From Artificial Neural Networks to Emotion Machines with Marvin Minsky

Jozef Kelemen

Institute of Computer Science, Silesian University, Opava, Czech Republic,
and VSM College of Management, Bratislava, Slovak Republic
kelemen@fpf.slu.cz or jkeleme@vsm.sk

Abstract: In 2007, one among the founders and internationally most recognized leading pioneers of the field of Artificial Intelligence and Cognitive Science, professor emeritus at Massachusetts Institute of Technology, Marvin Minsky celebrates his 80th anniversary. Exploiting thnis opportunity, the article overviews his contribution to the above mentioned fields (and might be also to some others), and sketches an (obviously incomplet) picture of the history of Artificial Intelligence and related disciplines, and a specific way of doing science, too.

Keywords: artificial intelligence, cognitive science, artificial neural network, theory of computation, emotion, frame representation of knowledge, society theory of mind, robot

1 Introduction on Two Anniversaries

In August 9, 2007, one among the founder and one among the internationally most recognized leading pioneer of (and professional in) the field of Artificial Intelligence and Cognitive Science, MIT professor emeritus Marvin Lee Minsky celebrated his 80th birthday. This is a good opportunity to overview his professional contribution to the above mentioned fields (and might be also some others), and to sketch a particular picture of the history of the disciplines he hbave contributed, esp. of the field of artificial neural networks, of Artificial Intelligence, and of Cognitive Science.

Another relevant reason to focus to Minsky's professional activities consists in the fact, that Artificial Intelligence (AI) celebrated in the year 2006 its 50th anniversary. That anniversary is closely related to Marvin Minsky's innitiatives a half century ago. AI is a discipline initialized thanks to a 1955 initiatives of a small group of enthusiastic young professionalsformed by Marvin Minsky, Claude Shannon, the founder of the modern quantitative theory of information, John McCarthy, the author of the famous programming language LISP, the *lingua franca* of advanced AI experimentations, and Nathalien Rochester.

Figure 1
Marvin Minsky at the Department of Cybernetics, Faculty of Electro-Engineering, Czech Technical University, Prague, June 2004 (photo by J. Kelemen)

The group wrote a letter to the *Rockefeller Foundation* in the summer 1955 – see (McCarthy et al., 2006) – applying for a support of a two month ten men summer study of artificial intelligence …*which will proceed on the basis of the conjecture that every aspect of learning or any other features of intelligence can in principle be so precisely described that a machine can be made to simulate it* (cited from the application letter).

The application resulted in the possibility to organize a two months research project (a summer school) at the Dartmouth College, Hanover, NH, in the period June 19-August 16, 1956.

During the summer school, the first, and up to that time isolated, research results and projects focused to study the possibilities of computers to simulate human level intelligence have been presented to a larger community of interested professionals, e.g. to (and by, as well) Herbert A. Simon, Allen Newell, John von Neumann, Oliver Selfridge, Marvin Minsky, John McCrathy and other important persons of the history of development of cybernetics, computer science, Artificial Intelligence, and Cognitive Science of the second half of 20th century.

Because of that, the 1956 *Dartmouth Summer School on Artificial Intelligence* is now recognized as the first professional meeting of the field of AI, and the starting-point of the development of AI as a well-constituted professional discipline, as a branch of scientific research which celebrated its golden anniversary past year.

2 Artificial Neural Networks

Minsky's interests in Artificial Intelligence started with study of (artificial) naural networks. As a PhD student in mathematics at Princeton University, and then as a fresh postdoc at Harvard, under the strong influence of works of Warren McCulloch and Walter A. Pitts concerning the cybernetic conceptual (mathematical) models of neuronal cells and the neural networks; see e.g. (McCulloch, Pitts, 1943), Minsky – in 1949 – started with an attempt to construct a working artificial neural network. He names the device *SNARC* (an abbreviation for *Stochastic Neuro-Analog Reinforcement Computer*). *SNARC* was a hardwired elecrotechnical device consisting of several identical artificial neurons connected in a network which was able to learn simple concepts (Minsky, 1952), (Minsky, 1954). It has been completed in 1951.

After the publication of the first monograph by Frank Rosenblatt (Rosenblatt, 1961) devoted to artificial neural networks, and under the rather unbounded Rosenblatt's optimism concerning the learning and recognition capabilities of the neural networks he proposed, Minsky and his collaborator from (and co-director of) the MIT Artificial Intelligence Laboratory, Seymour Papert, published the monograph (Minsky, Papert, 1969).

Minsky's and Papert's monograph reflected the authors' criticism of the Rosenblatt's concept of neural networks. It contains some mathematically proved theorems concerning some bounds of the learning and recognizing capacities of networks proposed by Rosenblatt. The criticism was based on the fact, that in certain cases it is impossible to learn or to recognize holistic concepts on the base of results of local observations. A simple illustration perhaps provides an intuitive picture of the core of troubles.

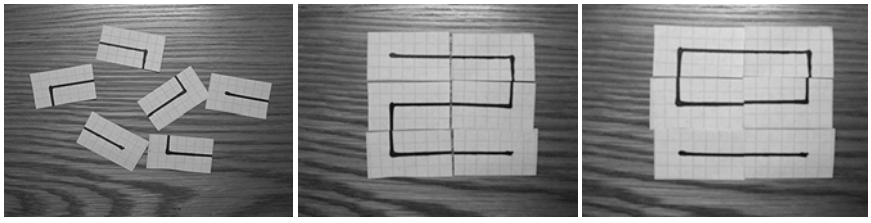


Figure 2
Two principally different 'global' drawings constructed from the same 'local' pieces; taked from (Kelemen, 2004)

Fig. 2 illustrates that on the base of the 'local observations' (represented by the isolated drawings in the picture on left), the drawings from the second and the third picture cannot be recognize. It is obvious, that the picture in left hand side contains the same 'local information' in both of cases of observing the parts of the second or of the third picture locally.

Minsky's and Papert's main contribution to better understanding of the recognition power of artificial neural networks consists in discovering of the *importance of the representation of whole concepts through local propositions*. This discovery then influenced Marvin Minsky's next scientific activities and professional views in many important ways.

Perhaps the first example of that influence is Minsky's relation and innovative views to computation and to abstract computating machines.

3 Finite and Infinite Machines

In 1967 Minsky published his influential and internationally recognized university text-book (Minsky, 1967).

The book starts with explaining finite-state machines as simple neural networks, so as devices constructed from large number of some small variety of simple computing units, as a distributed, multi-processor devices, in principle. Then Minsky continues to discuss the realism of the traditional Turing Machine (Turing, 1936) as a suitable formal model of computation and computers. He criticizes some of the more or less unrealistic features of the Turing Machines, and replaces the traditional thinking on computation as the function of the Turing Machine by another concept of the machinery, based on more realistic architecture.

The formal model of the Turing Machine memory in the form of a potentially infinite tape is in (Minsky, 1967) replaced, according his inspiration by the construction of the real hardwired (and because of that technical embodiment necessarily finite) computers working with partitioned memories, by a machine working with a finite number of so-called *registers*. The *Register Machine* is proposed as an alternative model of computing device, and it is proved its equivalent with the traditional Turing Machine with respect its computational power. In other words: What is computable by a Turing Machine is computable also by a Register Machine, and *vice versa*. The registers have infinite capacity each, but for performing the computation, only the question whether a register is empty or not is important! So, the Register Machine seems to be much more realistic, finite in certain sense, device, than the traditional Turing Machine which must deal with the infinity of its tape.

The Register Machine is a living concepts in the theoretical computer science up to now, and it makes many of proofs concerning the universality of the differently motivated computing devices in many cases much simpler; an example of such proof can be found e.g. in (Csuhaj-Varjú et al., 2006).

Another influence of the Minsky's conviction in *the power of well-mutable connections of simple parts* has been reflected in the architectural principle used

by Daniel Hillis, a post-doc at MIT AI Lab in that time, in the design of the famous *Connection Machine* (Hillis, 1985).

The first model CM 1 has been completed in 1986 as the collection of 65 536 simple one-bite processors organized in the form of a 12 dimensional hyper-cube with the memory bandwidth of 300 Gigabits per second. Different models of the Connection Machine have then been produced then by the Thinking Machines Corporation, Inc. founded by D. W. Hillis, M. Minsky, and others in 1986, in Cambridge, MA.

4 Knowledge Representation

The Minsky's conviction in the *importance of the representation of whole concepts through local proposition* is evidently present in his concept of the so-called *frame representation of knowledge*, a very influential knowledge representation concept in AI starting with its proposition by Minsky in June 1974 (in the form of the MIT AI Laboratory Memo No. 306). The first regularly published version of the proposal appeared then as (Minsky, 1975), and then many times in different wordings as a part of different edited books.

The core of the concept of *frame representation of knowledge* consists in integration of the positive sides of the three previously developed representational schemes.

The historically first, the *declarative* representational scheme, is based mainly on the idea of adaptation of traditional formalisms developed in the field of the formal logic for AI purposes. The representations are logical formulas, and the problem-solving procedures are based on automation of theorem-proving.

The second scheme, the *procedural* one, was inspired by the experiences in computer programming. It is focused to developing advanced programming environments enabling the end-users to solve problems by their declarative specification only, without the necessity of programming any procedural parts of the problem-solving procedures. Programming tools – like the Prolog programming languages family, for instance – are good examples of the influence of this representational ideology.

The inspiration for the third representational scheme, the *associative* one, originated in the circles of linguists converted to AI research because of their conviction that AI research is the best way to make computers able to understand natural languages and use them for communication with human beings and for mechanizing the translation of texts from one natural language into another one.

The problems concerning mechanization of translation leads immediately to the recognition of the importance of contexts in which human beings interprets natural

language texts in order to understand (and translate) them, so to the problems concerning the semantic meaning of the syntactic units forming the texts.

The traditional approaches to deal with meaning of formulas used in formal logic, as well as the philosophical conceptions of extensionality of meaning was recognized as not suitable enough, and the interest of linguists has been shifted to search for approaches which emphasize the intensionality. This attempt resulted in formalism of associative (often called also semantic) networks.

The renaissance of the approach is in present time recognizable in researches oriented to specification of web ontologies, for instance; see e.g. (Fensel, 2004).

Minsky in his approach to knowledge representation found a productive way of how to combine the positive sides of the previously developed representational schemes, and how to enlarge the resulting representational structure by some other practically required and psychologically reasonable properties, like representation of *expectations* of events and situation (as so called *default values* in some parts of the frame representational structures) and ways to implement the possibility of the so-called *non-monotone reasoning* recognized by psychologists in human problem-solving activities but not reflected in traditional formal logics up to the invention of frames (then the non-monotone reasoning became a part of research in formal logic, esp. thanks to initiatives of John McCarthy).

The influence of the idea of frames as representational structures to the practical programming is evident in the so-called object-oriented programming, and in object-oriented programming tools like C++ or Java, for instance, where the simplified concept of frames is present as the concept of objects; see e.g. (Coad, Nicola, 1993) for details concerning the object-oriented programming.

5 Architecture of Mind

Another consequence of Minsky's conviction in *the power of well-mutable connections of simple parts* is his concept of the mind as a society of interrelated and mutable society of simpler units called agents.

Agents in Minsky's understanding are organizational units with the ability to sense and act in their environments created by other agents, so be active in a multi-agent system as its parts, and create larger structures having the same behavioral properties as the agents have. Starting with the simplest agents having the power of simple computational devices Minsky proposes the way of the evolution of much more complicated societies of agents with the power comparable with the power of human minds.

This concept – the *society theory of mind* – is perhaps the most important Minsky's contribution to the development of cognitive science.

The first description of the structure of these agencies has been proposed in (Minsky, 1980), a paper focused to the problem of specific form of functioning of the human memory. Then human memory, in the contrast with the memories of artificial computing devices, enables reconstruct memorized events or objects incompletely, stressing only the contextually relevant (important) parts of the happened events or observed objects, without all of previously sensed but in the moment of re-construction of the passed events or sensed object contextually unimportant details.

The concept of mind as a society of interrelated developing system of agents is in the form of a monograph presented in (Minsky, 1986), and then it was massively propagated using different media and narratives. It has been explained in the form of an interactive multi-medial CD in (Minsky, 1996), In the context of a thrilling sci-fi story it was presented in the form of a techno-thriller in (Harrison, Minsky, 1992). In (Singh, 2003) the development and the influence of the theory is recapitulated with the complete professional rigor.

Roughly speaking, the importance of the society theory of mind is twofold. First, it consists in emphasizing the concept of agents in cognitive science as a useful and productive concept for explaining of the phenomenon of mind and some of the mind's capacities. Second, the concept of organized societies of agents became inspiring, broadly applicable, and also very trendy, during the 90ties of the past century. It happened in the field of practical applications of computing technology as one of the basic concept for dealing with computing machineries as specific types of multi-agent systems; see e.g. (D'Inverno, Luck, 2004).

5 Emotion Machines

Elaborating his society theory of mind Minsky discovered that there are no principal difference between the architecture of agents and agencies required by the rational side of intelligence, and that one required by the emotional one. The basic emotional states are results of activities of some simple agents sensing the chemical properties of an organism and evoking basic emotional states like hunger and thirst, or the concentrations of different hormones in the organism, and evoking emotional states of fear or wrath. Realizing this universality of the society theory of mind Minsky started at the beginning of 90^{ties} of the past century to work on his theory of emotions. The work resulted in the monograph (Minsky, 2006).

In (Minsky. 2006) emotions are treated in their abstract forms, so in the form, which does not emphasize and make no principal difference between the emotionality of human beings or machines. This view of emotionality is enabled thanks to the quit general view of the mind as a society of interrelated agents, and not as a specific phenomenological concept attributed exclusively to human beings.

The multi-agent (or resource instead of agent, as Minsky proposed instead of agents in his book) picture of the mind enables to formulate a specific approach to the functions of brains (human ones, but also artificial ones like computers, or advanced computer controlled robots). Minsky imagined brains as things which developed different ways to maintain parallel representations in multiple contexts, 'realms' as Minsky writes, so that each can apply its own resources to different aspects of the same word or topic. He names this multi-agent, or multi-resource, problem solving by a neologism *panalogy* (from *parallel analogy*).

For example, in the case described by the sentence 'John gave Mary a book' we recognize minimally three different interpretation (understanding) realms:

- the physical realm, where 'give' refers to book's motion through the physical space,
- the social realm, when 'give' refers to a specific social act, and
- the 'dominion realm', when if we heard that 'John gave Mary a book' we inferred that Mary then be holding the book and has gained permission to use it.

So, according Minsky's opinion, we have things well-represented if we have represented them in several realms at once, including the emotional realm(s). If we are watching a horror in a cinema for instance, we understood the movie not only in physical, social, common-sense etc. realms, but in certain moments necessarily also in realms of the surprise, fear, disgust, etc. The 'global' understanding of the movie emerges in certain sense from all these partial understandings, and this emerged understanding is the complete and psychically normal one.

7 Anothe Anniversary – Robots

As we have mentioned above, Marvin Minsky has contributed also to the development of the advanced cognitive robotics. Not only as the co-director of the MIT AI Laboratory, where one among the firsts projects of development of this kind of robots have been started during the end of 60ties of the past century, but also in the role of coordinator of research projects, and supervisor of many PhD projects oriented to different particular problems (e.g. in natural language understanding, manipulation with real objects, vision, hand-eye coordination, learning form experience, etc.) connected with the cognitive robots construction. Then, step-by-step he approached the cognition and emotions from a point of view from which the difference between the cognitive and emotional phenomena in biologically and culturally evolved human beings and in artificially engineered machines have no importance.

Minsky did that all as a rigorous scientist, however as a scientist with an almost unbounded ability to dream and intellectual creativity. Namely these his abilities connect him with other pioneers of robotics, and provide an opportunity for us to mention, after mentioning his anniversary, and the anniversary of AI, also a third anniversary, the anniversary of the birth of the word *robot*.

Figure 3
Karel and Josef Čapek. A photo from the twenties of the 20th century

In 2006, at least the world of theatre celebrated the 85th anniversary of the moment, when robots appeared first at the stage of the Prague's National Theatre, and started their successful march through numerous branches of culture, science and technology of the 20th century. The first night of the theatre play R. U. R. (Rossum's Universal Robots) by Karel Čapek, the play which evoked many of the appearing and appealing problems with numerous social and political dimensions of the co-existence of the human culture and the industrial production, of the co-existence of human and engineered, was in January 25, 1921.

This is also the date when the world *robot* – invented by Karel's brother Josef Čapek – was first used for denoting artificially created human-like beings. Perhaps it is not without interest that the play has been written during the summer mounts of 1920 in the Slovak spa Trenčianske Teplice, and that the neologism of the Slavonic origin, the word 'robot', has been proposed in the same place. However, this topic goes slightly out of the subject of present contribution. More about the cultural destiny of the word robot, and the cultural history of hunam-like creatures and human beings co-existence with such creatures see in (Horáková, 2006) or (Horáková, Kelemen, 2006).

Acknowledgment

This article is based on the author's talk delivered at the 9th International Conference *Informatics 2007* (Bratislava, Slovakia, June 21-22, 2007), and published in the conference proceedings as (Kelemen, 2007). The author's research is continously supported by the Gratex International, Inc., Bratislava, Slovakia.

References

[1] Coad, P., Nicola, J.: *Object-Oriented Programming*. Yourdon Press, Upper Saddle River, NJ, 1993

[2] Csuhaj-Varjú, E., Kelemen, J., Kelemenová, A., Paun, Gh., Vaszil, Gy.: Computing with Cells in Environment – P Colonies. *Journal of Multiple Valued Logic and Soft Computing* **12** (2006) 201-215

[3] D'Inverno, M., Luck, M.: *Understanding Agent Systems*. Springer, Berlin, 2004

[4] Fensel, D.: *Ontologies – a Silver Bullet for Knowledge Management and Electronic Commerce*. Springer, Berlin, 2004

[5] Harrison, N., Minsky, M.: *The Turing Option*. Warner Books, New York, 1992

[6] Hillis, W. D.: *The Connection Machine*. The MIT Press, Cambridge, MA, 1985

[7] Horáková, J.: Staging Robots – Cyborg-Culture as a Context for Robots Emancipation. In: *Cybernetics and Systems, Proc. EMCSR 06, Vol. 1* (R.

Trappl, Ed.). Austrian Society for Cybernetic Studies, Vienna, 2006, pp. 312-317

[8] Horáková, J., Kelemen, J.: From Golem to Cyborg – a Note on the Cultural Evolution of the Concept of Robots. *Human Affairs* **16** (2006) 83-98

[9] Kelemen, J.: The Minsky Way (From Artificial Neural Networks to Emotion Machines). In: *Proc. 9th International Conference Informatics 2007* (Plander, I., Ed.). Slovak Society for Applied Cybernetics and Informatics, Bratislava, 2007, pp. 131-138

[10] Kelemen, J.: Od umělých neuronů ke strojům s emocemi. *Computerworld* **15** (Czech Edition) No. 9 (2004) 24-25

[11] Lúčny, A.: Building Complex Systems with Agent-Space Architecture. *Computing and Informatics* **23** (2004) 1-36

[12] McCarthy, J., Minsky, M. L., Rochester, N., Shannon, C. E.: A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *The AI Magazine* **27** (Winter 2006) 12-14

[13] McCulloch, W., Pitts, W. A.: A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bull. Math. Biophys*. **5** (1943) 115-133

[14] Minsky, M.: *The Emotion Machine*. Simon & Schuster, New York, 2006

[15] Minsky, M.: *The Society of Mind* (Interactive CD Edition). The Voyager Company, Boston, MA, 1996

[16] Minsky, M.: *The Society of Mind*. Simon & Schuster, New York, 1986

[17] Minsky, M.: K-Lines – a Theory of Memory. *Cognitive Science* **4** (1980) 117-133

[18] Minsky, M.: A Framework for Representing Knowledge. In: *The Psychology of Computer Vision* (P. H. Winston, Ed.). McGraw-Hill, New York, 1975, pp. 211-277

[19] Minsky, M.: *Computation – Finite and Infinite Machines*. Prentice Hall, Englewood Cliffs, NJ, 1967

[20] Minsky, M.: *Theory of Neural-Analog Reinforcement System and its Applications to the Brain-Model Problem* (Ph. D. Thesis). University of Princeton, Princeton, NJ, 1954

[21] Minsky, M.: *A Neural-Analogue Calculator Based upon a Probability Model of Reinforcement*. Psychological Laboratories, Harvard University, Cambridge, MA, 1952

[22] Minsky, M., Papert, S.: *Perceptrons – An Introduction to Computational Geometry*. The MIT Press, Cambridge, MA., 1969

[23] Rosenblatt, F.: *Principles of Neurodynamics*. Spartan Books, Washington, D. C., 1961

[24] Singh, P.: Examining the Society of Mind. *Computing and Informatics* **22** (2003) 521-543

[25] Turing, A. M.: On Computable Numbers, with an Application to the Entscheidungsproblem. *Proc. London Math. Soc.* **42** (1936) 230-265; corrections **43** (1937) 544-546

Construction of Functions by Fuzzy Operators

József Dombi

Árpád tér 2, H-6720 Szeged, Hungary, e-mail: dombi@inf.u-szeged.hu

József Dániel Dombi

Árpád tér 2, H-6720 Szeged, Hungary, e-mail: dombijd@inf.u-szeged.hu

Abstract: In this paper we present a new approach to compose and decompose functions. This technology is based on pliant concept. We use the proper transformations of Conjunction of Sigmoid function to create an effect. We aggregate the effects to compose the function. This tool is also capable for function decomposition.

Keywords: Pliant concept, Sigmoid, Conjunction, Aggregation, Dombi operator

1 Introduction

Numerical analysis is the area of mathematics and computer science that creates, analyzes and implements algorithms for solving numerical problems of continuous mathematics. One of the main subfields of numerical analysis is interpolation. Interpolation is a method, where we determine a function (in most cases it is a polinom) which fits to the given data points and using this result we can determine the function value if new data points are given. In science in most cases we obtain samplings to determine the relations between inputs and outputs, which is called curve fitting, because usually we do not require the exact fit, only the approximation. We can say interpolation is a specific case of curve fitting, which case the function must go exactly through data points. There are lots of interpolations for example: linear, polynomial, spline and trignometric. Usually we fit data using a polynominal function of the form

$$y(x,w) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j \tag{1}$$

where M is the order of the polynomial, and x^j denotes x raised to the power of j. The polynomial coefficients $w_0,\ldots w_M$ are collectively denoted by the vector $\mathbf{w}$.

Note that, although the polynomial function y(x,**w**) is a nonlinear function of x, it is a linear function of the coefficients **w**. The values of the coefficients will be determined by fitting the polynomial to the training points. Curve fitting can be done by minimizing the error function that measures the misfit between the function for any given value of w and the data points. One simple and widely used error function is the sum of the squares of the errors between the predictions $y(x_n,w)$ for each data point x_n and the corresponding target values t_n, so that we minimize the energy function:

$$E(w) = \frac{1}{2}\sum (y(x_n, w) - t_n)^2 \tag{2}$$

We simply note that it is a nonnegative quantity that would be zero if and only if the function y(x,**w**) were to pass exactly through each training data point. The geometrical interpretation of the sum-of-squares error function is illustrated in Figure 1.

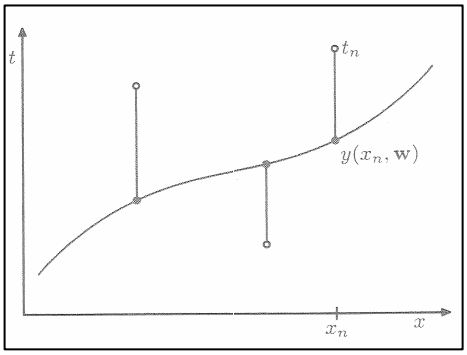


Figure 1
The error function

We can solve the curve fitting problem by choosing the value w for which E(w) is as small as possible. The main problem is how we choose the order *M* of the polynom and as we shall see this will turn out to the problem of model comparison or model selection. In Figure 2, we show four examples fitting polynomials having orders M = 0, 1, 3, and 9 to the data set, which sample getting from a sinus function with noise.

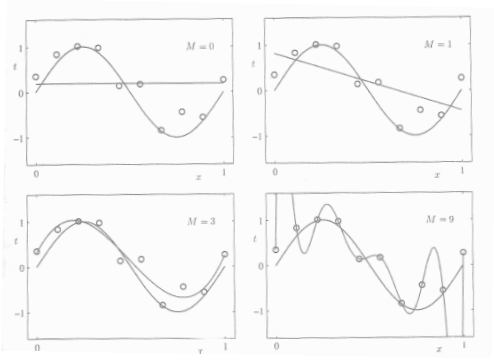


Figure 2
Plots of polynomials having various orders *M*

The problems of the curve fitting and interpolation are that we do not understand the meaning of the parameters. In our approach we drop the traditional concept. We define effects as a basic units, which have meaning. In Section 2 we define the basic units of the effects and we show how to create effect. In Section 3 we define how to compose and decompose functions. At the end we present some recent results.

2 Operations

In this section we present how to compose effect.

2.1 Sigmoid Function

A sigmoid function is a mathematical function that produces a sigmoid curve. We defined it by the following equation:

$$\sigma_a^{\lambda}(x) = \frac{1}{1+e^{-\lambda(x-a)}} \tag{3}$$

where λ is the sharpness of the function and *a* is the offset. It is easy to see that

- $\sigma_a^{\lambda}(a) = \frac{1}{2}$,
- $\sigma_a^{-\lambda}(x) = 1 - \sigma_a^{\lambda}(x)$.

2.2 Conjunction and Dombi Operator

To construct the effect we use fuzzy operators and sigmoid function. It is important to consider that the fuzzy operator and Dombi operator fits well. The operator is based on the solution of associativ function equation. The associative function equation is the following:

$$c(x,y) = f^{-1}\big(f(x) + f(y)\big) \tag{4}$$

This function $c:[0,1]\times[0,1]\to[0,1]$, which is coninuous and satisfied the following properties:

- Commutativity: $c(a,b) = c(b,a)$
- Monotonicity: $c(a,b) \le c(c,d)$, if $a \le c$ and $b \le d$
- Associativity: $c(a,c(b,c)) = c(c(a,b),c)$
- 1 is an identity element: $c(a,1) = a$

We get the Dombi operator by choosing:

$$f(x) = \frac{1-x}{x}, f^{-1}(x) = \frac{1}{1+x} \tag{5}$$

If we use the generator function in the conjunction case we get the following operator:

$$c(x,y) = \frac{1}{1+\dfrac{1-x}{x}+\dfrac{1-y}{y}} \tag{6}$$

We can extend this with power and weight:

$$c_\lambda(x,y,u,v) = \frac{1}{1+\left(u\left(\dfrac{1-x}{x}\right)^\lambda + v\left(\dfrac{1-y}{y}\right)^\lambda\right)^{\frac{1}{\lambda}}},\ u+v=1 \tag{7}$$

In this case we lose the associativity property.

2.3 Aggregation

Aggregation is not a logical operator. The form of an aggregation operator is:

$$a(x_1,..x_n) = \frac{1}{1+\frac{1-\upsilon_0}{\upsilon_0}\frac{\upsilon}{1-\upsilon}\prod_{i=0}^{n}\frac{1-x_i}{x_i}}, \tag{8}$$

where ν is the neutral value and ν_0 is the threshold value of the corresponding negation.

The aggregation operator satisfies the following properties:

- Defined on [0,1] interval: $\bigcup[0,1]^n \to [0,1]$,
- Identity when unary: $a(x) = x$, if $\upsilon_0 = \upsilon = \frac{1}{2}$,
- Boundary Conditions: $a(0,...,0) = 0$ and $a(1,...,1) = 1$,
- Non decreasing: $a(x_1,...,x_n) \le a(y_1,...,y_n)$ if $(x_1,...,x_n) \le (y_1,...,y_n)$.
- Selfduality: $a(n(x),n(y)) = n(a(x,y))$

3 Function Composition and Decomposition

We deal a function, which is defined on [0,1] interval. In order to working in [0,1] interval we have to transform the real function to [0,1]. This can be done by using a linear function or using the Sigmoid function.

To achieve our objective we will use Sigmoid function, Dombi operator, Conjunction operator and Aggregation operator. First we need natural effect, so we use two Sigmoid (3) function, and the conjunction operator (5):

$$eff_{a,b}^{\lambda_1,\lambda_2}(x) = \frac{1}{1+e^{-\lambda_1(x-a)}+e^{-\lambda_2(x-b)}} \tag{9}$$

In Figure 3 you can see this effect. We define positive effect, which means that this effect will increase the function value and we define negative effect, which is the opposite of the positive effect (ie. decrease the function value). We define the positve effect on [1/2,1] and negative effect on [0,1/2]. The neutral value is 1/2. To achieve this we need the proper transformation of (9):

$$P(x) = \frac{1}{2}(1 + eff_{a,b}^{\lambda_1,\lambda_2}(x)) \tag{10}$$

$$N(x) = \frac{1}{2}(1 - eff_{a,b}^{\lambda_1,\lambda_2}(x)) \tag{11}$$

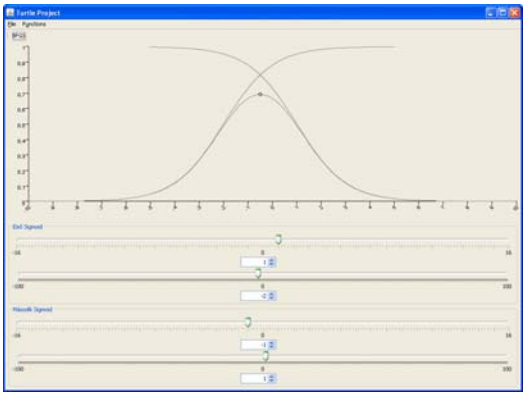


Figure 3
Conjunction of two Sigmoid function

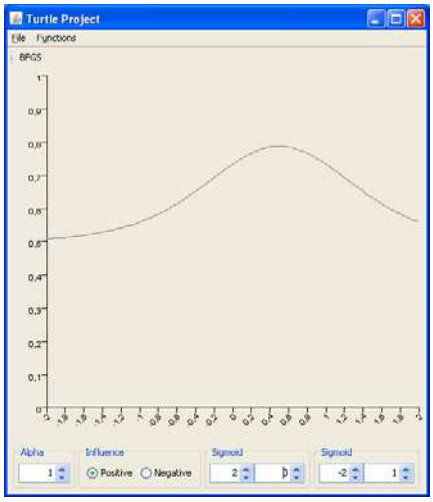


Figure 4
Positive Effect

In Figure 4 you can see that this function has four parameters. These parameters have semantic meanings: *a* is the starting point of the effect, *b* is the end point of the effect and lambdas are the sharpness of the effect.

The last step of function creations is to aggregate these effects as you can see in Figure 5.

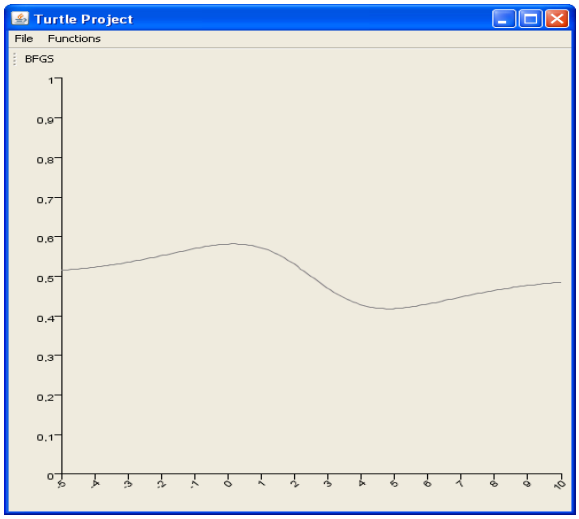


Figure 5
Aggregation of effects

In function decomposition we know the values. Our task is to divide real values into effects. It is an optimization problem. The critical point of the optimization is the initial values of the effects. If we define good initial points we have a high chance to find the global optimum.

In our case it is not impossible. We can define the number of the effect and each effect we can determine the following properties:

- Start/end point of the effect
- Sharpness of the effect.

4 Result

We create a Java program to create and analyze our approach. In our system we can create any arbitrary funcion as you can see in Figure 6.

