

52. IWK

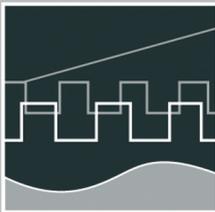
Internationales Wissenschaftliches Kolloquium
International Scientific Colloquium



PROCEEDINGS

10 - 13 September 2007

FACULTY OF COMPUTER SCIENCE AND AUTOMATION



COMPUTER SCIENCE MEETS AUTOMATION

VOLUME I

Session 1 - Systems Engineering and Intelligent Systems

Session 2 - Advances in Control Theory and Control Engineering

**Session 3 - Optimisation and Management of Complex
Systems and Networked Systems**

Session 4 - Intelligent Vehicles and Mobile Systems

Session 5 - Robotics and Motion Systems



Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der deutschen
Nationalbiografie; detaillierte bibliografische Daten sind im Internet über
<http://dnb.d-nb.de> abrufbar.

ISBN 978-3-939473-17-6

Impressum

Herausgeber: Der Rektor der Technischen Universität Ilmenau
Univ.-Prof. Dr. rer. nat. habil. Peter Scharff

Redaktion: Referat Marketing und Studentische Angelegenheiten
Kongressorganisation
Andrea Schneider
Tel.: +49 3677 69-2520
Fax: +49 3677 69-1743
e-mail: kongressorganisation@tu-ilmenau.de

Redaktionsschluss: 20. Juli 2007

Verlag: 
Technische Universität Ilmenau/Universitätsbibliothek
Universitätsverlag Ilmenau
Postfach 10 05 65
98684 Ilmenau
www.tu-ilmenau.de/universitaetsverlag

Herstellung und
Auslieferung: Verlagshaus Monsenstein und Vannerdat OHG
Am Hawerkamp 31
48155 Münster
www.mv-verlag.de

Layout Cover: Torsten Weilepp

Bezugsmöglichkeiten: Verlagshaus Monsenstein und Vannerdat OHG
Am Hawerkamp 31
48155 Münster
www.mv-verlag.de

© Technische Universität Ilmenau (Thür.) 2007

Diese Publikationen und alle in ihr enthaltenen Beiträge und Abbildungen sind
urheberrechtlich geschützt.

Preface

Dear Participants,

Confronted with the ever-increasing complexity of technical processes and the growing demands on their efficiency, security and flexibility, the scientific world needs to establish new methods of engineering design and new methods of systems operation. The factors likely to affect the design of the smart systems of the future will doubtless include the following:

- As computational costs decrease, it will be possible to apply more complex algorithms, even in real time. These algorithms will take into account system nonlinearities or provide online optimisation of the system's performance.
- New fields of application will be addressed. Interest is now being expressed, beyond that in "classical" technical systems and processes, in environmental systems or medical and bioengineering applications.
- The boundaries between software and hardware design are being eroded. New design methods will include co-design of software and hardware and even of sensor and actuator components.
- Automation will not only replace human operators but will assist, support and supervise humans so that their work is safe and even more effective.
- Networked systems or swarms will be crucial, requiring improvement of the communication within them and study of how their behaviour can be made globally consistent.
- The issues of security and safety, not only during the operation of systems but also in the course of their design, will continue to increase in importance.

The title "Computer Science meets Automation", borne by the 52nd International Scientific Colloquium (IWK) at the Technische Universität Ilmenau, Germany, expresses the desire of scientists and engineers to rise to these challenges, cooperating closely on innovative methods in the two disciplines of computer science and automation.

The IWK has a long tradition going back as far as 1953. In the years before 1989, a major function of the colloquium was to bring together scientists from both sides of the Iron Curtain. Naturally, bonds were also deepened between the countries from the East. Today, the objective of the colloquium is still to bring researchers together. They come from the eastern and western member states of the European Union, and, indeed, from all over the world. All who wish to share their ideas on the points where "Computer Science meets Automation" are addressed by this colloquium at the Technische Universität Ilmenau.

All the University's Faculties have joined forces to ensure that nothing is left out. Control engineering, information science, cybernetics, communication technology and systems engineering – for all of these and their applications (ranging from biological systems to heavy engineering), the issues are being covered.

Together with all the organizers I should like to thank you for your contributions to the conference, ensuring, as they do, a most interesting colloquium programme of an interdisciplinary nature.

I am looking forward to an inspiring colloquium. It promises to be a fine platform for you to present your research, to address new concepts and to meet colleagues in Ilmenau.



Professor Peter Scharff
Rector, TU Ilmenau



Professor Christoph Ament
Head of Organisation

Table of Contents

CONTENTS

	Page
1 Systems Engineering and Intelligent Systems	
A. Yu. Nedelina, W. Fengler DIPLAN: Distributed Planner for Decision Support Systems	3
O. Sokolov, M. Wagenknecht, U. Gocht Multiagent Intelligent Diagnostics of Arising Faults	9
V. Nissen Management Applications of Fuzzy Control	15
O. G. Rudenko, A. A. Bessonov, P. Otto A Method for Information Coding in CMAC Networks	21
Ye. Bodyanskiy, P. Otto, I. Pliss, N. Teslenko Nonlinear process identification and modeling using general regression neuro-fuzzy network	27
Ye. Bodyanskiy, Ye. Gorshkov, V. Kolodyazhniy, P. Otto Evolving Network Based on Double Neo-Fuzzy Neurons	35
Ch. Wachten, Ch. Ament, C. Müller, H. Reinecke Modeling of a Laser Tracker System with Galvanometer Scanner	41
K. Lüttkopf, M. Abel, B. Eylert Statistics of the truck activity on German Motorways	47
K. Meissner, H. Hensel A 3D process information display to visualize complex process conditions in the process industry	53
F.-F. Steege, C. Martin, H.-M. Groß Recent Advances in the Estimation of Pointing Poses on Monocular Images for Human-Robot Interaction	59
A. González, H. Fernlund, J. Ekblad After Action Review by Comparison – an Approach to Automatically Evaluating Trainee Performance in Training Exercise	65
R. Suzuki, N. Fujiki, Y. Taru, N. Kobayashi, E. P. Hofer Internal Model Control for Assistive Devices in Rehabilitation Technology	71
D. Sommer, M. Golz Feature Reduction for Microsleep Detection	77

F. Müller, A. Wenzel, J. Wernstedt A new strategy for on-line Monitoring and Competence Assignment to Driver and Vehicle	83
V. Borikov Linear Parameter-Oriented Model of Microplasma Process in Electrolyte Solutions	89
A. Avshalumov, G. Filaretov Detection and Analysis of Impulse Point Sequences on Correlated Disturbance Phone	95
H. Salzwedel Complex Systems Design Automation in the Presence of Bounded and Statistical Uncertainties	101
G. J. Nalepa, I. Wojnicki Filling the Semantic Gaps in Systems Engineering	107
R. Knauf Compiling Experience into Knowledge	113
R. Knauf, S. Tsuruta, Y. Sakurai Toward Knowledge Engineering with Didactic Knowledge	119
2 Advances in Control Theory and Control Engineering	
U. Konigorski, A. López Output Coupling by Dynamic Output Feedback	129
H. Toossian Shandiz, A. Hajipoor Chaos in the Fractional Order Chua System and its Control	135
O. Katernoga, V. Popov, A. Potapovich, G. Davydau Methods for Stability Analysis of Nonlinear Control Systems with Time Delay for Application in Automatic Devices	141
J. Zimmermann, O. Sawodny Modelling and Control of a X-Y-Fine-Positioning Table	145
A. Winkler, J. Suchý Position Based Force Control of an Industrial Manipulator	151
E. Arnold, J. Neupert, O. Sawodny, K. Schneider Trajectory Tracking for Boom Cranes Based on Nonlinear Control and Optimal Trajectory Generation	157

K. Shaposhnikov, V. Astakhov The method of ortogonal projections in problems of the stationary magnetic field computation	165
J. Naumenko The computing of sinusoidal magnetic fields in presence of the surface with bounded conductivity	167
K. Bayramkulov, V. Astakhov The method of the boundary equations in problems of computing static and stationary fields on the topological graph	169
T. Kochubey, V. Astakhov The computation of magnetic field in the presence of ideal conductors using the Integral-differential equation of the first kind	171
M. Schneider, U. Lehmann, J. Krone, P. Langbein, Ch. Ament, P. Otto, U. Stark, J. Schrickel Artificial neural network for product-accompanied analysis and control	173
I. Jawish The Improvement of Traveling Responses of a Subway Train using Fuzzy Logic Techniques	179
Y. Gu, H. Su, J. Chu An Approach for Transforming Nonlinear System Modeled by the Feedforward Neural Networks to Discrete Uncertain Linear System	185
 3 Optimisation and Management of Complex Systems and Networked Systems	
R. Franke, J. Doppelhammer Advanced model based control in the Industrial IT System 800xA	193
H. Gerbracht, P. Li, W. Hong An efficient optimization approach to optimal control of large-scale processes	199
T. N. Pham, B. Wutke Modifying the Bellman's dynamic programming to the solution of the discrete multi-criteria optimization problem under fuzziness in long-term planning	205
S. Ritter, P. Bretschneider Optimale Planung und Betriebsführung der Energieversorgung im liberalisierten Energiemarkt	211
P. Bretschneider, D. Westermann Intelligente Energiesysteme: Chancen und Potentiale von IuK-Technologien	217

Z. Lu, Y. Zhong, Yu. Wu, J. Wu WSReMS: A Novel WSDM-based System Resource Management Scheme	223
M. Heit, E. Jennenchen, V. Kruglyak, D. Westermann Simulation des Strommarktes unter Verwendung von Petrinetzen	229
O. Sauer, M. Ebel Engineering of production monitoring & control systems	237
C. Behn, K. Zimmermann Biologically inspired Locomotion Systems and Adaptive Control	245
J. W. Vervoorst, T. Kopfstedt Mission Planning for UAV Swarms	251
M. Kaufmann, G. Bretthauer Development and composition of control logic networks for distributed mechatronic systems in a heterogeneous architecture	257
T. Kopfstedt, J. W. Vervoorst Formation Control for Groups of Mobile Robots Using a Hierarchical Controller Structure	263
M. Abel, Th. Lohfelder Simulation of the Communication Behaviour of the German Toll System	269
P. Hilgers, Ch. Ament Control in Digital Sensor-Actuator-Networks	275
C. Saul, A. Mitschele-Thiel, A. Diab, M. Abd rabou Kalil A Survey of MAC Protocols in Wireless Sensor Networks	281
T. Rossbach, M. Götze, A. Schreiber, M. Eifart, W. Kattanek Wireless Sensor Networks at their Limits – Design Considerations and Prototype Experiments	287
Y. Zhong, J. Ma Ring Domain-Based Key Management in Wireless Sensor Network	293
V. Nissen Automatic Forecast Model Selection in SAP Business Information Warehouse under Noise Conditions	299
M. Kühn, F. Richter, H. Salzwedel Process simulation for significant efficiency gains in clinical departments – practical example of a cancer clinic	305

