classification* is a process of merge of the recognized pattern (feature vector) into predefined classes. Most practical recognition systems are using classificators based on statistical approaches for modeling the acoustic signal (e.g. Hidden Markov Models (HMM)).

The rest of paper is organized as follows. The next chapter presents a brief overview of the linear transformations used in speech processing. The third and fourth part attempt to describe the fundamentals of the mentioned linear transformations and present the mechanics of these processes. The fifth section highlights the main differences between LDA and PCA. And finally, the paper gives a short conclusion in section six.

II. Linear feature transformations and the dimensionality

According to [1], the methods for parameters extraction are:

- Principal Component Analysis - PCA;
- Linear Discriminant Analysis - LDA.

PCA seeks such vectors, which best describe the data set in terms of the largest variance, whereas LDA seeks vectors, which provide the best discrimination between classes [1] [3]. Specifically, in speech preprocessing, the acoustically similar sounds should be represented by feature vectors, which are geometrically close to each other (they are creating clusters of similar images). Generally, these clusters correspond to classes of sounds that are tied to each phoneme. Methods of selection and ordering the features with respect to this assumption are trying to seek a transformation reducing the dimension of space and ensuring sufficient distance between the phonetic classes at small distances of images within the classes. The linear transformations can be implemented by matrix multiplication [4].

A generally block diagram for recognition system using LDA or PCA is illustrated in the Fig. 1. The speech signal is at first processed (digitalization, emphasizing, and windowing is applied) to obtain the MFCC (Mel Frequency Cepstral Coefficients) features. Then, a dimension reduction method is applied on the MFCCs. The products of the analysis (LDA or PCA) are suitable for classification. The classificator can be based on HMM or Neural Network. The preprocessing and the linear transformation block form a so-called ASR front-end.

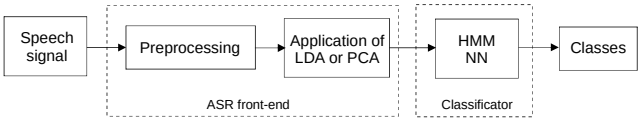


Fig. 1. The basic concept of speech recognition system using LDA or PCA.

A. The curse of dimensionality

Computational efficiency is an important problem for real-time continuous speech recognition systems. In practice, the amount of computations required for pattern recognition and the amount of data required for training systems grows exponentially with the increase of the dimensionality of the feature vectors so the system performance exponentially degrades [4]. This fact applies even to powerful computers. In practice, this effect is referred to as *curse of dimensionality*. The relation between the dimensionality and the performance models the graph in Fig. 2. As we can see from it, after exceeding a

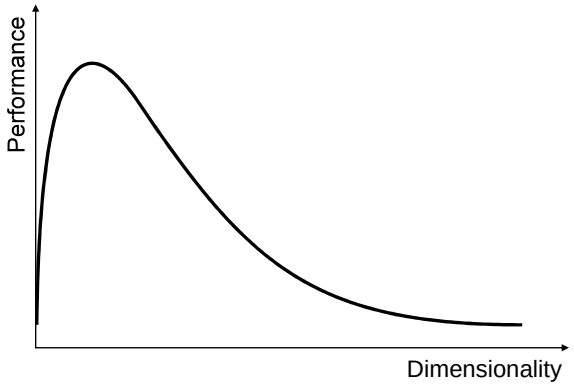


Fig. 2. Influence of the dimensionality to the performance.

certain dimension threshold, adding a further vectors causes performance decreasing. Optimal performance of the recognition system can be achieved when an appripriate dimension of vector is choosen.

B. The dimension reduction

Reducing the dimensionality of parameter vectors is the most direct way to solve the problems caused by high dimensionality. However, directly reducing the number of parameters will cause unpredictable information loss and consequently make the system performance unstable. We can overcome this problem using the mentioned linear transformations [4]. If we have an n-dimensional input vector $x = [a_1 a_2 \dots a_N]^T$ then the output vector $y = [b_1 b_2 \dots b_K]^T, K \ll N$, after the transformation has reduced dimension K. The basic idea of the dimension reduction can be written in mathematical (matrix) form as $Y = UX$, or $b_i = u_i^T a_i$, where $Y : k \times 1; U : k \times d; X : d \times 1$ and $k \ll d$.

III. Linear discriminant analysis

For succesfully classification and recognition of the vectors it is necessary to accomplish the requirements for their sufficient discrimination. These requirements may be executed by application of linear transformation such as LDA, which is applied on defined sequence of acoustic vectors. It extracts from these sequences a set of discriminant parameters maximalizing the classes separability [5].

Linear discriminant analysis is an optimal linear transformation that maps the input data set from n-dimensional space onto m-dimensional space, $m < n$ [5]. The LDA minimalizes the within-class distances and at the same time maximalizes the between-class distances such that the maximum class discrimination is achieved [6].

A. Different approaches to LDA

There are two approaches to data transformation [7]:

1) *Class-dependent (CD) transformation*. This approach involves maximizing the ratio of between class variance to within class variance (adequate class separability is obtained). CD-LDA uses two optimizing criteria for transforming the data sets independently.
2) *Class-independent (CI) transformation*. This approach involves maximizing the ratio of overall variance to within class variance. CI-LDA uses only one optimizing criterion. Each class is considered as a separate class against all other classes.

B. Mathematical description

Assume that x is a n-dimensional feature vector. The LDA seeks to find a linear transformation $\mathbf{R}_n \rightarrow \mathbf{R}_p, (p < n)$, in such form [8]:

$$y_p = W_p^T x, \tag{1}$$

where W_p is a $n \times p$ transformation matrix and y_p is the new transformed vector. Assume that W is a non-singular $n \times n$ square matrix, which is used to define a linear transformation $y = W^T x$. After its appropriate partition is obtained the following form:

$$W = [W_p W_{n-p}] = [\vec{W}_1 \dots \vec{W}_n], \tag{2}$$

where W_p consists of the first p columns of W and W_{n-p} consists of the remaining $n - p$ columns and $\vec{W}_i$ is the ith column of W. Then, feature dimension reduction can be viewed as a two step procedure. First a non-singular linear transformation is applied to x to obtain y. In the second step, only the first p rows of y are retained to give y_p.

Suppose now that the input data are partitioned into k pattern classes (k class problem) $\Pi = \{\pi_1 \dots \pi_k\}$. Let the ith observation vector from the class π_j be x_{ji}, where $j = 1, \dots, k$ and $i = 1, \dots, N_j$. N_j is the number of observations from class j. Then, the *within-class scatter matrix* (class-independent), *within-class scatter matrix* (class-dependent) for jth class and *between-class scatter matrix* are defined respectively [4]:

$$S_{W,CI} = \sum_{j=1}^{k} \sum_{i=1}^{N_j} (x_{ji} - \bar{x}_j)(x_{ji} - \bar{x}_j)^T, \tag{3}$$

$$S_{W,CD}^j = \sum_{i=1}^{N_j} (x_{ji} - \bar{x}_j)(x_{ji} - \bar{x}_j)^T \text{ and} \tag{4}$$

$$S_B = \sum_{j=1}^{k} N_j (\bar{x}_j - \bar{x})(\bar{x}_j - \bar{x})^T, \tag{5}$$

where $\bar{x}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} x_{ji}$ is the mean of class j, $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$ is the global mean vector and $N = \sum_{j=1}^{k} N_j$ is the number of all observations.

The corresponding transformed *within-class* and *between-class* scatter matrices are $\tilde{S}_W$ and $\tilde{S}_B$. It can be shown that

$$\tilde{S}_W = W_p^T S_W W_p, \text{ and } \tilde{S}_B = W_p^T S_B W_p. \tag{6}$$

Finally, the transformation matrices are computed by maximizing so-called Fisher's criteria [9]:

$$J_{CI}(W_p) = \frac{|W_{p,CI}^T S_B W_{p,CI}|}{|W_{p,CI}^T S_{W,CI} W_{p,CI}|}, \text{ or} \tag{7}$$

$$J_{CD}(W_{p,j}) = \frac{|W_{p,j,CD}^T S_B W_{p,j,CD}|}{|W_{p,j,CD}^T S_{W,CD}^i W_{p,j,CD}|}. \tag{8}$$

The solution is that ith column of an optimal W_p is the generalized eigenvector corresponding to the ith largest eigenvalue of matrix $S_W^{-1} S_b$. The final LDA transformed feature vectors represent the energy of a speech signal distributed along the eigenvector-spanned-coordinates on which the classes have the largest discriminants [4]. The application of LDA for two class problem is illustrated in Fig. 3.

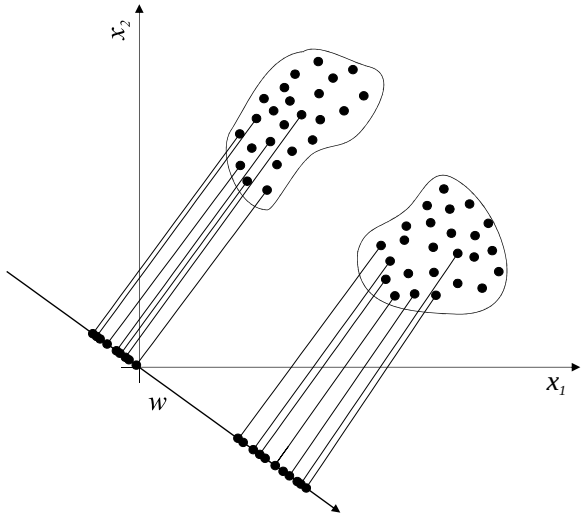


Fig. 3. Classes separation and dimension reduction by applying the LDA transformation.

C. LDA and automatic speech recognition (ASR)

As we mentioned in section II, each feature vector should be assigned to a class that are tied to each phoneme. This process can be described as follows [8]. Firstly, we estimate the HMM parameters using the n-dimensional training data and applying the Baum-Welch algorithm. We use these parameters to determine the most likely sequence of HMM states for the training data. Each state is treated as a different class and each feature is assigned to a class corresponding to the state it came from. Then, we can perform LDA (class labels are already available) to reduce the feature dimension. Finally, we estimate the HMM parameters for the new transformed and reduced features. To evaluate the performance on the test data the transformed features and the corresponding HMM parameters are needed.

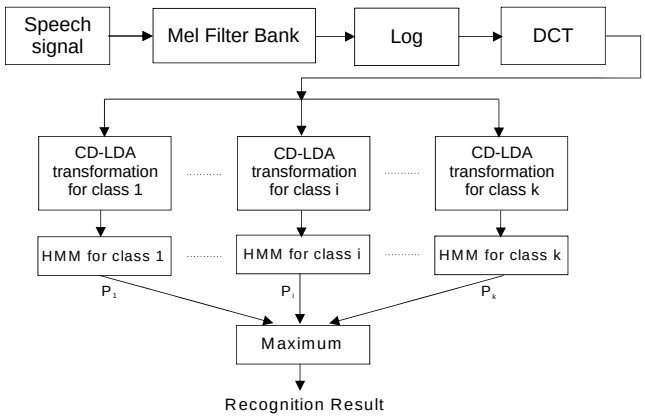


Fig. 4. Concept of ASR system using class dependent LDA.

In feature extraction, the choice of feature vector dimension is one of the important but not easy tasks. There are no formal method to do this. The best way is probably an experiment and comparison with existing features. Often are also used the cross validation way. For example, the MFCCs are typically 13-dimensional. If the dynamic coefficients are also included to them then are treated 39-dimensional feature vectors. Since speech recognizers operating with such a vectors constellation show very good performance, it is estimated that 39 dimensions is a suitable choice for the given recognizer [10]. A typically block diagram for ASR system using CD-LDA with HMM is depicted in Fig. 4.

D. Limitations of LDA

The implicit assumption of Gaussian distributions for the input data set is one of the general limitations of LDA. The objective function of LDA requires the nonsingularity at least one of the scatter matrices [11]. In some applications the overfitting of the data can be the outcome of LDA. The function of LDA can be also limited when the classes are not linearly separable.

