SCIENCE & MILITARY

No 1 | Volume 5 | 2010

Dear readers,

You are about to read the first issue of the Science & Military journal, volume 5, which, despite being one of the „youngest" periodicals, has been receiving increasing numbers of positive feedback from its readers.

Since the very beginning, our aim has been to publish attractive and high-quality scientific journal on military science, accessible to the expert public not only in Slovakia but also abroad. We are aware of the fact that the high-quality scientific journal cannot be created overnight; therefore, we are doing our best to constantly improve magazine articles and layouts to make the journal more interesting and enjoyable to you – our readers.

2009 was no exception and some changes were made in terms of the structure of the journal. Scientific papers submitted for publication must be in English only. International Editorial Board consists of experts in the field of Science and Research who are solely committed to the exploration for high-quality research papers that are suitable for publication in the Science & Military journal. They will work with the editor to achieve objectives of the journal; encourage developers to contribute their articles and they will also vote on the selection of high quality papers. In 2009, the Science & Military journal has been added to ProQuest Central full-text database and at the same time it has been accepted to take part in Thomson Reuters Journal Selection Process for a period of two years during which the journal will be evaluated for possible coverage in Web of Science. However, adding the journal to the Current Contents database remains the main long-term aim of the Editorial Board.

We have got a long and difficult way ahead of us but as an African proverb says „Smooth seas do not make skilful sailors", and therefore, we are ready to overcome all kinds of obstacles in order to achieve our goal. We will continue our efforts to make this journal a high-quality scientific journal that would take a rightful place among other significant scientific journals.

Dear readers, the first 2010 issue contains 16 new and no doubt interesting scientific articles submitted by researchers and experts from research institutes, Academies and Universities from Slovakia and also from foreign countries.

I would like to specifically point out to you the article submitted by Assoc. Prof. Ferdinand Chovanec, CSc. from the Armed Forces Academy in Liptovský Mikuláš. Assoc. Prof. Chovanec was awarded ***„Prize for three-year scientific response"*** *in technical science and geosciences* ***"as a public token for significant scientific response during the previous three years"****. During these years, Assoc. Prof. Chovanec and Assoc. Prof. František Kôpka, CSc. developed mathematical model to be applied in quantum physics and received scientific response from the International scientific community. Up till now they have received more than 250 scientific responses to their firs article on the topic, those including quotations in scientific papers, articles and studies.*

Lastly, allow me to extend my thanks to all authors, reviewers and editorial board members. Thanks to their efforts we were able to publish yet another issue of the Science and Military Journal.

I hope the reading of Science and Military Journal will be an enjoyable and worthwhile experience for you. I look forward to your responses and contributions to our next issue.

Assoc. Prof. Dipl. Eng. Pavel NEČAS, PhD.
Chairman of the Editorial Board

Reviewers

Assoc. Prof. Dipl. Eng. František **ADAMČÍK**, PhD.	Technical University of Košice (SVK)
Prof. Dipl. Eng. Mircea **BOSCOIANU**, Ph.D.	Air Force Academy of Brasov (RO)
Assoc. Prof. Dipl. Eng. Peter **DROPPA**, PhD.	Armed Forces Academy Liptovský Mikuláš (SVK)
Prof. Dipl. Eng. Milan **DŽUNDA**, CSc.	Technical University of Košice (SVK)
PhDr. Libor **FRANK,** Ph.D.	University of Defence Brno (CZE)
Assoc. Prof. JUDr. MUDr. Batyr **CHALMURADOV**, PhD.	National Aviation University Kiev (UKR)
Brig.Gen. Assoc.Prof.Dipl.Eng. Miroslav **KELEMEN**, PhD.	Armed Forces Academy Liptovský Mikuláš (SVK)
Prof. Dipl. Eng. Ján **KOLLÁR**, CSc.	Technical University of Košice (SVK)
Col. (GS) Dipl. Eng. Zbyšek **KORECKI**, Ph.D.	University of Defence Brno (CZE)
Prof. Dipl. Eng. Tobiáš **LAZAR**, DrSc.	Technial University of Košice (SVK)
Prof. Dipl. Eng. Igor **MOKRIŠ**, CSc.	Slovak Academy of Sciences in Bratislava (SVK)
Col. (GS, ret.) Assoc. Prof. Dipl. Eng. Pavel **NEČAS**, PhD.	Armed Forces Academy Liptovský Mikuláš (SVK)
Assoc. Prof. Dipl. Eng. Dušan **NEŠTRÁK**, CSc.	Technical University of Košice (SVK)
Prof. Dipl. Eng. Jozef **POVAŽAN**, CSc.	Technical University of Košice (SVK)
Prof. Dipl. Eng. Václav **PŘENOSIL**, CSc.	University of Defence Brno (CZE)
Prof. Dipl. Eng. Vladimír **SEDLÁK**, PhD.	The University of Security Management in Košice (SVK)
Prof. Dipl. Eng. Jiří **STODOLA, DrSc.**	University of Defence Brno (CZE)
Prof. Dipl. Eng. Ladislav **ŠIMÁK**, PhD.	University of Žilina (SVK)
Assoc. Prof. Dipl. Eng. Miroslav **ŠKOLNÍK**, PhD.	Armed Forces Academy Liptovský Mikuláš (SVK)
Assoc. Prof. PhDr. František **ŠKVRNDA**, CSc.	University of Economics in Bratislava (SVK)
BG (ret.) prof. Dipl. Eng. Rudolf **URBAN**, CSc. Dr. h. c.	University of Defence Brno (CZE)
RNDr. Šárka **VAVREČKOVÁ**, PhD.	The Silesian University in Opava (CZE)
Assoc. Prof. Dipl. Eng. Vladimír **VRÁB,** CSc.	University of Defence Brno (CZE)

CANTOR SETS

Ferdinand CHOVANEC

Abstract: The Cantor set is an interesting example of an uncountable set of measure zero and has many interesting properties and consequences in the fields of set theory, topology, and fractal theory. The principal aim of this paper is to introduce a generator of finite subsets of the basic Cantor (ternary) set and its generalization to the Cantor *n*-ary set. We compute the fractal dimension of these Cantor sets.

Keywords: Fractal, Cantor set, fractal dimension.

1 INTRODUCTION

Two years ago we have gained the grant project of Ministry of Defence of the Slovak Republic AGA-01-2008 „*Statistical analysis of the influence of the semiconductor system interface nanoroughness on its optical properties*“. In this project we have used fractal approach to the study of the surface of solid materials (see [5], [6], [7]). In this study we used the multifractal singularity spectrum function $f(\alpha)$ to describe the development of the surface fractal properties. We compared experimental $f(\alpha)$ curves with theoretical singularity spectra, obtained by computer simulation of surface structure based on Cantor numbers properties. This required the use of the Cantor set with the cardinality more than 10^5. Moreover, we applied different types of Cantor sets.

In the following paper we introduce a generator of the basic Cantor (ternary) set and the generalized Cantor (*n*-ary) set. We compute the Hausdorff-Besicovitch fractal dimension of the Cantor sets.

2 THE CANTOR SETS

The Cantor (ternary) set was first published in 1883 by German mathematician Georg Cantor [1]. The Cantor set plays a very important role in many branches of mathematics, above all in set theory, chaotic dynamical systems and fractal theory.

2.1 The Cantor ternary set

The basic Cantor (ternary) set is a subset of the interval $[0,1]$ and has many definitions and many different constructions. Although Cantor originally provided a purely abstract definition, the most accessible is the „middle-thirds“ or ternary set construction. Begin with the closed real interval $I_0 = [0,1]$ and divide it into three equal subintervals. Remove the central open interval $\left(\frac{1}{3},\frac{2}{3}\right)$ such that

$$I_1 = [0,1] - \left(\frac{1}{3},\frac{2}{3}\right) = \left[0,\frac{1}{3}\right] \cup \left[\frac{2}{3},\frac{3}{3}\right].$$

Next, subdivide each of these two remaining intervals into three equal subintervals and from each remove the central third and continue in the previous manner.

$$I_2 = \left(\left[0,\frac{1}{3}\right] - \left(\frac{1}{9},\frac{2}{9}\right)\right) \cup \left(\left[\frac{2}{3},\frac{3}{3}\right] - \left(\frac{7}{9},\frac{8}{9}\right)\right) =$$

$$= \left[0,\frac{1}{9}\right] \cup \left[\frac{2}{9},\frac{3}{9}\right] \cup \left[\frac{6}{9},\frac{7}{9}\right] \cup \left[\frac{8}{9},\frac{9}{9}\right]$$

In this way we obtain a sequence of closed intervals – one in the zero step, two after the first step, four after the second step, eight after the third step, etc. (2^k intervals of length $\left(\frac{1}{3}\right)^k$ after the k^{th} step). This process is visible in the Figure 1.

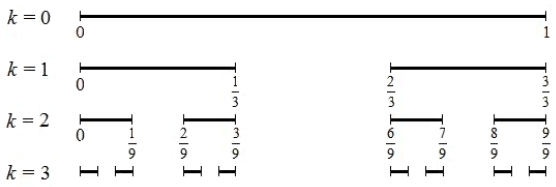


Figure 1 Initial steps of the construction of the Cantor ternary set

Finally, we define the Cantor ternary set $\mathbb{C}(3)$ as follows:

$$\mathbb{C}(3) = \bigcap_{k=0}^{\infty} I_k.$$

This construction does not provide a sufficient view of elements of the Cantor set. For these reasons we propose the construction of Cantor numbers, where subsets $\mathbb{C}_k(3)$ of the Cantor set $\mathbb{C}(3)$ are endpoints of the closed intervals creating I_k, $k = 0, 1, 2, \ldots$ (see Figure 1).
Denote by the symbol $|A|$ the cardinality of a set A. Then we have

$$\mathbb{C}_0(3) = \{0,1\} \text{ and } |\mathbb{C}_0(3)| = 2 = 2.2^0.$$

$$\mathbb{C}_1(3) = \left\{0, \frac{1}{3}, \frac{2}{3}, \frac{3}{3}\right\} = \bigcup_{j=0}^{3}\left\{\frac{j}{3}\right\}, |\mathbb{C}_1(3)| = 4 = 2.2^1.$$

$$\mathbb{C}_2(3) = \left\{0, \frac{1}{3^2}, \frac{2}{3^2}, \frac{3}{3^2}\right\} \cup \left\{\frac{6}{3^2}, \frac{7}{3^2}, \frac{8}{3^2}, \frac{9}{3^2}\right\} =$$

$$= \bigcup_{j=0}^{3}\left(\left\{\frac{j}{3^2}\right\} \cup \left\{\frac{j' \mid 6}{3^2}\right\}\right) = \bigcup_{j=0}^{3}\bigcup_{i_1=0}^{1}\left\{\frac{j' \mid 2(3i_1)}{3^2}\right\},$$

$$|\mathbb{C}_2(3)| = 8 = 2.|\mathbb{C}_1(3)| = 2.2.2^1 = 2.2^2.$$

$$\mathbb{C}_3(3) = \left\{0, \frac{1}{3^3}, \frac{2}{3^3}, \frac{3}{3^3}\right\} \cup \left\{\frac{6}{3^3}, \frac{7}{3^3}, \frac{8}{3^3}, \frac{9}{3^3}\right\} \cup$$

$$\cup \left\{\frac{18}{3^3}, \frac{19}{3^3}, \frac{20}{3^3}, \frac{21}{3^3}\right\} \cup \left\{\frac{24}{3^3}, \frac{25}{3^3}, \frac{26}{3^3}, \frac{27}{3^3}\right\} =$$

$$= \bigcup_{j=0}^{3}\bigcup_{i_1=0}^{1}\bigcup_{i_2=0}^{1}\left\{\frac{j + 2(3i_1 + 3^2 i_2)}{3^3}\right\},$$

$$|\mathbb{C}_3(3)| = 16 = 2.|\mathbb{C}_2(3)| = 2.2.2^2 = 2.2^3.$$

$$\vdots$$

$$\mathbb{C}_k(3) =$$

$$= \bigcup_{j=0}^{3}\bigcup_{i_1=0}^{1}\cdots\bigcup_{i_{k-1}=0}^{1}\left\{\frac{j + 2(3^{k-1}i_{k-1} + \cdots + 3i_1)}{3^k}\right\},$$

$$|\mathbb{C}_k(3)| = 2.|\mathbb{C}_{k-1}(3)| = 2.(2.2^{k-1}) = 2.2^k.$$

In this way we get a sequence $(\mathbb{C}_k(3))_{k=0}^{\infty}$ of finite subsets of the Cantor ternary set such that

$$\mathbb{C}_0(3) \subset \mathbb{C}_1(3) \subset \cdots \subset \mathbb{C}_k(3) \subset \mathbb{C}_{k+1}(3) \subset \cdots.$$

It is very well visible that the set $\mathbb{C}_k(3)$ is easily programmable for every nonnegative integer k.

If $|\mathbb{C}_k(3)| = x$ then $k = \frac{\ln x}{\ln 2}$. For example, if we need 1000 numbers of the Cantor set $\mathbb{C}(3)$, then it suffices to take the set $\mathbb{C}_9(3)$.

We should like to empasize that this method does not allow to construct any number of the Cantor set, only endpoints of the closed intervals remaining after removing the middle thirds. There are numbers in the Cantor ternary set which are not interval endpoints. One example of such number is $\frac{1}{4}$.

There is a natural question, how can we recognize elements (numbers) of the Cantor ternary set. It allows the triadic expansion of its numbers.

Let $x \in [0,1]$. Then its expansion with respect to base 3 (3-adic expansion) is given by the following expression

$$x = a_1 3^{-1} + a_2 3^{-2} + a_3 3^{-3} + \cdots + a_n 3^{-n} + \cdots$$

where $a_n \in \{0,1,2\}$ for very $n = 1,2,3,\ldots$. Then we write

$$x = 0.a_1 a_2 a_3 \ldots a_n \ldots|_3.$$

For example, $0.5 = 0.1111\ldots|_3$ $(= 0.\overline{1}|_3)$, because

$$1.3^{-1} + 1.3^{-2} + \cdots + 1.3^{-n} + \cdots = \sum_{n=1}^{\infty}\left(\frac{1}{3}\right)^n =$$

$$= \frac{1}{3}\sum_{n=1}^{\infty}\left(\frac{1}{3}\right)^{n-1} = \frac{1}{3}.\frac{1}{1-\frac{1}{3}} = \frac{1}{3}.\frac{1}{\frac{2}{3}} = \frac{1}{2}.$$

There are some numbers that have a terminating expansion and simultaneously an infinite expansion. Let us take $x = \frac{1}{3}$. Then $x = 0.1|_3$ and on the other hand we have

$$2.3^{-2} + \cdots + 2.3^{-n} + \cdots = 2\sum_{n=2}^{\infty}\left(\frac{1}{3}\right)^n =$$

$$= 2\left(\frac{1}{3}\right)^2\sum_{n=2}^{\infty}\left(\frac{1}{3}\right)^{n-2} = 2\left(\frac{1}{3}\right)^2\frac{1}{1-\frac{1}{3}} = \frac{1}{3}.$$

In ternary notation we have the similar equivalence that $0.1|_3$ equals $0.0222\ldots|_3$ $(= 0.0\overline{2}|_3)$. All numbers strictly between $\frac{1}{3} = 0.0\overline{2}|_3$ and $\frac{2}{3} = 0.2|_3$ must have a digit „1“ somewhere in the middle of the digit sequence. Therefore, these numbers are not in the Cantor ternary set.

In the general we can characterize any number of the Cantor ternary set in the following way.

Theorem 1 [10] *The Cantor ternary set is the set of numbers in [0, 1] for which there is a triadic expansion that does not contain the digit „1“.*

We are able easily to verify that $\frac{1}{4}$ = 0.02020202… (= $0.\overline{02}$) and hence $\frac{1}{4} \in \mathbb{C}(3)$.

From the mathematical point of view, the Cantor set has many interesting properties:

- The Cantor set is compact (i.e. closed and bounded).
- The Cantor set does not contain any open set.
- The Cantor set is perfect (and hence uncountable).
- The Cantor set has length zero.

We refer to readers the book [11] for detailed proofs of the above mentioned properties.

2.2 The Cantor quintuple set

Motivated by the ternary Cantor set $\mathbb{C}(3)$, we construct the Cantor quintuple set $\mathbb{C}(5)$. We begin with the closed real interval $I_0 = [0,1]$ again and divide it into five equal subintervals. Remove the open intervals $\left(\frac{1}{5},\frac{2}{5}\right)$ and $\left(\frac{3}{5},\frac{4}{5}\right)$ such that

$$I_1 = [0,1] - \left(\left(\frac{1}{5},\frac{2}{5}\right) \cup \left(\frac{3}{5},\frac{4}{5}\right)\right) =$$

$$= \left[0,\frac{1}{5}\right] \cup \left[\frac{2}{5},\frac{3}{5}\right] \cup \left[\frac{4}{5},\frac{5}{5}\right].$$

We subdivide each of these three remaining intervals into five equal subintervals and from each remove the second and fourth open subinterval, and continue in the previous manner. In this way we obtain a sequence of closed intervals I_k – one in the zero step, three after the first step, nine after the second step, etc. (3^k intervals of length $\left(\frac{1}{5}\right)^k$ after the k^{th} step). This process is visible in the Figure 2.

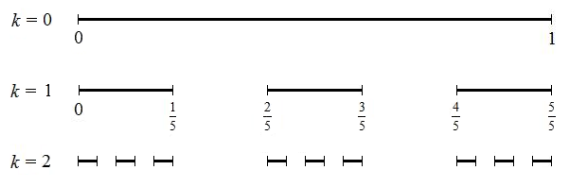


Figure 2 Initial steps of the construction of the Cantor ternary set

We define the Cantor quintuple set $\mathbb{C}(5)$ by the formula

$$\mathbb{C}(5) = \bigcap_{k=0}^{\infty} I_k.$$

Now we construct a generator of numbers of the set $\mathbb{C}(5)$. Let $\mathbb{C}_k(5)$ be sets of endpoints of closed intervals creating I_k, $k = 0, 1, 2, \ldots$ (see Fig. 2). Then

$$\mathbb{C}_0(5) = \{0, 1\} \text{ and } |\mathbb{C}_0(5)| = 2 = 2 \cdot 3^0.$$

$$\mathbb{C}_1(5) = \left\{0, \frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, \frac{5}{5}\right\} = \bigcup_{j=0}^{5} \left\{\frac{j}{5}\right\}$$

$$|\mathbb{C}_1(5)| = 6 = 3 \cdot |\mathbb{C}_0(5)| = 3 \cdot 2 \cdot 3^0 = 2 \cdot 3^1.$$

$$\mathbb{C}_2(5) = \left\{0, \frac{1}{5^2}, \frac{2}{5^2}, \frac{3}{5^2}, \frac{4}{5^2}, \frac{5}{5^2}\right\} \cup$$

$$\cup \left\{\frac{10}{5^2}, \frac{11}{5^2}, \frac{12}{5^2}, \frac{13}{5^2}, \frac{14}{5^2}, \frac{15}{5^2}\right\} \cup$$

$$\cup \left\{\frac{20}{5^2}, \frac{21}{5^2}, \frac{22}{5^2}, \frac{23}{5^2}, \frac{24}{5^2}, \frac{25}{5^2}\right\} = \bigcup_{j=0}^{5} \bigcup_{\ell_1=0}^{5} \left\{\frac{j + 2(5\ell_1)}{5^2}\right\}$$

$$|\mathbb{C}_2(5)| = 18 = 3 \cdot |\mathbb{C}_1(5)| = 3 \cdot 2 \cdot 3^1 = 2 \cdot 3^2.$$

$$\mathbb{C}_3(5) = \bigcup_{j=0}^{5} \bigcup_{\ell_1=0}^{5} \bigcup_{\ell_2=0}^{5} \left\{\frac{j + 2(5^2\ell_2 + 5\ell_1)}{5^3}\right\},$$

$$|\mathbb{C}_3(5)| = 54 = 3 \cdot |\mathbb{C}_2(5)| = 3 \cdot 2 \cdot 3^2 = 2 \cdot 3^3,$$

$$\mathbb{C}_k(5) =$$

$$= \bigcup_{j=0}^{5} \bigcup_{\ell_1=0}^{5} \cdots \bigcup_{\ell_{k-1}=0}^{5} \left\{\frac{j + 2(5^{k-1}\ell_{k-1} + \cdots + 5\ell_1)}{5^k}\right\},$$

$$|\mathbb{C}_k(5)| = 3 \cdot |\mathbb{C}_{k-1}(5)| = 3 \cdot (2 \cdot 3^{k-1}) = 2 \cdot 3^k.$$

We get a sequence $\left(\mathbb{C}_k(5)\right)_{k=0}^{\infty}$ of finite subsets of the Cantor quintuple set such that

$$\mathbb{C}_0(5) \subset \mathbb{C}_1(5) \subset \cdots \subset \mathbb{C}_k(5) \subset \mathbb{C}_{k+1}(5) \subset \cdots$$

If $|\mathbb{C}_k(5)| = x$ then $k = \frac{\ln x}{\ln 3}$. To have 1000 numbers of the Cantor set $\mathbb{C}(5)$ it suffices to take the set $\mathbb{C}_5(5)$.

There is a natural question: How can we characterize numbers of the Cantor quintuple set? Expected answer is – by means of their 5-adic expansion.

It is not difficult to verify that $\frac{1}{5} = 0.1|_5 = 0.0\overline{4}|_5$, $\frac{3}{5} = 0.3|_5 = 0.2\overline{4}|_5$ and for that reason all numbers strictly between $\frac{1}{5} = 0.0\overline{4}|_5$ and $\frac{2}{5} = 0.2|_5$, as well as between $\frac{3}{5} = 0.2\overline{4}|_5$ and $\frac{4}{5} = 0.4|_5$, must have digits „1“and „3“ somewhere in the middle of the digit sequence. Therefore, these numbers are not in the Cantor ternary set. For these reasons, the following assertion is true.

Theorem 2 *The Cantor quintuple set is the set of numbers in [0, 1] for which there is a 5-adic expansion that does not contain the digits „1“and „3“.*

2.3 The Cantor *n*-ary set

In this section we generalize the construction of the Cantor ternary and quintuple set.

Let $n = 2m + 1, m = 1, 2, 3, \ldots$. We start with the closed real interval $K_0 = [0,1]$ and divide it into n equal subintervals. Remove the open intervals $\left(\frac{1}{n}, \frac{2}{n}\right), \left(\frac{3}{n}, \frac{4}{n}\right), \ldots, \left(\frac{n-2}{n}, \frac{n-1}{n}\right)$ such that

$$K_1 = \left[0, \frac{1}{n}\right] \cup \left[\frac{2}{n}, \frac{3}{n}\right] \cup \ldots \cup \left[\frac{n-1}{n}, \frac{n}{n}\right].$$

We subdivide each of these $(m+1)$-remaining intervals into n equal subintervals and from each remove the 2nd, 4th, …, (2m)th open subinterval, and continue in the previous manner. In this way we obtain a sequence of closed intervals K_k – one in the zero step, $m+1$ after the first step, $(m+1)^2$ after the second step, etc. ($(m+1)^k$ intervals of length $\left(\frac{1}{n}\right)^k$ after the k^{th} step).

The Cantor n-ary set is defined by the formula

$$\mathbb{C}(n) = \bigcap_{k=0}^{\infty} K_k.$$

Now we construct a sequence $\left(\mathbb{C}_k(n)\right)_{k=0}^{\infty}$ of numbers of the the Cantor n-ary set such that

$$\mathbb{C}_0(n) = \{0, 1\} \text{ and } |\mathbb{C}_0(n)| = 2 = 2 \cdot (m+1)^0.$$

$$\mathbb{C}_1(n) = \left\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, \frac{n}{n}\right\} = \bigcup_{j=0}^{n} \left\{\frac{j}{n}\right\},$$

$$|\mathbb{C}_1(n)| = n + 1 = 2m + 2 = (m+1)|\mathbb{C}_0(n)| = 2(m+1)^1,$$

$$\mathbb{C}_k(n) = \bigcup_{j_0=0}^{n} \bigcup_{j_1=0}^{m-1} \ldots \bigcup_{j_{k-1}=0}^{m-1} \left\{\frac{j + 2(n^{k-1}j_{k-1} + \cdots + nj_1)}{n^k}\right\},$$

$$|\mathbb{C}_k(n)| = |\mathbb{C}_k(2m+1)| = (m+1)|\mathbb{C}_{k-1}(n)| = (m+1)2(m+1)^{k-1} = 2(m+1)^k.$$

If $|\mathbb{C}_k(n)| = x$ then $k = \frac{\ln \frac{x}{2}}{\ln(m+1)}$.

The sum of the lengths of the removed intervals is equal to 1, because

$$m\left(\frac{1}{n}\right) + m(m+1)\left(\frac{1}{n}\right)^2 + m(m+1)^2\left(\frac{1}{n}\right)^3 + \cdots + m(m+1)^{k-1}\left(\frac{1}{n}\right)^k + \cdots =$$

$$= \sum_{k=1}^{\infty} m(m+1)^{k-1}\left(\frac{1}{n}\right)^k = \frac{m}{n}\sum_{k=1}^{\infty}\left(\frac{m+1}{n}\right)^{k-1} = \frac{m}{n} \cdot \frac{1}{1 - \frac{m+1}{n}} = \frac{m}{n} \cdot \frac{n}{n - m - 1} = 1.$$

Videlicet, the Lebesque measure of the Cantor set $\mathbb{C}(n)$ is zero for every $n = 2m + 1, m = 1, 2, \ldots$.

3 CANTOR SETS AS FRACTALS

The Cantor set is the prototype of a fractal. A fractal is an object which appears self-similar under varying degrees of magnification. One of the typical features of fractals is their fractal dimension. The fractal dimension is essentially a measure of self-similarity (it is sometimes referred to as the similarity dimension). The fractal dimension is greater than the topological dimension. There are many specific definitions of fractal dimension. The basic type of fractal dimension is the Hausdorff-Besicovitch dimension, which is based on the definition of the Hausdorff measure [2]. One version of the Hausdorff-Besicovitch dimension is given by the formula

$$D = \frac{\log N}{\log \frac{1}{r}},$$

where N is the number of self-similar pieces and r is the contraction factor.

We note that there are several different ways of computing the fractal dimension (see [8], [9]).

Now we compute the fractal dimension of the Cantor sets. Let us assume the Cantor ternary set $\mathbb{C}(3)$. We have 2^k (self-similar) intervals of length $\left(\frac{1}{3}\right)^k$ after the k^{th} step, so $N = 2^k$ and $r = \left(\frac{1}{3}\right)^k$. Then

$$D(\mathbb{C}(3)) = \frac{\log 2^k}{\log 3^k} = \frac{k \log 2}{k \log 3} = \frac{\log 2}{\log 3} \approx 0.631.$$

In the case of the Cantor quintuple set $\mathbb{C}(5)$ we have 3^k (self-similar) intervals of length $\left(\frac{1}{5}\right)^k$ after the k^{th} step, therefore, $N = 3^k$ and $r = \left(\frac{1}{5}\right)^k$. Then

$$D(\mathbb{C}(5)) = \frac{\log 3^k}{\log 5^k} = \frac{\log 3}{\log 5} \approx 0.683.$$

In the case of the Cantor set $\mathbb{C}(n)$, $n = 2m + 1$, $m = 1, 2, 3, \ldots$, we have $(m+1)^k$ (self-similar)

intervals of the length $\left(\frac{1}{n}\right)^k$ after the k^{th} step, so $N = (m+1)^k$ and $r = \left(\frac{1}{n}\right)^k$. Then

$$D(\mathbb{C}(n)) = \frac{\log(m+1)^k}{\log(2m+1)^k} = \frac{\log(m+1)}{\log(2m+1)} < 1.$$

The Cantor set $\mathbb{C}(n)$ is an object with fractal dimensionality less than one, between a point (topological dimensionality of zero) and a line (topological dimensionality one), for every $n = 2m + 1$, $m = 1, 2, 3, \ldots$

Theorem 4 Let $\{\mathbb{C}(2m+1)\}_{m=1}^{\infty}$ be a sequence of Cantor sets. Then a sequence of their dimensions $\{D(\mathbb{C}(2m+1))\}_{m=1}^{\infty}$ is increasing and, moreover,

$$\lim_{m\to\infty} D(\mathbb{C}(2m+1)) = 1.$$

Proof We define three real functions

$$f:[1,\infty) \to R,\ f(x) = \log(x+1),$$
$$g:[1,\infty) \to R,\ g(x) = \log(2x+1),$$
$$h:[1,\infty) \to R,\ h(x) = \frac{\log(x+1)}{\log(2x+1)}.$$

To prove that the function h is increasing on the interval [1,1) we compute its first derivation.

$$h'(x) = \left(\frac{\log(x+1)}{\log(2x+1)}\right)' =$$

$$= \frac{\frac{(\log e)\log(2x+1)}{x+1} - \frac{(\log(x+1))\log e}{2x+1}}{(\log(2x+1))^2} =$$

$$= \frac{(\log e)[\log(2x+1)^{2x+1} - \log(x+1)^{x+1}]}{(\log(2x+1))^2(x+1)(2x+1)} =$$

$$= \frac{(\log e)\left[\log\frac{(2x+1)^{2x+1}}{(x+1)^{x+1}}\right]}{(\log(2x+1))^2(x+1)(2x+1)}.$$

Note that

$$\frac{\log e}{(\log(2x+1))^2(x+1)(2x+1)} > 0$$

for every x 2 [1,1).

We have

$$\log\frac{(2x+1)^{2x+1}}{(x+1)^{x+1}} = \log\frac{(2x+1)^{x+1}(2x+1)^x}{(x+1)^{x+1}} =$$

$$= \log\left(\left(\frac{2x+1}{x+1}\right)^{x+1}(2x+1)^x\right) =$$

$$= \log\left(1+\frac{x}{x+1}\right)^{x+1} + \log(2x+1)^x =$$

$$= (x+1)\log\left(1+\frac{x}{x+1}\right) + x\log(2x+1) > 0$$

for every x 2 [1,1).

We proved that $h'(x) > 0$ for every x 2 [1,1), so the function $h(x) = \frac{\log(x+1)}{\log(2x+1)}$ is increasing on [1,1) and hence, the sequence $\left\{\frac{\log(m+1)}{\log(2m+1)}\right\}_{m=1}^{\infty}$ is increasing too.

Let us calculate

$$\lim_{x\to\infty}\frac{\log(x+1)}{\log(2x+1)} = \lim_{x\to\infty}\frac{(\log(x+1))'}{(\log(2x+1))'} =$$

$$= \lim_{x\to\infty}\frac{\frac{\log e}{x+1}}{\frac{2\log e}{2x+1}} = \lim_{x\to\infty}\frac{2x+1}{2(x+1)} = \lim_{x\to\infty}\frac{2}{2} = 1.$$

Hence,

$$\lim_{m\to\infty} D(\mathbb{C}(2m+1)) = \lim_{m\to\infty}\frac{\log(m+1)}{\log(2m+1)} = 1.$$

5 CONCLUSION

The Cantor set has many interesting properties and consequences in the fields of set theory, topology, and fractal theory. An application of fractal theory to the theory of algebraic structures was presented on the Tenth International Conference on Fuzzy Sets Theory and Applications in Liptovský Ján (February 1–5, 2010) [3]. A fractal difference poset (a fractal D-poset, in short) was defined as a special pasting of MV-algebras [4]. In this sense, the Cantor fractal D-poset is the „0-1-pasting“ of MV-algebras.

This work has been supported in part by the Slovak Research and Development Agency under the contract No. APVV-0071-06, Grant VEGA 2/0032/09 and AGA-01-2008 Ministry of Defence of the Slovak Republic Grant Agency.

References

[1] CANTOR, G.: *Über unendliche, lineare Punkt-mannigfaltigkeiten V.* Mathematische Annalen 21, 1883, pp. 545-591.

[2] FRANCESCHETTI, G., RICCIO, D.: *Scatte ring, natural surfaces and fractals.* Elsevier Burlington USA, 2007.
[3] CHOVANEC, F., JUREČKOVÁ, M.: *Fractal difference posets.* Abstracts of the Tenth International Conference on Fuzzy Sets Theory and Applications, Liptovský Ján, February 1–5, 2010, (Klement, E. P., Mesiar, R., Struk, P., Drobná, E. – eds.), AOS Liptovský Mikuláš, ISBN 978-80- 8040-391-1.
[4] CHOVANEC, F., JUREČKOVÁ, M.: *MV-algebra pastings.* International Journal of Theoretical Physics, Vol. 42, No. 9, 2003, pp. 1913-1926, ISSN: 0020-7748.
[5] JUREČKA, S., PINČÍK, E., BRUNNER, R.: *Solution of the optical parameters of the thin film systems and interfaces.* Applied Surface Science 2008, Vol.254, No.12, pp.3672-3676.
[6] JUREČKA, S., JUREČKOVÁ, M., KOBAYA – SHI, H., TAKAHASHI, M., MADANI, M., PINČÍK, E.: *Statistical and fractal properties of semiconductor surface roughness*. Advances in Electrical and Electronic Engineering 2008, Vol.7, No.1-2, pp.377-381.
[7] JUREČKA, S., JUREČKOVÁ, M., CHOVA-NEC, F., KOBAYASHI, H., TAKAHASHI, M., MIKULA, M., PINČÍK, E.: *On the topographic and optical properties of SiC/SiO_2 surfaces.* Central European Journal of Physics, Vol. 7, 2009, No. 2, pp. 321-326. ISSN 1895-1082.
[8] JUREČKA, S., KOBAYASHI, H., TAKAHA-SHI, M., BRUNNER, R., MADANI, M., PINČÍK, E.: *On topographic properties of semiconductor surfaces and thin film systems.* Materials Science Forum, Vol. 609, 2009, pp. 275-279, Book Thin Films and Porous Materials, doi: 10.4028/3-908454-02-6.275.
[9] MASOPUST, P. R.: *Fractal functions, fractal surfaces, and Wavelets*. Academic Press, London, 1994.
[10] PEITKEN, H. O., JÜRGENS, H., SAUPE, J.: *Chaos and Fractals, New Frontiers of Science.* Springer Scienc+Business Media, LLC, 2004, ISBN 978-0-20229-7.
[11]ŠALÁT, T.: *Metrické priestory.* Bratislava, ALFA, 1981.

Assoc. Prof. RNDr. Ferdinand CHOVANEC, PhD.
Department of Iformatics
Armed Forces Academy of General M. R. Štefánik
Demänová 393
031 06 Liptovský Mikuláš
E-mail: ferdinand.chovanec@aos.sk

EXPERIMENTAL RESARCH OF THE MAGNETIC AURA OF A SMALL-SIZE JET-ENGINE AND THE POSSIBILITIES OF APPLICATION FOR DIAGNOSTICS AND CONTROL

František ADAMČÍK, Ján KABÁT, Jana MODROVIČOVÁ

Abstract: The contribution is presenting the results of experimantal measurement of the magnetic aura of a small-size jet-engine with focus on the effects of foreign bodies heated up. The experiments form part of the programe of the Laboratory of Intelligent Control Systems of the Faculty of Aeronautics of the Technical University Košice.

Keywords: The magnetic aura, jet-engine, diagnostics, magnetometer.

1 INTRODUCTION

The comprehensive targets of experimental measurements of the magnetic aura of an aviation engine je to investigate the possibilities of its application for diagnostings the given object, increasing the quality of its controll, to suggest its situational control and also for its safe operation [1]. Measurements were performed on an small-size aviation jet-engine, the MPM-20, installed in the Laboratory of the intelligent control systems of aviation engines of the Faculty of Aeronautics, Technical University Košice, making use of the existing measurement tools and a magnetometer designed and further improved by its authors [2, 3].

The aim of the measurements described were to find out and evaluate the changes in the magnetical aura of the object of measurement, i.e. the MPM 20 jet- engine, how it is affected by foreign, heated – up bodies. Measurements were conducted using the VEMA 030, a 4-channel magnetometer.

2 OBJEKT OF MEASUREMENT

The object subjected to measurement is an axial-flow, turbo-charged ejet-engine with a single-stage, one-sided radial compressor, joint combustion chamber, single.stage non-cooled gas-turbine and a fixed outlet nozzle of MPM-20 type. Originally, the engine was used as an air starter of a greater engine (TS-20) and as a stand-buy power generator for aircraft such as Su-7 and Su-22 [11].

Currently, it is used for experimental purposes as an ideal object of research oriented not only on magnetic aura or using elements of artificial intelligence in the field of control but also for further areas for example possibilities of using alternative fuels. Its size is quite small though, compared to a turbo-charged jet engine used to drive big aircraft, but its characteristics are comparable to other aviation engines in use.

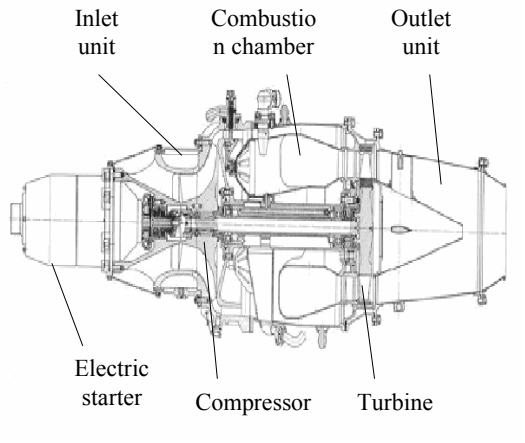


Figure 1 Functional diagram of the MP-20 engine

3 DESCRIPTION OF THE EXPERIMENTAL MEASUREMENTS CONDUCTED

The measurements were conducted in one-minute intervals with the engine in calm status (switched off) and with contents separated into three stages:

a) Measurement of the actual background of the MPM-20

Aimed at finding out the actual values of the magnetic aura of the environment of the measured object wothout external influences, on the basis of which it would be possible to evaluate the effect of external thermal sources - the light-bulb turned on or the piece of iron heated up. The probes were installed in the corners of a quadratic structure, directed gradually along the x, y and z axes (Figure 2a).

b) Measurement of the MPM 20´s magnetic aura wit a light bulb of 60W output turned on

The mode of thermal source influence - light bulbs – overall time of measurement 9 minutes, the light bulb was turned on in the 4th minute and turned off in the 7th minute of the measurement cycle.

c) Measurement of the MPM 20´s magnetic aura with a pice of iron heated

Overall time of measurement - 3 minutes, the picece of iron was being heated throughout the entire time interval.

Location of the light bulb and the heated piece of iron with reference to the MPM 20 is illustrated in Figure 2b.

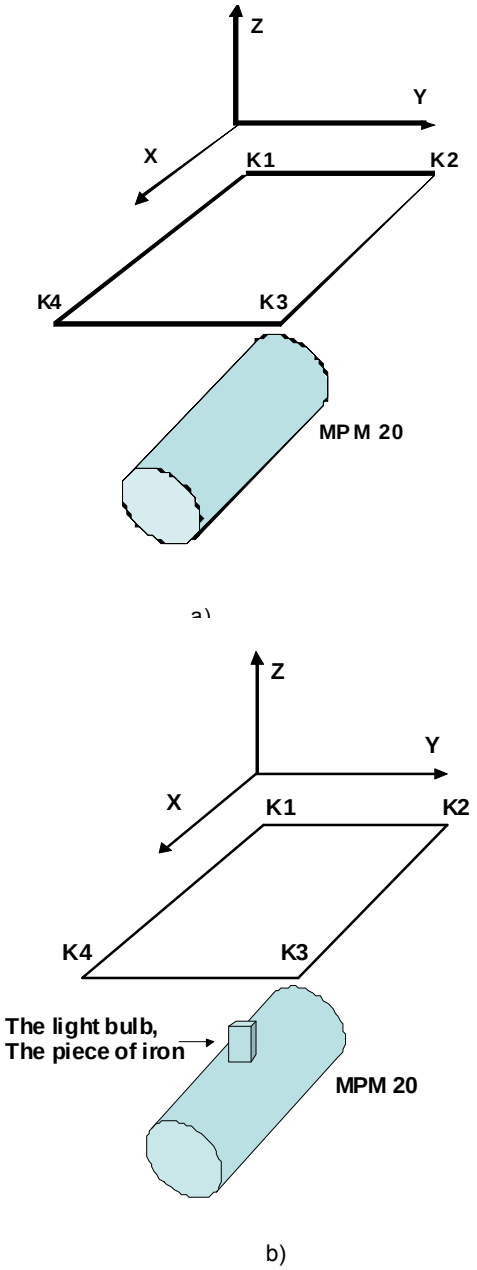


Figure 2 Location of probes, the light bulb and the piece of iron a) measurement of the background, b) measurement of the background with the light bulb and the piece of iron

The magnetic aura component can be horizontal and vertical, with horizontal components in direction of x and y axes and the vertical component in direction of the axis z.

Location of the probes is seen in Figure 2 marked asfollows: K1 – Channel 1 – Probe 1; K2 – Channel 2 – Probe 2; K3 – Cannel 3 – Probe 3; K4 – Channel 4 - Probe 4.

Actual meteorological conditions (dew point, atmospheric pressure, mean values of external and internal temperatures, ozone) in the time of measurements are given in Table 1.