D. Westermann, M. Kratz, St. Kümmerling, P. Meyer Architektur eines Simulators für Energie-, Informations- und Kommunikationstechnologien	311
P. Moreno, D. Westermann, P. Müller, F. Büchner Einsatzoptimierung von dezentralen netzgekoppelten Stromerzeugungsanlagen (DEA) in Verteilnetzen durch Erhöhung des Automatisierungsgrades	317
M. Heit, S. Rozhenko, M. Kryvenka, D. Westermann Mathematische Bewertung von Engpass-Situationen in Transportnetzen elektrischer Energie mittels lastflussbasierter Auktion	331
M. Lemmel, M. Schnatmeyer RFID-Technology in Warehouse Logistics	339
V. Krugljak, M. Heit, D. Westermann Approaches for modelling power market: A Comparison.	345
St. Kümmerling, N. Döring, A. Friedemann, M. Kratz, D. Westermann Demand-Side-Management in Privathaushalten – Der eBox-Ansatz	351
 4 Intelligent Vehicles and Mobile Systems	
A. P. Aguiar, R. Ghabchelloo, A. Pascoal, C. Silvestre , F. Vanni Coordinated Path following of Multiple Marine Vehicles: Theoretical Issues and Practical Constraints	359
R. Engel, J. Kalwa Robust Relative Positioning of Multiple Underwater Vehicles	365
M. Jacobi, T. Pfützenreuter, T. Glotzbach, M. Schneider A 3D Simulation and Visualisation Environment for Unmanned Vehicles in Underwater Scenarios	371
M. Schneider, M. Eichhorn, T. Glotzbach, P. Otto A High-Level Simulator for heterogeneous marine vehicle teams under real constraints	377
A. Zangrilli, A. Picini Unmanned Marine Vehicles working in cooperation: market trends and technological requirements	383
T. Glotzbach, P. Otto, M. Schneider, M. Marinov A Concept for Team-Orientated Mission Planning and Formal Language Verification for Heterogeneous Unmanned Vehicles	389

M. A. Arredondo, A. Cormack SeeTrack: Situation Awareness Tool for Heterogeneous Vehicles	395
J. C. Ferreira, P. B. Maia, A. Lucia, A. I. Zapaniotis Virtual Prototyping of an Innovative Urban Vehicle	401
A. Wenzel, A. Gehr, T. Glotzbach, F. Müller Superfour-in: An all-terrain wheelchair with monitoring possibilities to enhance the life quality of people with walking disability	407
Th. Krause, P. Protzel Verteiltes, dynamisches Antriebssystem zur Steuerung eines Luftschiffes	413
T. Behrmann, M. Lemmel Vehicle with pure electric hybrid energy storage system	419
Ch. Schröter, M. Höchemer, H.-M. Groß A Particle Filter for the Dynamic Window Approach to Mobile Robot Control	425
M. Schenderlein, K. Debes, A. Koenig, H.-M. Groß Appearance-based Visual Localisation in Outdoor Environments with an Omnidirectional Camera	431
G. Al Zeer, A. Nabout, B. Tibken Hindernsvermeidung für Mobile Roboter mittels Ausweichecken	437
5 Robotics and Motion Systems	
Ch. Schröter, H.-M. Groß Efficient Gridmaps for SLAM with Rao-Blackwellized Particle Filters	445
St. Müller, A. Scheidig, A. Ober, H.-M. Groß Making Mobile Robots Smarter by Probabilistic User Modeling and Tracking	451
A. Swerdlow, T. Machmer, K. Kroschel, A. Laubenheimer, S. Richter Opto-acoustical Scene Analysis for a Humanoid Robot	457
A. Ahranovich, S. Karpovich, K. Zimmermann Multicoordinate Positioning System Design and Simulation	463
A. Balkovoy, V. Cacenkin, G. Slivinskaia Statical and dynamical accuracy of direct drive servo systems	469
Y. Litvinov, S. Karpovich, A. Ahranovich The 6-DOF Spatial Parallel Mechanism Control System Computer Simulation	477

V. Lysenko, W. Mintchenya, K. Zimmermann 483
Minimization of the number of actuators in legged robots using
biological objects

J. Kroneis, T. Gastauer, S. Liu, B. Sauer 489
Flexible modeling and vibration analysis of a parallel robot with
numerical and analytical methods for the purpose of active vibration damping

A. Amthor, T. Hausotte, G. Jäger, P. Li 495
Friction Modeling on Nanometerscale and Experimental Verification

Paper submitted after copy deadline

2 Advances in Control Theory and Control Engineering

V. Piwek, B. Kuhfuss, S. Allers 503
Feed drivers – Synchronized Motion is leading to a process optimization

A. Yu. Nedelina / W. Fengler

DIPLAN: Distributed Planner for Decision Support Systems

INTRODUCTION

An analysis of the problems in the efficiency management reveals the necessity to introduce new intelligent technologies to the management process. The management of complex objects and processes of various natures in the context of rigid real-time requirements is a complicated problem. This is in particular the case in applications like energy and power management, water resource management, nuclear waste management, reactor dynamics, and fuel management etc.

Artificial Intelligence (AI) planning technologies form the basis of the intelligent management, because questions, which are related to the planning theory as a component of the task solution theory, occupy one of the leading positions in the field of intelligent systems research.

Planners are Artificial Intelligence applications for plans generating – optimal or admissible sequences of actions allowing to achieve the desired goal or to solve the problem situations. These programs are in many cases embedded into real time decision support systems (RTDSS) or analytical information systems that must show real-time behavior. Since this request usually can't be met by a complete systematic search within a search space of all possible solutions, such programs are usually equipped with heuristic functions and other types of guidance, in order to respond promptly.

A planning problem solution in complicated dynamic environments with real-time constraints requires the use of an integrated approach with parallel computing that allows to reduce the time of the problem solution sufficiently.

In the paper there are presented the concept and prototype of DIPLAN planner based on a distributed approach for solving AI planning problem.

PLANNING AS STATE-SPACE SEARCH

Most planning problems can be considered as searches through the space of world states.

Planning domains are characterized by a current world state or current state of the intellectual system environment along with regulations for state transitions within a state space. In DIPLAN, world states are described by boolean state variables (fluents).

A planning domain D is defined by the collection $D = \langle F, S, A, R \rangle$, where

- F is a finite set of fluents;
- $S \subseteq 2^F$ is a finite set of states;
- A is a finite set of actions, $A \subseteq M = \{M_1, \dots, M_n\}$ of RTDSS systems;
- $R \subseteq S \times A \times S$ is a transition relation.

Action $a \in A$ is executable in $s \in S$ if $\exists s' \in S : R(s, a, s')$.

A state space problem P for a planning domain D is defined by the collection

$$P = \langle D, I, G \rangle, \text{ where}$$

- $I \subseteq S$ is an initial state;
- $G \subseteq S$ is the set of goal states.

A plan π for a planning problem P in a planning domain D is a sequence of state-action pairs that transforms the initial state $s_0 \in I$ into one of the goal states $\in G$, where

- $\{(s, a) : s \in S, a \in A, a \text{ executable in } s\}$;
- $\exists (s_0, a) \in \pi : s_0 \in I, a \in A$;
- $\exists R(s, a, s') : s \in S, a \in A, s' \in G$.

PLANNING APPROACH

The proposed distributed planner (DIPLAN) introduced here, concerns state-space planning systems.

The state space in DIPLAN is represented as a tree $T = (V, E)$, where V is a set of nodes and E is a set of arcs (Fig. 1). Each node in the tree is a state of environment characterized by a number of parameters. Root of the tree is an initial state of environment. Arcs are labeled by environment activities that transform one state to another.

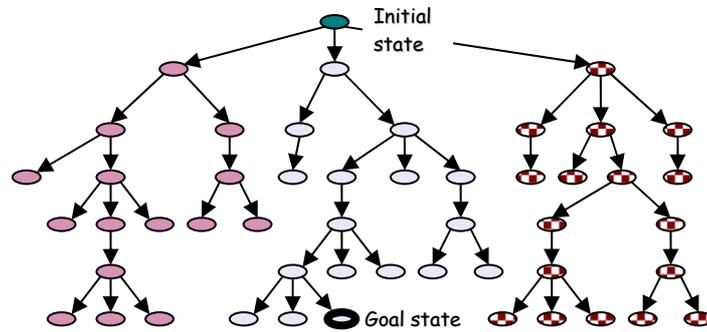


Figure 1. The DIPLAN state-space search tree

In that case, the state-space planning problem in DIPLAN is a search through the state-space for an optimal (shortest) path from initial state to some desired goal state of the goal state set.

The DIPLAN planning process flow diagram is presented on the Fig.2.

Basic methods and algorithms of state-space search for a planning system have been analyzed in the context of this research work [1,2]. The heuristic search algorithms allow to improve the efficiency of search by means of problem reduction. Hence, the current research work focuses on the heuristic search algorithms.

Unfortunately, the time calculating and memory requirements decrease the applicability of the considered heuristic methods in software tools such as RTDSS.

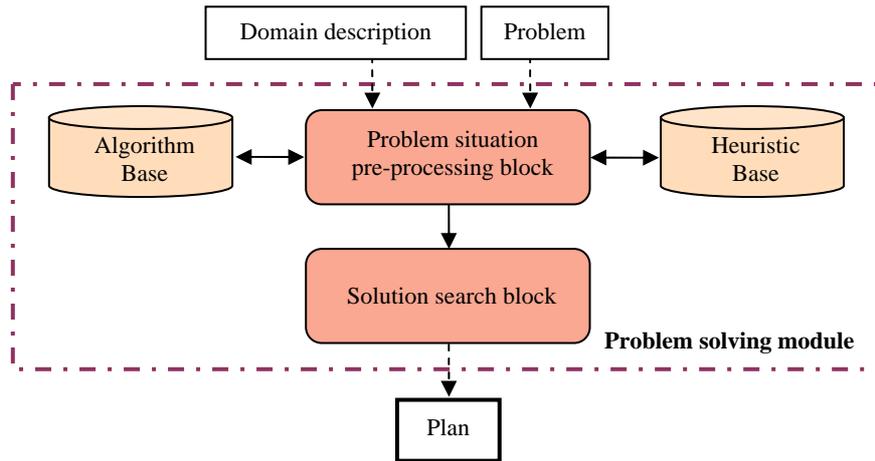


Figure 2. The scheme of DIPLAN blocks interaction

The following solution was proposed in the paper: parallel modifications of the heuristic search algorithms, optimized for the parallel hardware architecture [3]. Distributed computation approach allows to solve the planning problem by decomposition of a search problem into independent search spaces (OR- subtrees). Each search subtree may be allocated to separate processor (Fig. 1).