E. Extensions of LDA

There are problem areas, where the classical LDA can fail. For this reason the following extensions of LDA are used to achieve better results: Heteroscedastic Linear Discriminant Analysis (HLDA), Smoothed Heteroscedastic Linear Discriminant Analysis (SHLDA), Nonlinear Discriminant Analysis (NLDA), Kernel Linear Discriminant Analysis (KLDA), etc.

IV. Principal component analysis

Principal component alanysis is a linear orthogonal transformation that maps the original data set to the new coordinate system (Fig. 5). It provides an reduction of a large number of input variables (it is a dimension reduction technique) with undesirable redundancy (some variables are correlated because they are describing the same structure). The PCA is a linear nonparametric unsupervised method that also acquires relevant information from ambiguous data sets [4] [12].

The PCA performs a computation of a set of mutually independent output variables, called principal components, which are a linear combination of the original data. The first principal component represents the largest part of the variability, the second principal component represents the second largest part of variability in order, etc., until the whole variability of the data is described. Usually, the first few components represent about 80% of variability. The resulting components can be used as inputs to the following analyses.

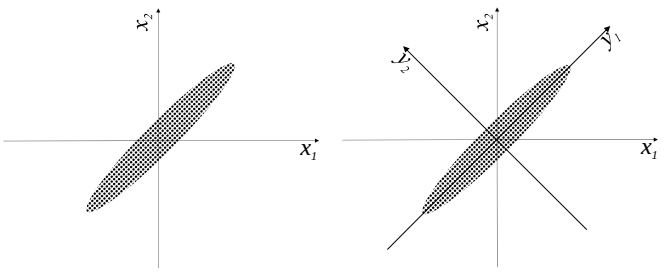


Fig. 5. The orientation of the original (x_1, x_2) and the transformed (y_1, y_2) coordinate axes with respect to the variability distribution.

The principal components can be determined by computing the covariance matrix of the data followed with finding their eigenvectors corresponding to the largest eigenvalues [4]. Thus, the principal components are the eigenvectors of the covariance matrix represented by the coordinate axes of the new transformed coordinate system. Note that the projection of the data to the eigenvectors associated with the largest eigenvalues is generally not always the most appropriate projection in terms of the following classification process.

A. Mathematical description

Suppose that $x_1, x_2, \ldots, x_M$ are $N \times 1$ vectors. According to [13], PCA can be performed in the following steps.

1) Computing the sample mean vector: $\bar{x} = \frac{1}{M}\sum_{i=1}^{M} x_i$.
2) Substracting the mean from each dimension: $\Phi = x_i - \bar{x}$.
3) Creating the $N \times M$ matrix $A = [\Phi_1 \Phi_2 \ldots \Phi_M]$.
4) Computing the $N \times N$ *sample* covariance matrix (AA^T):

$$S_G = \frac{1}{M}\sum_{n=1}^{M} \Phi_n \Phi_n^T = \frac{1}{M}\sum_{n=1}^{M} (x_i - \bar{x})(x_i - \bar{x})^T. \quad (9)$$

5) Computing the eigenvalues $\lambda_1 > \lambda_2 > \ldots > \lambda_N$.
6) Computing the eigenvectors $u_1, u_2, \ldots, u_N$, which form a basis (any vector x or actually $(x - \bar{x})$ can be written as a linear combination of the eigenvectors):

$$(x - \bar{x}) = b_1 u_1 + b_2 u_2 + \ldots + b_N u_N = \sum_{i=1}^{N} b_i u_i. \quad (10)$$

7) Dimensionality reduction - keeping only the terms corresponding to the K largest eigenvalues:

$$\hat{x} - \bar{x} = \sum_{i=1}^{K} b_i u_i, \quad \text{where } K \ll N. \quad (11)$$

The representation of $\hat{x} - \bar{x}$ into the basis $u_1, u_2, \ldots, u_K$ is thus $Y = [b_1, b_2, \ldots, b_K]^T$.

8) Finally, the linear transformation $\mathbf{R}_N \rightarrow \mathbf{R}_K$ is:

$$Y = U^T (x - \bar{x}), \quad (12)$$

where Y represents the transformed data and U^T is the PCA transformation matrix. The PCA projection is depicted in the Fig. 6 (but it is not the best projection for classification).

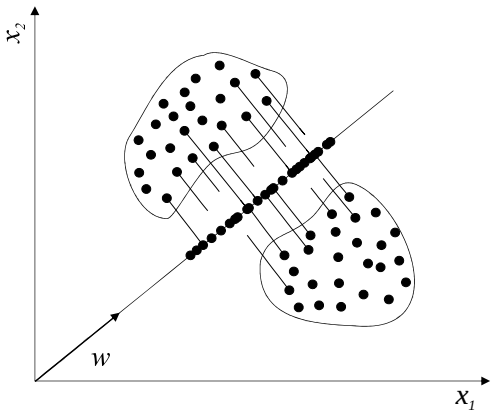


Fig. 6. PCA projection to 1D vector.

B. Limitations of PCA

If the species are not linearly related to each other, the linearity of PCA can be viewed as a limitation. It also works bad if there are too many isotropically distributed classes or meaningless variables with a high noise level [12]. The nonadaptibility and the static character (the monitored process is static) of PCA is also considered as their disadvantage.

C. Extensions of PCA

There are several extensions of PCA that overcome the traditional method. Some of these are the following: Dynamic PCA, Recursive PCA, Non-linear PCA, Multi-scale PCA, Partial PCA, Sparse PCA, etc.

V. Comparison of LDA and PCA

The main difference between LDA and PCA is that PCA changes the shape and the position of the original data to another space (but provides a poor discrimination [3]), whereas LDA changes not this position but provides much better discrimination. [7].

In some applications, the principal components of PCA are obtained firstly and these ones are then analyzed as an input to LDA [14]. This procedure is called *subspace LDA* and it overcomes some disadvantages of classical LDA.

VI. Conclusion

In this paper, we proposed two linear transformations used in speech processing. The article can be used as a teoretical overview. As a further work, experiments with real speech data using LDA and PCA are planned.

Acknowledgment

The research presented in this paper was supported by the Slovak Research and Development Agency under research projects APVV-0369-07 and VMSP-P-0004-09 and is the result of the project implementation Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] J. Juhár et al., "About development and innovation of the Slovak spoken language dialoque system," *In Journal of Electrical and Electronics Engineering*, vol. 2, No. 1, pp. 159–164, 2009, ISSN 1844-6035.

[2] S. M. Lee, S. H. Fang, J. w. Hung and L. S. Lee, "Improved MFCC feature extraction by PCA-optimized filter-bank for speech recognition," *IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU '01.*, pp. 49–52, 2001.

[3] D. Minnen, "Exploratory & Other Mini-Projects: PCA vs. LDA," November 2002. [Online]. Available: http://www.cc.gatech.edu/ccg/people/david/mini-proj.html.

[4] X. Wang and D. O'shaughnessy, "Improving the efficiency of automatic speech recognition by feature transformation and dimensionality reduction," *In Eurospeech 2003*, Geneva, pp. 1025–1028, September 2003.

[5] V. Fontaine, C. Ris and J. M. Boite, "Nonlinear discriminant analysis for improved speech recognition," *In Proceedings Eurospeech 1997*, 1997.

[6] J. Ye and Q. Li, "A two-stage linear discriminant analysis via QR-decomposition," *In IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, pp. 929–941, June 2005.

[7] S. Balakrishnama and A. Ganapathiraju, "Linear discriminant analysis - A brief tutorial," Institute for Signal and Information Processing, Mississippi State University, 1998. [Online]. Available: http://www.isip.piconepress.com/publications/reports/isip_internal/1998/linear_discrim_analysis/lda_theory.pdf.

[8] N. Kumar, "Investigation of Silicon Auditory Models and Generalization of Linear Discriminant Analysis for Improved Speech Recognition," Doctoral Thesis, The Johns Hopkins University, Baltimore, Maryland, ISBN: 0-591-40027-8, 1997.

[9] H. Abbasian, B. Nasersharif, A. Akbari, M. Rahmani and M. S. Moin, "Optimized Linear Discriminant Analysis for Extracting Robust Speech Features," *In 3rd International Symposium on Communications, Control and Signal Processing - ISCCSP 2008*, Malta, pp. 819–824, March 2008.

[10] S. Geirhofer, "Feature Reduction with Linear Discriminant Analysis and its Performance on Phoneme Recognition," *ECE272 - Individual Study in ECE Problems*, University of Illinois at Urbana-Champaign, 2004.

[11] J. Ye, R. Janardan and Q. Li, "Two-Dimensional Linear Discriminant Analysis," *In Book: Advances in Neural Information Processing Systems 17, Proceedings of the 2004 Conference*, pp. 1569–1576, July 2005.

[12] J. Shlens, "A Tutorial on Principal Component Analysis," Center for Neural Science, New York University, April 2009.

[13] G. Bebis, "Principal Components Analysis," University of Nevada. [Online]. Available: www.cse.unr.edu/~bebis/MathMethods/PCA/lecture.pdf.

[14] S. Prasad and L. M. Bruce, "Limitations of Subspace LDA in Hyperspectral Target Recognition Applications," *In Geoscience and Remote Sensing Symposium, IGARSS 2007, IEEE International*, pp. 4049–4052, July 2007.

Audio events detection and classification

Eva VOZÁRIKOVÁ

Dept. of Electronics and Multimedia Communications, FEI TU of Košice, Slovak Republic

eva.vozarikova@tuke.sk

***Abstract*—The problem of audio event detection and classification is nowadays an often solved problem. Audio and video information is used in the surveillance systems (surveillance of public places, stadiums, of public transport stations, etc.). Feature extraction is very important in the task of audio event detection and classification. Finding a suitable set of features which well represents spectral and time structure of events is the crucial task. Approaches based on HMM, GMM, SVM, Bayesian network, Neural network can be used as a classifier. This paper deals with the basic principles of audio event detection and classification.**

***Keywords*—Audio event, classifier, features.**

I. INTRODUCTION

As an audio event we consider a short audio segment which represents the potential threats. Audio events have the following properties: they have a rare occur, they are relevant for this task (may represent threat) and it is unknown whether and when they occur [1]. For example the sound of broken shop window is different from the sound of broken bottle, which can represent the threat. For the surveillance system it is very important to reduce false alarm, which arises as a result of incorrectly classified sounds.

For this task it is very useful to focus on specific groups of sounds. Sounds in one group (class) have similar features from group to group. It is possible to describe the sounds in the class by a common feature set, which has unique parameters, consequently parameters which represent sounds class well. Therefore a suitable feature set plays an important role for this task - detection and classification of events [2]. It is useful to divide the sounds depending on sound sources, for example the sounds which are produced by human or by artificial sources, see Fig. 1.

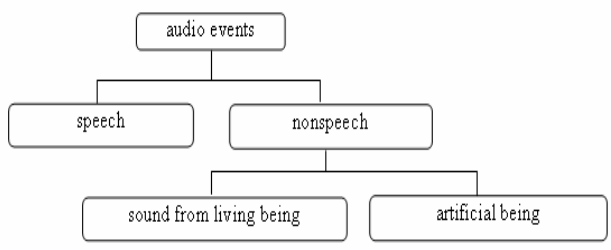


Fig. 1. Classification of the sounds

The most popular feature set consists of MFCC (Mel Frequency Cepstral Coefficients) and their time derivation. This feature sets well represents speech spectral structure but it has good performance for audio events too. We consider specific sounds from the category of sounds:

- living beings such as *shout*,
- artificial sounds such as *gun shot*, *explosion* and *broken glass*.

The detection system is built for these four audio events. Various learning methods can be use for this task for example classifier based on the Hidden Markov Models (HMMs) [1], [3], [4], Gaussian Mixture Models (GMMs) [3], [5], Support Vector Machines (SVMs) [3], [6], and Artificial Neural Networks [7] and Bayesian Network [3].