4 RESULTS OF THE MEASUREMENTS

Surveys on the measured values of the magnetic inductiion in the individual axes of x, y and z for the described measurements are shown in Table 2 while their graphical behaviours are illustrated in Figures 3, 4 and 5.

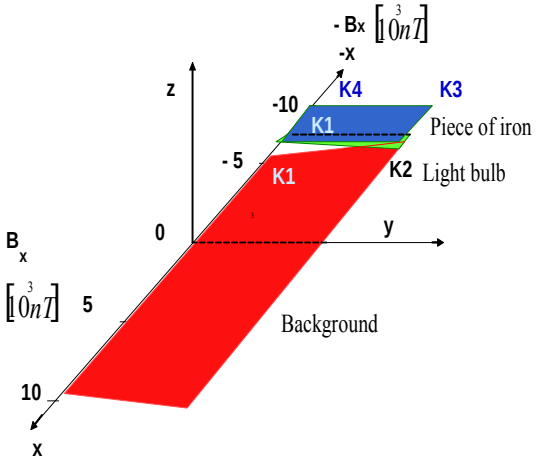


Figure 3 Behaviour of mean values of the magnetic induction B_x along the axis x

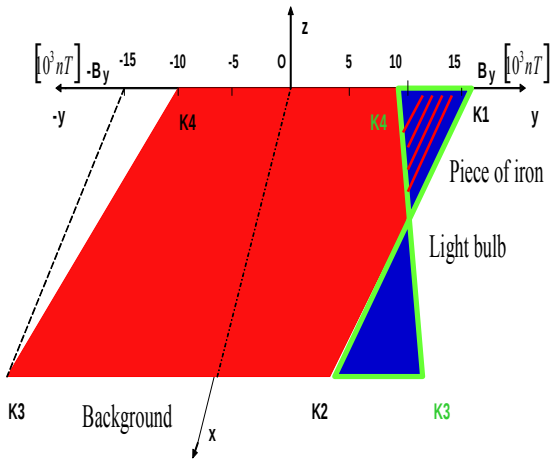


Figure 4 Behaviour of mean values of magnetical induction B_y along the axis y

Table 1 Meteorological condition of measurement

Dew point [°C]	Atmospheric pressure [hPa]	Average value of external temperature [°C]	Average value of internal temperature [°C]	Ozone [Dobs. Units]
4	1009	6	18	299 (0 %)

Table 2 Mean values of magnetic induction values measured in the separate axes [10^3 nT]

Measurement		**A**	**B**	**C**
Axis	Probe	Background	Light bulb	Iron
X	K1	-6,4	-7,1	-7,1
	K2	-7,8	-7,2	-7,0
	K3	10	-8,58	-9,8
	K4	9,5	-7,85	-9,57
Y	K1	16	16,2	16,2
	K2	9	9,3	9,3
	K3	-16	16,5	16,4
	K4	-10	10,3	10,2
Z	K1	24	22,8	22,8
	K2	25	24,2	24,1
	K3	-23	23,0	23,2
	K4	-25	25,1	25,0

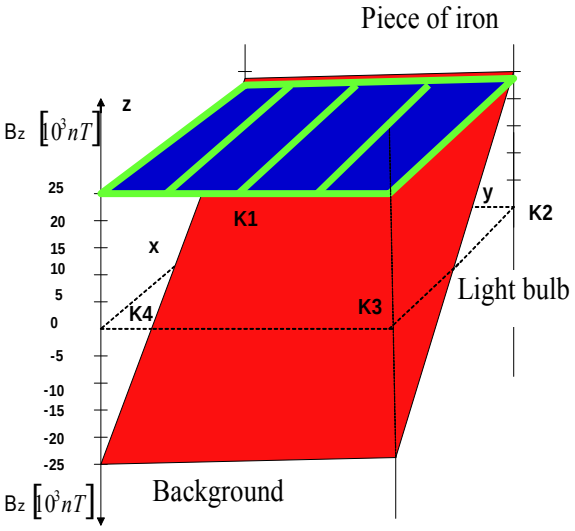


Figure 5 Behaviour of mean values of magnetic induction B_z along the axis

5 CONCLUSION

Based on the measured values and the individual graphical behaviours it follows that the effect of external thermal sources (light bulb, heated piece of iron) has been proved – both soruces of heat cause (almost equally) substantial changes in the MPM 20´s magnetic aura in the individual components.

To arrieve at higher accuracy of the measured values and the effect of external thermal sources exerted upon the magnetic aura of the measured object, further measurements are needed – with the engine in cold start–up and normal engine operation modes. The data obtained will be of use to establish and verify the system of diagnosting and situation control of aviation engine based on the information on the changes in its magnetic aura.

References

[1] ANDOGA, R.: *Hybridné metódy situačného riadenia zložitých systémov*. Doktorandská dizertační práca. FEI TUKE 2006. pp. 120.

[2] HUDÁK, J., ČOPÍK, J.: Magnetic field monitoring of areas. In: *Journal of Electrical Engineering,* Magnetic 1335-3632, Volume 55, 2004.

[3] ČOPÍK, J., HUDÁK, J.: 3 kanálový fluxgate magnetometer s CPLD Coolrunner- II. In: *Acta Avionica*. roč. 5, č. 8. 2003. p. 81-89. ISSN 1335-9479.

[4] ČOPÍK, J.: *Monitorovania magnetického poľa v záujmovom priestore*, Doktorandská dizertačná práca, LF TUKE, 2005, pp. 135.

[5] ČOPÍK, J.: *Autonómne meranie vybraných fyzikálnych polí*., Písomná práca k dizertačnej skúške, VLA Košice, 2002, pp. 49.
[6] FŐZŐ, L.: *Využitie matematického modelu rovnovážneho a nerovnovážneho chodu motora MPM-20 pri návrhu algoritmu riadenia v každom čase*, Doktorandská dizertačná práca, FEI TUKE, 2008, pp.142.
[7] HUDÁK, J.: *Magnetometria, vybrané problémy vývoja a využitia*, Habilitačná práca, VLA Košice, 1999, pp.107.
[8] KABÁT, J., MADARÁSZ, L., MODROVIČOVÁ, J.: Introduction to problem, basis and measurement of magnetic aura of turbojet aircraft engines, In: *Computional Intelligence and Informatics: Proceeding of the 9th International Symposium of Hungarian Researchers*, November 6-8, Budapest, ISBN 978-963-7154-82-9, 2008, pp. 379-386.
[9] MODROVIČOVÁ, J.: *Štúdia možnosti vplyvu magnetickej aury leteckého motora MPM 20 na jeho situačné riadenie.* Písomná práca k dizertačnej skúške, FEI TUKE, Košice, 2009, pp. 85.
[10]MODROVIČOVÁ, J., MADARÁSZ, L., KABÁT, J.: Magnetic aura study of small turbojet engine MPM 20. In: *SAMI 2009* - 7th International Symposium on Applied Machine Intelligence and Informatics: January 30-31, 2009, Herľany, Slovakia. IEEE, 2009. p. 25-28. ISBN 978-1-4244-3802-0.
[11]Technický popis MPM-20.

Assoc. Prof. Dipl. Eng. František ADAMČÍK, CSc.
Faculty of Aeronautics
The Technical University of Košice
Rampová 7
041 21 Košice
Slovak Republic
E-mail: frantisek.adamcik@tuke.sk

Dipl. Eng. Ján KABÁT
Faculty of Aeronautics
The Technical University of Košice
Rampová 7
041 21 Košice
Slovak Republic
E-mail: jan.kabat@tuke.sk

Dipl. Eng. Jana MODROVIČOVÁ
Faculty of Electrical Engineering and Informatics
The Technical University of Košice
Letná 9
040 21 Košice
Slovak Republic
E-mail: jana.modrovicova@tuke.sk

IMPROPER THRESHOLDING TECHNIQUES ELIMINATION

Ľuboš AUGUSTÍN, Milan LEHOTSKÝ, Jozef ŠTULRAJTER

Abstract: Image segmentation is an important stage in image processing. The purpose of segmentation is to detect and select significant image features. These can be suitably utilized in next image processing, such as classification, pattern recognition, texture analysis etc. From many segmentation methods contribution focuses on global thresholding. We suggest the algorithm detecting segmentation failure, based on resemblance of input and output image. It utilizes important image features in sense of significant edges according to Marr theory ideas. Implementation is realized by suggested edge operator, its application to input gray-valued image and definition of two functions describing detected edges distribution in binarized output.

Keywords: Successful Binarization, Edge operators, Edge Images, Optimal Thresholding, Image Segmentation.

1 INTRODUCTION

Segmentation techniques are usually divided into four categories (1):
- segmentation based on edge detection,
- segmentation based on region growing,
- segmentation using pattern matching,
- thresholding.

Thresholding is used in applications where brightness of object pixels is markedly different from background. It is an effective and fast tool for selection of optimal threshold – value dissociating image into two disjunct regions. The goal of thresholding is to find a set of pixels in image with attribute „object“. At present, thresholding is often limited. Because of excessive data quantity and complexity, thresholding techniques may fail and final image is unusable for following processing. However there are tasks where thresholding is an adequate and realizable choice. Recognition of written symbols is dependent on correct segmentation. Computer numeric control machines are equipped with optical sets allowing to detect product position, determine its shape and activate next automatic processing. Thresholding is used for component localization. In order to increase precision and effectivity there are many of optimizing methods, e.g. worktable adjustment, backlight, using non-transparency materials etc.

2 SELECTED THRESHOLDING TECHNIQUES ANALYSIS

Automatic thresholding algorithms can be categorized into six groups (2), based on type of extracted information:
- histogram-shape techniques where the peaks, valleys and curvatures of the smoothed histogram are analyzed,
- clustering techniques where the gray level samples are clustered in two parts as background and foreground (object) or alternately are modelled as two Gaussian distributions,
- entropic techniques which utilize entropy of foreground-background regions, the cross-entropy between the original and binarized image etc.,
- object attribute techniques which search a measure of similarity between the gray-level and binarized images, such as fuzzy similarity, shape, edges, number of objects etc.,
- spatial techniques which use the probability mass function models taking into account correlation between pixels on a global scale,
- local techniques which do not determine a single value of threshold but adapt the threshold value depending upon the local image characteristics.

The contribution analyses three thresholding techniques:
- clustering Lloyd,
- entropic Shanbag,
- attribute Pikaz technique.

2.1 Lloyd technique

It is assumed that the image can be characterized by a mixture distribution of foreground and background pixels (3). It is a modification of Riddler technique (4), but reflects the variance of the whole image. This eliminates iterative search and the threshold minimizes the total misclassification error:

$$T_{opt} = \arg\min_i \left[\frac{P(i)\cdot m_f(i) + (1-P(i))\cdot m_b(i)}{2} + \frac{\sigma^2}{P(i)\cdot m_f(i) + (1-P(i))\cdot m_b(i)} \cdot \log\frac{1-P(i)}{P(i)} \right] \quad (1)$$

where

$m_f(i)$ - mean of foreground pixels,

$m_b(i)$ - mean of background pixels,

$P(i)$ - foreground area probability on threshold i, $0 \le i < 256$.

2.2 Shanbag technique

Shanbag has considered a thresholding method that relies on a fuzzy membership coefficient, which indicates how strongly a gray value belongs to the

background or to the foreground (5). The membership value is based on the cumulative probability of that gray value. The optimal threshold is determined as

$$T_{opt} = \arg\min_i[H_f(i) - H_b(i)] \quad (2)$$

where

$$H_f(i) = -\sum_{j=0}^{i} \frac{p(j)\cdot\log(\mu_f(j))}{P(i)} \quad (3)$$

$$H_b(i) = -\sum_{j=i+1}^{255} \frac{p(j)\cdot\log(\mu_b(j))}{1-P(i)} \quad (4)$$

$$\mu_f(i-j) = 0{,}5 + \frac{p(i)+\ldots+p(i-1-j)+p(i-j)}{2\cdot P(i)} \quad (5)$$

$$\mu_b(i+j) = 0{,}5 + \frac{p(i+1)+\ldots+p(i-1+j)+p(i+j)}{2\cdot P(i)} \quad (6)$$

and p(i) is normalized histogram of image.

2.3 Pikaz technique

In this method offered by Pikaz and Averbuch (6), the objective is to binarize the image while establishing the correct size foreground objects. This is instrumented via the size-threshold function $N_s(T)$, parametrically dependent upon the object size. The threshold is established in the widest possible plateau of the graph of the $N_s(T)$ function

$$N_s(T) = \text{card}(A) \quad (7)$$

where A represents the set of objects with area greater than *s* pixels.

3 SELECTED TECHNIQUES APPLICATION

The proposed techniques were applicated on fingerprint image. Dactyloscopy is often conductive to conviction or disproof of potential culprit. The images of unique and invariant papillary lines in inner side of fingers are represented by thin line objects. The segmentation is optimal when output image contains maximum of utilizable information. Analyzed image and its binarized equivalents obtained by proposed techniques are displayed in Figure 1.

Top left is the original 256 gray-valued image. Top right image depicts result of application of Lloyd technique. There is a binarized version of source image obtained by Shanbag technique on the bottom left. As one can see, these techniques incorrectly segmented the image. Determined optimal thresholds were T_{opt}=254 and T_{opt}=11. The techniques produced images unusable for following processing. The result of Pikaz technique showed on bottom right produced optimal binarized image. Determined optimal threshold T_{opt}=153.

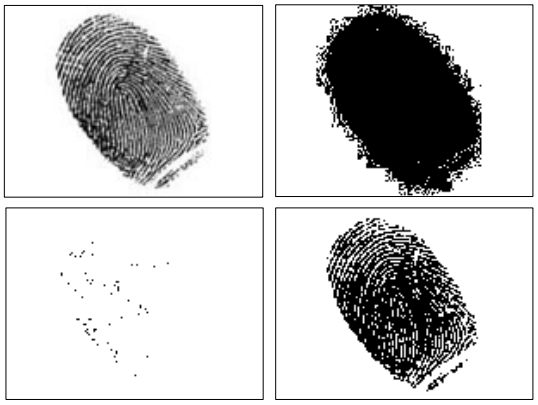


Figure 1 Thresholding techniques application

4 THRESHOLDING TECHNIQUE SUITABILITY ANALYSIS

Every thresholding technique is characteristic by its own algorithm for optimal threshold determination. Final binarized image is often improper for following automatic processing. In the next part the contribution deals with source and final image correlation analysis utilizing fundamental ideas of Marr theory (7, 13). Significant change of intensity function can be used for edge image definition.

Edge images can be formed using edge detectors, often categorized into two groups (8):

1. edge detectors based on discrete convolution, e.g. Roberts (9), Kirsch (10) and Prewitt (11) edge detector,
2. edge detectors finding pixels, where the second derivation of image function crosses zero, e.g. Canny (12) and Marr-Hildreth (13) edge detector.

To increase efficiency and robustness of suitability analysis we defined own edge detector. It is based on finding maxima of absolute values of image function gradient in reference pixel's neighborhood with coordinates (i, j)

$$\mathbf{e}(i,j) = \max(v1, v2, v3, v4), \quad (8)$$

where

$$\begin{aligned} v1 &= \text{abs}(\mathbf{a}(i\text{-}1,j) - \mathbf{a}(i\text{+}1,j)),\\ v2 &= \text{abs}(\mathbf{a}(i,j\text{-}1) - \mathbf{a}(i,j\text{+}1)),\\ v3 &= \text{abs}(\mathbf{a}(i\text{+}1,j\text{+}1) - \mathbf{a}(i\text{-}1,j\text{-}1)),\\ v4 &= \text{abs}(\mathbf{a}(i\text{-}1,j\text{+}1) - \mathbf{a}(i\text{+}1,j\text{-}1)), \end{aligned} \quad (9)$$

$\mathbf{a} \in \mathbf{Z}^{m\times n}$ is input 256 gray-valued image with dimensions $m\times n$,

$\mathbf{e} \in \mathbf{Z}^{m\times n}$ is output edge image.

Edge image for fingerprint is pictured in Figure 2.

Figure 2 Fingerprint edge image

Edge image can be used for suitability analysis of thresholding technique. We propose two functions describing detected edges distribution:

$$h_1(L) = \frac{h(L)}{h_e(L)}, \tag{10}$$

$$h_2(L) = \frac{h(L)}{N}, \tag{11}$$

where

$h(L)$ - number of detected edges of magnitude L,
$h_e(L)$ - edge image intensity histogram,
N - number of all detected edges, $L \in \{0..255\}$.

Function $h_1(L)$ defines ratio between number of detected edges of magnitude L and number of all present edges of magnitude L.

Function $h_2(L)$ defines dependence between all detected edges of magnitude L and all detected edges.

Suggested functions h_1 a h_2 for fingerprint image and proposed thresholding techniques are depicted in Figure 3 - 5.

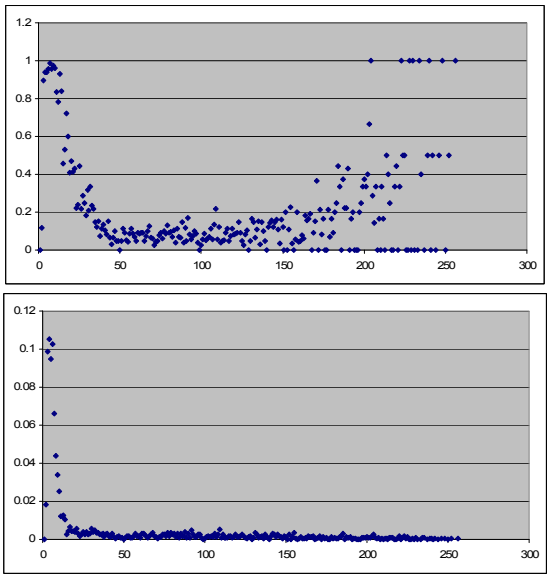


Figure 3 Functions h_1 a h_2 - Lloyd technique

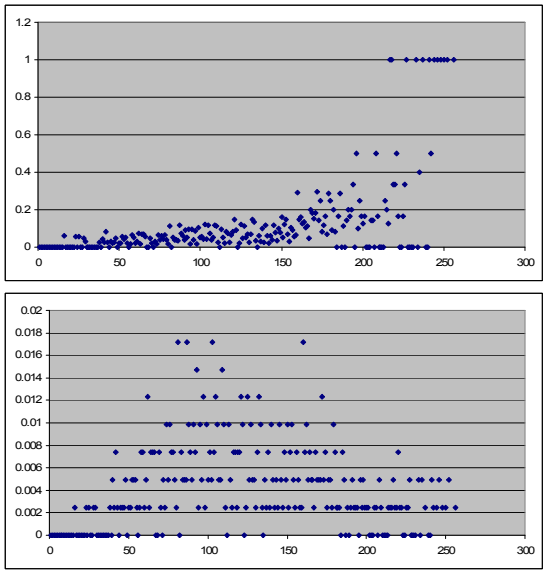


Figure 4 Functions h_1 a h_2 - Shanbag technique

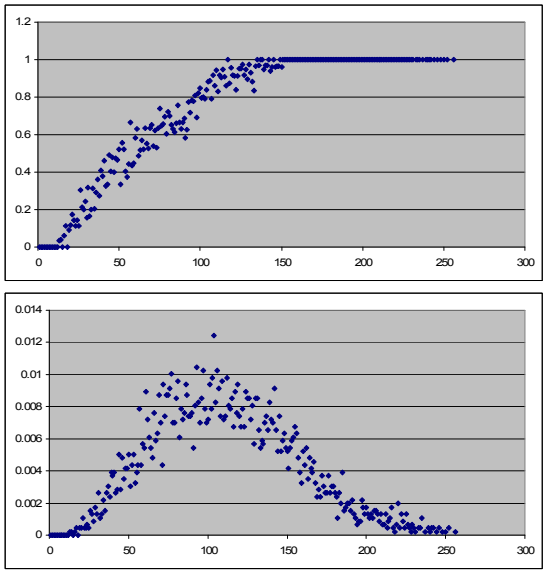


Figure 5 Functions h_1 a h_2 - Pikaz technique

As one can see from Figure 3, Lloyd technique detected expressive majority of edges with magnitude $L<25$. Output image is noisy and unusable for next automatic processing because of main features deficiency. Functions $h_1(L)$ a $h_2(L)$ for Shanbag technique are depicted on Figure 4. The technique preserved more important features as previous one, but $h_2(L)$ implies minimum amount of all detected edges. Binary result of thresholding isn't convenient for following processing. Pikaz technique selected significant edges. All edges with magnitudes $L \in \langle 150,256 \rangle$ were kept, as indicates function $h_1(L)$ on Figure 5. The distribution of

edges, visually described by function $h_2(L)$, is the most convenient of proposed techniques.

5 CONCLUSION

We have analyzed three segmentation methods based on thresholding in contribution. Their application was demonstrated on fingerprint image containing thin lines objects.

Different thresholding procedures produced different determined threshold values and output binary images. Lloyd and Shanbag techniques failed; only Pikaz algorithm resulted to usable output.

The contribution provides proposal how to eliminate improper segmentation. We defined new edge detector, invariant to gradient direction. It is suitable for edge image forming. Binary result of segmentation was compared with edge image in the next stage. Using suggested functions h_1 and h_2 we are able to find relationship between 256-gray level input ant its 2-level output. The functions h_1 and h_2 are convenient and robust descriptors for improper segmentation and may be helpful to optimal segmentation determination.

References

[1] HLAVÁČ, V., ŠONKA, M.: *Počítačové vidění*, Grada, 1992.

[2] SEZGIN, M., SANKUR, B.: Survey over thresholding techniques and quantitative performance evaluation, In: *Journal of Electronic Imaging*, 2004.

[3] LLOYD, D. E.: Automatic Target Classification Using Moment Invariant of Image Shapes, In: *Technical Report,* 1985.

[4] RIDLER, T. W., CALVARD, S.: *Picture thresholding using an iterative selection method*, IEEE Trans. System, Man and Cybernetics, pp. 630-632, 1978.

[5] SHANBAG, A. G.: *Utilization of Information Measure as a Means of Image Thresholding, Computer Vision Graphics and Image Processing,* pp. 414-419, 1994.

[6] PIKAZ, A., AVERBUCH, A.: *Digital Image Thresholding Based on Topological Stable State,* Pattern Recognition, pp. 829-843, 1996.

[7] MARR, D., VISION, A.: *Computational Investigation into the Human Representation and Processing of Visual Information*, 1982.

[8] MAŘÍK, V. et al.: *Umělá inteligence 2,* Praha: Academia, 1997.

[9] ROBERTS, L. G.: *Machine Perception Of Three-Dimensional Solids*, Lincoln Laboratory, 1965.

[10] KIRSCH, R.: *Computer determination of the constituent structure of biological images*, Comp. Biomed. Res, vol. 4., pp. 315-328, 1971.

[11] PREWITT, J. M. S.: *Object enhancement and extraction, Picture Processing and Psychotropictorics*, pp. 75-149, 1970.

[12] CANNY, J., A.: *Computational approach to edge detection*, IEEE Trans. Pattern Analysis and Machine Intelligence, pp. 337-346, 1986.

[13] MARR, D., HILDRETH E.: *Theory of edge detection*, Royal Society, 1980.

Dipl. Eng. Ľuboš AUGUSTÍN
Department of health and special training Lešť
Ministry of Defence of the Slovak Republic
962 63 Pliešovce
Slovak Republic
E-mail: lubos.augustin@mil.sk

Assoc. Prof. RNDr. Milan LEHOTSKÝ, CSc.
Faculty of Education
Catholic University in Ružomberok
Námestie A. Hlinku 56
034 01 Ružomberok
E-mail: milan.lehotsky@fedu.ku.sk

Prof. Dipl. Eng. Jozef ŠTULRAJTER, CSc.
Department of Infromatics
Armed Forces Academy of General M. R. Štefánik
Demänová 393
031 06 Liptovský Mikuláš
Slovak Republic
E-mail: jozef.stulrajter@aos.sk

A COMPREHENSIVE APPROACH TO DECISION-MAKING AND INTERACTION MECHANISMS THAT VALORIZE RELEVANT INFORMATION

Paul BECHET, Iulian ILIES, Amalia LEBU, Anca DINICU

Abstract: Effects-based organizational planning aims to synthesize the operational structure and strategies that enable organizations to achieve the desired effects in the context of a rapidly changing environment. The informational domain, as a link between the cognitive and physical domains of an organization sets forth to achieve "informational superiority" and valorize the valuable information of all organizational processes. Informational superiority does no longer refer to obtaining knowledge, but to how the valuable information produced within an organization is shared and applied by its members, as we are now witnessing a shift from a need-to-know, to a need-to-share approach to information. In this context, the competences of those actors participating in the "observe → monitor → decide → act" cycle as well as the interaction mechanisms established within or outside the organization become the key enablers. Our paper considers all the above mentioned aspects and introduces a basic model for the dynamic assignment of decision rights based on competences and interaction mechanisms.

Keywords: Collaborative environments, decision rights, comprehensive approach, knowledge management.

1 INTRODUCTION

Operational environments are getting more and more dynamic, being characterized by an infinite number of variables. The force that drives operational environments is change. It is adaptation to change and especially information superiority that makes the difference between success and failure. If modern technology is the key factor that gives us an advantage over the adversaries, information and interoperability are the key driving forces that help us operate efficiently in complex environments. If not before long „information superiority" used to be achieved by extracting the relevant information, i.e. by turning data into knowledge, nowadays we are witnessing a shift from the "need-to-know" to the "need-to-spread-and-share" approach to information. (Kristiina Rintakoski & Mikko Autti, 2008). Obtaining information, though still important, is no longer the main focus, emphasis being now placed on how information is shared between organizations, at a macro level, or between the members of an organization, at a micro level, the purpose being to achieve a comprehensive approach.

This article considers all the above mentioned aspects and sets forth to analyze the way decision rights could be assigned to the members of an organization, so as to have the decision-making process observe the principles of Comprehensive Approach.

2 COMPREHENSIVE APPROACH

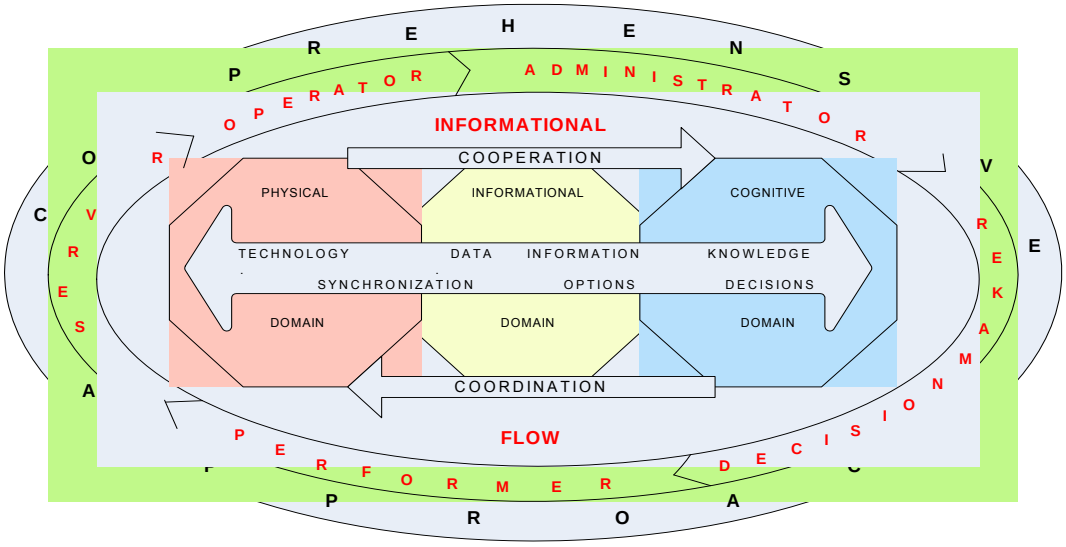
At present, NATO is focusing on developing a **comprehensive approach** as an operational, strategic and tactical concept, based on the **Effects Based Approach to Operations.** Whilst there is no commonly accepted definition for **Comprehensive Approach**, there is broad agreement that it implies an integrated effort of cooperation between the actors participating in a mission. We will introduce the principles underlying the concept of Comprehensive Approach and subsequently apply them within a flexible and robust organizational structure, with the purpose of identifying the individual characteristics that play an important part in the decision making process. Effects-Based Terminology was used to integrate the diplomatic, informational, military, and economic instruments to create the conditions for success. (Smith, 2002). Coordinating all these instruments into a comprehensive approach can be done with the help of several collaborative tools that will help us gain and maintain an integrated understanding of the problem and come up with comprehensive solutions.

The concept of comprehensive approach unifies all actors in an integrated manner making them act in close cooperation. Comprehensive Approach means that an organization should try to co-ordinate with other organizations, not coordinate them. Effective comprehensive approach requires unity of effort and shared vision among the agencies, organizations, institutions and forces participating in a mission. This requires forging a comprehensive approach with a shared understanding and appreciation for the intended end state.

Comprehensive approach is an approach that integrates all collaborative efforts of all entities to achieve unity of effort towards a shared goal. Comprehensive Approach is framed by three underlying tenets: understanding, cooperation and joint effort. The approach promotes a shared understanding of the situation towards accomplishing a shared goal and obtaining the desired end results. Understanding does not imply conformity, each actor contributing a distinct set of professional, technical or cultural values and competences to mission accomplishment. The approach is based on a cooperative effort reinforced by institutional trust and transparency.

This culture would deliver the best effects if it were implemented not only at the top level, but also at the lower (organizational and individual) level. Comprehensive Approach is not an end in itself but a means to an end. It is a tool that helps us obtain the desired effects, promote and develop a mechanism and culture based on trust, collaboration and cooperation both vertically between nations and international organizations and horizontally, within nations and organizations.

Applying comprehensive approach would mean implementing a mechanism and culture of cooperation that would allow information to flow smoothly both internally, between the members of an organization, and externally, between the organization and its operating environment (Figure 1).



People working in international organizations should make themselves familiar with the concept of Comprehensive Approach, which, after all, is more a mindset than a formalized way of working. The activities from the physical, informational and cognitive domains should be interdependent, and performed concurrently not sequentially, to identify coordinated, shared and comprehensive solutions. Optimizing the flow of information from the physical towards the informational and cognitive domain should not be an end in itself but a necessity because the activities in the three fields tend to be interdependent. Thus, achieving a culture of cooperation and coordination between actors at all levels – a comprehensive approach - becomes mandatory, the development of more structured relations between and within structures being a first step to that end. This is not an easy task to achievewe can, though, start by correctly assessing both competences and pitfalls of the organization and the members comprising it, and by identifying the most appropriate way for actors to act in order to complement each other's efforts. Comprehensive Approach should leverage the disparate capabilities of all actors, and not by compellingthem to work together towards a common goal, but by making them participate out of a shared understanding and appreciation of what that goal represents. Achieving this outcome will enable a smooth flow of information between actors who gain a comprehensive situational awareness, and an increased strategic planning and decision making capability.

3 ASSIGNING DECISION RIGHTS IN THE CONTEXT OF COMPREHENSIVE APPROACH

Assigning decision rights in an organization can fall between two diametrically opposing practices: on the one had there is total centralization – all decision rights are assigned to one actor alone- and on the other hand there is total decentralization – all the entities in an organization share equal decision rights. The implementation of modern concepts such as EBO (Effects- Based Organizations), KM (Knowledge

Management) or NEC (Networking Enabled Capabilities) has enabled a dynamic assignment of decision rights, where depending on the situation, various actors, can at different stages, gain access to the decision making process (Alberts et al., 2006). This becomes possible due to an increased level of situational awareness shared by all the members of an organization which expresses the self-synchronization capability of the organization referred to. It means that an actor can temporarily acquire decision-making rights of a flexible coverage, depending on the responsibility he/she assumes. From this perspective, the distribution of in-depth and in-width agents becomes decisive. An agent with a good in-width capacity will be an efficient decision-maker whereas an agent with a good in-depth capacity will be an efficient performer, able to put decisions into practice. (Levchuk et al., 2006).

Applying dynamic assignment to military organizations might as well be a viable solution, especially since such organizations are very homogenous and a combination of responsibilities between experienced and beginner agents can have significant consequences upon organizations operating in complex environments. This ensures plurality of perspectives and entails a constant combination of available pieces of information in ever newer ways, by valorizing the relevant information and the qualitative resources. This approach enables the accomplishment and maintenance of the observe-monitor-decide-act cycle, by engaging all the actors participating in the activity of an organization, at the three levels: physical, informational and cognitive. (Smith, 2002). In KM terms, the physical domain is characterized by the high capacity of state-of-the-art technology to observe and analyze, the informational domain by the capacity to process information, whereas the cognitive domain by actors' ability to extract the relevant information and make the optimum decision.

The key element is *interaction,* which on the one hand enables actors to recognize, distribute, use, access and share data, i.e. to achieve the informational cycle, and on the other hand ensures organizational cohesion, in that it generates knowledge. The quality of information and of the action itself is influenced by the *nature of interactions* occurring at the social, informational and cognitive levels, the purpose of which is to achieve collaboration towards a common goal. *Collaboration* refers to a wide spectrum of activities, including exchange of information, coordination, consultation, synchronization and integration. Increased interactions within or even outside the organization can significantly influence collaboration. Within the organization, an exchange of information arising out of an increased capacity to understand the information signals an increased level of cohesion. Outside the organization, information exchange expresses the capacity of an organization to adapt to complex environments characterized by an increased level of uncertainty and risk. (Bechet et al., 2009).

As far as information exchange is concerned, interaction within and outside the organization entails identifying those elements which give relevance to the informational gain, i.e. generate results, and achieve organizational objectives. The agent with the highest in-width capacity is the decision-maker, whereas at the opposite end there is the performer, with an increased in-depth capacity. Complex military actions require team members to have differentiated experience and understanding capacities. Team structure, in terms of the optimum ratio between in-depth and in-width competences as arising from a thorough analysis of the interactions occurring within the team, depends on the degree of complexity of team activities and on the distribution of these activities. Valuable information made available through individual actions may be turned into relevant information only in so far as the interaction mechanism allows team members to express themselves, by dynamically assigning rights and responsibilities. The collaborative environment is the perfect solution, as it constantly monitors the organizational state and signals whenever *a valuable piece of information becomes available*. All team members should have permanent access to the signaling channel, each member having to constantly update the available information. This does not rule out the assignment of certain levels of priority within the team.

Information exchange implies a source agent which generates the information and an addressee, a decision-making agent which understands and applies that information. We can speak of informational gain only when the source agent knows how to process the information and how to extract only those relevant bits of information the addressee needs in order to understand the message. In other words, the performer (specialist) adds value to information by "summarizing" it. We are dealing here with information processing useful for the transfer of information. Informational gain can also be analyzed from the perspective of the decision-maker and his capacity to process and understand the information he receives. This exchange of information between agents can also be analyzed in terms of loss, i.e. valuable information lost during the processing. As a consequence, for resource management to be as effective as possible, responsibilities among team members should be assigned depending not only on individual competences (suited for the specific activity to be performed) but also on team collaboration.

3.1 Competences and Interaction between Organizational Actors

The model we have used to quantify the value of the informational flow is the one based on the theory of multiple agents (Levchuk et al., 2006). To what extent agents can understand each other is an important part of the information transfer. The absolute value of the information processed by a specific agent represents the informational gain. To what extent is the absolute value of a piece of information, produced by an agent, relevant, depends on the interaction established among organization agents. It follows that an agent can "produce" extremely relevant information which can not be valorized due to a mismatch between the two communicating agents. We are dealing here with informational loss caused by an interaction between two agents with very different levels of understanding, a case in which the organizational resources can be said to have been used extremely erroneously. It can even be argued that the information created through individual actions can not be put to the common good use of the organization because the latter lacks an effective mechanism of interaction that would help organizational agents share a *common understanding of valuable information.* Such situations are frequent in static organizations where information is usually blocked because decision-making agents are unable to process and understand the valuable information received from the in-depth agents which do not have decision rights. There is a certain degree of congruence between each and every pair of organizational components, congruence which depends on the extent to which the needs, requirements, objectives and elements of one structure matches the needs, requirements, objectives and elements of another. Congruence can thus be said to measure the "matching" between each pair of organizational components.

The relation between actors' competence and interaction will be analyzed based on the theory of multiple agents (Levchuk et al., 2006), the interaction mechanism including 8 agents, one decision-maker, three in-width agents and four in-depth agents. For demonstration purposes, we will consider that only two competences are needed from an agent to perform an activity. It follows that the vector describing the experience of each agent will comprise only two components, as shown in the expression below:

$$A = [c_1, c_2] \tag{1}$$

where: c_1 represents the competence to process type 1 information and c_2 the competence to process type 2 information.

The in-depth and in-width capacity will be reduced to one unitary value (i.e. the sum of the two components will be the same for all agents). Thus, an agent having an equal/level distribution of the two competences will have a high in-width capacity, whereas an agent having a higher value for one of the two competences will have an increased in-depth capacity. Further on, we will discuss several situations depending on competence values, on the capacity of agents to process information (i.e. to reduce redundant information) and on decision-makers capacity to understand the information they receive. To that end, the agents of our analysis are either extremely specialized in one of the two competences (as for instance agent A_1 who has a very high capacity to process type 2 information) or relatively specialized in both competences, as is the case of agents A_3 and A_4 in expression 2. In the case of in-width agents/experienced agents (A_5, A_6 and A_7) their capacity to process information is the same for the two types of information.

$$\begin{cases} A_1 = \left[\frac{1}{9}, \frac{8}{9}\right], A_2 = \left[\frac{2}{9}, \frac{7}{9}\right], A_3 = \left[\frac{3}{9}, \frac{6}{9}\right] \\ A_4 = \left[\frac{4}{9}, \frac{5}{9}\right], A_5 = \left[\frac{1}{2}, \frac{1}{2}\right], A_6 = \left[\frac{1}{2}, \frac{1}{2}\right] \\ A_7 = \left[\frac{1}{2}, \frac{1}{2}\right] \end{cases} \tag{2}$$

As far as the decision-makers are concerned, their capacity to process the two types of information has been reduced as shown in expression 3 below, because the decision maker, as the key element of all organizations, will also have other types of information to process.

$$A_d = [\frac{1}{4}, \frac{1}{4}] \tag{3}$$

Simulations will reveal two situations, a first situation in which specialized agents, through their capacity to process information will extract the relevant information reducing the total flow of information to 20% and a second situation in which agents reduce the informational flow to 50% of the initial amount.

The analysis will then focus on the relation between the decision-maker and the other agents from three different perspectives. A first one in which the decision-maker makes decisions relying on the competence of a single agent, a second one in which the decision-maker considers the contributions of all agents for each type of information and a third one in which the decision-maker decides based on the competences of agents for both types of information.

Figures 2, 3, 4, and 5 present the results for the situation in which only 20% of the total amount of information is processed and forwarded to the decision making agent. It can be noticed that the best result in terms of informational gain is obtained by the in-depth agent 4. As expected, in-width agents 5, 6, and 7 have too a relatively good informational gain of a 2/1 ratio. Figure 3 suggests that the best suited agent to process information of the 2 type is agent 1, with a loss/gain ratio of approximately 4/1. This result remains valid provided no additional informational loss distorting the flow of information between the decision making and the in-depth agents, occurs in the interaction mechanism. Moreover, this situation also entails a high degree of trust between the two agents in that even if the decision-maker doesn't entirely understand all the information he receives, he should trust the agent to have given him enough relevant information, and thus successfully valorize the competence of agent 1. If, for various reasons, additional perturbations occur in the interaction mechanism, the above mentioned relevant pieces of information may be lost, and in that situation we are dealing with a piece of valuable information which is "produced" by the in-depth agent but which can not be valorized because the decision-making agent can not understand it as the competence levels of the two agents referred to are completely out of tune/sync. The informational gain obtained by grounding the decision on the agent which has the best competence is approximately 50 % higher than the informational gain obtained by grounding the decision on the contribution of all agents, regardless of their competence (Figures 4 and 5). However, this is only the case if the "communication channel" is very good and if there are no perturbations. If on the contrary, there are strong perturbations, then it is recommended to make the decision based on the contributions of all agents, regardless of their competence level, in which case the gain/loss ratio is higher (1/0.32 as compared to 1.44/0.61).