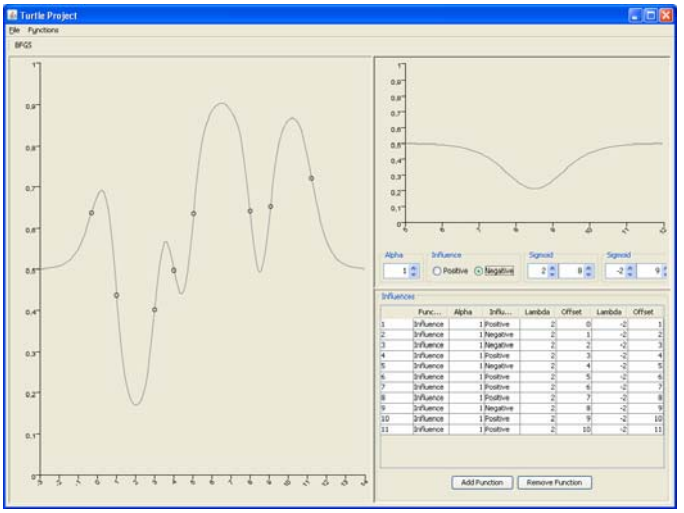


Figure 6
Our Java system

The global optimization is under development in our Java system, but we tried in Matlab and the solution calculates very fast the parameters and the error of the result is acceptable.

Conclusions

In this paper we develop a new tool, that is useful for interpolation. Our approach has good properties. Our model is based on fuzzy logic. We create effects and aggregate them. Our method is also capable to decompose a function. We are working on an global optimization algorithm that is working in our System.

References

[1] Dombi J.: A General Class of Fuzzy Operators, the De Morgan Class of Fuzzy Operators and Fuzziness Measures Induced by Fuzzy Operators, Fuzzy Sets and Systems, Vol. 8, 1982, pp. 149-163

[2] Dombi J.: Basic Concept for a Theory of Evaluation: the Aggregation Operator, Europian Journal of Operations Research, Vol. 10, 1982, pp. 282-293

[3] Detyniecki M.: Fundamentals on Aggregation Operators, asturias, 2001

Sensor-net for Monitoring Vital Parameters of Vehicle Drivers

Gergely Fördős, István Bosznai, Levente Kovács, Balázs Benyó, Zoltán Benyó

Department of Control Engineering and Information Technology,
Faculty of Electrical Engineering and Informatics,
Budapest University of Technology and Economics
Magyar tudósok krt. 2, H-1117 Budapest, Hungary
{fordos, bosznai, lkovacs, bbenyo, benyo}@iit.bme.hu

Abstract: Improving the safety of the traffic is a social interest. Accidents are not only caused by poor technical conditions of the vehicles, but also by tired, indisposed, or bad state-of-minded drivers. The managing of human factors needs the control, recording and monitoring of the most important vital parameters of the driver. The paper presents such a system, which is based on ECG recording and needs no or little cooperation of the driver.

Keywords: phiysiological signal measurement, on-line monitoring, vehicle on-board data collection, ECG, pulseoximetry, bluetooth communication

1 Introduction

In 2005, the Budapest University of Technology and Economics (BUTE) has started a interdisciplinary university-level grant, RET-04/2004, who's aim is to collect a know-how for the automotive industry. Among the 19 subprojects involved, the Control Engineering and Information Technology department's Biomedical Engineering research group coordinates an individual subproject titled RET 5.2 'Human factor in controlled vehicle systems'.

The aim of this subproject is dual: on the one hand to investigate the possibility of creating a sensor-net, which is suitable for recording the driver's vital parameters; on the other hand to develop a human model, which deals with the data collected by the sensors. This paper focuses on the first objective of the project.

Improving the safety of the traffic is a social interest. Accidents are not only caused by poor technical conditions of the vehicles, but also by tired, indisposed, or bad state-of-minded drivers. As a result it can be concluded that the safety and efficiency of today's road traffic over the technical and technological environment

is also highly dependent on human behavior. Road users play an essential role in regulating road traffic system. It is possible to influence driver behavior by structural changes in traffic control strategies, road design and vehicle characteristics. Specifications for intelligent driver support systems such as navigation and collision avoidance systems can be developed on the basis of knowledge about traffic participants' reactions.

The managing of human factors needs the control, recording and monitoring of the most important vital parameters of the driver. The most important human factors are the driver characteristics, driver health, mood states and personality factors, driver fatigue, driver decision-making, inattention /distraction and hazard perception.

During the first year of the RET 5.2 project an extensive review of technical literature related to the scope of this research has been carried out. The aim was to explore the stafe-of-art methods – or even ready-to-use systems – which are feasible for monitoring physiological parameters of drivers.

In the elaborated literature review, [1], we have structured the driver observing methods in the following categories:

1. Eye movements, PERCLOS (percentage of eyelid closure), tracking of gaze, EOG (elektro-okulogram);
2. Brain activity, EEG;
3. ECG, HRV (Heart Rate Variation);
4. Facial muscle activity, facial tracking;
5. Lane-related measures (the relative position of the car to the lane border);
6. Heading/lateral acceleration related measures;
7. Combined methods.

We have reviewed the publications in detail, [1], and we have rate them by five aspects:

- Realization: the results were obtained in real conditions or in simulated ones;
- Modeling: was there a model created or only a statistical analyze was made;
- Number of examined physiological parameters;
- Effectiveness of the presented (or created) system;
- Advantages and disadvantages of the article (from our point of view).

From the first three aspects we have summarized the results of the investigated articles and we have concluded the followings:

- It isn't still confirmed that a sensor network is created or researched which monitors the driver physiological parameters (in the sense what we have formulated as our aim for the 5.2 project);
- It is very useful to monitor the drivers physiological parameters from the safe driving point of view;
- It is worthy to start a research activity which investigates the connection between the traffic conditions and drivers physiological parameters.

According to these, two methods has been selected for our project, which involve measurement of two biological signals, [1], [2]:

- Drivers' ECG measurement;
- Oxygen saturation measurements.

The advantages of them, compared to other methods are: no or little cooperation of the driver is needed, they are unobtrusive and compared to image-processing methods require only low bandwidth and less computational power.

Consequently, our conception was to build a system based on signals obtained by the blood-pressure and heart rate measuring devices attached to the driver.

2 The Created Sensor Network

The measurement of the selected biological parameters described above is part of the experimental monitoring system. Figure 1 depicts a sketchy idea of the planned system. Recorded biological signals and supplementary information is sent over a Controller Area Network, which is a standard communication means in the automotive industry, as it is shown on the sketch. A state-of-art vehicle is equipped with one or more independent CAN[1] networks. Electronic components of the engine, transmission, chassis and brake are also connected via CAN. A dedicated network is used to connect components like lights, AC, central locks, seat and mirror adjustment. We can assume that if we want to establish communication between electronic components it is straightforward to use CAN bus; either to utilize one that is already available, or setup a dedicated channel.

[1] Controller Area Network is a broadcast, differential serial bus standard developed by Robert Bosch GmbH, for connecting electronic control units (ECUs).

2.1 The ECG Measurement Device

At the beginning we tried to use existing ECG devices form the market, like the CARD(X)PLORE ECG device from the Meditech Ltd. (which is a combination of holter ECG and ambulatory blood pressure monitor). However, we have realized that this and in generally the ECG devices which are on the market are not capable working on very noisy environments: driver's ECG signals not always could be interpreted (ex sudden movements).

As a result, we have developed a private ECG device to be able to eliminate the mentioned problems, [3], [4].

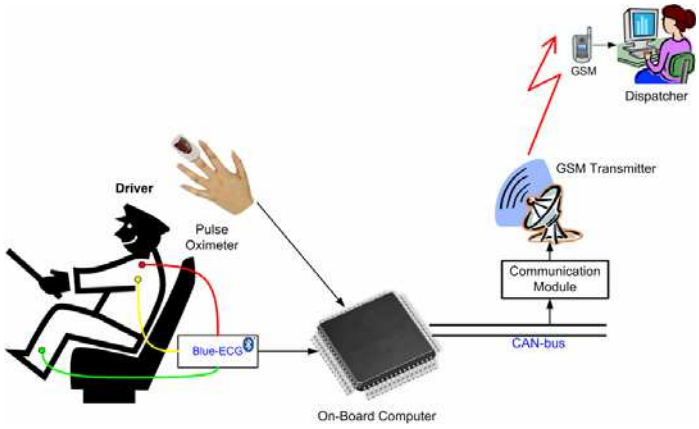


Figure 1

System for on-line monitoring of driver's key vital parameters

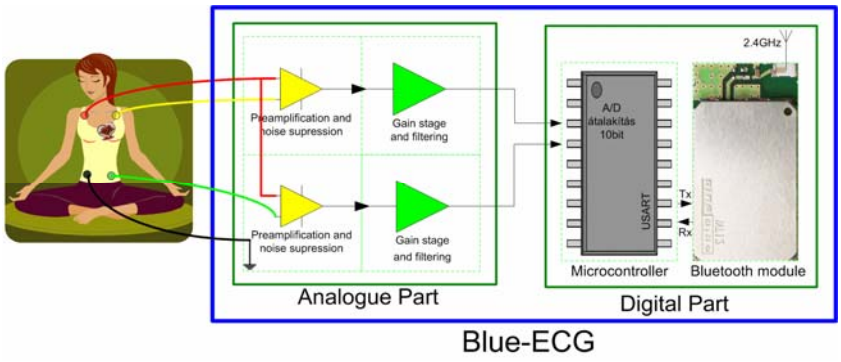


Figure 2

Schematic figure of the developed ECG device

The created system has two parts: one circuit is for amplifying, transforming and transmitting wireless the signal, while the other circuit is to collect and process the data. The schematic figure of the device is presented on Figure 2.

A modern ECG units must have a big common signal rejection ratio. Therefore we have calculated what is the necessary common-mode rejection ratio (CMRR) of the input amplifier, if we suppose that a 0.1 mA current goes through the driver and a maximum 10μV disturbance can be tolerated, while the resistance of the driver is around 1kΩ, [5], [6].

$$\begin{aligned} V_{CM} &\approx 0.1mA \cdot 1k\Omega = 100mV \\ V_Z{}^{\max} &= 10\mu V \\ CMRR &= \frac{V_{CM}}{V_Z{}^{\max}} = 10000 = 80dB \end{aligned} \tag{1}$$

Our requirements with the amplifier block were:

- At least 80dB CMRR, as mentioned;
- Working on single supply 3.3V;
- Max. 45 mA input current;
- +1.65V average output ECG;
- Working on telemetric principles;
- Even in case of 3m distance a sufficiently quick signal transfer;
- Capable of working continuously, min. 24 hours on battery;
- Small size.

In this way the ECG device can be structured on four main parts and three additional ones.

The main parts are:

- Preamplifier;
- High pass filter and output amplifier;
- Microcontroller;
- Bluetooth transmitter.

The additional parts are:

- Low drop linear stabilizer;
- Power voltage divider;
- High precision reference.

Briefly summarizing, the circuit is amplifying ten times the 1 mV ECG signal, then the high pass filter is cutting the DC component. After this, the output amplifier is amplifying 45 times the 10 mV signal, and the low pass filter is filtering the signals bigger then 106 Hz. The signal which is in this way amplified and filtered, is processed by the microcontroller, which is sampling the signal with 500 Hz, and based on serial protocol transmits it to the Bluetooth device and finally to the receiver. Consequently, the frequency interval is between 0.05 Hz-106 Hz, and the amplification is 450 (Figure 3).

The circuit was developed at the BME-ETT department, [4]. Figure 4 presents this circuit, where for better identification we have labeled the corresponding parts. The Bluetooth device is on the other part of the circuit, so it can't be seen on the figure.

The meaning of the different colors of electrodes used (Figure 6) are:

- Red: Right hand;
- Yellow: Left hand;
- Black: Right leg;
- Green: Left leg;

During the measurements we have used the standard Einthoven lead configuration presented on Figure 6 The electrodes were placed on the body of the driver.

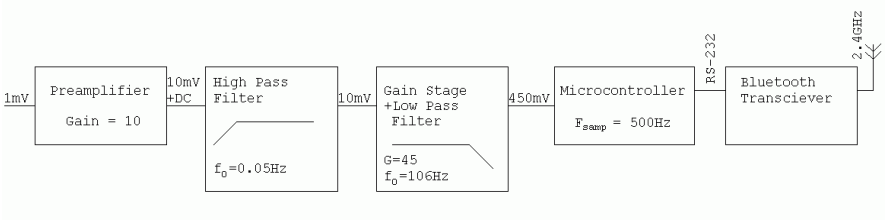


Figure 3

The block diagram of the system

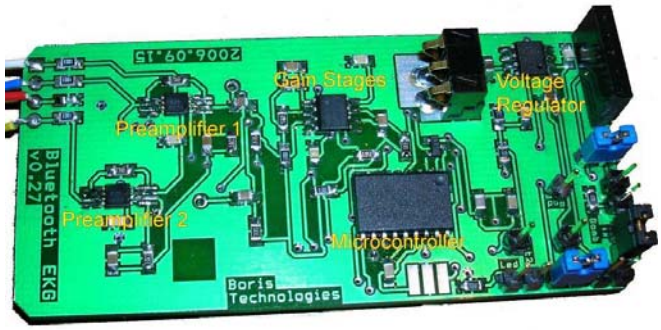


Figure 4

The developed ECG circuit

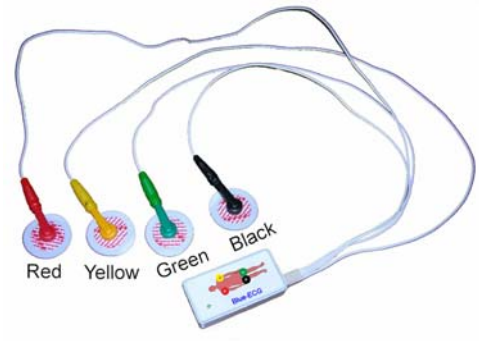


Figure 5

The final BLUE-ECG device

Figure 6

The presentation of the used convention

The two amplifier parts are measuring the signal of Einthoven I and II parts and and calculating the III[2], aVR, aVF and aVL[3] parts.

We have used 3M electrodes, which are so-called second-type electrodes, with metal core closed round with hardly soluble metal acid (like Ag/AgCl). Its advantage is that is not polarizing.

2.2 Brief Description of the Used Pulse Oximeter

The used pulse oximeter is a Nonin Avant 4100 type one (Figure 7) and it was described in the 2007 RET annual report, [3] Pulse oximetry provides estimates of arterial oxyhemoglobin saturation (SaO_2) by utilizing selected wavelengths of light to noninvasively determine the saturation of oxyhemoglobin (SpO_2). There is a tight relationship between oxygen saturation and brain blood perfusion, respiration and heart rate, [7]. The sensor additionally provides heart rate data of 1% relative precision, which can be utilized for heart rate variability analysis – an important measure described later on.

The selected sensor is a portable device and it comes with a finger clip sensor. Measurement data is sent to a computer over Bluetooth wireless protocol, which carries out necessary preprocessing operations. Oxygen saturation and pulse rate data is delivered once per a second. Sensor fault, low perfusion and other artifacts can be detected using the status word which is assigned to each data packet by the sensor.

Although, the pulse oximeter was used in our measurements all the time together with the ECG device, the measured data by it will be used only at the investigation of the second aim of the project (estimating the state of driver's driving capability). As a result in the followings we will restrict only on the ECG measurements.

3 The Procession of the Measurements

As a result, we have concluded that categorizing the driver safety states is necessarily arbitrary. Probably the most plausible category could be if the driver is or not capable to drive the vehicle (of course intermediary categories can be defined). Another division can be: awake, sleepy, insensible (due to an illness), drunk (probably also others).

[2] Leads I, II and III are the so-called limb leads because the subjects of electrocardiography had to literally place their arms and legs in buckets of salt water.

[3] Leads aVR, aVL, and aVF are augmented limb leads. They are derived from the same three electrodes as leads I, II, III.

By our assumption (supported by the big number of the collected and investigated articles) these driver states and the transition between them could be estimated by real-time monitoring of relevant physiological and other (ex. steering wheel angle) parameters.

The measurements were done in a car equipped with the BLUE-ECG device. The investigated traffic situations were:

- Normal urban traffic;
- Urban traffic jam;
- Driving on highway.

Figure 7

Nonin Avant 4100 pulse oximeter

The Einthoven I and II signals were both measured and they were transmitted to a notebook, where the data was processed and recorded. The driving conditions taken into account were:

- Normal driving (straightaway);
- Light truning and shifting;
- Truning

Our aim was to test our ECG device and to see if it is capable of working in these noisy situations. The data was recorded by a software also developed by us in .Net under C#. The main window of the software is presented by Figure 8, [3].

During the measurements, the previous step of starting to use the software is to set the serial port (and if it was necessary the type of the filter). The upper window of the main panel shows on-line the filtered ECG signal, while the lower one the unfiltered one (Figure 8). The filtered signal is needed to get the QRS complex from the ECG singal. The QRS signal represents the heartbeat, this way we could calculate the heartrate, and finally the HRV. A 10-16 Hz band-pass filter was used to filter the noise caused by body movements, the irrelevant parts of the ECG, and the 50 Hz line interference, [8].

The program automatically saves the presented data in a TXT file what is important for us, because the data will be processed and evaluated under MATLAB (for our second aim of the project).

On each measurement we have recorded at least 30 minutes driving in one of the above mentioned driving situations (Figure 9, Figure 10, Figure 11). It can be seen, that even in a sudden movement situation, the recorded ECG signal is interpretable, especially in the situation of determining the R-R[4] distances.

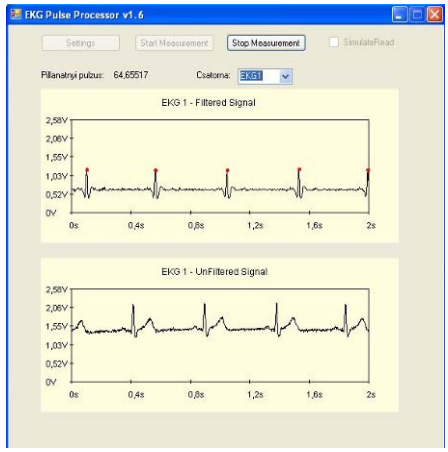


Figure 8

The main window of the ECG Pulse processor software

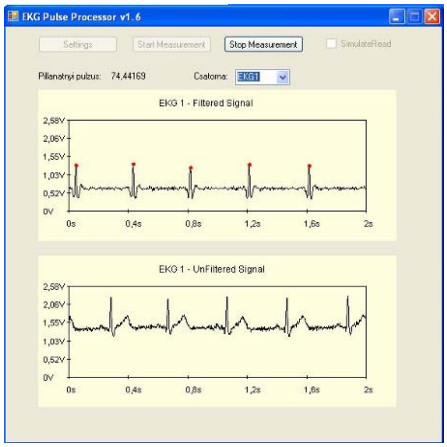


Figure 9

ECG measurements during straightaway driving

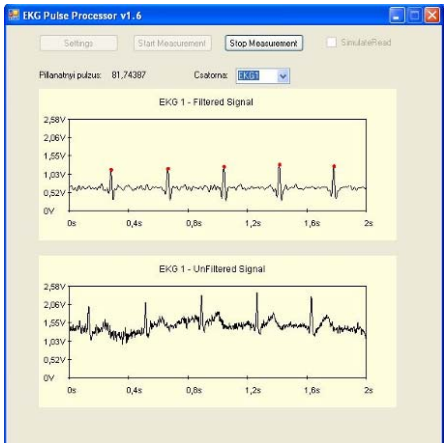


Figure 10

ECG measurements during small movements of the driver in the car

[4] R peak: appears, when the heart conrtacts, R-R distance: used to calculate the heart-rate

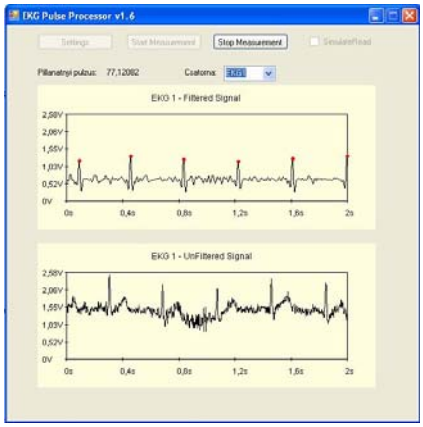


Figure 11

ECG measurements during sudden driving movements

Conclusions and Future Steps

We have realised a sensor-net, capable to record driver's vital parameters in noisy environments. Measurements demonstrated, that the net composed by an ECG device and a pulse oximeter can be efficiently used for our purposes. However, one of our future aims is to improve the sensor net in the sense of a wireless connection and transmission of the measured signals (e.g. no direct electrodes on the driver body).

Regarding the problem of estimating driver's driving capability based on the measured vital parameters and the developed sensor network, we will analyze drivers' stressors. Connected to this, we are planning to develop a classifying algorithm which is able to classify on-line the driving capability.

Finally, based on our results we are planning also to create a hypothetic model describing the interaction between the driver, his car and the surrounding environment. For this we will collaborate with our project's consortium partners: BUTE Department of Transport Automation (Faculty of Transportation Engineering) and BUTE Department of Road Vehicles (Faculty of Transportation Engineering).

Acknowledgement

This work was supported in part by Hungarian National Scientific Research Foundation, Grants No. OTKA T69055, F046726 and by Hungarian National Office for Research and Technology (RET-04/2004).

References

[1] Benyó Z., L. Kovács, L. Török, A. Reiss, B. Benyó: Human Factor in Controlled Vehicle Systems, 2005 annual report, 2005

[2] Benyó Z., L. Kovács, L. Török, A. Reiss, B. Benyó, L. Szilágyi, G. Fördős: Human Factor in Controlled Vehicle Systems, 2006 annual report, 2006

[3] Benyó Z., L. Kovács, G. Fördős, I. Bosznai, B. Benyó: 5.2 Human Factor in Controlled Vehicle Systems, 2007 annual report, 2007

[4] Bosznai I.: Designing and Implementing a Wireless Holter ECG device (in Hungarian), MSc thesis, Budapest University of Technology and Economics, 2007, consultants: Dr. Hunor Sátha, Norbert Stubán

[5] Raimo Sepponen: Biopotential Amplifiers - Medical instrumentation, 2000, http://sel18.hut.fi/166/Bioamplifiers-2005.pdf

[6] Prutchi D., M. Norris: Design and Development of Medical Electronic Instrumentation: A Practical Perspective of the Design, Construction, and Test of Medical Devices, Wiley, 2004

[7] Takatani S, P. W. Cheung, E. A. Ernst: Noninvasive Tissue Reflectance Oximeter: An Instrument for Tissue Hemoglobin Oxygen Saturation in Vivo. Annals of Biomedical Engineering, Vol. 8, pp. 1-15, 1980

[8] Jobbágy Á.: Biomedical Instrumentation I-II (in Hungarian), lecture notes, 1998

3D Image Sensor based on Parallax Motion

Barna Reskó[1,2], Dávid Herbay[3], Péter Krasznai[1], Péter Korondi[3]

[1] Budapest Tech
[2] Computer and Automation Research Institute of the Hungarian Academy of Sciences
[3] Budapest University of Technology and Economics
E-mail: rbarna@datatrans.hu

Abstract: For humans and visual animals vision it is the primary and the most sophisticated perceptual modality to get information about the surrounding world. Depth perception is a part of vision allowing to accurately determine the distance to an object which makes it an important visual task. Humans have two eyes with overlapping visual fields that enable stereo vision and thus space perception. Some birds however do not have overlapping visual fields, and compensate this lask by moving their heads, which in turn makes space perception possible using the motion parallax as a visual cue. This paper presents a solution using an opto-mechanical filter that was inspired by the way birds observe their environment. The filtering is done using two different approaches:using motion blur during motion parallax, and using the optical flow algorithm. The two methods have different advantages and drawbacks, which will be discussed in the paper. The proposed system can be used in robotics for 3D space perception.

Keywords: depth perception, parallax motion, rotating persicope, motion blur, optical flow

1 Introduction

Most herbivores, especially hoofed grazers, lack depth perception. Instead, they have their eyes on the side of the head, providing a panoramic, almost 360º view of the horizon - enabling them to notice the approach of predators from any direction. However both avian and mammalian predators have frontal eyes, allowing them to precisely judge distances when they pounce, or swoop down, onto their prey.

In modern terminology, stereopsis is depth perception from binocular vision through exploitation of the two slightly different projections of the world on the two retinas called parallax. Depth perception does indeed rely primarily on binocular vision, but it also uses many other monocular cues to form the final integrated perception. There are monocular cues that would be significant to a

'one-eyed' person, and more complex inferred cues, that require both eyes to be perceiving stereo while the monocular cues are noted. This 'third' group relies on processing within the brain of the person, as they see a full field of view with both eyes.

Stereopsis has been one of the most widely explored topics of 3D vision. A classical stereo system typically includes two cameras placed side by side to capture stereo image pairs. The depth information of the captured scene is calculated from the disparity map of these two images. Instead of using two or more cameras, one can also sequentially capture image pairs by repositioning a single camera. The advantage of using a single camera over the two-camera systems is that the identical intensity response of the stereo pairs captured with the same camera can improve the accuracy of correspondence matching.

Besides the approach of displacing a single camera, several other single-camera systems have also been investigated. Each of them has its own unique features. Adelson and Wang designed a plenoptic camera system which uses a lenticular array placed in front of the image plane to control the light path, and the depth information is extracted by analyzing a set of sub-images [5]. Lee and Kweon presented a bi-prism stereo camera system which forms the stereo pairs on the left and right halves of a single CCD sensor by a bi-prism [3]. Nene and Nayar investigated stereo imaging using different types of mirrors including planar, ellipsoidal, hyperboloidal and paraboloidal mirror [6]. The subject of this paper is a rotating periscope which is placed in the axis of the camera. By rotating the periscope in a vertical plane around the camera's optical axis, the image of the scene follows a circular shift on the vertical plane. In the presented experiment, two images corresponding to two fixed poses of the plate are captured. Feature points are extracted by looking for zero-crossings on images filtered by LOG (Laplacian of Gaussian) filter. In the proposed idea the depth information is extracted from the comparison of a sharp and a blurred image.

This paper is organized as follows: in Section 2 the proposed concept is described. This is followed by the computer based simulations of the concept in Section 3. Section 4 introduces the opto-mechanical filter and other hardware implementation. The paper is closed by the test results and conclusions about the results and future perspectives.

2 The Proposed Concept of 3D Perception

2.1 Depth Perception with Single Camera

Although herbivores have sacrificed the depth perception for a much larger view range with eyes on the sides of the head, they are able to judge approximately distances in another way. Actually it is realized by the parallax method as in case of predators, but in this case it is a parallax motion, so the slightly shifted points of view are realized by the motion of the head.

A well know example is the cock, which moves his head forward and backward synchronously with its stepping.

Another example in the nature which seems similar to the method presented in this paper is the method of the ostrich, which has a special method for the depth perception. From time to time it does a circular motion with its head in a vertical plan with the aim of getting better depth information on the great plain. The longer is the baseline (the distance of points of view), the more precise depth information the bird can perceive. This allows to judge the distance in a greater range.

2.2 Two Ways to Calculate Depth from Parallax Motion

Two different approaches are proposed for depth perception, both of them based on the motion parallax visual cue.

2.2.1 Using Motion Blur

There are several techniques for stereo vision with 2 or more cameras. The drawback of these techniques is the cost due to the multiple of high quality cameras, the sophisticated programming of image processing and consequently a big hardware background. Several problems turn up in these techniques: opposition of image quality (resolution, colour depth and frame per second) and data transfer and processing ability. The motion blur is proposed, which can be obtained by relatively moving the image sensor and the projected image during exposure. This effect will be used to obtain depth information.

2.2.2 Using Optical Flow

Another way of calculating the depth is to calculate the optical flow between two consecutive frames taken from the scene. The length of the optical flow vectors will be inverse proportional to the distance of the projected 3D point.

2.3 The Geometry of Stereo Vision in Case of a Two-Camera System

Consider the images *p* and *p'* of a point *P* observed by two cameras with optical centres *O* and *O'*. These five points all belong to the *epipolar plane* defined by the two intersecting rays *OP* and *O'P*. In particular, the point *p'* lies on the line *l',* where this plane and the retina P*'* of the second camera intersect. The line *l'* is the *epipolar line* associated with the point *p*, and it passes through the point *e'* where the baseline joining the optical centres *O* and *O'* intersects P'. Likewise, the point *p* lies on the epipolar line *l* associated with the point *p'*, and this line passes through the intersection *e* of the *baseline* with the plane P (Figure 1).

The points e and *e'* are called the *epipoles* of the two cameras. The epipole *e'* is the (virtual) image of the optical centre *O* of the first camera in the image observed by the second camera, and vice versa. As noted before, if *p* and *p'* are images of the same point, then *p'* must lie on the epipolar line associated with *p*. This *epipolar constraint* plays a fundamental role in stereo vision and motion analysis. Since the most challenging task in stereo vision is finding the correspondence between points on the two images, using the epipolar constraint for a stereo camera setup shrinks the search area for a point *p* being the pair of point *p'* from a plane to a line.

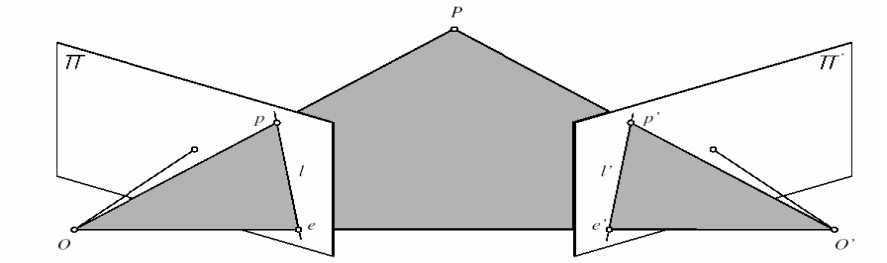


Figure 1

In the previous section it was shown that point P lies on the lines defined by Op and O'p'.

If the intrinsic and extrinsic parameters of the cameras are known, the above two lines determine the three-dimensional position of point P. The retrieval of three-dimensional information from two or more pieces of two-dimensional information is referred to as 3D reconstruction.

The reconstruction is realized by matrix transformations. There are several methods in which a common coordinate system is defined, in that coordinate system projection planes and the transformation matrix is to be defined for the 3D reconstruction.

But these are the vast calculations for a high resolution, great fps video.

The idea was inspired to get round this difficult image process by processing in an analog way.

2.4 Motion Blur

It is the apparent streaking of rapidly moving objects in a still image. When a camera creates an image, that image does not represent a single instant of time. Because of technological constraints or artistic requirements, the image represents the scene over a period of time. As objects in a scene move, an image of that scene must represent an integration of all positions of those objects, as well as the camera's viewpoint, over the period of exposure determined by the shutter speed. In such an image, any object moving with respect to the camera, will look blurred or smeared along the direction of relative motion. This smearing may occur on an object that is moving or on a static background if the camera is moving.

The mathematical form of the integration of the light captured on the sensor during the motion blur (Figure 2):

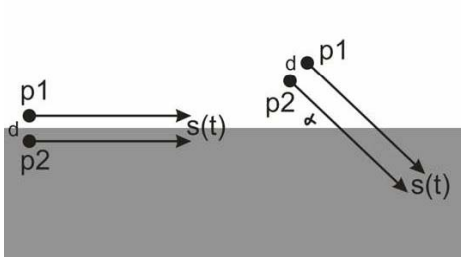


Figure 2

$\left.\begin{matrix} s_x(t) \\ s_y(t) \end{matrix}\right\} s(t)$ is the path of the point observed on the sensor

$m(p) = \int_0^{t_m} I(p + s(t))dt$, equation of the image of horizontal motion blur of any point is the time integration the light intensity on time with the motion of the point

$m_\alpha(p) = \int_0^{t_m} I(p + s_\alpha(t))dt$, the same with an inclination α

The equation of the whole image is the integration of the equation 3 on the entire surface:

$$M_\alpha(x_c, y_c) = \int_{p \in A_{x_c, y_c}} m_\alpha(p) dp$$

This way by using motion blur effect an analog integration of points can be gained without using any computer.