In this work we use HMMs for building the classifier. For each class of the audio events (shout, shot, glass and explosion) HMM model was trained.

The present paper has the following structure: section II describes the feature analysis, section III describes the corpus and the section IV contains methods of audio event classification. Our experiments are described in the section V. This is followed by the conclusion and future work.

II. FEATURE ANALYSIS

Many works deal with feature analysis for speech signal but only few works are focused on the feature analysis for acoustic events. The spectral structure of events and speech is different, see Fig. 2.

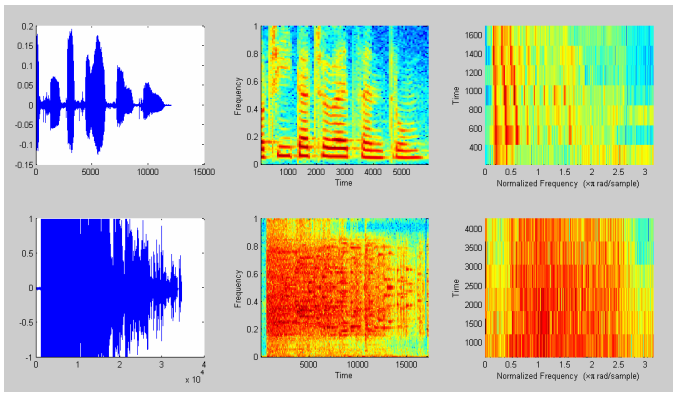


Fig. 2. Time - dependent frequency analysis for speech and audio event

In the upper part on the figure we can see the speech signal and its spectral characteristics. The speech signal has the periodic character and the most information is in low frequency part. Audio events are characterized by the nonperiodic nature and wider spectrum than the speech signal. Features for speech signal can neglect the frequency

parts that contain less speech discriminative information but may contain much discriminative information for acoustic events [2]. For these reasons, the feature set contains only features describing the speech signal may be not very suitable for some kinds of acoustic events. In addition to detecting acoustic events we often have to encounter with low value SNR. Too high levels of noise can significantly degrade the score of classification. The feature set may contain LPC, LPCC, MFCC, PLP. This set is very often used to describe the speech signal. LPC is based on a predictive analysis assuming that a speech sample at the current time is a linear combination of past n-speech samples. MFCC and PLP are computed after a transformation of signal in the spectral domain [8]. The set of features for acoustic events may contain speech features and another parameters for example Zero crossing rate (ZCR), pitch frequency, Fundamental frequency, Short-time energy, Sub-band log energies, Sub-band log energy distribution, Sub-band log energy correlations, etc. [9]. Features are computed from the frame of the signal where for each frame belongs a features vector. A large amount of parameters in the features vector is often reduced by the Principal Components Analysis (PCA) and Linear Discriminant Analysis (LDA). These methods are based on decorrelation features which describe the input signal. Reduction of vector size is achieved by PCA and LDA [10]. Due to these methods training and testing process can be more effective.

III. Corpus

In the experiment we work with the set of specific audio events. It contains audio events which are of interest for surveillance systems such as explosions, broken glass, shots and shouts [11]. The common problem was the lack of training data. Data for this work was obtained from the internet, movies and we also used our recordings. They were re-sampled to 24 kHz, 16 bit, mono-channel format. The recordings were cut and manually labelled. Particular events were not overlapped. In the training and testing process recordings with noise were used. The composition of corpus is the following.

TABLE I
Corpus

Acoustic events	Train files	Test files
Explosion	9	3
Broken glass	40	18
Shot	50	23
Shout	34	15
Other	20	6
Background	40	10
Silent	60	24

IV. Classifiers

In the process of audio event classification various methods can be successfully used, for example methods based on supervised learning, semi-supervised learning and unsupervised learning. The main difference between the supervised and unsupervised learning is in the algorithm which uses labelled or unlabelled input data. The supervised learning method uses labelled data and the non-supervised method works with unlabelled data. The lack of labelled training data is partially addressed by the semi-supervised method [1].

A. Semi-supervised HMM

One of the semi-supervised learning methods is the semi-supervised HMM. The basic principle is the following:

The model for usual event is trained by the large amount of data and model for unusual event is derived from a usual model in the iteration process via Bayesian adaptation Fig.3. In each step the samples are detected which have the lowest likelihood given the usual model. These samples are identified as outliers and outliers are used for training unusual event models [1].

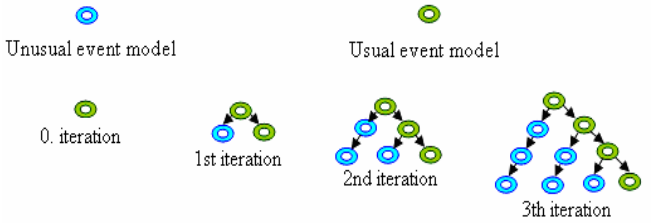


Fig. 3. Illustration of the semi-supervised learning algorithm. Unusual event model is created from usual event model using the outliers.

B. Large margin GMM

Another interesting algorithm for classification is the large margin GMM (Gaussian mixture model) which has many parallels to SVM. In the large margin GMM, sound classes are modeled by the ellipsoids instead of half spaces, see Fig. 4.

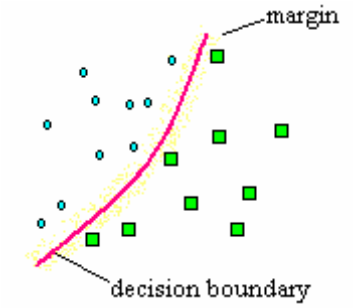


Fig. 4. The decision boundary in the large margin GMM separates the sound classes by the elipsoid.

Inspired by support vector machines (SVMs), the learning algorithm in large margin GMMs is designed to maximize the distance between labelled examples and the decision boundaries that separate different classes. In contrast to SVMs, however, large margin GMMs are very naturally suited to problems in multi way classification also, they do not require the kernel trick for non-linear decision boundaries. As the Kernel trick is not necessary to induce non-linear decision boundaries, large margin GMMs are more readily trained on very large and difficult data sets [4], [5].

C. Support vector machine

Support vector machine (SVM) is successfully used in the field of sound recognition. SVM classifier is a binary classifier but it can be used for multiclass classification too. The multiclass classification is achieved by many binary classifiers. SVM maximizes the margin, the boundary which separates one type of classified input data from another one. Nonlinear decision boundaries are supported by mapping the input feature space to a higher dimension where the features are linearly separable. This mapping is done using some kind of Kernel-trick. A very popular Kernel - trick is Gaussian (linear) which has these advantages: speed of training and good performances. Depending on misclassified examples it is possible to divide SVM to the „soft“ and „hard“ margin SVM, Fig. 5, 6. The soft SVM is more suitable for the classification in a noisy environment than hard SVM [3], [6], [8].

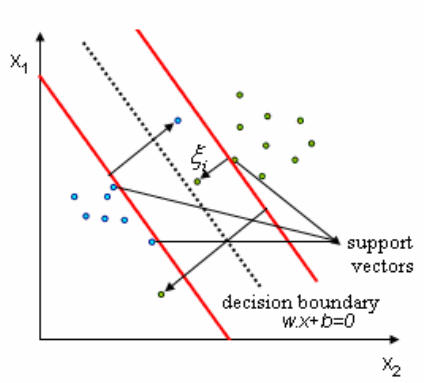


Fig. 5. Soft margin SVM

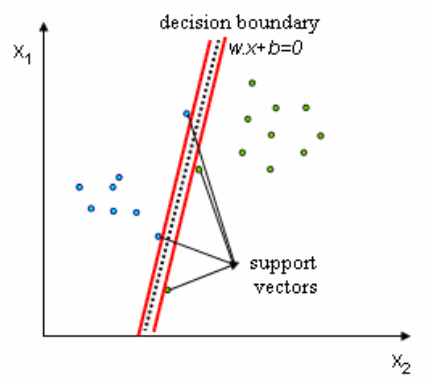


Fig. 6. Hard margin SVM

V. Experiments

In our experiment, the HMM classifier was used to classify sounds of audio events. The goal of the acoustic processing is provided by appropriate method for determination of conditional probability *P(O/W)*. It is probability, that word/event *W* will represent acoustical vector *O*. In our experiment were used continuous ergodic HMMs with 10 mixture and different numbers of state. The set MFCC is widely used in speech recognition and also in audio application. We used MFCC, their first and second time derivation. The vector size has 36 parameters. In the pre-processing the energy parameter was omitted, because every sound which had strong energy would be classified as an audio event, and other audio event with low energy would not be detected. Therefore the set of features should be independent from energy parameters. The input audio signal was processing in the frequency range of 50 Hz to 12 kHz, window size was set to 20 ms and preemcoeficient was set to 0.97. The high frequency contents of many types of sounds required a higher sampling frequency. In this work 24 kHz was used. The frequency 8 kHz was not enough to this task.

A. Test: Number of state

We tried to decide, which model is the most suitable for the task of acoustic event detection. The model with different number of state was trained for each class of the audio event. For testing were used recordings which contained only one acoustic event (shot, broken glass, shout and explosion). In the recognition process we achieved the following results.

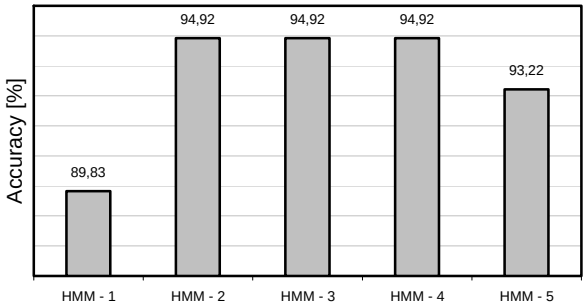


Fig. 7. HMMs results for the test with different number of states for separate audio events.

The models with 2, 3, 4 states obtained the accuracy 94,92%. The second best result 93,22% obtained five states HMM and one state model achieved the accuracy 89,83%.

B. Test: Detection of audio events

For this test another models for classes of events such as silent, background and for other sounds were trained. The model for the other sounds was trained from data, which not belongs to the remaining classes. The recordings were manually labelled. The data were described by MFCC, delta and acceleration coefficients. Recordings which contained more acoustic events were used for testing.

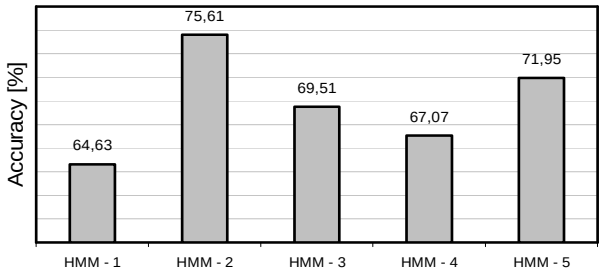


Fig. 8. HMMs results for the test with different number of states for the recordings with many audio events.

The best accuracy was reached by two states HMM models.

VI. Conclusion

These initial experiments with training and testing HMM models showed the basic principles of the task of acoustic event detection and classification. We used the HMM models for each audio event. The best results of models with two states were obtained. Many errors were caused by the lack of training data. A special attention should be devoted to the background model which covers everything sounds to given environment.

One of the most interesting tasks of this work will be built of a new acoustic models and extending the sound corpus. We will be focused on other classification methods which are described above in this paper.

Acknowledgment

This work has been performed in the framework of the EU ICT Project INDECT (FP7 – 218086) and as a result of the project implementation Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120020) supported by the Research & Development Operational Program funded by the ERDF.

References

[1] D. Zhang, D. Gatica-Perez, S. Bengio, I. McCowan, "Semi-supervised adapted HMMs for unusual event detection," Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005.

[2] X. Zhuang, X. Zhou, T.S. Huang, M. Hasegawa-Johnson, "Feature analysis and selection for acoustic event detection," Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Las Vegas, Nevada, USA, March-April 2008, pp. 17-20.