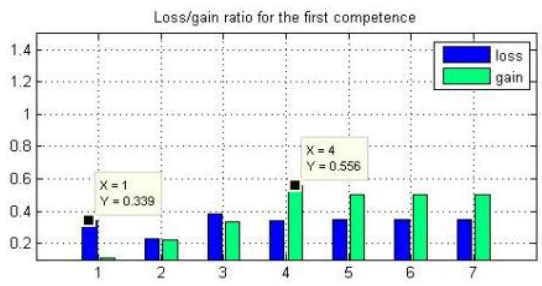


Figure 2 Informational loss/gain ratio for type 1 information when 20 % of the total amount of information is processed

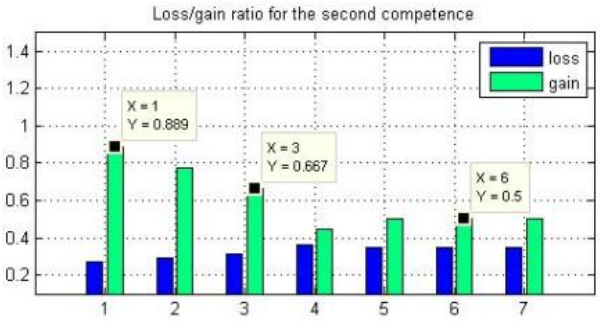


Figure 3 Informational loss/gain ratio for type 2 information, when 20 % of the total amount of information is processed

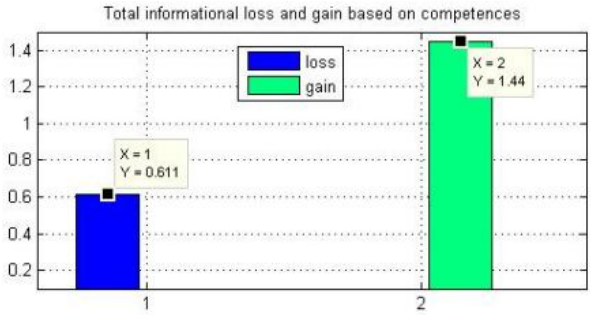


Figure 4 Total informational loss and gain when decisions are based on the agent that has the best competence and 20% of the total amount of information is processed

Figures 2, 3, 4, and 5 present the results obtained when 50 % of the total amount of information is processed. The best loss/gain ratio is obtained by in-width agents 5, 6 and 7. The increased competence of agent 2 to process type 2 information can no longer be valorized (Figure 7). This happens because the decision-making agent can not accurately understand the information sent to him by the in-depth agent, and will generate huge errors if he were to forward such information.

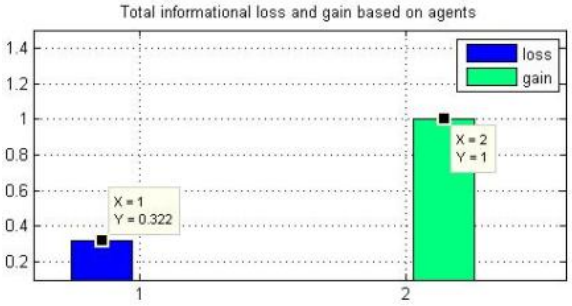


Figure 5 Informational gain and loss when decisions are based on the contributions of all agents and 20 % of the total amount of data is processed and turned into relevant information

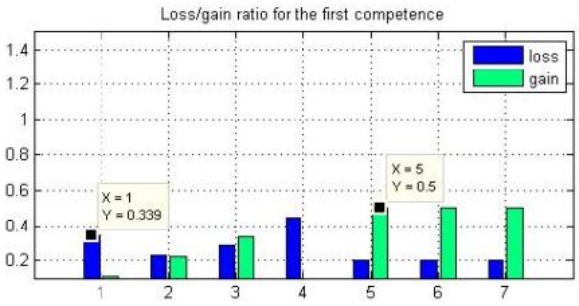


Figure 6 Informational loss/gain ratio for type 1 information when 50 % of the total amount of data is processed

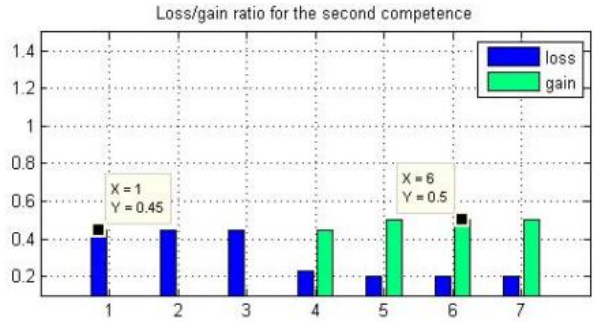


Figure 7 Informational loss/gain ratio for type 2 information, when 50 % of the total amount of data is processed

The interaction mechanism emphasizes the important part played both by the relevant information and by the decision-maker who should be at all times able to filter the information he receives both quantitatively and qualitatively. The results presented in figures 8 and 9 show that the decision-maker is even more incapable to valorize the information coming from the agent that has the best competence.

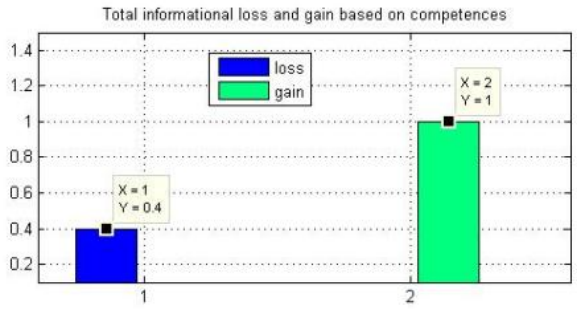


Figure 8 Total informational gain and loss when decisions are based on the agent that has the best competence and 50 % of the total amount of data is processed

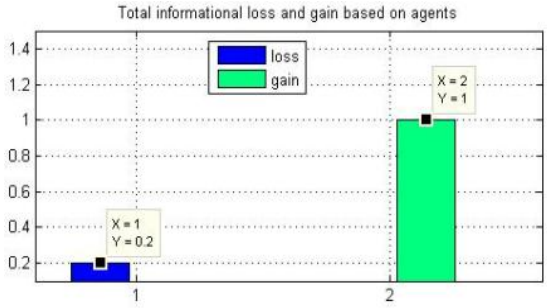


Figure 9 Total informational loss and gain when decisions are based on the contribution of all agents and 50 % of the total amount of data is processed

5 CONCLUSION

At present NATO is trying to develop a Comprehensive Approach based on the operational concept of effects based operations. Although there is no commonly accepted definition for Comprehensive Approach, the term can be said to imply a joint effort of cooperation between all the actors of an organization towards a common goal. In this context, the informational domain grows in importance, not just because smooth informational flows have to be ensured but also, more importantly, because relevant information has to be valorized and shared by all the members of an organization. It is along these lines that this paper has been developed to analyze the competences, interaction mechanisms and dynamic assignment of decision rights in organizations.

Valuable pieces of information made available through individual actions will be valorized and turned into relevant information in so far as the interaction mechanisms allow organization members to express themselves through a dynamic assignment of decision rights and responsibilities. As a consequence, we speak of informational gain in communication between agents if the valuable information produced by an in-depth agent (by processing it) can be transposed into a format that lies within the understanding range of the decision maker. If however, not all the pieces of information can be transposed into such a format, everything exceeding the decision maker's range of understanding is considered to be informational loss.

The issue of achieving organizational performance depends upon ensuring those operational pairs of agents within the team that work with maximum informational gain and minimum loss. A signaling mechanism could be really helpful in optimizing the interaction mechanisms by choosing those organizational pairs able to work to work at optimum parameters and produce maximum informational gain and minimum loss. All agents, independent of the informational flow should have

access to the signaling channel. Moreover, it is highly recommended that the priority to this channel should be dynamic, i.e. consideration should always be given to the agent with the best informational gain/loss ratio. Data gathering could thus be oriented towards those pairs of agents generating valuable information that can be shared among organization members.

The study has been developed on a simplified model and emphasized the importance of the interaction mechanism between the organizational pairs of actors by evaluating the informational loss/gain ratio in terms of information relevance within the decision-making process. The selection criteria we have used emphasized the need for a dynamic assignment of responsibilities and rights within an organization. The model can also be extended to larger organizations by dividing the organization into clusters, each cluster having a high dynamic and a smaller area of action, both the mechanisms within and between clusters being subject to optimization.

The work was supported by a grant financed by the Romanian Ministry of Education and Research, through Contract no. 367/2007 from CNCSIS (National University Research Council) - "Idei" Programme.

References

[1] ALTMAN, D. G., BLAND, J. M.: Statistics notes. Units of Analysis. In: *BMJ,* 1997, vol. 314; p. 1874.

[2] ŽIAKOVÁ, K. a kol.: *Ošetrovateľstvo - teória a vedecký výskum.* Martin : Osveta, 2003. 319 s. ISBN 80-8063-131-X.

[3] RYBÁR, L.: *Základy rétorickej komunikácie.* Bratislava : Obzor, 1980. 304 s. ISBN 65-014-80.

[4] ALBERTS, D. S., HAZES, R. E., [2006], *Understanding Command and Control*, CCRP, Washington .

[5] LEVCHUK, G. M., YU, F., LEVCHUK, Y, PATTIPATI, R.: [2006], *Networks of Decision-Making and Communicating Agents: A New Methodology for Design and Evaluation of Organizational Strategies and Heterarchical Structures*, Research Report, Contract # N00014-02-C-0233, University of Conneticut, USA.

[6] BECHET, P. I., ILIEŞ, J., LEBU A., BABOŞ A.: [2009], *Effects-Based Knowledge Management in Changing Environments Meant to Improve the Information Domain of Communicating Agents*, The 10th European Conference on Knowledge Management. Università Degli Studi Di Padova, Vicenza, Italy, 3-4 September 2009.

[7] RINTAKOSKI, K., AUTTI, M.: Comprehensive Approach, Trends, Challenges and Possibilities for Cooperation In: *Crisis Prevention and Management,* In: *Comprehensive Approach Seminar,* 17 June 2008, Helsinki.

[8] SMITH, E. A.: *Effects Based Operations: Applying Network Centric Warfare in Peace*, Crisis and War, CCRP, Washington, 2002.

Assoc. Prof. Dipl. Eng. Paul BECHET, PhD.
SA Amalia LEBU
Dipl. Eng. Iulian ILIES
Lect. univ. Anca DINICU
"Nicolae Balcescu" Land Forces Academy Sibiu
3-5 Revolutiei Street
Sibiu
Romania
E-mail: toaderamalia@yahoo.com

APPLICATIONS AND COMPUTATIONAL ASPECTS REGARDING THE COANDĂ EFFECT

Mircea BOŞCOIANU, Vasile PRISECARIU, Ionică CÎRCIU

Abstract: Coanda effect is the phenomena in which a jet flow attaches itself to a nearby surface and remains attached even when the surface curves away from the initial jet direction.In free surroundings, a jet of fluid entrains and mixes with its surroundings as it flows away from a nozzle. When a surface is brought close to the jet, this restricts the entrainment in that region. As flow accelerates to try balance the momentum transfer, a pressure difference across the jet results and the jet is deflected closer to the surface - eventually attaching to it.

Key words: Slot, attached jet, static pressure, centrifugation zone, suction zone.

1 INTRODUCTION IN PHYSICAL ASPECTS REGARDING THE COANDA EFFECT

The Coandă effect is a natural phenomenon with action on the flow attached to a divergent wall (volet or airfoil) characterized by a high assimmetry. It is posible to remark the following aspects (Figure 1):

1. The depressured zone determines:
   a) *flow acceleration upstream in the slot*, without increasing upstream pressure or temperature,
   b) *the displacement of the local fluid.*

2. Detaching and re-attaching is caracterized by histerezis (the reattaching is produced at smaller angles than the detaching).

3. The global flow that results from the mixing between the main flow and the displaced one is situated in the depressure zone and is characterized by lower temperature.

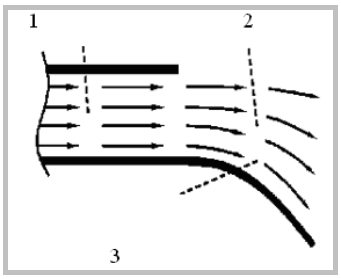
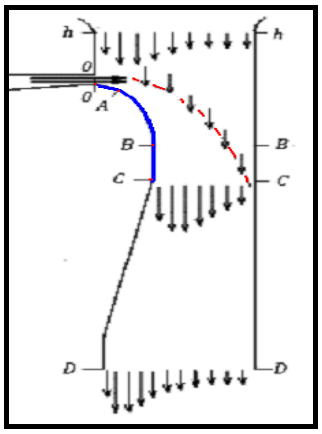


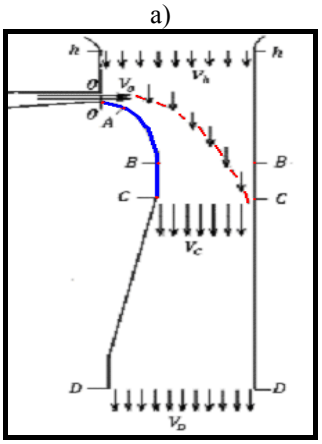
Figure 1 Coandă flow (2D)

2 A GLOBAL ANALYSIS OF THE MIXING PROCESS IN COANDĂ EJECTION EVICE

Let consider an ejection device that we are going to analyse from the point of view of the mixture between the primary flow , the active one, through which energy is introduced into the system, and the secondary flow.



a)



b)

Figure 2 Coanda ejector with: a) non uniform speed distribution, b) uniform speed distribution

In the inlet (Section **0-0)**, the primary flow is introduced by compression, acceleration or through absorbtion directly from the environment. The absorbtion section (h-h) through which the resulting inflow moves only and is characterized by the fact that the total enthalpy i^* of the flow is the same with that of the environment i_H^*. The place around A is supposed to be the spot where the depressurization flow is maximal. Section B-B shows the end of the Coanda profile (line OAB). Section C-C is where the absorption section ends and the thickness of the

mixin region equals that of the C-C section. D-D is the exit section from the ejection disposal and is characterized through the fact that the static pressure is equal with that of the environment static pressure p_H. The area **h-0-C-B-h** is considered to be the absorption area where the total enthalpy i^* of the flow is the same as that of the environment i_H *. Area **0-ABC-C-0** is considered to be that of the mixture where the whole quantity of generated flow is received through the permeable surface **C0**. Area **C-D-D-C** is the area of acquiring uniformity for aerothermogazodynamic parameters in section C-C and it usually has a divergent form which contributes to the increase of efficiency of the ejection device. Its existence leads to the increase of the generated flow but it doesn't necessarily mean an increase of the propulsion force. The research on the force increase will have to take into consideration the entire geometry of the ejection device. The known factors are the geometry of the ejection device in its sections (**Ah, A0, AB=AC, AD**), the fuel conditions in the slot (p^*, P_0), and environmental conditions (p_H, ρ_H, i_H^*). Also, for this global analysis of the mixture in the ejection device the values of the energetic performance ηC, ηD on sections **00-CC**, **00-DD**, are considered as known.
In fig.3 is presented the distribution of speed in a section of the Coanda ejection device with two different regions, an asymmetrical one (d width), and a uniform one (D-d width) where the length of the boundary layer at the wall being s.

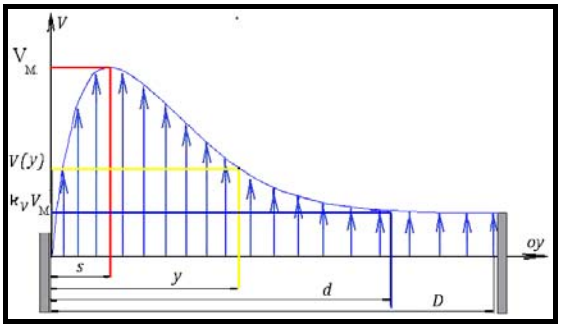


Figure 3 Distribution of speed in a section

3 A GLOBAL ANALYSIS OF: COANDA EJECTION DEVICE WITH UNEVEN SPEED

Let a Coanda ejector with non uniform and variable speed distribution. In the D exit section, the static pressure p_D equals the environment pressure p_H. The power transferred to th fluid in D section is: $P_0 = \eta P_D = \int_{A_D} \rho_H V_D(y) \left(i_D^* - i_H^*\right) dA_D =$

$$\frac{\rho_H V^3_{MD} A_D \chi_{3D}}{2} \quad (1)$$

The gain force is given by the difference between the two force distributions, with a maximal value corresponding to A (Figure 4):

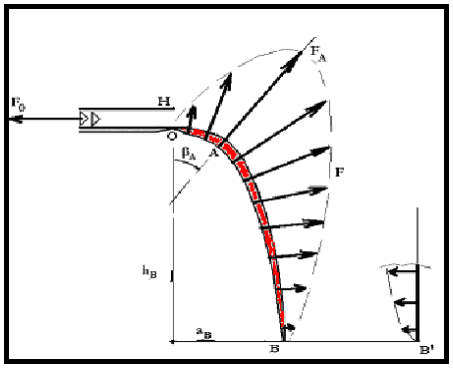


Figure 4 Force distributions on Coanda airfoil

Let detail Coanda flow by using two zones with special properties, the centrifugation zone and the suction zone (Figure 5).

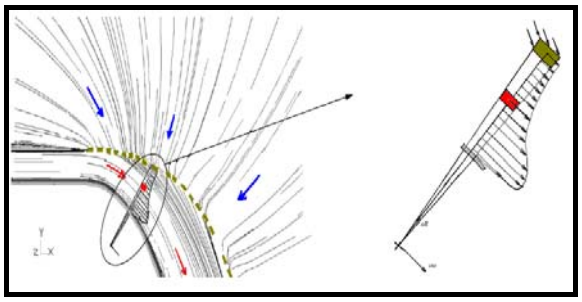


Figure 5 Detailed analysis of Coandă flow

The equations for the centrifugation zone which is associated to the mixing region 0-ABC-C-0 with C0 permeable are:

$$\frac{1}{r} \cdot \frac{\partial(\rho \cdot u_\omega)}{\partial \omega} = 0 \quad (2)$$

$$-\frac{u_\omega^2}{r} = -\frac{1}{\rho}\frac{\partial p}{\partial r} \quad (3)$$

$$u_\omega \frac{\partial u_\omega}{\partial \omega} = -\frac{1}{\rho}\frac{\partial p}{\partial \omega} \quad (4)$$

$$i^* = i_H^* \left(\frac{p}{p_H}\right)^{\frac{k-1}{k}} + \frac{u_\omega^2}{2} \quad (5)$$

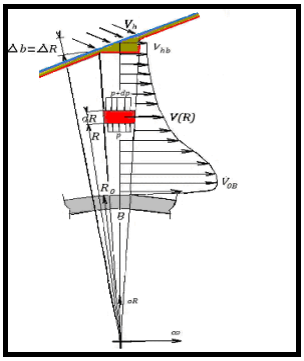


Figure 6 Element of jet

For a small element of jet (fig.6) flow, the radial movement equation is:

$$\frac{dR}{R} = \frac{dp}{\rho u_{\omega}^2} \tag{6}$$

For $\boldsymbol{B_i}$ on the profile:

$$u_{\omega} = u_{\omega 0} f_u(R) \qquad u_{\omega 0} = u_0 f_{u0} \tag{7}$$

and the total enthalpy is conserved:

$$i^*(R) = \frac{[u_{\omega}(R)]^2}{2} + \int \frac{[u_{\omega}(R)]^2}{R} dR\big|_R + i_c^* \tag{8}$$

The static pressure is expressed by:

$$p(R) = p_H \left(1 + \frac{1}{i_H^*} \int \frac{[u_{\omega}(R)]^2}{R} dR\big|_R\right)^{\frac{k}{k-1}} \tag{9}$$

and the static density ad static temperature are:

$$\rho(R) = \rho_H \left(1 + \frac{1}{i_H^*} \int \frac{[u_{\omega}(R)]^2}{R} dR\big|_R\right)^{\frac{k}{k-1}} \tag{10}$$

$$T(R) = T_H \left(1 + \frac{1}{i_H^*} \int \frac{[u_{\omega}(R)]^2}{R} dR\big|_R\right)^{\frac{k}{k-1}} \tag{11}$$

The gain in force at $\boldsymbol{B_i}$

$$\phi_{Bi} = \frac{1}{b_0} \int_{R1}^{R2} \left(1 + \frac{1}{i_H^*} \int \frac{[u_{\omega}(R)]^2}{R} dR\big|_R\right)^{\frac{k}{k-1}} f_{u0}^2 f_u^2(R) dR \tag{12}$$

and the corresponding efficiency is:

$$\eta_{Bi} = \frac{1}{b_0} \int_{R1}^{R2} \left(1 + \frac{1}{i_H^*} \int \frac{[u_{\omega}(R)]^2}{R} dR\big|_R\right)^{\frac{k}{k-1}} f_{u0}^3 f_u^3(R) dR \tag{13}$$

We note that the flow attached is situated in the depressure zone defined by the exit from slot, 0-0, B-B section and D-D exit with a maximal value in A.

4 EXPERIMENTAL RESEARCH: NOTAR HELICOPTERS

Coanda Effect Small Appliance experimentally validated results Coanda effect on the tail of the helicopter beams are obtained by linking the theoretical and computational study and realization of profiles that require permanent adjustments, consisting of: - changes in flow air induced by the main rotor (here we have 2 variables that blades gave up their rotation speed), - changes in flow induced by fan beam tail in blood flow adjustable - profile geometry (in this case is the beam tail) - Coanda slot number, their geometry (including width and length l). We should also note that good and realistic fluid flow in Coanda effect has been obtained from practical experiments using hundreds of adjustments and amendments agreed to variables. The experiments in real physical scheme (Figure 7) is the model helicopter IAR316B, 1:10 scale, which are installed drive components that measure and control equipment. - It is considered a technical unit that has the structure:

A structure helicopter consists of two main components: a cabin and tail beams of composite materials (glass and carbon fiber).

Figure 7 hub and blades of carbon fiber highlighting the Coanda effect and performance measurement on the beam parameters of the experimental device queue (v. Figure 5.14) seeks a lateral force F as possible. He looked fluid flow along the tail beam, the desired length to be studied, and of the lateral force **F** due to Coanda effect. This was attempted obținerera a Coanda profile optimally adjustable, depending on the flows data load-bearing rotor, fan tail, position and geometry of the slot. This optimization allows the replacement of the rotor anti-torque eliminating its disadvantages and lead to the advantage given to obtain high lateral forces (maximum) with low energy consumption, which is found in formulas elements performance helicopters.

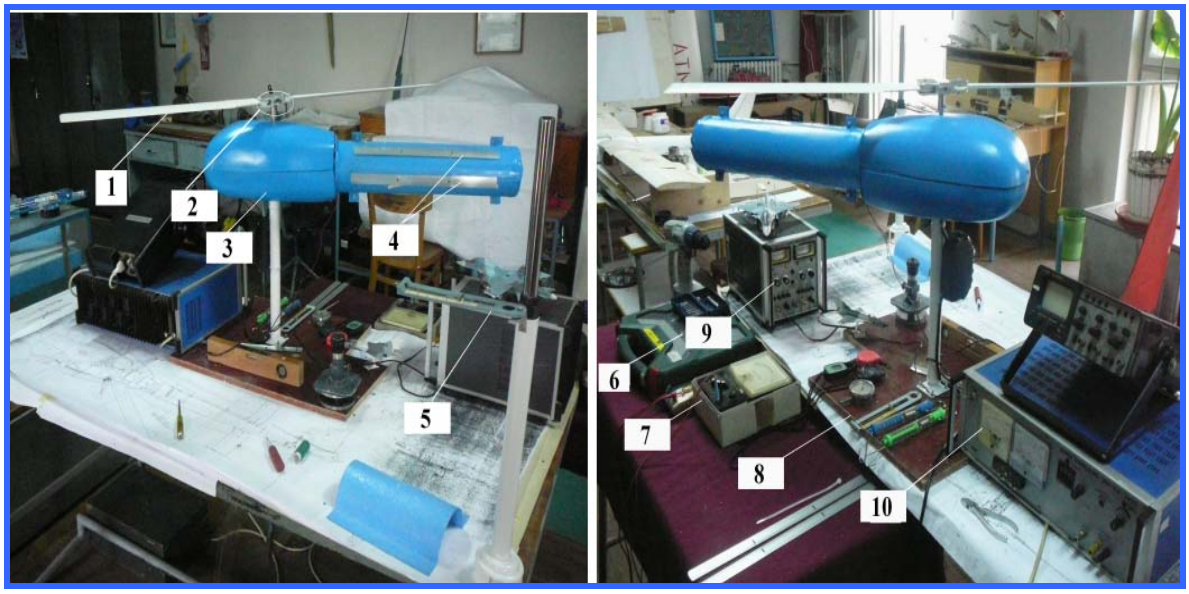


Figure 7 Experimental device: 1-NOTAR helicopter blades carbon fiber, 2 -hub pitch, 3-structure helicopter fiberglass, 4-Coanda slots, 5-device measuring dynamometer force F, 6-tool kit with 7-measuring and control equipment (multivoltampermetre) 8-instruments for making measurements (timer, dynamometers, Anemometers electronic, mechanical comparator, roulette), 9 - dual source DC power, 10-stabilized power source and electronic oscilloscope.

Table 1 Results from the experimental device with dimensions: length l = 30cm gap width = 2 mm and up blades p = 4^0

Nr. crt	P	p	P_c	V_{GC}		V_f	V_R	Q_R	Q_{gc}	$2Q_f$	E_{fc}	F
UM	[W]	[θ_p ^{0}]	[W]	Km/h	[m/s]	[m/s]	[m/s]	[m^3/s]	[m^3/s]	[m^3/s]	[N/W]	[N]
1	70	4	6	19,4	5,30	5,74	7,22	8,161	4,242	0,00688	0,0783	0,47
			7,2	21,6	6,00	5,93			4,804	0,00710	0,0708	0,51
2			8,4	25,3	7,02	7,24			5,621	0,00868	0.0678	0,57
			9,6	28,3	7,86	7,35			6,294	0,00882	0,0666	0,64
3			10,6	30,8	8,55	8,65			6,846	0,01038	0,0660	0,70
			12	33,4	9,27	9,35			7,420	0,01122	0,0658	0,79
4	95		6	19,4	5,30	5,91	8,31	9,393	4,242	0,00709	0,0950	0,57
			7,2	21,6	6,00	6,22			4,804	0,00746	0,0916	0,66
5			8,4	25,3	7,02	7,44			5,621	0,00892	0,0904	0,76
			9,6	28,3	7,86	8,12			6,294	0,00974	0,0875	0,84
6			10,6	30,8	8,55	8,89			6,846	0,01066	0,0896	0,95
			12	33,4	9,27	9,38			7,420	0,01125	0,0875	1,05
7	145		6	19,4	5,30	5,98	9,56	10,806	4,242	0,00717	0,1133	0,68
			7,2	21,6	6,00	6,37			4,804	0,00764	0,1083	0,78
8			8,4	25,3	7,02	7,88			5,621	0.00945	0,1047	0,88
			9,6	28,3	7,86	8,39			6,294	0,01006	0.1041	1,00
9			10,6	30,8	8,55	9,12			6,846	0,01094	0,1037	1,10
			12	33,4	9,27	9,43			7,420	0,01131	0,1033	1,24

Interpretation of results was performed in graphical form as follows: A first step is to highlight the evolution of lateral force F due to Coanda effect depending on the power consumed by the rotor bearing the six-power variation of the fan beam intubated in the tail , was added to this step change blades with three values. Figure 8 highlights the changes in lateral force due to Coanda effect to adjustments of power tools (three values: P = 70 [W], 95 [W], 125 [W]) applied to the rotor portal, maintaining constant width = 2mm to slot the same value of step blades $p = 4^0$. For each amount of useful power applied to the rotor bearing was varied six-speed adjustable fan P_C Power intubated in the beam tail.

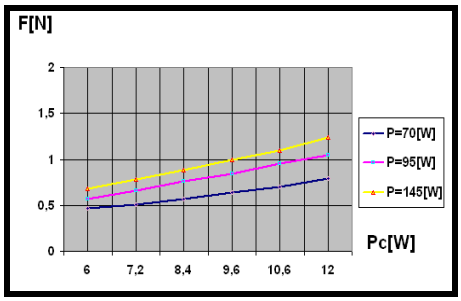


Figure 8 Variation lateral force F due to Coanda effect to maintain constant: $p = 4^0$ and = 2mm

It can be observed approximately linear increase in lateral force F for each value of power P, and could approximate the family of linear functions useful GN-power applied force determining correspondence.

Example: - P = 95 [W] to maintain constant step $p = 4^0$ and = 2 mm is obtained with a linear function that can find value GN lateral force F due to Coanda effect: Where do I noted the power of P_C fan beam intubated tail with index variable.

E_f efficiency of a propulsion device as the ratio of force produced (generated) of propellant and power consumed by it: in our case F is the lateral force developed by Coanda effect and Po is power consumed by the fan tail of beam Pc (Figure 9).

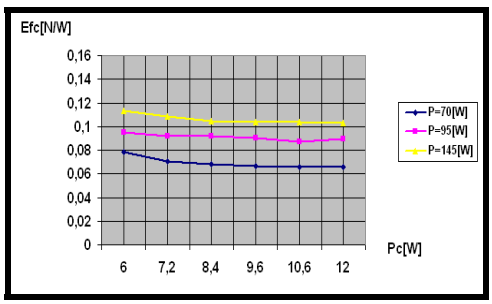


Figure 9 Variation of tail-beam efficiency E_{FC} by lateral force F and power consumed in the three steps of PC operating rotor blades portant step in the value of $p = 4^0$

5 CONCLUSIONS

In the conclusion we can state that for the same energy available P_0, the D_f *force* gain can be obtained by decreasing the speed $V_D < V_M$, similarly to an increase by ejection of the mass flow *evacuated.* In order to obtain the highest force possible for an available used energy it is preferable to put into motion the highest amount of fluid possible with the lowest speed possible instead of a small amount of fluid put into motion with a high speed.

This study was conducted in the idea of highlighting the usefulness of the devices developed force using Coanda effect, the main aim being to obtain a lateral force inthe Coanda effect, the possibility of control where needed, time required to stabilize the portal created by flying helicopter rotor monorotor. results were obtained with values close to those obtained numerically by computational, respecting the geometric dimensions of the beam tail and fluid velocities in beam tail rotor that product flow rate carrier - shows the smooth growth index values E_{FC} a sharp increase in force F which requires a finding of finding optimal positioning slot; can do a pretty fair approximation of a full scale helicopter.

References

[1] BERBENTE, C., CONSTANTINESCU N. V.: *Dinamica gazelor şi Dinamica fluidelor vâscoase*/Aerotermochimie, Bucuresti:UPB, 1985.

[2] CARAFOLIE, E., CONSTANTINESCU N. V.: *Dinamica fluidelor compresibile*, Ed. Academiei RSR, Bucuresti, 1984.

[3] COANDĂ, H.: *Air - source du progres* [Book]. - [s. l.] : qqq, aprox1961.

[4] COANDĂ, H.: *Fluid propulsion system* [Patent] : US3337121 – 1967.

[5] COANDĂ, H.: *Programme d'etude et de realisation d'un aerodyne actionne par un systeme thermosouflant*, 1960.

[6] Constantinescu, N. V., GALETUSE, S.: *Mecanica fluidelor si elemente de aerodinamica,* Bucuresti, Ed. didactica si pedag., 1983.

[7] DINEA, S.: *Contribuţii la studiul efectului Coandă,* Ph.D Thesys, Bucureşti 2009.

[8] CÎRCIU, I.: *Îmbunătăţirea performanţelor de zbor ale elicopterelor prin aplicarea efectului Coandă-sistemul NOTAR* ,Ph.D Thesys, Braşov, 2009.

Prof. Dipl. Eng. Mircea BOŞCOIANU, Ph.D.
Dipl. Eng. Ionică CÎRCIU, Ph.D.
Dipl. Eng. Vasile PRISECARIU
"Henri Coanda" Air Force Academy in Braşov
160, Mihai Viteazu St,
500 183 Brasov, Romania
E-mail: mircea_boscoianu@yahoo.co.uk
circiuionica@yahoo.co.uk, aerosavelli73@yahoo.com

SOME ASPECTS REGARDING THE FLIGHT DYNAMICS AND STABILTY OF QUAD-ROTORS MICRO AERIAL VEHICLES

Ionică CÎRCIU, Mircea BOŞCOIANU, Stanislav SZABO

Abstract: The actual interest is to develop micro aerial vehicles with VTOL (vertical take off and landing) capabilities that could better respond to the new D3 missions (dull, dirty, dangerous missions). The focus on downsizing aspects is essential because in this case it is possible to adopt new missions that are not possible for the conventional UAV systems. First we present the benefits of RW-MAVs together with a comparative analysis of the matching between the configuration and different mission scenarios. According to the new missions we define the basic requirements that RW-MAV that should be satisfied in order to successfully complete urban and indoor missions. We define the new 4RW-MAV architecture and we propose a comparative analysis with the characteristics and performances of different classic configurations. Based on a better maneuverability, portability and agility, the 4RW-MAV architecture is promising but depending on the geometry there are some differences regarding the performances, stability and the payload capacity. In urban or indoor missions the maneuverability is crucial and thus the new architecture should provide better movement capabilities. The 4RW-MAV configuration is effective in indoor narrow space with a capability to maneuver in a very fast and effective way, impossible for other configurations. The net effect relevant for control during autorotation landing is analyzed by adding a vertical offset relative to the vertical position predicted in the absence of ground effect. This vertical offset is estimated from flight data and taken into account accordingly.

Keywords: Micro aerial vehicle (MAV), quad-rotors MAV (4RW-MAV), VTOL capabilities.

1 INTRODUCTION

Autonomous RW-MAV provide unique capabilities, such as in place hover and low-speed flight, important for many applications but their flight dynamics represents a challenging control problem because their control is significantly difficult comparative to fixed-wing category (FW-MAVs).

Recently, there has been considerable progress in autonomous RW-MAV flight. Examples range from basic upright hovering and forward flight to inverted hovering, and even to extreme aerobatic maneuvers. All of this prior work pertains to helicopters operating with normal engine power.

The starting point for this kind of research is based on a new definition, a new classification of scenarios and missions proposed for the smallest category of UAVs. Based on the new technologies and the downsizing of the payload and sensors, the effective envelope of civil applications for UAV systems (UAS) is extended and the research is focus on the ways to find new architectures, new solutions for reducing the costs of missions. According to their special capability to hover, there are different types of missions for RW- UAVs: urban law enforcement, special operations and information gathering; coastal patrol, on-shore border patrol and maritime surveillance; civil security (search & rescue and avalanche survivor search); fire brigade; civil security and police (contamination measurement and natural disaster monitoring); environmental (crop monitoring and local science mission); flight services (training, terrain mapping, photography and monument inspection). The aim of the analysis of scenarios and the capability to respond to different possible profiles of the mission is to obtain new solutions, more robust and more effective. The main obstacles in the development of small size rotary wing vehicles are related to the following aspects: it is very difficult to develop control laws in an environment in which the flow induced by rotors in the vicinity of walls generates strong nonlinear aerodynamic ground effects; the problem of obstacle avoidance is difficult for small size objects; the autonomous navigation in a GPS-denied environment is not very accurate for small systems; the design of an airframe that can protect the vehicle against collision is possible only for dedicated configurations (ISAE concept). There are made of course new steps in video compression and real time monitoring, in navigation and control of micro vehicles. It is also necessary to reduce the weight, size and power consumption of payload (analyzing sensors technology, optics, housing and cabling and connectors), to adopt innovative sense and avoid systems, to test some new platform configurations that allow an extended envelope of operation for such miniaturized systems.

The basic performance parameters are presented in the following list:

- **Maximum Take Off Weight**: it represents the overall value of the vehicle mass. It is calculated considering the whole RW-MAV with every kind of device or instrument installed on it at the moment of the start of the mission (take off).
- **Payload:** for this parameter different definitions can be found. In our case, considering the modular conception of the platforms that will be designed, the payload can be seen as the maximum weight of the module applied to the

vehicle. The difference between maximum take off weight and payload represents just the weight of the vehicle with the only devices strictly needed to make it fly.

- **Maximum speed:** this is the highest value of speed that the RW-MAV can reach during the fly.
- **Endurance:** it represents the time that the air vehicle can spend flying before a new landing is required (for changing batteries, refueling, recharging, downloading collected data, …).
- **Range**: it represents the maximum distance from the starting point that the vehicle can reach considering that it must come back and land. This parameter could change depending on the mission. In fact if it is not required the RW-UAV to come back, the maximum range could be theoretically double.
- **Ceiling:** due to the changing of the air characteristics with the altitude with respect to the sea level, the RW-UAV can reach a maximum height depending on its characteristics (power, efficiency, etc.).

2 BASIC REQUIREMENTS AND INNOVATIVE SOLUTIONS

The basic requirements to satisfy in order to successfully complete an urban mission are:

- **Safety:** is for sure the most relevant topic when any kind of vehicle, especially if it is a flying vehicle, has to operate near to human beings. In case of accident, due to an external factor (system failure, too strong wind, etc) or a mistake during the mission (wrong manual maneuver, bad mission definition to the autopilot, etc), the contact between any rotating part of the vehicle and people in the surrounding has to be prevented and avoided.
- **Agility:** in urban environments it is common to find buildings very close one to each other, with different height, trees, electric cables, poles and a huge number of other fix or moving obstacles. For this reason, once took for granted that any "urban-UAV" must have its own collision avoidance system, the vehicle needs a great agility in terms of rapidity in changing speed, direction or altitude. The controllability must be improved as the speed of the platform increases as at high speed there is less time for decision and command of escape maneuvers. This characteristic would be probably more relevant than other, like the maximum speed value, because in narrow spaces it can become strictly necessary. Form this point of view the best platform would be the smallest and lightest one (for instance a small 4 rotors) or in general the one with the higher power/mass ratio.
- **Autonomy:** in order to satisfy this particular requirement, the general platform layout or shape is not so relevant. More relevance has to be given to all the vision sensors (cameras, IR, thermal or sonar sensors, etc) and the flight control software (autopilot, collision avoidance, etc.). So during the design phase of an UAV for urban applications, all these devices must be taken into account and must be developed very carefully.

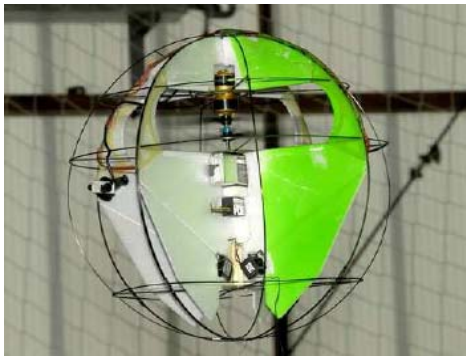


Figure 1 Concept from ISAE

a)Vision concept from ISAE (Figure 1) is a new tilt body bi-rotor concept based on two propellers facing each other and surrounded by a series of circular carbon rods. Each propeller is driven by a simple out runner brushless motor which has the advantage to avoid the complexity of a hollow shaft system. Speed control of both motors is carried out through a pair of speed controllers electrically connected through the carbon rods.The ultimate goal of the Vision is to be used as a hand-launch projectile which could be thrown through a window, roll on the floor and take-off to complete its indoor spying mission.

b) AirRobot (Figure 2) is a micro UAV with autonomous flight and navigation capabilities and modular payloads for use in reconnaissance, surveillance, search and rescue, documentation, inspection and also other scenario.

Figure 2 AirRobot

AirRobot AR family is a concept based on quadro-rotor solution and has a compact size of only 700 mm in diameter and utilizes a new (patented) propulsion system.

Figure 3 Fancopter

Fancopter (Figure 3) is a close range aerial reconnaissance micro- system. The compact dimensions and collision avoidance system enable this RW- MAV to be used even inside buildings.

The conventional **4RW-MAV** (Figure 4a) is suited for very small payloads and can eventually fly inside buildings. Due to the small size of the platform part of the structure or internal components can be shared with model aircraft industry, this allows the widespread utilization of COTS for the design and development of such systems. Quad-rotors can be up scaled to higher maximum take off weight mass and payload mass to fulfill payload requirements for other missions, such as those coupled with the small shrouded rotor platform (Figure 4b). **SIERRA Cargo Plus** allows the utilization of bigger sensors for missions inside buildings. It results a capability to perform some missions that involve surveillance thanks to the availability of professional camera and video camera in the new payload range

a)

b)

Figure 4 a) The conventional Concept , b) Sierra concept "4RW Cargo Plus"

3 AN ANALYSIS OF THE FLIGHT DYNAMICS FOR 4RW-MAV CONCEPT

The 4RW-MAVs configurations are well-known to have complex dynamics. For instance, to completely capture the state of the "4RW-MAV system" one would have to include the state of the air around the 4RW-MAV into the dynamics model. However, various prior work done on a conventional RW-MAV has shown it is possible to build a sufficiently accurate model for control by treating the 4RW-MAV as a rigid-body, possibly including the blade-flapping dynamics and the main rotor speed.

Let $\{e_N, e_E, e_D\}$ the inertial axes and $\{x_B, y_B, z_B\}$ the body axes. Euler angles of the body axes are $\{\varphi, \theta, \psi\}$ with respect to the e_N , e_E and e_D axes (roll, pitch, yaw). Let *r* the position vector from the inertial origin to the vehicle center of gravity and ω_B the angular velocity. The current velocity direction is referred to as e_v in inertial coordinates.

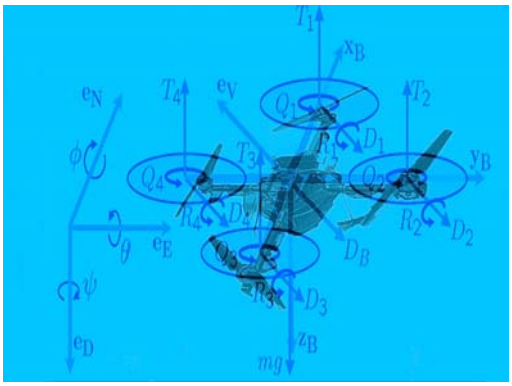


Figure 5 Diagram of a quad-rotor aircraft

The rotors, numbered 1−4, are mounted outboard on the x_B , y_B , $-x_B$ and $-y_B$ axes, respectively, with position vectors r_i with respect to the CG. Each rotor produces an aerodynamic torque, Q_i, and thrust, T_i, both parallel to the rotor's axis of rotation, and both used for vehicle control.