Figure 3

We can see also that the edges are reinforced which are oriented to the direction of the motion blur and those are blurred which are oriented in other direction. So we can reinforce the contour of object in a certain orientation with a straight motion (Figure 3). But if we want to reinforce the contour in all directions then we have to apply a circular motion.

This leads to the final idea, that is by moving one camera on a circular path creates a blurred image with reinforced contours from which spatial information can be extracted based on that consideration that the closer the object is the more it is blurred, and so the further the object is the less it is blurred. It means the more intense the contour is the closer the object is.

It is already a technical consideration that instead of moving a camera, it is advisable to use an optical device creating a turning image. The most suitable such a device is the periscope. It shifts the image parallel as much as it is designed. Fortunately it does not rotate the image as it turns around just shift the view point and describe a circular path. In an idealistic case it does not cause even any image distortion.

An image may contain edges of different strengths. Applying the circular blur on the images may result in similar edge strength values, if the blurred edges were not similar and they were in a different distance. This explains why the resulting blurred image has to be compared to a reference image containing the non-blurred edge values (Figure 4).

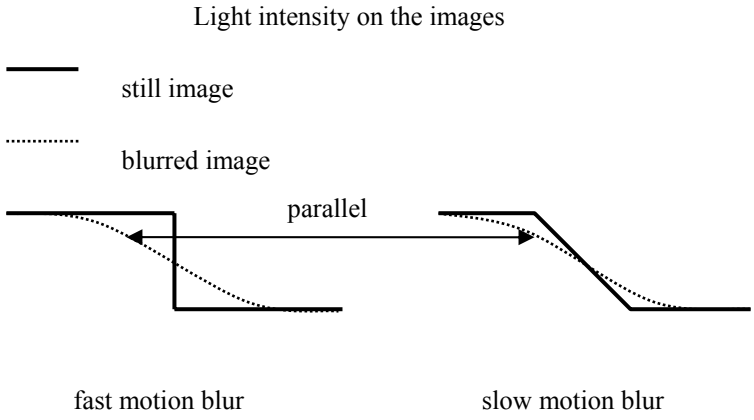


Figure 4

2.5 Optical Flow

Optical flow is the velocity field which warps one image into another. This can be described by a vector field, which contains vectors representing the relative position of corresponding pixels (features).

Estimating the optical flow is useful in pattern recognition, computer vision, and other image processing applications. It is closely related to motion estimation and motion compensation. Often the term optical flow is used to describe a dense motion field with vectors at each pixel, as opposed to motion estimation or compensation which uses vectors for blocks of pixels, as in video compression methods such as MPEG.

2.5.1 Optical Flow Estimation by Lucas-Kanade Method

The Lucas-Kande method for optical flow is described in [8]. The problem statement of optical flow is as follows: let A and B be two 2D greyscale images. The two quantities A(x) =A(x; y) and B(x) =B(x; y) are then the greyscale values of the two images at the location $x = [x,y]^T$, where x and y are the two pixel coordinates of a generic image point x. The images A and B are discrete functions (or arrays), and the upper left corner pixel coordinate vector is $[0,0]^T$. Let w and h be the width and height of the two images. Then the lower right pixel coordinate vector is $[w-1,h-1]^T$.

Consider an image point $u = [ux, uy]^T$ on the image A. The goal of feature tracking is to find the location $v = u+d = [ux +dx\ uy +dy]^T$ on the image B. The vector $d = [dx, dy]^T$ is the image velocity at x, also known as the optical flow at x.

Because of the aperture problem, it is essential to define the notion of similarity in a 2D neighbourhood sense. Let winx and winy be two integers.

We define the image velocity d as being the vector that minimizes the residual function defined as follows:

EQUATION 1

$$\varepsilon(dx,dy) = \sum_{x=ux-winx}^{ux+winx} \sum_{y=uy-winy}^{uy+winy} \left(A(x,y) - B(x+dx, y+dy)\right)^2$$

The similarity function is measured on a image neighbourhood of size (2winx +1)X(2winy +1). This neighbourhood will be also called integration window.

Minimising ε with least squares method

At the optimum, the first derivative of ε(dx,dy) with respect to d=(dx,dy) is zero:

EQUATION 2

$$\frac{\partial \varepsilon(d)}{\partial d} = [0,0]$$

After expansion of derivate:

EQUATION 3

$$\frac{\partial \varepsilon(d)}{\partial d} = 2 \sum_{x=ux-winx}^{ux+winx} \sum_{uy-winy}^{uy+winy} (A(x,y) - B(x+dx, y+dy)) \begin{bmatrix} \frac{\partial B}{\partial x} & \frac{\partial B}{\partial y} \end{bmatrix}$$

Let now substitute B(x+dx,y+dy) by its first order Taylor expansion about the point $d = [0,0]^T$.

EQUATION 4

$$\frac{\partial \varepsilon(d)}{\partial d} \approx 2 \sum_{x=ux-winx}^{ux+winx} \sum_{uy-winy}^{uy+winy} (A(x,y) - B(x,y) - \begin{bmatrix} \frac{\partial B}{\partial x} & \frac{\partial B}{\partial y} \end{bmatrix} d) \begin{bmatrix} \frac{\partial B}{\partial x} & \frac{\partial B}{\partial y} \end{bmatrix}$$

The meaning of A(x,y)-B(x,y) can be interpreted as the temporal image derivative at the point $[x\ y]^T$:

EQUATION 5

$$\delta I = A(x,y) - B(x,y)$$

The matrix $\begin{bmatrix} \frac{\partial B}{\partial x} & \frac{\partial B}{\partial y} \end{bmatrix}$ is the image gradient vector. Now make a new notation:

EQUATION 6

$$\nabla I = \begin{bmatrix} Ix \\ Iy \end{bmatrix} = \begin{bmatrix} \frac{\partial B}{\partial x} & \frac{\partial B}{\partial y} \end{bmatrix}$$

If we use central difference operator to make derivation, we can express Ix, Iy:

EQUATION 7

$$Ix(x,y) = \frac{B(x+1,y) - B(x-1,y)}{2}$$

EQUATION 8

$$Iy(x,y) = \frac{B(x,y+1) - B(x,y+1)}{2}$$

Following this new notation, the equation can be written:

EQUATION 9

$$\frac{1}{2}\frac{\partial \varepsilon(d)}{\partial d} = \sum_{x=ux-winx}^{ux+winx} \sum_{uy-winy}^{uy+winy} (\delta I - \nabla I d)\nabla I$$

EQUATION 10

$$\frac{1}{2}\frac{\partial \varepsilon(d)}{\partial d} = \sum_{x=ux-winx}^{ux+winx} \sum_{uy-winy}^{uy+winy} \left(\begin{bmatrix} \delta IIx \\ \delta IIy \end{bmatrix} - \begin{bmatrix} Ix^2 & IxIy \\ IyIx & Iy^2 \end{bmatrix} d \right)$$

Make new notations:

EQUATION 11

$$G = \sum_{x=ux-winx}^{ux+winx} \sum_{uy-winy}^{uy+winy} \begin{bmatrix} Ix^2 & IxIy \\ IyIx & Iy^2 \end{bmatrix} \quad and \quad b = \sum_{x=ux-winx}^{ux+winx} \sum_{uy-winy}^{uy+winy} \begin{bmatrix} \delta IIx \\ \delta IIy \end{bmatrix}$$

Then the equation can be written:

EQUATION 12

$$\frac{1}{2}\frac{\partial \varepsilon(d)}{\partial d} = b - Gd \qquad and\ at\ optimum \qquad d = -G^{-1}b$$

This equation can be solved if G is invertible. This means that the gradient is large enough under point u, which is the corresponding position on the other image we are looking for.

Problems with this approximation

This method has two key assumptions:

- **colour constancy:** a point in frame A looks the same in B (for greyscale images, this is brightness constancy)
- **small motion**: points do not move very far

By using first order Taylor series we can track only small motion between frames. By using image gradients we are assuming, that corresponding pixels have the same intensity.

3 Computer Simulations of Motion Blur-based Depth Perception

In the first step with the purpose of ensuring these theories, a computer simulation was elaborated.

With a Matlab based program, two different distances were simulated by applying two different motion radiuses on a given image.

After applying the Sobel operator and comparing the changes in intensity, the following results were obtained.

Figure 5

The Figure 5 shows the still, reference image of the original test object.

The first motion blur was made on a circle of a radius 10 units (Figure 7), the second on a radius of 30 units (Figure 6) which are equivalent with two different distances from the camera.

On the right of the Figures 6 and 7 the index numbers show the difference of the light intensity changes.

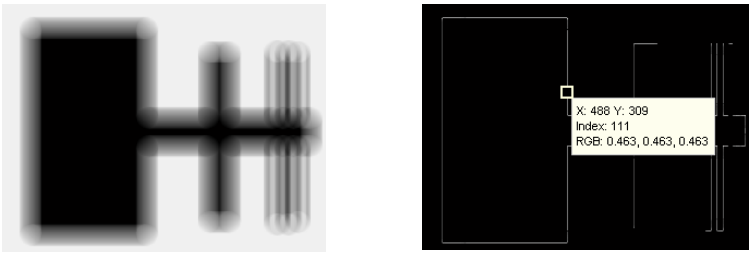


Figure 6

After checking several point pairs on the two results, we got that the index number is always larger in the second case. This means that the proposed method is able to compare distances. We can say that the theory is proved by the simulation.

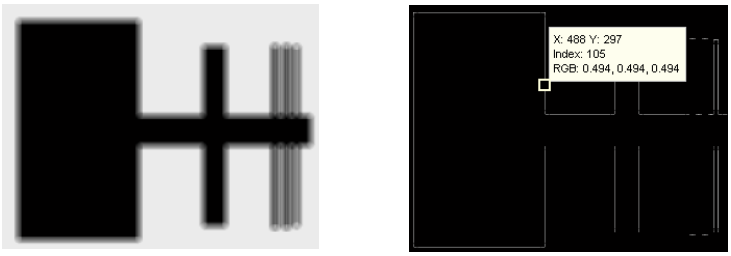


Figure 7

4 The Rotating Periscope

A mechanical frame was designed to implement the rotating periscope concept. During the design a compromise had to be made between the lengh of the periscope and the angular width of the field of view. The result was a 20 degrees symmetrical field of view and a periscope length 68 millimeters. This length has been chosen because it coinsides the average distance between the eyes of a human.

The camera is positioned in front of the periscope, which is mounted using a large diameter bearing around the optical axix of the camera. This allows the periscope to be rotated using an external motor, and also reduces vibration and any other mechanical noise.

4 mm thick mirrors have been used, which have a very low distortion and bending radius. The mirrors have to be parallel with a very low error causing a disalignment of the image less than one pixel.

The drawings and rendered view of the periscope are shown in Figure 8, while the implemented

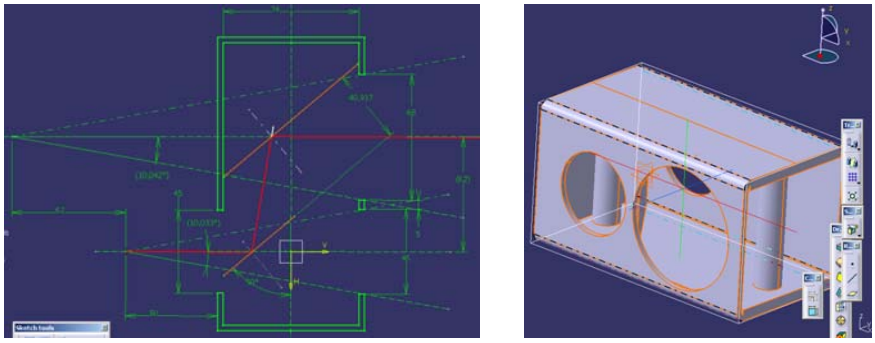


Figure 8

Figure 9

5 Testing and Evaluation

5.1 The Motion Blur-based Method

To set up the test, it is only needed to give the supply voltage for the motor and the camera and connect the camera to a PC by a fire-wire cable (Figure 10).

On the camera's lens the aperture and the focal distance can be adjusted. The user interface supplied for the camera the exposure time, the resolution, the colour depth, etc. are adjustable.

Figure 10

At the first experiments the mechanism worked well, but the image was distorted. The point of the test was that if the camera looks at an infinite distance (600 meters across the river) the image should not change at all. But the result was a fluctuating distortion of the image. It was obviously the consequence of the too low quality mirrors. This way the evaluation of the image is not possible because the blurred image is not a consequence of the relative motion.

Evidently the mirrors were to be changed.

After changing to 4 mm thick much more rigid mirrors the result was better, not distorted, but still not perfect. The test was the same as previously, a still image was expected to appear. The resulting image was however blurred again (Figure 11). It was realized that there was also a misalignment of the mirrors. If they are not perfectly parallel then the axis of the periscope does not coincide with the axis of the camera and the rotation axis, so the image is blurred proportionally to the misalignment, and not to the displacement of the view point.

Short estimation of the misalignment: the distance from the buildings on the other side of the river is about 700 m, the relative displacement of them is about 2 m, arctan (2/700)=0,1 degree. So the periscope is to be redesigned a bit to be able to regulate the inclination of the mirrors. Until the submission of this paper it was not worked out.

After designing and fabricating the periscope the theoretical conception was tested. It was realized that the concept is relatively very sensitive to the precision of the optical device but this precision requirement is still far from the precision of general optical device like for e.g. a handy cam. It can be stated that the image quality depends from the applied camera, and the image processing is faster than in case of the binocular stereo matching techniques.

Figure 11

If good test results will be obtained in the near future, then this concept has a great chance to be worked-out in a more compact and applicable form, and has a good chance for interest in the industrial applications, or for e.g. intelligent space applications and robotics.

5.2 The Optical Flow-based Method

The method using optical flow has been implemented in a C++ software. Figures 12-14 show the obtained results. The first three images were taken with a camera, below left shows the optical flow vectors, and on the right their length is encoded in color, blue meaning the far points, red meaning the close points.

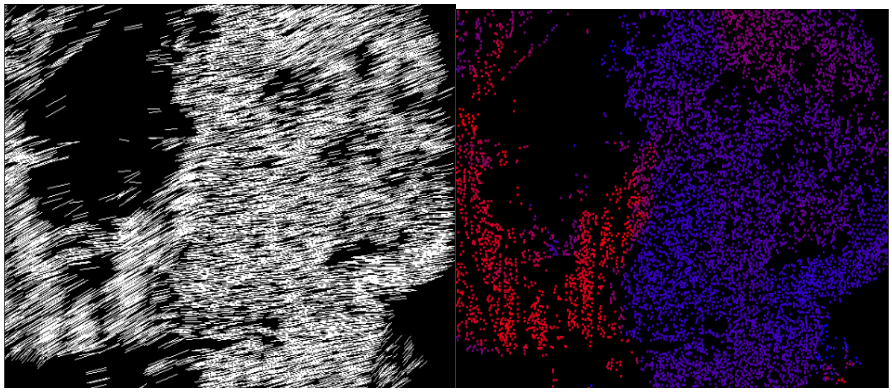


Figure 12

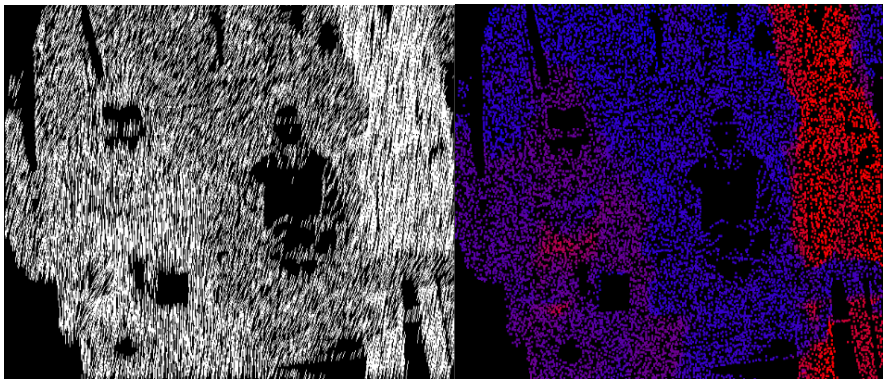


Figure 13

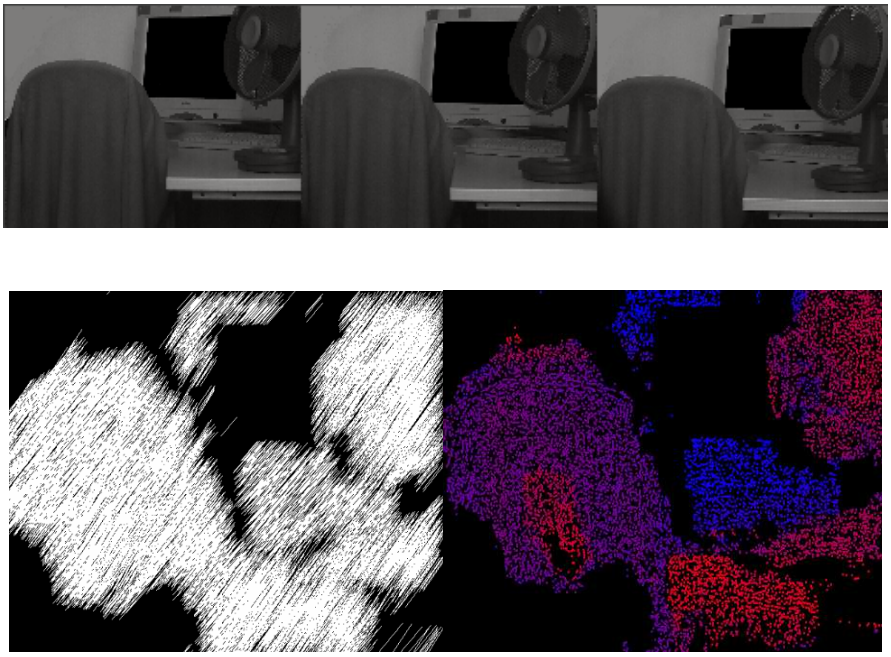


Figure 14

Conclusion

The obtained results show that the porposed methods are able to extract spatial information from motion parallax cues. The motion blur based method is computationally very unexpensive, but it is very sensible to the accuracy of the optical devices installed in the hardware equipment. The optical flow based method is on the other hand very robust to inaccuracies of the mirrors, but computationally more complex. This complexity is due to the optical flow algorithm. The applicability of the proposed method could be enhanced by using special hardware based optical flow calculation devices. Such devices have recently emerged on the market, and the authors intend to test their results using such hardware tools.

Acknowledgement

The authors wish to thank the National Science Research Fund (OTKA K62836) for their financial support, and the Rancz Ltd. for the support in manufacturing the periscope presented in this paper.

References

[1] SKF.com, http://www.skf.com/portal/skf/home

[2] Introduction to stereo vision system, http://www.cs.mcgill.ca/~gstamm/P644/stereo.html

[3] Doo Hyun Lee, In So Kweon, Roberto Cipolla, "A Biprism-Stereo Camera System," *cvpr*, p. 1082, 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'99) 1, 1999

[4] Stereovision, http://axon.physik.uni-bremen.de/research/stereo/

[5] Adelson, E. H., Wang, J. Y. A., "Single Lens Stereo with a Plenoptic Camera", PAMI,14(2), pp. 99-106, 1992

[6] S. A. Nene, S. K. Nayar. Stereo with Mirrors. In Proc. ICCV, pp. 1087-1094, 1998

[7] Barna Reskó. Optical Motion Tracking for Industrial Robot Programming, Master's thesis, 2004

[8] Lucas B D, Kanade: An Iterative Image Registration Technique with an Application to Stereo Vision

Computer-aided Analysis of Phisiological Systems

Balázs Benyó

Department of Control Engineering and Information Technology
Faculty of Electrical Engineering and Informatics
Budapest University of Technology and Economics
Magyar tudósok krt. 2, H-1117 Budapest Hungary
bbenyo@iit.bme.hu

Department of Informatics, Széchenyi István University
Egyetem tér 1, H-9026 Győr, Hungary
benyo@sze.hu

Abstract: This paper presents the recent biomedical engineering research activity of the Medical Informatics Laboratory at the Budapest University of Technology and Economics. The research projects are carried out in the fields as follows: Computer aided identification of physiological systems; Diabetic management and blood glucose control; Remote patient monitoring and diagnostic system; Automated system for analyzing cardiac ultrasound images; Single-channel hybrid ECG segmentation; Event recognition and state classification to detect brain ischemia by means of EEG signal processing; Detection of breathing disorders like apnea and hypopnea; Molecular biology studies with DNA-chips; Evaluation of the cry of normal hearing and hard of hearing infants.

Keywords: biomedical systems, physiological models, image analysis, signal processing, detection algorithms

1 Introduction

Education, research and development related to the biomedical engineering are the main activities of the Biomedical Engineering Laboratory at the Budapest University of Technology and Economics. Several research and development project has been carried out in the Lab during last fifteen years. The main research areas of the last few years are briefly presented in this article.

2 Computer-aided Identification of Physiological Systems

The computer analysis is a set of theoretical processes, allowing us to describe the behaviour of biological systems in terms of a mathematical model. Therefore, in medical-biological researches the computer analysis carries the same importance as that of the state equation in modern control theory [1].

2.1 Computer Evaluation of Measurement Data Using Compartment Analysis

The compartment model of clearance test is described generally by a linear differential equation system of the first order. Evaluation of the test consists essentially of determined parameters X_i and λ_i in Eq. (1). And then solving this system of differential equations.

$$y = \sum_{i=1}^{m} X_i e^{\lambda_i t} \tag{1}$$

where $\lambda_1=0$.

Thereafter, a curve of the sum of exponential terms has to be fitted to the set of measurement points bearing test results, and after having achieved and appropriate accuracy of a given degree, values of coefficients X_i and of exponents λ_i have to be printed out.

For the computer evaluation of clearance-type tests, the algorithm relying on the least squares principle, developed by D. W. Marquardt, proved excellent [2], [3].

Notation used in this chapter:

n	number of measurement data;
k	number of assessed parameters;
$t_1,\dots t_n$	sampling times (independent variables);
$y_1,\dots y_n$	measurement data at the time instants above;
$\hat{y}_1 \dots \hat{y}_n$	computed assessment values;
$b_{10},\dots b_{k0}$	initial values of the parameters to be assessed be given.

Let us determine parameter values $b_1,\dots,b_k$ for which the sum of the square of differences between *measured data* y_i and *computed assessments* $\hat{y}_i$ is minimum, that is, minimize.

$$\Phi = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{2}$$

To determine assessment, let us rewrite (1):

$$\hat{y}_i = b_1 + \sum_{l=2,4,\ldots}^{k-1} b_l \exp(b_{l+1} t_1) \tag{3}$$

To minimize (2) parameter vector b has to changed stepwise, according to some convenient strategy. After some experimentation we used an ***iterative computing method*** [3].

In conformity with this strategy, repeating iteration steps yields the minimum for Φ**.** Condition of stopping the iterative computation may be defined either by specifying a given value for Φ minimum, to stop after its achievement, or by observing the rate of conver-gence, and to accept the result after a certain slowing down in the rate of convergence.

Namely, the rate of convergence abruptly drops after starting with very high value.

2.2 Evaluation of Liver Flow Tests

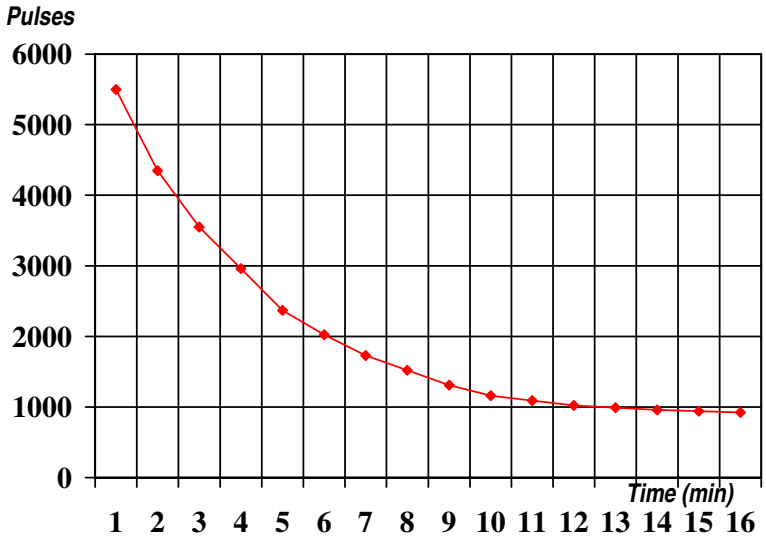


Figure 1
An example of measurement evaluation

The test relies on colloids of a given particle size being screened out from the plasma by (reticulo-endothelial system) cells of the organism. Most of these cells are found in the liver, but there are some in the spleen and the medulla.

The test procedure: Colloid 198 Au is administered to the patient. After the tracer has been assimilated, plasma activity (counts per minute) is determined. Theoretically, measurement points lie on a exponentially decreasing curve.

Structure of the mathematical model of the test is a so-called **mammilary system** of four compartments (Fig. 2).

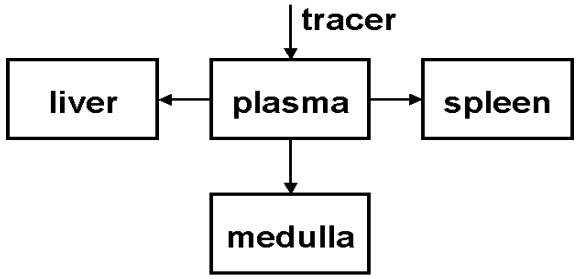


Figure 2
Mamilary system

After input of measurement data the computer program computes the parameter of the mathematical model, then physiological data much more informative for the doctors (biological halving times for liver, spleen, medulla, and blood flow in the liver as a percentage of the volume per minute of circulation).

Exponential regression outputs:

b_1=223, b_2=4729, b_3=-0.32, b_4=1243, $b5$=-0.142, b_6=784, b_7=-0.036

Halving time for liver RES cells: 2.16 min.

Halving time for the spleen RES cells: 4.9 min.

Halving time for the medulla RES cells: 19.36 min.

Blood percentage flowing through the liver V=32.03%

2.3 Non-Homogeneity

The concept of homogeneous compartments is fundamental to most of compartmental analysis. However, if a non-homogeneous compartment is suspected to be present, it can usually be represented by two sub-compartments [1]. One of these exchanges rapidly with the rest of the system, and the second exchanges slowly with the first. Thus the non-homogenous compartment is divided into section 2a and 2b (Fig. 3).

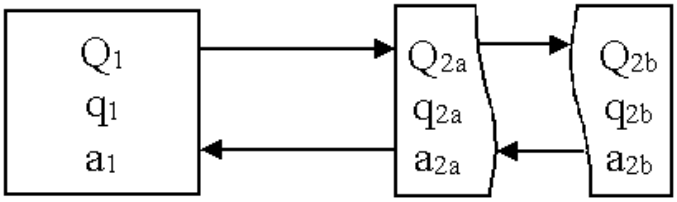


Figure 3
Non-homogenous system

For compartment 1 the amount of tracer present is given by

$$q_1 = a_1 Q_1 \tag{4}$$

For compartment 2 the total amount of tracer is

$$q_2 = q_{2a} + q_{2b} = a_{2a} Q_{2a} + a_{2b} Q_{2b} \tag{5}$$

If is the mean specific activity of the substance in compartment 2, then

$$q_2 = \bar{a}_2 Q_2 \tag{6}$$

Becauseand are exchanging rapidly, then a_{2a} is almost equal to a_1, so that

$$\begin{aligned} \bar{a}_2 Q_2 &= a_1 Q_{2a} + a_{2b} Q_{2b} \\ Q_{2b} a_{2b} &= \bar{a}_2 Q_2 - a_1 Q_{2a} \end{aligned} \tag{7}$$

If $\bar{a}_2$ and a_1 are expressed as functions of time, then

$$a_{2b}(t) = \bar{a}_{2(t)} \frac{Q_2}{Q_{2b}} - a_1(t) \frac{Q_{2a}}{Q_{2b}}. \tag{8}$$

Now, if the tracer is initially added to compartment 1, $a_{2b}(0)$ will be zero. During the experiment the specific activities in compartments 1 and 2 can be determined and plotted as functions. Extrapolation of these functions to t=0 gives values for $\bar{a}_2(0)$and $a_1(0)$. Thus, at t=0

$$\bar{a}_2(0) Q_2 - a_1(0) Q_{2a} = 0 \tag{9}$$

$$\frac{Q_{2a}}{Q_2} = \frac{\bar{a}_2(0)}{a_1(0)}. \tag{10}$$

Hence, the sizes of the sub-compartments can be calculated.

2.4 Study of Non-Homogeneity

Our study concerning non-homogeneity are focussed on the detection of liver diseases. The detection is based on the CLEARANCE OF COLLOID 198 Au test. Human experiments have been made at Semmelweis Medical University of Budapest, the clearance process of over 500 subjects have been identified. The detection of liver diseases has been performed using the following model (Fig. 4).

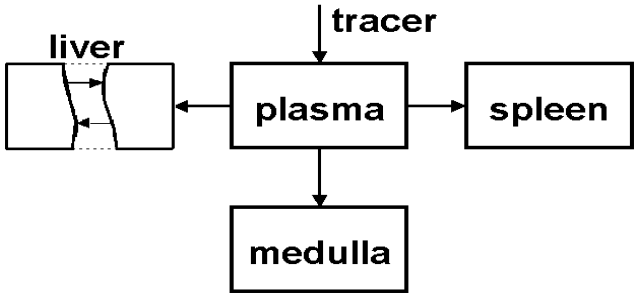


Figure 4
Non-momogenous mamilary system

We have found, that in case of liver diseases the 4th exponential factors b_8 and b_9 have significant values. Among the over 500 subjects, 45 such cases have been found. To approve the real liver disease, further examinations proved to be necessary: biopsy, liver enzyme tests, etc. These further examinations lead to 42 real liver disease cases.

Consequently we can affirm, that the non-homogeneous compartment models consist a useful tool for the identification of damages in physiological systems.

With further theoretical examinations we'll hopefully be able to extend this method for other practical applications.

2.5 Use of Event Recognition Method for Determining the Illnesses of the Physiological Systems

Monitoring and quick detection of illnesses of physiological systems are of great importance. An inverse event recognition method [2] has been used for such purpose.

As a first step the parameters of a linear process model should be identified (as terminal variables). After that the monitoring of state variables and using a proper event recognition method, the outer variables (nature of illnesses) can be recalculated by an identified linear model. In this way the possible illnesses of the system can be located.

2.5.1 Linear Model

Let us consider the following linear multivariable dynamic model:

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{11}$$

where ***x(t)*** is the measurable ***state deviation vector*** characterising the deviation from the normal steady state, ***u(t)*** is the so-called ***event indicator vector***. Whenever the ***j***-th **event** takes place $u_j=1$, otherwise $u_j=0$. The elements of the ***A*** and ***B*** matrices can be identified by the "**teaching**" process.

2.5.2 Teaching Process

The identification of the matrices, the elements for the linear model can be accomplished by direct measurements. During this identification scheme the events to be recognised are numerically simulated one by one and the matrix elements a_{ij} and b_{ij} can be computed according to the *x(t)* trajectories provided by the physiological hybrid model that minimises the following multi-variable non-linear function:

$$I(a_{ij}, b_{ij}) = \sum_{n=1}^{N} \int_0^T \left(\dot{x}_m^n - Ax_m^n - Bu_m^n \right)^2 dt \tag{12}$$

where $x_m^n(t)$ is the measured trajectory (in our case computed by the non-linear model) in the *n*-th experiment, when only the *n*-th element of the vector *u* differs from zero.

2.5.3 Event Recognition

The event recognition is based on the measurement of state variables and the inverse solution of (11). The solution of Eq. (11) for discrete time step ΔT in recursive form [3] is:

$$x^{k+1} = \Phi(\Delta T)x^k + \frac{\Delta T}{2} Bu^{k+1} + \frac{\Delta T}{2} \Phi(\Delta T) Bu^k \tag{13}$$

where the trapezoidal rule was used and

$$\Phi(\Delta T) = e^{A\Delta T} \; and \; x^k = x(t), t = k\Delta T \tag{14}$$

Reordering Eq. (13) for the event indicator-vector, we get

$$u^{k+1} = \frac{2}{\Delta T} B^{-1} x^{k+1} - \frac{2}{\Delta T} B^{-1} \Phi T x^k - B^{-1} \Phi(\Delta T) Bu^k \tag{15}$$

In the case when disturbance of the *j*-th type has occurred, the state deviation vector *x(t)* will change accordingly and the value of the event indicator computed by the recursive algorithm (15) will approach (11).