[3] I. Trancoso, J. Portêlo, M. Bugalho, J. Neto, A. Serralheiro, "Training audio events detectors with a sound effect corpus," Proc. INTERSPEECH, Brisbane, Australia, September 2008, pp. 2546-2549.

[4] F. Sha,. L. K. Saul, "Large margin hidden Markov models for automatic speech recognition," In B. Schoelkopf, J. Platt, and T. Hofmann (eds.), Advances in Neural Information Processing Systems 19, MIT Press, pp. 1249-1456.

[5] F. Sha, L. K. Saul, "Large margin Gaussian mixture modeling for phonetic classification and recognition," Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Toulouse, France, 2006, pp. 265-268.

[6] J. Portêlo, M. Bugalho, I. Trancoso J. Neto, A. Abad, A. Serralheiro, "Non-speech audio event detection," Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Taipei, Taiwan, April 2009, pp. 1973-1976.

[7] M. Katrak, J. Juhár, Training Slovak Phonemes for Speech Recognition Using Neural Network, Proc. ECMS 2007, Liberec, Czech Republic, May 21 – 23, 2007, pp. 73-78.

[8] J.-L. Rouas, J. Louradour, S. Ambellouis, "Audio events detection in Public Transport vehicle," Proc. IEEE Intelligent transportation systems conference, Toronto, Canada, September 2006, pp. 73 -738.

[9] A. Temko, "Acoustic event detection and classification," Phd. thesis, Universitat politècnica de Catalunya, Brcelona, December 2007.

[10] Jieping Ye, Qi Li, "A two-stage linear discriminant analysis via QR-decomposition," Proc. IEEE Transactions on pattern analysis and machine intelligence, June 2005, Vol. 27, No.6.

[11] M. Pleva, E. Vozáriková, S. Ondáš, J. Juhár, A. Čižmár, "Automatic detection of audio events indicating threats", IEEE International Conference on Multimedia Communications, Services and Security, to be published.

Analyses of Knowledge Creation Processes

[1]*Jozef WAGNER,* [1]*Ján PARALIČ*

[1]Dept. of Cybernetics and Artificial Intelligence, FEI TU of Košice, Slovak Republic

[1](jozef.wagner, jan.paralic)@tuke.sk

***Abstract* – Current socio-technical trends in the learning include the use of Computer Supported Collaborative Learning tools based on the social constructionism theories and the trialogical approaches to the learning. They have put some new requirements on the analyses of students activities. This paper presents the analyses of the practices that result in the creation of a new knowledge. The current state of the art is described together with the suggested approach to the modeling and analyses of the knowledge creation processes.**

***Keywords* - computer supported collaborative learning, knowledge-creation practices, process analysis, trialogic learning.**

I. Introduction

We live in a knowledge society, where the knowledge is the driving force of advancement. Ability to grasp the information at hand, to create and make use of the knowledge is still very exceptional. [1] In spite of the investments to the learning and education, the advancements in this area are far from revolutionary.

Knowledge management has transformed from the field of artificial intelligence [2] into the multidisciplinary field composing practices from AI, information systems, management and psychology. The emergence of the Computer Supported Collaborative Learning (CSCL) tools helps to systematize some parts of the learning process, while giving an opportunity to record and analyze the activities that were performed in the courses. This is crucial to the learning process, as by capturing and analyzing such activities, one can identify the practices that led to the creation of the new knowledge. By examination, inspection and reflection on these practices, both teacher and student can better understand how the learning process was performing. [3]

II. Background

A. Constructionism

In the past years, the use of CSCL tools in the education grew rapidly. This and the other modern trends in the learning are based on the constructionism theory, developed by Seymour Papert, a pioneer of artificial intelligence [4]. In his theory, people are actively acquiring knowledge by practice and interaction with the environment and within the group. Constructionism moreover emphasizes the notion of "learning by making", focusing on the object of the activity, a public entity on which everybody is cooperating. [5] The term "knowledge artefact" is emerging, representing the man made object (real or digital), which is the object of the collaborative activity. A typical example of such knowledge artefact is a wiki page on a Wikipedia.

B. Trialogical Learning

Scandinavian researchers Sami Paavola and Kai Hakkarainen have tried to bring together various approaches to the process of acquiring knowledge and introduced three metaphors of learning [6]:

- Monological approach – learning is the process of knowledge acquisition by an individual, "student – book"
- Dialogical approach – learning through active participation on the activities in the given group and environment
- Trialogical approach – learning through collaborative creation of shared objects, knowledge artefacts

Concept of trialogical learning emerged from the finding, that in todays knowledge society, it is not sufficient to just mediate knowledge to the individual learners, nor to just support group learning. The third approach, also called the metaphor of knowledge creation, is focusing on the process of how people collaborate on the creation of the knowledge artefacts.

The Trialogical learning is not a brand new approach. It builds on the Constructionism theory and there are existing examples of practical and widely used learning methods based on trialogical learning, such as knowledge creation in organizations (Nonaka, Takeuchi) [7], knowledge building (Bereiter) [8], Activity Theory [9] and theory of Expansive learning (Engeström) [10]

III. Existing approaches

In a trialogical approach to the learning, the focus is on the knowledge artefacts, which were created during learning in a given CSCL tool. The tool should thus be able to explain and describe processes involved with a given artefact. This leads to the better understanding of the learning process.
Following sources can serve as an input to the analyses:

- Current data from the CSCL tool
- Historic versions of the artefacts
- Event logs
- Process models and descriptions

First two sources of the information can be used for the statistical analysis, nowadays used in many applications. It is the third source and the last one, which are of great importance for the analyses of knowledge processes.

A. Traditional approaches to the event log analyses

Traditional event log analyses support following features:

- Browsing the event logs, filtering
- Searching with regex support, custom parsers
- Real time support
- Finding critical events, support for notifications

- Aggregation functions (min, max, count)
- Statistical functions (average, mean, frequency, histogram)

Mainly in the web environment, where the event logs are omnipresent, following web-specific features are present:

- Numbers of visitors, unique vs. recurrent ones
- IP address translation, geolocation
- Peak hours, entry and exit pages
- Browser and OS charts, used search engine

Classic example of such event log analyses is the Urchin, or the Google Analytics [11], shown on the Figure 1.

Fig. 1. Urchin, Google analytics

Process which is being analyzed in these traditional tools is static, its model seldom changes. While the process can be identified and is executed repeatedly, it is defined a priori and tool does not support its modifications or customization.

B. Process Mining

Process mining tool, called ProM [12], is developed in the Eindhoven University of Technology, Netherlands. This analytical tool represents state of the art in the process analysis.

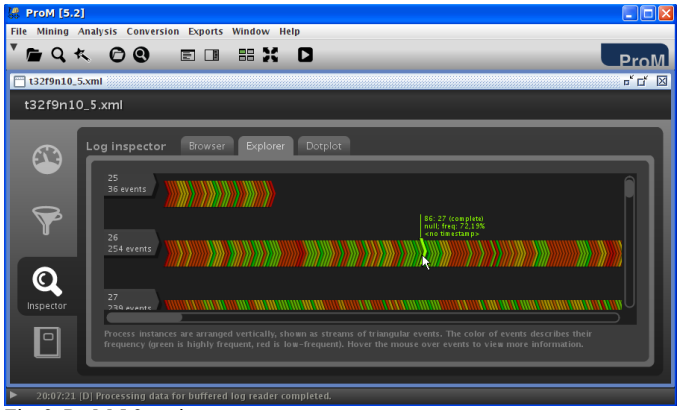


Fig. 2. ProM 5.2 environment

The input format if the event logs for this tool is the MXML format, and it is also possible to import a wide range of existing log formats (e.g. from web servers, version control systems, ERP, relational databases). Actual processes can also be in various formats, including Petri Nets, YAWL, BPEL and event-driven process chains. ProM environment is depicted in a Fig. 2.

ProM tools offers vast number of analyses of the event logs themselves, or analyses where also the process model is present. For the purposes of analyses, the ProM tool expects that it is possible to divide activities from the event log into the separate sets, where each set represents one instance of a process, called 'case'. Given activity presents in the event log must belong to exactly one process instance. This requirement is critical for the purposes of the subsequent analyses in the ProM tool.

Some of the analyses offered by the ProM tool are:

- Process discovery – using control-flow algorithms (α-algorithm, fuzzy miner, genetic miner)
- Organizational algorithms – Social network analysis, analysis of the organizational model
- Conformance and LTL checker
- Extension – process model extension with the results from performance and decision mining
- Simulation – event log simulation based on the execution of a given process model

C. CSCL tools

CSCL tools usually contain some type of analytical tools, used mainly to analyze activities performed within the system. This analysis is designed to be used by the students, teacher or the administrator of the given CSCL tool. These analytical tools support a lot of functionalities from a traditional event log analyses, mainly:

- Activity chart, summary of visits and contributions
- Temporal view on a particular course or a group, basic aggregation capabilities
- Report on the students presence and activities

Figure 3 shows an output of the analysis in the Moodle [13], as a representative example of a CSCL tool.

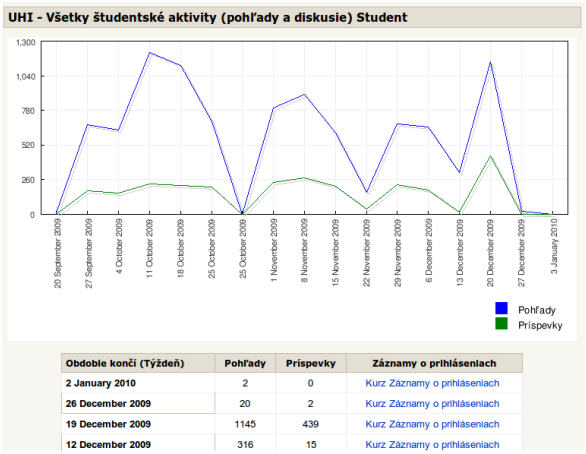


Fig. 3. Temporal view of the group activities in the Moodle tool

Similar temporal views and aggregation analyses are available in the Knowledge Process Environment, a CSCL tool developed within the 6. FP IST project called KP-Lab [14]. Stand-alone tool called Visual analyzer (depicted on Fig. 4) is able to provide an aggregation analysis in an interactive and user friendly way. User is able to choose, filter and restrict the types of events he/she wants to analyze. The result is presented in a graph and it is also possible to export the results into the spreadsheet applications.

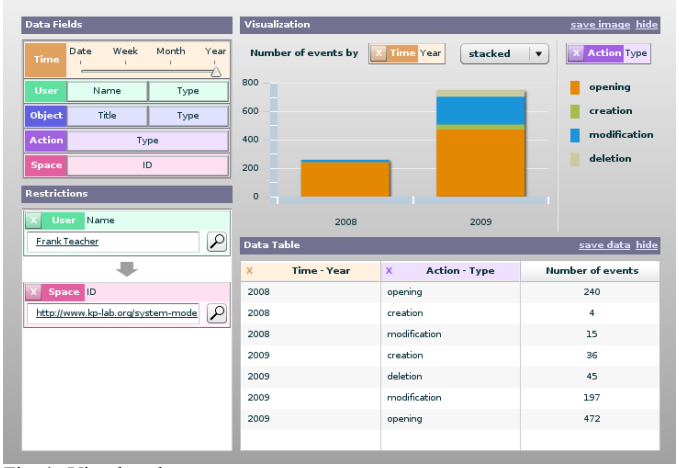


Fig. 4. Visual analyzer

The last type of analysis presented in this paper is called a Timeline based analysis (Figure 5). It provides a detailed report on the activities performed on a given knowledge artefacts. The information is presented on a timeline, emphasizing the temporal aspect of the analysis. Each participant is displayed on a separate row, and the tool provides clear view on who, when and how participated on the activities around given knowledge artefact. It is this type of analysis, which provides the clearest view on the processes that were performed in the given course. With a proper visualization, a teacher can discover and better understand these processes. However, the processes are still not explicitly and formally defined, so it is on the user, how the result of the timeline view analysis will be interpreted.