Here, $T_i \approx u_i \frac{k_t}{1+0.1s}$, (1)

where u_i is the voltage applied to the motors, as determined from a load cell test. In flight, T_i can vary greatly from this approximation. The torques, Q_i, are proportional to the rotor thrust, and are given by $Q_i = k_r T_i$. Rotors 1 and 3 rotate in the opposite direction as rotors 2 and 4, so that counteracting aerodynamic torques can be used independently for yaw control. Horizontal velocity results in a moment

on the rotors, R_i, about $-e_v$, and a drag force, D_i, in the direction, $-e_v$. The body drag force is defined as D_B, vehicle mass is m, acceleration due to gravity is g, and the inertia matrix is $I \in R^{3x3}$. A free body diagram is depicted in Figure 2. The total force, F, and moment, M, can be summed as,

$$F = -D_B e_v + mge_D + \sum_{i=1}^{4}\left(-T_i z_B - D_i e_v\right)$$
$$M = \sum_{i=1}^{4}\left(Q_i z_B - R_i e_v - D_i\left(r_i \times e_v\right) + T_i\left(r_i \times z_B\right)\right) \quad (2)$$

The full nonlinear dynamics can be described as,

$$m \cdot \ddot{r} = F$$
$$I\dot{\omega}_B + \omega_B \times I\omega_B = M \quad (3)$$

where the total angular momentum of the rotors is assumed to be near zero, because they are counter-rotating. Near hover conditions, the contributions by rolling moment and drag can be neglected in Equations (1) and (2). Define the total thrust as $T = \sum_{i=1}^{4} T_i$. The translational motion is defined by,

$$m\ddot{r} = F = -R_\psi \cdot R_\theta \cdot R_\varphi T z_B + mge_D \quad (4)$$

where R_φ, R_θ, and R_ψ are the rotation matrices for roll, pitch, and yaw, respectively. Applying the small angle approximation to the rotation matrices,

$$m\begin{bmatrix}\ddot{r}_x\\ \ddot{r}_y\\ \ddot{r}_z\end{bmatrix} = \begin{bmatrix}1 & \psi & \theta\\ \psi & 1 & \varphi\\ \theta & -\varphi & 1\end{bmatrix}\begin{bmatrix}0\\ 0\\ -T\end{bmatrix} + \begin{bmatrix}0\\ 0\\ mg\end{bmatrix} \quad (5)$$

Finally, assuming total thrust approximately counteracts gravity, $T \cong \overline{T} = mg$, except in the e_D axis,

$$m\begin{bmatrix}\ddot{r}_x\\ \ddot{r}_y\\ \ddot{r}_z\end{bmatrix} = \begin{bmatrix}0\\ 0\\ mg\end{bmatrix}\begin{bmatrix}0 & -\overline{T} & 0\\ \overline{T} & 0 & 0\\ 0 & 0 & 1\end{bmatrix} + \begin{bmatrix}\varphi\\ \theta\\ T\end{bmatrix} \quad (6)$$

For small angular velocities, the Euler angle accelerations are determined from Equation (3) by dropping the second order term, $\omega \times I\omega$, and expanding the thrust into its four constituents. The angular equations become,

$$\begin{bmatrix}I_x\ddot{\varphi}\\ I_y\ddot{\theta}\\ I_z\ddot{\psi}\end{bmatrix} = \begin{bmatrix}0 & l & 0 & -l\\ l & 0 & -l & 0\\ K_r & -K_r & K_r & -K_r\end{bmatrix}\begin{bmatrix}T_1\\ T_2\\ T_3\\ T_4\end{bmatrix} \quad (7)$$

The 4RW-MAV control is based on a 4-dimensional action space: the cyclic pitch controls i_{lon}, i_{lat}, which cause the 4RW-MAV to pitch forward/backward or sideways; the tail rotor (rudder) control i_{rud}, which affects tail rotor thrust, and can be used to yaw (turn) the 4RW-MAV; the main rotors collective pitch control i_{col}, which changes the main rotors thrust by changing the pitch of the rotor blades.The interest is to use a dynamics model with a relatively small number of parameters to be estimated from flight data. In this case we first subtracted the effects of inertia and gravity, and then learn a model from data to predict accelerations in a coordinate frame attached to the 4RW-MAV. We integrate the accelerations over time to obtain position, velocity, orientation, angular rate and main rotor speed. The simplified dynamics uses the following parameterization:

$$\dot{u} = v \cdot r - w \cdot q - g_u + C'_u \cdot [u] \quad (8)$$

$$\dot{v} = w \cdot p - u \cdot r - g_v + C'_v \cdot [v] \quad (9)$$

$$\dot{w} = u \cdot q - v \cdot p - g_w + C'_w \cdot \left[1; w; i_{col} \cdot \Omega; \sqrt{u^2 + v^2}\right] \quad (10)$$

$$\dot{\Omega} = C'_\Omega \cdot \left[1; \Omega; i_{col}; w; \sqrt{u^2 + v^2}; \left(i_{lat}^2 + i_{lon}^2\right)\right] \quad (11)$$

The velocities (u, v, w) and angular rates (p, q, r) are expressed in the 4RW-MAV's reference frame. Here g_u, g_v, g_w refer to the components of gravity in the 4RW-MAV's reference frame; is the main-rotor speed.

4 CONCLUSIONS AND FUTURE WORK

The 4RW-MAV configuration has similar characteristics to the traditional shrouded configuration. The main differences are in the payload entity, the maneuverability and the portability. The payload, instead of being from 3 to 20 kilograms, is less than one kilogram. In this way it is the best choice for scenarios whose payload is an optical camera or a simple IR camera. This is the case of the different indoor missions. In this case the maneuverability is crucial and thus the configuration must provide a very high level of movement.

The 4RW-MAV configuration is able to maneuver in a very fast and effective way, moving in a way not possible for the other configurations. The portability can fulfill the needs of a typical mission, in which the rotorcraft should be carried on by a single person and should become operational in a very short time. The weaknesses of the 4RW-MAV configuration are the low speed, low endurance and short range.

We found that the net effect relevant for control during an quad rotor autorotation landing was sufficiently well captured by adding a vertical offset relative to the vertical position predicted in the absence of ground effect. This vertical offset was easily estimated from flight data and taken into account accordingly.

References

[1] BAGNELL, J., SCHNEIDER, J.: *Autonomous helicopter control using reinforcement learning policy search methods*. In International Conference on Robotics and Automation. IEEE, 2001.

[2] La CIVITA, M., PAPAGEORGIOU, G., MESSNER, W. C., KANADE, T.: *Design and flight testing of a high-bandwidth H1 loop shaping controller for a robotic helicopter*. Journal of Guidance, Control, and Dynamics, 29(2): 485–494, March-April 2006.

[3] NG, A. Y., COATES, A., DIEL, M., GANAPATHI, V., SCHULTE, J., TSE B., BERGER, E., LIANG, E.: *Autonomous inverted helicopter flight via reinforcement learning*. In: ISER, 2004.

[4] SEDDON, J.: *Basic Helicopter Aerodynamics. AIAA Education Series.* America Institute of Aeronautics and Astronautics, 1990.

[5] BOSCOIANU, M., PAHONIE, R.: "Some Aspects Regarding the Cooperative control problem for flying wing micro aerial vehicles" , 10th WSEAS Int. *Conf. on Mathematical Methods and Computational Techniques in Electrical Engineering (MMACTEE'08)* , Sofia, Bulgaria, May 2-4, 2008.

[6] PAHONIE, R., CÎRCIU, I., BOŞCOIANU, M.: *"An Analysis of Different Rotary Wing Micro Air Vehicles Solutions",* Metalurgia Internațional vol XIV, no.7 Special ISSUE.

[7] ROBERTS, J. M., CORKE, P. I., BUSKEY, G.: *"Low-cost flight control system for a small autonomous helicopter*". In IEEE Int'l Conf. on Robotics and Automation, 2003.

Dipl. Eng. Ionică CÎRCIU , PhD.
"Henri Coanda" Air Force Academy in Braşov
160, Mihai Viteazu St,
500 183 Brasov
Romania
E-mail: circiuionica@yahoo.co.uk

Prof. Dipl.Eng. Mircea BOŞCOIANU , PhD.
"Henri Coanda" Air Force Academy in Braşov
160, Mihai Viteazu St,
500 183 Brasov
Romania
E-mail: mircea_boscoianu@yahoo.co.uk

Assoc. Prof. Dipl. Eng. Stanislav SZABO, PhD.
Faculty of Aeronautics
The Technical University of Košice
Rampová 7
041 21 Košice
Slovak Republic
E-mail: stanislav.szabo@tuke.sk

THE EVALUATION OF INTERIOR CAR´S AIR QUALITY AND SAFETY OF TRAFFIC

Štefan ČORŇÁK, Pavel BRAUN

Abstract: Looking at the traffic accidents statistics which has been collated for long time period it is remarkable that the main contributors to the traffic accidents are drivers (more than 90 %). An environment quality it is one of the essential contributors to the driver's performance apart of the good life style (feeding, corporal activities, sleeping, resting, etc.). The air quality of the driver's environment is the main contents of the contribution.

Keywords: Traffic safety, driver's environment, driver's microclimate, carbon dioxide.

1 INTRODUCTION

A number of means of transport have been increasing constantly which has a negative effect upon the number of road accidents. Main causes of accidents are well known and might be generally divided into the accidents caused by technical and human errors. It turns out that systematic reliability and safety growth of technical parts and vehicles systems as a whole can significantly reduce the purely technical errors of transport means. Up to 50% of all traffic accidents have been caused due to drivers´ attention decrease [1]. All drivers regardless of their age, sex and previous experience might face the attention decrease. It is generally known that a human being (a driver) can focus his attention on a certain number of items in his view range for a limited period of time only. After some time, usually after an hour, the level of attention of most human beings drops. Thus the capability of performing a required activity in reliable and safe way is significantly limited. The attention decrease of drivers is a common phenomenon and of course each driver is subject to it in a different way. The research of driver attention is therefore observed systematically. Researchers all around the world focused their work on the development of the systems which help to control, support and limit driver actions.

Methods measured parameters of driver's attention might be generally divided into the technical (LDW systéme, vehicle velocity, steering wheel turbiny, reaction time, quality of required tasks fulfilment) and the psycho-physiological (eye ball motion and blinkány, EEG signals analysis, face observation, muscular aktivity, blood pressure, heart beating frequency, skin resistence, breathing frequency, blood alcohol measurement) [2]. It results from the information stated above that car manufacturers pay considerable attention to the driver's attention issue.

A driver's working environment, or its final condition in a vehicle's interior in terms of affecting the human being, has also significant effect onto the driver's attention level. The environment condition in the homosphere (i.e., the place where the human stays) is called a microclimate. It is the microclimate of the vehicle´s interior then. The vehicle environment modification enables the passengers to be transported in nice climate conditions and significantly decreases the driver's tiredness, thereby increasing the total vehicle safety.

The vehicle microclimate is determined by [6]: air temperature, air humidity, air flow velocity, air quality (air change, O_2 content, concentration of CO, CO_2, NO_X, dust, etc.).

The air temperature and air humidity are the main contributors to the microclimate quality. Systems like ventilation, heating and air conditioning help to provide thermal comfort in a vehicle. Harmful substances elimination produced by passengers, i.e., CO_2 and the air humidity as a result of breathing, CO as a result of smoking, different smells etc., are meant by the air cleanness. Keeping the air cleanness in a vehicle interior is often neglected even in the most modern cars and is limited only to the monitoring of CO which gets into the vehicle from the outside. It turns out that poor attention is paid to the fact that a lot of drivers spend quite a long time in the vehicle's closed space. Closed (airtight) spaces have many advantages, e.g. they are resistant to the noise incursion from the outside, etc. The main disadvantage is the automatic air change impossibility which causes the air quality worsening. In the graph in Figure 1 there is an approximate composition of the "pure" outside atmospheric air and an approximate composition after breathing out by a human.

It results from the graphs shown above that oxygen is changed to carbon dioxide due to respiration. The inhaled air contains approximately 0,04 % of carbon dioxide (CO_2), which is approx 400 ppm of CO_2, however the exhaled air of an adult contains approx 4 % of CO_2 on average, which is approx 40 000 ppm of CO_2 (which is approx 100 times higher concentration than in the surrounding air). When breathing in the open air the 4 % concentration of CO_2 is likely to be insignificant.

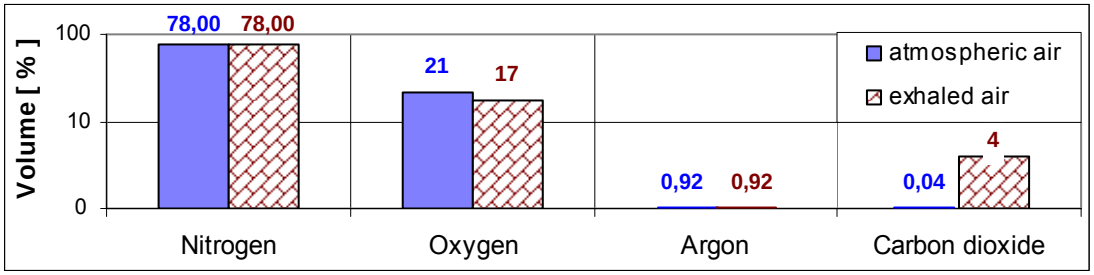


Figure 1 Approximate composition of the „pure" outside atmospheric air and approximate composition after breathing out by a human [5]

However, a different situation might occur in the closed (almost airtight) space. Driver environment is a typical example where a small space and insufficient ventilation might be the main cause of the increase in CO_2 concentration. A driver's sleepiness, lethargy, tiredness might increase when a CO_2 level is higher. Driver attention decline might be finally the main cause of a traffic accident. The microclimate in a vehicle and measuring of CO_2 in vehicles are supposed to be the main subject of the article.

2 CARBON DIOXIDE CONCENTRATION MEASUREMENT

A number of approaches are used for measurement of CO_2 concentration in the air. The most frequently used method is using sensors working with infra red emission absorption (so called NDIR – Non Dispersive Infra Red Method), then sensors working on the electro-acoustic principles and sensors working on the electro-chemical principle. Each of the principles mentioned above has its own advantages and disadvantages [4], [5], [6].

The NDIR sensors are the most frequently used and their principle of work is based on the Lambers Beer law [5]:

$$A = \varepsilon_0 .c.l \tag{1}$$

where: ε_0 -molar absorption coefficient,
l - absorption environment thickness,
c - observed substance concentration,
A - absorbance.

It results from the formulas stated above that the absorbance is directly proportional to the concentration of the observed substance ***c*** when the environment thickness ***l*** is known. A overview of the infrared absorption a wave lenght of gases is introduced in the Figure 2.

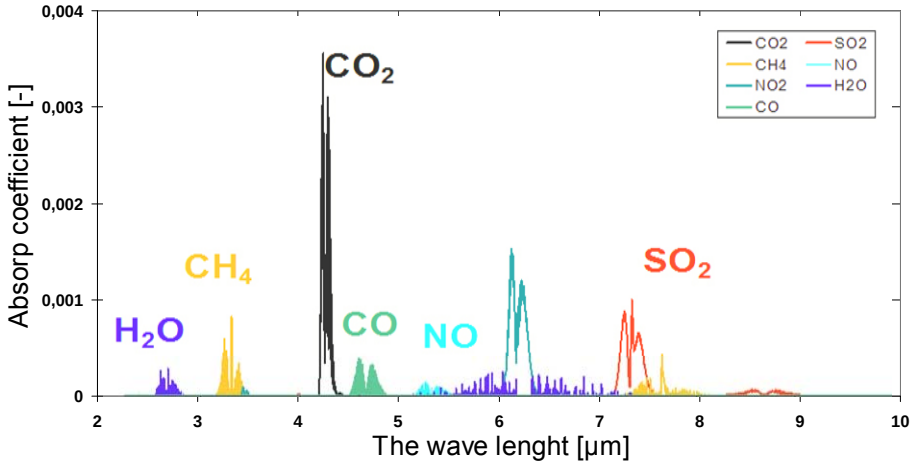


Figure 2 The infrared absorption reprsented by wave length of gases [5]

The non-dispersive infrared sensor (NDIR) works with so called one channel double ray method. In a miniature box there are placed two infrared detectors (see Figure 3, position 4 and 5) and two different optical filters (a measuring position 3 and a reference filter 2).

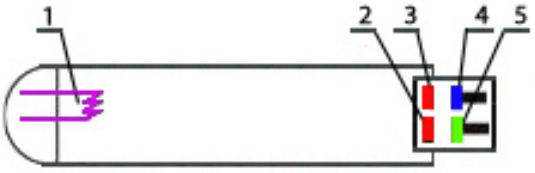


Figure 3 Non-dispersive infrared sensor (NDIR) carbon dioxide (CO_2) infrared sensor [6] where: 1 - Infrared projector, 2 - Interference filter I = 4,00 µm Absorption CO_2, 3 - Interference filter I = 4,27 µm, 4 - Infrared detector I = 4,27µm Absorption CO_2, 5 - Infrared detektor I = 4,0 µm

The measuring filter (see Figure 3, position 3) transmits only the emission with the wave length of approx 4,27 µm. The reference filter (see Figure 3, position 2) transmits the emission with the wave length of approx 4,0 µm. The emission with such wave length is not absorbed by gas. By comparison of both signals the electronic device evaluates the concentration of CO_2 in gas.

3 RESULTS AND DISCUSSION

3.1 Experiment conditions

The sampling spot was placed at the breathing zone level, which means the place where a driver and a passenger inhale its content. It was a low middle class car with internal volume of 2 550 dm^3 with two passengers on board. The vehicle was equipped with manually controlled air conditioning. The total cumulative mileage was 40 508 km and the vehicle did not crash before the measuring. The concentration values were recorded every second during the measuring. During the whole period of measurement the ventilator velocity switch was adjusted to the middle level (see Figure 4, position 2).

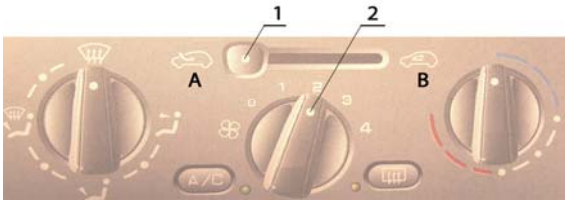


Figure 4 A/C kontrol and vehicle ventilator control[3] where: 1- the air throttle switch, 2- the ventilator velocity switch, A – position of the outside air come in. , B – positron of the inside air come in

The device Testo 435 with a registration part for measured values was used to determine CO_2 concentration in the air. The measuring sampler IAQ of the Testo 435 device has besides a CO_2 sensor a humidity sensor, a temperature sensor, and a pressure sensor too [6]. The record evaluation was carried out with the application of the Komfort-Software Testo V 3.4.

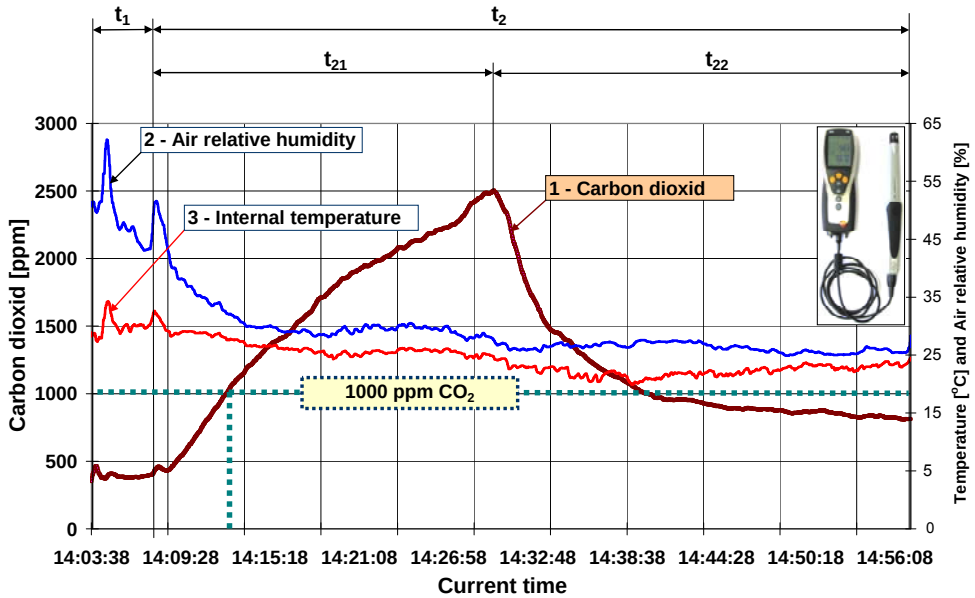


Figure 5 Records of climate conditions measurement inside the vehicle using the Testo 435 device where: 1 – time course of carbon dioxide, 2 – time course of the humidity , 3 - time course of the vehicle internal temperature, t_1 – the period with no people on board, t_2 – the period with passengers on board, t_{21} – the period with passengers on board and with internal air circulation, t_{21} – the period with passengers on board with external air circulation

3.2 Records of climate conditions inside the vehicle

In Figure 5 there is a record of climate conditions inside the vehicle using the Testo 435 device. The whole record might be divided into two main phases: the phase with no people on board (t_1), and the phase with the passengers on board (t_2).

The first phase, in the time period of t_1, started by switching on the Testo 435 device, then the opening of the vehicle and installation of the device inside the vehicle came after. The second phase, in the period of t_2, started with two passengers getting into the vehicle, then the starting of the engine, the closing of the doors and the switching on air conditioning came after. In the time period t_{21} the air throttle was adjusted to the position of internal air circulation in the vehicle, and in the time period t_{22} the air throttle was adjusted to the position of outside air income. During the whole entire period of t_2 the ventilator velocity switch was adjusted to the middle level (see Figure 4, position 2).

It results from Figure 5 that after switching on air conditioning a nice temperature and humidity climate was created (the indoor temperature dropped from 32 °C to approx 28 °C in six minutes and the humidity settled on the approximate level of 31%). The sharp progressive rise of CO_2 which was caused by breathing of the passengers inside the car contrasts with the settled state. The edge limit of 1 000 ppm of CO_2 was reached approximately 4 minutes after getting into the car. CO_2 was rising continuously. The air throttle (see Figure 4, position 1) was adjusted to the position of external air income to the "A" position when reaching the CO_2 concentration of 2 495 ppm. The ventilator velocity switch was left in the middle level (see Figure 4, position 2).

It results from Figure 5 that the CO_2 concentration inside the vehicle dropped down rapidly when opening the air throttle to the "A" position and letting the outside air come in. However, the air temperature and humidity conditions almost did not change. With the view of this information it is obvious that for keeping good air quality in a vehicle (without CO_2) the air ventilation throttle should be closed for a limited period of time only (e.g., when entering tunnels, driving in heavy traffic in cities, etc.)

4 CONCLUSION

The results of the research showed that small space in a vehicle and insufficient ventilation are the main cause of CO_2 rise. High CO_2 concentration can lead to a driver's drowsiness, tiredness, lethargy, etc. Decrease of driver's attention might be the main cause of traffic accidents. Therefore we suggest developing this matter deeper as well as cooperating with vehicles manufacturers, research institutes, and medical institutions. Implementing an on-board diagnostics for CO_2 concentration inside a vehicle could be a good solution too.

This work was completed as a part of the research project of the Faculty of Military Technologies Nr. MO0FVT0000401 and project ME 949 in frame of the programme KONTAKT MŠMT "Analysis of the negative impact onto the drivers´ attention".

References

[1] NOVÁK, M., VOTRUBA Z.: *Human Factor Influencing the Reliability of Car Driving*. Proceeding of the conference: Wearing, dependability and diagnostics 2007, Czech republic, Brno 2007, Nr. 1, pp. 197 -204, ISBN 978-80-7231-294-8. In: Czech language.

[2] TICHÝ, T.: *Methods for exact detection and level of driver attention*. Second book thesis, ČVUT. Praha 2008, In: Czech language.

[3] ČORŇÁK, Š., BRAUN, P.: *On selected aspects of driver's microclimate.* Proceeding of the conference: Wearing, dependability and diagnostics 2008, Czech republic, Brno: 2008, Nr. 1, pp. 21-24, ISBN: 978-80-7231-558-1, In Czech language.

[4] DROPPA. P., SUSEDIK, I.: Thermovision and military mobile technology. Proceeding of the conference: *International Conference on Military Technologies (ICMP 2007)*. Brno : University of Defence, Brno, 2007, pp. 155-162. ISBN: 978-80-7231-238-2.

[5] *Carbon dioxide sensors.* TOP Instruments, s.r.o., Dukelská 367, 534 01 Holice, http://topinstruments.cz/ke-stazeni/prumyslove-a-vzduchotechnicke-snimace-co2.pdf, In Czech language.

[6] Sites www: http://www.testo.cz

Assoc. Prof. Dipl. Eng. Dr. Štefan ČORŇÁK
Assoc. Prof. Dipl. Eng. Pavel BRAUN, CSc.
University of Defence
Kounicova 65
612 00 Brno
Czech Republic
E-mail: stefan.cornak@unob.cz
pavel.braun@unob.cz

DOMAIN-SPECIFIC LANGUAGES FOR COMMAND AND CONTROL SYSTEMS

Ľubomír DEDERA

Abstract: The paper tries to show the parallels in syntactical and semantic processing between traditional high-level programming languages and domain-specific languages. We also demonstrate the possibility of utilization of domain-specific languages in C2 systems.

Keywords: Domain-specific language, programming language, context-free grammar, parser, semantic routines, syntax, command and control system.

1 INTRODUCTION

In the beginning let us look at classical programming languages. The idea of high-level programming languages (PLs) (like C, Java, Lisp, Prolog) is widely known not only in the technical community. They play inevitable role in the process of software development. "A programming language is an artificial computer language designated to express computations that can be performed by a machine. PLs can be used to create programs that control the behavior of a machine, to express algorithms precisely, or as a mode of human communication [1]." Most PLs describe computation in an imperative style, i.e. as a sequence of commands and support object-oriented paradigm of programming (OOP). However, there are PLs supporting declarative programming paradigms such as functional (Lisp) or logical (Prolog) paradigms.

Most PLs belong to the group of general-purpose programming languages. It means that they can be used to code software applications for many various application domains. Typically they are Turing-complete; loosely speaking, according to the Church-Turing Thesis it means that they are capable of describing solutions of all algorithmically solvable problems.

In order to be able to be processed by machines PLs need to have exactly and unambiguously specified their syntax and semantics. Syntax of PLs is usually specified by means of context-free grammars (CFGs) or from CFGs derived Backus-Naur forms (BNFs).

A context-free grammar is a 4-tuple

$$G = (N, T, P, S), \tag{1}$$

where N is a finite set of non-terminal characters (or variables), T is a finite set of terminal characters (or terminals), S, $S \in N$ is the starting symbol of the grammar from which each derivation starts, and P is a finite set of productions (or rewriting rules) of the form

$$B \rightarrow \alpha, \tag{2}$$

where $B \in N$ is the left-hand side and $\alpha \in (N \cup T)^*$ is the right-hand side of the production. An example of a CFG generating a simplified Pascal-like sequence of statements is given in Figure 1.

$G = (N, T, P, S)$
$N = \{$<program>, <stmts>, <stmt>, <expr>$\}$
$T = \{$**begin**, **end**, **if**, **then**, **e**, **s**, **;**$\}$
$P = \{$
<program> → **begin** <stmts> **end**
<stmts> → <stmt> **;** <stmts>
<stmts> → ε[1]
<stmt> → **if** <expr> **then** <stmt>
<stmt> → **s**
<expr> → **e**
$\}$
S = <program>

Figure 1 Example of a context-free grammar

CFGs describing syntax of PLs often need to satisfy additional properties (LL(1), LR(1), LALR(1) grammars) in order to be able to be deterministically parsed by a corresponding type of parser. The role of a parser (or syntactic analyzer) is to determine whether the program being parsed is syntactically correct and to build a parse tree of the program. An example of the parse tree of the "program" **begin if e then s**; **s**; **end** using the grammar from Figure 1 is in Figure 2.

Typically parsers play roles of central controlling units of compilers and interpreters of PLs: They call lexical analyzers (or scanners), which recognize lexical elements (tokens) such as keywords, identifiers, constants, operators and separators. Actually these tokens form the set of terminal symbols T in the CFG (1). In our example, lexical elements are depicted in **bold**. When a parser recognizes a particular syntactical structure (for example, a particular statement or its important part), then it calls a particular semantic routine to interpret the structure that has been recognized. The output of the interpretation might be, for example, "invisible" processing of a declaration of a variable

[1] Symbol ε denotes the empty string.

modifying the symbol and attribute table of the compiler, or a sequence of intermediate (or machine) code instructions realizing the conditional statement, or, as it is in the case of a Prolog interpreter, the direct result of a query.

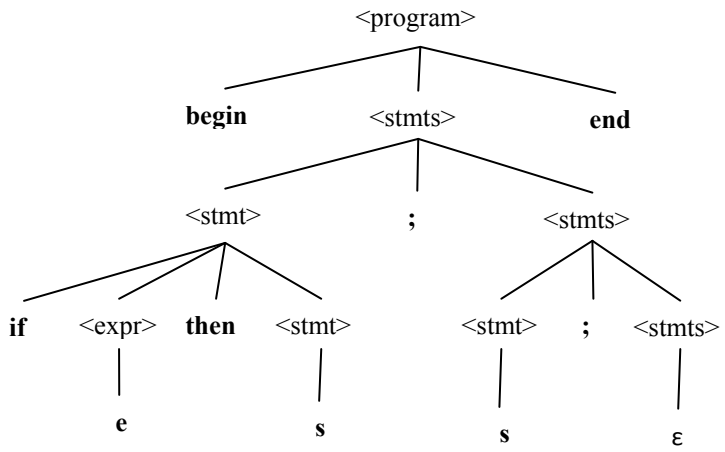


Figure 2 Parse tree of the "program" ***begin if e then s; s; end***

It is not the aim of this article to go into deeper details of the compiler/interpreter construction problems. However, we want to point out that analogical principles regarding syntactical and semantic processing can be utilized in the case of domain specific languages specially tailored for the needs of modern C2 systems. Now we can introduce the idea of domain-specific languages.

2 DOMAIN-SPECIFIC LANGUAGES

2.1 The idea of domain-specific languages

Domain-specific languages (DSLs) are computer languages designed for a specific class of problems and for particular application domains. They might be designed with the intention to be [2, 3]:

- Programming languages dedicated for a particular problem domain, or
- Specification languages dedicated for a particular problem domain.

DSLs can be dedicated to a particular problem domain, a particular problem representation technique or a particular solution technique. The basic idea behind DSLs is to offer means which would allow expressing solutions in the idioms and at the abstraction level of the problem domain. For example, a DSL designed for banking systems could contain language constructions containing keywords such as debit, credit, loan, etc. The consequence is that domain experts (or qualified users) can express, validate or modify solutions described in a particular DSL. It is not necessary for the DSLs to be Turing-complete as it has been mentioned above regarding general-purpose programming languages. On the other side, general-purpose languages are created to solve problems in many application domains.

Examples of widely-used DSLs include the educational programming language LOGO designed for education of basic algorithmic principles at the children's level, hardware description languages (VHDL) for description of digital circuits, Mathematica for symbolic mathematical manipulations, SQL for relational database queries and manipulations, Yacc for compiler construction, etc.

DSLs can have both textual and graphical (visual diagrams) forms. The latter one is popular due to an increasing number of supporting tools for its creation (e.g. Generic Eclipse Modeling System, or Microsoft Visual Studio DSL), the former one usually brings higher productivity.

DSLs can be classified as either internal or external. Internal DSLs are only extensions of existing general-purpose computer languages. They are sets of functions, data structures, and conventions applied to existing languages, such as C++ or Java. On the contrary, external DSLs are independent languages that have been entirely designed for their specific purpose. Generally, a DSL program is a text file, which is then interpreted (or compiled) by the corresponding engine or subsystem.

The great advantage of properly-designed DSLs is that they are both

- **Human-readable** and understandable (in comparison with, for example, XML-based languages, which are also sometimes considered to be human-readable), and

- **Machine-processible**, since they have formally defined syntax and semantics.

DSLs are primarily used in software engineering where they can help to overcome the gap between the worlds of domain experts and implementers of software systems. Their design and implementation are challenging tasks since they require expert knowledge of both the problem domain and the area of computer languages, language processors, compilers, and interpreters.

2.2 Syntax, semantics, and processing of domain-specific languages

Similarly as in the case of high-level programming languages, the syntax of DSLs can be described by CFGs (1). When designing syntax of a particular DSL it is necessary to cope with the following challenges at the same time:

- The syntax should be based on natural language constructions commonly used in the particular context in the problem domain; otherwise there would be a high risk that the users would refuse to use the DSL. For that reason, the knowledge of the problem domain is very important.
- In order to be parseable, the CFG must be a deterministic context-free grammar of a type corresponding to the particular type of parser used.
- Syntax of individual productions has to be designed with respect to semantic processing of the productions.

Generally, there are two main strategies of parsing – top-down, where the parse tree of the given input is constructed from the root towards the leaves and bottom-up, where the parse tree is constructed in the opposite direction. A typical type of top-down parser is a LL(1) parser, which is supported by the ANTLR (ANother Tool for Language Recognition) or JavaCC parser generators. Most commonly-used bottom-up parsers are LALR(1) parsers, which are supported by Yacc/Bison parser generators. In general, bottom-up parsing techniques are more powerful than the top-down ones in the sense of the class of recognized languages.

For the reason of semantic processing we can extend the definition of CFG (1) with semantic action symbols. A CFG with semantic action symbols (CFGSAS) is a 5-tuple

$$G = (N, T, A, P, S), \tag{3}$$

where the sets of variables N, terminals T and the starting symbol $S \in N$ have the same meaning as in the case of an ordinary CFG (1), A is a finite set of action symbols and P is a finite set of productions of the form

$$B \rightarrow \alpha, \tag{4}$$

where the right-hand side $\alpha \in (N \cup T \cup A)^*$; it means that the right-hand sides of productions can contain, besides terminals and variables, semantic action symbols. An example will be given in the next section.

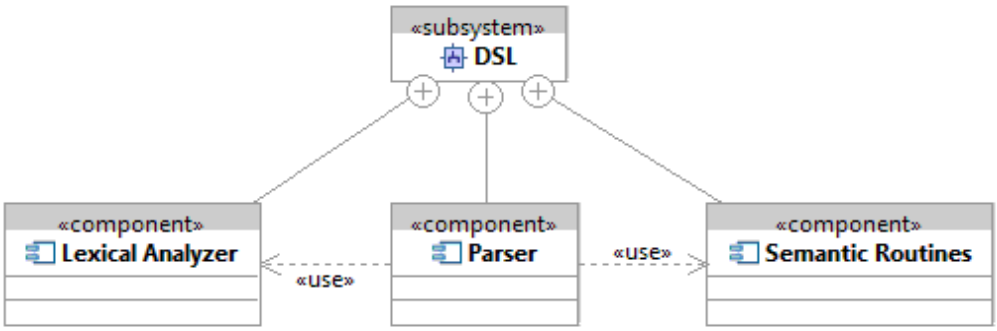


Figure 3 Component architecture of the DSL subsystem

Semantic action symbols play an important role in the DSL subsystem architecture (Figure 3). The parser component (e.g. of LALR(1) type) is the central controlling unit of the whole DSL subsystem. It uses the lexical analyzer to recognize lexical elements from the input. When the parser comes across an action symbol (while processing a particular production or its part), it calls the corresponding semantic routine [4]. Mutual communication among semantic routines can be implemented by means of a semantic stack maintained by the parser. Concrete techniques for implementing a semantic stack depend on the type of parser.

3 UTILIZATION OF DOMAIN-SPECIFIC LANGUAGES WITHIN C2 SYSTEMS

DSLs have promising potential to be utilized within modern C2 systems. This potential comes

from the fact that military application domains have established their own terminology with quite formal syntax and semantics [5]. We consider DSLs to be a supplementary technology that could be used together with the mainstream ontology and semantic web technologies [6]. In the case of C2 systems we see the following potential areas of utilization of DSLs:

- Specification and modeling tools in the process of development of C2 systems, and
- Subsystems of C2 systems themselves.

Let us look at the latter possibility of utilization. A DSL subsystem can be incorporated into the architecture of a C2 system (Figure 4). The DSL subsystem is controlled by the User Interface Subsystem (alternatively it could be an internal part of it) and its role is to process user inputs in the form of documents prepared in the DSL. These documents can have both textual and graphic forms containing, for example, observed and human-processed information about the battlespace awareness and knowledge [5] that can influence the Common Operational Picture (COP), or they could also contain direct commands that could be processed by the Executive Control Subsystem, which could disseminate them to the appropriate entities.

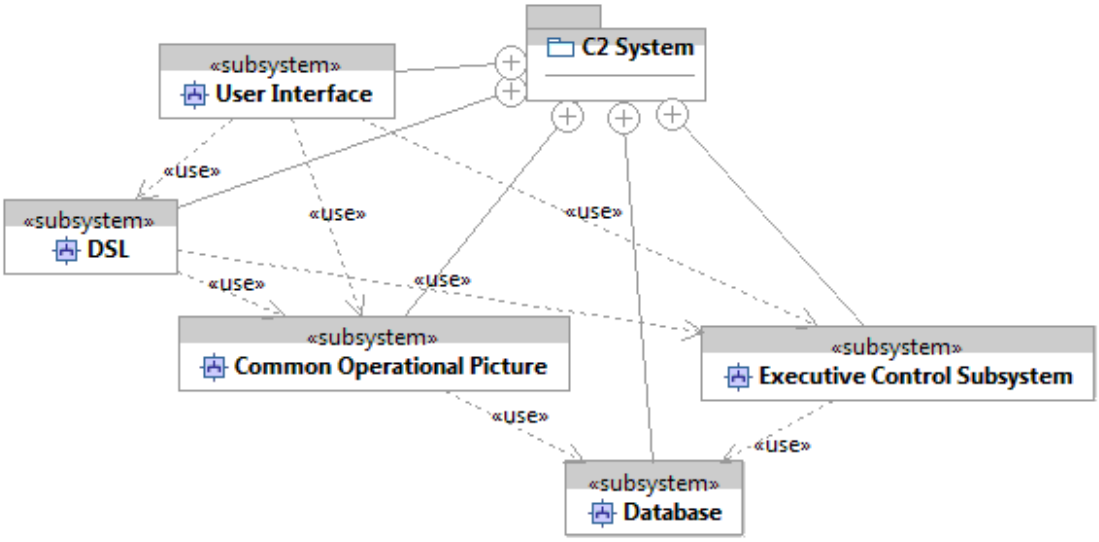


Figure 4 Elements of an architecture of a C2 system that are interesting from a DSLs point of view

To demonstrate the approach described, let us have a look at the following example. We would like to describe syntax and semantics (by means of a CFGSAS and associated semantic routines) of a very simplified form of a sequence of commands, which could be processed by a C2 system. Each command in the sequence can be either a direct command (which immediately triggers an action) or a conditional command (where the action must be triggered by a particular event, in our example such an event occurs only in the moment when our own military unit is being attacked). Each action can be either a destroy action or move action. A destroy action is in fact a command for a particular actor (either a particular military unit or a weapon) to destroy a particular target (either a ground target specified by its coordinates or an attacker). In the case of a move action a particular actor is directed to move its position to the specified coordinates.

The syntax of the sequence of commands mentioned can be described by the CFGSAS with productions in Figure 5. Lexical elements (terminals) of the grammar are depicted in bold and consist of reserved words (**if**, **then**, **destroy**, **move**, **ground**, **target**, **attacker**, **own**, **unit**, **attacked**), open parentheses, commas, semicolons, GPS coordinates constants, military unit and weapon identifiers. All lexical elements mentioned can be described by regular expressions and are recognized by the Lexical Analyzer component. Variables of the grammar are depicted in angle brackets, the variable <commands> is considered to be the starting symbol of the grammar. The grammar is of both LL(1) and LALR(1) types.

The semantics of the language is expressed by semantic action symbols and corresponding semantic routines. Semantic actions symbols (depicted as identifiers starting with #) are included in the right-hand sides of the productions of the grammar. They do not belong to either variables or terminals and their role is to call the corresponding semantic routines by way of the parser. In our example, all action symbols are placed at the ends of productions so the corresponding semantic routines are called when the whole right-hand sides of productions are recognized by the parser.