As the linear model estimates the states of the non-linear process in integral or global way (see Eq. (12)), therefore using an integral (global) indicator vector *U(t)* is more efficient [3] [4] [5].

3 Diabetic Management and Blood Glucose Control

Symbolic computation has been applied to design a procedure for blood glucose control of diabetic patients under intensive care. The used model has been verified and tested in clinical environment. From the engineering point of view, the treatment of diabetes mellitus can be represented by outer control loop to replace the partially or totally failing blood glucose control system of the human body. We have used symbolic computation to design multivariable modal control based on the space representation of a verified nonlinear model. To simulate the insulin-glucose interaction in human body, a two-compartment model was employed, which was verified by parameter estimation technique based on clinical test data. Insulin injection was given to the test person and the glucose concentration of blood has been measured in time. According to the results, the system is expected to provide a useful help to the control of blood glucose level in diabetics under intensive care, and to the optimization process of diabetic administration [10], [20].

4 Remote Patient Monitoring and Diagnostic System

The goal of this research project was to develop a patient monitoring system supporting remote access of the patient information. The traditional patient monitoring systems do not support the remote access (i.e. through internet connection) of the managed patient information. The remote access of the system makes possible the implementation several useful services:

- patients can be monitored by their doctors through the internet;
- remote diagnostic systems can be connected to the monitoring system to process the patient data.

The remote patient monitoring system is developed by the collaboration of the Department of Informatics, Széchenyi István University and the Department of Control Engineering and Information Technology, Budapest University of Technology and Economics (DCEIT-BUTE).

The structure of the system is shown in Fig 5. The system integrates a local patient monitoring system shown at the left hand side of the figure. This subsystem is responsible for the data acquisition. This a system had been developed previously at the BUTE in the framework of the INCO/Copernicus Project (INCO 960161) project with the collaboration of the University of Munich, University of Vienna, University of Prague, and University of Wroclav [11], [12].

The data is extracted from the local monitoring system by a data access server which transfers it to the data converter server. The role of the data access server is to increase the portability by isolating the other parts of the system from the local patient monitor. If somebody want to use the remote patient monitoring system with other local patient monitor only the data access server has to be modified.

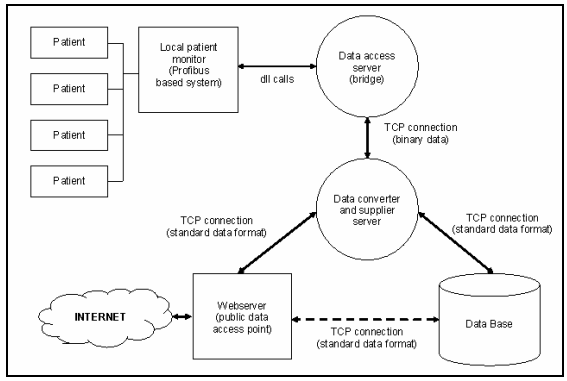


Figure 5
Structure of the remote patient monitoring and diagnostic system

The internal data representation of the system is standardized: all the system components exchange the information in the same format. The format of the internal data representation is defined based on international standards in order to make easy the interconnection of the remote patient monitor to other medical information systems such a HIS. The responsibility of the data converter server is to convert the binary data (from the data access server) to the standard internal representation.

The external interface of the system is supplied by a web server. The system stores the patient data in an SQL data base to provide access to the historical patient data as well.

Many free software components are integrated to the remote patient monitor in order to reduce the cost and the time of the development.

The core remote patient monitoring system has been already implemented. Currently we are working on the development of diagnostic systems to connect them to the remote patient monitor.

5 Automated System for Analyzing Cardiac Ultrasound Images

Ultrasound echocardiography is a widely used clinical technique, but the images obtained using the current technology, are still processed manually with semiautomated methods. In contrast to this, the newly developed system works in an automated way, first obtaining a series of long and short axes views of the heart synchronized by the ECG in real time, then processing them offline. After the detection of the internal edges of the left ventricle, the system determines the short/long axes areas, diameters, calculates the volume of the left ventricle frame by frame and, based on this, the ejection fraction for each cardiac cycle [13].

6 Event Recognition and State Classification to Detect Brain Ischemia by Means of EEG Signal Processing

EEG is an electric signal, which is the spatial sampling of an electric field around the active neurons of the brain. In general, we observe both temporal and spatial characteristics of this signal ensemble. Due to its origin from nerve cells, it is strongly correlated with the primary component of the metabolic activity of these cells: oxygen supplied by cerebral blood flow. Physiological inputs modify the state of the brain, which is reflected in parametric changes based on the EEG. The available processing methods are automatic event recognition or state classification, or semi-quantitative human observation. Two series of animal experiments were carried out with cats and rats, where the occlusion of the middle cerebral artery was used as a model of the ischemic damage. For human registrations, data was collected during open-heart surgery and carotid endarterectomy; in these cases the observation of the neurologist served as reference. Three different methods for event recognition were developed: (1) wavelet-based semi-quantitative spectral analysis; (2) automatic detection by AR model fitting and observation of dominant frequency and peak power shifts; (3) multi-channel activity correlation method based on the eigenvalue representation of a correlation matrix [14]. On the other hand, two classification methods were developed: (1) using a distance measure derived from the Karhunen-Loéve transform; (2) discriminant analysis using multi-channel autoregressive model parameters [15].

7 Single-Channel Hybrid ECG Segmentation

Our new system presents a time domain single-channel ECG segtation method based on beat type and a priori morphological information. The propesed hybrid method combines the advantages of several well-known ECG signal processing techniques with several new solutions. The denoising of the ECG signal is provided by a wavelet-based filter [16]. The R-peaks are detected on a differential basis. A new adaptive beat isolator was developed, which separates the samples belonging to individual ECG beats. The separated beats are classified by a neural network, which can currently distinguish the normal sinus rythms, ventricular and premature atrial contractions. We used a modified version of the PLA algorithm, which produced the linear approximation of the beat. The rule-based morphological segmentation is carried out by individual segmentation units specific to the beat type. The test and evaluation of the system was carried out using reference ECG databases. The segmentation precision was over 90%.

8 Detection of Breathing Disorders like Apnea and Hypopnea

The most common breathing disorder is the apnea, which means the lack of breathing for certain amount of time. Episodes of apnea can happen commonly during sleep, but also at intensively cared patients and new-born babies. Frequent episodes of apnea can disturb the normal breathing process and can cause the development of metabolic, organic and nervous disorders. The diagnosis of apnea has several standardized steps, including the detection of apnea and its milder version (hypopnea), the determination of their type (central vs. obstructive) and the calculation of frequency. The applied signal processing methods evaluate the respiration signals and provide over 80% apnea detection accuracy. The determination of event type is provided only by the off-line analysis of multi-channel signal assays (e.g. polysomnography). We developed a new neural-network based apnea/hypopnea detection method, which uses the sophistically preprocessed nasal airflow signal as network input. We tested two-hour-long airflow excerpts of 16 different patients. The entire detection precision was over 90% [17].

9 Evaluation of the Cry of Normal Hearing and Hard of Hearing Infants

Is there any difference between the cry of a normal hearing and a heard of hearing infant? What is the earliest time in the infant's life when these differences are manifested? These are the questions to answer during the investigations. The recently developed technique consists of four steps: recording, pre-processing, processing, and evaluation. Recording is performed in a quiet room, at 44100 Hz sampling frequency, the distance between the microphone and the mouth of the infant was 1 meter. Pre-precessing phase contains an intelligent segmentation of stationary signal parts. The processing phase consists of extracting statistic information on the strength of the speech chorus from the clarified signal. FFT is applied to obtain the spectrum of the statistical signal. In order to perform the evaluation, the fundamental frequency of the cry has to be determined. The fundamental frequency isn't the most dominant one, but the lowest useful component in the spectrum of the infant cry. This frequency differs from infant to infant, and in most cases is situated between 400 Hz and 600 Hz. The fundamental frequency is determined using the Fourier transformated spectrum called cepstrum. The diagnosis of the infant cry is based on the following criterion: the dominant frequency of the normal hearing infant is an odd multiple of its fundamental frequency, while those of hard of hearing ones are even multiples [19].

10 Molecular Biology Studies with DNA-Chips

This topic is focused on a DNA-sequencer system for translation of genetic information in to protein-structure. Several problems have been studied in this domain: (1) making electronic DNA-chips, developing new electronic hybridization and immobilization methods with polipirrol-copolimerization and with tiol-subgroups on gold-monolayer; (2) tumor development and propagation, several methodological questions, such as producing reference-RNA for microarray analysis and describing a simple recovery method of cyanine dye labeled nucleotide triphosphates; (3) Monitoring mRNA-expression changes and chromosomal aberrations in specific tumor-tissues and to build up a database. Using spotted cDNA-chips he performs Comparative Genomic Hybridization and mRNA-expression monitoring examinations on metastatic lung and hepatocellular tumors and multiplex cancers associated with sclerosis tuberosa, MEN-syndrome and adrenal gland adenomas. The main goal is to gain insights and to determine genes and pathways playing role in metastasis-development and multiplex cancer development [21].

Acknowledgement

The research topics presented above were supported by the National Science Foundation (OTKA), Grants No. T04299, F046726 and K69055.

References

[1] G. L. Atkins: Multicompartment Models in Biological Systems, Methuen and Co., 1969

[2] Z. Benyó: Experimental Determination and Analysis of the Process Models, Medical-Physiological Application, Doctor of Sciences Thesis, Hungarian Academy of Sciences, 1993 Budapest, 138 p.

[3] Z. Benyó, I. Benyó, S. Benedek, B. Paláncz: Use of an Event Recognition Method for Determining the Illnesses of the Physiological Systems. Proceedings of the Twelfth Annual International Conference IEEE EMBS, pp. 1198-1199, 1990

[4] Z. Benyó, I. Villányi, E. Berki, I. Benyó: Application Techniques of Comartment Analysis for Physiological Processes and System Identification. 10th World Congress on Automatic Control. Preprints Volume 5, pp. 89-96, Munich, 1987

[5] Z. Benyó: Computer Analysis of Physiological Processes, Multicompartment Models, System Identification Event Recognition, Computational Statistics. Journal on Communications, Medical Electronics, Volume XLIII, July 1992, pp. 21-29

[6] Holtum D.: Education in Biomedical Engineering in Europe, Stuttgart, 1992

[7] Benyó Z.: Hungarian Education System and its Transfer Abroad, International Federation of Automatic Control (IFAC) 9th World Congress, IFAC'84, Case-Study No. 5, XI, 1984, pp. 107-121

[8] Benyó Z., S. M. Szilágyi, P. Várady, B. Benyó: Biomedical Engineering Education in Hungary, 20th Annual International Conference of IEEE/EMBS, Hong Kong, VI, 1998, pp. 3359-3360

[9] Benyó Z.: Computer Analysis of the Physiological Processes, Multicompartment Models, System Identification, Event Recognition, Computational Statistics, Journal on Communications, Medical Electronics, XLIII, 1992, pp. 21-29

[10] Juhász Cs., B. Asztalos: AdASDiM: An Adaptive Control Approach to Diabetic Management, Innovationet Technologie en Biologie et Medicine, 1996, 17 (1)

[11] Benyó Z., B. Benyó, P. Várady: Patient Monitoring on Industry Standard Fieldbus, IEEE First Joint BMES/EMBS Conference Serving Humanity, Advancing Technology, Atlanta, 1999, p. 704

[12] Benyó, Z., P. Várady, B. Benyó: Remote Patient Monitoring System based on an Industry Standard Fieldbus, 2nd World Congress on Biomedical Communication, Amsterdam, 1999, p. 51

[13] Várady P.: New Methods for Computerized Analysis of Vital Parameters, and their Integration into Diagnostic System, PhD Thesis, Budapest University of Technology and Economics, 2002

[14] Czinege, L.: Event Recognition and State Classification to Detect Brain Ischemia my Means of EEG Signal Processing, PhD Thesis, Budapest University of Economics and Technology, 1998

[15] Benyó, Z., L. Czinege: Computer Analysis of Dynamic Systems with Application in Physiology, 15th World Congress of IMACS on Scientific Computation, Modeling and Applied Mathematics, Berlin, III, 1997, pp. 663-668

[16] Szilágyi S. M., L. Szilágyi, L. Dávid: Comparison between Neural-Network-based Adaptive Filtering and Wavelet Transform for ECG Characteristic Points Detection, 19th Annual International Conference of IEEE/EMBS, Chicago, I, 1997, pp. 272-274

[17] Várady P., T. Micsik, S. Benedek, Z. Benyó: A Novel Method for the Detection of Apnea and Hypopnea Events in Respiration Signals, IEEE Transactions on Biomedical Engineering, 2002, Vol. 49 (9), pp. 936-942

[18] Várady P., L. Wildt, Z. Benyó, A. Hein: An Advanced Method in Fetal Phonocardiography, ELSEVIER Computer Methods and Programs in Biomedicine, 2003, Vol. 71, pp. 283-296

[19] Várallyay Jr. G., Z. Benyó, A. Illényi, Zs. Farkas: Methods for the Analysis of Acoustic Biomedical Signals. Proc. IASTED International Conference on Biomedical Engineering (BioMED 2005), Innsbruck, 2005, pp. 434-438

[20] Kovács L.: New Principles and Adequte Control Methods for Insulin Dosage in Case of Type I Diabetes Mellitus, PhD Thesis (in Hungarian), Budapest University of Technology and Economics, 2007

[21] L. G. Puskás, A. Zvara, L. Hackler Jr, T. Micsik, P. van Hummelen: Production of Bulk Amounts of Universal RNA for DNA Microarrays, Biotechniques, 2002 Oct, 33 (4), pp. 898-904

Industrial Component-based Sample Mobile Robot System

Péter Kucsera

Budapest Tech
kucsera.peter@kvk.bmf.hu

Abstract: The mobile robot development can be done in two different ways. The first is to build up an embedded system, the second is to use 'ready to use' industrial components. With the spread of Industrial mobile robots there are more and more components on the market which can be used to build up a whole control and sensor system of a mobile robot platform. Using these components electrical hardware development is not needed, which speeds up the development time and decreases the cost. Using a PLC on board, 'only' constructing the program is needed and the developer can concentrate on the algorithms, not on developing hardware. My idea is to solve the problem of mobile robot localization and obstacle avoidance using industrial components and concentrate this topic to the mobile robot docking. In factories, mobile robots can be used to deliver parts from one place to another, but there are always two critical points. The robot has to be able to operate in human environment, and also reach the target and get to a predefined position where another system can load it or get the delivered product. I would like to construct a mechanically simple robot model, which can calculate its position from the rotation of its wheels, and when it reaches a predefined location with the aid of an image processing system it can dock to an electrical connector. If the robot succeeded it could charge its batteries through this connector as well.

Keywords: Autonomious Mobile Robot, Localization, Image processing, Industrial, PLC system, Servo Drive, Mobile robor docking

1 The Aim of the Development

The Phoenix Contact is a multinational manufacturer of industrial components and one of the few companies who has almost all the components to build up a mobile robot controlling system. They produce PLC[1] CPU[2] and I/O modules, uninteruptable power supplies, wired and wireless communication components.

[1] Programmable Logical Controller
[2] Central Processor Unit

Phoenix Contact makes big effort to innovate and develope, and they also organize competitions to involve universities and colleges into the development. This experimental system is going to be built up for a competition, called XPLORE, organized by Phoenix Contact in 2007. My team is one of the few who got the opportunity to realize its project in the real life. For the chosen teams, Phoenix Contact covers the cost of the development up to 3000 Euros.

2 Experimental System Structure

On Figure 1, the planned system structure is shown. The mobile part of the project is a three wheel vehicle. Two of the wheels are driven by two DC motors with rotation encoders. The third wheel only chocks the platform and follows its movement. With the help of the encoders each turn of the wheels can be detected. By connecting the encoder signals to fast counters and by processing this data the position can be calculated. This calculation in never accurate, because the small errors from the skip of the wheels are integrated and produce a vast error after a while. To specify the position in critical places an image processing system can be used, which can detect the robot position and send the information to the robot. To make this task easier, a mark is placed at the top of the robot and is recognized by the camera posing system.

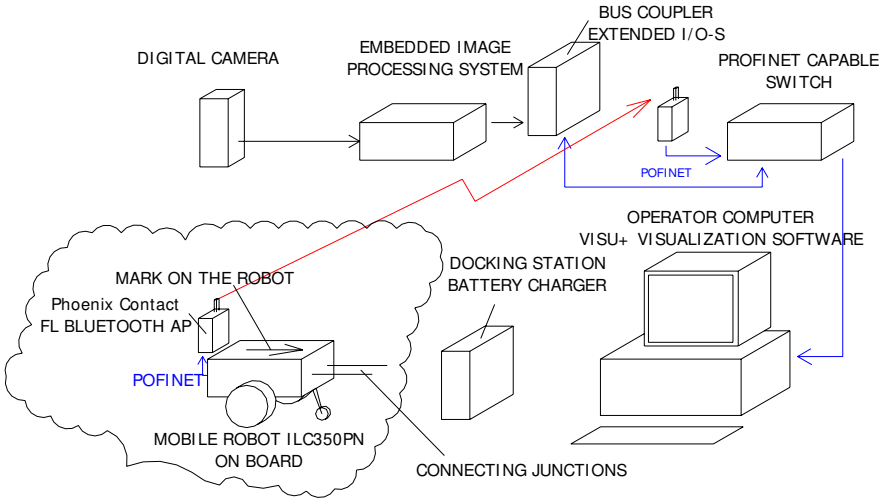


Figure 1
System Structure

2.1 PLC CPU Unit

The mobile part has built on an ILC350PN PLC CPU unit. This unit is based on an ARM32 processor and a Windows CE operation system. The program development is possible through a special software, called PCWORX. The programming languages are the standard IEC61131 PLC programming languages. The speed of the controller is 0.5 ms for 1 K instructions, typically. The communication with the controller is possible through an Ethernet based industrial bus called ProfiNet. The controller's main task is to communicate with the operator, to calculate the position information and control the servo drives. It also has to monitor the power supply voltage and handle the placed sensors. The mobile platform block diagram is shown on Figure 2.

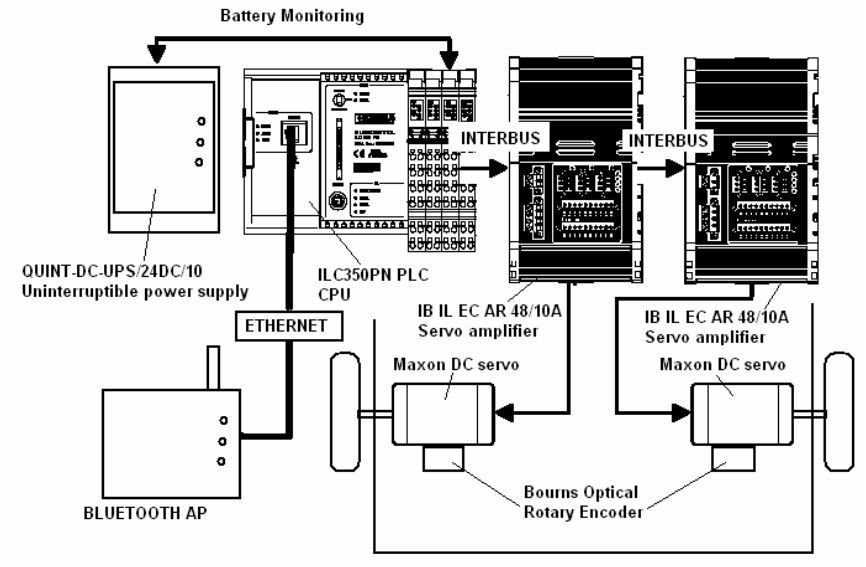


Figure 2
Mobile platform block diagram

2.2 Calculating the Position from the Encoder Signals

Using the simplest mechanical structure has some advantages [8]. If there is only two driven wheels and a chooks wheel, determining the position needs only an easy mathematical calculation ((1) - (5) equations). At the simplest case the two wheels turn at the same speed at the same direction, so the robot moves straight. If the two wheels turn to the opposite direction the robot can turn back in a very small place. By using only two driven wheels, it is easy to build up a robot in a lab from a few components.

Calculating this way only gives proper result if the robot moves in two dimensions, in an absolutely flat ground. If the wheels skip the error is integrated and it adulterates the calculated result. To avoid errors caused by the skipping wheels, the motors have to speed up and down gradually. It is also recommended to tune a controller to keep the wheel speed constant. These problems can easily be solved by using IB IL EC AR 48/10A servo drives.

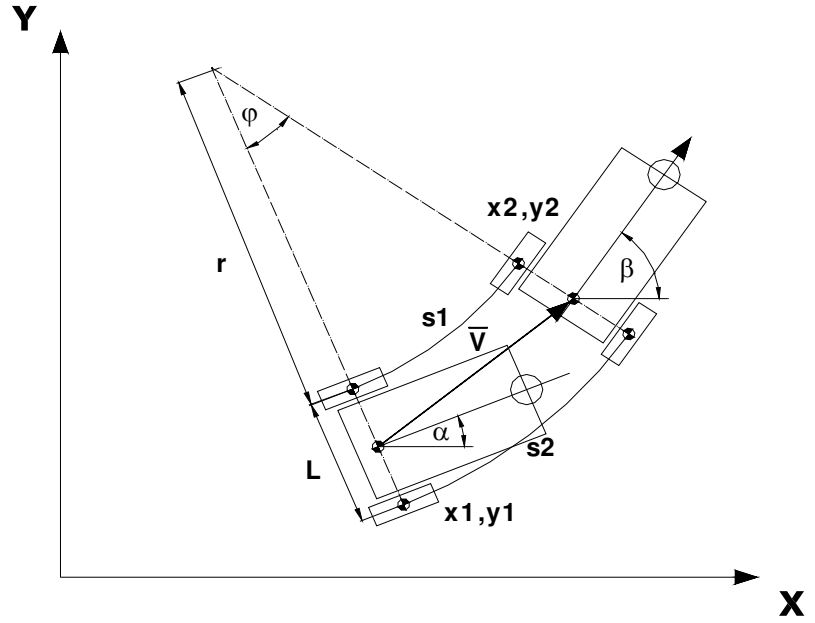


Figure 3
Calculating the position from the turn of the wheels

$$\varphi = \frac{s1 - s2}{L} \tag{1}$$

$$\beta = \alpha + \varphi \tag{2}$$

$$r = \frac{s1}{\varphi} \tag{3}$$

$$\left|\overline{V}\right| = 2 \cdot \left(r + \frac{L}{2} \right) \sin \frac{\varphi}{2} \tag{4}$$

$$x2 = x1 + \left|\overline{V}\right| \cdot \cos\left(\frac{\varphi}{2} + \alpha \right) \tag{5}$$

$$y2 = y1 + \left|\overline{V}\right| \cdot \sin\left(\frac{\varphi}{2} + \alpha\right) \tag{6}$$

2.3 IB IL EC AR 48/10A Servo Drive

The used servo drive can be connected to an industrial bus-system called INTERBUS. This bus is available on the used PLC, so additional gateway is not needed. The extension modules like additional I/O-s can be connected to the CPU unit, through INTERBUS, by using special sliding connectors. The modules are connected serially and one module can be connected to the other like LEGO peaces. That is why it is called INLINE PLC system.

The drive commands to the servo drive are sent through control words on the local bus, and information can be gained at the same way from the status word. To be able to set the parameters of the servo drive a new communication channel, called PCP channel, has to be used. Fortunately the used programming software has ready to use blocks to help the PCP communication. The programmer's task is 'only' to set the proper registers (about 120 different registers and about 30 has to be set) of the drive.

The servo drive has 8 different operation modes. It is possible to use Brushed and Brushless DC motors. Position, speed and toque control modes are available on both types of motors. The speed can be calculated from an incremental encoder or from the motor voltage at the case of DC motors.

The servo drive has its own controller so controlling the motors does not need resources from the PLC CPU. Tuning the controller is also possible through the PCP channel. As it was mentioned, the skipping of the wheels causes errors in the position calculation, so the motors have to be accelerated and braked gradually. This servo drive has the option to set the speed up and down on a specified ramp. The modular structure of the IB IL EC AR 48/10A servo drive can be seen on Figure 4.

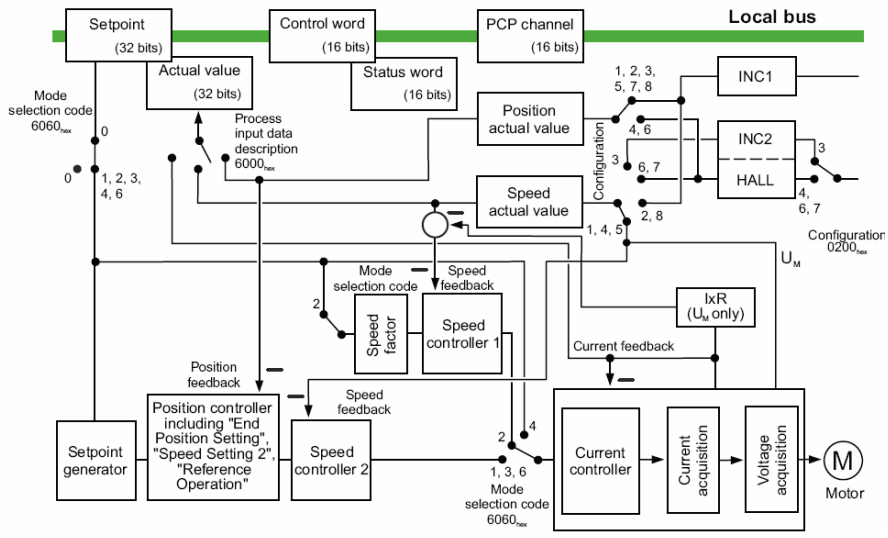


Figure 4
Structure of the IB IL EC AR 48/10A servo drive [7]

2.4 Sensors on the Mobile Robot Platform

One of the most important tasks of a mobile robot system is to acquire knowledge about its environment. This can be done by using different sensors which take measurements and translating the measured information to meaningful data to the control system.

Sensors can be classified to proprioceptive and exteroceptive sensors [1].

Proprioceptive sensors measure internal values like battery voltages, wheel speed.

Exteroceptive sensors acquire information from the environment, like distance from an obstacle, or global position. The sensors collecting information from the real world environment can be arranged to two other groups according to their functions.

Obstacle avoidance – sensing dynamic or static obstacles.

Localization – collecting data to determine the accurate position of the robot.

Furthermore, sensors can be classified to active, energy emitting (for example scanners) and passive energy receiving sensors (for example CCD cameras).

2.4.1 Time of Flight Active Ranging

The time of flight ranging makes use of the propagation speed of an emitted wave and measures the traveling time. The transmitted wave can be sound, light or electromagnetic wave. In the case of mobile robots, the measuring range is usually between 5 cm and 10-20 m. If sound is transmitted, the propagation speed is 342 m/s, so the time of the flight at a distance of 3 m is about 20 ms (to reach the target and get back) which is a measurable value. If the transmitted wave is a laser beam, the propagation speed is 3 x 108 m/s which is about 106 times faster than it was in the case of sound transmitter. The time of flight is 20 ns! If the target is even closer than 3 m it is a very challenging task to measure so short times with small error. On the other hand, if we need short scanning time and measuring huge amount of points, the propagation speed of the sound does not suit our requirements.

2.4.1.1 Ultrasonic Sensors

The ultrasonic sensor's basic principle is to transmit ultrasonic wave packages and measure the time it takes to get back to the receiver. The ultrasonic wave's typical frequency is between 40 and 180 kHz and it is generated by piezo transcuders. The main disadvantage of the ultrasonic sensors is that it can only give information about whether there is an obstacle in an area or not, but it can not give information about the exact position. It is because the sound propagates in a cone with an angle of about 30 degrees. It cannot be determined from which direction the reflected wave came back within the distribution angle. There are several other drawbacks of ultrasonic sensors. The speed of the sound depends on the temperature, and the strength of the reflected signal depends on the acoustic behavior of the obstacle's material. Ultrasonic sensors have relatively slow cycle time. As it was described, at a distance of 3 m, the traveling time is 20 ms. If more sensors are used, and interference needs to be avoided the scanning time multiplies by the number of sensors. Experiments were made using a TURCK TP30UDPB Ultrasonic sensor. The sensor has 30-200 cm sensing range and the ultrasonic cone angle is 150. The sensor is made for a prepared industrial environment so it is not appropriate to sense small size obstacles like a long and slim stick. If there is no obstacle in the sensing range the sensor output is undetermined, even if the obstacle is in 30 cm the sensor can give and value at the output. This sensor can only be used as a complementary sensor to a bigger sensor system [9].

2.4.1.2 Laser Rangefinder

There are two methods to measure distance with laser light beam. One way, like in the case of ultrasonic sensors, laser pulses are transmitted and the reflection time is measured. The other easier method is to transmit 100% amplitude modulated

light beam at a defined frequency and compare the phase shift between the transmitted and the reflected light. This scanner has much higher resolution than the ultrasonic sensor. By pointing the measuring beam to rotating mirror a fast scanning can be accomplished in a vertical dimension, but the whole scanner has to be able to move horizontally if 3D scanning is needed. A very common used sensor is the Proximity Laser Scanner. An example the SICK PLC Proximity sensor has 4 meter protective range and 30 meter warning range. The sensor can scan a 2 dimensional area in 80 ms. If a 3 dimensional scanning is needed a mechanic rotation has to be used to move the scanned area vertically.

2.4.2 Video-based Sensors

The capturing sensor can be placed on the mobile platform or on a fixed place. A fixed camera can give position information to the robot, but can not help in the obstacle avoidance. The camera on board can detect obstacles or marks in the surrounding environment and adopt the movement of the robot to the dynamic environment. If the mobile robot system is used in an indoor environment like in a manufacturing plat, an installed image processing system can recognize a specific mark on the board so it can determine the accurate position of the robot as well. This system could be a complementary sensor to correct the calculated error of the wheel encoders or the gyroscope.

On the lab I have built an experimental image processing system for determining the position of a predefined symbol on a picture. The system consists of a digital camera and an FPGA board, which transmits the camera signal through Ethernet with the help of UDP packages. The processing software is running on a PC. The program first calculates the average brightness value of the picture. In the next step proportional with this value filters the picture dots, giving “1” value if the brightness of the dot is bigger than the average brightness of the picture and giving “0” if the value is smaller. After this processing, a predefined matrix (Figure 5) is running through the filtered image, as a mask, and checks if the rived off image part is the same as the predefined matrix.

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0							0	0
0	0		1	1	1	1		0	0
0	0		1	1	1	1		0	0
0	0		1	1	1	1		0	0
0	0		1	1	1	1		0	0
0	0							0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

Figure 5
Predefined image masking matrix

At this step we have all the position of the predefined size dots on the picture. The last thing we have to check is if the dots are at the same geometrically coded position. On the Figure 6 a sample mark is recognized by the system. The written program can give the position of the given mark, and send this data trough an RS232 serial line to the PLC system.