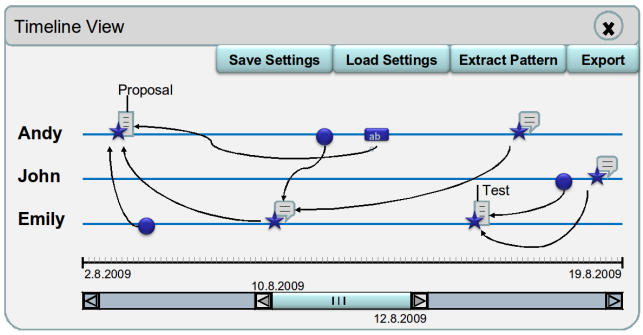


Fig. 5. Timeline view

IV. Suggested approach to the analysis

Suggested approach to the analysis of knowledge creation practices is motivated by the trialogical learning. According to the third metaphor of learning - the knowledge creation metaphor - the learning is done through the collaborative creation of shared knowledge artefacts. Analysis should thus focus on the processes of creating and building of new materials, objects, artefacts. It is with the help of these processes that the knowledge is deliberately and consciously discovered, created and enriched.

Traditional approaches to the process modelling assume that the process can be clearly and distinctly defined. This is in a direct contradiction with the methods of trialogical learning, in which the learning is understood as a complex process, often unique, ill defined and not easily formulated. Applying methods from traditional branches of process modeling would not lead to the better understanding of the knowledge creation practices.

What should knowledge creation process look like? Model of the knowledge process must satisfy following requirements:

- Must formalize the notion of time
- Orientation on the subject, or the group
- Orientation on the object and activities
- Ability to cover ill defined cases
- Ability to dynamically build, modify and customize the description of the process

This however does not solve the problem of process discovery and identification. It is not possible to explicitly define a process where its instances are different in some parts and are always unique. To cope with this problem, I suggest generalizing processes in the CSCL tools into so called knowledge-creation patterns. These patterns would not completely describe the knowledge process as such, but they will be able to formally and explicitly define at least some parts of the process. Having formal description of the pattern, analytical tools could search the subsequent event logs for the next occurrences of this pattern, greatly helping user in the identification and comprehension of the knowledge practices in the course.

Pattern will be formalized as a sequence of activities. Moreover, following requirements have to be taken into the account in the process of formalizing the knowledge creation patterns:

- Generalization of the activities or their parameters
- Ability to define relationships between activities on the level of relationships between their generalized parameters
- Beyond simple sequence, ability to define multiplicity and intervals between activities
- Ability to define weights of importance for the parts of a pattern, with the intention of more accurate pattern definition and matching

It is clear that the analytical tools could not define the patterns automatically. The process of pattern definition must be interactive and iterative, where analytical tool helps user to express his/hers intentions by providing user friendly environment to define, customize and apply patterns.

V. Conclusion

Social constructionism and trialogical learning theories represent modern approaches to the learning in schools and organizations. Computer supported collaborative learning tools are quickly finding way into the learning process and improve the management and understanding of the courses. Teachers and students benefit from various kind of analysis provided by such tools. However, analytical tools available within a given e-learning tool mostly offer only traditional features. This paper proposes to analyze events from the CSCL tools by creating and finding knowledge-creation patterns. This process generalization would formalize parts of the knowledge processes taking place in the CSCL tool and help users in identification and comprehension of the knowledge practices. Requirements for the formal definition of such patterns are outlined and the paper suggests using the timeline visualization to aid the user during analysis.

Acknowledgment

The work presented in this paper was supported by: European Commission DG INFSO under the IST program, contract No. 27490; the Slovak Grant Agency of Ministry of Education and Academy of Science of the Slovak Republic

under grant No. 1/0042/10. This publication is the result of the project implementation Development of Centre of Information and Communication Technologies for Knowledge Systems (project number: 26220120030) supported by the Research & Development Operational Programme funded by the ERDF

The KP-Lab Integrated Project is sponsored under the 6th EU Framework Programme for Research and Development. The authors are solely responsible for the content of this article. It does not represent the opinion of the KP-Lab consortium or the European Community, and the European Community is not responsible for any use that might be made of data appearing therein.

References

[1] Jonathan Rosenberg: From the height of this place, available at <http://googleblog.blogspot.com/2009/02/from-height-of-this-place.html>
[2] Paul M. Hildreth, Chris Kimble: The duality of knowledge. Information Research, Vol. 8 No. 1, October 2002
[3] Scardamalia, M., & Bereiter, C.: "Knowledge Building". Published in: J. W. Guthrie (Ed.), Encyclopedia of Education. 2nd edition. New York: Macmillan Reference, 2003
[4] Seymour A. Papert, biography. Available at <http://www.media.mit.edu/people/papert>
[5] Seymour Papert: Mindstorms: children, computers, and powerful ideas. Basic Books, 1980, ISBN 0465046290
[6] Sami Paavola, Kai Hakkarainen: The Knowledge Creation Metaphor – An Emergent Epistemological Approach to Learning. Science & Education, Springer Netherlands, 2005, ISSN 0926-7220
[7] Ikujirō Nonaka, Hirotaka Takeuchi: The knowledge-creating company: how Japanese companies create the dynamics of innovation. Oxford University Press US, 1995, ISBN 0195092694
[8] Scardamalia, M., & Bereiter, C.: "Knowledge Building". Published in: J. W. Guthrie (Ed.), Encyclopedia of Education. 2nd edition. New York: Macmillan Reference, 2003
[9] Engeström, Y.: Perspectives on activity theory. Cambridge, MA: Cambridge University Press, 1999, ISBN 052143730X
[10] Engeström, Y.: Learning by expanding: An activity-theoretical approach to developmental research. Helsinki, Orienta-Konsultit, 1987, ISBN 9519593322
[11] Google Analytics, available at <http://www.google.com/analytics/index.html>
[12] Process Mining: Project page for ProM and MXML, available at <http://prom.win.tue.nl/research/wiki/>
[13] Moodle, available at <http://moodle.org/>
[14] KP-Lab, avaialble at <http://kp-lab.org/>

Adaptation Techniques of Domain-Specific Languages

Ľubomír WASSERMANN, Michal FORGÁČ

Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

(Lubomir.Wassermann, Michal.Forgac)@tuke.sk

***Abstract*—This paper presents current overview of the area of domain-specific languages, their utilization, division and ways of their adaptation. Domain-specific languages (DSLs) are computer languages tailored to the specific application domains. They can have textual, graphical or combined form. Their evolution can be based on metamodel transformation or grammar adaptation. This paper describes some existing approaches to DSL evolution including approach for runtime evolution of language grammar.**

***Keywords*—software engineering, domain-specific languages, language-oriented programming, adaptation of language.**

I. Introduction

Nowadays, there is significant demand for software systems designed for various domains. Complexity and size of software projects are gradually increasing and with them is also increasing pressure on software engineers. Software engineering is facing various challenges. The main reason is permament change to which software system must dynamically adapt during its life cycle (mainly after its deployment).

However due to various problems, wrong choice of programming paradigm and its tools during software development causes, that few software projects are successful in the result. The reason of failure often is unfulfilled original specification, budget or time horizon. One of the aims of software engineering is the effort for the creation of new paradigm, which would solve existing problems and simplify and accelerate the development of software systems.

During last years in software engineering emerged various trends and approaches to improve the existing situation. Use of general purpose languages for solving specific problems is in some cases unappropriate. Therefore, the importance and popularity of new approach to software development called language-oriented programming [1] is increasing. This approach consists of creating own high-abstract languages oriented at a specific domain, called domain-specific languages (DSL) [1]. Language-oriented programming and usage of DSLs is promising approach to increase productivity of the development of software systems. This is achieved by raising the level of abstraction and usage of effective generative techniques.

In language oriented programming, a language is defined by its structure, editor and semantics. Its structure defines abstract syntax, its editor defines concrete syntax and its semantics define behaviour [2]. Language oriented programming introduces different aproach to development of a complex software system. This approach starts with development of formal, specific, domain-oriented, high-level programming languages which can be used to develop software system or program.

Domain-specific languages represent significant increase in productivity in the development of software systems thanks to their focus on solving problems from a given application domain. The interconnection with the domain has also some disadvantages. Since the real word is not static, domain for which the language is intended is dynamically changing. Therefore, to preserve usability of a domain-specific language it is necessary to adapt it to domain changes.

Adaptation[1] is important part of life-cycle of domain-specific languages. Therefore, one of the objectives of research in this area is support of evolution of a domain-specific language in the form of a special execution environment or architecture, which provides this support. Since the largest costs are associated with language evolution, there is strong effort for automatization of a process as much as possible and thus reduce the costs associated with evolution.

II. Domain-specific languages

Domain-specific languages (DSLs) are computer languages tailored to the specific application domains. These languages are usually small, with restricted suite of notations and abstractions [3]. They have also another names, e.g. application-oriented languages [4], special-purpose languages [5], specialized languages [6], or task-specific languages [7]. Famous representatives of DSLs are for example SQL, or LaTeX.

Benefits of DSL usage are e.g. conciseness of source code, productivity, reliability, maintainability and portability enhancement. Domain-specifing languages are aimed to solve problems from specific domain with special constructs and notations and thus offer increased productivity when compared to general-purpose language. Source code written in DSL is easily readable and understandable also by domain experts [8].

There are also some disadvantages when using DSLs. It is namely, cost of designing, implementing and maintaining of a DSL, cost of education for DSL users, limited availability of DSLs. To reduce overall cost connected with usage of DSL, when deciding whether to use DSL or not, reusability of newly created DSL should be considered.

Since DSLs are intended to address problems of a target domain, the form and type of designed DSL depends on the nature of problems, which will be solved using given DSL. Based on the way of DSL creation and its representation is possible to divide DSLs on some groups:

[1]Term adaptation is used in the meaning of evolution

- textual DSLs - program is represented through textual form
- visual (graphical) DSLs - program is represented by graphic objects

There are various ways of DSL creation. According to the [9] it is possible to divide these approaches to three groups:

- internal DSLs - an internal DSL is dependent on the selected programming language using its syntax
- external DSLs - an external DSL has its own syntax defined independently of any other language
- language workbenches - language workbench is a toolkit, which supports creation of DSL, this toolkit is included in integrated development environment (IDE)

III. Adaptation of Domain-Specific Languages

Traditionally, programming languages are perceived as finite set of black box abstractions [10]. End users of programming language have only limited or even no control over implementation and semantics of these abstractions. This point of view is misleading because it is based on a fact that languages are immutable artifacts [11].

Programs are dependent on language, in which they are written, and tools which this language offers (e.g. interpreters, compilers, virtual machines etc.). When we admit that program is subject of evolution and is tightly coupled with language in which is written we can assume that language can be subject of evolution, too [12].

Evolution is important part of the life cycle of DSL. Because of dynamically changing environment, change is permanent driving force of evolution of language. When considering the evolution of DSL, stress is mostly laid on the automatization of the evolution process. This evolution process should be connected with evolution and adaptation of DSL programs and tools used to process them.

A. Evolution of DSL based on metamodel transformation

Evolution of DSL used in process of domain-specific modeling requires specific approach because modeling DSLs are mostly graphical languages expressed by subset of UML in form of metamodel. Metamodel defines DSL and its elements based on elements of UML. Therefore evolution of modeling DSL can be achieved by transformation of its metamodel.

Vermolen presented proposal of architecture of system for support of coupled evolution of metamodel and migration of corresponding models[2] [13], [14]. Evolution of DSL is defined by basic transformations. Transition between different version of DSL is defined by evolution specification which consists of individual operations (transformations). Evolution specifications are then mapped to the model migrations to establish automatic migration.

Evolution specification is output from automatic process of detection of DSL evolution. In [13] are mentioned two approaches to evolution detection - detect changes with help of editor or detect all changes after the modification of language. First approach has one disadvantage that dependency on specific editor reduces flexibility of DSL and its evolution.