A challenging feature of the modern Artificial Intelligence applications is the ability to distribute the workload among several processors, in order to increase the execution speed. Although the technology of parallel architectures is quite mature and a large number of parallel systems are available at a reasonable cost, there are not many software products that can exploit these capabilities.

The classification of parallel hardware architectures is given on Fig. 3.

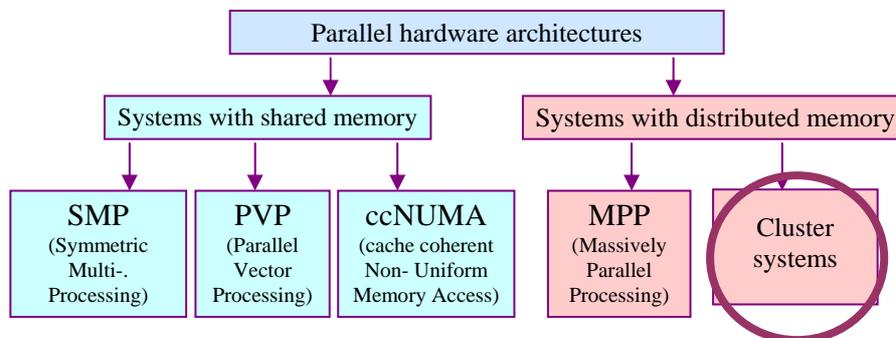


Figure 3. The classification of parallel systems by memory architecture

The cluster systems are chosen for parallel implementation of DIPLAN planner due to best correlation between cost / performance [4].

DIPLAN PROTOTYPE

The DIPLAN architecture contains the following modules (Fig. 4):

- **Interface module** – uses graphical tools for data representation;
- **Information processing module** – responsible for input data gathering, goal and task specifications, and also for interaction with the external systems;
- **Solution search module**- provides a state-space search using a decomposition of the general task to local disjunctive subtasks, distributes

the subtasks between the processors, it also controls the lifecycle of the processors (i.e. it can generate and modify the task and initial data).

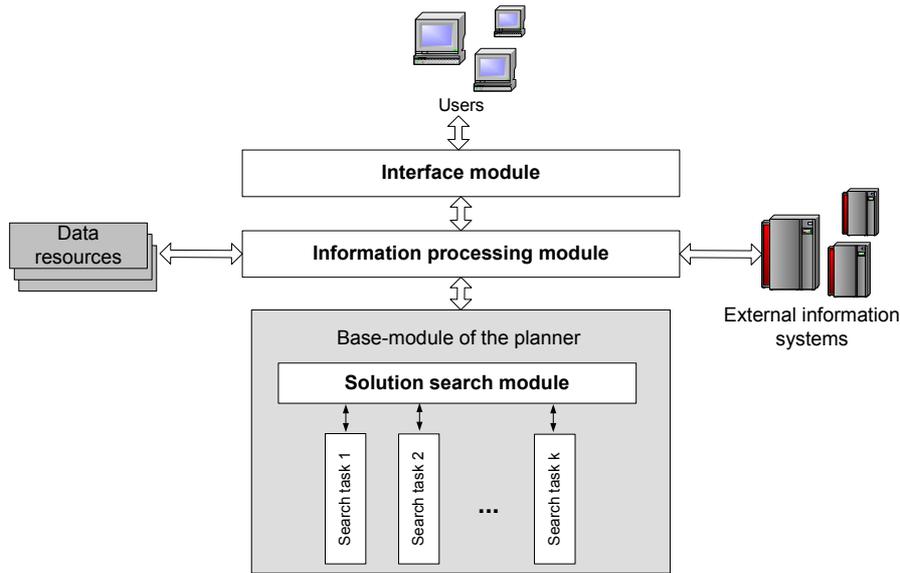


Figure 4. The generalized architecture of the DIPLAN planner

As shown in Fig. 4, the main functional component of DIPLAN is the problem solving module, which based on a distributed-memory Multiple Instruction Multiple DATA (MIMD) model.

The distributed environment of the Solution search module of DIPLAN is based on message-passing model with the usage of the communication package MPI (Message Passing Interface) for cluster systems [5]. During the parallel processing of solution search several individual processes work in a distributed system and communicate via data streams to perform a common task [6].

For the problem solving module was proposed and implemented parallel version of IDA* (Iterative-Deepening A*) algorithm. This algorithm is described in details in [3].

The DIPLAN was developed using C++ in the Microsoft Visual C++ 6.0 environment. The Solution search module was implemented in Visual C++ 6.0 with the usage of a parallel MPI technology and the package mpich.nt.1.2.5.[5], in a mode of parallel computation modeling on one processor.

EMPIRICAL RESULTS

For measuring the performance of solution search module we take a benchmark model – the 15-Puzzle [7,8]. It is simple, but has a combinatorial large problem space of 10^{13} states. Finding a solution to the 15-puzzle in terms of the state-space problem is a finding an optimal (shortest) solution path in a tree of possible puzzle configurations from an initial to goal state of the puzzle.

We used our parallel IDA* algorithm with the classic heuristical evaluation function for the sliding tile puzzles, the Manhattan distance [2]. This evaluation function estimates a sum of the minimum displacement of each tile from its goal position.

The average of iterations, received on base of the empiric results on a 100 experiments, is about 53 iterations for the domain of the 15-Puzzle problem. The 15-Puzzle problem was solve using IDA* on one processor and using parallel IDA* version on 2 and 4 processors. The results of a 100 experiments for Intel Pentium IV 2,8 GHz divided to the 5 sets (20 experiments in each set) are given on Fig.5.

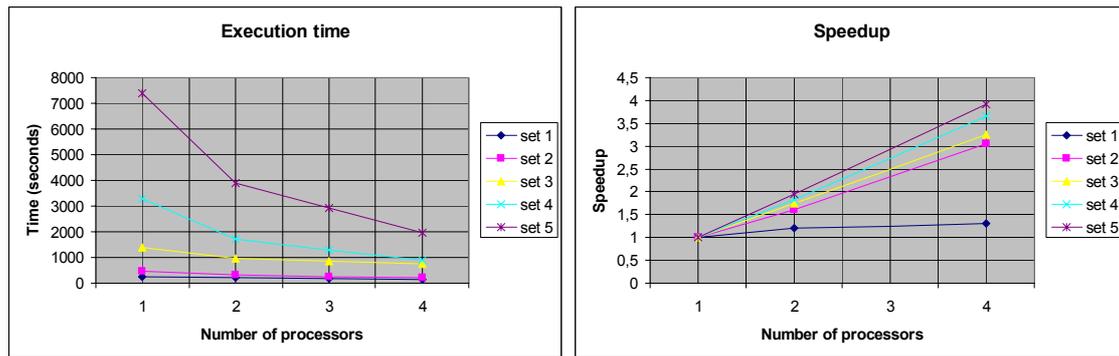


Figure 5. The diagrams of speedup and time execution for parallel IDA* version

CONCLUSION

The testing of the parallel IDA* algorithm for 15 Puzzle problem with the Manhattan distance heuristic has improved the speedup results in comparison with the consecutive version.

Later on, the developed distributed planner will be implemented on a cluster environment.

Furthermore, some real benchmark data will be tested for DIPLAN planner.

The possibilities of implementing the Solution search module of DIPLAN on symmetric multi-processing (SMP) systems will be analyzed.

References:

- [1] N. Nilsson, "Principles of Artificial Intelligence", Tioga Publishing Company, Palo Alto, CA, 1980.
- [2] R.E. Korf, "Depth-first iterative-deepening: An optimal admissible tree search", Artificial Intelligence 27, 1985, pp. 97-109.
- [3] A.P. Ereemeyev, A.Yu. Nedelina, "Methods and algorithms of strategic planning for decision support systems", Proc. of the 6 Joint Conference on Knowledge-Based Software Engineering, IOS Press, 2004, pp. 247-252.
- [4] R. Buyya, "High Performance Cluster Computing: Architectures and Systems", Prentice Hall PTR, NJ, USA, 1999.
- [5] W. Gropp, E. Lusk, R. Thakur, "Using MPI-2: advanced features of the message-passing Interface", MIT Press, Cambridge, USA, 1994.
- [6] M. Knapik, J. Johnson, "Developing intelligent Agents for distributed systems", McGraw-Hill, New York, USA, 1998.
- [7] A. Felner, R.E. Korf, S. Hanan, "Additive Pattern Database Heuristics", J. of Artificial Intelligence Research 22, 2004, pp. 279- 318.
- [8] A. Yu. Nedelina, A. P. Ereemeev, W. Fengler, "Development of distributed planner for decision support systems", Proc. of the 2nd International Conference on Dependability of Computer Systems (DepCoS - RELCOMEX 2007), IEEE, 2007, pp. 324-330.

Authors:

Dipl.-Ing. Anna Nedelina
 Prof. Dr.-Ing. habil. Wolfgang Fengler
 Moscow Power Engineering Institute (Technical University)
 Russian Federation, Moscow
 Phone: +7 916 3269342
 Fax: +7 495 3174663
 E-mail: anna@sytech.ru

O. Sokolov / M. Wagenknecht / U. Gocht

Multiagent Intelligent Diagnostics of Arising Faults

Decision Support Systems

This paper is devoted to fault diagnosis problems using decision making mechanism based on multiagent approach. We investigate dynamic systems which can be diagnosed by symptom-fault rule bases. The main question to be answered is what faults produce observable symptoms in the first moments of their appearance. To solve this task we propose to use a set of agents for making hypothesis about symptoms and coming to an agreement.

1 Symptom-fault model

We use $F = \{f_l\}_{l=1}^{N_F}$ and $S = \{s_m\}_{m=1}^{N_S}$ to denote the finite sets of all possible faults and symptoms, respectively. There is also multi valued mapping

$$\psi : S \rightarrow F \quad (1)$$

that can be presented as a **binary diagnostic matrix** like in Table 1. This matrix reflects the influence of relations between elements of sets S and F as numbers from $I = \{0,1\}$.

Table 1: Example of binary diagnostic matrix

S / F	f_1	...	f_{N_F}
s_1	1	...	0
...	0	...	1
s_{N_S}	1	...	1

Let a set of agents are responsible for observing of symptoms and making hypothesis about possible faults [2].

The main mission of agents in diagnostic task is to perceive the symptoms on first stages of their arising and recognize caused faults.