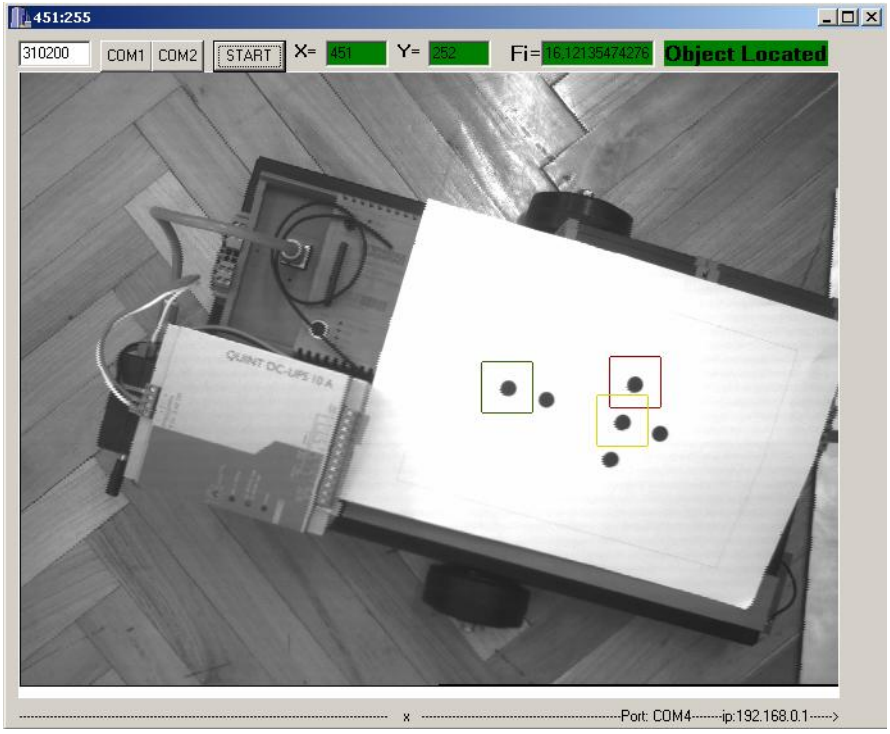


Figure 6
Image processing software, the recognized symbol on the picture

2.5 Communication

To build up this system the whole communication is based on PROFINET and INTERBUS. INTERBUS is used to communicate with the I/O modules, and PROFINET to connect the different modules to each other. The mobile platform has to be able to communicate with the image processing system. It can be done in two ways.

– Using a Bus Coupler prom PROFIBUS to INTERBUS and using a serial I/O card which can send and receive data trough RS232 or RS485 (the image processing system is basically PC based so it is not a problem to establish a serial interface).

– Sending the image processing systems data through Ethernet in UDP packages, and establishing an OPC server in the operator PC, so the visualization software could get the data, and send back to the mobile platform (this solution seems to be more difficult and slower than the first one).

Because there is more than two clients in this system a PROFINET switch is also needed. The communication between the modules uses most of the recent BUS systems, and so gives a good presentation how to connect different parts to each other. Using only standard communication makes the system flexible and extendable.

2.5.1 Wireless Communication

There is a wide range of commercial and industrial communication products which operate in the 2.4GHz ISM band. For example the Phoenix Contact Bluetooth AP (Figure 7) is developed to industrial environment. This device has configurable 0.1 – 250 meter range outdoor. It can be connected to Standard Ethernet network. It has 500kBps data speed, which match to the Phoenix Profinet industrial Ethernet bus.

Figure 7
Phoenix Contact Bluetooth AP [5]

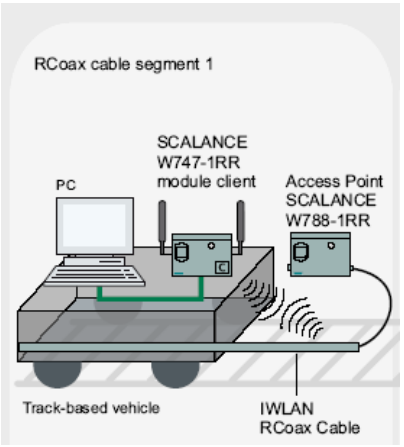


Figure 8
Siemens WLAN mobile communication [6]

Just to show another example, the Siemens also offers wireless solutions for most of the possible mobile platforms. For example if the Mobile platform is track based or working in a very well defined area The IWAN RCoax cable can be a good solution to establish reliable wireless connection (Figure 8).

3 Mobile Robot Docking

During the robots task, which could be anything from transfering goods in a factory to observing an area, the robots battery will drop below a presetted threshold. During the docking process the robot has to be able to connect to a power supply to charge its battery. This coupling can be electrical, electromagnetical or mechanical. The simplest way is to establish electrical connection with the help of a mechanical terminal (Figure 9). The prestressed flexible connector which is placed on the mobile robot can establish a connection with the fixed slope junction with the help of mechanical stress when the robot drives to the docking station. It is a commonly used method to charge a mobile robot battery. If the robot is working with high charging current, it is recommended to first dock in and establish the electrical connection and just after that switch on the charging current, to avoid sparks.

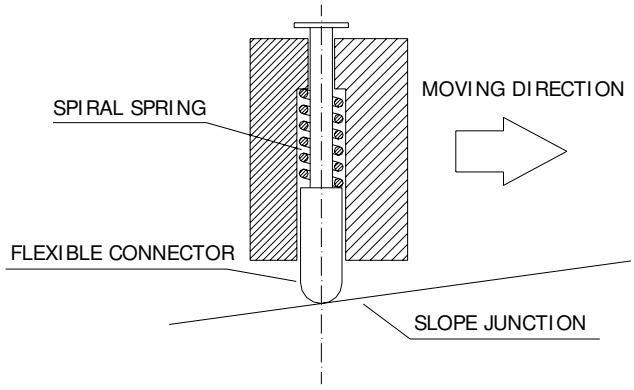


Figure 9
The flexible docking terminal

4 The Built up Mobile Robot System and the Future Development Plants

Now the mobile part is ready and moving (Figure 10). The PLC can calculate the position and can follow a preprogrammed path. The components were delivered 1 month ago and it took a lot of time for us to put together the robot and operate the motors. The docking station is also ready and the next step is to inteface the image processing system to the PLC. We are still waiting some components to accomplish thies task.

It is also planned to develop a visualization system, using VISU+, the Phoenix Contact SCADA[3] program. SCADA softwares are used to communicate with the PLC and give a simple interface to the operators to let them change parameters in the system, or monitor the system state. We want to use this system for monitoring the momentary position of the robot, the battery voltage, reading out the ultrasonic sensor's measurements. This program could be also be used to give the path points to the robot, and control it manually.

Figure 10
The built up mobile robot system

Conclusions

The mobile robot systems have more and more importance these days, so dealing with them in the higher education is necessarily. Autonomous mobile robots can be used to deliver parts in factories, being complementary platforms in a security system and they also can be used in hazardous areas where humans can not stay. This paper is written from a model robot system built for a competition. The system consists of a mobile part, a docking station, an image processing system and an operator computer. At the end of the development the mobile robot has to be able to run cyclically in a predefined path, and if it is needed, it can dock to a

[3] Supervisory Control And Data Acquisition

docking station and charge its batteries. When the described object is reached, the system can be a good base for additional research and development as well.

References

[1] R. Siegwart, I. R. Nourbahsh: Introduction to Autonomous Mobile Robots, The MIT Press Massachutsetts institute of Technology Cambridge 2004

[2] J. Borestein, H. R. Everett, L. Feng: Where Am I? Sensors and Methods for Mobile Robot Positioning, University of Michigan, Michigan 1996

[3] http://pdb.turck.de/media/Anlagen/d100660.pdf (2007.02.15)

[4] http://www.mysick.com/saqqara/pdf.aspx?id=im0011885 (2007.02.15)

[5] http://eshop.phoenixcontact.com/ (2007.02.15)

[6] http://www.automation.siemens.com/download/internet/cache/3/1392310 /pub/en/03_Industrial-Mobile-Communication.pdf (2007.10.15)

[7] www.phoenixcontact.de (2007.10.15)

[8] Gyula Mester: Modeling of the Control Strategies of Wheeled Mobile Robots, Proceedings of The Kandó Conference, pp. 1-4, Budapest, 2006

[9] István Matijevics: Microcontrollers, Actuators and Sensors in Mobile Robots, in Proceedings of 4th Serbian-Hungarian Joint Symposium on Intelligent Systems (SISY 206) September 29-30, 2006, Subotica, Serbia

Digital Electronic Control of a Small Turbojet Engine MPM 20

Rudolf Andoga, Ladislav Főző, Ladislav Madarász

Department of Cybernetics and Artificial Intelligence, Faculty of Informatics and Electrical Engineering, Technical University of Košice
Letná 9, 04200 Košice, Slovakia

E-mail: rudolf.andoga@tuke.sk, ladislav.madarasz@tuke.sk, ladislav.fozo@tuke.sk

Abstract: Small turbojet engines represent a special class of turbine driven engines. They are suitable for scientific purposes and research of certain thermodynamic processes ongoing in turbojet engines. Moreover such engines can be used for research in the area of alternative fuels and new methods of digital control and measurement. Our research, which is also presented in this article, is headed toward these aims. We evaluate and propose a system of digital measurement of a particular small turbojet engine – MPM 20. Such engine can be considered as highly non-linear large scale system. According to obtained data and experiments we propose different model models of the engine and design of situational control algorithms for the engine with use of certain methods of artificial intelligence as new methods of control and modeling of large scale systems.

1 Introduction

The state of present technologies in technical and also non-technical practice implies creation of growing complexity of systems. Turbojet engine as a complex system is multidimensional highly parametric system with complex dynamics and non-linearities. Its particular property is operation in a broad spectrum of changes in environment (e.g., temperatures from -60 to +40 °C). If we want to secure optimal function of such system, it is necessary to develop models and control systems implementing the newest knowledge from the areas of automation, control technology and artificial intelligence (AI). The present control systems and dynamic models are often limited to control or modeling of a complex system in a certain (often operational) state. However, in practice the turbojet engine finds itself in very different working conditions that influence parameters of its operation and characteristics of behavior. By creation of algorithms of control, it is necessary to create models in the whole dynamic spectrum of the modeled system (turbojet engine) and also its erroneous states. Furthermore we need to design a

control system that will secure operation converging to optimal one in all eventual states of environment and also inner states of the system represented by its parameters. Except implementation of classic algorithms of control, it is possible to design such systems of control and models by use of progressive methods of artificial intelligence [1]. In the article, we will focus on implementation methods of situational control as a framework method, which is suitable for use in design of dynamic models and systems of control of turbojet engines with use of intelligent elements that comply with full authority control digital systems standards.

2 Control Systems of Turbojet Engines

The main global aim of control of turbojet engines is similar to other systems and that is increasing their safety and effectiveness by reduction of costs. This demands application of new technologies, materials, new conceptions of solutions [7] and also development in systems of control and regulation of aircraft turbojet engines and processes ongoing in them.

Demands for control and regulation systems result mainly from specific properties of the object of control – a turbojet engine. Among the basic functions of control systems of turbojet engine belong the following ones – manual control, regulation of its parameters and their limitation. Manual control and therefore choice of regime of the engine is realized by a throttle lever according to a flight situation or expected maneuver. By regulation of a turbojet engine we understand such a kind of control where the chosen parameters of the engine are maintained on certain set levels, thus keeping its regime.

In the past the classical control systems of turbojet engines were implemented mainly by hydro-mechanical elements, which however suffered from deficiencies characteristic for such systems. Among such deficiencies were, high mass of such systems, inaccuracies due to mechanical looses and low count of regulated parameters. However development of electronic systems and elements is ongoing, which will allow to increase precision of regulation of parameters of turbojet engines and their count to secure more complex and precise regulation of turbojets.

Use of electronics and digital technologies in control systems of turbojet engines has brought: [7]:

- lowering of mass of control system
- higher complexity of control – The count of regulated parameters used to be 3 to 7 by hydro-mechanical systems, however the digital systems operate with 12 to 16 parameters;

- increasing of static precision of regulation of different parameters (for example, precision of rotations from ±0.5% to ±0.1%, precision of regulation of temperature from ±12K to ±5K
- increase in reliability, service life and economics of operation of the driving unit of an aircraft;
- easier backup, technology of use and repairs, possibility of use of automatic diagnostics.

By design of solution of a control system for a turbojet engine, it is necessary to build an appropriate mathematical model of the engine. The ideal approach to design of electronic systems is a modular one, from hardware or software point of view. This implies use of qualitative processing units that are resistant to noises of environment and also realization of bus systems with low delays is very important in this approach. Further improvement in quality of control can be achieved by implementation of progressive algorithms of control, diagnostics and planning in electronic systems. These algorithms have to be able to asses the state of the controlled system (turbojet engine in our case), then parameterize action elements and they have to be able to control the engine under erroneous conditions represented in outer environment or as errors in subsystems of the engine itself. Prediction of such states represents an area to incorporate predictive control system. Methods of situational control bound with elements of artificial intelligence supply many robust tools for solution of afore mentioned problems and sub-problems.

2.1 Basic Types of Control Systems of Turbojet Engines

From the point of view of use of electrical and electronic systems in controls the turbojet control systems can be roughly divided into following sets: [2]:

- electronic limiters,
- Partial Authority Flight Control Augmentation (PAFCA,
- 'High Integration Digital Electronic Control' (HIDEC); 'Digital Engine Control' - (DEC); 'Full Authority Digital Electronic Engine Control' – (FADEEC)).

The division of control systems into these three levels is not absolutely distinct, as systems on higher level as for example HIDEC system can utilize control mechanisms as electronic limiters. For example FADEC systems are often realized as single or double loop control systems with utilization of PI control algorithms or electronic limiters with estimation filters [6,13]. Example of such FADEC algorithm is shown in Figure 1 [6].

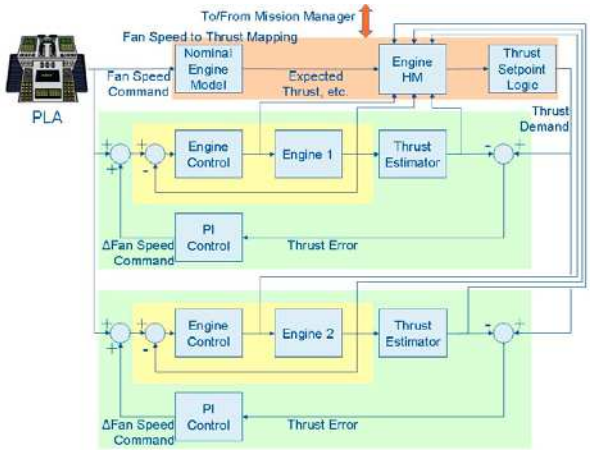


Figure 1
FADEC control system with implemented PI electronic controllers

Such engine control systems are often integrated into the whole framework of an aircraft control system.

2.2 FADEC Control Systems of Turbojet Engines

There are of course many possibilities and methodologies applicable to control systems of turbojet engines, which are FADEC compliant. Such application has to cope with strong non-linearity and changing structure of models and constants during operation of a turbojet engine. Such intelligent system should also be able to form decisions and predict faults either in control circuit or the object of turbojet engine itself. Therefore intelligent turbojet engine control is often bound with design of intelligent diagnostics systems [15] that also deal with control of an engine during its long-term deterioration. Example of such control based on diagnostics modules is shown in Figure 2.

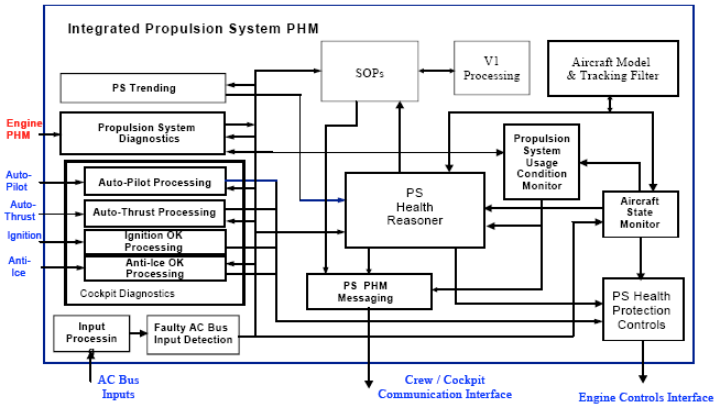


Figure 2
Diagnostic FADEC control system of a turbojet engine [15]

The control system in this case is based on intelligent PHM (Prognostics Health Management) of the engine. Diagnostic systems of turbojet engines can be further realized by means of artificial intelligence. In design of diagnostic and control system which would control the engine in its erroneous states and act long before actual critical states develops itself; we need to form exact dynamic models of the engine. In design of classic control systems only first to second order linear models are commonly used. Methods of AI however offer possibilities of modeling the dynamic parameters of an engine in multi variable space with great precision in the whole range of operation of engine. Such models can have precision within 2% of standard error in whole area of operation of a jet engine [1,2]. Integrated model used for control of a turbojet engine can be seen in Figure 1. Importance of modeling during operation of a turbojet engine can be further extended to fault detection of sensors and other parts of control system and the engine itself. In design of control system, the architecture also plays a significant role. Two common architectures can be presently found in design of turbojet engine FADEC control systems [13]. The first one is the centralized one, which is reliable and well understood, but on the other hand has many drawbacks like inflexibility, high weight, complicated fault detection, etc. This architecture is shown in Figure 3.

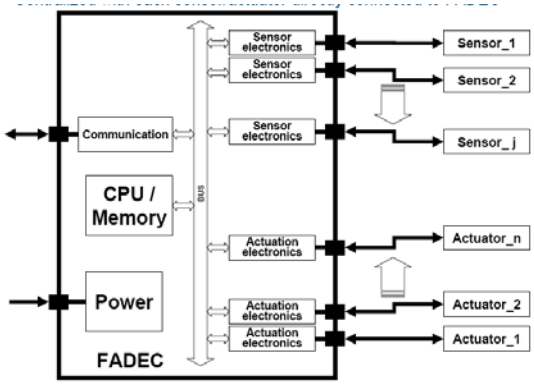


Figure 3
Centralized FADEC architecture [13]

The other usable architecture for design is the distributed architecture (Fig. 4). Its main advantage is high flexibility, easier fault detection and isolation, its cons are mainly higher complexity, communication unknowns and deterministic behavior and it requires new technologies, i.e. high temperature electronics for use in turbojet engines.

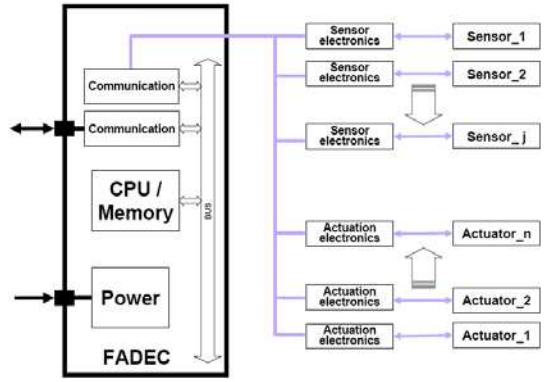


Figure 4
Decentralized FADEC architecture [13]

2.3 Methodology of Situational Control in Turbojet Engines

Methodology of situational control is one of the possible methods of control of complex systems [1,9], however it is necessary to say that as a frame method, it integrates in itself also other methods of control of complex systems. Situational control can incorporate methods of decentralized control, multi-agent system, catastrophe theory, 'anytime control', etc. The basic paradigm of situational

control lies in the fact that a complex system can find itself principally in endless count of state, however resources, strategies and means of control of such system are usually limited. Solution of this problem lies in design of time invariant situational classes (frames) that have their strategies of control assigned.

Because, the creation of exact mathematical model describing all dynamic properties of a complex system like a jet engine is, use of methodology of situational control and modeling bound with data oriented methods of sub-symbolic artificial intelligence is a usable approach to design of models and systems of automatic control for turbojet engines. Neural networks and fuzzy inference systems show up here as appropriate methods implemented in hybrid architectures with paradigm of situational control. For example use of fuzzy inference systems by design of algorithm for control of cooling of turbine blades lowers absolute mean error in simulations from 5°C to 0.2°C compared to use of a classic relay regulator [1,3]. Similarly by use of a hybrid neuro fuzzy architecture bound with methodology of situational modeling a dynamic model of a small turbojet engine MPM 20 has been created with absolute percent error <2% across whole area of operation of the engine. [1,5]. The results show that use of these methods bound with modern algorithms of learning, we will be able to create a situational system of control of a turbojet engine with full authority.

3 Design of FADEC Control System for MPM 20 Engine

Small turbojet engine MPM 20 is constructionally derived from turbo-starter TS-20, which has been used for rotating a rotor of a normal sized engine by its start-up. The small turbojet engine MPM 20 is a single stream, single shaft turbojet engine with single stage one sided radial compressor, bound combustion chamber, single stage un-cooled gas turbine and fast exhaust nozzle [7].

3.1 Principle of Operation of a Small Turbojet Engine MPM 20

Air is flowing into rotating rotor of compressor. After pressing the air, the whole temperature is increasing, static temperature, static pressure and speed of air are also increasing, The air is flowing from the compressor in radial direction and enters diffuser that changes the kinetic energy of the air into pressure. The diffuser also divides air uniformly for the combustion chamber. In the combustion chamber, chemical energy of fuel is change into temperature. The gas turbine changes the pressure and temperature energy of gases into kinetic energy and further into mechanical work. The stator is created as a system of narrowing

channels, where potential energy of gasses. The gasses flowing through the rotor change their moment of movement and create work of the turbine which drives the compressor. Beyond the turbine, the gasses have still higher pressure than the outer atmosphere and are flowing into the outlet. In the exhaust nozzle the temperature and pressure energy is changed into kinetic energy that creates thrust of the engine [5].

3.2 System of Automatic Control of MPM 20

3.2.1 Stabilization System of MPM 20

The control system of MPM 20 is corresponding to its purpose as a single regime turbo-starter of a normal sized turbojet. This control system stabilizes the work of the engine on one preset regime. (that means one level of rotations per minute of its rotor). Control of the engine is realized by aggregate, which can have preset level of fuel flow by start-up of the engine and for the regime of its stable operation. In the regime of stable operation, the fuel system is influenced by an action element, which works as a proportional transfer element (P element) according to pressure P_2 (pressure beyond the compressor). That means that by lowering of pressure supplied by the compressor the fuel from the fuel pup is run into backline and fuel supply is thus lowered into engine, what bring decrease in rotations per minute. The function of this element can be described as:

$$Q_{pal} = f(P_2) \tag{1}$$

This system is stabilization system of the engine, because it proportionally increases fuel supply by startup and secures stable and safe operation of the engine, because it is a robust hydro-mechanic element. This system however doesn't allow dynamic changes in regime during operation of the engine, because the constant of amplification of the element given by stiffness of the spring is preset to a constant level and the engine works on constant rotations. Operation of the engine is possible in the constructionally most straightforward variant by implementation of a system of releasing the air from this action element by a controlled release vent.

3.2.2 Digital Control of MPM 20 Engine

Because our aim is hardware and software implementation of a digital system of automatic control of MPM 20, control of the air release from the action element is realized by a circuit including gate valve, oil-fuel pump 414AF-3 and PIC 16F84A microcontroller. To manipulate with the gate valve a servo-vent is used, and this servovent is controlled by PIC 16F84A microcontroller. The basic structure of this control circuit loop is shown in Figure 5.

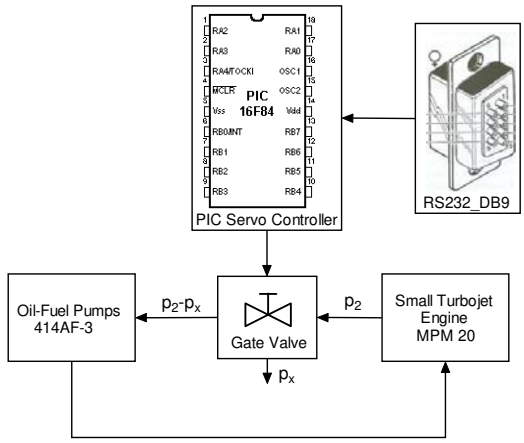


Figure 5
Control circuit loop for digital control of the MPM 20 engine

The gate valve acts as an air vent, which releases a certain amount of pressure, thus decreasing the pressure on the feedback fuel line in oil fuel pump 414AF-3, so the resulting pressure can be expressed as (P_2-P_x) *[at]*. Because the action element acts as a proportional transfer element by decreasing pressure in front of the oil-fuel pump, we decrease the amount of fuel flowing into the engine. Full digital control system with digital realtime measurement system is shown in Figure 6.

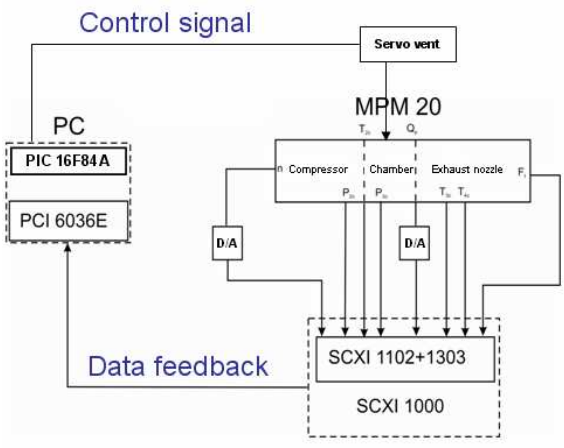


Figure 6
Fully digital control circuit for MPM 20 engine

3.2 Situational Control System Design for MPM 20 Engine

By design of control system for the MPM 20 engine, we consider two basic approaches. In broader scope it will be the situational control methodology and control by a single parameter and in more focused scope it will be anytime control and control by two parameters. The proposed structure of situational control system uses paradigm and schemes described in [3] or in [8]. The whole conception of the situational control system is decomposition of operational states into time spaced situational frames (classes) and every situational frame has one corresponding control algorithm (or controller) assigned to it. Anytime control techniques are more focused because their aim is to control the system in specific critical states or situational frames. By those critical states we mean the deficiency in data for the controller due to system's overload or failure of certain components in control circuit. Anytime control techniques offer possibilities how to avoid these critical states, so that the controlled system would flexibly react to changes of outer environment and could survive deficiency of processing time, data and resources [16].

In development and design of FADEC (Full Authority Digital Engine Control) compliant control system, situational control methodology approach has been used. It is similar in to the one described in previous chapter, what means we use a gating neural network as a classifier of situational frames and system of controllers to handle those situational frames. We use concepts of traditional situational control and formatter control of complex systems [3]. The system has been described in [2] or in [8]. The resulting physical architecture including analyzers of input (X), state (Z), output (Y) and desired (R) parameters is shown in the Figure 7.

Blocks designated as $S_{i,j,k}$ represent controllers for different situational frames, which result from the situational decomposition of operation of the engine in three levels [1]. Within the frame of anytime control methods, we will deal with proposal of a simplified dynamic model of the small turbojet engine. The model will incorporate two input parameters compared to the more complex situational model, which is dependant on the fuel supply parameter only. The model will result from measurements of change to fuel supply and different cross-sections of exhaust nozzles. Simulation of temporary failure in input data (sampling errors) of the system (or designed dynamic model). Design and implementation of anytime control algorithm for the constructed mathematical model and critical states. This will include design of multi-parametric system of automatic control, in our case with two inputs and multiple outputs. Research and observation of flexibility and quality of regulation of the designed system according to measured data and other possible critical states of action units (for example in case of blockage of outlet nozzle or total failure in data measuring current cross-section, etc.). Approaches that lie under the terms 'anytime control', 'anytime processing' serve for proposal

of decisions in real time. They are intelligent supervisor systems using interactive algorithms and have a modular structure [16].

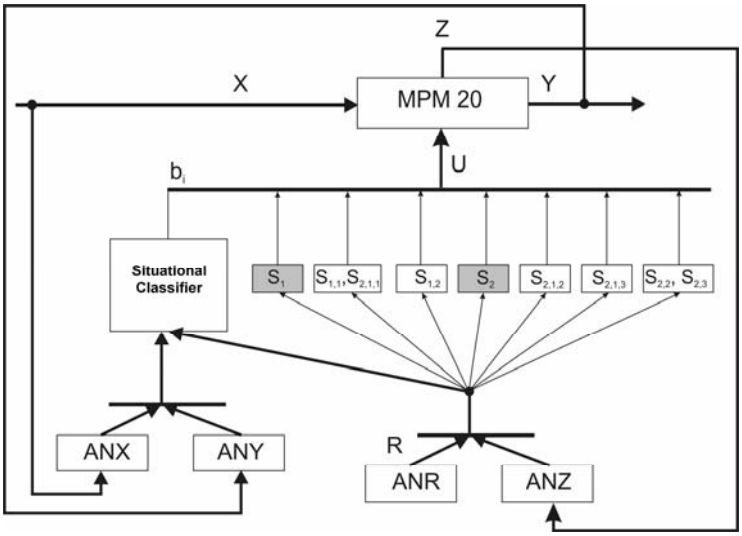


Figure 7
Situational control system architecture for MPM 20 engine

Conclusions

The object of a small turbojet engine MPM 20 provides us an ideal test bed for research of methods in the areas of non-linear dynamic systems modeling and design of advanced control algorithms. Further research will be done in the area of situational modeling that will be headed towards broadening of input parameters of the situational model of the engine and further refinement of situational classes designation. In this area we will be aimed at use of automatic algorithms to find boundaries between situational frames within multivariate space of parameters contrary to their setting by an expert. Anytime control algorithms represent other area of our interest with great possibilities of application of intelligent algorithms that will deal with critical states of operation of the engine and will be further embedded in the whole system of situational control of the engine. Design of such algorithms demands also further refinement of proposed models. All research in the areas of situational modeling, situational control and anytime algorithms should bring new quality of control and modeling in the area of turbojet engines and we expect this knowledge to be also expanded to other areas of technical systems.

Acknowledgement

The work was supported by projects: VEGA no. 1/2183/05 Multi-agent hybrid control of complex systems and VEGA no. 1/2185/05 – Intelligent and information technologies in object recognition.

References

[1] Andoga, R., Madarász, L., Főző, L.: Situational Modeling and Control of a Small Turbojet Engine MPM 20, in Proceedings of IEEE International Conference on Computational Cybernetics, August 20-22, 2006, Tallinn, Estonia, ISBN 1-4244-0071-6, pp. 81-85

[2] Andoga, R.: Hybrid Methods of Situational Control of Complex Systems, Dissertation Thesis, Technical University of Košice, 2006, p. 120

[3] Beneš, J, *Teorie systémů (řízení komplexů)*, 200 pp. Academia, nakladatelství ČSAV, 1974

[4] Harris, Ch., Hong, X., Gan, Q.: Adaptive Modelling, Estimation and Fusion from Data, Springer, ISBN 3-540-42686-8, p. 323, 2006

[5] Hocko, M., *Hodnotenie stavu LTKM na základe zmeny termodynamických parametrov,*, Dissertation Theses, Vojenská Akadémia Gen. M. R., Štefánika, Košice, 2003

[6] Jonathan S., L., Turso, J., A., Shah, N., Sowers., T., S., Owen, K., A.: A Demonstration of a Retrofit Architecture for Intelligent Control and Diagnostics of a Turbofan Engine, NASA/TM -2005-214019, 2005

[7] Lazar, T. et al., *Tendencie vývoja a modelovania avionických systémov*, Ministerstvo Obrany SR, p. 160, 2000, ISBN 80-8842-26-3

[8] Madarász, L., Andoga, R., The Proposal of use of Hybrid Systems in Situational Control of Jet Turbo-compressor Engines, *In Proceeding of SAMI 2005 Conference*, January 21-22, 2005, Slovakia, p. 479, ISBN 963-7154-35-3

[9] Madarász, L., *Inteligentné technológie a ich aplikácie v zložitých systémoch,* Vydavateľstvo Elfa, s.r.o., TU Košice, 349 pp., ISBN 80 – 89066 – 75 - 5, 2004

[10] Moller, M. F., A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning, Neural Networks, Vol. 6, pp. 525-533, 1993

[11] Pospelov, D.A, *Situacionnoje upravlenije. Teoria i praks. Nauka*, Moskva, 1986, 284 pp

[12] Ružek, J., Kmoch, P., *Teorie leteckých motoru I.*, 373 pp, 1979

[13] Sanjay G.: Fundamentals of Aircraft Engine Control Design Course, Lecture, September 15, 2007, NASA Glenn Research Center, 2007

[14] Sanjay, G.: NASA Glenn Research in Controls and Diagnostics for Intelligent Propulsion Systems, NASA/TM 2005-214036, 2005

[15] Wiseman, M.: Intelligent Engine Systems, NASA CR/-2005-213964, 2005

[16] Zilberstein, S.: Using Anytime Algorithms in Intelligent Systems. AI Magazine, 17(3):73-83, 1996

Detecting Renamings in Three-Way Merging

László Angyal, László Lengyel, Hassan Charaf

Department of Automation and Applied Informatics
Budapest University of Technology and Economics
Goldmann György tér 3, H-1111 Budapest, Hungary
{angyal, lengyel, hassan}@aut.bme.hu

Abstract: Teamwork is the typical characteristic of software development, because the tasks can be splitted and parallelized. The independently working developers use Software Configuration Management (SCM) systems to apply version control to their files and to keep them consistent. Several SCM systems allow working on the same files concurrently, and attempt to auto-merge the files in order to facilitate the reconciliation of the parallel modifications. The merge should produce syntactically and semantically correct source code files, therefore, developers are often involved into the resolution of the conflicts. However, when a general textual-based approach reports a successful merge, the output can still be failed in compile time, because semantic correctness cannot be ensured trivially. Renaming an identifier consists of many changes, and can cause semantic errors in the output of the merge, which subsequently have to be corrected manually. This paper introduces that matching the identifier declarations, e.g. class, field, method, local variables, with their corresponding references in the abstract syntax trees of the revisions, and considering the detected renamings during the merge takes closer to semantic correctness. The problem is illustrated and a solution is elaborated in this work.