Process of evolution itself consists of detection of evolution and creation of evolution specification and its mappings. This is input of the architecture (Fig. 1) that controls and realizes DSL evolution by:

- automatic derivation of transformation language for particular domain (DSL),
- automatic derivation of transformation interpreter written in transformation language,
- automatic migration of programs based on transformations.

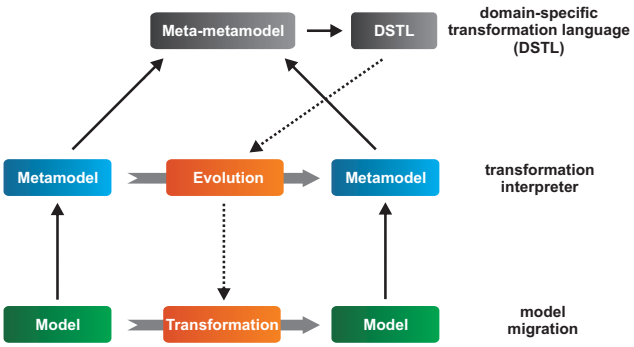


Fig. 1. Architecture of system for support of coupled evolution of metamodel and model by Vermolen

Similar approach was used by design of system proposed in [15]. Proposed system for support of DSL evolution consists of:

- representation of metamodel of DSL by domain classes, properties and relations between classes,
- comparison algorithm that calculates differences between DSL definitions,
- convertor generator that generates convertors between DSL versions based on detected differences.

Coupled evolution of metamodels and models is subject of research of more research groups, e.g. [16], [17], [18].

B. Evolution of DSL based on grammar adaptation

Evolution of DSLs that are defined formally by context-free grammars lies in the adaptation of the grammar and its language processor and existing programs. This approach is used in work of Jürgens and Pizka [19], [20]. They are presenting their tool for development of DSLs with support of their evolution. Tool *Lever* provides languages to support DSL development and evolution. Adaptation of syntax, language processor and program is automatized and its based on evolution description expressed with help of languages provided by Lever.

Architecture of the tool Lever provides functionality to ensure DSL evolution and transition of the programs to new version of DSL. Architecture comprises of:

- *DSL history* - Contains evolution operations (steps) that defines all DSL versions and difference between them. This information is used for automatic adaptation of compiler and DSL programs to specific version.
- *DSL specification* - Specification of syntax and translational semantics of the current version of DSL. It is available during the execution and controls the compilation.
- *Syntax tree* - Representation of the program in the memory. It is abstract syntax tree extended by concrete syntax and semantic attributes.

[2]From point of view of domain-specific modeling, model represents program.

- *DSL program* - Represents an input for the compilation process. DSL programs include information about the version of DSL, in which they were written.
- *Target code* - Represents the result of compilation process.

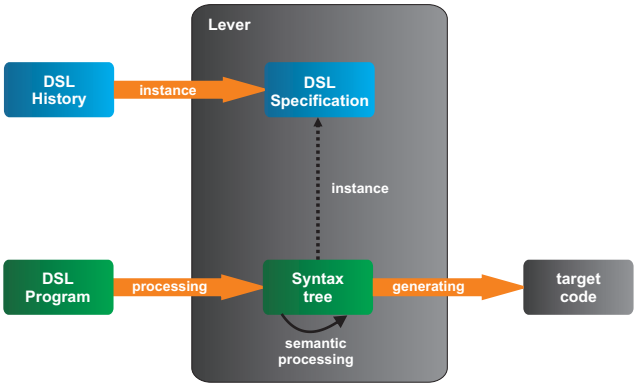


Fig. 2. Architecture of Lever tool

Tool *Lever* allows compilation of any version of DSL included in DSL history. The process of translation of any version of DSL is as follows:

1) Identification of the version of DSL.
2) Implementation of the evolution operations based on DSL history to create DSL specification for corresponding DSL version.
3) Generation of the language processor according to DSL specification.
4) Language processor processes the DSL program and creates a syntactic tree.
5) Transformation of DSL specifications and syntax tree to the latest version.
6) Semantic processing according to semantics contained in the DSL specification. In the output is target code generated for a given syntactic tree.

Among the disadvantages of this tools include a lack of support for adaptation of additional tools such as an editor with syntax highlighting or debugger. They must be modified manually. Another disadvantage is the focus only on the textual DSL.

IV. Runtime Evolution of Grammar

Evolution of language during program runtime requires special tools or architecture, which supports this type of evolution. In the PhD. thesis [21] was described experimental execution environment, which supports combined dynamic modification (evolution) of programs and languages. This solution is based on the principles of metaprogramming, reflection, and aspect-oriented programming.

Proposed method of combined modification is based on the statement, that any execution environment on the metalevel 2, which allows modification of program during its execution is a mestasystem of execution environment on the metalevel 1, which allows modification of a language for the program, which is in the environment on the metalevel 1 interpreted (Fig. 3).

The method of combined modification of programs and languages allows four types of modification:

- dynamic modification of program,

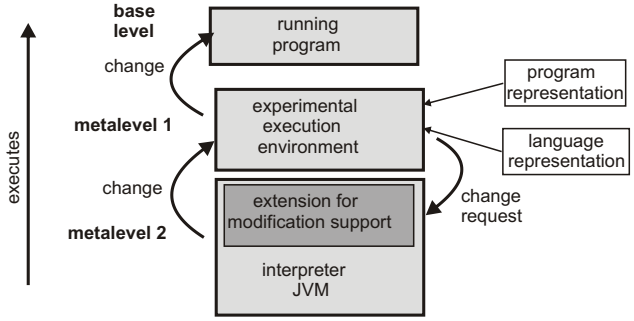


Fig. 3. Principle of method for combined evolution

- dynamic modification of semantics actions independently from program modification,
- dynamic modification of lexical and syntactic specification of language and semantic actions and consecutive dynamic modification of program,
- dynamic modification of lexical and syntactic specification of language and consecutive modification of program without modification of semantic actions.

Listed possibilities can be combined properly. Objective of this method is reaching of modification of interpreted functionality through properly implemented programmed solution.

Designed experimental execution environment offers all types of modification. This environment is implemented in Java programming language with utilization of Javassist class library for editing bytecodes. Another utilized mechanism is HotSwap mechanism, which allows dynamic reloading of required class file to update the class definition. Lexical and syntactic analysis is performed by program, generated through ANTLR parser generator [22].

This solution can be used in some specific application domains. For example there could be a control program for an industrial machine and there would be a need for optimization of its manufacture process during its runtime without device stopping. Such optimization would require new language elements with appropriate semantics.

V. Conclusion

Based on the analysis of state-of-the-art in area of evolution of software systems and languages in our research group, the aims of next research were established.

1) Analysis of existing language workbenches for DSL development with support of DSL evolution.
2) Design system architecture for support of evolution of DSL and its tools.
3) Create mechanism for support of dynamic evolution of DSL.
4) Create mechanism for automated coevolution of DSL and its programs.
5) Implementation and experimental verification of designed system.

The aim is to create system for support of DSL evolution coupled with evolution of corresponding programs. DSL evolution should be connected with adaptation of tools used to process programs of DSL (language processor, target code generator). Designed system should be flexible and provide functionality to support automatized process of DSL evolution.

VI. Acknowledgment

This work is the result of the project implementation: Centre of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] M. P. Ward, "Language-oriented programming," *Software — Concepts and Tools*, vol. 15, no. 4, pp. 147–161, 1994.

[2] S. Dmitriev, "Language oriented programming: The next programming paradigm," http://www.onboard.jetbrains.com/articles/04/10/lop/, 2004.

[3] A. van Deursen, P. Klint, and J. Visser, "Domain-specific languages: an annotated bibliography," *SIGPLAN Not.*, vol. 35, no. 6, pp. 26–36, 2000.

[4] J. E. Sammet and D. Hemmendinger, "Programming languages," in *Encyclopedia of Computer Science*. Chichester, UK: John Wiley and Sons Ltd., 2003, pp. 1470–1475.

[5] S. Kamin, S. N. Kamin, and D. Hyatt, "A special-purpose language for picture-drawing," in *USENIX Conference on Domain-Specific Languages*, 1997, pp. 297–310.

[6] T. J. Bergin and R. G. Gibson, Eds., *History of programming languages—II*. New York, NY, USA: ACM, 1996.

[7] B. A. Nardi, *A Small Matter of Programming: Perspectives on End User Computing*. The MIT Press, 1993.

[8] M. Mernik, J. Heering, and A. M. Sloane, "When and how to develop domain-specific languages," *ACM Comput. Surv.*, vol. 37, no. 4, pp. 316–344, 2005.

[9] M. Fowler, "Domain specific languages," http://martinfowler.com/dslwip/, 2009.

[10] G. Kiczales, J. D. Rivieres, and D. G. Bobrow, *The Art of the Metaobject Protocol*. Cambridge, MA, USA: MIT Press, 1991.

[11] J.-M. Favre, "Languages evolve too! changing the software time scale," in *IWPSE '05: Proceedings of the Eighth International Workshop on Principles of Software Evolution*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 33–44.

[12] Ľ. Wassermann, "Metaobject protocol as tool for language evolution," *Journal of Computer Science and Control Systems*, vol. 2, no. 1, pp. 79–82, 2009.

[13] S. Vermolen, "Software language evolution," in *WCRE '08: Proceedings of the 2008 15th Working Conference on Reverse Engineering*. Washington, DC, USA: IEEE Computer Society, 2008, pp. 323–326.

[14] S. Vermolen and E. Visser, "Heterogeneous coupled evolution of software languages," in *MoDELS '08: Proceedings of the 11th international conference on Model Driven Engineering Languages and Systems*. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 630–644.

[15] G. de Geest, S. Vermolen, A. van Deursen, and E. Visser, "Generating version convertors for domain-specific languages," in *WCRE '08: Proceedings of the 2008 15th Working Conference on Reverse Engineering*. Washington, DC, USA: IEEE Computer Society, 2008, pp. 197–201.

[16] M. Herrmannsdoerfer, S. Benz, and E. Jürgens, "Cope - automating coupled evolution of metamodels and models," in *ECOOP*, 2009, pp. 52–76.

[17] D. Balasubramanian, T. Levendovszky, A. Narayanan, and G. Karsai, "Continuous migration support for domain-specific languages," in *OOPSLA '09: Proceedings of The 9th OOPSLA Workshop on Domain-Specific Modeling*. New York, NY, USA: ACM, 2009.

[18] B. Gruschko, D. S. Kolovos, and D. F. Paige, "Towards synchronizing models with evolving metamodels," in *Workshop on Model-Driven Software Evolution at CSMR 2007*, 2007.

[19] E. Jürgens and M. Pizka, "The language evolver Lever," in *Sixth Workshop on Language Descriptions, Tools and Applications – LDTA 2006*, Vienna, Austria, 2006.

[20] M. Pizka and E. Jürgens, "Automated language evolution," in *First IEEE/IFIP International Symposium on Theoretical Aspects of Software Engineering*. Shanghai, China: IEEE Computer Society, Jun. 2007, pp. 305–315.

[21] M. Forgáč, "Metóda kombinovanej dynamickej modifikácie programov a jazykov," Ph.D. dissertation, Technická univerzita v Košiciach, 2009.

[22] T. Parr, *The Definitive ANTLR Reference: Building Domain-Specific Languages*. Pragmatic Bookshelf, 2007.