Let the set of agents given by $A = \{a_i\}_{i=1}^{N_A}$, $N_A \leq N_F$ and let exists a map

$$\xi : A \rightarrow F \quad (2)$$

Let each agent is responsible for several faults but one fault can be recognized by only one agent. The way in which the set F is divided by agents from A is a state of art of the subject diagnostic area. The natural manner is to divide faults according to subsystems of diagnostic objects [5].

2 Multi-agent diagnosis model

Let $S_o \subseteq S$ be a subset of symptoms observed by some management application. It reduces to the generation of an appropriate logical expression from (1) in Conjunctive Normal Form, i.e.

from $\psi|_{S_o} : S \rightarrow F$ we can get

$$\sigma = \bigvee_{i=1}^{D(S_o)} \left(\bigwedge_{j \in C_i(S_o)} f_j \right) \quad (3)$$

where $D(S_o)$ is a number of disjunctions in the formula, $C_i(S_o)$ is a set of indexes in the i^{th} conjunction expression.

Formula (3) gives the *minimum number of disjunctions* of faults that cause the set of observed symptoms. It is a necessary condition for getting S_o from diagnostic matrix.

Let $C = \{a_{1_c}, \dots, a_{N_c}\}$ be a set of agents responsible for σ . For each agent a_{j_c} there is a set of appropriate faults $Af_{j_c} = \{f_{1_{j_c}}, \dots, f_{N_{j_c}}\} \subset F$.

We propose the following heuristic recurrent procedure for choosing the set of testing faults from each element of $F_C = \{Af_{1_c}, \dots, Af_{N_c}\}$.

Let $\sigma_0 = \bigvee_{i=1}^{D_0} \left(\bigwedge_{j \in C_i^0} f_j \right)$ be an initial formula (3) where $D_0 = D(S_o)$ is the number of disjunctions in the formula; $C_i^0 = C_i(S_o)$ is the set of indexes in the i^{th} conjunction expression. Let $F_0 = \{F_{1_c}^0, \dots, F_{N_c}^0\} = F_C$ be the set of possible faults that caused the observed symptoms distributed on agents from C . Further we propose the procedure that performs by each agent Ag from $C = \{a_{1_c}, \dots, a_{N_c}\}$ at the k^{th} step simultaneously. We call this an α -**procedure**.

1. Refresh

$$\sigma_k = \bigvee_{i=1}^{D_k} \left(\bigwedge_{j \in C_i^k} f_j \right)$$

according to the result of faults checking at the previous step and abbreviate by cancellation excessive variables and false conjunctions (executed one time at each step).

2. **If** $F_{Ag}^k \neq \emptyset$ **then** choose one fault f_{Ag} from F_{Ag}^k that belongs to the conjunction in (3) with minimal length $\min \text{card}(C_i^k)$. If there are several conjunctions with the same length we apply a random choice. Set $F_{Ag}^{k+1} = F_{Ag}^k \setminus f_{Ag}$ **else stop** testing process for the current agent.
3. Check f_{Ag} and set $f_{Ag} = \{true, false\}$.
4. Set global state $k = k + 1$.

The proposed algorithm is converging due to the reliable set of symptoms, the truthful binary diagnosis matrix and the step-by-step reducing of the set of possible faults.

It is possible to improve the procedure and do not consider all agents simultaneously but one after another. In cooperation logic such model is called β -coalition contrary to the α -coalition for the case described above [5]. The β -coalition foresees the conditional coalition when the current decision strongly depends on the previous one.

The β -**procedure** is as follows.

1. Determine

$$\sigma_k = \bigvee_{i=1}^{D_k} \left(\bigwedge_{j \in C_i^k} f_j \right)$$

from the previous step or after modification at the current step.

2. **If** $F_{Ag}^k \neq \emptyset$ **then** choose one fault f_{Ag} from the set F_{Ag}^k that belongs to the conjunction in (3) with minimal length $\min \text{card}(C_i^k)$. Set $F_{Ag}^{k+1} = F_{Ag}^k \setminus f_{Ag}$ **else** stop testing the process for the current agent.
3. **If** $\min \text{card}(C_i^k) = 1$ and $F_{Ag}^{k+1} = \emptyset$ **then** set $f_{Ag} = \text{true}$ without checking. **else** check f_{Ag} and set the appropriate value $f_{Ag} = \{\text{true}, \text{false}\}$.
4. Refresh

$$\sigma_k^{\text{mod}_{Ag}} = \bigvee_{i=1}^{D_k} \left(\bigwedge_{j \in C_i^k} f_j \right)$$
 according to the value $f_{Ag} = \{\text{true}, \text{false}\}$ and abbreviate by cancelling excessive variables and false conjunctions.
5. Send to the next agent (from $C = \{a_{1c}, \dots, a_{Nc}\}$) the modified formula $\sigma_k^{\text{mod}_{Ag}}$. If this agent is last in the list then set $\sigma_{k+1} = \sigma_k^{\text{mod}_{Ag}}$.
6. After testing (4.9) by the remaining agents set global state $k = k + 1$.

Step 5 demands to set the schedule of agents for sending information. This schedule may be organized by

- A list in $C = \{a_{1c}, \dots, a_{Nc}\}$;
- randomly;
- in inverted range of cardinality of elements in $F_k = \{F_{1c}^k, \dots, F_{Nc}^k\}$.

3 Example

Let us consider the example of three-tank system (Fig.1).

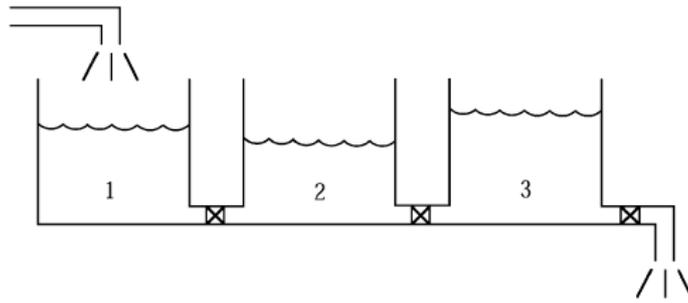


Fig.1. Three-tank system

The list of possible faults in this system is given in Table 2 [2].

Table 2: Set of faults of three-tank system

f_k	Fault description
f_1	fault of the flow sensor F
f_2	fault of the level sensor L_1
f_3	fault of the level sensor L_2
f_4	fault of the level sensor L_3
f_5	fault of the control path U
f_6	fault of the control-valve
f_7	fault of the pump
f_8	lack of medium

f_9	partial clogging of the channel between the tanks Z_1 and Z_2
f_{10}	partial clogging of the channel between the tanks Z_2 and Z_3
f_{11}	partial clogging of outlet
f_{12}	leaking from the tank Z_1
f_{13}	leaking from the tank Z_2
f_{14}	leaking from the tank Z_3

Let us assume that fault detection is realized with the use of five residuals generated on the grounds of the physical equations of the system [2]. These residuals generate the symptoms $S = \{s_i\}_{i=1}^{N_s=5}$.

Let us represent symptom-fault mapping with following binary diagnostic matrix (Table 3) [2].

Table 3: Binary diagnostic matrix

S/F	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}
s_1	1				1	1	1	1						
s_2	1	1	1						1			1		
s_3		1	1	1					1	1			1	
s_4			1	1						1	1			1
s_5	1	1	1	1							1	1	1	1

Let us create agents that are response for faults. The possible distribution of agents are presented in Table 5.3.

Table 4: Assignment of agents to faults

Unit	Agent	List of faults
Tank Z_1	a_1	f_2, f_9, f_{12}
Tank Z_2	a_2	f_3, f_{10}, f_{13}
Tank Z_3	a_3	f_4, f_{11}, f_{14}
Pump	a_4	f_1, f_5, f_6, f_7, f_8

Our model has following body:

$$A = \{a_1, a_2, a_3, a_4\}, F = \{f_1, \dots, f_{14}\}, Af_1 = \{f_2, f_9, f_{12}\}, Af_2 = \{f_3, f_{10}, f_{13}\},$$

$$Af_3 = \{f_4, f_{11}, f_{14}\}, Af_4 = \{f_1, f_5, f_6, f_7, f_8\},$$

Let the following event takes place – $S_o = \{s_2, s_3, s_5\}$.

$$\text{Then we have } \sigma = f_2 \vee (f_9 \wedge f_{13}) \vee (f_9 \wedge f_{12}) \vee (f_{12} \wedge f_{13}),$$

Let us apply α -procedure.

We have now two agents responsible for faults - $C = \{a_1, a_3\}$.

$$\text{Besides, } F^0 = \{F_1^0, F_3^0\} = \{\{f_2, f_9, f_{12}\}, \{f_4, f_{11}, f_{14}\}\}.$$

Step	Action	
0	$\sigma_0 = f_2 \vee (f_9 \wedge f_{13}) \vee (f_9 \wedge f_{12}) \vee (f_{12} \wedge f_{13})$ $C_1^0 = \{2\}, C_2^0 = \{9,13\}, C_3^0 = \{9,12\}, C_4^0 = \{12,13\}$	
0	$Ag = a_1;$ $\min \text{card}(C_i^0) = 1;$ $f_{Ag} = f_2;$	$Ag = a_3;$ $\min \text{card}(C_i^0) = 2;$ $f_{Ag} = f_{12};$
0	Checking: $f_2 = \text{false}$	Checking: $f_{12} = \text{true}$

1	$\sigma_1 = f_9 \vee f_{13}$ $C_1^1 = \{9\}, C_2^1 = \{13\}$	
1	$Ag = a_1;$ $\min card(C_i^0) = 1;$ $f_{Ag} = f_9;$	$Ag = a_3;$ $\min card(C_i^0) = 1;$ $f_{Ag} = f_{13};$
1	Checking: $f_9 = false$	Checking: $f_{13} = true$

As it easy to see in this case we check all possible faults from the initial list and do not get a profit in compare with consecutive search except parallel calculation.

Let us apply β - procedure now.

Step	Action	
0	$\sigma_0 = f_2 \vee (f_9 \wedge f_{13}) \vee (f_9 \wedge f_{12}) \vee (f_{12} \wedge f_{13})$ $C_1^0 = \{2\}, C_2^0 = \{9,13\}, C_3^0 = \{9,12\}, C_4^0 = \{12,13\}$	
0	$Ag = a_1;$ $\min card(C_i^0) = 1;$ $f_{Ag} = f_2;$ Checking: $f_2 = false$	
0	$\sigma_0^{\text{mod}_{a_1}} = (f_9 \wedge f_{13}) \vee (f_9 \wedge f_{12}) \vee (f_{12} \wedge f_{13})$ $C_1^{0\text{mod}_{a_1}} = \{9,13\}, C_2^{0\text{mod}_{a_1}} = \{9,12\}, C_3^{0\text{mod}_{a_1}} = \{12,13\}$	
0		$Ag = a_3;$ $\min card(C_i^0) = 2;$ $f_{Ag} = f_{12};$ Checking: $f_{12} = true$
1	$\sigma_1 = f_9 \vee f_{13}$ $C_1^1 = \{9\}, C_2^1 = \{13\}$	
1	$Ag = a_1;$ $\min card(C_i^0) = 1;$ $f_{Ag} = f_9;$ Checking: $f_9 = false$	
1	$\sigma_1^{\text{mod}_{a_1}} = f_{13}$	
1		$Ag = a_3;$ $\min card(C_i^0) = 1;$ $f_{Ag} = f_{13};$ Setting: $f_{13} = true$

Using β - procedure for this case reduce number of fault checking to three times and substitute checking process by setting.