Keywords: Three-way Merge, Abstract Syntax Tree, Refactoring, Renaming Identifiers, Semantic Errors

1 Introduction

There are two traditional concurrency models among the source code management (SCM) systems: lock and merge models. The lock model prevents the concurrent modification on the same files, but the merge model allows the parallel editing, and performs a merge to reconcile the changes. A three-way merge engine is a usual part of SCM systems, and some of them attempt to auto-merge the files, but they often fail due to textual-based approaches or semantic conflicts. The best methods should treat modifications as semantic changes in high abstraction level, rather than atomic changes. The atomic changes do not reflect the intentions of the developers at all, therefore, discovering those intentions can significantly improve merge approaches.

Refactoring [1] means restructuring the code of a system without modifying its run-time behaviour. Refactorings are composite changes in higher abstraction level. In contrast to simple low-level atomic changes, they aim at improving several characteristics of the software source code e.g. understandability, maintainability. For instance, renaming an identifier to a better name, can help the understability, while this renaming activity consist of numerous atomic changes in the code.

The reconcilation of two modified revisions of a source code file is referred to as merge. Merge includes (i) finding the last incorporated changes by differencing, (ii) conflict detection and resolution, and finally (iii) the propagation of changes in order to produce the reconciled version. The differencing matches the corresponding elements (e.g. lines, syntax elements) of the original and the altered file. The modifications are derived from the non-matched elements. The output of the differencing is the list of edit operations that are applied to the original file provides the modified one. The three-way merge approach (illustrated in [2]) is an unambiguous way of detecting the modifications in the altered revisions, and where the original file is also taken into account. Change propagation is performed by replaying the detected edit operations.

The granularity of the merge means the size of the smallest indivisible changes that can be detected and then propagated. Obviously, the fine-grained methods have slower execution time over coarse-grained ones, but better conflict resolution can be achived by a fine-grained merge. For example, the widespread line-based textual algorithm [3] detects even the smallest change as the line changed. More changes within the same line became invisible and the source of further merge conflicts. Usually, there are relations among the independent changes, which involve certain semantic meanings as well. These relations should be considered while merging revisions of files. The refactorings affect many lines of the file. The typical merge engines handle the composite changes as set of independent atomic changes. This makes them unfeasible for merging files after refactoring.

Fine-grained approaches like abstract syntax tree (AST)-based approaches (e.g. [4], [5]) are more suitable for source code differencing and merging, because, with the knowledge of the language's syntax, they always produce syntactically correct output contrary to line-based textual approaches e.g. the diff3 tool [6]. However, semantic correctness is not ensured at all by considering the syntax. Furthermore, after the merge, the reconciled AST has to be serialized. The technique that visits all the nodes of the tree to emits source code is called pretty-printing [7]. An AST-based merge is language dependent and works with lower performance, therefore, it is rarely used in general versioning systems.

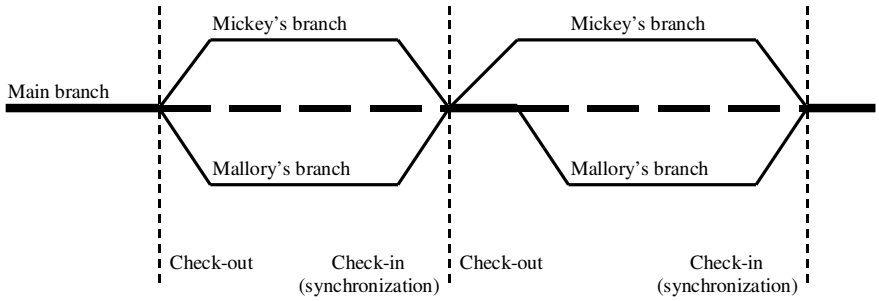


Figure 1
Developing process of the running example

Consider the evolution of a software, which is developed by two users, Mickey and Mallory. They use a versioning system that supports the merge concurrency model. A part of the development process can be followed in Figure 1. After checking-out the files, both of them can modify the same files. Mickey renames a class without telling Malory to do the same. Mallory uses a reference to that class in her inserted lines. After a successful merge performed by the versioning system, they found that the merged file contains some semantic errors. The inserted class references were not corrected with the new name of that class. This paper discusses a solution for these problems.

The renamings should be detected before the merge and should be considered while reconciling the changes. The identifier declarations and their corresponding reference nodes have to be collected by traversing the ASTs of the files. Using the matches from the differencing, a mapping has to be found between the identifiers in different ASTs, which can be used to the correction of the renamed identifiers.

The remainder of the paper is organized as follows. We discuss the semantic conflict related to identifier renaming in Section 2. A solution for that problem is described in Section 3. This is followed by the introduction of the existing approaches and tools, and finally conclusions and future work are elaborated.

2 Problem Statement

Consider the situation, where the developers checked-out a source file to modify independently at the same time. The used file versioning system performs an AST-based three-way merge on the concurrently altered files after file check-ins. Figure 2 depicts the original file and both modified revisions by Mickey and Mallory as well. The merge first compares the revisions to the original file, to detect changes. The difference analysis of the Mickey's version produces the following differences as atomic AST node operations.

```
using System.Drawing;

public interface Widget {
   void Paint(Graphics g);
   void SetLocation(Point p);
   void SetSize(Size s);
}

public class Label : Widget {

   string label;
   Point location;
   Size size;

   public void Paint(Graphics g) {
      g.DrawString(this.label, this.location);
   }

   public void SetLocation(Point p) {
      this.location = p;
   }

   …

   public override string ToString() {
      string str;
      str = this.label;
      return str;
   }
}
```

Original file

Mickey's revision

```
using System.Drawing;

public interface Control {
   …
}

public class StaticText : Control {
   …
   public void Paint(Graphics
graph) {
      graph.DrawString(…);
   }

   …

   public override string
ToString() {
      string value;
      value = this.label;
      return value;
   }
}
```

Mallory's revision

```
using System.Drawing;

   …

public class Label : Widget {
   …
   public override string
ToString() {
      string str;
      str = this.label;
      return "[" + str + "]";
   }
}

public class Button : Widget {
   Label label;

   public void Paint(Graphics g) {
      g.DrawRectangle(…);
      this.label.Paint(g);
   }
   …
}
```

Figure 2
The original and the modified files

OP	Name	Type of the AST node	Name of the parent node	New value
UPD	*Widget*	TypeDeclaration	Global_Types	Control
UPD	*Label*	TypeDeclaration	Global_Types	StaticText
UPD	*g*	ParameterDeclarationExpression	Paint_Parameters	graph
UPD	*g*	VariableReferenceExpression	DrawString	graph
UPD	*str*	VariableDeclarationStatement	CodeStatementCollection	value
UPD	*str*	VariableReferenceExpression	Assign (str=label)	value
UPD	*str*	VariableReferenceExpression	return str	value

Table 1

Mickey's version contains some updates

There are relations among the identified edit operations (Table 1): (i) parameter declaration of *g* has been changed to *graph*, and consequently, the reference to *g* has also been changed to *graph*, (ii) local variable *str* has been changed to *value* and their corresponding references as well. From the high abstraction level semantical point of view, these are two composite changes, not a list of independent atomic changes.

Mallory has not updated anything existing (Table 2), but she has inserted a new class *Button* and reused the existing interface *Widget* and the class *Label*. She has changed an expression with a previously declared local variable *str*.

OP	Name	Type of the AST node	Index	Name of the parent
INS	*Button*	TypeDeclaration	2	Global_Types
INS	*label_0_Label*	MemberField	0	Button
INS	*Paint_Public_Graphics*	MemberMethod	1	Button
INS	*Add*	BinaryOperatorExpression	0	return
INS	*Add*	BinaryOperatorExpression	0	Add (left side)
INS	*"["*	PrimitiveExpression	0	Add (left side)
INS	*"]"*	PrimitiveExpression	1	Add (right side)
MOV	*str*	VariableReferenceExpr	1	Add (right side)
...				

Table 2

Mallory's version contains inserts

The merge does not detect any conflicts between the two list of operations. Executing the edit operations of Mickey's file on Mallory's version, without any semantic considerations, produces the output depicted in Figure 3. The file is still syntactically correct, but contains several semantic errors, which have to be corrected manually after the merge.

```
using System.Drawing;

public interface Control {
   …
}

public class StaticText : Control {
   …

   public override string ToString() {
       string value;
       value = this.label;
       return ("[" + str + "]");
   }
}

public class Button : Widget {
   Label label;

   public void Paint(Graphics g) {
       g.DrawRectangle(this.location, this.size);
       this.label.Paint(g);
   }
   …
}
```

Figure 3
Merged version with certain semantic errors

The variable *str* has been renamed to *value*, class *Label* has been renamed to *StaticText* and interface *Widget* to *Control*, according to the edit operations. However, the newly inserted reference to *str* remains *str* and the class *Button* tries to implement the already renamed interface *Widget*. A merge relies only on the detected edit operations and replays them without sense, this can easily produce compile time errors. The merge should correct the errors by detecting the renames and applying the new names in the newly inserted references.

3 A Renaming-Aware Extension

The purpose of this extension is to extend a three-way merge approach with the ability to be renaming-aware. When reconciling two source files with a renaming that has been performed in one of them, then the newly inserted references with the old identifier names must be renamed as well in order to ensure the semantic correctness. Previous section has shown that merge engines should take the identifier renaming into account and this section proposes a solution, that is illustrated via AST nodes of Microsoft's .NET i.e. CodeDOM [8] nodes.

The two major points of our approach are

(i) discovering the identifier dependencies and building a lookup table of the identifier declarations and the corresponding references with fully qualified names,

(ii) while executing the edit operations, the identifier dependencies are taken into account.

Before describing point (i), we take a closer look at the different types of identifier declaration nodes and their dependencies.

Declaration node	Place of the declaration	References node
VariableDeclaration	In method bodies with unique name	*VariableReferenceExpression*
ParameterVariable-Declaration	In method signatures: method parameter block	*VariableReferenceExpression*

Table 3
Local variable declaration nodes

The union of the visibility scope of local variables with the same name is prohibited within a method body, and a variable (Table 3) with the name of a parameter variable in the method signature cannot be declared, since Java or C# compilers report error. A global lookup table with fully qualified variable names is enough to unambiguously select a certain identifier declaration node.

Declaration node	References nodes
Namespace	In fully qualified *TypeReference* or *VariableReferenceExpression*
Class/Structure *TypeDeclaration*	Base class in class declaration (*TypeReferenceExpression*) Static method invocation (*VariableReferenceExpression*) Static field reference (*VariableReferenceExpression*) Field type (*TypeReference*) Variable type (*TypeReference*) Object creation (*ObjectCreateExpression*) Array type (*ArrayCreateExpression*) Casting (*CastExpression*) Generics (*TypeReference*)
MemberField	*FieldReferenceExpression*
MemberMethod	Method invocation (*MethodReferenceExpression*)
MemberEvent	*EventReferenceExpression*
MemberProperty	*PropertyReferenceExpression*

Table 4
Identifiers with global visibility

The full name comprises the namespace, the name of the class, the method that contains that local declaration, and finally, the variable name as well as the order of its declaration if there are more variables with the same name within a method.

Table 4 summarizes the identifiers with global visibility beside some possible reference nodes that are offered by CodeDOM.

Figure 4 illustrates the partial AST of the running example with its lookup table, and the relations between the nodes and the rows in the table. The identifier lookup table contains the identifiers with their fully qualified names, the link to the declaration node (red arrows), and the list of links to the correcponding reference nodes (blue arrows) as well. This lookup table can be built by traversing the AST before the merge.

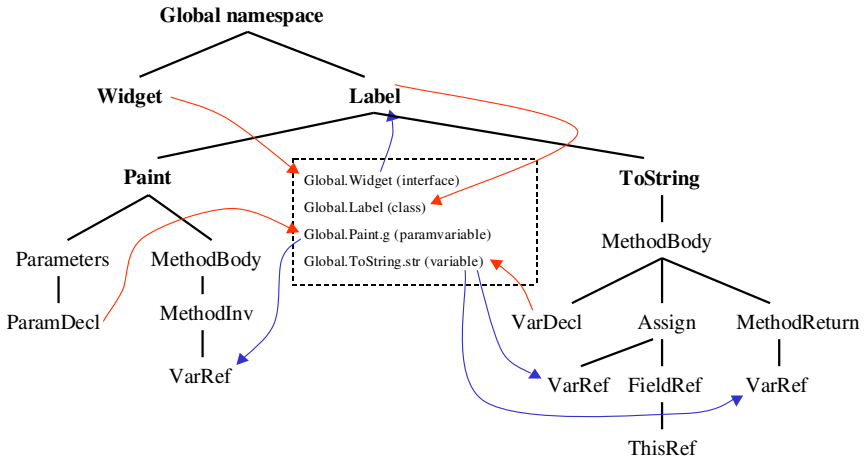


Figure 4

Partial AST of the original version of the code in the running example and its identifier lookup table

According to item (ii), identifier dependencies are considered while doing the merge. We distinguish between two kinds of operation: (a) insert a new node and (b) update an existing node.

First of all, we need a mapping between the lookup tables of the different ASTs. The common point of these different tables is that the difference analysis of the two ASTs matches the corresponding nodes in different trees. For instance, in Figure 4 variable declaration node *str* is matched with variable declaration node *value* in Figure 5, thus, even if their fully qualified name is different, there is a mapping between these nodes. This aim needs a differencing with an identifier name independent match.

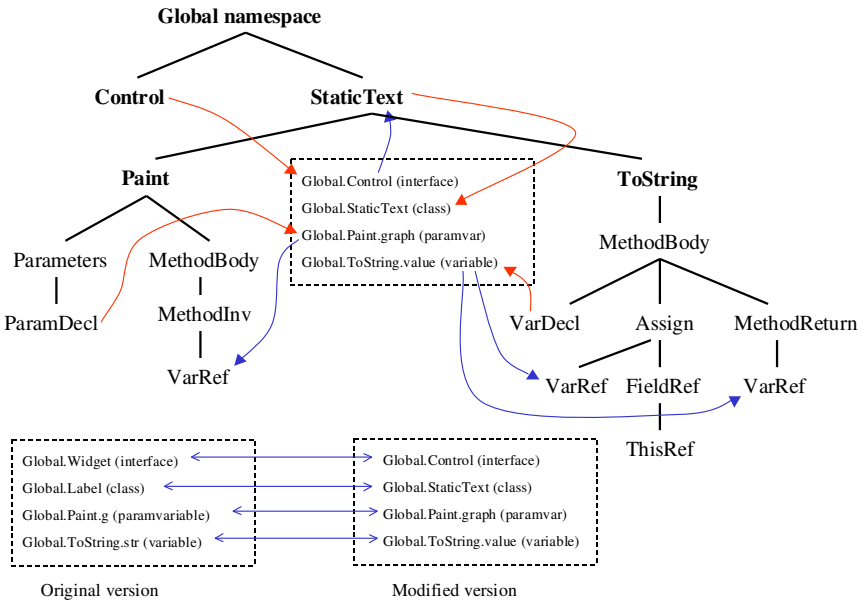


Figure 5
Mickey's version and the mapping between lookup tables

In case (a) when inserting a new reference node, the dependency table should be looked at. If the reference name differs from its declaration name, then the reference name that is going to be inserted must be renamed. Figure 3 illustrates that a local variable *str* is inserted without checking its declaration name, which has been changed to *value* meanwhile, due to Mickey's work. The mapping between the tables allows to look up the matching between the declaration node *str* and the declaration node *value*. Along these connections, we can found that the declaration name is different. Therefore, the algorithm should rename the new variable reference node that is to be inserted, and the new name has to be *value*.

In case (b), when updating an indentifier declaration node, the corresponding references, which also store the name of the identifier have to be changed as well. These reference nodes can be looked up from the table. For example, if we want to execute the edit operations from Mickey's version on Mallory's file, updating the interface declaration *Widget* to *Control* should involve the change of the class reference in the declaration of *Button* from *Widget* to *Control*.

4 Existing Tools and Approaches

This section introduces some of the most relevant tools and approaches that are related to our work.

The well-known *CVS* [9] uses line-based textual merge. Due to its coarse-grained granularity, it detect atomic changes together with their context, for instance, renaming a variable in an expression indicated as the whole line has been changed. After a successful merge of files that were edited in parallel, syntactical and semantical errors can remain in the source code. These problems must be corrected manually after the merge, which was reported to be finished without conflicts. The errors that are revealed in compile time are better than run-time errors, because they are hidden e.g. unintended method overrides and can cause the malfunction of the software.

A common characteristic of textual and AST-based differencing is that they detect several atomic changes without connection between them, abstraction of the changes should be extracted to guess the intentions of the developer. [10] presented that identifying the relations of the atomic changes is important to improve the comprehension of the source code evolution. Small changes can be grouped together into high-level abstract operations. Other advantage of the abstraction is that the changes become reusable on other files.

The state-of-the-art approaches handle source code changes as semantic actions, because they present more information and reflect the intentions of the developers. The differencing techniques that detect changes in lines or in ASTs provide the list of atomic changes e.g. insertion or deletion of a node, but these changes have no abstract information value. The modern integrated development environments (IDE) have the ability to log the semantic changes in high abstraction level and the corresponding low-level details as well. For instance, *Eclipse* [11] has a refactoring engine that logs the changes performed by refactoring actions, like renaming a variable or a class, if they were done via that engine. These logs can be utilized further during the merge process.

Molhado is a refactoring-aware SCM system, that includes an *Eclipse* plugin *MolhadoRef* [12], which captures and stores the performed refactorings on Java files. Its underlying data model is flexible and allows representing programs in any language. It performs only lightweight parsing due to performance reasons, the method bodies in string format are handled as attributes of methods. *MolhadoRef* use the *Eclipse* built-in differencer engine to perform textual difference analysis, the changed lines are examined if the changes were caused by the refactoring operations, and if so, they are removed from the change list. After that, *Molhado* can perform a textual merge by replaying the recorded refactorings together with other edit operations in order to propagate changes. Authors of *MolhadoRef* exhibits better merge results with less human intervention compared to *CVS*.

The merge based on previously saved logs is called operation-based merging. These logs recorded by the source code editor contain text edit operations and informations on other high level source code transformations. Operation-based approaches can be very precise in recording the changes and replaying them, but this technique heavily relies on the editor, and sometimes the log files are unavailable or inconsistent with the changes, because it cannot be ensured that every developer uses the predefined editor. *RefactoringCrawler* [13] is a tool that can reconstruct with good reliability some kinds of applied refactorings by comparing the original and the modified version of a Java file. It uses user adjustable parameters to match the method bodies of the classes. Its matching algorithm is based on an approach that uses fingerprints of the tokenized method bodies. After matching, it performs semantic analysis. *RefactoringCrawler* is limited to examine API interfaces, it does not deal with local variables and has some shortcomings with fields.

In [14], a tool is presented that detects and reports the name and type changes in identifiers of different versions of a C program. The purpose of this tool is to improve the understanding of software evolution with higher level abstract information about the name and type changes. It uses a *TypeMap* for typedefs, structures and unions, a *GlobalNameMap* for global variables, and *LocalNameMap*s per function bodies to collect the matched identifiers. Types and functions are matched if they have the same name. The AST traversal within function bodies is performed by parallel and the local variables are mapped by their syntactical position.

In our approach, there is AST-based differencing and execution of atomic node operations, where the related identifier declarations and references are connected together and taken into consideration while applying those detected operations on the other AST. If any of the identifier nodes are changed, it should also affect the others. If the name of a variable is changed in the declaration, we modify every references that have to be refreshed. If the code, which is taken as input is semantically correct i.e. can be compiled, the identifier references must be consistent with their corresponding declarations. The advantage of our method over other object-oriented tools is that we support local variables.

Conclusions and Future Work

The importance of the comprehension of the committed changes has been pointed out via an illustrative example. The detection of the renamings should be considered in any three-way merge systems, to avoid tedious and error-prone manual code reviews and fixing the code after merging. The presented approach takes us closer to a semantically correct merge. Although, there are several other semantic related problems that were not addressed, however, huge number of compile-time errors can be reduced by the presented approach, and it moves toward an automatic merge without human interaction. Our future work involves the reseach of the solutions to other semantic problems.

The approach can work in merging generated code with manually written code, where refactorings are not explicitly intended by the developers, but caused by the code generator, because some parameters have changed. As future work, we also plan to improve the presented approach to create an efficient code generation tool with round-trip engineering support. The tool can be used in a visual designing environment, which applies bi-directional validated model to source transformations. Round-trip engineering of custom models lacks tool support due to the complexity and the difficulty of the generalization of all specific cases.

Acknowledgement

The fund of 'Mobile Innovation Centre' has partly supported the activities described in this paper. This paper was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

References

[1] Martin Fowler et al., Refactoring: Improving the Design of Existing Code. Addison-Wesley, 1999, ISBN 0201485672

[2] Tom Mens, A State-of-the-Art Survey on Software Merging, IEEE Transactions on Software Engineering, 28(5), May 2002, pp. 449-462

[3] Eugene W. Myers, An O(ND) Difference Algorithm and its Variations, Algorithmica, 1(2), 1986, pp. 251-266

[4] Wuu Yang, How to Merge Program Texts, Journal of Systems and Software, Vol. 27, No. 2, 1994, pp. 129-135

[5] Ulf Asklund, Identifying Conflicts During Structural Merge, Proceeding of the Nordic Workshop on Programming Environment Research '94, Lund University, 1994, pp. 231-242

[6] Sanjeev Khanna, Keshav Kunal, and Benjamin C. Pierce. A Formal Investigation of Diff3, Manuscript, University of Pennsylvania, 2006

[7] Derek C. Oppen, Prettyprinting ACM Transactions on Programming Languages and Systems, 2(4), 1980, pp. 465-483

[8] Microsoft's CodeDOM Web Site, http://msdn2.microsoft.com/en-us/library/ system.codedom.aspx

[9] CVS Wikipedia Web Site, http://ximbiot.com/cvs/wiki/

[10] Peter Ebraert, Jorge Antonio Vallejos Vargas, Pascal Costanza, Ellen Van Paesschen, Theo D'Hondt, Change-Oriented Software Engineering, 15th International Smalltalk Joint Conference, Lugano, Switzerland (to be published), 2007

[11] Eclipse Web Site, http://www.eclipse.org

[12] Danny Dig, Kashif Manzoor, Ralph Johnson, and Tien N. Nguyen, Refactoring-Aware Configuration Management for Object-Oriented Programs. International Conference on Software Engineering. IEEE Computer Society, Washington, DC, pp. 427-436

[13] Danny Dig, Can Comertoglu, Darko Marinov, Ralph Johnson, Automated Detection of refactorings in evolving components, European Conference on Object-Oriented Programming, Nantes, France, 2006, pp. 404-428

[14] Iulian Neamtiu, Jeffrey S. Foster, and Michael Hicks, Understanding source code evolution using abstract syntax tree matching. ACM SIGSOFT Software Engineering Notes 30(4), July 2005, pp. 1-5

Robust Convex Hull-based Algoritm for Straightness and Flatness Determination in Coordinate Measuring

Gyula Hermann

Budapest Tech
Bécsi út 96/B, H-1034 Budapest, Hungary
hermann.gyula@nik.bmf.hu

Abstract: Algorithms to determine the minimum zone straightness and flatness have been successfully established by a number of researchers. This paper after presenting algorithms based on techniques borrowed from computational geometry focuses on the robustness and simplicity of the mathematical techniques. As the most resource consuming part of these algorithms is the determination of the convex hull, both in two and three dimensions, emphasize is given to them. Subsequently the complexity and implementation issues are discussed. The paper outlines an application using the described algorithms.

Keywords: Minimum zone methode, straightness, flatness, coordinate measuring, evaluation software

1 Introduction

The geometric features of manufactured components are generally simple shapes like straight lines, planes, circles, spheres, cylinders and cones. These features can be checked by using specific gauges, but currently these tasks are more and more performed by coordinate measuring machines, because they are powerfull and flexible devices. Obtaining a set of datapoint representing accuratly the workpiece to be checked proved to be difficult. The measurement accuracy depends on the sampling schema, systematic errors, measurement uncertainty and the robustness and efficiency of the verification algoritms.

Form tolerances attached to single features control how close to ideal form the feature must be. Form tolerances most commonly attached to planar surfaces are straightness and flatness. Accurate algorithms determining the deviation from the ideal feature will reduce the possibility to reject good workpieces.

2 Literature Overview

Direct search methods for determining the minimum zone were first mentioned by Murthy and Abdin, who describe a simplex search method starting at the least-square solution and terminating after a number of iteration. Anthony et al. discuss the theoretical basis behind the exchange algorithms, which are essentially geometrically based non-linear programming algorithms, and develop a feasible-descent data exchange algorithm. They also describe a subset algorithm that starts with the solution to a subset of datapoints and iteratively adds points until all points are inside the solution boundary. Carr and Ferreira proposed a linear program model using the small displacement screw matrix to linearize the non-linear constrains, and they used this technique to solve the minimum zone problem. Huang, Fan and Wu proposed the smallest parallelepiped enclosure method for spatial straightness error evaluation from the composition of two orthogonal planer straightness errors (vertical and horizontal). This solution cannot guarantee that the evaluated error has the same value in every direction. Kaiser and Krishnan developed a 'brute force' approach, which finds the minimum zone using a simple iterative algorithm by rotating the initial guess of the envelope lines. The process stops as the condition mentioned in the definition of straightness is reached. This algorithm was extended to determine flatness of a plane face. Samuel and Shunmugam used the computational geometric technique to solve the minimum zone and function-oriented solutions for the straightness and flatness problem. However, the flatness obtained from the 2-2 case was not taken into account in their algorithm. The paper by Lee deals extensively with the 2-2 model and gives a solution based on the convex hull. Suen and Chang propose the application of neural networks to determine the minimum zone straightness and flatness using interval regression analysis. Zhang et al. state in their paper that finding the spatial straightness error based on the minimum zone condition can not be found using the simplified linearized model. They propose evaluation using the original nonlinear model.

Algorithms to determine the minimum zone straightness and flatness have been successfully established by a number of researchers. Because the number of datapoints involved in the evaluation is not extremely large (usually less than 50) this paper focuses on the simplicity of mathematical techniques but at the same time keeps an eye on the efficiency as well.

The paper begins with preliminaries presenting the definition of straightness and flatness as it is given in the standard [ANSI/ASME Y14.5M] followed by the description of the proposed algorithms based on techniques borrowed from computational geometry. Subsequently the performance of these algorithms and implementation issues are discussed. Finally an application using the described algorithms is presented.

3 Preliminaries

In this paper the evaluation of straightness and flatness is considered in accordance with the ANSI/ASME Y14.5 and the ISO/R1101 standards and other related papers listed in the references.

Straightness is a condition where an element of a surface is a straight line. A straightness tolerance specifies that each line element must lie in a zone bounded by two parallel lines separated by the specified tolerance and that are in the cutting plane defining the line element (Fig. 1).

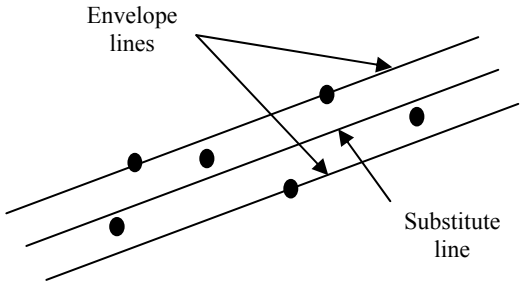


Figure 1
Definition of the straightness with given datapoints

Flatness is the condition of a surface having all elements in one plane. A flatness tolerance specifies a zone defined by two parallel planes separated by the specified tolerance within which the surface must lie (Fig. 2).

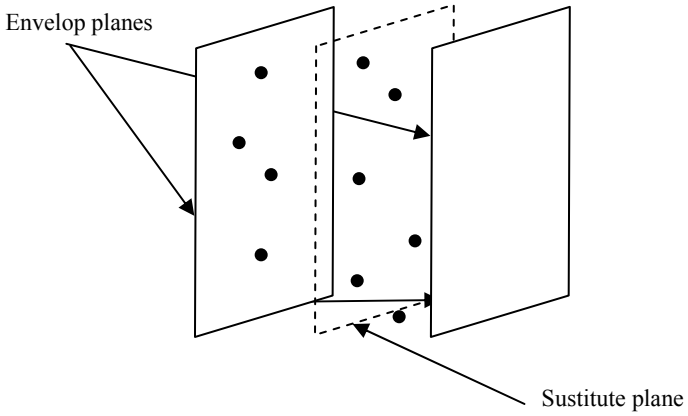


Figure 2
Definition of the flatness with given datapoints

The L_p-norm solution minimizes the following function:

$$L_p = \sum_{i=1}^{n} |d_i|^p$$

Where n is the number of data points, d_i is the distance between the i-th datapoint and the ideal feature and $1 \leq p \leq \infty$. The L_2 norm the so-called least square-fit, widely used in coordinate measuring machine software, minimizes the sum of the squares of the residual errors d_i ; $\text{Min}\sum|d_i|^2$. The Chebyshev best-fit (L_∞-norm solution) minimizes the maximum deviation from the ideal feature and results in a $\text{Min}\{\text{Max}|d_i|\}$ objective function. The Chebyshev best-fit problem can be transformed into a non-linear constrained optimization problem. The solution of it requires iterative process whose convergence depends to a large extend on the initial guess.

In case of straightness and flatness result from computational geometry based on the convex hull can be used to obtain the exact solution in a number of definitive steps.

4 The Proposed Algorithm

Because the problem we are considering here are essentially discrete from its nature we are looking for a robust solution in the context of combinatorial geometry.

The following conditions must be met by the minimum zone straightness solution:

- The minimum zone envelop lines must contact at least three datapoints.
- These datapoints must be in a upper-line/lower-line/upper-line or a lower-line/upper-line/lower-line sequence.

The algorithm determining the substitute line and the straightness error consists of two main steps:

1 Computing the convex hull of the datapoints, resulting in a list of points in boundary traversal order.

2 Determination of the substitute line for the convex pointset and the width of the convex hull in one step.

For the determination of the convex hull of a planar pointset a number of algorithms have been invented in the recent past. Most widely known are: the gift wrapping, the quickhull invented by Preparata and Shamos, the Graham's scan, the divide and conquer and the incremental algorithms. While the Graham's scan provides the solution in optimal time, O(nlogn) it has no obvious extension to

three dimension. Therefore other solution having similar execution property was taken into consideration. It is the modified incremental algorithm:

Let $H_2 \leftarrow$ convex_hull$\{p_0,p_1,p_2\}$
for k = 3 to n-1 do
$H_k \leftarrow$convex_hull$\{H_{k-1} \cup p_k\}$

The first convex hull is the triangle p_0,p_1,p_2 . Let $Q=H_{k-1}$ and $p=p_k$. The computation of H_k falls into two cases: if $p_k \in Q$ (even on the boundary) it can be discarded. p_k is not in Q than the convex hull should be modified. Therefore we need only to find the two tangent lines from p to Q. p_i is a tangency point if for two subsequent edges the LeftOn test (on which side of the line the point is) gives different results. This is shown in the subsequent figure.