Composition and Integration of Domain-Specific Languages into Development Process

Ľubomír WASSERMANN

Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

Lubomir.Wassermann@tuke.sk

***Abstract*—Language-oriented programming and the usage of domain-specific languages is promising approach to increase productivity of the development of software systems. This is achieved by raising the level of abstraction and usage of effective generative techniques. Domain-specifing languages are aimed to solve problems from specific domain. To develop large software system, developers are using more domain-specific languages. To preserve their advantage of increased productivity, they need to be properly integrated and used in development process. This paper proposes the mechanism for composition and integration of domain-specific languages into common programming environments.**

***Keywords*—software engineering, domain-specific languages, language-oriented programming, integration and composition of languages**

I. Introduction

Since their creation, programming languages have undergone dynamic evolution. From usage of Assembler it gradually moved to usage of the third-generation language (3GL languages) which represented a significant step in terms of productivity. Significant increase in productivity was due to increased abstraction of 3GL languages over Assembler. This enabled programmers to create a program with the same functionality easier and with shorter and more readable source code. Next progress in programming languages and switching to another paradigm has not contributed to increase of productivity in such measure.

During the development of a software system are meeting two domains. Problem domain, which is represented by domain expert[1]. On the other hand, the solution domain, which is represented by a software system to be developed itself.

Description of the problem called requirements specification is expressed in the language of the problem domain. The language of the solution domain is a programming language chosen by the programmer. It represents a tool which can describe the solution of the problem from particular domain in a form understandable to computer. Since these languages are very different, finding a solution to the problem is a complex process. The actual creation of a software system can thus be defined as the transformation of concepts of the problem domain into concepts of the solution domain.

Nowadays, software system developers are mainly using general-purpose languages (GPL) which are intended to solve various problems in any domain. But this widens the gap between problem and solution domain due to differences of domain languages which results in communication barrier between domain expert and system developer. A possible solution is to bring the problem and solution domains and their respective languages closer. This is one of the goals of *language-oriented programming* (LOP) and its approach.

[1]*Domain expert is a person who has practical experience in the domain of the problem and the necessary knowledge in this domain needed to solve the problem.*

II. Language-oriented programming

Term *language-oriented programming* was used and explained in detail in the paper of M. P. Ward [1]. LOP raises the abstraction level with help of small languages with a specific purpose created by programmers. These languages are called domain-specific languages and are tailored to a particular domain or particular problem. The great advantage of LOP is that it allows the programmer to use and work with concepts from the domain of the solved problem without being forced to transform his ideas and solutions to concepts and constructs used by general-purpose languages (e.g. classes, functions, branching, cycles, etc.) [2].

Since the real software systems are large and include multiple domains, usage of the principles of LOP for the development of such systems is appropriate. Ideally, to solve problems in each domain, DSL should be created, and the resulting software system would be composed of programs written in them. These are the main advantages of using the principles of LOP (Fig. 2) compared to using GPLs (Fig. 1).

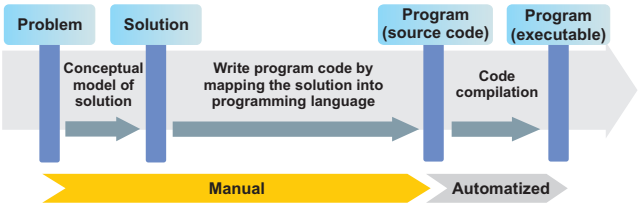


Fig. 1. Programming with general-purpose language

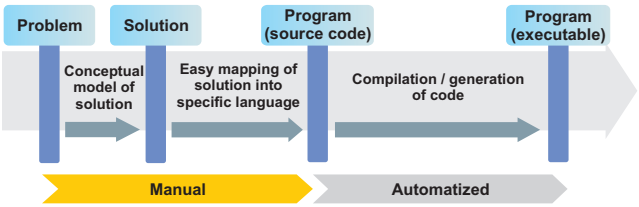


Fig. 2. Language oriented programming with domain-specific languages

A. Domain-specific languages

Domain-specific languages (DSL) are computer languages aimed at solving problems from particular domain. They are essentially a small languages with limited expressiveness, offering programmer limited set of concepts, notations and abstractions from the target domain [3]. Thanks to the higher level of abstraction by using the domain concepts, the productivity of programmers and their communication with the domain experts is improved when compared to using GPL [4]. When design of DSL is simple and suitable for the domain, it is even possible that DSL can be used by the domain expert or end user (end-user programming [3]).

In combination with an application library, any GPL can be perceived as DSL. The librarys Application Programmers Interface (API) constitutes a domain-specific vocabulary of class, method, and function names that becomes available by object creation and method invocation to any GPL program using the library [4].

The main advantages of DSLs are [5], [6], [3]:

- *Capturing domain knowledge* - DSL uses concepts of the problem domain and defines their mutual relations. DSL represents the domain knowledge and allows their reuse in other projects and maintains them in a consistent state.
- *Participation of domain expert* - Thanks to the use of concepts from the language of domain, the domain expert can easily understand, validate, edit or even write DSL program itself. Improving communication with domain expert and his closer involvement in the software development process helps prevent errors in the design and implementation of requirements.
- *Improving productivity* - DSLs improve productivity of programmers thanks to the use of higher level abstractions. DSL programs are shorter, more readable, easier to mantain and less prone to errors. Since language of problem domain is closer to the language of the solution domain, write solution to the problem is easier as opposed to the use of GPL.
- *Reusability* - When DSL is well designed, its reusability presents an advantage. Due to increased costs associated with implementation of DSL, it is not suitable to create and use DSL for single solutions. But coping with more same or similar problems, use of appropriate DSL brings great benefits.
- *Changing Paradigms* - Using DSL to solve problems can help to overcome deficiencies of GPL and find simpler solution of the problem with appropriate DSL, which is designed for this domain [7].

On the other hand, the use of DSL also brings disadvantages. However, most of these shortcomings can be addressed to the fact that use of DSL is not yet established and most programmers do not know how to properly design and implement DSL or do not see the importance of DSL on a wider scale in software development [7]. The disadvantages of DSL are:

- *Learning curve* - Every new technology brings additional costs with it because it is necessary to get familiar with its principles and gain experience how to apply them correctly. Gained knowledge can be subsequently applied to other projects but the initial use of new technology is connected with higher costs [7].
- *Cost of building* - Despite the knowledge of DSL technology, additional costs for creating the DSL are needed. Good design, implementation and maintenance of DSL requires knowledge and experience in language construction. When designing a DSL it is necessary to keep in mind that DSL should have only limited expressiveness [7].
- *Lack of tool support* - There is no existing widespread, fully functional, integrated environment to support the construction and use of DSL in the development of software systems [5]. Several research groups and development teams are working on the environment with such a support as described in [2], [8], [7].
- *Difficulty of migrating DSL programs* - When it is necessary to change DSL due to new requirements, it rises a problem of migration and compatibility issues of already created programs. Due to the absence of a tool support for the evolution of DSL, it is necessary to ensure the migration or to avoid it with support of older versions of DSL [7].
- *Loss of domain specificity* - During its life cycle, DSL is a subject to a number of changes that leads to the evolution of DSL. With longer use of DSL, there is a risk of slipping to generality and specialized DSL can become language with expressiveness of GPL [3], [9].

III. Composition of domain-specific languages

As mentioned before, real software systems are large and include multiple domains. *Multiparadigm programming* requires that each problem should be dealt with the most suitable language [10], [11]. According to this definition, to solve problems in each domain, DSL should be created, and the resulting software system would be composed of programs written in them. To preserve their advantage of increased productivity, they need to be properly integrated and used in development process.

System of composition and integration of multiple DSLs into the development process should use common tools used by programmers and in this way avoid a need for special environment or additional library. It should be easily used with tools offered by IDE (Integrated Development Environment) and constructs of programming language.

Ideally, the programmer could choose which language to use based on the problem he is currently solving. This leads to combination of GPL code with code of multiple specialized DSLs to create a complex software system. Composition mechanism has to ensure correct compilation and/or interpretation of embedded DSL code in the project.

IV. System for composition of DSLs

In this paper is proposed system for composition and integration DSLs into software development process. Programming language Java is used as a general-purpose language of proposed system but used constructs and tools make .NET implementation possible, too.

The principle of composition is that each method can be implemented in selected DSL or in Java. Language selection is based on decoration of method with Java annotation that specifies the name of used language for implementation (Listing. 1). If no annotation is present the method is implemented in Java and no special processing takes place. Java class can then contain methods annotated with different DSLs and methods

without annotations with Java code together. DSL code of annotated methods is stored in external file with same name as method.

```
@Language("UIDsl")
public void drawUI() {
  ...
}
```

Listing 1. Annotated method with specified DSL

The DSL composition system itself is not responsible for providing tools for processing DSLs (language processor, generator, etc.) but its task is to compose code of different languages together, transparently for the programmer. The architecture of the DSL composition system (Fig. 3) consists of three main parts.

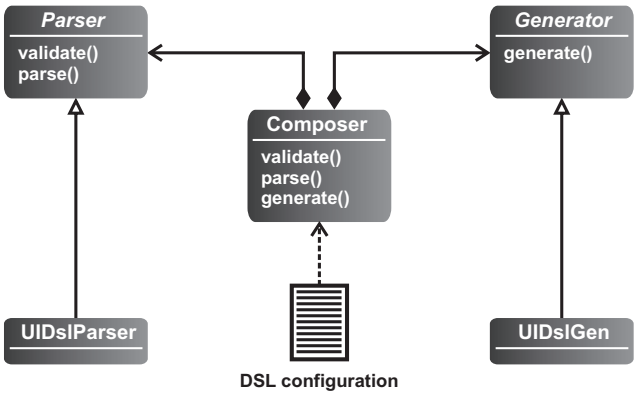


Fig. 3. Architecture of DSL composition system

Configuration file contains list of DSLs that can be used in composition. For each registered DSL are specified names of classes representing DSL parser and generator used to process DSL programs and identifier of DSL used in the method annotations.

To ensure compatibility between DSL processing tools which is needed to composition, system contains abstract classes *Parser* and *Generator* to be inherited and implemented for each DSL to be composed. This also allows for easier extendability of the system with support for additional DSLs.

The core of whole composition mechanism is the *Composer* class which is working as preprocessing before starting the whole system. First, *Composer* initializes all needed tools for the registrated DSLs listed in configuration file to be able process any DSL program. Then looks through all the methods with help of reflection to find annotated methods and corresponding DSL. For each annotated method, *Composer* gets DSL code from external file and calls corresponding DSL parser which validates and processes the DSL code. Result of the parsing process is an object model which serves as memory representation of DSL program. In the next step, object model is input for the generator which generates target Java code with same semantics as DSL code.

Correct execution of generated method from DSL code is achieved by generating code into the common class with the same name as in source class. Body of the annotated methods representing DSL code contains call of the generated method by dynamic invocation (Listing. 2). Use of dynamic invocation of the method is necessary because generated method is not available at compile time but generated at runtime. After the processing of annotated methods, the system itself is started.

```
@Language("UIDsl")
public void drawUI() {
  Class gen = Class.forName("sk.tuke.Generated");
  Method m = gen.getMethod("drawUI", null);
  m.invoke(new sk.tuke.Generated());
}
```

Listing 2. Example of dynamic invocation of generated method

With this approach is ensured the integration of DSL code with GPL Java code transparently from the programmer's point of view.

V. Conclusion

Proposed DSL composition system helps the programmer to use multiple DSLs along with chosen GPL in development process of software system. Composition and integration is based on annotated methods containing DSL code which are processed by corresponding parser and generator. Generated method in GPL code is then called by dynamic invocation.

Proposed system can be extended with new functionality e.g. support for interpretation of DSL code as an alternative to compilation/generation of target code, passing parameters usable in DSL code and combination of more DSLs and GPL code within one method. Also, the integration of DSL composition system with used IDE can be improved by creation of a plugin used to edit external DSL code within IDE with syntax highlighting and code completion.