1. <commands> → <command> **;** <commands>
2. <commands> → ε
3. <command> → <direct command> #process_direct_command
4. <command> → <conditional command> #process_conditional_command
5. <direct command> → <action> #copy
6. <conditional command> → **if** <event> **then** <action> #process_if
7. <action> → **destroy (**<actor>**,** <target>**)** #process_destroy
8. <action> → **move (**<actor>**,** <coordinates>**)** #process_move
9. <actor> → **weapon_id** #process_weapon_id
10. <actor> → **unit_id** #process_unit_id
11. <target> → **ground target** <coordinates> #process_ground_target
12. <target> → **attacker** #process_attacker
13. <event> → **own unit attacked** #process_own_unit_attacked
14. <coordinates> → **coordinates** #process_coordinates

Figure 5 Productions of the context-free grammar with semantic action symbols

For the majority of the semantic routines it is necessary to pass some data (called semantic attributes) to other semantic routines. For example, the semantic routine #process_coordinates needs to pass the GPS coordinates recognized by the lexical analyzer to the semantic routines #process_move or #process_ground_target. This task can be accomplished using a semantic stack. A semantic stack is a second stack maintained by the parser which contains semantic records associated with the terminals and variables of productions being parsed. Each semantic record can contain arbitrary information (semantic attributes) that needs to be passed between individual semantic routines. The semantic records associated with the variables and terminals of individual productions are depicted in the pseudocode of semantic routines in the same way as the corresponding terminals and variables, but using the Courier font. In our example, the semantic routine #process_coordinates stores the coordinates recognized by the lexical analyzer in the semantic record associated with the left-hand side of the production (<coordinates>), where the semantic routines #process_move and #process_ground_target can find it (as semantic records associated with the variable <coordinates> in their right-hand sides) and further process. Structural character of some of the semantic attributes is expressed using C struct/union-like syntax; constant values are in capital letters.

Next we introduce the semantic routines itself. For simplicity we do not include such aspects of real semantic processing as error handling and validity checking. All but last two semantic routines just collect information (semantic attributes) from the lexical analyzer or from preceding semantic routines and pass it (via semantic records associated with the left-hand sides of productions) to the subsequent semantic routines. Last two semantic routines utilize the information collected and call the Executive Control Subsystem (Figure 4) to disseminate the commands processed to the particular end actors (in fact, it might be military personnel and/or machines).

```
#process_coordinates
{
<coordinates> = coordinates
      recognized by the lexical
      analyzer;
}

#process_own_unit_attacked
{
  <event>.type = OWN_UNIT_ATTACKED;
}

#process_attacker
{
  <target>.type = ATTACKER;
}

#process_ground_target
{
  <target>.type = GROUND_TARGET;
  <target>.coordinates =
   <coordinates>;
}

#process_unit_id
{
  <actor>.type = UNIT;
  <actor>.unit_id = unit_id
   recognized by the lexical
   analyzer;
}
```

```
#process_weapon_id
{
  <actor>.type = WEAPON;
  <actor>.weapon_id = weapon_id
  recognized by the lexical
  analyzer;
}

#copy
{
   <direct command> = <action>;
}

#process_destroy
{
  <action>.type = DESTROY_ACTION;
  <action>.actor = <actor>;
  <action>.target = <target>;
}

#process_move
{
  <action>.type = MOVE_ACTION;
  <action>.actor = <actor>;
  <action>.coordinates =
  <coordinates>;
}

#process_if
{
  <conditional command>.condition =
  <event>;
  <conditional command>.action =
  <action>;
}
#process_direct_command
{
  Call Executive Control Subsystem
  to process direct command of type
  <direct command>.type with
  additional parameters depending
  on the type of command and
  contained in <direct command>;
}

#process_conditional_command
{
  Call Executive Control Subsystem
  to process conditional command
  with action
  <conditional command>.action
  and condition
  <conditional command>.condition;
}
```

Finally we will demonstrate the processing of a particular sequence of commands. Using productions in Figure 5 we can formulate the following sequence of two commands, where the commander commands platoon 1 to move its position to the particular coordinates and in the case when the own unit is attacked, he commands tank 2 to destroy the attacker:

move(platoon_1, N49°4'1,21" E19°35'53,62");
if own unit attacked then destroy (tank_2, attacker);

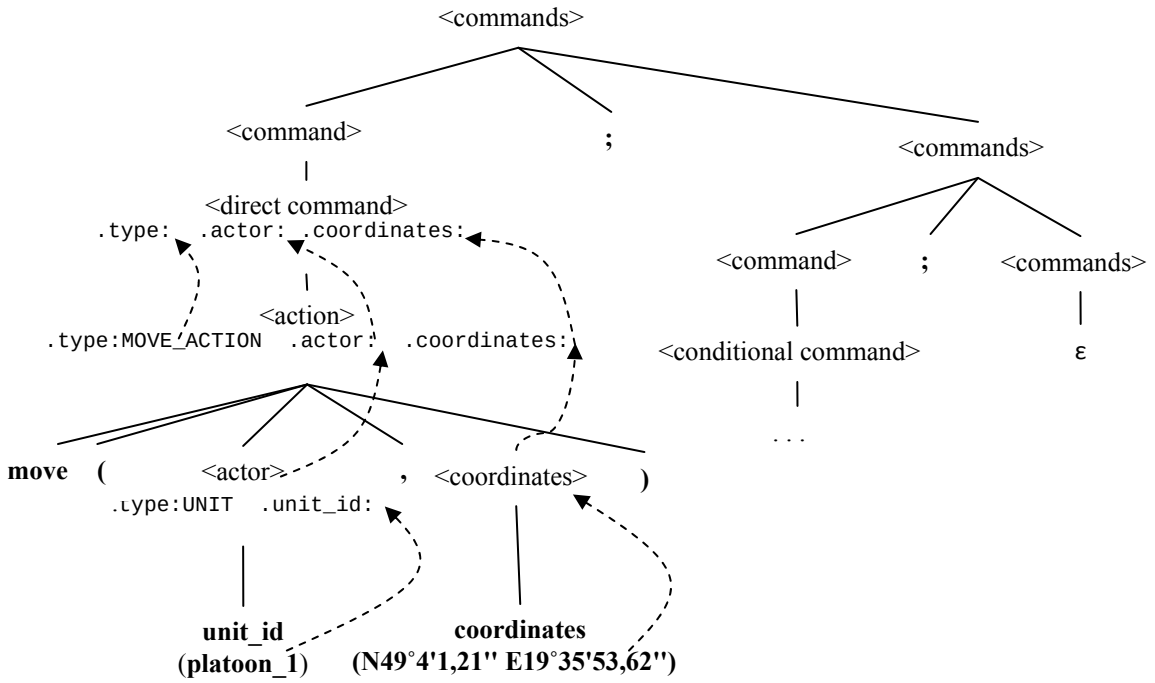


Figure 6 Part of the parse tree of a command sequence with the flow of semantic attributes

A part of the parse tree of these commands depicting the syntactical and semantic processing of the command

move(platoon_1, N49°4'1,21'' E19°35'53,62'');

is in Figure 6. The flow of semantic attributes based on the semantic routines presented above is depicted with arrows. We can observe how individual semantic routines "collect" individual semantic attributes from the lexical elements recognized by the lexical analyzer (**unit_id** with the corresponding sequence of characters **platoon_1**, and **coordinates N49°4'1,21'' E19°35'53,62''**) and pass them up the tree until they are utilized by the semantic routine `#process_direct_command.`

4 CONCLUSION

The aim of this paper was to introduce the notion of DSLs in the context of modern C2 systems and to point out the parallels between programming languages and DSLs regarding their syntactical and semantic processing. We have tried to give a simple example of using DSLs in a C2 system. The role of a computer programmer in relation to a programming language can be analogical to the role of a trained military commander in relation to a DSL. The grammar described in the example was designed for demonstration purposes only and should not be considered as a result of deeper research in this area.

Regarding the utilization of DSLs within C2 systems we can see the following areas for further research:

- Detailed analysis of existing military C2 systems with the aim to identify potential areas of utilization of DSLs;
- Design of a pilot DSL for military C2 systems and verify its utilization during a practical military exercise using simulation tools;
- Combining DSLs and ontology technologies within C2 systems;
- Due to multinational character of military units deployed in various parts of the world we find it interesting to study DSLs with different syntax, but with mutually related semantic processing [7, 8];
- Finally, due to potentially different ways of information processing at various levels of command and control we find it interesting to study the ways how to bind syntax and semantic processing.

References

[1] *Programming language*. (2010, February 18). In Wikipedia, The Free Encyclopedia. Retrieved 10:10, February 22, 2010, from http://en.wikipedia.org/w/index.php?title=Programming_language&oldid=344909286

[2] DEURSEN, A., KLINT, P., VISSER, J.: *Domain-Specific Languages: An Annotated Bibliography*. In: SIGPLAN Notices, vol. 35, no. 6, 2000, pp. 26–36.

[3] FREUDENTHAL, M.: *Domain-Specific-Languages in a Customs Information System.* IEEE Software Vol. 27, Iss. 2 (2010); pp. 65-71.

[4] FISHER, J., LEBLANC, R. J.: *Crafting a Compiler with C.* Benjamin – Cummings Publishing Co., New York, 1992.

[5] ALBERTS, D. S., GARSTKA, J. J., HAYES, R. E., SIGNORI, D. A.: *Understanding Information Age Warfare.* Command and Control Research Program, 2001.

[6] BUŘITA, L.: *Interoperabilita v informačních systémech.* In: Science&Military, 2007, Vol. 2 No. 2; p. 11-18.

[7] PORUBÄN, J.: *Návrh a implementácia počítačových jazykov (Habilitačná práca).* Technická univerzita v Košiciach, 2008, 102 s.

[8] KOLLÁR, J., PORUBÄN, J.: *Building Adaptive Language Systems*. In: INFOCOMP - Journal of Computer Science, Vol. 7, No. 1 (2008), pp. 1-10, ISSN 1807-4545.

Assoc. Prof. RNDr. Ľubomír DEDERA, PhD.
Department of Informatics
Armed Forces Academy of General M. R. Štefánik
Demänová 393
031 06 Liptovský Mikuláš
Slovak Republic
E-mail: lubomir.dedera@aos.sk

CATALOGUE OF ACCOUTREMENTS ALLOWANCES

Petr HARAŠTA, Jiří DVOŘÁK

Abstract: The primary goals and final outcomes of the project "Catalogue of Accoutrements Allowances" are presented in the paper. A virtual catalogue was developed as an application to be employed in the military information network with the aim to provide professional soldiers of any posts, ranks and specializations with a comprehensive survey of the accoutrements allowances contained in all relevant standards. The project was part of the research plan at the Department of Logistics of the Faculty of Economics and Management of the University of Defence in Brno filed under VZ04-FEM-K05-02-SED.

Keywords: Catalogue, standards, criteria, databases, materiel, allowances.

1 INTRODUCTION

The goal of this project was to develop a virtual "Catalogue of Accountrements Allowances" (hereinafter Catalogue) to make it clearly organized into all categories of soldiers, civil employees, workplaces and offices.

During the project the hypothesis was verified whether it is possible to create a network application enabling the database of accoutrements allowances to be easily operated and provide an individual with independent, complex information on accoutrements allowances.

The Catalogue should enable authorized persons to quickly and easily obtain a comprehensive survey of the accoutrements allowances contained in all relevant standards related to their area of interest. The basic goal was connected with the intended implementation of individual parts of information system about the accoutrements standards into the domain "Accoutrements" within CADS.

2 BRIEF PROCEDURE OF TASK SOLUTION

The project started at a time when military regulations, aids, instructions and orders determining and amending the materiel of individual use of soldiers on active duty were considerably fragmented. There was a high number of them and the prescribing, issuing and checking of allowances became disorganized. At the same time, the restructuring of supply system was in progress. The supply system should, among other things, also incorporate civil supply elements in the near future. The above described problem became even more topical, when we realize the fact that newly admitted people often worked in the supply structures at all levels. Some of them were not properly trained for such work. Based on the above mentioned reasons it seemed to be urgently needed in 2004 to introduce a tool for the supply practice which would help the legislative standards in the supply of materiel of individual use to be better arranged.

First the Accoutrements regulation was thoroughly analyzed and all the approaches granting the accoutrements allowances were studied. Then the expert opinions were elaborated in order to correct some inaccuracies and system errors.

Subsequently, the project team started fulfilling the basic task, i.e. the determination of detailed Catalogue classification criteria, which enable a particular person to correctly look up all the accoutrements he/she is entitled to in the database of accoutrements standards.

The following search paths have been selected as the basic groups of Catalogue classification criteria:

1. Type of activity;
2. Gender;
3. Classification category;
4. Army branch – military specialization;
5. Specialization – appointment being held;
6. Acquired education;
7. Rank.

After setting all the classification criteria it was necessary to determine graphically all their mutual connections in order to develop a particular catalogue database.

When the Web administrator v. 1.0 software was developed, it was necessary to load the database with all accoutrements standards and to allocate individual search criteria to individual store items.

3 THE CONTENT OF WEB ADMINISTRATOR V. 1.0 PROGRAM

3.1 Description of Working Cards

The working cards form the basic working environment for all important program functions. All working cards are controlled by file handles functioning similarly to MS Windows.

Figure 1 An Introductory Screen of the Program Called The Catalogue of Accoutrements Allowances in the Army of the Czech Republic

In specific cases the working cards are supplemented with special control elements, which are described below. They are divided into headings, fields and footings.

Card Heading

The name of the card can be found in the heading, e.g. STANDARD CHART – CHANGE OF RECORD and common file handles for displaying a window (as in MS Windows). Then there is a tool bar with basic file handles SAVE (records the changes in the content of the card and saves them) and CANCEL (closes the card without changes).

Card Field

This area displays the data, which are classified into one or more sub folders. The number of folders differs according to the type of card.

Folders

The folder GENERAL is located in all types of working cards except for SEARCHING CRITERIA DATA SETTING. This folder is set as the basic one for displaying the data. In the upper part of the folder there is a chart with data concerning the location of record in the database and the edit box called NAME, into which it is possible to write the name of data entry. There are two fields in the bottom section of the card. The first one is called PICTURE (for locating the image data), the second one is called DESCRIPTION (for recording the text data).

The folder MATERIEL (located only in the working card of the type STANDARD CHART – CHANGE OF RECORD) serves for recording the items belonging into specific accoutrements allowances. It has a folder serving as a file handle for displaying another folder (only if there is more than one folder). In the upper part of the folder there is a tool bar containing file handles NEW ENTRY, EDIT, DELETE, SELECT, and UPDATE. The above mentioned file handles are of standard functioning. The bottom part of the folder is divided into a large CHART WITH ITEMS OF ACCOUTREMENTS ALLOWANCES on the left and the SEARCH TOOLS located on the right edge. The CHART WITH ITEMS OF ACCOUTREMENTS ALLOWANCES includes the data arranged into columns; display not only the location of records in the database, but also the name of materiel, amount, measuring unit and standard lifetime.

The folder DATA is located only in the working card SEARCHING CRITERIA DATA SETTING - STANDARD. The mentioned folder serves to set the classification criteria, which have a fundamental effect on the selection of required items from the database when a certain type of accoutrements allowance is chosen.

In the upper part of the folder there is a tool bar containing file handles SAVE AND CLOSE, SAVE, GROUP OF CRITERIA, and UPDATE. The functions of the file handles are clear from their names except for the GROUP OF CRITERIA. This file handle serves us to select the required group of criteria, which is then displayed in the bottom part of the folder. The chart GROUP OF CRITERIA is

located in the bottom part of the folder. It includes the data arranged into columns, displaying not only the location of records in the database, but also the names of materiel and individual classification criteria related to individual items.

There is a tick box under each criterion in all chart lines. Thus the criterion is selected for each particular item.

The folder NOTES is located only in the working card ENTRY HEADINGS STANDARDS CHART – CHANGE OF RECORD. This folder serves us to create, edit, mark, delete and update the notes to the standards.

In the upper part of the folder there is a tool bar containing file handles NEW ENTRY, EDIT, DELETE, SELECT, and UPDATE. The mentioned file handles have standard functions. The bottom part of the folder is divided into a large CHART OF NOTES on the left, and SEARCH TOOLS located on the right edge. The CHART OF NOTES INCLUDES the data arranged into columns, displaying not only the locations of records in the database, but also the orders of notes and the texts of notes.

Description of the Types of Working Cards without Folders

The working card MATERIEL – LIST functions as a specific code list of data. It includes all types of accoutrements used in the Catalogue of Accoutrements Allowances. Items from the list can be easily selected and added into the standards if a new standard is developed or the current standards are edited.

In the upper part of the card, there is a tool bar located belo the folder containing file handles NEW ENTRY, EDIT, DELETE, SELECT, and UPDATE. The functions of the file handles are clear from their names. The bottom part of the folder is divided into a large CHART OF MATERIEL on the left, and SEARCH TOOLS located on the right edge. The CHART OF MATERIEL includes the data arranged into columns, displaying not only the locations of records in the database, but also the catalogue number of materiel and the name of materiel.

The working card ITEM CODES – LIST functions as a specific code list of data. It includes all types of items used in the Catalogue of Accoutrements Allowances. Items from the list can be easily selected and added into the standards if a new standard is developed or the current standards are edited.

The description of the card is similar to the card MATERIEL - LIST. The only difference is that items are displayed in the chart instead of materiel.

3.2 The Description of Structural Tree of the Catalogue of Accoutrements Standards

The structural tree of the Catalogue of Accoutrements Standards is located on the card REGULATIONS AND STANDARDS - TREE.
In the card heading there is a tool bar including the following file handles:

- UPDATE (it records current data);
- MAXIMIZE/MINIMIZE;
- NEW ENTRY (it creates a new entry in a data tree);
- EDIT (it edits the present entries);
- DELETE (it deletes selected entries);
- SEARCH CRITERIA.

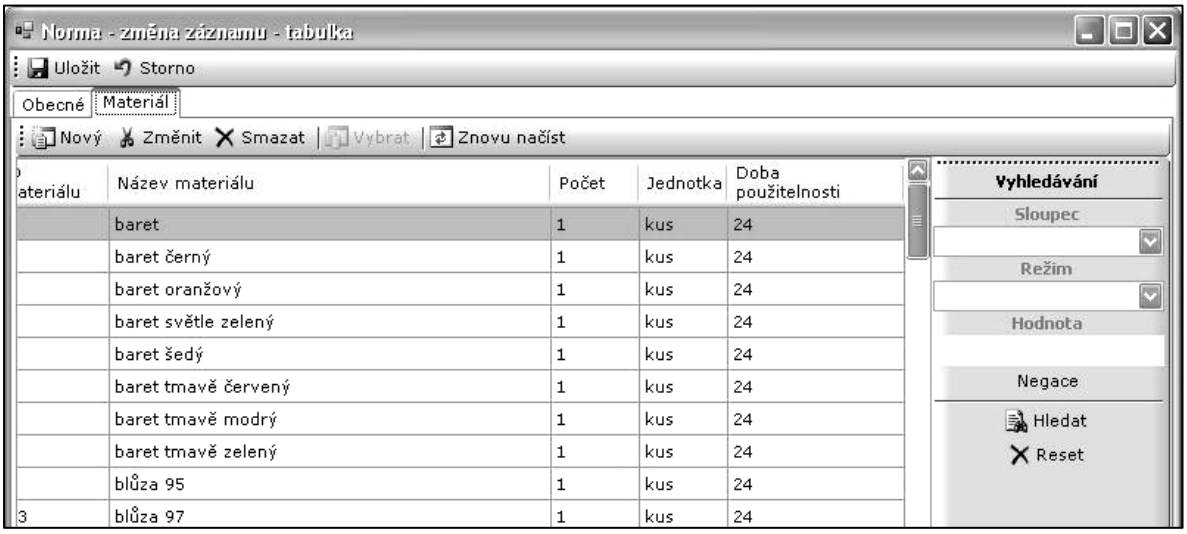


Figure 2 Demonstration of the Screen STANDARD - CHANGE OF RECORD

The basic display of entries in the card includes the following columns: TREE, NAME – TEXT, TYPE, ORDER, and KEY. The column TREE includes a file icon with a file handle (+ / -) for opening or closing the tree structure of data. The column NAME – TEXT includes the name/text of item and the item can be opened for scanning, or editing by either a double-click on the text, or by choosing the file handle EDIT. The column TYPE includes the type of entry (heading, introduction, text, content, annex, and chart). The columns ORDER and KEY include numeric data related to the location of record in the database.

4 CONCLUSION

The following main outcomes were achieved in the course of the project:

1. Web application „Catalogue of Accoutrements Allowances for the Professional Soldiers of the Czech Army".
2. The database structure consisting of 14 mutually connected charts, including, among others, materiel code lists (more than 900 entries), measuring units and notes to individual standards.
3. Digitized data in the range of the Regulation called Vševojsk-16-11, including the 1st and 2nd annexes, selected passages from the amended Regulation Všeob-P-47 and Government Decree No 495/2001 Col.
4. Publications on individual partial observations.

References

[1] SEDLÁK, B., HARAŠTA, P.: *Final Report on the Partial Project of the Research Goal of the Faculty of Economics and Management (VZ04-FEM-K05-02-SED).* Brno: University of Defence, 2008. 249 p.

Maj. Eng. Petr HARAŠTA
Department of Logistics at the
Faculty of Economics and Management
University of Defence
Kounicova 65
662 10 Brno
Czech Republic
E-mail: petr.harasta@unob.cz

Mgr. Jiří DVOŘÁK
Language Training Centre
University of Defence
Kounicova 65
662 10 Brno
Czech Republic
E-mail: jiri.dvorak@unob.cz

MOMENTS HAVING EFFECT ON A FLYING MISSILE

Peter LIPTÁK, Milan JOZEFEK

Abstract: The article describes effects of aerodynamic moments on the antiaircraft missile during the flight. Decomposition of the overall aerodynamic moment on the controlling, dampening and stabilization one enables to analyze their effect on the antiaircraft missile.

Keywords: Aerodynamic force, aerodynamic moment, antiaircraft missile, aerodynamic deployment of missiles.

1 INTRODUCTION

To analyze forces and moments which have effect on a flying antiaircraft missile we can apply basic relations of external ballistic. Fundamental knowledge of the forces and moments as well as its precise formulation has principal importance on securing the encounter between the antiaircraft missile and the target or homing the missile into the safe destruction zone. During the flight the effective forces are:

- overall aerodynamic force R,
- thrust of the rocket engine (reactive force) P,
- gravity G.

The gravity acts in the center of gravity, the reactive force acts in the flight direction and the overall force acts in a specific distance from the center of gravity.

The point where the aerodynamic power acts is called a point of aerodynamic force application and it is always off the center of gravity and thus creates a resultant (overall) aerodynamic moment "M". The point of aerodynamic force application can be placed before or behind the center of gravity according to the aerodynamic arrangement of the missile. The point of aerodynamic force application is a function of the aerodynamic arrangement of the missile and is essential for dynamic parameters of the missile during the flight.

2 CHARACTERISTIC OF MOMENTS ACTING ON THE MISSILE DURING THE FLIGHT

Moments acting on the missile are result of aerodynamic forces acting off the center of gravity.

The resultant aerodynamic power comprising of the sum of particular aerodynamic powers acts always off the center of gravity and thus creates the resultant (overall) aerodynamic moment "M".

The missile is in terms of external ballistic considered as an ideal solid and thus centrifugal and inertial force will not have effect on its movement except the Coriolis force. The latter one emerges at the combustion of the propellant and due to the change of the body weight everywhere where composition of the forward and rotational movement of the missile in a stream of running-out gases can be found. Moments of powers which have effect on the rotational part of the movement can be divided into external forces moments and internal forces moments.

Moments having effect on the missile

- **of external forces**
  - ∗ Basic
    - + Aerodynamic moment
    - + Engine thrust moment
  - ∗ Complementary
    - + Periodic
      - • Magnus force moment
      - • Surface friction moment
    - + Non periodic
      - • Reactive
      - • Aerodynamic
- **of internal forces**
  - ∗ Coriolis force moment at the propellant combustion
  - ∗ Moment of the force appearing by movement of the center of gravity of the missile

Out of the three basic powers the effect on rotational movement has only the thrust of the rocket engine and the air resistance moment, the gravity acts in the centre of gravity and thus does not create any moment.

The engine thrust moment – this moment is a result of deviation of jets from the longitudinal axis of the missile making an angleσ, the jets being symmetrically dispersed along the symmetrical axis of the missile. The thrust of the rocket engine is a sum of partial thrusts along axes of particular jets and its resultant force acts along the axis of the missile. The only deviation is caused by rotation of the missile. A deviation of jets creates a moment given as:

$$M_F = F_t \cdot \sin \sigma r_t \qquad (1)$$

Where σ is an angle of jets deviation

r_t is a radius of the jet span

The air resistance moment is used to stabilize the missile; it means to decrease the angle between the missile axis and the tangent to a missile trajectory. Its effect can be seen on the Figure 1.

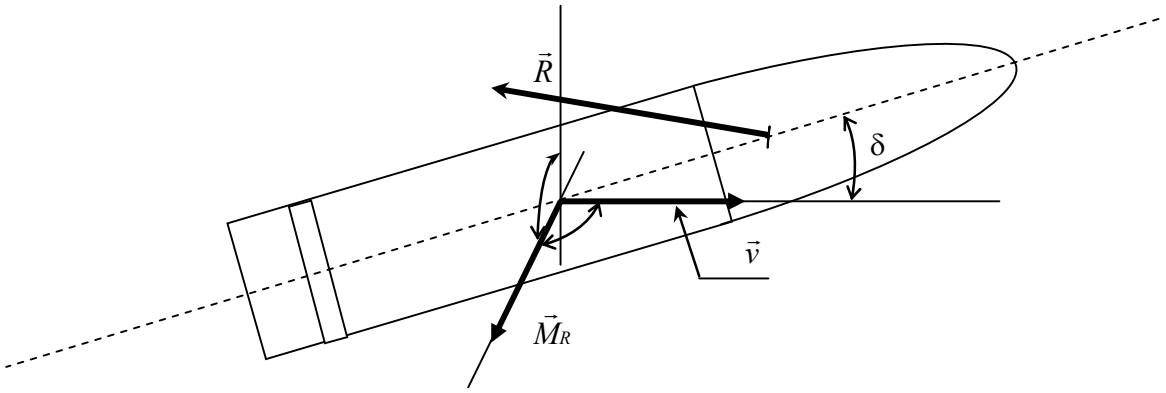


Figure 1 The Air resistance moment

Knowing the air resistance in tangent and line normal to the missile trajectory R_t and R_n we can calculate the air resistance moment by the following formula:

$$M_R = (R_t \sin\delta + R_n \cos\delta)h \qquad (2)$$

where h is a distance of the point of aerodynamic powers application from the center of gravity.

More usually can be M_R calculated through the coefficient of the banking moment acquired by measurement in a wind tunnel. This coefficient is marked $C_M(M)$ and it enables to calculate the air resistance moment through the equation [4]:

$$M_R = \frac{d^2 h}{g} 10^3 v^2 C_M(M)\delta \qquad (3)$$

The Coriolis force moment at the outflow of combustion gases appears - according to the dynamics - inside the accelerated system where the vector of rotational movement contains an angle different from zero with the vector of forward speed. The rotational movement during movement of the missile along the active trajectory is presented by stabilization oscillation around the tangent to the trajectory (Figure 2) and the forward relative movement of the missile regarding to the atmosphere.

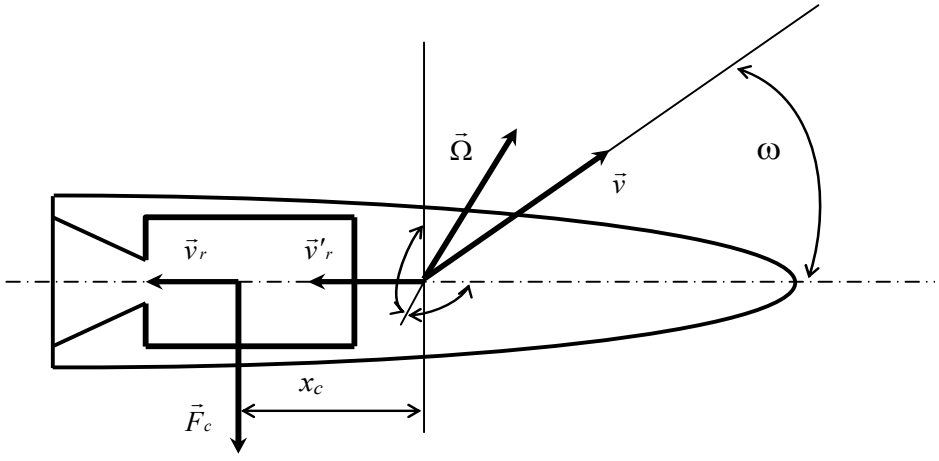


Figure 2 The Coriolis force appears at the outflow of combustion gases

The Coriolis force is given by the vector equation

$$\vec{F}_c = -2m(\vec{\Omega} \times \vec{v}_r) \qquad (4)$$

where $\vec{\Omega}$ is angular speed of the missile oscillation around the tangent to the trajectory, $\vec{v}_r$ is relative speed of running-out gases.

The moment of Coriolis force is then given by the formula

$$M_c = F_c x_c \qquad (5)$$

Where x_c is a distance of the point of application of Coriolis force from the center of gravity.

In all cases when the missile is constructed according to the scheme 5, the moment of Coriolis force acts against the moment of rotational movement which induces the force, it means that it improves the stability of the flight of the missile and thus it is not needed to consider it in the calculation.

The moment of the center of gravity movement force – it is a kind of Coriolis force moment, it appears as a result of movement of the center of gravity at the combustion of the propellant. According to Figure 3 during movement of the center of gravity there exists simultaneous oscilation of the missile around the center of gravity and the forward movement given by the vector of relative speed of the center of gravity v_t. During this process there appears a force given by formula:

$$\vec{F}_{te} = -2m\left(\vec{\Omega} \times \vec{v}_t\right) \tag{6}$$

where $\vec{v}_t$ is relative speed of the center of gravity.

According to the Figure 3 the mentioned force acts also against missile oscilation and it helps to stabilize the missile. That is why we usually do not include it into the calculation.

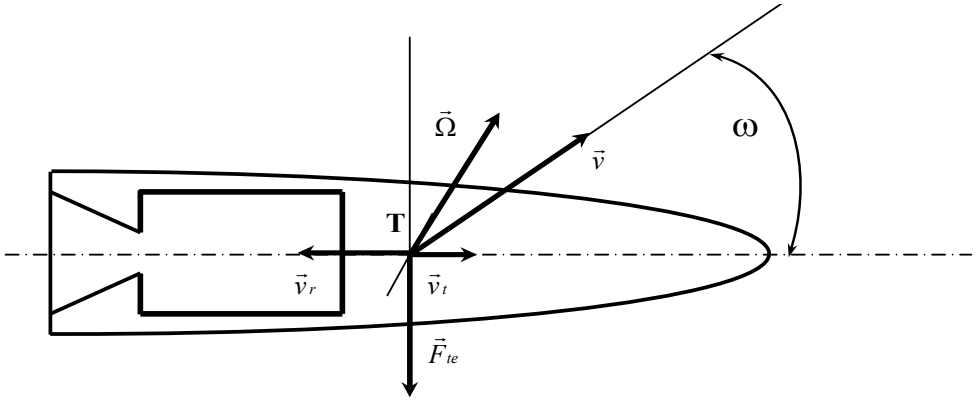


Figure 3 The moment of the center of gravity movement force

It is demanding to find forces and moments mentioned in this lecture and it also depends on experimentally acquired data, of which only the most important are usually taken into solutions of the external ballistics. The object of this lecture is to present overall information about factors having effect on missile movement in the atmosphere. The extent of presentation reflects this requirement.

Moments of complementary forces

All complementary forces act generally in points placed off the center of gravity. Besides having effect on the forward movement they also act on the rotational movement of the missile by means of the moments. They have effect on missile stability; nevertheless its value is insignificant in comparison to basic moments, that is why we consider them only at overall evaluation of the missile stability during the flight.

3 CONSIDERATION OF EFFECT OF THE RESULTANT AERODYNAMIC MOMENT ON THE ANTIAIRCRAFT MISSILE

The resultant aerodynamic moment acting on the antiaircraft missile is principal on its movement along the precalculated trajectory. This moment can be decomposed intothe board coordinate axes system $\vec{M}_{xr}$, $\vec{M}_{yr}$, $\vec{M}_{zr}$. The resultant moment is given by:

$$\vec{M} = \vec{M}_{xr} + \vec{M}_{yr} + \vec{M}_{zr} \tag{7}$$

The moment $\vec{M}_{xr}$ causes rotation of the missile along the longitudinal axis and we call it the banking moment. Moments $\vec{M}_{yr}$ and $\vec{M}_{zr}$ cause missile turning around corresponding axes y_r and z_r and we call them moment of direction and banking (transverse). Each one of moments is analyzed as a sum of three moments: the stabilization $\vec{M}_{ST}$, controlling $\vec{M}_{RIAD}$ and dampening $\vec{M}_{TL}$ one. The stabilization banking moment appears only at asymmetric airflow around the missile when angles $\alpha \neq 0$ and $\beta \neq 0$, that means the resultant aerodynamic force is not along the longitudinal axis of the missile. Since its value is small it is usually not considered. Effect of moments acting along particular axes is analogical. Let's analyze components of the longitudinal banking moment:

$$\vec{M}_{zr} = \vec{M}_{zrST} + \vec{M}_{zrRIAD} + \vec{M}_{zrTL} \quad (8)$$

where M_{zrST} – stabilization moment
M_{zrRIAD} – controlling moment
M_{zrTL} – dampening moment

The stabilization moment $\vec{M}_{ST}$ appears when the missile longitudinal axis is deviated from the speed vector $\vec{v}_r$ in the vertical plane, it means that there is an angle of pitch (angle of attack) of the missile "α". It is caused by the uplift force of wings and the body of the missile with regard to its center of gravity. The uplift force depends on the aerodynamic arrangement of the missile, speed of the flight, density of air and the angle of pitch of the missile. The value of stabilization moment is given by the formula:

$$\vec{M}_{zrST} = m_{zr}^{\alpha} \cdot \frac{\rho v_r^2}{2} \cdot S \cdot b \cdot \alpha \quad (9)$$

where: m_{zr}^{α} – is a coefficient of the stabilization moment dependent on geometry characteristics, shape of the missile and the number "M",
b – is a characteristic linear dimension of the missile, usually the medium aerodynamic chord,
S – characteristic surface.

The stabilization moment is a function of speed, flight altitude and angle of attack, it can be expressed by a formula:

$$\vec{M}_{zrST} = M_{zr}^{\alpha}\left(v_r, H\right) \cdot \alpha \quad (10)$$

The dampening moment $\vec{M}_{TL}$ appears at rotation of the missile in the air stream around the axis "0,z_r" at specific angular speed "ω_{zr}".

$$\vec{M}_{zrTL} = m_{zr}^{\omega} \cdot \frac{\rho v_r^2}{2} \cdot S \cdot \frac{b^2}{v_r} \cdot \omega_{zr} = M_{zr}^{\omega} \cdot \omega_{zr} \quad (11)$$

Its value depends on the speed of missile rotation, air density, speed of the flight of the missile, geometric characteristics of the missile and shape of the missile. It is defined by the formula:

$$\vec{M}_{zrTL} = M_{zr}^{\omega}\left(v_r, H\right) \cdot \omega_{zr} \quad (12)$$

The dampening moment is always directed in the opposite direction to angular speed of missile rotation.

The controlling moment $\vec{M}_{zrRIAD}$ appears at deviation of top rudders (deflectors) in consequence of the uplift force acting on rudders which are situated off the center of gravity. Its value is given by the formula:

$$M_{zrRIAD} = m_{zr}^{\delta} \cdot \frac{\rho v_r^2}{2} \cdot S \cdot b \cdot \delta_v \quad (13)$$

where: m_{zr}^{δ} – is a coefficient of the controlling moment which depends (at given aerodynamic arrangement of the missile) on a number "M",
δ_v – is an angle of deviation of top rudders,
b – is a characteristic linear dimension of the missile, usually the medium aerodynamic chord.
S – is a characteristic surface.

A character of dependency of the coefficient m_{zr}^{δ} on the number "M" is proportional in the same way as the coefficient m_{zr}^{α}, thus we can write:

$$\vec{M}_{zrRIAD} = M_{zr}^{\delta}\left(v_r, H\right) \cdot \delta_v \quad (14)$$

Considering aerodynamic arrangement of the missile we distinguish following types of the missile.

4 AERODYNAMIC ARRANGEMENT OF MISSILES

The arrangement of anti aircraft guided missile control deflectors usually is:
- normal (Figure 4)
- without tail deflectors (Figure 5)
- canard (Figure 6)
- with turning wings (Figure 7)

At the normal aerodynamic arrangement the uplift caused by turning of deflectors should be subtracted from the uplift of the body and wings which appears when the speed vector contains an angle α with the longitudinal axis of the missile. That means that there is certain loss of the control force due to the balancing of the missile. For normal aerodynamic arrangement there stands:

$$\left(\frac{\alpha}{\delta}\right)_{st} < 0 \quad (15)$$

Usage of the aerodynamic arrangement without tail deflectors is suitable for missiles destroying targets at high altitudes and high approaching speeds.

4.

5.

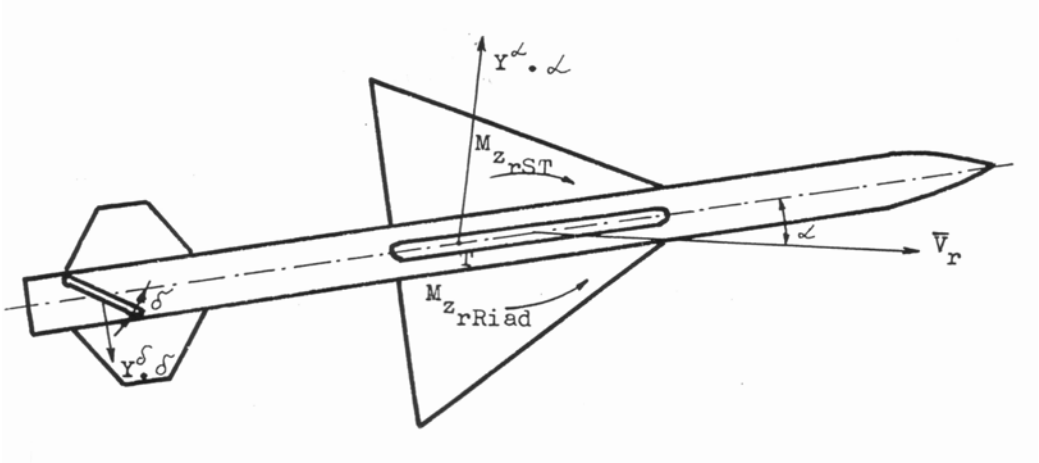


Figure 4 The missile with the normal aerodynamic arrangement

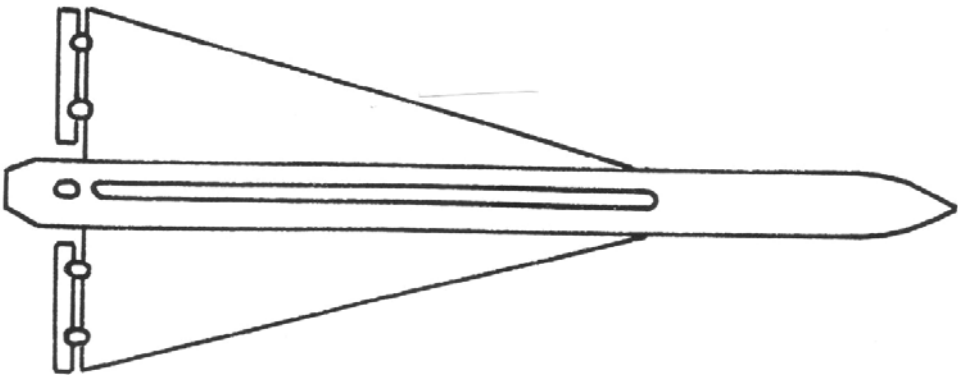


*Figure 5 The aerodynamic arrangement without tail deflector*s

Deflectors of the aerodynamic arrangement Canard are situated in the front part of the missile and the uplift on them is as important as the uplift of the wings and body caused by the pitch angle α . Stabilization pitch angles are corresponding with the angle of deflectors deviation:

$$\left(\frac{\alpha}{\delta}\right)_{st} > 0 \qquad (16)$$

A placement of deflectors in the front part of the missile causes incurvation of the air stream during their deviation, that causes a loss of uplift on wings. That may appear as a disadvantage in comparison to the normal aerodynamic arrangement. Regardless of disadvantages of the canard arrangement this arrangement is generally used. The missile with deflecting wings is depicted on the Figure 7. At this arrangement the function of deflectors and wings is common. In the rear part of the body there may be situated stabilization winglets.

The principle of guiding the missile with help of aerodynamic forces means deviating of deflectors by given angle after receiving an order. There appears the uplift force on wings which induces the controlling moment with regard to the centre of gravity. Under the effect of the controlling moment the missile starts to move around the longitudinal axis"0, z_r", thus making an angle of pitch (angle of attack) α.