Remark. We do not draw attention now on procedure of fault checking but we understand that this is some deeply investigation of symptoms dynamic contrary to

identification of symptom to obtaining the set of observed symptoms S_o . Moreover the checking procedure may include additional actions (like computational and even mechanical) to recognize suspected fault. That is why even though big reduce of checking procedure is more than is required.

Conclusion

Due to the distributional character of fault diagnosis and necessity of taking into account different uncertainties during the diagnosis process the multi-agent methodology has been proposed. We showed that the principle of assigning the agent to subsystems in this case is also useful. Then the local knowledge base of agents and rules for communications can precise the possible diagnosis in addition to the ordinary binary or fuzzy diagnostic matrix.

As a result of this investigation we could propose engineering methods for applying multi-agent methods to fault diagnosis problems.

Acknowledgement

This research has partially been supported by the Saxon Ministry of Sciences and Art (Germany).

References:

- [1] R. Isermann. Fault-Diagnosis Systems. An Introduction from Fault Detection to Fault Tolerance. Springer, 2006.
- [2] Soo Young Eo, Tae Suk Chang, Dongil Shin, En Sup Yoon. Cooperative problem solving in diagnostic agents for chemical processes. Computer & Chemical Engineering 24 (2000) 729 – 734.
- [3] M. Wooldridge. An introduction to multi-agent systems. John Wiley&Sons. 2001.
- [4] W. Hoek, M. Wooldridge. On the logic of cooperation and propositional control. Artificial intelligence 164 (2005) 81-119.
- [5] B.D. Pomeroy, H.A. Spang, M.E. Dausch. Event-based architecture for diagnosis in control advisory systems. Artificial Intelligence in Engineering 5 (1990) 174-181.

Authors:

Prof. Dr. Oleksandr Sokolov
National Aerospace University,
Tchkalov str., 17
61070 Kharkiv, Ukraine
Phone: ++38 057 707 43 04
Fax: ++38 057 315 11 31
E-mail: asokolov@xai.edu.ua

Dr. Michael Wagenknecht
University of Applied Sciences Zittau/Goerlitz, IPM
Theodor-Koerner-Allee 16
02763 Zittau, Germany
Phone: ++49 (0) 3583 61 1546
Fax: ++49 (0) 3583 61 1288
E-mail: m.wagenknecht@hs-zigr.de

Dr. Ulrike Gocht
University of Applied Sciences Zittau/Goerlitz, IPM
Theodor-Koerner-Allee 16
02763 Zittau, Germany
Phone: ++49 (0) 3583 61 1546
Fax: ++49 (0) 3583 61 1288
E-mail: u.gocht@hs-zigr.de

V.Nissen

Management Applications of Fuzzy Control

Introduction

Fuzzy control [1] is one of the oldest and best established application areas of fuzzy set theory. In this paper, the concept of a fuzzy controller as a knowledge-based system is transferred from the well-established technical to the management domain. Similarities and differences are highlighted and sample applications in management are given.

Fuzziness is a form of uncertainty. It embraces in particular linguistic and informational uncertainty [2]. Linguistic uncertainty is characterized by a lack of precision and the indefinite nature of human language. (*“high”* interest rates). Informational uncertainty exists when many descriptors are necessary to describe a term clearly (*“creditworthy”* company). In contrast to stochastic uncertainty, it is not adequate to model these forms of uncertainty based on probability theory.

Since the mid 1960s, fuzzy set theory has been developed by Zadeh [3], Zimmermann [2] and others as a theoretical basis in order to model fuzziness. Fuzzy Set Theory is well-known in the technical domain. In 1975, a cement kiln built in Denmark became the first industrial application of fuzzy control. Later, the idea of fuzzy control was particularly successful in Japan. Today, we find fuzzy control applications in many aspects of daily life, such as automatized chemical process control, gear shifts that adapt to the car driver, and driverless train operations. In addition, many applications of other fuzzy techniques, such as fuzzy data analysis and hybrid approaches, such as Neuro-Fuzzy systems, can be identified.

Fuzzy systems are successful, because they allow for a relatively straightforward modelling and transparent model structure even in complex tasks [4]. Moreover, they demonstrate robust behaviour, for instance in dynamic environments.

The adequate treatment of fuzziness also has great significance in management. For example, qualitative expert judgements, potential information overflow and vague relationships characterize important management domains such as knowledge management, strategic foresight and customer relationship management. Nevertheless, productive management applications of fuzzy set theory are still rare.

Fuzzy Set Theory and Fuzzy Control

Fuzzy systems and fuzzy methods have a solid mathematical basis. In classical set theory, an element x from a basic set X ($x \in X$) either definitely belongs to a set A or it definitely does not belong to A . However, for many real circumstances such a strict distinction does not render an appropriate representation. In fact, gradual membership prevails in reality. Thus a fuzzy set \tilde{A} is characterized by the fact that the membership of an element x to \tilde{A} can be indicated by a real number which is usually standardized on the range of values $[0,1]$, thus describing formally a fuzzy set \tilde{A} by a real value membership function $\mu_{\tilde{A}}$ is: $\mu_{\tilde{A}} : X \rightarrow [0,1]$. Herein, a value $\mu_{\tilde{A}}(x) = 0$ means that x does not belong to the fuzzy set \tilde{A} , while a value $\mu_{\tilde{A}}(x) = 1$ indicates full membership. Values within the interval $0 \leq \mu_{\tilde{A}}(x) \leq 1$ indicate a partial membership of x in the set \tilde{A} . The classical, non-fuzzy set A can be interpreted as a special fuzzy set, where only two alternatives, no membership or full membership, exist.¹

The application domain of technical control is characterised by the goal to automatize the supervision and correcting activities for complex (non-linear) technical processes. In classical control theory, the design of a controller is based on a mathematical model of the technical process, frequently in the form of differential equations that define the system response to its inputs. Fuzzy controllers differ from that by modeling the know how of a human control expert for the technical process [4]. Usually, this know how is expressed in logic rules (IF-THEN statements) where fuzzy sets are employed to model qualitative terms like “high” and “low” that the expert uses in his rules.

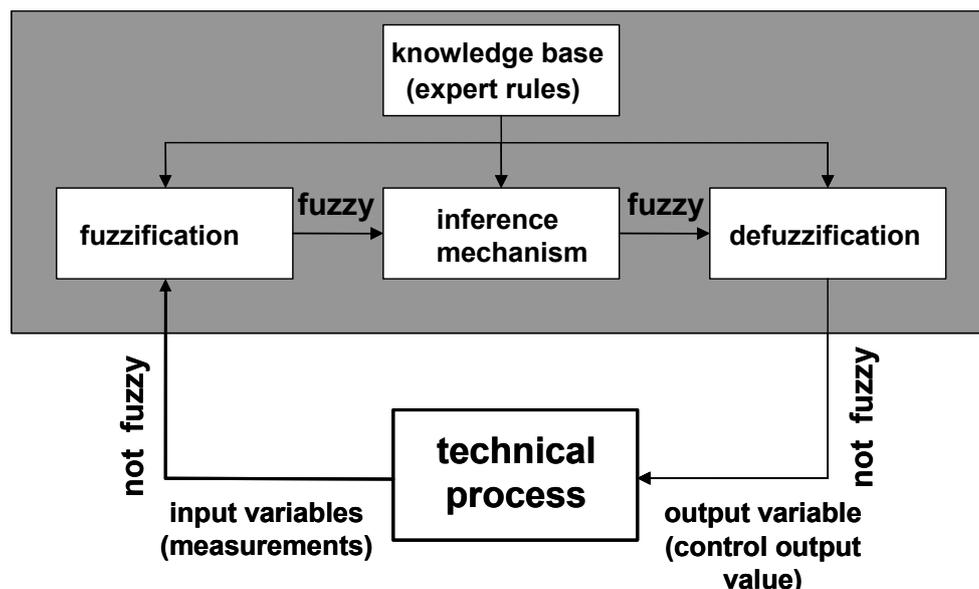


Fig. 1: Structure of a fuzzy controller

¹ For a more detailed introduction to fuzzy set theory see [2], [4], and [5].

Figure 1 outlines the general structure of a fuzzy controller. The input and output of the controller are generally crisp (not fuzzy) values, while the inference mechanism is based on fuzzy data. Thus, initially the input data is individually mapped from crisp to fuzzy values, a step that uses sets of membership functions. The result is for each fuzzy set a real value in the interval $[0, 1]$. This is called “fuzzification”. Because neighbouring fuzzy sets overlap, an input variable’s state does not jump abruptly from one state to the next. Instead, it loses value in one membership function while gaining value in the next. Through fuzzification the compatibility of the facts (measurements) with rule antecedents is determined. In contrast to classical expert systems, rule antecedents of fuzzy controllers may only be fulfilled partly. All rules from the knowledge base with antecedents that have strictly positive membership values are activated in parallel. Even conflicting rules may be simultaneously active.