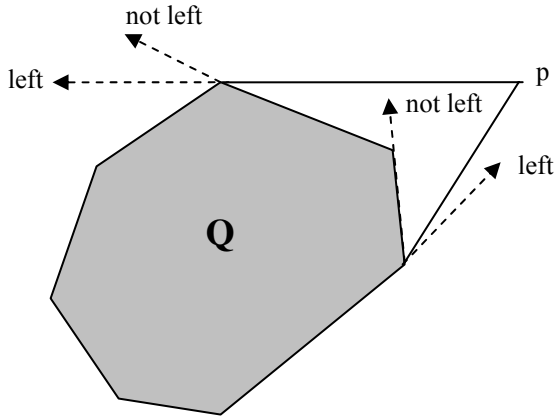


Figure 3
Detemination of the tangency lines from p to Q

The algorithm runs in $O(n^2)$ time. However by sorting the points by their x coordinates this can be decreased to O(nlogn).

Given the points of the convex hull the next step is the computation of the substitute line and the width of the convex hull. The boundary lines of the minimum zone are determined by three points: two of them defining one envelope line is colinear with one of the edges of the convex hull, the third one is a point with the maximum perpendicular distance to this line. The algorithm next does the job in $O(n^2)$ time, where n is the number of extreme points on the convex hull:

```
d=0
for i = 1 to n
        {c=0, j=i+2
        while j<2n dist(e_i,p_j)>d  {c= dist(e_i,p_j), j=j+1}
        if d<c {d=c, l=i, m=j-l}
        }
```

The substituting line is defined by the points $(p_l+p_m)/2$ and $(p_{l+1}+p_m)/2$. The envelop lines are given respectively by p_l, p_{l+1} and $p_l,p_m-p_l+p_{l+1}$.

Analogically the minimum zone flatness solution must satisfy the following conditions:

- The minimum zone lines must contact at least four datapoints.
- If the four contact points form a 2-2 model (two points contacting each plane), the line connecting the two upper plane points will intersect the line connecting the two lower plane points when the points are projected onto the same plane.
- If the four contact points form a 3-1 model (three points contacting one plane and one point contacting the other plane), the single point must be inside the triangle formed by the three points on the other plane when the points are projected onto the same plane.

From the above properties it be came obvious that the datapoints defining the minimum zones are vertex points on the convex hull of the original dataset. The straightness error is equal to the distance between the one data point and the line spanned by the two other points while the flatness error is given by the distance between the two lines (in case of the 2-2 model) or by the distance between the point and the plane determined by the three points (3-3 model).

The overall structure of the algorithm is the same as in two dimensions. In each step the computation of the new convex hull is again the main problem. If p is inside Q then it is discarded. If not the cone with apex p_i and tangent to Q is computed. It is clear that those faces are to be discarded which are visible from p_i.

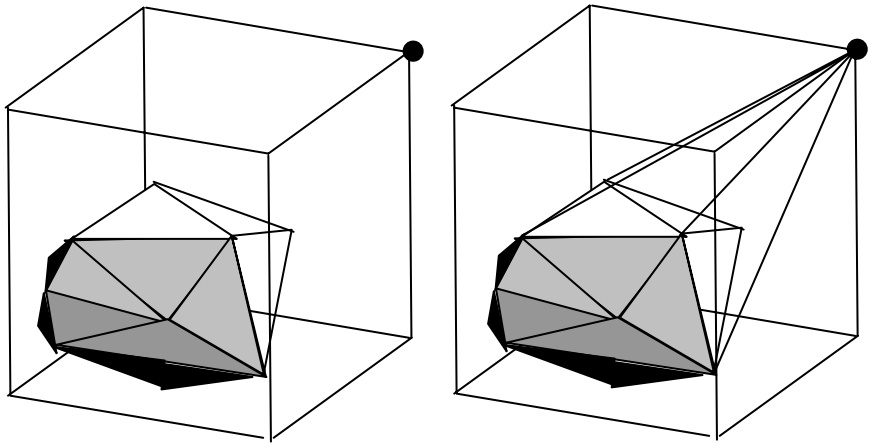


Figure 4
Convex hull before and after adding a point

The incremental convex hull algorithm is summerized as follows:

```
Initialize H3 to tetrahedron (p0,p1,p2,p3)
for i = 4 to n-1 do
for each face f of H_{i-1} do
        Mark f if visible
if no faces are visible
        then dicard p_i (it is inside H_{i-1})
        else
            for each border edge e of H_{i-1} do
                Construct cone face determined by e and p_i
            for each visible face f do
                Delete f
            Update H_i
```

Since the loops marking the visible faces and constructing the cones are inbedded inside a loop that iterates n times the complexity of the algorithm is $O(n^2)$.

Given the points of the convex hull the next step is the computation of the substitute planes and the width of the convex hull. The boundary planes of the minimum zone are determined by the 3-1 model: the minimum distance between the face defined by three points and one point with the maximum perpendicular distance to this line. The algorithm next does the job in $O(m^2)$ time, where m is the number of edges on the convex hull:

```
d=0
for i = 1 to n
        Compute the equation of the face f_i from it's cornerpoints
        for j = 1 to n
        Determine the distance of the point p_j from the face
        if  dist(p_j,f_i) > d then d = dist(p_j,f_i)

for i =1 to m
    for j = 1 to m
         if dist (e_i, e_j) > d then d = dist (e_i, e_j) > d
```

5 Implementation Issues

The computation of straightness is straightforward, it does not require any particular datastructure. In the three dimensional case the topology both of the polytope and the convex hulls should be explicitly represented in order to speed up the search processes. For its effectiveness the socalled quad-edge data structure invented by Guibas and Stolfi for the representation of any subdivision 2-manifolds was selected.

In the data structure each edge record is part of four circular lists. Two list are the lists for the endpoints and two lists reperesent the adjacent faces. Members of the list are linked by pointers. By introducing convention for the loops representing the faces boundary the direction of the surface normal vector are defined implcitly. Using this information the hidden faces can be easily be determinated and eliminated.

6 The Host Application

As the problem was raised in connection with the calibration on straight edges and on a coordinate measuring machine it was an obvious choise to implement the above described algorithm within the calibration software. It is written in Visual Basic and receives data from an Excel table. The output is the final calibration certificate while the data values and the evaluation result are archived.

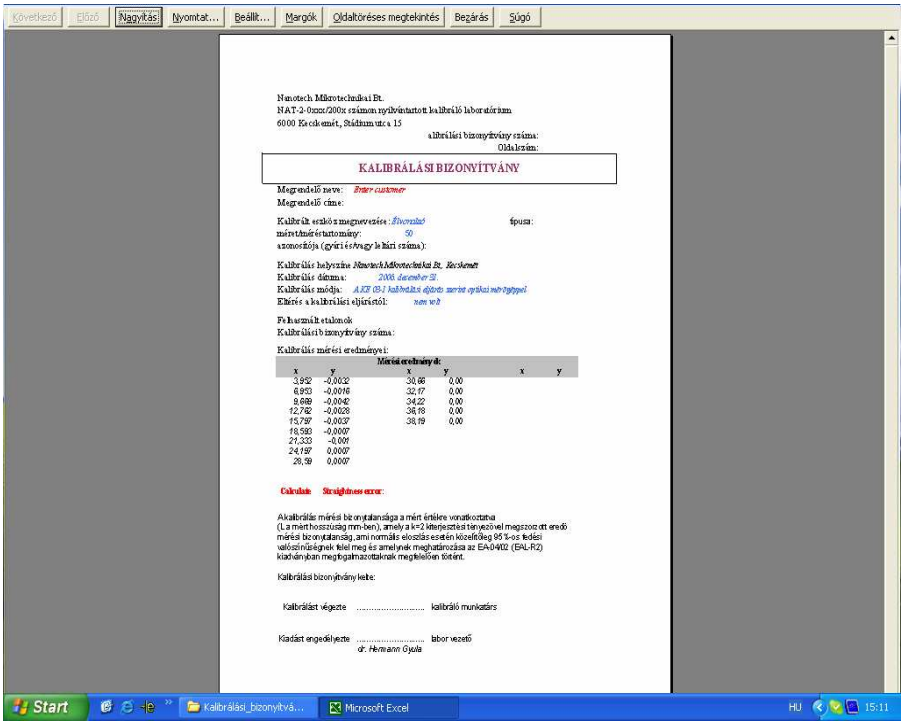


Figure 5
The user interface of the calibration software

Conclusions

The paper describes a simple robust algorithm for determining the straightness and flatness of surface features. The algorithm was implemented as part of a software package supporting the calibration of gauges and it is currently in usage. It's price/performance ratio is superior to other software currently available on the market.

References

[1] G. T. Anthony, H. M. Anthony, B. Bittner, B. P. Buttler, M. G. Cox, R. Driescher, R. Elligsen, A. B. Forbes, H. Gross, S. A. Hannaby, P. M.

Harris, J. Kok: Reference Software for Finding Chabysev Best-Fit Geometric Elements, Precision Engineering 19 (1996) 28-36

[2] ANSI/ASME Y14.5M Dimensioning and Tolerancing. New York, American Society of Mechanical Engineers, 1994

[3] K. Carr, P. Ferreira: Verification of Form Tolerances Part I: Basic Issues, Flatness, and Straightness, Precision Engineering 17 (1995) 144-156

[4] K. Carr, P. Ferreira: Verification of Form Tolerances Part II:Cylindricity and Srtraightness of a Median Line, Precision Engineering 17 (1995) 131-1

[5] L. J. Gubia, J. Stolfi: Primitives for the Manipulation of General Subdivisions and the Computation of Voronoi Diagrams, ACM Transactions on Graphics 4, 74-123

[6] J. Huang: Evaluation of Angular Error between Two Lines, Precision Engineering 27 (2003) 304-310

[7] Huang, K. C. Fan, J. H. Wu: A New Minimum Zone Methode for Evaluating Straightness Errors, Precision Engineering 15 (1993) 158-165

[8] M. J. Kaiser, K. K. Krishnan: Geometry of the Minimum Zone Flatness Functional: Planar and Spatial Case, Precision Engineering 22 (1998) 174-183

[9] M. K. Lee: A New Convex-Hull-based Approach to Evaluating Flatness Tolerance, Computer Aided Design 29 (1997) 861-868

[10] T. S. R. Murty, S. Z. Abdin: Minimum Zone Evaluation of Surfaces, Intenational Journal of Machine Tool Design and Research, 20 (1980) 123-136

[11] J. O'Rourke: Computational Geometry in C, Cambridge University Press 1998

[12] F. P. Preparata, M. Shamos: Computational Geometry, New York Springer Verlag, 1985

[13] G. L. Samuel, M. S. Shunmugam: Evaluation of Straightness and Flatness Error Using Computational Geometris Techniques, Computer Aided Design 31 (199) 829-843

[14] D. S. Suen, C. N. Chang: Application of Neural Network Interval Regression Methode for Minimum Zone Straightness and Flatness, Precision Engineering 20 (1997) 196-207

[15] Qing Zhang, K. C. Fan, Zhu Li: Evaluation Methode for Spatial Straightness Errors Based on Minimum Zone Condition, Precision Engineering 23 (1999) 264-272

Evolutionary Strategy for Feeding Trajectory Optimization of Fed-batch Reactors

Tamás Varga, Ferenc Szeifert, János Abonyi

Department of Process Engineering,University of Pannonia
Egyetem u. 10, H-8200 Veszprém, Hungary
e-mail: {vargat, szeifert, abonyij}@fmt.uni-pannon.hu

Abstract: Safe and optimal operation of complex production processes is one of the most important research and development problems in process engineering. This problem is the most relevant at the design of the optimal feeding profile of fed-batch chemical reactors due to the nonlinear and unstable dynamical behavior of the processes. This paper shows that how the optimal feeding policy can be determined in fed-batch reactors by sequential quadratic programming, classical evolutionary strategy (ES) and the advanced version of ES that is based on covariance matrix adaptation. A multi-objective function was created and the search space was constrained in case of all of the three applied algorithms. The switching times between states in the feeding trajectory and the feed rates in each state were manipulated to find the global minima of the objective function. To obtain the optimal feeding policy the first-principle model of a pilot fed-batch reactor was implemented in MATLAB and applied as a dynamic simulator of the process. Off-line optimization process was carried out in case of different dosing time distribution. As the results show a significant improvement can be achieved in process performance applying advanced ES based optimization algorithms to generate feeding trajectories.

Keywords: optimization, fed-batch reactor, quadratic programming, evolution strategy

1 Introduction

As optimal operating conditions of production processes are getting closer and closer to physical constraints, the development of knowledge based expert systems for supporting the operators to keep the operating conditions in this narrow range becomes more and more important [1, 2].

This problem is the most relevant at the design of the optimal feeding profile of fed-batch chemical reactors due to the nonlinear and unstable dynamical behavior of the involved processes, because in case of exothermic reactions thermal runaway can also occur [3-6]. Reactor runaway means a sudden and considerable change in the process variables that is a serious problem in many industrial

chemical processes, like oxidation and polymerization. For example in case of a highly exothermic reaction thermal runaway occurs when the reaction rate increases due to an increase in temperature, causing a further increase in temperature and hence a further increase in the reaction rate while the reactants are depleted.

Unfortunately the systems with low complexity are rare in production process, hence the analytical determination of optimal operation variables is rather complicated or in some cases it is impossible. An excellent summary of problems of the operation of batch and fed-batch reactors can be found in [7], where the author focuses on the safety, the product quality and the scale-up. To calculate the optimal temperature profile or the optimal feeding policy in such system a lot of optimization algorithm can be applied. Particle Swarm Optimization (PSO) algorithm is applied to estimate the optimal control feeding profile in a fed-batch fermentation process where heat effects have less importance in the changes of state-variables and cubic splines are used instead of linear ones to design a more smooth operation [8]. Different algorithms belonging to three main groups - Evolutionary Algorithms (EA), PSO and Differential Evolution (DE) - are applied to solve the same optimization problem and compared based on their performance in [9]. It is shown that PSO is the worst choice from the group of the compared optimization algorithms while they suggest applying DE algorithms instead of EA because of the fast convergence. These methods can be applied in those cases when the time of calculation is not an important aspect, e.g. to solve off-line or batch-to-batch optimization problems.

In this paper three optimization algorithms are used for the estimation of optimal feeding trajectory for a pilot-plant fed-batch reactor and the reached optimization performances are shown and compared at different distribution of feeding time. The paper is organized as follows: in Section 2 the applied optimization methods are briefly introduced; it is followed by the introduction of the investigated pilot-plant fed-batch system and its mathematical model. Further sections show the results and the comparison of optimization algorithms.

2 Applied Optimization Techniques

In these days sequential quadratic programming (SQP) methods are routinely applied to solve practical nonlinear programming problems. SQP methods are stabilized by a monotone line search procedure subject to a suitable merit function. In each iteration step of the SQP algorithm the solution of the quadratic program (QP) must be found. That makes the complex optimization problem simpler. The optimization algorithm has a quadratic objective function and applies linear constraints for manipulated variables. The SQP algorithm converges very rapidly, therefore it requires few iterations (hence few QP solutions) to find an

approximate solution with a good precision. The accuracy can be improved by using second derivatives in QP. The SQP algorithm is an appropriate approach to solve nonlinear optimization problems defined by the evaluation of functions and to calculate their derivatives is time consuming. Most industrial applications of SQP methods are found in optimization of mechanical structures [10-12].

The Evolutionary Algorithm (EA) is an optimization method which is based on the natural selection and survival of the fittest as it works in the biological world. EAs consistently perform well at approximating solutions for all types of problems because they do not make any assumption about the underlying fitness landscape that makes EAs applicable from different fields of engineering to social sciences [13]. EAs differ from more traditional optimization techniques in that they involve a search from a 'population' of solutions in each iteration step, instead of a single point. In each iteration step a competitive selection is made based on the 'fitness' value of solutions and the weak ones from the population are eliminated. The solutions with high fitness are 'recombined' with other solutions by swapping parts of a solution with another. Solutions are also 'mutated' by making a small change to a single element of the solution. Recombination and mutation are used to generate the new population of solutions that are diverted towards regions of the searching space where good solutions have already been seen.

However, it should be noted that EA has several different implementations: Evolutionary Programming (EP) which focuses on optimizing continuous functions without recombination, Evolutionary Strategy (ES) which focuses on optimizing continuous functions with recombination, Genetic Algorithm (GA) which focuses on optimizing general combinatorial problems and Genetic Programming (GP) which mainly evolves programs [10]. The selection of the proper technique and the tuning of the parameters of the selected technique require some knowledge about these techniques. In this article two different ES algorithms are applied to obtain the optimal feeding trajectory, a not modified one and an ES which applies covariance matrix adaptation (CMA) [14, 15].

Learning the covariance matrix in the CMA-ES is analogous to learning the inverse Hessian matrix in a quasi-Newton method. At the end, any convex-quadratic (ellipsoid) objective function is transformed into a spherical function. This can improve the performance on ill-conditioned and/or non-separable problems by orders of magnitude. The CMA-ES overcomes typical problems that are often associated with EAs, e.g. the poor performance on badly scaled and/or highly non-separable objective functions.

3 Fed-Batch Reactor and its Model

A highly exothermic reaction system was investigated and an optimal feeding policy was determined using SQP and ES techniques. First the investigated reactor and its first-principle model will be shortly introduced. The main product, 2-octanone is produced in a two-phase reaction system in a fed-batch reactor from 2-octanol. The 2-octanol is continuously fed into the organic phase, which does not exist before the start of the feeding. Hence, the organic phase is the dispersed one while the aqueous nitric acid phase where all the reaction steps take place is the continuous phase. The two phases are connected by the mass and heat transfer phenomena. Based on Castellan et al.'s work [16] which shows that the oxidation processes in room temperature mostly takes place according to ionic mechanism, Woezik and Westerterp [17] determined a reaction kinetic describing the way of oxidation of 2-octanol to form 2-octanone in nitric acid, which is proved to be feasible.

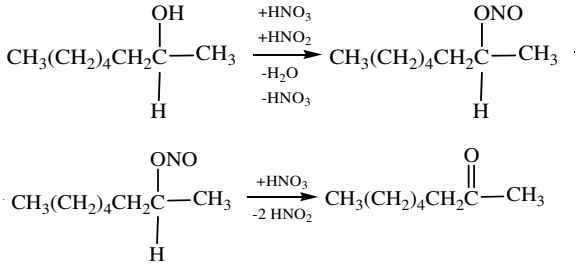
In the first step of this reaction mechanism the nitrosonium ion forms through this reaction:

$$HNO_3 + HNO_2 \rightleftharpoons NO^+ + NO_3^- + H_2O \tag{1}$$

The above reaction has a very long induction time, which can be shortened by adding a small amount an initiator like $NaNO_2$ to generate the necessary nitrous acid faster:

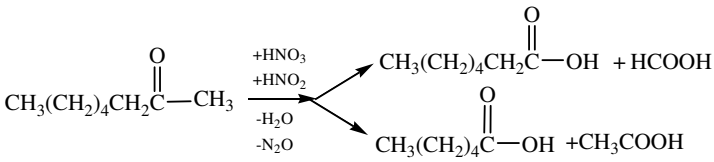
$$NaNO_2 + H_3O^+ \longrightarrow HNO_2 + Na^+ + H_2O \tag{2}$$

The nitrosonium ion forms only in the aqueous phase. It is followed by the oxidation of 2-octanol forming 2-octanone:



(3)

and the oxidation of 2-octanone producing different carboxylic acids:



(4)

This process can be considered as an advanced benchmark problem in the field of process engineering. In [18] the control of this process has been analyzed, while in [19] the safety issues related to its operation were discussed. A report how the simulator of this process can be used in process engineering education has been published in [19], and from the viewpoint of this paper the most relevant work is published in, where the Hazard and Operability Analysis (HAZOP) of this process is developed also based on the application of process simulator.

Woezik and Westerterp introduced a mathematical model based on a simplified form of the kinetic introduced above to predict the dynamic behavior of the reactor [14]. In the next part just a brief introduction will be given about this model; a well detailed description can be found in [17, 18]. The simplified reaction mechanism:

$$A + B \rightarrow 2P + B \tag{5}$$

$$P + B \rightarrow X \tag{6}$$

where A is 2-octanol, B is nitrosonium ion, P is 2-octanone and X represents all the byproducts. The rate of reactions are calculated by the following equations

$$r_1 = \left(1 - \varepsilon_{org}\right) \cdot k_{eff,1} \cdot m_A \cdot c_A^{org} \cdot c_B^{aq} \tag{7}$$

$$r_2 = \left(1 - \varepsilon_{org}\right) \cdot k_{eff,2} \cdot m_P \cdot c_P^{org} \cdot c_B^{aq} \tag{8}$$

where the effective reaction rate constants are depend on the temperature and the acidity of the aqueous phase, ie.

$$k_{eff,i} = k_{eff,i}^0 \cdot \exp\left(-\frac{E_{A,eff,i}}{R \cdot T^R} - m_{H_0,eff,i} \cdot H_0 \right) \tag{9}$$

where the H_0 represents the Hammett-acidity function. To describe the concentration trajectories during the operation the following ordinary differential equations must be solved

$$\frac{d\left(V^R \cdot c_A^{org}\right)}{dt} = B^{R,in} \cdot c_{A,in} - r_1 \cdot V^R \tag{10}$$

$$\frac{d\left(V^R \cdot c_B^{aq}\right)}{dt} = V^R \cdot \left(r_1 - r_2\right) \tag{11}$$

$$\frac{d\left(V^R \cdot c_P^{org}\right)}{dt} = V^R \cdot \left(r_1 - r_2\right) \tag{12}$$

$$\frac{d\left(V^R \cdot c_X^{org}\right)}{dt} = V^R \cdot r_2 \tag{13}$$

$$\frac{d\left(V^R \cdot c_N^{aq}\right)}{dt} = -V^R \cdot \left(r_1 + r_2\right) \tag{14}$$

To describe the change in the temperature of the reactor and the jacket the following equations must be solved

$$\frac{dT^R}{dt} = \frac{1}{HC^R} \cdot \left(Q_r + Q_{in}^R - Q_{cool} + Q_{stir} - Q_{loss}\right) \tag{15}$$

$$\frac{dT^C}{dt} = \frac{1}{HC^C} \cdot \left(Q_{in}^C + Q_{cool}\right) \tag{16}$$

To illustrate the complex dynamical behavior of the process the following two experiments are performed. Trajectories of state-variables on Fig. 1 show the normal operation of the reactor when 2-octanone is the main product and the total quantity of byproducts is low. A small, only 3 °C change in the inlet temperature of the jacket results in a reactor runaway (see on Fig. 2). In this case byproducts are mainly generated during the operation while the conversion of 2-octanol is significantly decreased at the end of the operation. On Figs. 1 and 2 grey areas represent the instability region of the process. In another work we analyzed the stability of the model of this reactor using Ljapunov's indirect stability analysis method [20]. Our ultimate goal is to find an optimal feeding trajectory when reactor runaway doesn't occur.

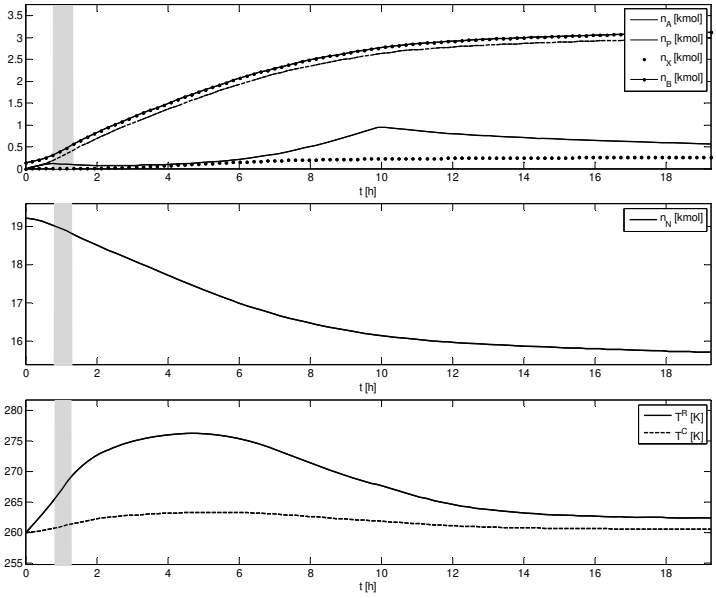


Figure 1
Trajectories of state variables (normal operation)

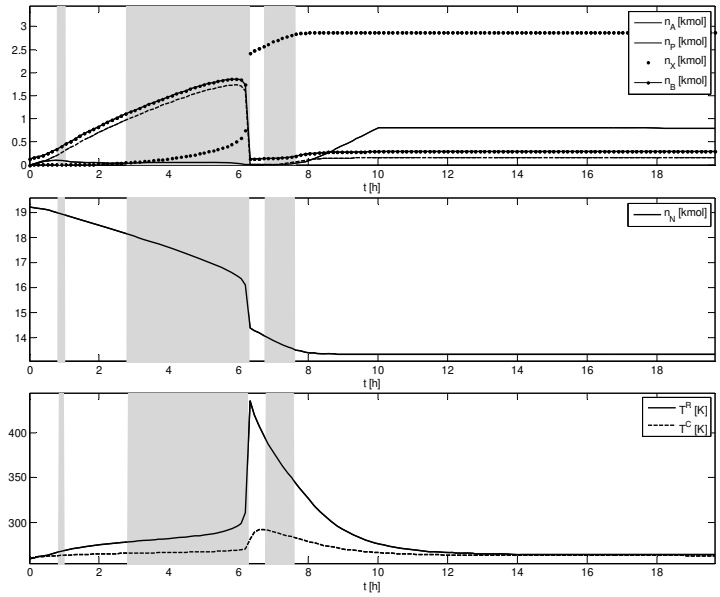


Figure 2
Trajectories of state variables (runaway occurs)

4 Optimization Results

In the objective function the purity of the product and the conversion of the 2-octanol was included with different weights. All the optimizations runs were stopped at the same tolerance for the change in objective function value. Initial values of the manipulated parameters were the same in every experiment.

Fig. 3 shows optimization results obtained with SQP. The feeding time domain is distributed into 4-9 parts. The value of the objective function can be seen next to the number of time segments. The first curve shows the trajectory of reactor temperature without optimization and applying constant feed rate. The best objective function value belongs to the fifth run, when the number of time segments is 8 while the worst belongs to the second run. As it shown the final value of the objective function is changing between -26.0 and -26.7 without convergence to any value of the objective function. As it can be seen in case of the second run (5 time segments) the algorithm can't find the global extrema. Only a small improvement can be achieved with the substitution of SQP with ES as it can be seen on Fig. 4. Otherwise, a significant difference can be noticed in the final value of the objective function at each run with the same time segments. The function value changes in a narrower range than previous one.

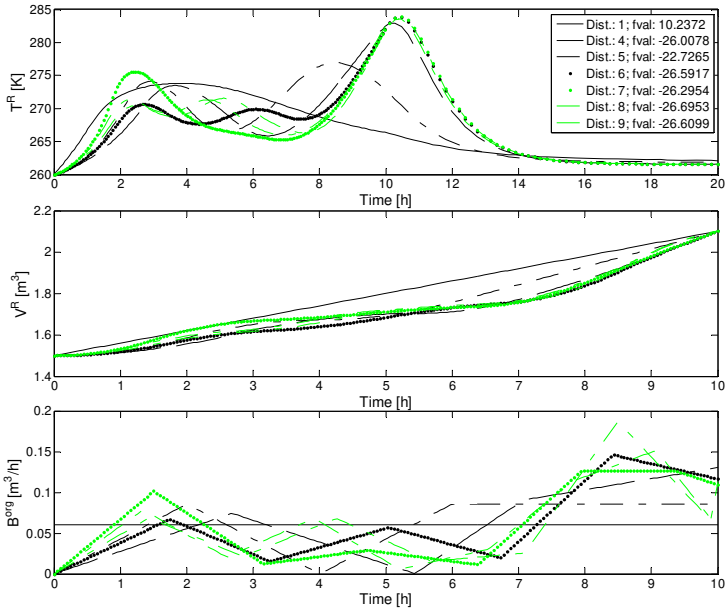


Figure 3
Optimization results with SQP

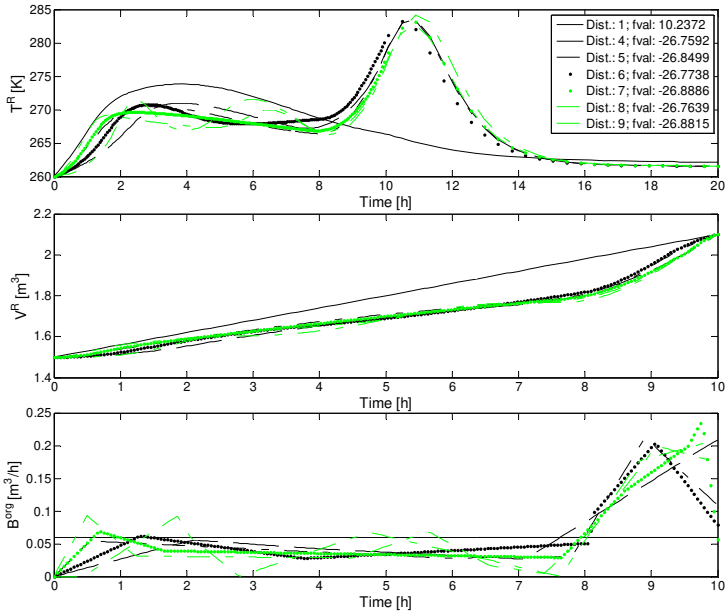


Figure 4
Optimization results with ES

The best results were obtained by the CMA-ES algorithm as it can be seen on Fig. 5. Beside to the 'highest' objective function values the CMA-ES prove the most reliable from the tested algorithms. The results for all methods are collected in Table 1. As it can be noticed two algorithms find the optimal number of the feeding time distribution at 8 and the other one at 7, thus it can be stated that it no need to further increase the distribution number in this optimization problem.

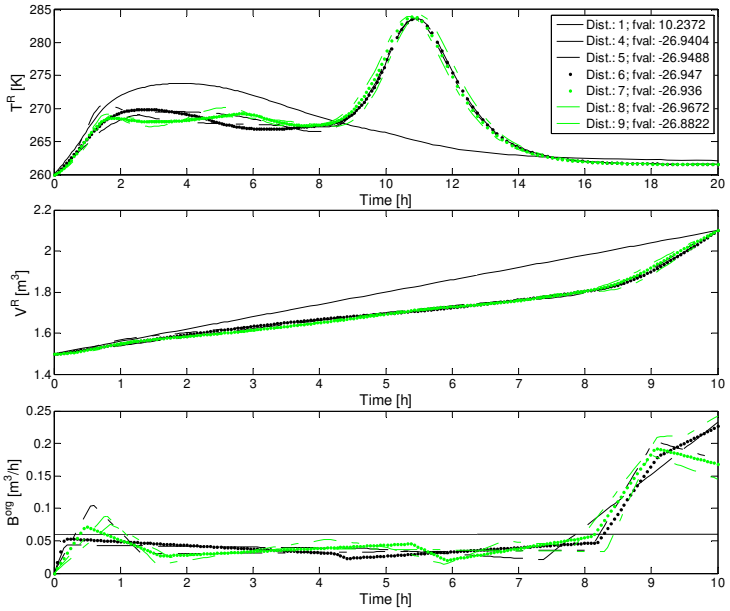


Figure 5
Optimization results with CMA-ES

Table 1
Optimization results with CMA-ES

Dist.	SQP	ES	CMA-ES
4	-26.0078	-26.7592	-26.9404
5	-22.7265	-26.8499	-26.9488
6	-26.5917	-26.7738	-26.9470
7	-26.2954	-26.8886	-26.9360
8	-26.6953	-26.7639	-26.9672
9	-26.6099	-26.8815	-26.8822

Conclusions

The operation of complex production processes is one of the most important research and development problems in process engineering. This problem is most relevant in the design of the optimal feeding profile of fed-batch chemical reactors

due to the nonlinear and unstable dynamical behavior of these processes. The optimal feeding profile was generated with SQP, ES and CMA-ES algorithms at different feeding time interval distribution. The obtained results show that the objective function value can't be significantly improved with ES and CMA-ES algorithms but the reliability of results is increased. The best results are achieved by the CMA-ES algorithm in the investigated optimization problem. From the structure of process model can be extracted some a priori information, which can be applied in the improvement of optimization performance and sped up the algorithm. In the following research the possible ways of utilizing this kind of extracted knowledge will investigated and will focus on how the extracted knowledge can be transformed into a set of constraints on the process variables and how these constraints can be applied in batch-to-batch optimization.