When we have a certain angle of attack, on wings appears the uplift which induces the stabilization moment tending to decrease the angle of attack. Rotation of the missile around the transverse axis is stopped when the resultant moment acting on the missile equals zero. Each deflection δ means corresponding angle and the value of the control force under given circumstances. The control force changes the direction of the missile flight, it sets the value and direction of the normal acceleration which is perpendicular to the vector of the missile speed.

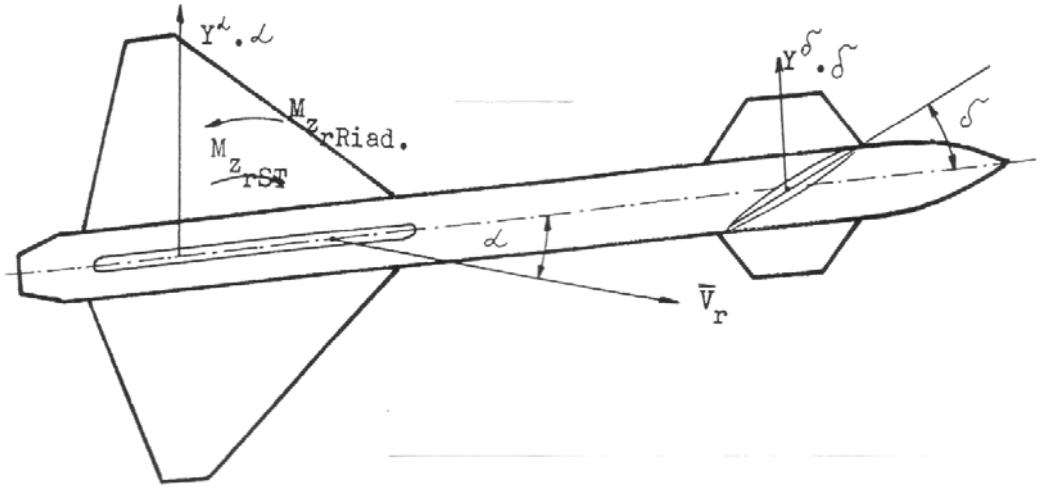


Figure 6 The aerodynamic arrangement "Canard"

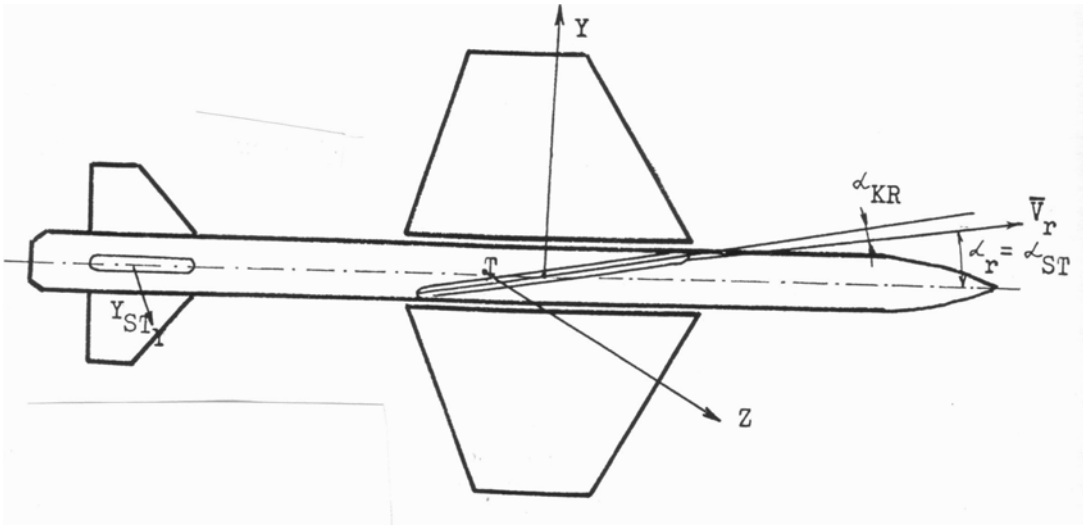


Figure 7 The missile with deflecting wings

The maneuverability of the missile at high altitude is realized by increasing of the wings surface, thus increasing weight and dimension of the missile. That is why by surpassing the certain height the missile is not navigated by aerodynamic forces, but by gas rudders in the stream of running-out gases from the missile jet. This enables effective navigation of the missile in thin atmosphere.

The maneuverability of the missile is usually characterized by rate of change of direction and value of the speed vector; it means the value of normal tangent acceleration.

5 CONCLUSION

Moments acting on the missile during the flight are principal for navigation of the missile along the precalculated trajectory considering mutual motion of the missile and target. Those moments together with forces acting on the missile during the flight are initial values for calculation of flight commands.

References

[1] JOZEFEK, M.: *Teória streľby prostriedkov* PVPŠ. VVTŠ, Liptovský Mikuláš : 1983.
[2] NEUPOKOJEV, F., K: *Streľba zenitnymi raketami* .Vojenizdat Moskva : 1984.
[3] JIRSÁK, Č., KODYM, P.: *Vnejší balistika a teorie střelby*. Naše vojsko, Praha : 1985.
[4] *Spravočnik oficera protivozdušnoj oborony*. Vojenizdat, Moska : 1987.
[5] SKOUPÝ, A.: *Vnejší balistika.* 1. část VAAZ Brno: 1988.
[6] SKOUPÝ, A.: *Vnejší balistika.* 2. část VAAZ Brno: 1989.

[7] JOZEFEK, M.: *Automatizované systémy špeciálnej techniky II.* TnUAD Trenčín : 2004, ISBN 80–8075–023–8.
[8] LIPTÁK, P., JOZEFEK, M.: Efektívnosť bojového pôsobenia prostriedkami PVO. In: *Vedecké práce a štúdie Fakulta špeciálnej techniky*. ISBN 80-8075-170-6. ISSN 1336-9008. - č. 5 (2006), s. 96-103.
[9] LIPTÁK, P., JOZEFEK, M.: Destroing of air targets under elektronic jaming. In: *Proceedings of the International conference on military technologies* 2 – 4 May 2007. s 174 – 179. University of Defence. Brno 2007.

Assoc. Prof. Dipl. Eng. Peter LIPTÁK, CSc.
Alexander Dubček University of Trenčín
Študentská 1
911 50 Trenčín
Slovak Republic
E-mail: liptak@tnuni.sk

Assoc. Prof. Dipl. Eng. Milan JOZEFEK, CSc.
Alexander Dubček University of Trenčín
Študentská 1
911 50 Trenčín
Slovak Republic
E-mail: jozefek@tnuni.sk

INTELLIGENT AGENTS IN MILITARY DECISION MAKING

Florin MOISESCU, Mircea BOŞCOIANU, Gabriela PRELIPCEAN, Mariana LUPAN

Abstract: Military decision making demands an increasing ability to comprehend and structure the critical information on the battlefield. As the military evolves into a networked force, strain is placed on headquarters and others to collect and utilize information from across the battlefield in a timely and efficient manner. Human capability in analyzing all the data is not sufficient because the modern battlefield is characterized by dramatic movements, unexpected evolutions, chaotic behavior and non-linearities. It results a great need of powerful AI assistance in military decisions.

Keywords: Artificial intelligence (AI), AI algorithms, military decision making, course of action (CoA).

1 INTRODUCTION

Military decision should consider information about the huge range of assets and capabilities (human resources combat and support vehicles, helicopters, sophisticated intelligence and communication equipment, artillery and missiles) that may perform complex tasks of multiple types: collection of intelligence, movements, direct/ indirect fires, infrastructure, and transports.

Military decision should also respond to the emerging evolutions that have increased under the pressure of multiple factors (new military-political realities, the asymmetric warfare, the rise in operations other than conventional war) in a new, more effective way. The decisional factor needs an integrated framework capable to perform the critical steps, from capturing a high-level course of action (CoA) to realizing a detailed analysis/ plan of tasks (Hayes, Schlabach 1998, Atkin, 1999, Tate, 2001, Kewley, Embrecht) and one possibility is to be based on different AI techniques, ranging from qualitative spatialinterpretation of CoA diagrams to interleaved adversarial planning/ scheduling.

Given the logistics consumption and the complexity of time/space analysis, the classic decisional process is time and manpower consuming (Bohman 1999, Paparone 2001) and is dramatically limiting the number and diversity of options able to explore and analyze (Banner 1997). The AI ingredient permits to explore a greater range of options, enabling the staff to analyze more possible options in the same amount of time, together with a deeper analysis of these options.

The military planning process is typically composed on the following steps: *initiation:* corresponds to mission trigger and task reception; *orientation:* includes mission assessment, mission statement and decision maker's planning guidance; *concept development:* includes staff's analysis, friendly and enemy courses of action development and analysis, and decision maker's estimate; *decision:* includes courses of action comparison and selection, course of action approval, decision maker's direction, review of critical assumptions; *plan development:* mainly concerned by synchronization and finalization; *plan review:* includes analysis and revision of plans.

Elaboration, mitigation and evaluation of different CoAs are significant steps in planning process. CoA development and analysis are exercises in which are simulated different situations. Time constrains the process to generate a complete range of CoAs, and evaluate them according to significant point of views, before selecting and executing the optimal one.

2 A REVIEW OF THE POSSIBILITIES TO INTRODUCE AI ALGORITHMS IN MILITARY APPLICATIONS

AI based military decision behavior models can be classified into the following groups: models based on neural networks (NN), Bayesian belief networks (BBN), fuzzy logic (FL), geneticalgorithms (GA) and expert systems (ES).

a) Neural Networks applications

NN philosophy is based on the concept of a neuron as a unit for information storage and mapping input to output. NNs are based on the connection of sets of simple processing elements/ nodes, where a weight is associated to each connection between nodes. Weights are initialized randomly at the beginning, and as the network begins to learn, the weights change. The neuron receives a numerical input vector (binary or part of a continuum) and each element of the input vector is scaled by a weighting constant, which assigns the importance rank to each input. The result of the dot product is used into a squashing function whose output is used as the input to another neuron.

Other types of networks are self-organizing maps (SOMs) in which neurons are connected together in a grid such that each neuron is connected only to its neighbors, receiving input from the bottom and giving output at the top. SOM-like networks excel at picking out features from images. Other network types include recurrent Hopfield networks and stochastic Boltzmann machines.

NNs can be trained to produce specific outputs for specific inputs and also to produce specific answers for specific kinds of inputs. This leads to their most common usage: pattern recognition. Their status as a decision algorithm rests on their ability to classify inputs for which they have not been previously trained. The greatest disadvantage of NNs is that they are exceedingly slow to train because they are usually run on a single processor computer and do not take advantage of their massive parallel processing potential - the potential that nature maximizes in human brains. We see the same problem later in GAs.

NNs are usually used for pattern recognition or classification but they are poor in decision-making applications because they lack computational efficiency and tend to act as a black box unless a laborious query-and-response procedure is undertaken to develop rules after training is complete. NNs have been successfully applied to automatic target recognition (Rogers, 1995), data fusion (Bass, 2000; Filippidis, 2000), agent-based, recognition-primed decision models (Liang, 2001) and determining decisive points in battle planning (Moriarty, 2001).

In Bayesian belief networks (BBN), the architecture is designed in accordance with expert knowledge instead of trained. BNN allow users to develop a level of confidence that a particular object will be in a particular state based on certain available information. Belief networks add probability to facts and inferences that indicates how much credence the fact lends to an inference. In (Starr, 2004), BNN are "directed acyclic graphs over which is defined a probability distribution". Each node in the graph represents a variable that can exist in one of several states. A node could be ground forces with different states (attack, withdrawal, defending). The network is set up to represent causal relationships. For example an enemy intention node might be the parent of a ground forces node. Bayesian networks can be solved using conditional probability methods. Bayesian networks are suitable when variables have a small number of states. They could be useful in multi-resolution models where smaller networks can be connected into larger ones and treated as black boxes. They are not a good choice for maneuver or force allocation because of their scalability limitations (probabilities are difficult to assign).

b) Genetic Algorithms

The classic genetic algorithm (GA) begins as a search technique for tackling complex problems. Through the process of initialization, selection, crossover, and mutation, GA repeatedly modifying a population of artificial structures in order to chose an appropriate structure for a particular problem. GAs are useful when the fitness landscape contains high, narrow peaks and wide stretches of barren waste between them, GAs. If the area covered by fitness peaks approaches zero compared to the number of bad solutions in the landscape (good solutions are exceedingly rare) a random problem solver will rarely find a good solution. Real world fitness landscapes correspond to the difficult problems where traditional algorithms fail, and GAs should be applied to these problems.

Some researchers have attempted to use GA in assisting military decisions. Packard (1990) used GAs in time-series prediction. Allen, Karjalainen (1993) used genetic programming (GP) to find new decision rules. Bauer (1994) suggested a decision selection method based on GAs in which one or more variables are defined to determine an attractive strategy, and a GA finds thresholds for these variables, above or below which a strategy is attractive.

GAs design requires only few heuristics and their input and output design is highly configurable and more intuitive. GA's discover the rules that create good solutions, and these rules are often ones that humans would rarely consider. NN input must be in a vector format, and certain input configurations may be better than others; NN outputs must be in a vector format, numbers between 0 and 1. Using heuristics, the designer must convert solutions and input data into a format that may not be either intuitive or optimal. GA's input should be defined as the parameters of a fitness function whose output is a single number. The fitness function has an intuitive interpretation describing how good a solution is.

GA it is fast, flexible, intuitive and transparent, and lends itself to the discovery of a variety of options. GA begins with a seed population of trial solutions and then evolves this population over several generations to find better and better solutions. The process is analogous to natural selection: Solutions are grouped by similarity, combined to form new possibilities, varied to allow for incremental improvement and evaluated against each other to find the best of each generation to pass to the next. This principle can be repeated a fixed number of times or until the solutions stop improving appreciably

c) Fuzzy Logic

FL architecture consists of a set of fuzzy rules that expressed the relationship between inputs and desired output. In these models inputs are fuzzified, membership functions are created, association between inputs and outputs are denned in a fuzzy rule base, and fuzzy outputs are restated as crisp values. Fuzzy rules in such a model could be provided by the decision maker (subjective fuzzy logic) or elicited from raw data (objective fuzzy logic).

Wong, Wang (1992) developed a fuzzy-neural system for decision selection. Yuize (1991) applied fuzzy logic approach to a decision support system. Ye, Gu (1994) developed a hybrid neuro-fuzzy model in which fuzzy logic enhances a neural trading system.

Fuzzy Associative Memories (FAM) proposed by Kosko (1992) is used to determine decision rules. In this method, the weight vector of a network trained by input-output data is considered as the membership function of input-output space.

Benachenhou, (1994) developed a fuzzy rule extraction tool (FRET) that extracts fuzzy rules from input-output data by FAM method, and then uses them in a fuzzy decision support system. A fuzzy rule set derived from sample data is then used as a fuzzy expert system.

Man, Bolloju (1995) implemented a prototype of a fuzzy rule based decision support system. To extract and transfer decision maker's expertise, they employed unstructured interviews with some experienced decision makers. Fuzzy rules representing the commander's decision making process are quite close to the terminology used by the experts and the rules are easily interpretable. The use of FL for knowledge representation has facilitated a high level of abstraction of the experts' knowledge. Moreover, the flexible relationship represented by membership functions and fuzzy rules, between the variables in the model have provided a robust model of the decision making process.

In maneuver planning and force allocation, FL's usefulness comes from its capability to synthesize easy to understand statements from complex data, a kind of fusion. This leads to the judgment that they ought to be closer together. In this case FL allows facts to be translated into judgments quite easily but is not suitable in telling a unit to go to a particular point or a specific coordinate. The performance of global judgments based on FL unless a GA ingredient is added

d) Expert Systems

Expert systems (ES) use a knowledge base including a set of rules and an inference mechanism that provides computer reasoning through inductive, deductive, or hybrid inductive- deductive reasoning. Knowledge base rules usually are undertaken through interview with traders. Rules in such knowledge-based systems are represented in the form of computer readable sentences. Checking for consistency and validity of rules is essential for a knowledge-based system, which is complex and difficult in the financial field, even when it is a system with only a dozen rules.

Lee, Jo (1999) developed an expert system based on candlestick analysis to determine the timing of action. In candlestick analysis there are several patterns which can imply future battlefield movements. Various such patterns were used to construct the knowledge base. Several aspects, such as recognition of patterns, formulization of pattern definition, rule generation based on the patterns, performance evaluation of the rules, should be considered, which requires much effort.

3 A COMPARATIVE ANALYSIS OF THE CANDIDATE AI TECHNIQUES

Regarding the learning capacity of various AI techniques ES, FL, NN, GA can be ordered from low to high. ES and FL as suggested by Zadeh are not capable of learning anything. NN and GA have learning capability, although on average, pure GA usually need a longer learning time (Russo, 1998), but when a priori knowledge is concerned, the order is inverted. GA need no a prior knowledge; NN need very little; FL and ES need quite detailed knowledge of the problem to be solved.

NNs are capable of learning and can therefore be used when all that is available are some significant examples of the problem to be solved, rather than a solution algorithm. NNs are capable of learning from examples, but what is learned is not easy for humans to understand. Complexity and interactions between the hidden nodes of a NN make it unattainable to understand how a decision is made. The outputs have to be trusted blindly, and this is what does not endear the NN to decision makers.

GAs are affected much less than NNs by the problem of local optima and has far less likelihood than a NN of finding a local optimum rather than a global one; this is likely to correspond to a less significant learning error. Their learning speed is generally slower and they are computationally intensive requiring much processing power.

Greedy algorithms are useful in determining paths through cost topologies. The A* algorithm combines the Best-First-Search (BFS) algorithm, a quick algorithm, and Dijkstra's route finding algorithm, which is an optimal solution-finding algorithm. A* algorithm works by starting at a node and using a heuristic to determine the best node to move to and this heuristic choice depends on many aspects. In the case of certain choices, A* is similar with Dijkstra's algorithm, testing every path. The downside of A* is that, while it is an excellent route finder, it requires that the designer choose a heuristic, which can lead to rather suboptimal paths if those paths are chosen poorly.

ES are more flexible to modification than neural or genetic based systems because rules can be adjusted over time, and when the system doesn't perform properly but it is impossible to build in the absence of experts and a priori knowledge.

In comparison with FL, more rules are needed in expert systems to cover possible outcomes.

Subjective FL's linguistic representation is very close to human reasoning. It is much less complex in terms of computational effort. Unlike in ES, overlap or ambiguity between rules can be managed in FL. It is not capable of learning and it is impossible to use when experts are not available.

The objective FL (Takagi, Sugeno, 1985), inherits all the advantages of subjective fuzzy logic, but not the less desirable features. It possesses good learning capacity and can therefore be used when all that is available are some significant examples of the problem to be solved, rather than a solution algorithm. The system generates a fuzzy knowledge base, which has a comprehensible representation. Therefore, one can easily understand how a decision is made. It is independent of experts and it has a low degree of computational complexity. The optimization of a fuzzy model requires some effort in order to arrive at the optimal mix of membership functions and the number of fuzzy rules. Lack of available tools that optimize these functions is the main bottleneck.

4 THE USE OF DSS- COA IN OPERATION PLANNING

CoA design is based on the understanding of the situation assessment, mission analysis, resources status assessment. According on the time available, the decision staff should develop different CoAs that answer to some critical questions (when, who, what, where, why and how), each of them suitable, feasible, acceptable, exclusive, complete. The analysis of these CoAs could be based on war gaming simulations even if some authors considered that war gaming could be a frustrating tool for the military since the selected CoA is never wargamed sufficiently to achieve synchronization. Based on the fact that the staff has to deal with huge volume of information in a very short time period DSS would be helpful in any step of the operation planning process.

DSS-CoA should be based on a detailed investigation of how the staff perform CoAs evaluation, analysis, selection. Since the evaluations of the CoAs according to the different criteria might include uncertainty, ambiguity, fuzziness, subjectivity, is necessary to minimize the risk component introduced during the evaluation process. A graphical and intuitive tool could balance the relative importance of the set of criteria. A stability interval analysis tool could be the answer to the increase of the awareness of the decision-maker about the role of relative importance coefficients.

The design, development, implementation of DSS- CoA is based on formal models of a CoA designed in special processes of knowledge acquisition. The event model uses operational information required by the evaluation and analysis tools and contextual information (socio- political aspects). Even if the event model would have been a lot simpler without this contextual information, this information is critical in the CoAs generation process.

DSS should be integrated to the organization workflow, and should be designed in a way to facilitate the acceptance and the transition. DSS should interact with other information, planning and decision systems.

DSS-CoAs selection support the following functions: description of the event, development/description of possible CoAs, identification of criteria to be used in the evaluation process, evaluation of the CoAs according to the selected criteria, analysis and comparison of these CoAs, and post-execution analysis that are performed sequentially or simultaneously. Decision staff is in charge of describing the events and the capability to support this function should allow the creation of new events, the upgrade of the description of an existing event, the retrieval of old events to trigger the CoAs development or the selection processes. The event description should be based on a framework that include information related to situation review, assumptions about the enemy, enemy forces and CoAs, planning guidance, other consideration aspects, theatre of operation features, and own forces capabilities. This information is essential to fully understand the problem, essential for a better assessment of the situation.

CoAs facility should include the creation of a new CoA, the update of existing ones, and the verification of CoA feasibility. In this case a model that include information related to action items that describe the actions to be performed by the resources (what, who, how, when) should be used to represent a CoA. As soon as the CoAs description is completed, the planning officer needs a communication channel to trigger the evaluation and selection processes.

The automated evaluation of each CoA is made according to each criterion. Heuristics may be used or subjective assessment may be directly provided by the users. A selection facility must allow automated CoAs comparison and the decision-maker considers different criteria when comparing CoAs. This facility should then, according to different types of situations, propose different criteria to be considered in the evaluation process, and predefined weights and thresholds accordingly. Even if the proposed criteria should be considered, the decision-maker should have the possibility to select those he considers most appropriate for the actual situation

This should be performed in an interactive way. When the criteria are selected, the CoA comparison should be done automatically, using Multiple Criteria Decision Analysis (MCDA) procedure, and different types of results must be presented to the decision-maker. A graph may represent the ranking of the CoAs. It is essential that information about the quality of each CoA should be presented since this graph only indicates only the rank of CoAs. Among the analyses that can be provided to help a decision-maker, there is a dominance check which verifies if a CoA is better than all other CoAs on all the criteria, no matter the value assigned to the different thresholds. A weight stability analysis offer to the decision-maker information on the sensitivity of the criteria when weights changes. A what-if analysis on the model parameters or on the CoAs evaluations allows the decision-maker to foresee the effects of the actual settings on the prioritization of the CoAs. This enables the user to either select any CoA while providing justifications, or ask for more satisfactory CoAs and information.

A post-analysis facility should allow the reconsideration of the relevance of the choice made while the event is completed. Once a CoA has been selected and executed, the commander could then re-evaluate if its decision was the best one or not, and why. This precious knowledge should be archived for reference to future operations. This knowledge will be used to learn from experience.

Finally, the functional facilities must allow the management of the criteria, and the default parameters used within the different decision analysis procedures. This facility must support an analyst in creating new criteria, updating existing ones and associating criteria with generic instances of events. Also, this facility should enable him to set default values for different parameters.

Since the processes of defining events and CoAs, evaluating and comparing CoAs, and selecting the most appropriate one are realized through a team effort, it is important to be able to assign different facilities to different people by defining user's profiles (event editor, responsible to describe an event; CoA editor, responsible to define and describe appropriate CoAs for a specific event; commander to select the most appropriate CoAs; analyst for managing the criteria and to set the parameters according to the preferences of the decision-maker; system administrator, responsible to define who can have access to the system to do what.

DSS-CoA must have a facility to manage the user's profile, and maintain the databases on event, CoA and criteria and this ingredient could be used by a system administrator to create new users, assign privileges, and update user's profile.

5 SOME ASPECTS REGARDING AI BASED PRODUCTS FOR MILITARY APPLICATIONS

COA Creator (Qualitative Reasoning Group) is a tool that allows a user to sketch a CoA into the computer (Ferguson, 2000). The semantic knowledge based representations stored into the computer for each item added to the CoA sketch. The *CoA statement tool* (Alphatech Inc.) is an upgrade that allows the staff planner to enter the COA statement and it equiped with an interface that allowed users to produce natural language sentences to build an intelligent CoA statement.

Once the digital representation of the sketch and statement information was fused and translated, Carnegie Group Inc. proposed *CADET*, an instrument capable to transforms the sketch and statement into a detailed plan/schedule of the operation. *CADET* expands critical tasks, determines the necessary supporting relations, allocates/ schedules tasks to friendly assets, takes into account dependencies between tasks and availability of assets, predicts enemy actions, devises friendly counter-actions, estimates paths of movements, timing requirements, logistics consumption, attrition and risk (Kott, Ground, Budd, 2002). The resulting output can display as a synchronization matrix or as an animated map. Although the resulting plan still requires the experience of a careful review done by the decision maker, it represents an important starting point for further analysis and an efficient time saver. Once the CoA is digitized, CADET automatically performs the detailed planning, including the traditionally time-consuming tasks for the plan (time-distance analysis, logistics elements, and potential attrition estimators).

The input comes from the decision maker's intent, concept of operation and the target end-state for the operation. The information is used to develop CoA sketches that comprise a set of high-level actions, goals, and sequencing, referring to the movements of the friendly forces. Based on this input, the members of the planning staff analyze in detail the elements of the friendly CoAs. The *war-gaming* process involves planning of the detailed tasks required to accomplish the specified CoA, allocation of tasks to the diverse forces, assignment of suitable locations, estimates of battle losses (attrition), predictions of enemy actions.

The output is presented in a synchronization matrix format, a generalized Gantt chart in which time periods constitute the columns and the functional classes of actions are the rows (maneuver, combat service support, intelligence). The content of the plan, is expressed in the matrix that includes the tasks and actions of the multiple subunits and assets

of the forces, their objectives and manner of execution, expected timing, dependencies, routes and locations, availability of supplies, attrition, enemy situation.

6 CONCLUSIONS

The AI ingredient in military decision making offers a strong support capable to create natural sketch-based interfaces that domain experts can use with low training. The users expressed the desire for a single integrated framework that captures CoA sketches and statements simultaneously and capable to provide a unified map-based interface to do both tasks. The interest is to design a framework capable to express CoA sketches equipped with visual understanding. It is also possible to use the rich representations in the CoA Creator as inputs to other instruments (pattern completion and access to previous plans via analogy).

DSS-CoA offers the inspiration to define a set of facilities appropriate for any DSS developed for the evaluation and selection of CoAs: event management facility; CoA management facility; CoA evaluation facility; CoA comparison, analysis and selection facility; Post-analysis facility; criteria management facility; system administration facility. DSS- CoA users should be aware about the limitations and the level of trust and an explanation facility providing result explanations adapted to the user (background, experience, knowledge, preference) and the context (time available) are important in military decision training.

References

[1] ABI-ZEID, I., BELANGER, M., GUITOUNI, A., MARTEL, J. M.: *The criteria for an advisor tool in Canadian airspace violations situations,* Monterey, California, 1998.

[2] ATKIN, M. S., WESTBROOK, D. L., COHEN, P. R.: *Capture the Flag: Military simulation meets computer games*, AAAI Spring Symposium Series on AI and Computer Games, 1999.

[3] BANNER, G .T.: Decision Making -A Better Way, *Military Review*, Headquarters, Department of the Army, vol. LXXVII, no.5, 1997.

[4] BOHMAN, W. E.: June, STAFFSIM, *An Interactive Simulation for Rapid, Real Time Course of Action Analysis by U.S. Army Brigade Staffs*, Monterey, CA, 1999.

[5] COHEN, P. R., JOHNSTON, M., McGEE, D., OVIATT, S., PITTMAN, J., SMITH, I., CHEN, L., CLOW, J.: *QuickSet: Multimodal interaction for distributed applications*, The Fifth Annual International Multimodal Conference, ACM Press, 1997.

[6] FERGUSON, R.W., RASCH, R. A., TURMEL, W., FORBUS, K. D.: *Qualitative Spatial Interpretation of Course-of-Action Diagrams,* The 14th International Workshop on Qualitative Reasoning, Morelia, 2000.

[7] FORBUS, K., FERGUSON, R., USHER, J., *Towards a computational model of sketching*, Santa Fe, 2001.

[8] FORBUS, K.: Exploring analogy in the large. In Gentner, D., Holyoak, K., Kokinov, B.: Eds: *Analogy: Perspectives from Cognitive Science*, MIT Press, 2001.

[9] FORBUS, K., USHER, J., *Sketching for knowledge capture: A progress report*, San Francisco, California, 2002.

[10] GUITOUNI, A., MARTEL, J. M., BELANGER, M.: *A multiple criteria aggregation procedure for the evaluation of courses of action in the context of the Canadian Airspace Protection,* DREV TR 1999-215.

[11] HAYES, C. C., SCHLABACH, J. L.: *FOX-GA: A Planning Support Tool for assisting Military Planners in a Dynamic and Uncertain Environment*, AAAI Press, 2002.

[12] HUNG, C. L., YONG, K., *Multiple Attribute Decision methods and Applications, A State-of-the-Art Survey,* Springer, 1991.

[13] Innocenti III, Charles W., *Abbreviated Military Decision Making for Brigade Combat Teams*, Fort Leavenworth, KS, 2001.

[14] KEWLEY, R., EMBRECHT, M.: A Computationally Intelligent Military Tactical Planning System, *IEEE Transactions on Systems, Man and Cybernetics*, 2003.

[15] KLEI, G. A.: *Sources of Power*, Cambridge, Massachusetts, MIT Press, 1999.

[16] KOTT, A., GROUND, L., BUDD, R.: 2002. *Toward Practical Knowledge-Based Tools for Battle Planning and Scheduling*, 2003.

[17] MARR, J.: *The Military Decision Making Process: Making Better Decisions Versus Making Decision Better*, Fort Leavenworth, KS, 2001.

[18] PAPARONE, C. R.: US Army Decision making: Past, Present and Future, *Military Review*, Headquarters, Department of the Army, vol.LXXXI, no. 4, 2001.

[19] PRELIPCEAN, G**.,** MOISESCU, F., BOSCOIANU, M., CIUFUDEAN, C., CIOBAN, G.: *A Hybrid Trading System for Defense Procurement Application*, Simulation, Modelling and Optimization (SMO '09), Budapest, Hungary, p. 173–178, WSEAS Press, 2009.

[20]PRELIPCEAN, G., BOŞCOIANU M., POPOVICIU, N.: The role of predictibility of financial series in emerging market application, *WSEAS TRANSACTIONS on MATEMATICS*, Issue 7, volume 7, July 2008, pp. 472-481, 2008.
[21]PRELIPCEAN, G., BOŞCOIANU M.: *Computational framework for assessing decisions in energy investments based on a mix between real options analysis and artificial neural networks*, in Proceedings of the 9th WSEAS International Conference on Mathematics & Computers in Business and Economics, pp. 179-184, 2008.
[22]BOŞCOIANU, M., PRELIPCEAN, G.: *An Analysis of the Innovative Financial Support of Space- Defence Investments*, in Recent Advances in Space Technologies, RAST'07. 3rd International Conference on, 14-16 June, Istanbul, Turcia, pp. 74 - 77, 2007.
[23]TATE, A., LEVINE, J., JARVIS, P., DALTON, J.: Using AI Planning Technology for Army Small Unit Operations, T*he Fifth International Conference on Artificial Intelligence Planning Systems,* 2001.
[24]WASS de CZEGE, H., BIEVER, J. D.: Six Compelling Ideas On the Road to a Future Army, *Army Magazine,* Vol.51, 2001.

Col. Prof. Dr. Florin MOISESCU
„Henri Coanda" Air Force Academy in Brasov
81-83 George Cosbuc
500 183 Brasov
Romania
E-mail: secretariat@afahc.ro

Prof. Dipl. Eng. Mircea BOŞCOIANU, PhD.
„Henri Coanda" Air Force Academy in Brasov
81-83 George Cosbuc
500 183 Brasov
Romania
E-mail: mircea_boscoianu@yahoo.co.uk

Univ. Professor Gabriela PRELIPCEAN, PhD.
"Stefan cel Mare" University of Suceava
13 Universitatii Street
720229 Suceava
Romania
E-mail: gprelipcean@yahoo.com

Univ. Lecturer Mariana LUPAN, PhD.
"Stefan cel Mare" University of Suceava
13 Universitatii Street
720229 Suceava
Romania
E-mail: mariana_lupan@yahoo.com

IRANIAN WAR ON INSECURITY: ROADMAP FORWARD?

Pavel NEČAS, Miroslav KELEMEN, Jaroslav UŠIAK, Peter SPILÝ

Abstract: At present, the attention of the world is turned towards the Middle East. The Middle East is an area of global strategic importance and of considerable economic and political interest. Especially Iran is today a potential source of threat to the whole world. After numerous verbal attacks against Israel, Iran has shown strong determination to continue its uranium enrichment program, despite pressure from the United States, the EU and Russia. Iran has therefore been subject to several series of economic sanctions in order to leverage prevention of the proliferation of weapons of mass destruction, countering international terrorism and promotion of democratic government, all named major goals of the international community. In order to pursue the world developed countries' foreign policy goals, these have been increasingly using the tool of economic sanctions, as in the cases of Iraq, Libya and Syria or the still remaining case of North Korea. The international community does not look for but insists on the sanctions as such for the vein only. But in the absence of a negotiated solution, the process of imposed international sanctions will extend by damaging the Iranian economy for a long time and preventing Iran from enjoying all the benefits of international cooperation.

Keywords: Instruments of Power, International Pressure, Rules of Engagement, Business Conditions, Sanctions and Restrictions, National Caveats.

1 INTRODUCTION

Great European companies are struggling to cope with the considerable pressure applied by their national governments, but also by the United States as such. Although they do not officially admit the political interest as the reason to withdraw from Iran or to freeze their investment, like for example the corporation Total, the national caveats certainly count a lot. For example, President Sarkozy believed that France was asking its own companies "to refrain from going to Iran." France has already urged its oil giant Total and its gas firm Gaz de France not to bid for new projects in Iran and urged French banks to stop doing business there. Similarly, Angela Merkel said German companies could do much more in reducing their commercial ties to Iran. "Germany needs to look somewhat closer at the existing business ties with Iran."

2 CORPORATIONS: READY TO INVEST

But the strong "must" also comes from the US, who wants the European Union to push forward with further sanctions against Iran, and urges Iran's major trading partners to cut back business with Tehran to send a strong message. Moreover, the US sanction laws are quite strict and enable the administration to fine non-American companies investing heavily in Iran. Informally, companies are afraid of loosing lucrative business in the US in case of being involved in Iran.

Apparently, the sanctions have now an effect on many aspects of Iran's economy and make everything more complicated, particularly in financial sphere. Private non state actors proved to be very sensitive to sanctions, as they need stabile environment to make investments. Sanctions against

Iranian banks and the withdrawal of major Western banks make the transactions more difficult and therefore more expensive, which is in the end reflected in the costs of production. Moreover, all the supply chain is impeded, as suppliers also have to fight against higher costs and less support by their western partners. Suppliers of big European companies in Iran such as Renault or PSA therefore cannot meet their deadlines which leads to considerable production delays.

Generally, the preconditions for business seem to deteriorate. Iranian government is losing patience and exerts much more pressure on the grand Western companies, especially involved in the energy sector, which is vital for Iran. If the companies do not accelerate and utilize effectively their investments, the Iranian government is ready to offer the business to Asian or Russian companies at once.

Finally, the strict regulations and state-dominance do not encourage companies to invest in Iran. An improvement in this field, such as privatizations, incentives to invest, tax reductions and others would attract companies, or at least modify their plus or minus balance, while considering predictive positive and active involvement in Iran.

3 ALTERNATIVE SOLUTIONS

Although during the last few years conditions for the US entrepreneurs and European companies have worsened and became more complicated to achieve and the trade between EU and Iran has decreased, most of the big companies, especially in the energy sector want to keep positions in the Iranian business spectrum. So far, no Western oil company has definitively said NO to future investments in the country. In reality, many of them have frozen new investments or even withdrawn from several big projects (Total in South Pars, Royal Dutch Shell, Repsol YPF, Eni, StatoilHydro), but they cannot

afford leaving and abandoning their expensive investments performed during the last twenty years. Due to its exceptional natural richness, Iran is strategically important for most of the companies to keep their relationship with the country's national oil corporation. The corporations still continue their current activities such as oil buyback projects, gas development in existing fields and lubricants supply. For example, Shell presence in Iran is not growing spectacularly, but is not shrinking either; at the moments Shell top management is hesitative and cautious of taking any decisions on further investment in Iran.

The big players - oil giants look for delayed investments and may even trade their current phases in the big projects such as South Pars for a later one, hoping that in short or middle terms, when the ulterior phases will enter into force, the sanctions will not be effective any more. Once the geopolitical situation is stabilized, the conditions to continue investments and development in the country will resume.

It is definitely very important for companies to keep their presence in Iran. With its 70 million inhabitants, its geographical position and its exceptional natural resources, Iran's strategic importance for future remains. Consequently, the companies face many difficulties while doing business or investing in Iran, few of them withdraw completely from Iran and they are definitely aware of the great potential Iran might offer once the embargo is lifted.

4 EFFECTIVE APPROACH AHEAD

President Mahmud Ahmadinejad comments on economic instruments of power imposed: "This weapon is not even effective against small and deprived countries, let alone the great people of Iran." Although Iranian authorities continue to criticize the efficiency and the „raison d'être" of the sanctions, as a matter of fact especially on the financial level, the sanctions subvert the financial support companies need for the overall effective functioning.

Firstly, the UN, the US and EU sanctions target the three biggest Iranian banks, which cover 80 % of the Iranian international transactions - Bank Melli, Bank Saderat and Bank Sepah and sanctions are being considered also against the Iranian National Bank Markazi. Other banks are „listed" by the US Treasury Department's Financial Crimes Enforcement Network and international organizations such as Financial Action Task Force, which is a strong signal to Western entities to be cautious in the transactions with Iranian institutions. These actions have placed increased burdens on theInternational banking community in dealing with these targeted Iranian entities. The three sanctioned banks are in fact cut off from the international system and their assets are frozen and their international operations severely restricted. It is thus practically impossible to manage any transactions with these banks, which causes inevitable problems for the casual business with Iranian partners.

Secondly, after several urgent calls and measures by American and European authorities, many major Western banks have withdrawn from Iran or have at least frozen their activities there (they do not open new accounts and keep minimal presence only), being concerned about their reputation and commercial interests in the big markets, such as the US market. Britain, France, Germany, Japan and Italy have been slowly reducing banking exposure and government credit guarantees for exports to Iran. The companies' commodities export is thus not comprehensively covered. Big banks such as HSBC (UK), UBS (Switzerland), Barclays (UK), Société Générale (France), ABN (Netherlands), Standard Chartered (UK), Deutsche Bank (Germany), BNP Paribas (France) or Commerzbank (Germany) curbed business with Iran and stopped offering letters of credit, a standard payment guarantee in international trade. By November 2008, more than 80 banks have suspended business with Iran. Today it is nearly impossible to secure project financing by the Western banks. Without the banks' support, which is very important, corporations have enormous difficulties to conduct and manage their everyday business activities.

The imposed financial restrictions are certainly not easing the system of wireless and credit payments and operations. Transactions in US dollars have also become more difficult. Companies are obliged to pay in cash or to use alternative channels of money transfer. However, that is much more expensive. Because of government's extensive spending, cash is strictly controlled in Iran, and it may be quite difficult to raise important sums of money. Moreover, the transactions in cash generate substantial fees, such as commissions for the middle-men or cost of travel while handling cash. Informal money transfers are called "Hawala."

These brokers are largely unregulated merchants who arrange with counterparts in other countries to deliver money across borders quickly, usually for a fee. On one hand, such transactions inflate the service requirement of the companies, increase the costs and consequently slow down, damage and curtail business, but on the other hand they make the informal money channels developed and grown. It is becoming more difficult for regulators in the region to track suspicious transactions, as these are opaque and outside of the regulated global banking system. As nobody really knows where money flows, it is

almost impossible to trace money laundering, drug smuggling and terror financing in the region.

All these banking restrictions, lack of US dollars in transactions and the reluctance of Western banks to support business in Iran definitely do annoy companies and increase their costs.

5 ILLEGAL COUNTERMEASURES

Several big European companies, such as banks or gas giants, have frozen their activities in Iran; others try to circumvent the measures. Due to the increasing US pressure on Iran‘s financial system, most international banks have stopped dollar transactions with Iran. Transactions, especially in the oil business, are conducted increasingly in other currencies, such as Euro or yen. Moreover, foreign currency holdings of Iranian central bank are being switched from dollars to Euros. This moving from dollar transactions to other currencies seems to work quite well, but the cost, of course, is not negligible.

Another common technique how to circumvent the sanctions is to export through a third country in Asia or in the Persian Gulf. Many companies are now searching and identifying in the United Arab Emirates and other Persian Gulf commercial hubs for business partners to help them skirt the sanctions. The banking restrictions have forced Iranian companies that previously exported to and imported directly from European and Asian suppliers, to open affiliates in the UAE or to work through front companies and middle-men; on the contrary, Western companies do the same in the opposite direction. Most of the UAE exports to Iran originate in the third countries regionalized in Asia, United States and Europe. UAE and Iran have very strong historical and economic ties and their population is strongly mixed and the crucial transactions between the both nations have not been strongly regulated and disaffected so far. One may learn that UAE abide by UN sanctions, but it is not the case for US and EU sanctions. Gradually, even this channel is becoming difficult to explore, because UAE is under strong pressure from the US and has recently improved its customs and transshipment laws.