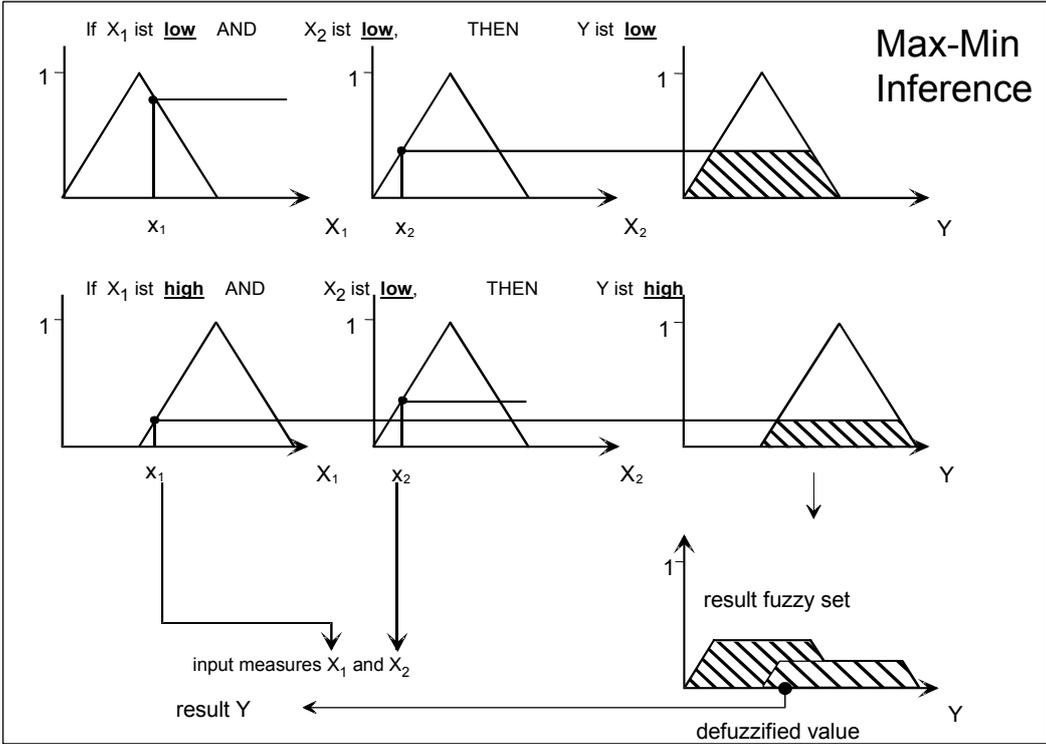


Fig. 2: Principle of max-min inference and center of gravity defuzzification

Usually, rules have several antecedents that are aggregated using fuzzy operators, such as AND, OR, and NOT. Moreover, rules may be weighted to express the confidence of the process expert in the rule. Then, by applying fuzzy inference the fuzzy values of input variables are mapped on the output variable of the fuzzy controller. Frequently, the fuzzy minimum operator (AND) is used for this purpose. Figure 2 highlights, for the example of the AND-operator, how fuzzy sets of the output variable are cut off at the level of the minimum of fuzzified inputs. As several rules can be active simultaneously,

their results must be accumulated. Often this is achieved by using the fuzzy maximum operator (OR). The combination of max and min operators is called max-min-inference and a common form of inference mechanism to determine the fuzzy output of the controller. At last, a “defuzzification” step provides a crisp output value, often by calculating the center of gravity for the combined output fuzzy set of all active rules (figure 2).

Applying Fuzzy Control in Management

On an abstract level, fuzzy controllers are used to model human knowledge about complex, non-linear processes in a transparent, formal way that allows for automated execution. In the management domain during the 80ies and 90ies, there was a strong movement to model human expertise in rule-based “expert systems”. These generally suffered from brittleness at the border of their domain of expertise, as well as from often great complexity in the rule base and inference mechanism that prompted design and performance problems. Fuzzy controllers are also rule-based systems, but contrary to classical expert systems, the structure of their knowledge base is much simpler, and they rely on a simple feedforward inference mechanism without backtracking.

Potential applications of fuzzy control-like systems in management can be found where:

- analytical models of the domain are impossible or require a prohibitive effort,
- the results of analytical models are intransparent and have no acceptance,
- qualitative judgements and/or vague relationships are important,
- human expertise about the complex application domain is available in the company.

Based on these criteria, many areas of management qualify for the application of fuzzy rule based systems. In our research group, amongst others, the following successful applications were developed:

- weather-dependent production planning in an industrial bakery [6],
- fuzzy analysis of company balance sheets [7],
- forecasting consulting effort for IT-projects [8],
- modelling corporate strategy in a fuzzy balanced scorecard [9],
- modelling qualitative information in management simulation games [10].

The system structure is always basically equivalent to that of a fuzzy controller as given in figure 1. Again, the rule base contains expert knowledge (if-then statements) about important relationships between variables of the application domain. However, some important differences between technical fuzzy controllers and similar fuzzy rule-based systems in management applications exist:

- Contrary to technical process control, where in short intervals repeated measurements are performed and a control variable is adapted, the cycle of determining inputs, reasoning, and deciding about the output is only performed once in management applications. Thus, the output result is of greater importance than in traditional fuzzy control, as it can not be corrected by a consecutive inference cycle. (However, the fuzzy system may be used for simulations with different input combinations [9].)
- The rule base is generally more complex, because the number of relevant input (and sometimes output) variables is higher in management, so that expert knowledge cannot be summarized in only five to ten rules, as can be done in many technical control applications. To avoid problems with large knowledge bases, fuzzy rule-based systems in management applications frequently take the form of a hierarchical fuzzy controller, where several interrelated rule bases exist that individually solve parts of the overall problem. The total output is then determined by a hierarchical cascade of intermediate results.
- Hence, the system design process can be more complex than in technical fuzzy controllers. Moreover, determining the output quality during the design phase of the rule-based system is sometimes difficult and may rely on expert judgements.
- A true automatization of decision making is often not the goal with fuzzy rule-based systems in management. It is more a device for supporting human decisions.
- Input data is frequently quantitative, as in technical control, but it may suffice or render desired extra information to have the output also in qualitative form, giving all positive membership values for output fuzzy sets. Thereby, information about the risk of the decision (output) is available to management.
- To increase management acceptance for the fuzzy rule-based output in decision support, it is desirable (and possible) to construct an explanation component. Such a component makes use of intermediate results of the hierarchical fuzzy system to explain how the final output was generated [11].

A brief example [8] will help to make these differences more accessible. Figure 3 outlines the general structure of a fuzzy rule-based system that takes certain characteristics of a planned consulting project as input (left side of fig. 3) and generates an estimation of required consulting effort (man-days), separated for different parts of the project, as output. The result can then be used in the quotation of the consulting company to win the respective customer order. The idea in this case was to support and eventually even automatize the otherwise laborious manual process of estimating consulting effort. Know

how of experienced consultants was structured in several rule bases to allow for a hierarchical decomposition of the overall estimation problem. Every triangle in fig. 3 corresponds to one fuzzy controller, roughly analogous in structure to fig. 1. A cascade of intermediate results (only an excerpt of the whole system is given in fig. 3) then leads to the final effort estimation. This tool showed good performance in practical tests when compared to manual results. Of course, effort estimation is always domain-specific, but the general idea can be quickly transferred from one consulting domain to another.

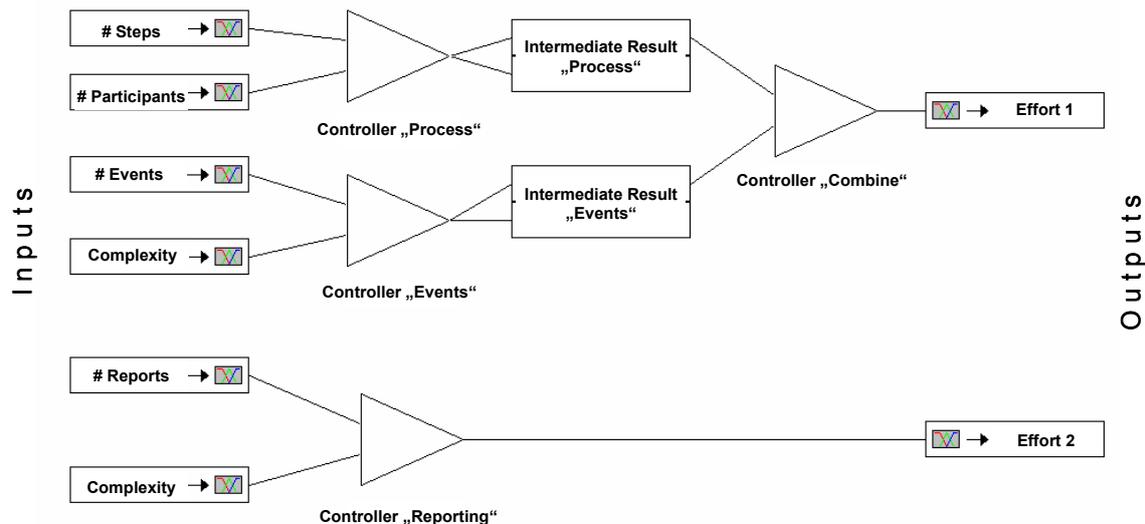


Fig. 3: Structure of a hierarchical fuzzy controller that estimates consulting effort, based on project characteristics that sales staff provided (excerpt) [8]

References:

- [1] Mamdani, E.H.; Assilian, S.: An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller, in: Int. Journal of Man Machine Studies 7 (1975), 1 – 13.
- [2] Zimmermann, H.-J.: Fuzzy Set Theory – and Its Applications, 4. ed., Boston: Kluwer, 2001.
- [3] Zadeh, L.A.: Fuzzy Sets, in: Information and Control 8 (1965), 338 – 353.
- [4] Nauck, D. und Kruse, R.: Fuzzy-Systeme und Soft Computing, in: Biethahn, J.; Hönerloh, A.; Kuhl, J.; Nissen, V. (eds.): Fuzzy Set Theorie in betriebswirtschaftlichen Anwendungen, Vahlen: München 1997, 3 – 21.
- [5] Sivanandam, S.N.; Sumathi, S.; Deepa, N.: Introduction to Fuzzy Logic using MATLAB, Berlin et al.: Springer, 2006.
- [6] Helmeke, J.; Hönerloh, A.: Wetterabhängige Produktionsplanung in einem Konditoreibetrieb, in: Biethahn, J.; Hönerloh, A.; Kuhl, J.; Nissen, V. (eds.): Fuzzy Set Theorie in betriebswirtschaftlichen Anwendungen, Vahlen: München 1997, 143 – 153.
- [7] Leisewitz, M.-C.: Leisewitz, M.-C.: Das Problem der Unschärfe in der Unternehmensbewertung: Ein Fuzzy-Expertensystem zur Findung des Grenzpreises bei Unternehmenskäufen, Göttingen: Unitext, 1999.
- [8] Nissen, V.: Wissenskonservierung mittels eines fuzzy-regelbasierten Systems für die Aufwandsplanung von Beratungsprojekten, in: Zeitschrift für Planung & Unternehmenssteuerung, 14 (2003) 3, 243 – 258.
- [9] Nissen, V.: Modelling Corporate Strategy with the Fuzzy Balanced Scorecard. In: Hüllermeier, E.; Kruse, R.; Nürnberger, A.; Strackeljan, J. (eds.): Proceedings Symposium on Fuzzy Systems in Computer Science FSCS 2006, Magdeburg, 2006, 121 – 138.
- [10] Nissen, V.; Ananidze, G.: Modelling Qualitative Information in a Management Simulation Game, in: Waldmann, K.-H.; Stocker, U.M. (eds.): Operations Research Proceedings 2006, Berlin u.a.: Springer, 2007, 389 – 394.
- [11] Kuhl, J.: Angepaßte Fuzzy-Regelungssysteme. Entwicklung und Einsatz bei ausgewählten betriebswirtschaftlichen Problemstellungen, Göttingen: Unitext, 1996.