Acknowledgements

The authors acknowledge the financial support of the Cooperative Research Centre (VIKKK, project 2004-III/2), the Hungarian Research Found (OTKA 49534), the Bolyai János fellowship of the Hungarian Academy of Science, and the Öveges fellowship.

References

[1] C. S. Kao, K. H. Hu: Acrylic Reactor Runaway and Explosion Accident Analysis, Journal of Loss Prevention in the Process Industries, 2002, 9, pp. 213-222

[2] J. Albert, G. Luft: Runaway Phenomena in the Ethylene/Vinylacetate Copolymerization under High Pressure, Chemical Engineering and Processing, 1998, 37, pp. 55-59

[3] C. H. Barkelew: Stability of Chemical Reactors, Chemical Engineering Progress Symposium Series, 1959, 25, pp. 37-46

[4] J. Adler, J. W. Enig: The Critical Conditions in Thermal Explosion Theory with Reactant Consumption, Combustion and Flame, 1964, 8, pp. 97-103

[5] A. A. Lacey: Critical Behaviour of Homogenous Reacting Systems with Large Activation Energy, International Journal of Engineering Science, 1983, 21, 501-515

[6] C. Potter, S. Baron: Kinetics of the Catalytic Formation of Phosgene, Chemical Engineering Progress, 1951, 47(9), pp. 473-515

[7] D. Bonvin: Optimal Operation of Batch Reactors – a Personal View, Journal of Process Control, 1998, 8(6), 355-368

[8] A. Ismael, F. Vaz, E. C. Ferreira, A. M. T. Mota: Fed-Batch Trajectory Optimal Control Problems with Cubic Splines, WSEAS Transactions on Systems and Control, 2006, 2(1), pp. 207-213

[9] R. Mendes, M. Rocha, I. Rocha, E. C. Ferreira: A Comparison of Algorithms for the Optimization of Fermentation Processes, In Proceedings of the 2006 IEEE Conference on Evolutionary Computation, 2006, 7371-7378

[10] J. F. Bonnans, J. Ch. Gilbert, C. Lemaréchal, C. A. Sagastizábal: Numerical Optimization: Theoretical and Practical Aspects (2nd edition), Springer, 2006

[11] K. Schittkowski: Optimization in Industrial Engineering: SQP-Methods and Applications, report, *http://www.uni-bayreuth.de/departments/math/~kschittkowski/radioss.pdf*

[12] P. E. Gill, W. Murray, M. A. Saunders: SNOPT: An SQP Algorithm for Large-Scale Constrained Optimization, Society for Industrial and Applied Mathematics, 2005, 47, pp. 99-131

[13] A. E. Eiben, J. E. Smith: Introduction to Evolutionary Computing (2nd printing), Springer, 2007

[14] N. Hansen: Invariance, Self-Adaptation and Correlated Mutations in Evolution Strategies, In Proceedings of the 6th International Conference on Parallel Problem Solving from Nature, 2000, pp. 355-364

[15] N. Hansen: References to CMA-ES Applications, www.bionik.tu-berlin.de/user/niko/cmaapplications.pdf, 2005

[16] A. Castellan, J. C. J. Bart, S. Cavallaro: Nitric Acid Reaction of Cyclohexanol to Adipic Acid, Catal. Today, 1991, 9, pp. 255-283

[17] B. A. A. van Woezik, K. R. Westerterp: The Nitric Acid Oxidation of 2-Octanol. A Model Reaction for Heterogeneous Liquid-Liquid Reactions, Chem. Eng. and Proc., 2000, 39, pp. 521-537

[18] B. A. A. van Woezik, K. R. Westerterp: Runaway Behavior and Thermally Safe Operation of Multiple Liquid-Liquid Reactions in the Semi-Batch Reactor. The Nitric Acid Oxidation of 2-Octanol, Chem. Eng. and Proc., 2001, 41, pp. 59-77

[19] S. Eizenberg, M. Shacham, N. Brauner: Combining HAZOP with Dynamic Simulation – Applications for Safety Education, J. of Loss Prev. in the Proc. Ind., 2006, 19, pp. 754-761

[20] T. Varga, F. Szeifert, J. Réti, J. Abonyi: Analysis of the Runaway in an Industrial Heterocatalytic Reactor, Computer Aided Chemical Engineering, 2006, 24, pp. 751-756

Plausible Reasoning in Modular Robotics and Human Reasoning

Claudiu Pozna

University Transilvania of Brasov, Departament of Product Design and Robotics, Bd. Eroilor 28, 500036 Brasov, Romania
E-mail: cp@unitbv.ro

Radu-Emil Precup

Politehnica University of Timisoara, Department of Automation and Applied Informatics, Bd. V. Parvan 2, 300223 Timisoara, Romania
E-mail: radu.precup@aut.upt.ro

Abstract: Present paper continues the researches on cognitive system design. The goal of the paper is to illustrate the variety of models which can be constructed using the Bayesian plausible reasoning theory. The first case study develops a classical inverse kinematical model into a Bayesian model. The second case study models the human reasoning presented by the famous story of Sun Tzu: 'Advance to Chengang by a hidden path'.

Keywords: model, Bayesian theory, plausible reasoning

1 Introduction

Present paper continues the author's researches on cognitive system design. These researches have been started by a phenomenological analysis of AI collocation [1] and have continued by researches on modeling with Bayesian plausible reasoning. The goal of this paper is to illustrate the variety of phenomenon which can be modeled using the mentioned theory. For this reason we have structured our paper in three parts. The presentation starts with a briefly introduction of the plausible reasoning theoretical background. The second part illustrates the transformation of a classical inverse kinematics problem into a Bayesian model. The third part tried to model a human reasoning example. More precisely, our intention was to explain (by modeling) the famous story of Sun Tzu: 'Advance to Chengang by a hidden path'.

The Principles of Plausible Reasoning [6]:

1 The representation of degree of plausibility is given by the plausibility function:

$$p:\Phi \to [0 \quad 1];\ p(A \mid X) = y \tag{1}$$

where: Θ is a set of sentences; $p(A \mid X)$ is a continuous and monotonic function which associates a particularly degree of truth for the sentence A in the condition that sentence X is true;

2 The consistence of the commune sense requires the following property for the p function

$$p(AB \mid X) = p(A \mid X)p(B \mid AX) \tag{2}$$

$$p(A \mid X) + p(\neg A \mid X) = 1 \tag{3}$$

$$p(A + B \mid X) = p(A \mid X) + p(B \mid X) - p(AB \mid X) \tag{4}$$

$$p(A_i \mid X) = \frac{1}{n} \qquad i = 1...n \tag{5}$$

where $\{A_i\}|_{i=1...n}$ is a complete set of mutual excusive sentence

Some comments are necessary:

- by consistence we mean:
  - every possible way of reasoning a sentence must lead to the same result;
  - the equivalent sentences have an equal degree of plausibility;
- in order to obtain the degree of plausibility for a sentence we must take into account all the evidence available;
  - $p(AB \mid X)$ means the plausibility of sentence A **and** B in the condition that sentence X is true;
  - $\neg A$ means **non A;**
  - $p(A + B \mid X)$ means the plausibility of sentence A **or** B in the condition that sentence X is true;

Theoretical Results:

Analyzing the mentioned postulates, theoretical results can be deduced. From the beginning we will mention that because the probability function has the same properties (1…5) it can be accepted that the plausibility function is synonymous with the probability function. This is the only reasons that theoretical results from probability theory can be transferred to the theory of plausible reasoning [1].

We will resume presenting the Bayesian theorem which can easily deduce from (1-5). If we name by d the evidence of an experiment and by $h_{i=1...n}$ a set of mutual exclusive hypotheses the Bayesian theorem tells us that the plausibility of hypothesis h_i in the condition of evidence d is equal with the plausibility of hypothesis h_i multiplied by the plausibility of evidence d in the condition that hypothesis h_i is true and divided by the sum of the same product made for all the hypotheses of the set.

$$p(h_i \mid d) = p(h_i) \frac{p(d \mid h_i)}{\sum_{k=1...n} p(h_k) p(d \mid h_k)} \tag{6}$$

The plausibility of hypothesis h_i in the condition of evidence d is named the a posteriori knowledge, the plausibility of hypothesis h_i is named the a priori knowledge and the plausibility of evidence d in the condition that hypothesis h_i is trough is named the likelihood. The sum from the denominator is named the marginalization sum.

2 The First Case Study: the Robots Inverse Kinematics Problem

There are many methods to solve the inverse kinematics problem. A very good classification of these methods is presented in [2]. The inverse kinematics methods can be divided in:

1 Analytical, which allows the computation of all possible solutions:

   a) Closed – form, where the solutions can be expressed as a set a closed – form equations. This method is restricted to a 6 degree of freedom robots;

   b) Algebraic elimination based, where the results can be expressed as a solution to a system of multivariable polynomial equations;

2 Numerical, which converge iteratively to a single solution. This solution depends on the initial searching value:

   a) Newton – Raphson methods, where the solution is the root of a nonlinear equation. The convergence is slow and is sensitive to robot singularity;

   b) Resolved motion rate control method which is an improvement of Newton – Raphson method [3];

   c) Methods based on pseudo-inverse computation of the Jacobian, allow the computation for redundant manipulators, cases when the Jacobian is not square [4];

d) Control theory based methods, which transform the differential equation associated to the inverse kinematics problem in to a control problem [5];

e) Optimizations methods, which transform the problem in to a nonlinear optimization problem [6]

We can add to this list the Artificial Indigence methods. This direction can be exemplified by the Adaptive Neuro-Fuzzy solution presented by MathWorks. All the combinations: joint variables (inputs) – robot gripper positions and orientations (outputs) are computed with the direct kinematics methods. The results of this computation are training data collections. The learning algorithm which, use the training data, tunes the membership functions of a Sugeno-type Fuzzy Inference System using the training input-output data.

From [2] where the mentioned methods are analyzed we know that:

- The numerical methods can suffer from numerical instability which prevent an algorithm to converge to a solution;
- The numerical methods have poor reliability near the Jacobian singularities;
- The numerical methods can be generalized to solve additional constraint and objective function;
- The analytical method are restricted to 6 degree of freedom system when additional constraint are impose;
- The Newton–Raphson methods are the slowest and the analytical algorithms are the fastest;
- The Jacobian transpose method and the optimization method are sensitive to the local minima;

If we analyze the Adaptive – Neuro – Fuzzy solution we can remark the number of computation needed to obtain the training data collection. The proposed method belongs to the Artificial Intelligence and numerical methods because obtaining the inverse kinematics solution will be transformed into a decision problem and these decisions are based on numerical computations [12].

We will start our presentation by the inverse kinematics problem definition. For the robot structure presented in Figure 1 the mentioned problem input is the gripper position and orientation and the solution (output) is the joint variable value.

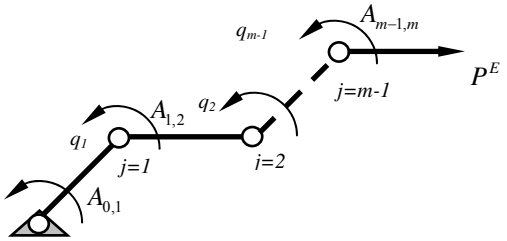


Figure 1
The robot structure; the input data P^E and the output data q_j

The Problem Definition

The mathematical description of the problem is (7):

$$P^E = \prod_{j=0}^{m-1} A_{j,j+1} \tag{7}$$

where: P^E is the homogenous matrix of the robot end point (gripper) position and orientation (the input data); $A_{j,j+1}$ are the homogenous transformation from joint j to joint $j+1$; m is the joint number.

We mention that matrixes $A_{j,j+1}$ depend on the links parameter and also on the joints variable q_j which are the problem solution (output data). In [10] there are more details on the homogenous matrixes construction when the robot configuration is modular.

Proposed Solution

We will start by mentioning that each actuator has a finite resolution. More precisely there are joint variable which are technical impossible to be realized.

For a 2 degree of freedom (DOF) robot the combination of possible joints variable combination can be expressed with the direct kinematical problem in (8).

$$\Phi\left(\begin{bmatrix} (q_0^1,q_1^1) & (q_0^1,q_1^2) & . & (q_0^1,q_1^{n_1}) \\ (q_0^2,q_1^1) & (q_0^2,q_1^2) & . & (q_0^2,q_1^{n_1}) \\ . & . & . & . \\ (q_0^{n_0},q_1^1) & (q_0^{n_0},q_1^2) & . & (q_0^{n_0},q_1^{n_1}) \end{bmatrix}\right) = \begin{bmatrix} (x,y)_{1,1} & (x,y)_{1,2} & . & (x,y)_{1,n_1} \\ (x,y)_{2,1} & (x,y)_{2,2} & . & (x,y)_{2,n_1} \\ . & . & . & . \\ (x,y)_{n_0,1} & (x,y)_{n_0,2} & . & (x,y)_{n_0,n_1} \end{bmatrix} \tag{8}$$

where: Φ is an operator which transforms the joint variable into the robot end point position.

The consequence of this observation is that there are positions of the robot end point which are technical impossible to be obtained by the robot and in these cases the only solution is to approximate them. The generalization of this observation means that for each desired position and orientation of the robot end point we must obtain a suitable combination of the joint variable. This combination will lead to an approximation of the desired position and orientation.

If we resume to the end point position, the approximation means that the distance between the robot end point $(x,y,z)_{i_1,i_2,\ldots i_{m-1}}$ and the desired point (input of the inverse kinematics problem) must accomplish the following condition.

$$d\left((x,y,z)_{i_0,i_2,\ldots i_{m-1}},(x,y,z)_D\right) \le \rho \tag{9}$$

where: ρ is the approximation distance, this means the radius of the desired point vicinity inside which the approximation is accepted (see Figure 2); $i_0,\ldots i_{m-1}$ are the joint variable index, m is the number of joints.

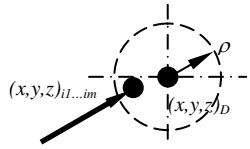


Figure 2
The desired point approximation

It is clear that we can transform the condition (9) into an optimization problem (10). In this case the appropriate combination of angle is the combination which minimizes the distance between the desired point and robot end point.

$$\min_{i_k \in 1\ldots n_i}\left(d\left((x,y,z)_{i_0,i_2,\ldots i_{m-1}},(x,y,z)_D\right)\right) \tag{10}$$

From this point of view the inverse kinematics problem can be transformed into a search problem. By searching we will analyze the joints variables combinations. More precisely the mentioned combinations will be transformed by direct kinematics method in to the robot end point positions, and then these positions will be used in (9) or (10). For this problem we must find a method which:

- Decreases the number of searching steps;
- Avoids the search divergence.

Our idea starts from this point. The proposed searching algorithm will be managed by the plausible reasoning theory. In order to do this we will use the Bayesian theorem. We will rewrite the mentioned theorem:

$$p(h_i^{+,-} \mid o) \propto p(h_i^{+,-}) p(o \mid h_i^{+,-}) \tag{11}$$

where: "$h_i^{+,-}$ "is the sentence: *"In order to obtain (9) the joint variable from joint i must be increased, (decreased)"*; "o" is the observation of approximation in term of (9) when the action of increasing (decreasing) tack place; $\propto$ means proportional; $i = 0...m-1$ is the current number of joints.

Some comments are necessary:

- Searching the appropriate combination for a desired point means an iterative process;
- The plausibility $p(h_i^{+,-})$ is an a priori known: for the initial step of the iteration we have no knowledge in order to consider one sentence more plausible the other. This means that all the sentence have the same plausibility;
- Usually the likelihood $p(o \mid h_i^{+,-})$ is an a priori knowledge, in this case we must compute, by direct kinematics method, the position of the robot end point and to compare this value with de desired point. Therefore we have defined the likelihood by (12):

$$p(o \mid h_i^{+,-}) = \frac{1}{d((x,y,z)_{i_0,i_2,\ldots i_{m-1}}, (x,y,z)_D)} \tag{12}$$

- The a posteriori plausibility $p(h_i^{+,-} \mid o)$ will be obtained by normalizing (11). In the next iteration step this plausibility will become the a priori plausibility;
- At each step (epoch) we must compute $p(h_i^{+,-} \mid o)$ these means 2xm plausibility computation. From this set we will choose the sentence which has the maximum plausibility and we will perform the increasing (decreasing) mentioned in this sentence;

- In order to avoid the searching divergence we will impose a maximum number of epochs. If condition (9) is accomplished, the searching algorithm will end. If not, the algorithm will perform the maximum number of epochs and will output the joints variables which lead to the minimum distance between the desired point and the robot end point;
- A possible divergence cause is the fact that all possible end points are out from the vicinity of the desired point (see Figure 3). For this situation usually we say that the desired point is out of the working volume. Actually we deal with a set of possible points '*points cloud*'. These points intersects or not the mentioned vicinity;

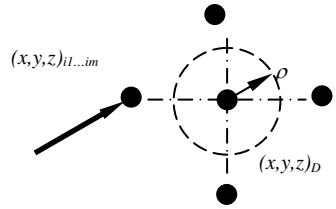


Figure 3
The desired point is out of the *points cloud*

- A desired trajectory is a set of desired points. The algorithm performs searching for each desired point and output the joints variables. Performing the direct kinematics for these joints variable we will obtain a trajectory which approximate the desired trajectory;
- In order to increase the smooth of the approximation we have develop our algorithm by introducing a dynamic vicinity of the desired points. More precisely if the algorithm converges for radius ρ_i we will perform a new search for radius $\rho_{i+1}=0.5\ \rho_i$. The dynamic vicinity search will end when a divergence occurs or when the radius is smaller than an imposed value, this means $\rho_k<\rho_{min}$.

The Algorithm

If we summarize these considerations we can propose the following algorithm for the inverse kinematics problem:

1 The initial (input) data are:

- the robot configuration, here this means the necessary data for compute the homogenous transformations ($A_{j,j+1}$);
- the robot resolution for each joint ($\Delta_{0\ldots m-1}$);
- the approximation radius (ρ);

- the initial position of the robot ($q_{0\ldots m-1}$);
- a set of desired points $(x,y,z)_{D0\ldots r}$;
- the epochs number (N_{epochs});

2 For each desired point a searching algorithm can have N_{epochs}, and start with:

- setting the a priori plausibility: $p(h_i^{+,-})$=1/2m, *i=0…m-1;*

3 The searching algorithms epoch means:

- computing the likelihood by (12), for all *2*x*m* sentences; the end point position is computed by (1) for $q_i = q_i \pm \Delta_i$;
- computing the a posterior plausibility with (11) for each *2*x*m* sentences;
- finding the maximum plausibility $p(h_k^{+,-} \mid o) = \max_{i=1\ldots m} p(h_i^{+,-} \mid o)$;
- set the angel values: $q_i = q_i; i = 0\ldots m-1; i \neq k \quad q_k = q_k \pm \Delta_k$;
- set the a priori plausibility: $p(h_i^{+,-}) = p(h_i^{+,-} \mid o)$;
- if condition (9) is accomplished then:
  - set $q_{i,j}^S = q_i$ where $q_{i,j}^S$ means the solution for joint *i* at the desired point *j*; $i = 0\ldots m-1; j = 0\ldots r$ *;*
  - go to the next desired point;
- if condition (9) is not accomplished then
  - set $q_i^{PS,k} = q_i$; $d^k = d$; $q_i^{PS,k}$ means the possible solution at epoch k , *k=1…* N_{epochs}
  - start a new searching epoch;

4 If after N_{epochs} the condition (9) is not accomplished then set $q_{i,j}^S = q_i^{PS,l}$ where $d^l = \min_{k=1\ldots N_{epochs}}(d^k)$;

5 In the end the problem solution (output) consists on a set of angular values for each desired point:

$$Q = \left\{ q_{i,j}^S \mid j = 0\ldots r; i = 0\ldots m-1 \right\}$$

In the next section results simulation of the proposed algorithm are presented.

The Simulations

The proposed algorithm was transformed into a computer program. We have used this program for several robots configurations (several DOF) and several desired trajectories.

After obtaining the inverse kinematics solution the program animates the robot for each desired point. In this way we have observed the continuity of the robots movements.

In the simulation we have used Δ_i=0.005[rad]; ρ=3[mm]; N_{Epochs}=100;

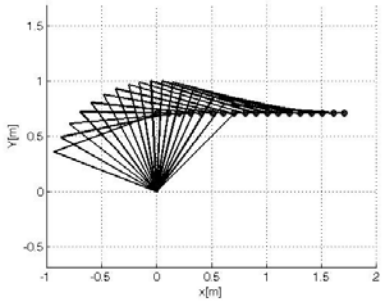


Figure 4
Simulation for a 2 DOF robot

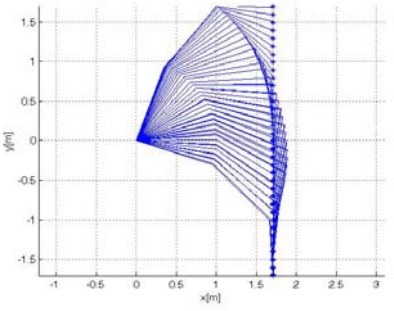


Figure 5
Simulation for a 3 DOF robot

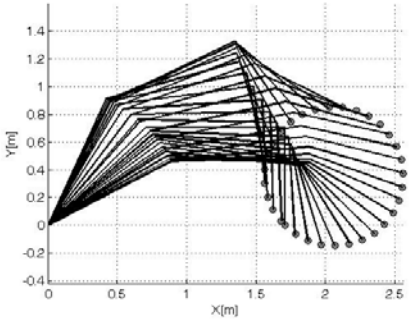


Figure 6
Simulation for a circlular desired trajectory

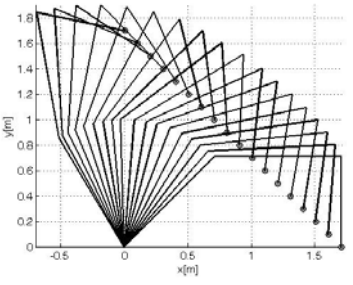


Figure 7
Simulation for a sloping desired trajectory

In Figure 4 a 2DOF robot which performs a horizontal trajectory is presented. In Figure 5 a 3 DOF robot performs a vertical trajectory. This trajectory starts at (1.7,0) goes to (1.7,-1.7) returns at the first point and end at (1.7,1.69).

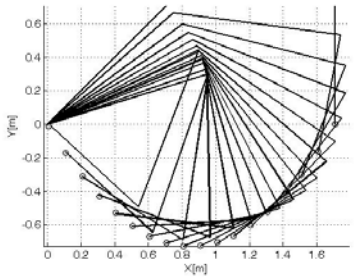


Figure 8
Simulation for a parabolic desired trajectory

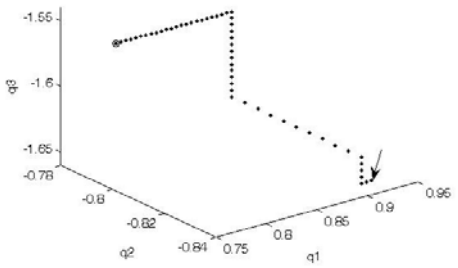


Figure 9
An epoch: the convergence

In Figure 6, 7 and 8 the 3DOF robot performs a circular, sloping, respectively a parabolic trajectory. In the end, in Figure 9 the convergence processes is presented. More precisely the robot joints variables starting point is marked with "o" and during an epoch these variables converges to the final joints variables, marked by the arrow. The final joints variables are the solutions $q_{i,j}^{S}$ and will be memories in the problem solution set.

Conclusions

Present work continues the research program, about the modular robots by proposing a new solution for the inverse kinematics problem. This is a numerical solution based in plausible reasoning theory that is the Bayesian theorem. The mentioned theorem was adapted for a searching algorithm in order to obtain an accepted approximation of the desired robot end point.

The proposed solution's limitations are that we haven't considered the possible conditions (avoiding obstacles etc.) and the fact that the solution depends on the initial values. In future works we intend to generalize the method for optimization problems and also to replace the desired points (xyz) position by the analytical form of the desired trajectory.

3 The Second Case Study: a Strategically Problem

The intention of the third case study is to prouve the ability of Plausible Reasoning in human reasoning modeling. For this pourpouse we will tray to model the famous story of Sun Tzu: 'Advance to Chencang by a hidden path'.

The story that we intend to explain by Bayesian model is the following:

This stratagem took place towards the end of the Qin dynasty. Xiang Yu appointed Liu Bang as king of Hanzhong, effectively making him leave China. To further ensure that Liu Bang does not return to China from the East, Xiang Yu divided

Guanzhong into three principalities and put three people in charge, informing them to be alert against Liu Bang.

Liu Bang said, 'In order to placate Xiang Yu and the three kings, we must destroy the mountain plank road to show that we've no intention of returning to China.'

After nine years of preparations, Liu Bang's army became powerful and was ready to march eastwards. Liu Bang ordered his generals to take 10,000 men and horses and repair the plank road within three months.

Meanwhile, his enemies were greatly perturbed. One of the kings even led his forces to block the plank road exit.

Liu Bang then led his generals and several thousand troops to overrun Guanzhong by the old roundabout route through Chencang.

We intend to model this story by using the Bayesian theorem (6). At first sight the victory of Liu Bang is based on his ability to increase the plausibility of the likelihood that he will attack on the plank road.

If we analyze more deeply the story we will find that there are two stage of the conflict: the first when Liu Bang must decide about the reaction concerning the Xiang Yu actions, and the second when Liu Bang shows his attack intention but he must choose the attack direction.The story scenario is presented in Figure 10. It can be see that in the first stage of the conflict, by destroying the road Liu Bang have increased the peace (non attack) likelihood and in this way manipulate Xiang You. In the second stage of the conflict by restoring the road Liu Bang have increased the mountain direction attack (Am) and manipulate once again his enemy. From mathematical point of view this scenario can be describe in the following way:

- in the initial moment Xiang You can not decide the intention of Liu Bang:
  - $P(A) = P(\neg A) = 50\%$; where A is the sentence 'Liu Bang will attack'
- after seeing that Liu Bang destroyed the road Xiang You decides that:
  - $P(O_1 \mid \neg A) > P(O_1 \mid A)$; where O_1 is the observation of the destroyed road;
  - in consequence (6) $P(\neg A \mid O_1) > P(A \mid O_1)$;
- in the initial moment Xiang You can not decide the attack direction of Liu Bang:
  - $P(Am) = P(\neg Am) = 50\%$; where Am is the sentence 'Liu Bang will attack from the mountain';
- after seeing that Liu Bang constructs the road Xiang You decides that:

- $P(O_2 \mid Am) > P(O_2 \mid \neg Am)$; where O_2 is the observation of the constructed road;
- in consequence (6) $P(Am \mid O_2) > P(\neg Am \mid O_2)$

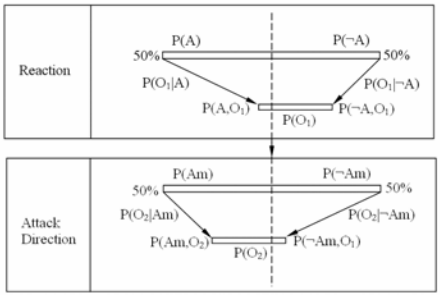


Figure 10
The story scenari

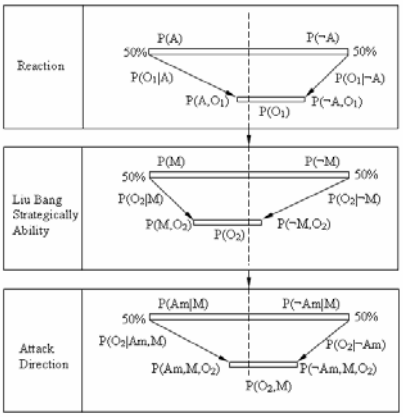


Figure 11
A posible solution

The famous story can be continued with a problem: have had Xiuag You the chance to react at his opponent ability? There are several solutions of this problem the first consist on increasing the number of hypothesis of attack direction and find new observations (spy). The second solution is presented in Figure 11 and is based on changing the causal network by introducing a new decision step. More precisely it can be see that after the second observation O_2 Xiuang You becomes able to decide the tactic that Liu Bang will use. This observation increases the likelihood that his opponent uses his ability to manipulate him.

From mathematical point of view this solution can be described in the following way:

- in the initial moment Xiang You can not decide the intention of Liu Bang:
  - $P(A) = P(\neg A) = 50\%$; where A is the sentence 'Liu Bang will attack'
- after seeing that Liu Bang destroyed the road Xiang You decides that:
  - $P(O_1 \mid \neg A) > P(O_1 \mid A)$; where O_1 is the observation of the destroyed road;
  - in consequence (6) $P(\neg A \mid O_1) > P(A \mid O_1)$;
- in the initial moment Xiang You can not decide about the strategically ability of his opponent:

- $P(M) = P(\neg M) = 50\%$; where M is the sentence 'Liu Bang is able to manipulete'

- after seeing that Liu Bang intends to attack Xiang You decides that:
  - $P(O_2 \mid M) > P(O_2 \mid \neg M)$; where O_2 is the observation of the constructed road; M is the sentence 'Liu Bang is able to manipulate'
  - in consequence (6) $P(M \mid O_2) > P(\neg M \mid O_2)$;

- in the initial moment Xiang You can not decide the attack direction of Liu Bang:
  - $P(Am) = P(\neg Am) = 50\%$; where Am is the sentence 'Liu Bang will attack from the mountain';

- after seeing that Liu Bang constructs the road, and knowing that his opponent can manipulate Xiang You decide that:
  - $P(O_2 \mid Am, M) < P(O_2 \mid \neg Am, M)$;
  - in consequence (6) $P(Am \mid O_2, M) < P(\neg Am \mid O_2, M)$

Conclusions

The second case study illustrates the ability of the plausible reasoning in modeling the human reasoning. It was presented the analysis of a famous strategically story and the synthesis of a problem. New solutions obtained by developing the causal network are very attractive.

4 Final Conclusions

Present paper continues the author researches on cognition system design by presenting two different case studies which use the same theory: the Bayesian theory of plausible reasoning. The first case study develops classical inverse kinematics problems in to a Bayesian model by proposing a searching algorithm. In the second case study end we have tray to model by Bayesian theory one of the famous stories of Sun Tzu.

Acknowledgement

This research work is supported by the Romanian Ministry of Education and Research trough CNCSIS project 895/2007.

References

[1] C. Pozna, A Phenomenological Analysis Trial of the AI syntagma, In Proceedings of the 7th International Symposium of Hungarian Researchers, Budapest 2006, pp. 159-165

[2] D. Tolani, A. Goswami, N. I. Badler, Real-Time Inverse Kinematics Techniques for Anthropomorphic Limbs Graphical Models 62(5): pp. 353-388, 2000

[3] D. E. Whitney, Resolved Motion Rate Control of Manipulators and Human Prostheses, IEEE Trans. Man–Machine Systems 10, pp. 47-63, 1969

[4] C. Klein, C. Huang, Review of Pseudoinverse Control for Use with Kinematically Redundant Manipulators, IEEE Trans. Systems Man Cybernet. 7, pp. 868-871, 1997

[5] L. Sciavicco, B. Siciliano, Modeling and Control of Robot Manipulators, McGraw–Hill, NewYork, 1996

[6] J. Zhao, N. Badler, Inverse Kinematics Positioning Using Nonlinear Programming for Highly Articulated Figures, Trans. Comput. Graph. 13(4), pp. 313-336, 1994

[7] Jaynes, E. T., Probability Theory with Application in Science and Engineering, Washington University, 1974

[8] http://yudkowsky.net/bayes/bayes.html Yudkowsky, E., An Intuitive Explanation of Bayesian Reasoning

[9] J. Pearl. Causality: Models, Reasoning, and Inference, Cambridge Univ. Press, New York, 2000

[10] C. Pozna, Modular Robots Kinematics, Journal of Applied Science at Budapest Tech, Vol. 3, Nr. 2, pp. 5-18, 2007

[11] Rudas, I., Evolutionary Operators New Parametric Type Operators Family, Fuzzy Sets and Systems 23, pp. 149-166