In any case, transferring the re-export and altering the currency strongly increase the final costs for the company, just like the other measures to circumvent the sanctions. In addition, such illegal procedures might be risky and certainly punished by the national or international authorities.

Western corporations, especially oil and industrial giants, have been under considerable political pressure as for their investments and activities in Iran. Many of them are following the decisions of international community and although their sector is not specially targeted by the sanctions, they freeze their investments in Iran. This loyalty to their national governments or to their business partners (such as the US) is respectable, but does it really compel Iran to abandon its nuclear program?

Large companies like Royal Dutch Shell or Total which have adopted this difficult decision, have been clearly threatened by the Iranian authorities by handling over the business to other companies, namely from Russia or Asia, if they do not proceed their planned investments. This shift towards East has already started and Russian and Asian companies slowly fill in the gap left by the Western companies. These actors have extraordinary capital and are happy to invest in Iran, as evidenced by their massive investments during the last years. Russian Gazprom is signing multi-billion dollar projects in order to become a world-wide gas giant. Chinese Sinopec and CNPC see great opportunities to respond to Chinese growing demand on energy in Iran’s natural resources. Besides, India’s ONGC and Malaysia’s SKS Ventures, as well as Persian Gulf’s companies, want to get their share of the pie too.

In the banking sector, second and third-tier banks in places such as Pakistan, but also in Europe (Malta, Turkey) are competing to provide financing to Iranians frozen out of the Western banking system. It would be naive to think that European companies are irreplaceable in Iran, especially in today’s globalised and diverse world. On the short term, Iran will certainly have to shelve its ambitious liquefied natural gas (LNG) plans due to lack of access to the Western technology required to develop the necessary infrastructure.

The Western companies do have some technological advance and are historically established in Iran, but these advantages are rapidly erased and Russian and Asian companies will easily entrench in the region’s business environment in the middle and long term. The decline of European companies’ presence in a country which has the second largest gas reserves and the third largest oil reserves in the world might have serious effects for them in the long term. Although their present investments in Iran are frozen, most of them keep some activities in Iran and they wish to fully develop their investments once the sanctions episode is over. In fact, the situation might improve, as in the case of Libya, and companies must be ready to react immediately. If they do not want to completely lose their previous investments and their market share in Iran, it is very important to keep good relations with the NIOC and not withdraw completely.

In order to continue doing business with Iran, European companies are obliged to find new ways to adjust their transactions to cope with the sanctions. Firstly, they switch the currency of the transactions from dollar to Euro or yen. Secondly, some of them try to circumvent the sanctions by exporting their goods through a third country. If the European

companies do not continue developing their business in Iran, their Russian or Asian competitors rapidly try to replace them and benefit from the opportunities of Iranian rich natural resources.

The US, EU and UN economic sanctions against Iran are especially aimed at the banks, companies and individuals alleged of nuclear proliferation or of cooperation with nuclear proliferators. The assets of these entities have been frozen and their representatives are banned from travelling. On the contrary, the use of economic sanctions and measures, negatively influence an open world economy and undermine the very foundations of free trade. The companies have to face these political directives and national caveats, which can often be very harmful. The issue of economic sanctions in such a troubled and rough country is inevitably worrying the foreign companies interested or already present in the region.

6 CONCLUSIONS

One may acknowledge that the newspapers information on President Ahmadinejad, Iranian nuclear proliferation, sensitive relations with Israel and on specific information about particular corporations and companies is incomplete. In fact, the issue is often presented from the political point of view and rarely from the economic one. And to prove the fact, the only elaboration is being done on the issue of oil. Few companies express their business strategy in Iran in the media, they rarely talk about their concrete problems and they never publish official press release or annual report describing their activities in Iran. Firms are very cautious about their communication regarding Iran; they do not want to question their relations with Iranian authorities or with Western governments.

Finally, the episode of economic sanctions against Iran is a current issue. The situation on the politic and economic stage is changing rapidly and new circumstances and facts occur almost every day. The US President Barack Obama promised a metamorphosis in US-Iranian relations: more direct talks, persuasive diplomatic incentives, including a credible offer of improved relations and security guarantees.

References

[1] ADAMČÍK, F., KELEMEN, M.: Entity condition change of an asymmetrical opponent. In: *Obrana a strategie*. Roč. 6, č. 1 (2006), s. 31-43, ISSN 1214-6463.

[2] ALI, A. J., CAMP, R. C.: Economic Sanctions: Obstruction or Instrument for World Trade? In: *Managerial Finance***,** Volume 25 Number 3/4 1999, pp. 66-75.

[3] BAINS, E.: Sanctions benefit local firms. In: *MEED: Middle East Economic Digest,* Vol. 52, Issue 40, March 10, 2008.

[4] BUČKA, P.: *Significant changes in NATO´s Command Structure.* 5th International Conference New Challenges in the Field of Military Science 2007, Hungary : Budapest 13. – 14. 11. 2007, p. 8.

[5] BUJŇÁK, J., FAJMON, P., LÖRINCZOVÁ, M.: Asymmetric Warfare: The Slovak Perspective. In: *Vojenské reflexie*, 6/2009, roč. IV., č. 1/ 2009, pp. 50-66, ISSN 1336-9202.

[6] DENNIS, R.: Talk Tough With Tehran. In: *Newsweek*, Vol. 152, Issue 23, August 12, 2008.

[7] HADFIELD, W.: State sell-off boosts Tehran bourse. In: *MEED: Middle East Economic Digest*, Vol. 52, Issue 35, August 29, 2008.

[8] LAWRENCE, J.: European carmakers hit by UN's Iran sanctions. In: *Automotive News Europe*, Vol. 13, Issue 8, April 14, 2008.

[9] McDOWALL, A.: The price of sanctions. In: *MEED: Middle East Economic Digest;* Vol. 51 Issue 15, April 13, 2007, p. 4-5.

[10] NEČASOVÁ, V.: *La prosperite des entreprises europeennes en Iran: sanctions ou libre-echange?* Reims Management School, Mai 2009.

[11] RECKNAGEL, Ch.: Europe Gets Increasingly On Board with Iran Sanctions. In: *Radio Free Europe*, July 11, 2008, available on: http://www.rferl.org/content/Europe_Gets_ Increasingly_On_Board_With_Iran_Sanctions/1183057.html.

[12] SEDLAK, V., LOSONCI, P., KISS, I.: *Bezpečnostné informačné technológie,* VŠBM Košice (vyd.), Košice, 2008, 133s., ISBN: 978-80-89282-26-5.

[13] SCHOTT, J. J.: Economic Sanctions, Oil, and Iran. In: *Testimony before the Joint Economic Committee, United States Congress Hearing on Energy and the Iranian Economy*, July 25, 2006, available on: ww.iie.com/publications/papers/paper.cfm?ResearchID=649.

[14] TEREM, P.: *Globalisation and its Environmental Dimension*. Siwek,T.-Baar,V.(ed.): Globalisation and its Geopolitical, Cultural, Economic and Ecological Context. 8.-10. 10. 2003 University of Ostrava, 2004, str. 304-314, AEC. ISBN 80-7368-022-X.

[15] VYROSTEK, J., KELEMEN, M.: Synergia občianskej, národnej, medzinárodnej bezpečnosti a nové výzvy vojenskej vedy. In: *Bezpečné Slovensko a Európska únia*. Zborník príspevkov z 2. medzinárodnej vedeckej konferencie, 13. – 14. november 2008, Košice : Vysoká škola bezpečnostného manažérstva, s. 397-399, ISBN 978-80-89282-28-9.

[16] WEISMAN, S. R.: Big Europe banks curb Iran deals. In: *International Herald Tribune*, May 22, 2006.

[17] ZAKARIA, T.: Iran Biggest Mideast Challenge for Obama - W. House. In: *Reuters*, January 7, 2009, available on: www.reuters.com/article/vcCandidateFeed2idUSN07493706.

Col. (GS, ret.) Assoc. Prof. Dipl. Eng. Pavel NEČAS, PhD.
Armed Forces Academy of General M. R. Štefánik
Demänová 393
031 06 Liptovský Mikuláš
Slovak Republic
E-mail: pavel.necas@aos.sk

Brig. Gen. Assoc. Prof. Dipl. Eng. Miroslav KELEMEN, PhD.
Armed Forces Academy of General M. R. Štefánik
Demänová 393
031 06 Liptovský Mikuláš
Slovak Republic
E-mail: miroslav.kelemen@aos.sk

Dr. Jaroslav UŠIAK, PhD.
Faculty of Political Sciences and International Affairs
of the Mathias Bel University in Banská Bystrica
Kuzmányho 1
974 01 Banská Bystrica
Slovak Republic
E-mail: jaroslav.ušiak@umb.sk

Assoc. Prof. Dipl. Eng. Peter SPILÝ, PhD.
Armed Forces Academy of General M. R. Štefánik
Demänová 393
031 06 Liptovský Mikuláš
Slovak Republic
E-mail: peter.spily@aos.sk

KERNEL DEPENDENCIES IN A MODERN GENERAL-PURPOSE GPU ARCHITECTURE

Miloš OČKAY, Marcel HARAKAĽ, Miroslav LÍŠKA

Abstract: CUDA (Compute Unified Device Architecture) allows programmers to access massive power of graphics processing unit (GPU) and offload several compute intensive portions to the GPU. Massive parallel coprocessor is now available for central processing unit (CPU) to make computation process faster and accurate, even across the huge datasets (vector, matrix). In this paper we highlighted main aspects of GPU parallel program (kernel) sizing to achieve high performance results and we described dependencies of the main parts of GPU computational process. Kernel sizing theory is compared to the results of our benchmark.

Keywords: Compute Unified Device Architecture, CUDA, Graphics Processing Unit, GPU, Parallel computing, Kernel, Application Programming Interface, API, High Performance Computing, HPC.

1 INTRODUCTION

Several questions arise from decision regarding to use GPU chip for general purpose computing. The first one is "What do we need, if we want to bust computation power with GPU?". There are several products available from the different vendors, offering GPU power for general purpose computing. Because there is still a lack of standardization across the vendors, we will focus our interest to Nvidia's solution. Basic hardware requirements must be met. We need host system with CUDA compatible graphics accelerator [1]. A driver, a software development kit (SDK) and an application interface (API) installation have to be done properly. Software installation is straight forward and all needed multiple platform software can be downloaded from the Nvidia's homepage for free [2]. To overcome possible hardware configuration problems, it is reasonable to buy complete host with accelerator solution from authorized vendors. Otherwise, it is necessary to study hardware specification of hardware components, which you are planning to include in your system. You should be careful with this, it is not always easy to get statements or recommendations from vendors support site for complex hardware problems.

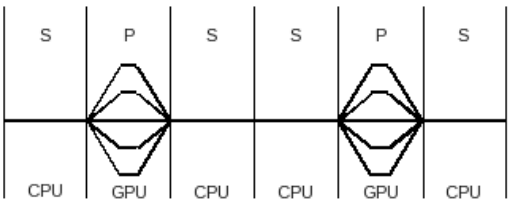


Figure 1 Portioning of an algorithm
S – sequential, P – parallel portion

Hardware and software installation is just the first step to include GPU in your applications. After the installation, you have a powerful tool for parallel programming, but a successful use of this tool is hidden behind the profound understanding of its potential. It is quite easy to create parallel program for GPU, but you need to consider far more aspects to achieve really high performance results. Your application algorithm partitioning must be done accordingly. Not every part of the algorithm can be paralleled and computed on a parallel system, but if you choose your parallel partitions wisely, you can experience massive speed-up (Figure 1). Parallel portions of an algorithm must operate on the dataset large enough to achieve proper saturation of computational resources. Data portioning has to respect hardware architecture to achieve the best possible results. In this paper we present the description of GPU parallel program (kernel) and we describe aspects which must be considered in a process of kernel definition and invocation. There are many aspects we should consider to achieve the best possible speed-up. Proper kernel configuration is just one of them, dealing with high performance results.

2 KERNEL

In CUDA programming model, parallel parts of the program are termed as kernels. Software kernel is executed by the host on the device (graphics adapter with GPU). Strictly speaking, kernel is defined as a C programming language function and invoked by the special C directive, which includes configuration for the kernel launch as a set of parameters (Code 1).

2.1 GPU architecture

If we want to understand kernel configuration dependencies, we need to have a look at the GPU core architecture briefly. Nvidia GPU consists of multiple stream processors (SP) and these stream processors are grouped for a purpose of sharing some other functional units and memory segments (Figure 2). These groups are called a multi-stream processor (MSP).

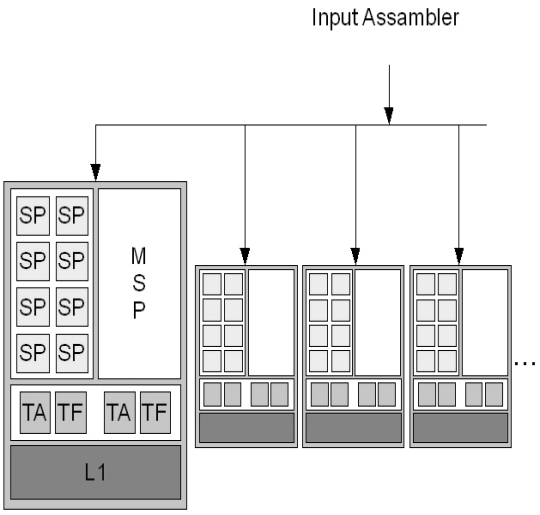


Figure 2 Nvidia GPU core, unified architecture

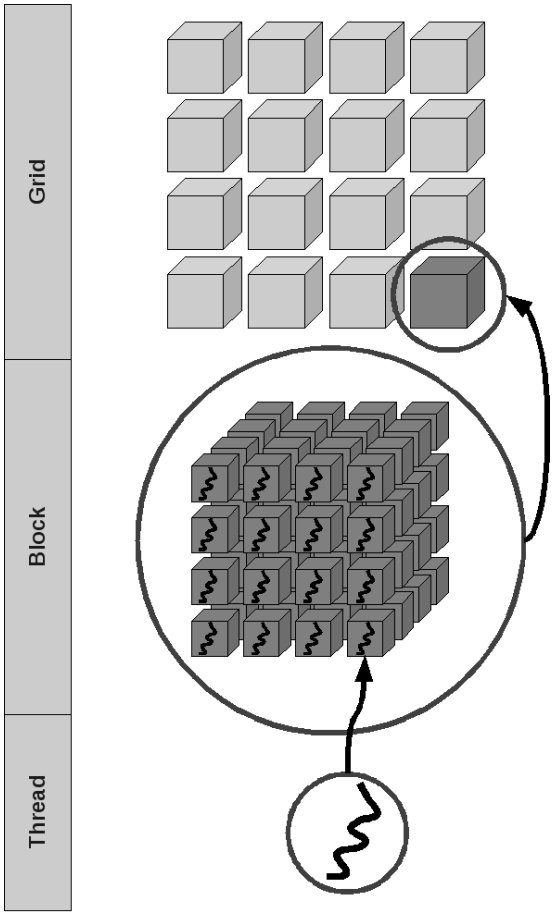


Figure 3 Kernel configuration

There are many different versions of GPUs, but all of them share this common configuration idea. More powerful graphics adapters have more multi-stream processors included in their GPU. This solution is scalable and the point of diminishing returns can be pushed up by the optimization of other parts of graphical adapter's architecture. Scalability is also based on kernel configuration flexibility.

2.2 Kernel configuration

Kernel configuration a is complex multi-dimensional structure. This structure reflects hardware architecture of the GPU. Thread is a fundamental element of the parallel computation. In a parallel programming, main task is divided into the subtasks and we refer to them as the threads. Threads are grouped into the block in purpose of intercommunication and memory resources sharing. Block can be one, two or three dimensional structure. Block has its limitations and to overcome them it is possible to group blocks into a grid. The grid is a one or two dimensional structure (Figure 3). Because GPU is a single instruction multiple data (SIMD) hardware architecture, each thread in the grid will compute the same kernel function on different part of dataset. Its main advantage is its simultaneity. The thread count which will operate simultaneously is limited by the hardware architecture.

2.3 Kernel definition and invocation

Two parts are required for kernel integration to the C code. The first is a definition of the kernel function and the second is the kernel invocation. Kernel definition looks like ordinary C function with a specific qualifier (__global__) at the beginning which declares function as being a kernel. Kernel executes on device (graphics adapter) and the kernel is callable from the host. The best way is to keep kernel lightweight. You should avoid branching if possible or use it according to the branching limitations [5].

```
//--------------------------------------------------------------------------
// Kernel definition
__global__ void MatAdd(float* A, float* B, float* C)
{
      int i = threadIdx.x;
      C[i] = A[i] + B[i];
}

int main()
{
// Kernel dimensions configuration
dim3 dimBlock(16, 16);
dim3 dimGrid((N + dimBlock.x – 1) / dimBlock.x,
             (N + dimBlock.y – 1) / dimBlock.y);
// Kernel invocation
MatAdd<<<dimGrid, dimBlock>>>(A, B, C);
}
//--------------------------------------------------------------------------
```

Code 1 Kernel definition and invocation example

According to the kernel definition, each thread is given a unique ID which is generated by using the built-in threadIdx variable. ThreadIdx is a 3-component vector, so the threads can be identified using a one, two or three dimensional thread index. As a result, these threads form a one, two-dimensional or three dimensional thread block [3]. Thread ID for the one dimensional block is the same as thread index. For two dimensional block of the size (Dx, Dy) can be applied the following scheme. ID(x,y) =(x + y Dx). For three dimensional block of the size (Dx, Dy, Dz) we can use ID(x,y,z)=(x + y Dx + z Dx Dy). Of course, any other mapping model which generates unique IDs can be applied. Threads form single or multidimensional block. Threads within the block share the low latency memory resources and they can cooperate by sharing data through this memory. The number of threads per block is limited by memory resources of GPU core. Currently it is possible to have 512 threads in one block. Of course, this is not a limitation of total amount of threads on the device. If the count of threads exceeds this limit, we can form equally sized blocks to overcome the block limitation. The multiple blocks are organized to one or two dimensional grid (Figure 3). Each block within the grid can be identified through the built-in blockIdx variable. The same rules apply to a block ID generation and the thread ID generation, but the grid is only of one or two dimensional structure. Current kernel size limitation is 65535 blocks in each dimension. All indexing and limitations related facts have to be used in a process of kernel dimensions configuration [4].

Dim3 is an integer vector type based on unit3 that is used to specify dimension variables. dimBlock specifies how many threads will be in the block and if they are built into one, two or three dimensional block. dimGrid specifies how many blocks are within the grid. It is usually calculated using dataset dimensions and block dimensions (Code 1.). There is no universal answer to the question "What is the best block and grid dimension configuration?". Kernel dimensions must suit processed dataset and kernel needs to keep hardware occupied in the best possible way, with respect to all configuration limits.

To find the balance between these factors needs a little experimentation and practice. We will mention some recommendations in the following section.

2.3 Kernel execution model

Execution model may help you better understand the kernel configuration context. Execution model is constructed as follows. The host invokes the kernel with initial configuration *(kernelName<<<dimensionsOfGrid,dimensionsOfBlock>>>(parameters);).* Kernel is executed on grid. The blocks of the grid are enumerated and distributed to multi-stream processors. Threads of thread block are executed on one multi-stream processor concurrently in warps of 32 threads (Figure 4). As thread block is processed, new block is launched on idled multi-stream processor. There should be at least as many blocks as there are multi-stream processors in the device. Running only one block per multi-stream processor will force the multiprocessor to idle during thread synchronization and also during device memory reads, if there are not enough threads per block to cover the load latency. It is therefore usually better to allow for two or more blocks to be active on each multi-stream processor to allow overlap between blocks. For this to happen, not only should there be at least twice as many blocks as there are multi-stream processors in the device but also the amount of registers and shared memory required per block must be low enough. The number of blocks per grid should be at least 100, if one wants it to scale to future devices and 1000 blocks will scale across several generations of these devices [3].

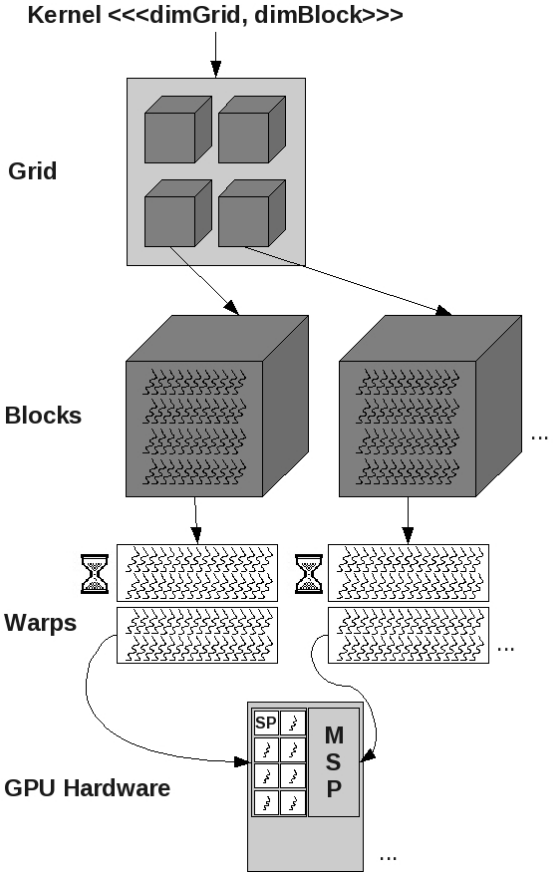


Figure 4 Kernel execution model

Table 1 Kernel configuration test

	1	2	3	4	5	6	7	8
Blocks / 1D Grid	60000	30000	15000	7500	3750	3840	3200	66207
Threads / 1D Block	32	64	128	256	512	500	600	29
Threads	1920000	1920000	1920000	1920000	1920000	1920000	1920000	1920000
Blocks	60000	30000	15000	7500	3750	3840	3200	66207
Warps	1875	938	469	235	118	120	100	2069
Blocks / MPS	3750	1875	938	469	235	240	200	4138
Warps / MPS	118	59	30	15	8	8	7	130
Runtime [ms]	0.59286	0.46394	0.43389	0.42808	0.50308	2.72451	Launch failed	Launch failed

The number of threads per block should be chosen as a multiple of the warp size to avoid wasting computing resources with under-populated warps. Allocating more threads per block is better for efficiency. Increasing number of threads per block reduces registers availability per thread. 64 threads per block is minimal and makes sense only if there are multiple active blocks per multiprocessor. 192 or 256 threads per block is better and usually allows enough registers to compile. The ratio of the number of active warps per multiprocessor to the maximum number of active warps is called the multiprocessor occupancy. In order to maximize occupancy, the compiler attempts to minimize register usage while keeping the number of instructions and local memory usage to a minimum. For more accurate description of the computational capabilities for the specific model of GPU, please look at the reference [3 Appendix A].

2.4 Kernel configuration test

We composed kernel configuration test program to explore runtime differences across different kernel configurations. The free download of this program is available on http://academic.mnemonix3d.com website.

The program computed increment of elements of the large array (vector). Each element of the vector is handled by a single thread. This problem is suitable for computation on the GPU. There is a one dimensional dataset represented by the vector, and a simple kernel can be constructed to operate across this dataset. Dataset is one dimensional, so we need a one dimensional grid which is composed of one dimensional blocks. We set up different thread count per block and measured runtime of the kernel.

We started with 32 threads per block and we raised this number to 512. More threads per block decrease number of blocks per grid. We also tested three other cases which didn't respect block, grid limitations and recommended warp size.

As results clearly display (Table1), we achieved the best runtime with configuration of 256 threads per block. Block with 32 threads showed up longer runtime because minimal recommended block size (64 threads per block) was not set up. Case with 500 threads per block showed significantly longer runtime because the number of threads per block was not set as multiple of the warp size to avoid the lack of hardware saturation. In the cases with 600 and 29 threads per block kernel failed to launch because block (600 threads case) and grid (29 threads case) limit has been reached.

3 CONCLUSION

The use of the GPU in non graphics application is no longer long lasting and painful process. Experienced programmer can manage it in a short time. But we have to draw a line between GPU implementation and high performance GPU implementation. Firstly, decomposition of the problem which we want to solve on GPU has to be done. Few important questions have to be answered. Can I solve this problem using GPU? Is there a part in my problem which is parallel or can be paralleled? Is dataset large enough? How many dimensions are involved? All the answers can be used to build effective, properly sized kernel.

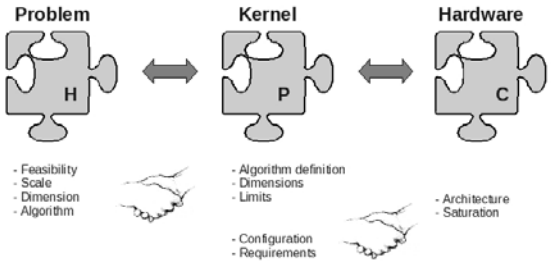


Figure 5 Problem, kernel and hardware dependencies

If we want to bust performance, not just make it work on GPU, we should understand limitations of hardware's computational and memory resources, which need to be considered during the kernel configuration process to keep hardware occupied as much as possible without crossing the limitation borders.

Program is compatible across different GPUs with unified architecture because it is a scalable architecture. However, in order to achieve the best performance results it is worth configuring kernel more specifically for your hardware. More powerful GPUs with more multi-stream processors need to be more saturated to keep them occupied. This is a process of finding balance between problem & kernel and kernel & hardware at the same time (Figure 5). This process requires a little experimentation to find the balance and identify which configuration is the best.

References

[1] Nvidia: Graphics adapters supporting CUDA http://www.nvidia.com/object/cuda_learn_products.html
[2] Nvidia: Software download for CUDA http://www.nvidia.com/object/cuda_get.html
[3] Nvidia: nVidia CUDA, Programming Guide. version 2.2, 4.2.2009.
[4] Nvidia: CUDA Technical Training, (2008). Volume I: Introduction to CUDA programming, http://www.nvidia.com/docs/IO/47904/VolumeI.pdf
[5] Wen-mei W. Hwu, David Kirk: Programming Massively Parallel Processors, University of Illinois, (2008), Course

Dipl. Eng. Miloš OČKAY
Armed Forces Academy of General M. R. Štefánik
Demänová 393
031 06 Liptovský Mikuláš
Slovak Republic
E-mail: milos.ockay@aos.sk

Assoc. Prof. Dipl. Eng. Marcel HARAKAĽ, PhD.
Department of Informatics
Armed Forces Academy of General M. R. Štefánik
Demänová 393
031 06 Liptovský Mikuláš
Slovak Republic
E-mail: marcel.harakal@aos.sk

Prof. Dipl. Eng. Miroslav LÍŠKA, CSc.
Department of Informatics
Armed Forces Academy of General M. R. Štefánik
Demänová 393
031 06 Liptovský Mikuláš
Slovak Republic
E-mail: miroslav.liska@aos.sk

ADAPTIVE FILTERING FOR NAVIGATION SYSTEMS OF ROBOT-AEROCRAFT

Andrey Victorovich PROLETARSKY, Konstantin Avenirovich NEUSIPIN

Abstract: The objective of this research are to design and implement an adaptive algorithm with Kalman filter that aimed to increase accuracies for measurement navigation systems of robot-aerocraft.

Keywords: Kalman filter, navigation system.

1 INTRODUCTION

Researches had been carried out with Kalman Filter and its errors due the lack of information and divergences of evaluation. The divergence of Kalman filter is result from the absence of reliability of input disturbance's statistics, not quite exact matrix that described models as well as of the errors of calculation when applied in computers. One result of research had been given is the normal or classical Kalman filters are not suitable to increase the accuracies of robot systems.

Some methods had been suggested to prevent the filter from its divergence such as: freezing method, comparison method, adaptive method. Since these discussion, the author had chosen an prospective adaptive algorithm. This algorithm used the method of continuously replaced and updated with signals and measurements.

Authors have been researched an improved adaptive algorithm that could operate well under condition that lack of initial information of stable characteristics of disturbances. Different from the known algorithms, with the improved algorithm, the full information of process or objects are taken into account. This information are derived from measurements when calculating the state of a specific variable. There are also many coefficients that had been referred in those measurements. Off course, only necessary and believable coefficients could be used with this method.

Analyzing all results from models of suggested evaluation methods, it appeared that the most efficient method is to average method applied on all periods of measurements of errors and to average in frame of sliding window Iazwinxki to determine errors in computing velocities.

Based on all analyze above, it could be come to conclusion that, an modified scalar adaptive evaluation algorithms are suitable for practical applications.

2 ADAPTIVELY EVALUATION ALGORITM

The invention of the modern digital computers has created new directions in optimal evaluation and control. One of the most widely used method is Kalman filters [1,2]. The method of Kalman garantees a "khongdich" evaluation with the minimum variance of state vectors of dynamic systems in which, input and measured disturbances are all the Gaussed- white noises. The Kalman filters are now widely applied because theirs filtering equations are quite simple in practice and not necessity of knowledge of optimal evaluation and control.

One main disadvatageous of Kalman filter is that its equation for a optimal filtering required precise mathematic model of dynamic systems and also exact statistics of accidental (random) values or signals. In some cases, it is necessary to know exactly the transient matrices of systems and covariance matrices between the white-noises. However, an adaptive Kalman filter enabes us to deal with this.

Adative Kalman filters are divived into some different groups but with following principles:
- Recognition of unknown parameters
- Have no covariance between recovering serries.

On the first direction, the state vectors in any kind of forms could be acessed, then those are sent to filtering equation of Kalman [3, 4]. In [4], the authors reseach the situation the value of trasient matrices are not known. These unkown parameters are related to the state vectors of systems and those vectors are improved with the use of improved filter called Kanman - Bius. In [3] suggested an convergence algorithm to evaluate the unknown parameters of linear systems when we did not completely known above the statistic characters of errors.

On the second direction, we use one of important characters of Kalman filter as the recovering serries are not conjugate with times. The basis of this method is that we can access the conjugation of times in recovering serries, then build up amplifier matrices that aimed to reduce the conjugation.

The methodology of evaluation uses complete mathematical models of researched processes. Therefore, it is very difficult to use classical algorithms to evaluate processes under limited condition of computation.

2.1 The improved adaptive algorithms for evaluation

Based on the algorithms to evaluate the time-variable components of the state vectors, in cases the information of these variable states could be directly measured, we should use the improved adaptive Kalman filters as follows [3, 4]:

$$\widehat{x}^i_{nk+1} - s_{ii}\widehat{x}^i_{n(k-1)+1} + u^i_k + k^i_{k+1}(z^{*i}_{k+1} - s_{ii}\widehat{x}^i_{n(k-1)+1} - u^i_k) \quad (1)$$

$$where: u^i_k = s_{i1}z^{*1}_k + s_{i2}z^{*2}_k + .. + s_{in}z^{*n}_k \quad (2)$$

The covariance of evaluated errors and the filtering gain matrix will be:

$$p^i_{k,k-1} = s^2_{ii}p^i_{k-1} + (k^i_{k-1})^2(v^i_k)^2 \quad (3)$$

$$k^i_k = \frac{p^i_{k,k-1}}{p^i_{k,k-1} + \widehat{r}^i_k} \quad (4)$$

$$p^i_k = (1 - k^i_k)p^i_{k,k-1} \quad (5)$$

where, the covariance of the measuring disturbances $\widehat{r}^i_k$ is determined by adaptive evaluation:

$$\begin{cases} \widehat{r}^i_k = (v^i_k)^2 - p^i_{k,k-1}; (v^i_k)^2 > p^i_{k,k-1} \\ 0; (v^i_k)^2 \le p^i_{k,k-1} \end{cases} \quad (6)$$

where v^i_k is the recovering series in adaptive - scalar algorithm, calculated in formula:

$$v^i_k = z^{*i}_k - s_{ii}\widehat{x}^i_{k-1} - u^i_{k-1} \quad (7)$$

or

$$k^i_k = \begin{cases} \dfrac{p^i_{k/k-1}}{M[(\widehat{v}^i_k)^2]}; if: M[(\widehat{v}^i_k)^2] > r^i, \\ and: M[(\widehat{v}^i_k)^2] > p^i_{k/k-1} \\ 0; if: M[(\widehat{v}^i_k)^2] \le r^i \\ 1; if: M[(\widehat{v}^i_k)^2] > r^i, \\ and: M[(\widehat{v}^i_k)^2] \le p^i_{k/k-1} \end{cases}$$

The above given algorithms of evaluation could be applied in case we lack of initial information about input and measured disturbances. The use of inversed-hardness relation by the recovering series v allows us to evaluate high-degree kinetics systems. With that, the accuracies of evaluation (in case covariance of measuring disturbances is bigger of equal to that of evaluation norms) will be reduced several times. In that time, when the information about kinetics system are not exactly, the inversed-hardness relation with v will enable us to get the non-divergence evaluation of the i - components of the state vectors.

3 EVALUATION IN UNUSUAL CONDITIONS

The measurement that used in evaluation algorithms, in principal, is modified due to noises. When they are use to be the source of information, the real information of objects will be changed by the high frequencies noises. The value of these noises are different from the measurement noises, and this different are depending on the working conditions of the objects. The sudden change of measuring noises due to switching from one working condition to another will create non-normal conditions called " humps " pulses in measurements. The measurements that contains those "hump" pulses are called non-normal measurements.

Solving problems or preventing the presence of non-normal pulses from evaluating the state vectors is very important in recognition theory. This work will be more important in adaptively evaluating that use characteristics of recovering series. This could be explained that when tuning method applied in recovering series (contains non-normal measurements), evaluation of the state vectors could be seriously damaged. With the sudden increases of recovering series, the covariance of the input disturbances will be sudden increased as a result while the input disturbances are actually unchanged. The non-normal measurements could cause the increase of gain matrix, then that led to presence of "random" pulses in evaluating the state vectors. In such cases, the divergence of evaluation process are often reached as a result.

Avoiding the influences of non-normal measurements in evaluating components of the state vectors, the evaluations of parameters of marine navigation systems are often implemented as follows:

First, it need to construct criterions for appearances of non-normal measurements. This criterion could be given in form:

$$tr(v_k v^T_k) \le k.tr[H.P_{k/k-1}.H^T + R_k] \quad (8)$$

where k - coefficient of non-normal measurements; $k = 7 \div 9$ and often chosen as value of v is not 3 times bigger than its means square value.

Criterions are chosen based on both conditions of means square values of recovering process and its theoretical estimated covariance.

The recovering series have the covariance matrix:

$$M[v_k v^T_k] = HP_{k/k-1}H^T + R_k \quad (9)$$

At any k -sampling interval, the chosen v_k must be satisfied equation (8). If this equation are not satisfied, the unusual pulse will be appeared in measurements. The frequency of measurement can be considered as unusual and this can be tuned by changing the coefficient k.

When non-normal measurement appeared, the value of gain matrix should be chosen to be zero.

The above criterion to recognize the presence of unusual measurement need to have information about correlation matrices of measured disturbances. When we do not have this information, we could apply criterion that based on implementing the recovering series.

Using formula in (8) and (9), we can derived the correlation matrix of that series:

$$k_k.M[v_k v_k^T].k_k^T = k_k H P_{k/k-1}$$

So that, we have criterion of appearance of unusual measurements with condition of accuracies:

$$tr[k_{k-1}.v_k v_k^T.k_{k-1}^T] \le k.tr[k_{k-1}.H.P_{k/k-1}]$$

All above criterions could be easily implemented with computer systems. The choice of the gain matrix in form of relay will allows us to eliminate all the influences of unusual measurements on evaluation of the state vectors.

4 MODIFY ADAPTIVE ALGORITHMS IN UNUSUAL CONDITIONS OF MEASUREMENT

So that, adaptive algorithms for evaluation (discussed in previous chapter) under unusual measurements will be improved as:

$$v_k = z_k - H_k \phi_{k/k-1} \hat{x}_{k-1} \quad (10)$$

$$P_k = \phi_{k/k-1} P_{k-1} \phi_{k/k-1}^T + K_{k-1}(v_k v_k^T) K_{k-1}^T \quad (11)$$

$$P_k = (I - K_k H_k) P_{k/k-1} \quad (12)$$

$$K_k = \begin{cases} P_{k/k-1}.H_k^T.[H_k P_{k/k-1} H_k^T + R_k]^{-1}, if. \\ equation(8) satisfied. \\ 0; if. equation(8) disatisfied \end{cases} \quad (13)$$

$$P_k = (I - K_k H_k) P_{k/k-1} \quad (14)$$

The special different of previous algorithms compared with this algorithms in equation (11) to (14) is the presence of relay form of the filtering matrix K_k.

5 SIMULATION

The application ability of above evaluation algorithm could be tested by method of simulation. On simple example is application in inertial navigation systems (INS). The error equation of INS are in form:

$$x_k = \phi.x_{k-1} + w_k$$

where:

ϕ - the system matrix and

$$\phi = \begin{bmatrix} 1 & 0 & 0 & -gT & f_y T & 0 & 0 & 0 \\ 0 & 1 & gT & 0 & -f_x T & 0 & 0 & 0 \\ 0 & -\frac{T}{R} & 1 & 0 & 0 & T & 0 & 0 \\ \frac{T}{R} & 0 & 0 & 1 & 0 & 0 & T & 0 \\ \frac{T}{R} tg\varphi & 0 & 0 & 0 & 1 & 0 & 0 & T \\ 0 & 0 & 0 & 0 & 0 & 1-\beta T & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1-\beta T & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-\beta T \end{bmatrix}$$

$$x_k^T = \begin{bmatrix} \delta V_x & \delta V_y & \phi_x & \phi_y & \phi_z & \varepsilon_x & \varepsilon_y & \varepsilon_z \end{bmatrix}$$

$$W_k = \begin{bmatrix} B_x & B_y & 0 & 0 & 0 & W_1 & W_2 & W_3 \end{bmatrix}$$

where:

$-\delta V_i$ is the error of speed (i=x, y),

$-\phi_i$ is the slanted angle of compass. (i=x,y,z),

$-\varepsilon_i$ is angle velocity (i=x,y,z),

$-B_i$ is the error of acceleration measurement device.

In order to access quality of measurement, we use the difference between two measurement equipment when calculating the speed of object (INS and DISS)

The measurement equation:

$$z_k = Hx_k + V_k$$

where: - W_k, V_k are white noises;

- $V_k = \begin{bmatrix} V_1 & V_2 \end{bmatrix}$

- $H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

- sampling period: T=0.2 min

- The initial values:

$B_i = 5.10^{-4}$,

$x_0^T = 100 \quad 100 \quad 1E-4 \quad 1E-4 \quad 1E-4 \quad 0 \quad 0 \quad 0$

In order to function algorithms, it need to assign the initial value of the correlation matrix of

evaluation errors: $P_0 = M[x_0 x_0^T]$. The initial value of correlation matrix of measuring disturbance. Frequency of random overflow $\beta = 1E-3$ (1/min).

We can see that, if UHC working with automation condition in 01 hours (in principle), then we start tuning signal of marine parameters together with using tuning equipment on land. So it is enough time (01 hours) for us to stimulate algorithms of evaluation.

The evaluation with above algorithms are not much different from one another. Therefore, results here are derived from one evaluation algorithm.

The adaptive filtering process will start its satisfaction in evaluating errors of UHC with speed since the 5th steps, with angle error since the 30th steps, slanted angle of gyro ($x_6 .and. x_7$) since 50th steps. The overflow in vertical lines during this period are not considered, and often very weak and this needed to observe in longer period of time.

6 ANALYSIS WITH EXAMPLES

With application of above adaptively algorithms, the accuracies of marine navigation systems had been drastically improved and it enabled us to revise the accuracies of the output information's.

These adaptive algorithms could be functioning well without enough information of the correlation matrices of input disturbances. Whereas, we can derived these information from recovering processes of input disturbances. In the given algorithms, all information of the recovering processes are used instead of the correlation matrices.

In order to compensate the influences of evaluation when working with these algorithms, all unusual pulses are eliminated by the improvement of the adaptive algorithm where the relay style had been applied for the gain matrices.

These designed algorithms require small amount of memories, easy to implement with computers, ensure high accuracies with lack of information of input noises.

The Kalman filter in inertial systems has required the outside supplemental information's, normally using equipment called DISS. Information of speed of object are sending from INS to the N^02 adder. The second output signal is sent to DISS. The output of the N^02 adder, we receive signal that proportional to speed measured in INS and DISS. These signal are sent to Kalmal filter to get better quality. The measuring statistics are driven to the error calculation block of UHC where the block of filtering gain matrix calculation and block of error evaluation are created to make equivalent model of INS.

The filtering gain coefficients are calculated based on deduced information about statistics of measured disturbances (R) and the correlation matrix of evaluated disturbances. The evaluated errors are sent from the block of error evaluation, where the correlation matrices are calculated by using the input disturbances Q and gain matrix K_k in previous step.