Author:

Prof. Dr. Volker Nissen
 TU Ilmenau, FG Wirtschaftsinformatik für Dienstleistungen,
 PF 10 05 65, D-98684 Ilmenau
 Phone: +49 (0)3677 69-4047
 Fax: +49 (0)3677 69-4219
 E-mail: volker.nissen@tu-ilmenau.de

O. G. Rudenko / A. A. Bessonov / P. Otto

A Method for Information Coding in CMAC Networks

1. Introduction

An artificial neural network in form of a **CMAC** net (**Cerebellar Model Articulation Controller**), which was developed firstly for the regulation of robot manipulators [1,2], is characterized by large learning speed, simple realization and the possibility of influencing the necessary memory capacity purposefully [3-6]. The main difference between CMAC nets and the majority of other network types lies in the possibility of selection of a relatively large number of free parameters of the net (methods of the information coding, number of the quantization stages and level of quantization, kind of the activation functions of the neurons). While the problem of the effective choice of the basic functions of the neurons has been given large attention [7-9], the problem of the optimal selection of the other parameters of the net has so far hardly been attended. This work is therefore dedicated to the investigation of methods for selection of patterns to information coding in a CMAC net.

2. Structure of a CMAC Network

The net consists of the input -, a hidden and an output layer, denoted by L1, L2 and L3 respectively (fig. 1). The layer L1 has R_i quantization level with a quantization step r_i for each input i . In addition all quantization levels of the i^{th} inputs have the same number of ρ quantization stages $(C_1^i, C_2^i, \dots, C_\rho^i)$, from which each includes $\rho^* \leq \rho$ quantization ranges A, B, \dots, a, b, \dots to which a priori selected basic functions are assigned. Therefore the quantization ranges can be taken as neurons of the input layer and the basic functions can be regarded as their activation functions. A n -dimensional input vector x activates $N\rho$ neurons of L1.

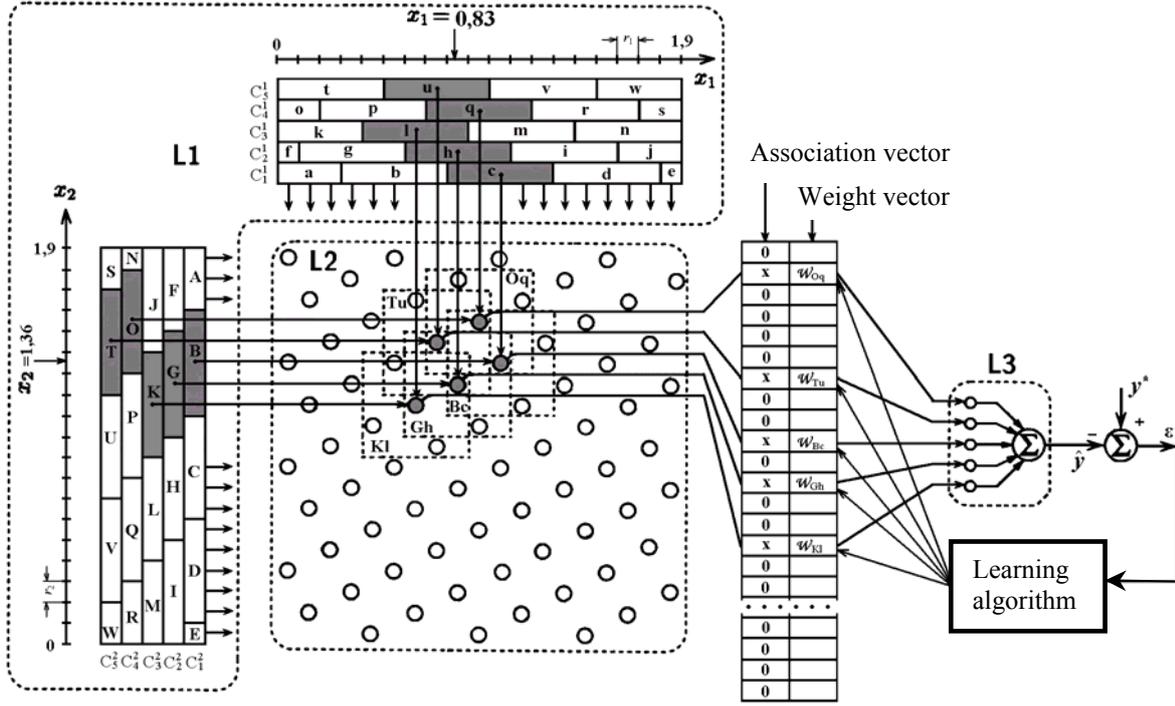


Fig 1: Structure of a CMAC Network

The layer L2 consists of associative neurons, which are connected with certain neurons of the layer L1 and whose working ranges combine. The maximum number of associative neurons is given by the equation

$$n_{\max} = \left\lceil \rho \left(\frac{R-1}{\rho} + 1 \right)^N \right\rceil, \quad (1)$$

where R – is the Number of quantization levels of the input signals,

N – is the dimension of the input vector,

$\lceil \cdot \rceil$ - designates the round to the next larger whole number.

Each neuron in L2 has its own receptor field, which includes further $(\rho - 1)$ associative neurons. A signal deviating from zero leads to the excitation of ρ associative neurons, which supply output signals, which correspond to the product of the signals arriving from the neurons of the layer L1. All neurons of the output layer L3 are connected with the associative neurons of the second layer. Each connection has its own adaptive weight w_i ($i = \overline{1, n}$). The output signal of the third layer arises therefore as weighted sum of the output signals of the second layer. The transformation $\mathbf{a} = S(\mathbf{x})$ codes the information (layer L1 and L2) and the transformation $\hat{\mathbf{y}} = P(\mathbf{a}) = \mathbf{a}^T \mathbf{w}$ computes the output signal (layer L3).

3. Choice of Activation function

By a non- rectangular activation functions, the output signal is computed using

$$\hat{y} = H(\mathbf{a}^T \Phi(\mathbf{x}))\mathbf{w}, \quad \text{with } \Phi(\mathbf{x}) = \text{diag}(\Phi_1(\mathbf{x}), \Phi_2(\mathbf{x}), \dots, \Phi_n(\mathbf{x})), \quad (2)$$

and $\Phi_i(x) = \prod_{j=1}^N \phi_{ij}(x_j)$; $\phi_{ij}(x_j)$ - is the value of the activation function in x_j .

In CMAC nets with rectangular basic functions is $\Phi(\mathbf{x}) = \mathbf{I}$. The elements of the association vector have only the values 0 or 1. However, for high requirements of the precision of the models they are not suitable. Learning occurs in CMAC nets in the change of the weight vector \mathbf{w} by comparison of the reaction of the net $\hat{y}(k)$ with for a learning example $\{\mathbf{x}(k), \mathbf{y}^*(k)\}$ given output value $\mathbf{y}^*(k)$. The learning algorithm reads:

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \frac{\mathbf{y}^*(k) - \hat{y}(k)}{\|\Phi(\mathbf{x}(k))\mathbf{a}(k)\|^2} \Phi(\mathbf{x}(k))\mathbf{a}(k). \quad (3)$$

Frequently Gauss' curves are used as activation functions. It is however difficult to specify the excitation borders exactly what is important for coding the information. The cosine function avoids this disadvantage

$$\Phi_i(x_j) = \begin{cases} \cos\left(\frac{\pi}{\rho r_j}(x_j - \lambda_i)\right); & x_j \in \left(\lambda_i - \frac{\rho r_j}{2}, \lambda_i + \frac{\rho r_j}{2}\right]. \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

4. Information coding

A special coding makes a substantial decrease in memory requirement for CMAC nets possible. In the layers L1 and L2 a coding is done, with which for each N - dimensional input vector $\mathbf{x}(i)$ using the association layer a n - dimensional association vector $\mathbf{a}(i)$ is formed. Its elements are assumed to be from the interval $[0, 1]$. Only $\rho \ll n$ elements of this vector have values not equal zero, i.e. only ρ memory elements are active. Coding takes place in two steps: First in the layer L1 a discretization is made via quantization of the input signals and secondly the code is computed in the layer L2. Thereby the mutual arrangement of the neurons of the input layer is important, which belong to different quantization stages, because herewith the coding pattern is clearly determined.

With non rectangular basic functions a problem develops with the selection of the coding pattern. The simplest coding method is the use of a table, in which $\rho=1$ is selected. A maximum memory volume of R^N memory locations is required. This coding pattern is suitable only for rectangular affiliation functions, because for all other types the receptive fields of the associative neurons do not overlap. Learning is unnecessary thereby because simply the value of the function can be registered in the appropriate place of the association vector. If the quantization patterns are alike concerning all components of the vector of the output signals, the associative neurons will be arranged diagonally. Here the maxima are likewise for non-rectangular basic functions on the diagonals, which makes learning in points between the diagonals more difficult and substantially worsens the filter characteristics of the net.

Exchanging some neuron layers in the quantization matrices (fig. 1), permits a change of the coding pattern. With the same signal the same neurons will be excited, while the receptive fields of the associative neurons are completely differently arranged. Since no pronounced diagonal arrangement of the associative neurons is to be registered here, the approximation accuracy becomes substantially larger. At present there are still no recommendations regarding the optimal choice of the coding pattern. It is experimentally determined. It must be paid attention nevertheless to the fact that any neuron of a certain stage C_k^i of the input layer should not be connected to more than $(\rho - 1)$ neurons of the neighbour layers C_{k-1}^i and C_{k+1}^i . This corresponds to a delimitation of the association vector on maximally $(\rho-1)$ components for coding of two different vectors of the input signals, for which a recognition is still possible.