Acknowledgment

This work is the result of the project implementation: Centre of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120020) supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] M. P. Ward, "Language-oriented programming," *Software — Concepts and Tools*, vol. 15, no. 4, pp. 147–161, 1994.
[2] S. Dmitriev, "Language oriented programming: The next programming paradigm," http://www.onboard.jetbrains.com/articles/04/10/lop/, 2004.
[3] A. van Deursen, P. Klint, and J. Visser, "Domain-specific languages: an annotated bibliography," *SIGPLAN Not.*, vol. 35, no. 6, pp. 26–36, 2000.
[4] M. Mernik, J. Heering, and A. M. Sloane, "When and how to develop domain-specific languages," *ACM Comput. Surv.*, vol. 37, no. 4, pp. 316–344, 2005.
[5] D. Spinellis, "Notable design patterns for domain specific languages," *Journal of Systems and Software*, vol. 56, no. 1, pp. 91–99, Feb. 2001. [Online]. Available: http://www.spinellis.gr/pubs/jrnl/2000-JSS-DSLPatterns/html/dslpat.html
[6] A. van Deursen and P. Klint, "Little languages: little maintenance?" *Journal of Software Maintenance*, vol. 10, no. 2, pp. 75–92, 1998.
[7] M. Fowler, "Domain specific languages," http://martinfowler.com/dslwip/, 2009.
[8] J. Greenfield, K. Short, S. Cook, and S. Kent, *Software Factories: Assembling Applications with Patterns, Models, Frameworks, and Tools*. Wiley, August 2004.
[9] J. Gray, K. Fisher, C. Consel, G. Karsai, M. Mernik, and J.-P. Tolvanen, "Dsls: the good, the bad, and the ugly," in *OOPSLA Companion '08: Companion to the 23rd ACM SIGPLAN conference on Object-oriented programming systems languages and applications*. New York, NY, USA: ACM, 2008, pp. 791–794.
[10] V. Karakoidas and D. Spinellis, "J%: Integrating domain specific languages with Java," in *PCI 2009: Proceedings of 13th Panhelenic Conference on Informatics*, V. Chrissikopoulos, N. Alexandris, C. Douligeris, and S. Sioutas, Eds. IEEE Computer Society, Sep. 2009, pp. 109–113.
[11] J. Placer, "Multiparadigm research: a new direction of language design," *SIGPLAN Not.*, vol. 26, no. 3, pp. 9–17, 1991.

Sources of Project Knowledge and Their Integration Into Software Architecture

[1]*Peter Žárský*

[1]Department of Computers and Informatics, FEI TU of Košice, Slovak Republic

[1]peter.zarsky@tuke.sk

***Abstract* — This paper presents main advantages and disadvantages of two main sources of project knowledge, which can be integrated into software architecture. The first part describes the concept of software architecture with integrated project knowledge. The second describes reverse engineering and project documentation as source of project knowledge.**

***Keywords* — knowledge, project documentation, reverse engineering, software architecture.**

I. INTRODUCTION

Life cycle of information system doesn't end with its deployment. IEEE Standard 1219 defines software maintenance as: "The modification of a software product after delivery to correct faults, to improve performance or other attributes, or to adapt the product to a modified environment." [1]

Reasons of software maintenance can be categorized into four classes:

- Adaptive – changes in the software environment,
- Perfective – new user requirement,
- Corrective – fixing errors,
- Preventive – prevent problems in future.

With rising complexity of information system declines its understandability and become less easy to handle it. Project knowledge have therefore important role in management, maintenance and the modification of complex software systems. Easier maintenance means less cost. "Maintenance typically consumes 40 percent to 80 percent of software costs. Therefore, it is probably the most important life cycle phase of software." [2]

II. MODEL DRIVEN MAINTENANCE

Program comprehension, impact analysis and regression testing are the most challenging problems of software maintenance. [3]

Model-driven maintenance (MDM) [3, 4, 5, 6] process is one useful aspect of knowledge-based software life cycle oriented to better usability of all analysis, design and implementation models in maintenance of systems. It tries to streamline and speed up the maintenance process of software system together with keeping system consistency with the help of knowledge acquired from software system's abstract models.

Models are cornerstones of MDM, and their consistency with other artifacts of the system, especially with the code is crucial. This is the reason for the proposal of a modified system architecture, which supports a conjunctive preservation of the program code and its models. [6]

III. CONCEPT OF KNOWLEDGE BASED ARCHITECTURE

Concept of knowledge based architecture described in [7], integrates project knowledge directly into structure of software architecture and binds them with other artifacts of information system.

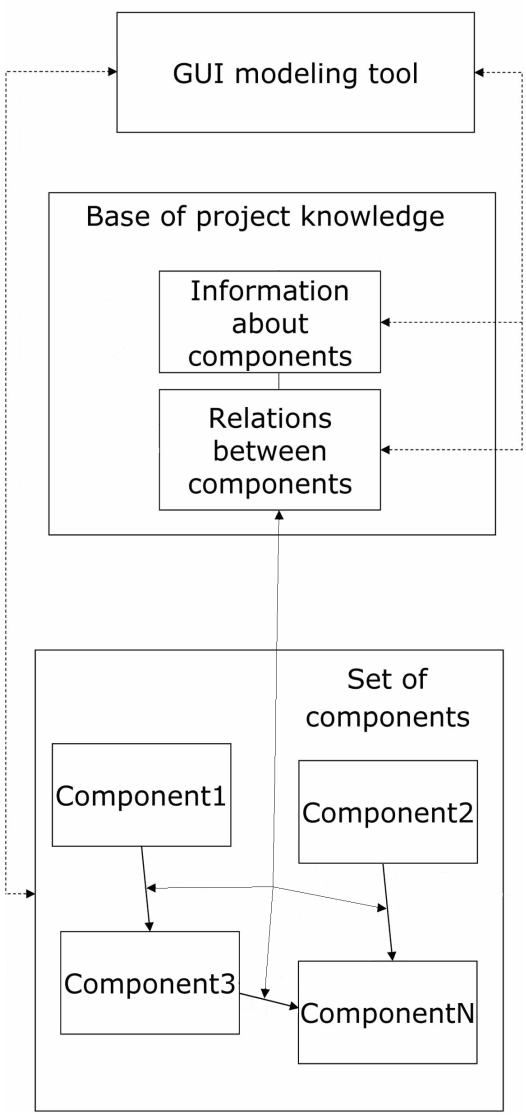


Fig. 1. Structure of architecture described in concept consisting from three parts: set of components, modeling tool, and base

Structure of this concept consists of three main parts. Set of components, that is the core of the information system, base of project knowledge and GUI (Graphical User Interface) based tool. This provides access to the base of the project knowledge for the designer or user, as shown on Fig. 1.

The advantage of this concept is that knowledge in the knowledge base is bound with other artifacts of information system. It ensures consistency between software artifacts, knowledge base and models generated from this knowledge, which can be used during software maintenance.

In advantage, changes of information system that doesn't require development of new components or change in implementation of used components can be done by simple change of model and updating of knowledge base.

IV. Reverse engineering

Chikofsky and Cross defined reverse engineering [8] as "analyzing a subject system to identify its current components and their dependencies, and to extract and create system abstractions and design information". Reverse engineering has been traditionally viewed as a two step process: information extraction and abstraction.

Research of reverse engineering [8, 9, 10, 11] has produced a number of capabilities for analyzing code, including subsystem decomposition, concept synthesis, design, program and change pattern matching, program slicing and dicing, analysis of static and dynamic dependencies, object-oriented metrics, and software exploration and visualization.

Main advantage of the knowledge mining at code level is the best accuracy of gained knowledge.

On the other hand, code does not contain all the information that is needed. Typically, knowledge about architecture and design tradeoffs, engineering constraints and the application domain can be found in the project documentation or in the worst case, in the mind of software engineer only.

V. Documentation

The documentation of software system is a set of all artifacts whose purpose is to communicate information about the software system to which it belongs [11, 12]. It contains information about the system in more abstract level and therefore it is easier to analyze it for developer comparing to the software code.

Another advantage of the knowledge mining from the documentation is that it also contains information that cannot be found in the source code.

Main problem of the documentation as a source of knowledge about software system is the possibility of inconsistency between the documentation and other artifacts of software system. CASE systems have been helping designers and programmers to design, build and maintain large information systems for many years. Despite this fact, incorrect use of CASE tools leads to inconsistent state between abstract models and real state of the information system.

VI. Conclusion

Due to character of knowledge gained from both, reverse engineering and documentation, the best way of knowledge mining seems to be its combination. Reverse engineering at the code level is necessary because components in the described concept of architecture [7] need to be bound with knowledge base at code level to ensure consistency between these artifacts. Knowledge gained from both sources, reverse engineering and verified knowledge from documentation, provide more complex view about software system. Accurate knowledge about software system shortens the time needed for maintenance, prevents spreading of new bugs into the software and therefore lowers the costs.

Acknowledgment

This work is the result of the project implementation: Centre of Information and Communication Technologies for Knowledge Systems (ITMS project code: 26220120020) supported by the Research & Development

Operational Programme funded by the ERDF; and also is supported by VEGA Grant project No. 1/0350/08: "Knowledge-Based Software Life Cycle and Architectures".

References

[1] IEEE Std. 1219: Standard for Software Maintenance, Los Alamitos CA., USA, IEEE Computer Society Press, 1993.

[2] R.S. Pressman: Software Engineering: A Practitioner's Approach, 7th ed. McGraw Hill, 2005.

[3] Havlice Z., Kunštár J., Adamuščínová* and Pločica O.: Knowledge in Software Life Cycle, SAMI 2009 Proceedings, 7th International Symposium on Applied Machine Intelligence and Informatics, Herľany, Slovakia, 30.-31.1.2009, 2009, IEEE Catalog Number CFP0908E-CDR, ISBN 978-1-4244-3802-0

[4] Kunštár,J., Adamuščinová, I., Havlice, Z.: The use of development models for improvement of software maintenance, Acta Universitatis Sapientiae, Informatica, 1,1, 2009, ISSN 1844-6086.

[5] Kunštár, J., Adamuščinová,I., Havlice,Z.: Model-Driven Life Cycle, CSE'2008 International Scientific Conference on Computer Science and Engineering, High Tatras – Stará Lesná, Slovakia, Semtember 24-26, 2008, ISBN 978-80-8086-092-9.

[6] Samuelis L., Havlice Z., Telepovská H., Kunštár J., Adamuščinová I., Pločica O., Varga M., Železník O., Révés M., Huňady M.: Software Processes Based on Knowledge, Computer Science and Technology Research Survey, Košice, Department of Computers and Informatics, FEII TU of Košice, 2008, 1, 3, pp. 11-18, ISBN 978-80-8086-100-1

[7] Žárský P., Lakatoš M., Havlice Z.: The Role of the Project-Knowledge in IS Development, Journal of Computer Science and Control Systems, Academy of Romanian Scientists, University of Oradea, Faculty of Electrical Engineering and Information Technology, Vol. 2, Nr. 2, 2009, pp. 80-83, ISSN 1884-6043

[8] Chikofsky E., Cross J.: Reverse engineering and design recovery: A taxonomy. IEEE Software, 7(1):13–17, January 1990.

[9] Muller H. A., Jahnke J. H., Smith D. B., Storey M.-A., Tilley S. R., Wong K.: Reverse engineering: a roadmap. , ICSE '00: Proceedings of the Conference on The Future of Software Engineering. New York, NY, USA: ACM, 2000, pp. 47-60.

[10] Tilley S. R.: The Canonical Activities of Reverse Engineering, Baltzer Science Publishers, The Netherlands, February 2000.

[11] Canfora G., Harman , Penta M.: New Frontiers of Reverse Engineering, 2007 Future of Software Engineering, p.326-341, 2007.

[12] Forward A. and Lethbridge T. C., "The relevance of software documentation, tools and technologies: a survey," presented at ACM Symposium on Document Engineering, 2002.

Author's index

10th Scientific Conference of Young Researchers
of Faculty of Electrical Engineering and Informatics
Technical University of Košice

Proceedings from Conference

Published: Faculty of Electrical Engineering and Informatics
Technical University of Košice
I. Edition, 381 pages, the number of CD Proceedings: 120 pieces

Editors: prof. Ing. Alena Pietriková, PhD.
Ing. Jana Modrovičová

ISBN 978-80-553-0423-6