Signal of evaluated errors are sent to the negative terminal of the N^01 adder which is the overall signal INS (real signal plus errors due to INS). The information above speed and position of objects are derived at the output terminal of N^01 adder.

The Kalman algorithms require small amount of memories, easy to implement with large computer systems. However, if the information about input and measured disturbances Q and R, also without exact model of systems, the Kalman filter will function with notable errors.

In practice of application, they are not used in object's desk because of the divergence character of these Kalman filters.

7 CONCLUSION

1. Researches had been carried out with Kalman Filter and its errors due the lack of information and divergences of evaluation. The divergence of Kalman filter is result from the absence of reliability of input disturbance's statistics, not quite exact matrix that discribed models as well as of the errors of calculation when applied in computers. One result of research had been given is the normal or classical Kalman filters are not suitable to increase the accuracies of robot systems.
2. Some methods had been suggested to prevent the filter from its divergence such as: freezzing method, comparison method, adaptive method. Since these discussion, the author had chosen an prospective adaptive algorithm. This algorithm used the method of continuously replaced and updated with signals and measurements.
3. Author had been researched an improved adaptive algorithm that could operate well under condition that lack of initial information of stable characteristics of disturbances. Different from the known algorithms, with the improved algorithm, the full information of process or objects are taken into account. This information are derived from measurements when calculating the state of a specific variable. There are also many coefficients that had been referred in those measurements. Off course, only necessary and believable coefficients could be used with this method.

References

[1] КУЗОВКОВ, Н. Т., САЛЫЧЕВ, О .С.: Инерциальная навигация и оптимальная Фильтрация. М.: Машиностроение, 1982, 215 с.
[2] НЕУСЫПИН, К. А., ПРОЛЕТАРСКИЙ, А. В., ЦИБИЗОВА, Т. Ю.: *Системы управления летательными аппаратами и алгоритмы обработки информации.* – М.: Изд. МГОУ, 2006, 262 с.
[3] САЛЫЧЕВ О. С.: *Скалярное оценивание многомерных динамических систем* - М.: Машиностроение,1987, 216 с.
[4] JAZWINSKI, A. H.: *Stochastic Processes and filtering Theory.* N.Y., 1970, 376 p.

Prof. Andrey Victoriovich PROLETARSKY, Ph.D.
Prof. Konstantin Avenirovich NEUSIPIN, Ph.D.
Departament Automatic Control Systems
Bauman Moscow State Technical University
2 nd. Baumanskaya, 5
105 005 Moscow
Russia
E-mail: pav_mipk@mail.ru

CONTRIBUTION TO ESTIMATION OF ECONOMIC CONSEQUENCES OF DISASTER

Jaroslav SLEPECKÝ

Abstract: In this paper I deal with the economics of disasters and related problems to this subject. The paper is divided into two main parts. First part provides a perspective toward disaster related research with focus on theoretical consequence of disasters. In recent years significant progress has been made in economics of disasters worldwide. The review is carried through new findings from recent studies and research. The second part point on losses of disasters and is dividing them into three types, direct losses, indirect losses and secondary effects of natural disasters. Each type is characterized by individual characteristics.

Keywords: Economics of Disasters, Economics of Catastrophes, Crisis Management.

1 INTRODUCTION

At any time, some natural disaster has just struck or is about to strike somewhere in the world, taking a toll of human lives and material losses. Thus, natural disasters are human, social and economic disasters. Whether it has a disastrous impact or not depends upon where the natural event strikes. An earthquake or hurricane that strikes a densely settled region can have disastrous consequences, whereas a similar natural event might cause little havoc if it hit an unpopulated area. Today's populated areas–cities and agricultural zones–constitute an increasingly valuable asset base. Potential human, social and economic losses from natural disasters therefore grow year by year, independently of Nature's forces. Increased vulnerability requires that natural disaster management and economics of disasters be at the heart of economic and social development policy of disaster prone countries.

2 DISASTERS

As a part of scientific research we can focus on theoretical consequence of disasters as it is in Figure 1.

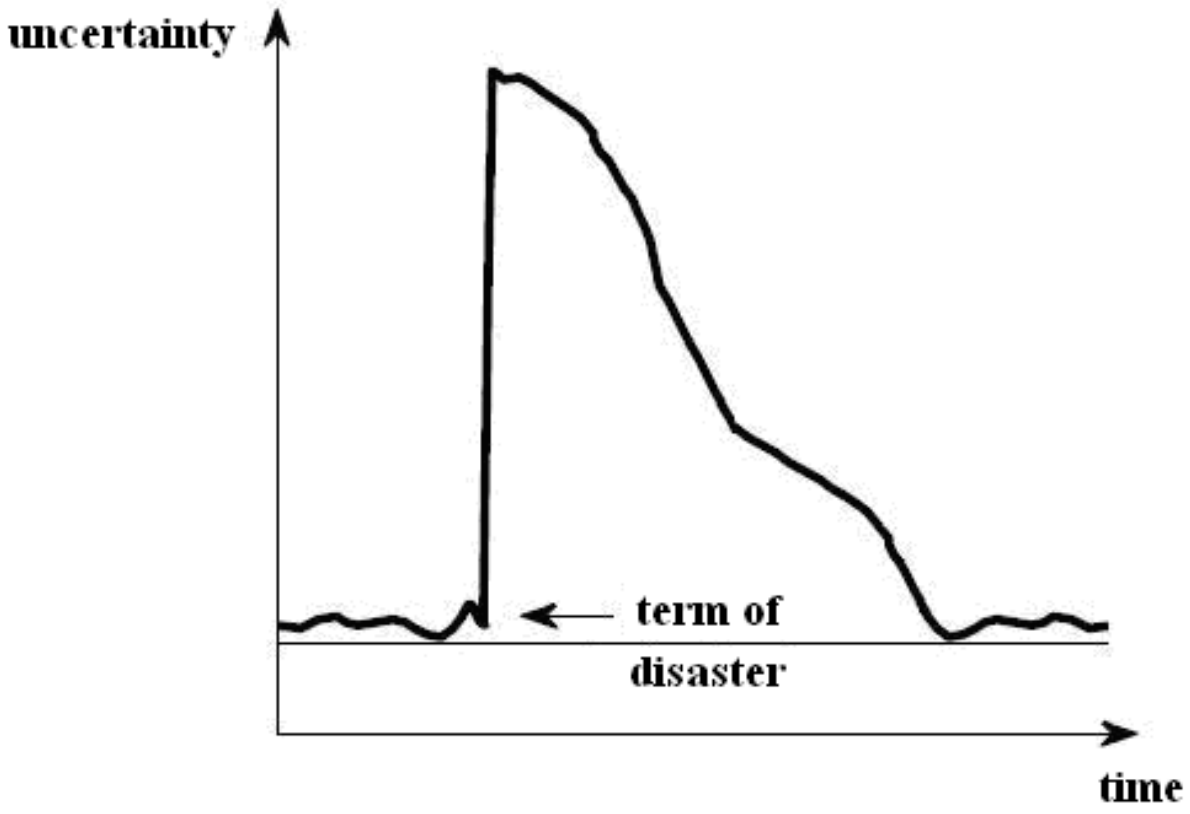


Figure 1 Theoretical consequence of disasters

Disasters are by definition extreme events. Over the last 30 years 6367 natural disaster killed more than 2 million people and caused 1.4 trillion USD worth of damages. 255 million people affected each year, average number between 1997 and 2006. During the last decade disasters caused damage of an estimated 67 billion USD per year on average. (Numbers provided by EM-DAT 2007) Therefore data on natural disasters and their impacts on economies pay a very important role in understanding the factors that increase human vulnerability and the importance of preparedness and prevention. In this, part of this research is also work with EM-DAT - International Emergency Disaster Database, which presents core data on the occurrence and effects of over 15000 disasters from 1900 to present, including natural a technological disasters.

3 ECONOMICS OF DISASTERS

Economic reality belongs to objects with difficult research areas. Economics of Disasters is one of them, because it is focusing on processes that are very difficult to predict and accrue from step changes. Economics of Disasters search for problems with instability of dynamic economic system (environment) and it is reaction to step changes caused by external or internal conditions.

Subject of Economics of Disasters is identification of threats of instability in economic system, specification of priorities and possibilities to mobilize resources for stabilization or elimination of losses.

Position of Economics of Disasters in Traditional Classifying of Economic Sciences environment is in Figure 2.

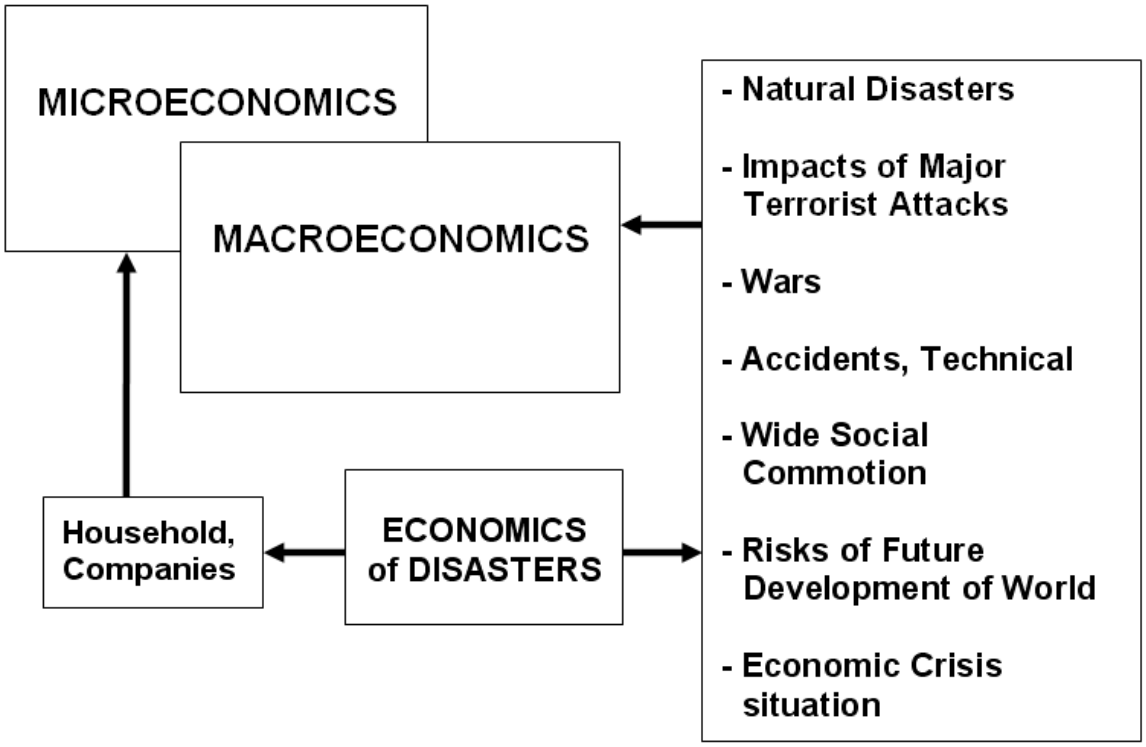


Figure 2 Position of Economics of Disasters in traditional classifying of economic sciences

4 LOSSES

According to this research a several published papers from representatives of World Bank and other institutions we can divide losses of disasters into three main types, direct losses, indirect losses and secondary effects of natural disasters. Each type is characterized by individual characteristics.

Direct losses are accounted for by the financial value of damage to and loss of capital assets– buildings, infrastructure, industrial plant, and inventories of goods including crops. Direct losses are usually the most readily assessed after a natural disaster has struck. They refer to readily visible effects of disasters that are widely disseminated through the mass media and which make up governments' initial damage assessments. Past experience with measuring direct losses of natural disaster events make estimating the direct loss potential from future natural disasters a relatively straightforward exercise, albeit based upon probability assessments of the likelihood of a natural

disaster strike. In economic terms, direct losses such as these can be equated with stock losses.

Indirect losses arise from interrupted production and services, measured by loss of output and earnings. Damage to roads and ports, for example, can hold up exports, imports, distribution of basic necessities affecting health and education, as well as other productive sectors. Natural disasters in geographically large countries may not have easily observable impacts on national GDP figures, but they can still affect provincial economies significantly. On the other hand, effects of natural disasters can spread beyond national borders. When an earthquake destroyed the central telephone exchange in Mexico City (in 1985), for instance, Central American countries were affected as their transmission lines ran through Mexico City. In economic terms, such indirect losses can be equated with flow losses.

Secondary effects of natural disasters are felt through longer-term impacts upon economic performance. There are many such effects. In economic terms, secondary losses such as these can be counted among the negative externalities of natural disasters. There can also be positive effects of disasters, which provide unexpected opportunities to upgrade plant and machinery or renew aging infrastructure. In most cases, however, benefits are unlikely to outweigh the costs of the losses. Whether positive or negative, secondary effects are not easily to estimate in practice.

Even the impacts can be seen from longer point of view as opportunities (Although natural disasters spread destruction and economic pain to a wide variety of businesses, for some, it can mean a burst of activity and revenue.), we are still left with the hard economic reality that disasters impose losses on economies now. The destruction of resources reduces GDP by reducing productive capacity: fewer inputs mean fewer outputs.

5 CONCLUSION

"The Economics of Disasters" has become even more important in recent years because the economic damage caused by various degrees and types of disasters are constantly increasing and has a strong influence on and overloads the financial budget of individual states that, for obvious reasons, do not create reserves for uncertainty in the future. A number of industrial breakdowns, which happened in our territory recently as well as natural disasters from floods to snow calamities could have been successfully solved by us only with use of effective integrated systems (Simak-Ristvej, 2009). In many cases the results of unexpected events are greater than the affected state is able to cope with. Previously unknown reinsurance companies based in so-called tax havens have taken advantage of stricter conditions for addressing the risk of potential losses from natural disasters and successfully offered their services to insurance companies. Many of them invest large amounts in the development and operation of complex systems to provide advance warning of hurricanes etc.

References

[1] ARNOLD, V. I.: 1986, *Catastrophe Theory*, (Slovak version), Bratislava : Alfa, 1986, ISBN 63-204-86.

[2] EM-DAT. 2007. *The OFDA/CRED International Disaster Database*, www.em-dat.net, Université Catholique de Louvain, Brussels – Belgium, [online] <www.em-dat.net>, 2009–05–05.

[3] GILBERT, R., KREIMER, A.: 2005, *Learning from the World Bank's Experience of Natural Disaster Related Assistance.* [online] < http://www.worldbank.org/hazards/files/learningfromwb.pdf >, 2009–11–09.

[4] LOVECEK, T.: *Security systems – Safety of information systems*. EDIS, 2008. ISBN 978-80-8070-767-5.

[5] KUNREUTHER, H. ROSE, A.: *The Economics of Natural Hazards Volume I.-II.* Northampton-MA USA, An Elgar, 2004, ISBN 1-84064-815-5.

[6] RISTVEJ, J.: *Disasters and Economic Impacts*, In: zborník z konferencie Transcom 2007, 25.-27. jún 2007, Žilina : Žilinská universita, 2007, str. 77-80, ISBN: 978-80-8070-698-2.

[7] RISTVEJ, J.: Ekonomické dôsledky katastrof a krízový manažment. In: zborník z konferencie "*Riešenie krízových situácií v špecifickom prostredí*," 9.-10. november 2006 SPU, Nitra: 2006, str. 112-116, ISBN: 80-89162-24-X.

[8] RISTVEJ, J.: Možnosti výpočtov makroekonomických vplyvov katastrof na ekonomiku, In: zborník z konferencie *"Riešenie krízových situácií v špecifickom prostredí"*, 20.-21. jún 2007, Žilina: FŠI ŽU, 2007, str. 561-566, ISBN 978-80-8070-702-6.

[9] SIMAK, L. RISTVEJ, J. (2009) *"The Present Status of Creating the Security System of the Slovak Republic after Entering the European Union,"* Journal of Homeland Security and Emergency Management: Vol. 6 : Iss. 1, Article 20. ISSN 1547-7355. DOI: 10.2202/1547-7355.1443, Available at: http://www.bepress.com/jhsem/vol6/iss1/20

[10] SLEPECKY, J. RISTVEJ, J.: *Economic Impacts of Disasters*, Zilina : EDIS 2008. 136 p. ISBN 978-80-8070-830-6.

[12]WHITTOW, J.: Disasters: *The Anatomy of Environmental Hazards*. Athens, GA, USA : University of Georgia Press, 1980, ISBN 08-2030-542-1.
[13]The World Bank Group: 2005, Lessons from Natural Disasters. [online] < http://www.worldbank.org/oed/disasters/lessons_from_disasters.pdf >, 2008–05–05.

Assoc. Prof. Dip. Eng. Jaroslav SLEPECKÝ, PhD.
Institute of Economic and Enviromental Security
University of Security Management in Košice
Ul. Kukučinova 17
040 01 Košice
Slovak Republic
E-mail: jaroslav.slepecky@gmail.com

INTERACTIVE VISUALIZATION OF ABSTRACT DATA

Martin ŠPERKA, Peter KAPEC

Abstract: Information visualization is a large research area. Currently with more powerful computers and graphic accelerators more and more visualization techniques become part of daily use. In this paper we discus visualization of abstract data – data that is difficult or impossible to manually grasp. Using visualization of abstract data we can gain better insight. We present several experimental methods for visualizing graphs and show possible applications in the software visualization field.

Keywords: Information visualization, visual data mining, software visualization.

1 INTRODUCTION

With the exponential growth of data caused by the penetration of information and communication technologies to every branch of human activities, grows the demand on extracting facts and knowledge from the flood of these data by means of business intelligence and data mining. Visual data mining is the method which is often used in many fields by using different data and science visualization methods and tools. Special category is visualization abstract data in form of graphs.

Abstract data visualization is used in different application areas as organizational structures, family trees, geographical data, transport, communication and social networks, ontologies, information systems – hardware interconnection, software, distributed and collaborative systems and modelling (for example UML).

The most cited visual investigation technique, and not only related to abstract data, is the visualization seeking mantra - *Overview first, zoom, filter, and then focus details-on demand* [22]. Users first see data in general view, then select subset of his/her interest and finally focus on particular attributes of selected data objects. Visualization mantra is an interactive process requiring fast response of computer, which is especially in case of large data sets not trivial. However new multi-core processors and GPU allows to render complex scenes and new forms of interactive visualization are possible. In this paper we discus the visualization process, show several experimental methods of graph visualizations and present applications in software visualization.

2 INFORMATION VISUALIZATION

The visualization process is a transformation of data in one representation to another, mostly to a representation better observable by humans. The following steps of a visualization process can be found in any problem area [21]:

- data preparation;
- encoding;
- presentation and interaction.

Preparation, the first step of the visualization process, is used to identify relevant entities and events that the visualization will deal with.

The second step encoding deals with problems how the data will be displayed. The questions to be considered are oriented to efficiency, aesthetic, understanding, similarity etc. These aspects of visualization play major role for humans that will be involved in the visualization process, because when the visualization is not understandable or uses non-standard visuals, the benefits of visualization may get lost.

The final phase is presentation and interaction and should answer questions about how the visualization objects are displayed and which interaction possibilities are offered to the user. For each specific visualization the vocabulary used is also important. We can identify many aspects for graphical elements of a vocabulary [16].

Visual exploration techniques can be classified according to three orthogonal criteria [13]:

- data to be visualized;
- the visualization technique;
- interaction and distortion technique used.

Data type may be one-, two- and even n-dimensional data, text and hypertext, hierarchies and graphs, temporal data and recently also algorithms and software. Visualization techniques may be classified into standard 2D/3D displays, geometrically transformed displays, icon-based displays, dense pixel displays and stacked displays. Interaction and distortion techniques may be classified into interactive projection, filtering, zooming and distortion techniques

These techniques can be freely combined to form new exploration techniques suitable for specific applications.

3 VISUALIZING GRAPH STRUCTURES

Visualizing graphs is not trivial and involves several theoretical and practical problems. The main problem is the size and density of the visualized graph – when denoting the number of nodes as |N|

and the number of edges as |E|, we can categorize graphs into [16]:

- sparse |E| <= |V|
- normal |V| < |E| <= 3|V|
- dense |E| > 3|V|

The graph size directly affects scalability of graph layout algorithms [2] – they often work for small/sparse graphs but are to slow to be usable or even fail to finish the layout when applied to large/dense graphs. To lower the graph size we can layout only the spanning tree of a graph, ignore edges with edge weight lower than some limit or create clusters.

The problem of display area is also related to graph size and graph layout algorithms. Displaying the whole graph in a limited display area is often not very comprehensible or even technically not possible. This problem can be solved using appropriate visualization technique, e.g. *focus+context* or *distortion techniques*.

Another problem is related to incremental and dynamic changes in graph structures often caused by user interaction. The graph layout algorithm should be capable to handle local graph modifications without the need to modify already finished graph layouts – force-based layout algorithms are very suitable for these cases.

Moving into 3D space may offer "more" space for visualization, however it may introduce more problems. Typically objects in 3D space overlap after projection making them difficult to observe. This can be solved by finding best views in which aesthetic criteria are fulfilled. The most obvious problem is that users often look at 2D projections of 3D space, thus interaction and navigation becomes more complex. The current trend of using 3D displays may provide better insight into visualizations, but effective user interfaces are still an open challenge.

3.1 Criteria for graph visualization

An important criterion for understanding visual information is ordering elements. Basic aesthetic principles and rules for visualization of graphs, according to [1], are related to

- positioning of nodes:
  - balanced node placement - symmetry
  - not overlapping nodes
  - related nodes create clusters
  - nodes are not to close to edges
  - maximize node orthogonality
- edge placement:
  - minimize edge crossing and bending
  - equal edge lengths
  - maximize angles between edges
  - maximize edge orthogonality
- whole graph layout:
  - maximize graph global and local symmetry
  - minimize layout area
  - adjust layout area to display area

In case of trees there are additional rules

- nodes at the same level should be on same horizontal lines
- equal distance among node's children

Several of these criteria are related and can be applied together, but some conflict, thus when developing a graph layout algorithm we have to follow aesthetic criteria that will provide best graph layout for specific applications. Of course there are specific requirements for different graphs and applications. One of them is to minimize the graph area.

Another requirement is predictability; it means that when we render two topologically similar graphs, the results should not be very different. In case of different results, the user could be confused – the observer's mental map would become deformed [17]. In general not all rules and aesthetic criteria can be fulfilled in case of especially large graphs. Interactive 3D graphics and virtual reality helps to overcome the difficulty in visualizing large data sets.

3.2 Graph layout methods

Graph layout algorithm can be categorized into two categories: *deterministic* and *non-deterministic*.

Deterministic layout algorithms use exact equations to place nodes. Typical examples of this approach are layered/hierarchical views [9], Reingold-Tilford's views [19], cone trees [20] and radial views [9], tree-maps [10] etc.

Non-deterministic graph layout algorithms use some physical model in which nodes are positioned by applying forces and their final position is reached when the whole system reaches minimal energy state. These force-based methods consist of a model and algorithm. The model defines graph nodes as physical objects that react on forces. The algorithm then iteratively reassigns node positions according to forces until an equilibrium state is reached. Force-based layout methods have many advantages:

- easily implementable
- very parametrizable
- modifiable by adding new forces
- effective for small graphs
- produce symmetrical layouts
- animation of layout preserves mental map
- easily expendable into 3D

Disadvantages are:

- slow for large graphs
- final layout is not predictable

The first force-based layout method was developed by Eades [6] in which adjacent nodes are attracted by spring forces defined by edges and nodes are repelled when no edge connects them. Various modifications of this approach have been developed [12], each scaling better for larger graphs. Recently implementations of force-based layout algorithms on GPUs allow to layout very large graphs in very short time [7].

3.3 Example graph visualizations

In one project we deal with using virtual reality in graph visualization [3]. The goal of this project was to try to increase the visual perception and help to discover facts about mutual connections between different trees, representing hierarchical structures for example companies, families or clans. Trees are displayed on different planar and semitransparent layers as displayed in Figure 1.

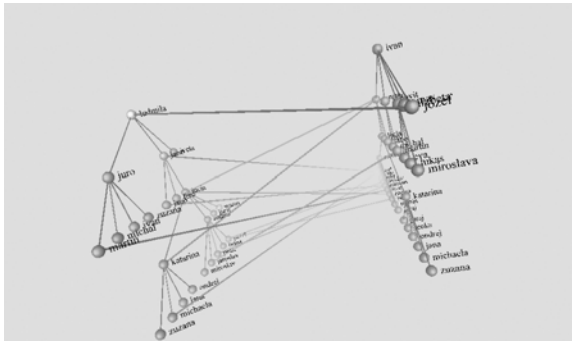


Figure 1 Trees displayed on two layers

Observer can move in Z direction (forward and back) and see different trees individually. It is possible to zoom and see overview or details of the tree. The layer can be opaque and then no other layers (trees) are visible or transparent so other layers are visible and the context with selected tree is observable. The transparency is arbitrary adjustable by the formula $I = I_a * alpha + I_r * (1 - alpha)$ where I_a is luminance of selected layer, I_r is luminance of other layers and alpha is variable in the interval <0,1> . Depth cueing (layered fog) helps to increase visual separation of different trees. When the user wants to see connections between different trees the view is rotated 90 degrees around X or Y axis and different trees are then represented as horizontal lines or vertical lines. The view can be rotated about all axes in arbitrary angle so different axonometric and perspective projections can be used for visual inspection. This tool was implemented in

Java language and the resulting trees are exported to VRML 97 language.

In another project we experimented with a metaphorical visualization of graphs that displays graphs as soap bubble clusters [24]. Nodes are displayed as soap bubbles and edges as thin soap membrane between adjacent nodes. The visualization is expected to produce rather dense cluster, therefore actual edges are displayed as straight lines between soap bubble centres. The bubble layout is based on force-directed placement as described in section 3.2. An example graph visualization using soap bubble metaphor is shown in Figure 2.

To create thin bubble membranes between adjacent bubbles we used the approach by Sunkel et al [23]: firstly intersection planes of colliding bubbles are identified and afterwards vertices of a sphere exceeding into another sphere are projected onto the intersection plane along the membrane normal. To accelerate membrane creation these calculations were implemented as vertex shaders.

Due to the requirements of real-time visualization the visual properties of real soap bubbles are limited to simulating the Fresnel effect, interference and environment reflection and are based on existing approaches [8].

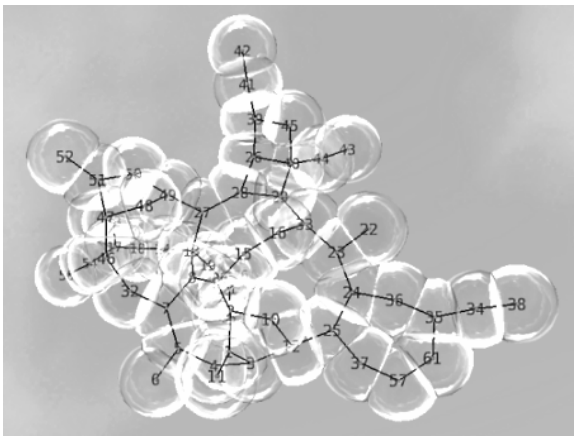


Figure 2 Graph visualized as soap bubble cluster

One problem, which is also in focus of our research, is 3D presentation of trees with perspective stereoscopic projection [18]. Stereo pair of images can be viewed on the desktop computer with the synchronized shutter glasses or red and cyan colours filters. Several VRML or X3D plug-in viewers have this option. Figure 3 illustrates results of an experimental application using WEB3D technology with stereo output – visualization of the binary tree showing results from the ice hockey championship.

Figure 3 Stereo pair of interactive binary tree visualization in 3D space

To accelerate the overview-zoom-detail cycle process we explored methods that map 2D images onto a sphere. The sphere can be interactively rotated and zoomed, allowing to see the whole graph and its details without loosing context. When looking from the outside of the sphere, projection is similar to the fish eye projection, where the observer can concentrate on details of the object at the centre. Distant objects are scaled down and situated at the periphery of the sphere. When looking from inside of the sphere, the centre is far from the viewer and objects on the periphery are bigger. Mapping image stereo pairs on the sphere is one form of visual exploration in an immersive virtual environment – half sphere dome or spherical CAVE. Figure 4 shows graph seen from the inside and outside of the sphere.

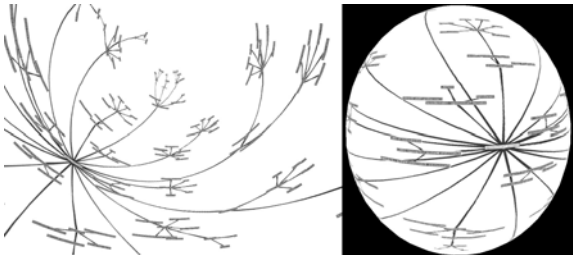


Figure 4 Objects seen from the inside (left) and outside (right) of the sphere

Graph visualizations have found many applications in many areas. Our research applies known and experimental graph visualizations in the software visualization field.

4 SOFTWARE VISUALISATION

Software can be considered as a special data type that is very suitable for visualization. The intangibility of software components makes it very difficult to comprehend all aspects of software systems – especially today when we look at enterprise systems and their increasing complexity. Software is not only source code, but consist of many artefacts including data, algorithms, documentations, user interfaces etc. and all possible documents related to software development. These software artefacts occur in the whole development process. Graphs and their visualizations are often used in the software visualization field.

Software visualization can focus on three main aspects: *software structure, behaviour of executing processes* and *evolution of software development* [4]. In our experiments we focused on visualizing structure of existing software systems. To visualize class inheritance we implemented a modified cone-tree visualization that displays namespaces, classes and their inheritance on layered circles [14]. For each class inheritance relations are shown as links to higher layers. Methods and attributes of classes are positioned on the outside of each cone. A visualization of an existing system containing nearly two thousand software artefacts is shown on Figure 5.

For visual data-mining the implemented visualization system allows to filter out not important elements and to sort them according to user preferences.

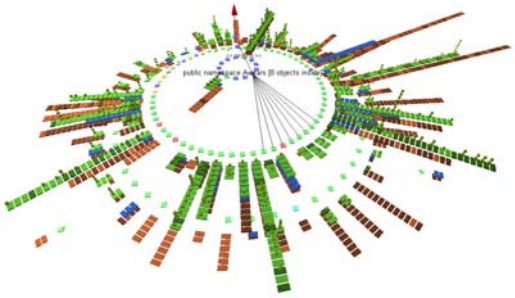


Figure 5 Visualization of an existing system

The Figure 6 displays the visualization of the same system, but with only classes and their methods and sorted by the decreasing number of methods a class has.

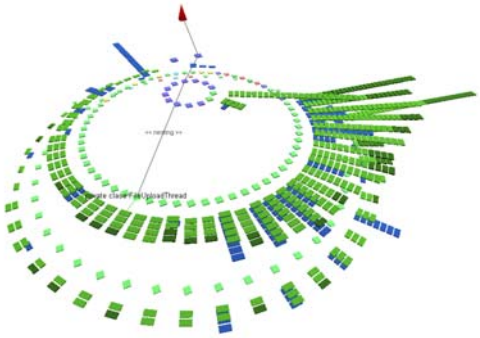


Figure 6 Sorted view of classes and their methods

To represent relations between various types of software artefacts we developed a hypergraph-based model that allows to store relations between semantically different software artefacts [11]. The hypergraph model uses hyperedges that allow to connect more than two nodes with one hyperedge.

This allows storing complicated relations between artefacts with one relation for which standard graphs would need several edges. A typical example of such relations can be found in a call-graph that represents calling relations between functions or methods. Functions can be represented by hypergraph nodes. Using one hyperedge we can connect all those functions that are called by the function of our interest. Based on this hypergraph model we developed a software visualization system that utilizes hypergraphs in the whole visualization process: hypergraphs are used to store software artefacts and their relations, they are used to query and filter the analysed software systems and finally are used in 3D visualization. Force-based graph layout algorithms can be used also for hypergraphs due to a mathematical transformation of hypergraphs into bipartite graphs.

A hypergraph-based visualization of a small open-source system is shown in Figure 7. The visualization shows nearly the whole extracted hypergraph containing more than thousand software artefacts, mostly source code elements and related documentations, and more than five hundred relations related to these artefacts.

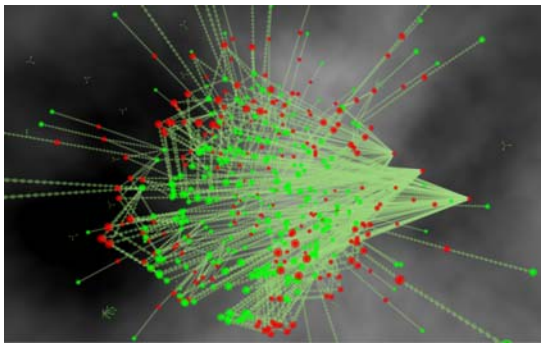


Figure 7 Hypergraph visualization of an existing open-source system

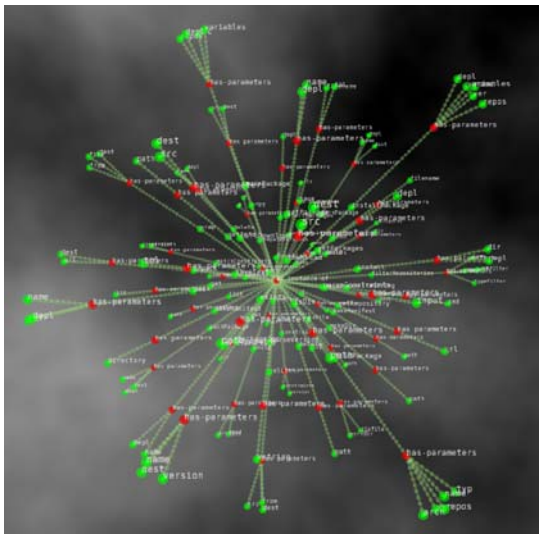


Figure 8 Results of a hypergraph query

The hypergraph model allows us to look at software artefacts as knowledge repository with query functionality. A hypergraph representation of software can be queried by hypergraph queries – the results of these queries are also hypergraphs, thus making this concept transparent. A result of a hypergraph query applied to the whole hypergraph representation of the system in our study is shown in Figure 8. The query extracted all functions, their parameters and return values and related documentation from the whole hypergraph.

Dynamic aspects of software visualization are also in our focus. We developed a visualization system that captures the execution of JAVA programs and stores it for post-mortem analysis. This record is then used for visualization in a separate application that allows to playback this record, fast-forward playback or jump to specific time. The record contains all important information suitable for debugging and analysis of executed programs. The user can step the animation forward and also backward, thus providing advanced debug functionality not available in current development environments. Figure 9 displays a snapshot showing a class in figure centre and living instances of this class at a specific moment during program execution. Important events during program execution are highlighted.

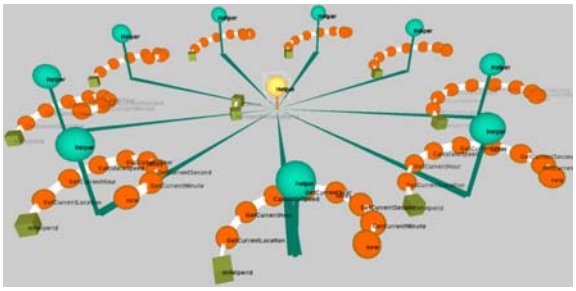


Figure 9 Class and instances during program execution

All presented experimental software visualization systems use 3D space for visualization in which users can interactively navigate using a virtual camera. The virtual camera we use allows fly-mode for free graph exploration and also orbiting around selected nodes. This way the user can zoom to nodes of interest and by orbital movements explore relations originating from selected node.

5 CONCLUSION

In this paper we presented several experimental graph layout algorithms. Our main research is however more focused on software visualization where we focus on structural and behavioural aspects of software systems. For this purpose we developed several visualization systems that display software artefacts and their relations as graphs or

hypergraphs. Proper evaluation of these visualization systems in practice is currently in our main focus.

Future work will be oriented on more effective layout algorithms capable to layout very large graphs containing hundred-thousands of nodes and (hyper)edges, thus allowing to visually analyse enterprise size software systems. Also work is oriented to provide not only visualizations for analysis, but also provide a visual programming environment that seemingly integrates hypergraph visualizations with textual programming in 3D virtual environment.

This work was partially supported by the grant KEGA 244-022STU-4/2010: Support for Parallel and Distributed Computing Education. We would like to thank master degree student Ivan Ruttkay-Nedecký for developing the program runtime visualization system.

References

[1] BENNETT, C., RYALL, J., SPALTEHOLZ, L., GOOCH, A..: The aesthetics of graph visualization. In: *Computational Aesthetics*, pages 57–64. Eurographics Association, 2007. ISBN 978-3-905673-43-2.

[2] CHEN, C.: *Information Visualization*, chapter 3: Graph Drawing Algorithms. Springer, 2004. ISBN 978-1-84628-340-6.

[3] ČÍKOVÁ, Z.: *Interactive Visualization of Large Graphs with Virtual Reality*. Bachelor thesis. FIIT, Slovak University of Technology (supervisor M. Šperka), Bratislava, 2009.

[4] DIEHL, S.: *Software Visualization: Visualizing the Structure,* Behaviour, and Evolution of Software, 1 ed. Springer, May.

[5] EADES, P.: Drawing Free Trees. In *Bulletin of the Institute for Combinatorics and its Applications*, 1992, s. 10– 36.

[6] EADES, P.:. *A heuristic for graph drawing. Congressus Numerantium*. 1984, vol. 42, is. 1, s. 149-160.

[7] FRISHMAN, TAL.: *Multi-level graph layout on the GPU*. IEEE transactions on visualization and computer graphics, vol. 13, no. 6, 2007.

[8] GLASSNER, A.: *Soap Bubbles: Part2. IEEE Computer Graphics and Applications*, (2000), vol. 20, no. 6, pp. 99-109.

[9] HERMAN, I. et al.: Graph *Visualization and Navigation in Information Visualization*. 2000.

[10] JOHNSON, B., SHNEIDERMAN, B.: Tree-Maps: A Space-filling Approach to the Visualization of Hierarchical Information Structures. In *Proceedings of the 2nd conference on Visualization (Vis'91)*, pp. 284–291, 1991.

[11] KAPEC, P.: *Hypergraph-based software visualization, International Workshop on Computer Graphics*, Computer Vision and Mathematics, GraVisMa'09, pp 149-153, 2009

[12] KAUFMANN, M., WAGNER, D.: *Drawing Graphs: Methods and Models.* 2001.

[13] KEIM, D. A.: *Information Visualization and Visual Data Mining*, IEEE Transactions on visualization and computer graphics, vol. 7, no. 1, pp 100-107, 2002.

[14] KOMOROVSKÝ, F.: *Visual data-mining in software, Master thesis.* FIIT, Slovak University of Technology, Bratislava, (supervisor P. Kapec) 2009.

[15] MACKINLAY, J.: *Automating the design of graphical presentations*. ACM Transactions on Graphics, 5(2):110–141, 1986.

[16] PAJNTAR, B.*: Overview of algorithms for graph drawing.* Conference on Data Mining and Data Warehouses. 2006.

[17] PARKER, G., FRANCK, G., WARE, C.: *Visualization of Large Nested Graphs in 3D: Navigation and Interaction,* Journal of Visual Languages and Computing, vol. 9, pp 299-317, 1998.

[18] POZOR, M.: *Rendering Graphs in Virtual Reality Environment*. Master thesis. FIIT, Slovak University of Technology Bratislava, (supervisor M. Šperka) 2006

[19] REINGOLD, E. M., Tilford. J. S.: Tidier *Drawing Of Trees*, IEEE Transactions on Software Engineering, 1981, Vol. SE-7, No. 2, pp. 223-238.

[20] ROBERTSON, G. G., MACKINLAY, J. D., CARD, S. K.: Cone trees: Animated 3D visualizations of hierarchical information. In *Proceedings of ACM Conference on Human Factors in Computing Systems.* 1991. s. 189-194.

[21] SCHUMANN, H., MULLER, W.: *Visualisierung: Grundlagen und allgemeine Methoden. Springer-Verlag*, Berlin, 2000.

[22] SHNEIDERMAN, B., PLAISANT, C.: *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (4th Edition). Pearson Addison Wesley (2004).

[23] SUNKEL, M., KAUTZ, J., Seidel, H. P.: Rendering and Simulation of Liquid Foams In: *Proceedings of the Vision, Modelling and Visualization*, (2004), pp. 285-294.

[24] UKROP, Ľ.: Visualizing graph structures using soap bubbles. In: *IIT.SRC: 6th Proceedings in Informatics and Information Technologies*, STU

in Bratislava FIIT, vol. 2, pp 467-474, 2010.
ISBN 978-80-227-3266-3.

Assoc. Prof. Dipl. Eng. Martin ŠPERKA, PhD.
Faculty of Informatics
Bratislava School of Law
Tematínska 10
851 05 Bratislava
Slovak Republic
E-mail: martin.sperka@uninova.sk

MSc. Peter KAPEC, PhD. student
Faculty of Informatics and Information Technologies
Slovak University of Technology
Ilkovičova 3
842 16 Bratislava
Slovak Republic
E-mail: kapec@fiit.stuba.sk