5. Simulative Experiments

A goal is the investigation of the influence of the coding pattern on the identification and the control of the nonlinear system expressed by equation (5) when disturbed by white noise in the interval $[-0,3 \ 0,3]$.

$$y(k+1) = 0.725\beta \sin\left(\frac{16u(k) + 8y(k)}{\beta(3 + 4u^2(k) + 4y^2(k))}\right) + 0.2u(k) + 0.2y(k), \quad (5)$$

After the identification with 30000 examples with $\beta = 0,8$, basic functions according to equation (4) and the desired controlled variable

$$y^*(k) = 0.48 + 0.07 \sin(\pi k / 200) - 0.05 \cos(\pi k / 100),$$

stationary, evenly distributed random signal $u(k)$ from the interval $[-1, 1]$ and the parameters of the net $R=50$, $\rho = 20$, which requires 238 memory locations, the control was performed using the following algorithm:

$$u(k+1) = u(k) + \gamma \nabla_u \hat{y}(k) e_y(k), \quad (6)$$

with γ – the incrementation (affects the convergence speed);

$$\nabla_u \hat{y}(k) = \frac{\partial \hat{y}}{\partial u(k)} = \sum_{i=1}^n \left[a_i w_i \left(\prod_{\substack{j=1 \\ j \neq k}}^N \phi_{ij}(\hat{y}(k-1)) \right) \frac{\partial \phi_{ik}}{\partial u(k)} \right]; \quad e_y(k) = y^*(k) - \hat{y}(k).$$

Fig. 5a shows the results of the identification (upper diagram) and the control (lower diagram) by using the "diagonal" and fig. 5b by using the random coding pattern accordingly. The desired output signal is represented by a solid and the actual by a broken line. The control variable according to equation (6) is characterized by a line with circular markers. From fig. 5 it can be seen that the random coding pattern performs better.

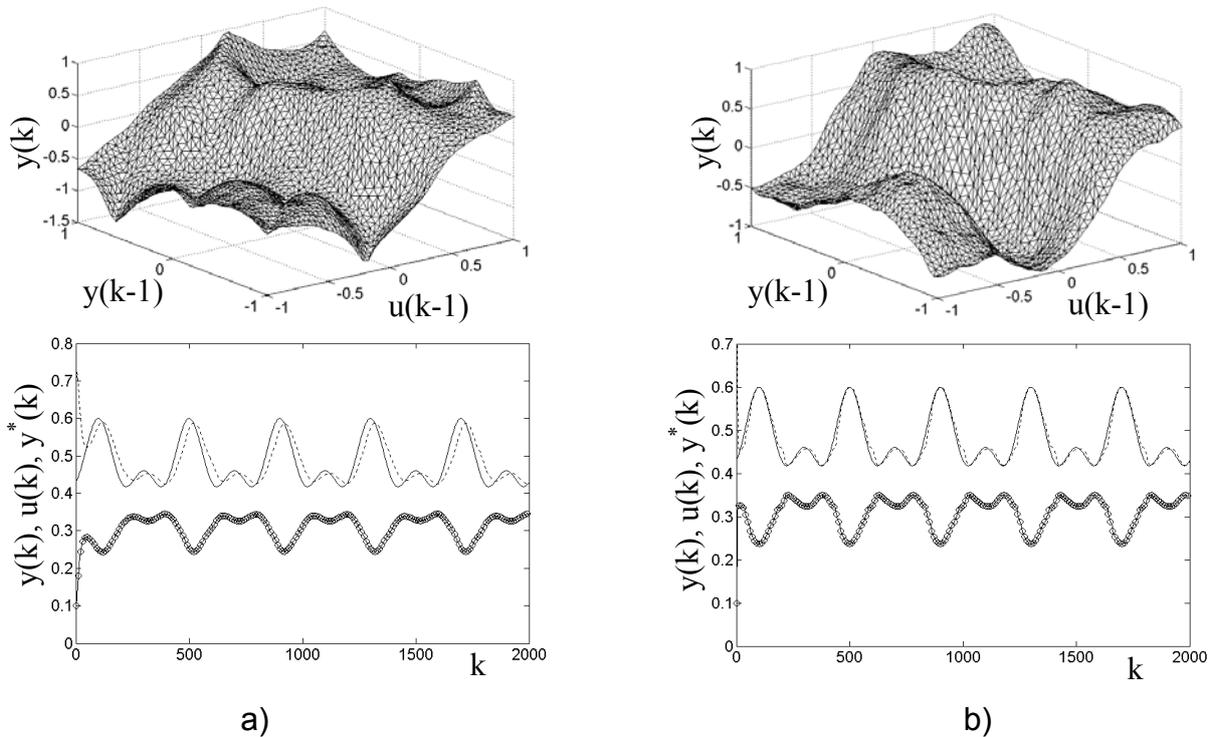


Fig 5: Results of the identification and control of a system with disturbances

6. Conclusions

On the basis of the obtained results the following conclusions can be drawn:

The approximation characteristics of a CMAC net depend in contrast to many other neural network types not only on the learning algorithm and the basic functions, but also on the choice of the coding pattern of information, i.e. the arrangement of the neurons in the input layer.

The choice of the distribution pattern of the associative neurons affects also the quality of control of nonlinear systems.

The coding pattern in which the neuron layers of the quantization matrices are randomly arranged has proven to be most effective.

References:

- [1] Albus J.S. A new approach to manipulator control: the cerebellar model articulation controller (CMAC) // ASME Trans., J. Dynamic Systems, Measurement and Control, 1975. – Vol. 97. – №3. – P. 220-227.
- [2] Albus J.S. Data storage in cerebellar model articulation controller (CMAC) // ASME Trans. J. Dynamic Systems, Measurement and Control, 1975. – Vol. 97. – №3. – P. 228–233.
- [3] Commuri S., Lewis F.L. CMAC neural networks for control of nonlinear dynamical systems, structure, stability, and passivity // Proc. IEEE Int. Symp. Intell. Control, 1995. – Pp. 123-129.
- [4] Commuri S., Lewis F.L., Jagannathan S. Discrete-time CMAC neural networks for control applications // Proc. of the 34th Conf. on Decision & Control. – New Orleans, LA-1995. – Pp.2420-2426.
- [5] Iiguni Youji. Hierarchical image coding via Cerebellar Model Arithmetic Computers (CMAC) // IEEE Trans. on Image Processing, 1996. Vol.5. №10. P. 1393-1401.
- [6] Руденко О.Г., Бессонов А.А. Адаптивное управление нелинейными объектами с помощью нейронной сети CMAC // Проблемы управления и информатики. –2004. –№5. – С. 16-30.
- [7] Lane S.H., Handelman D.A., Gelfand J.J.. Theory and development of higher-order CMAC neural networks // IEEE Control Systems. – 1992. – Vol. 12. – №2.– Pp. 23-30.
- [8] Chiang Ch.-T., Lin Ch.-Sh. CMAC with General Basis Functions // Neural Networks, 1996. – v.9. - №7. – Pp. 1199-1211.
- [9] Руденко О.Г., Бессонов А.А. О выборе базисных функций в нейронной сети CMAC // Проблемы управления и информатики. – 2004. – №2. – С. 143-154.

Authors:

Prof. Dr.-Ing. habil. Oleg G. Rudenko
Dr.-Ing. Alexander A. Bessonov
Staatliche Universität für Funkelektronik, Leninpr. 14
61166 Charkov Ukraine
Phone: +38 0577021354
E-mail: rudenko@kture.kharkow.ua

PD Dr.-Ing. habil. Peter Otto
TU Ilmenau, Kirchhoffstr. 1
98693 Ilmenau
Phone: 03677 692773
Fax: 03677 1434
E-mail: peter.otto@tu-ilmenau.de

Ye. Bodyanskiy / P. Otto / I. Pliss / N. Teslenko

Nonlinear process identification and modeling using general regression neuro-fuzzy network

INTRODUCTION

At present time, artificial neural networks are widely used for solving the problems of identification, prediction and modeling of nonlinear processes and systems. However, when the data are fed in real time and their processing must be simultaneous with functioning of the plant, in the case of nonstationary plant the problem becomes difficult. The so-called "optimization-based networks" such as Multilayer Perceptron, Radial Basis Functions Networks (RBFN) or Normalized Radial Basis Functions Networks (NRBFN) can be ineffective to solve such a problem through their slow convergence rate, curse of dimensionality, appearance of regions where all neurons of the network are inactive and possibility of getting to the local minima.

"Memory-based networks", such as General Regression Neural Network (GRNN), proposed by D. F Specht [1], can be referred to, so-called, just-in-time models [2], which are learned by one-pass learning algorithm by the principle "neurons at data points" [3]. These properties are the cause of GRNN high learning rate.

For the solving of nonlinear plant identification problem

$$y(x) = F(x(k)),$$

where $y(x)$, $x(k)$ – scalar and $(nx1)$ -vector of output and input signals correspondingly in the instant time $k=1,2,\dots$, $F(\bullet)$ – unknown nonlinear operator of the plant, it is necessary to form learning sample $\{x^*(k), y^*(k)\}$, $k=1,2,\dots,l$, whereupon it is possible to get the estimate $\hat{y}(x)$ of the plant response $y(x)$ to arbitrary input signal x in the form

$$\hat{y}(x) = \sum_{k=1}^l y^*(k) \varphi(D(k)) \left(\sum_{k=1}^l \varphi(D(k)) \right)^{-1} \quad (1)$$

where $D(k)$ – distance measure in accepted metrics between x and $x^*(k)$, $\varphi(\bullet)$ – some kernel function, usually, Gaussian. GRNN converges asymptotically to optimal

nonlinear regression surface with the growing of learning sample size [4] and its learning process can be organized easily in real time. But the main problems connected with GRNN using are defined by possible curse of dimensionality, when the number of data l is large.

Neuro-Fuzzy Systems (NFS) [5-6] combine the neural networks learning abilities with transparency and interpretability of the Fuzzy Inference Systems (FIS). Having approximating abilities of RBFN [6-7], NFS subject to curse of dimensionality with less degree, that provides them advantage in comparison with neural networks because of using univariate Fuzzy Basis Functions (FBF) instead of multidimensional RBF.

Among NFS Adaptive Network-based Fuzzy Inference System (ANFIS) have got wide spread [8]. ANFIS has five-layer architecture and is typical representative of the optimization-based networks family, which are characterized by insufficient learning rate. Lattice-based Associative Memory Networks (LAMN) [9-10] are the representatives of memory-based networks, whose output signal is formed on basis of univariate bell-shaped functions uniformly distributed on axes of n -dimensional input space. As a result of aggregation operation multidimensional FBFs are formed, whose centers are also uniformly distributed in multidimensional space, and their layout doesn't depend on characteristics of learning sample.

The goal of this work is solving the problem of nonlinear process identification and modeling using General Regression Neuro-Fuzzy Network (GRNFN), which represents by itself NFS and learns as GRNN that provides approximating properties of ANFIS with learning rate of memory-based networks.

THE GENERAL REGRESSION NEURO-FUZZY NETWORK ARCHITECTURE

The architecture of General Regression Neuro-Fuzzy Network is illustrated on Fig. 1 and consists of five sequentially connected layers. First hidden layer is composed of l blocks with n FBF in each and realizes fuzzification of the input variables vector. Second hidden layer implements aggregation of membership levels that are computed in first layer, and consists of l multiplication blocks. Third hidden layer – the layer of synaptic weights that are defined in special way. Fourth layer is formed by two summation units and computes the sums of output signals from the second and third layers. Finally, normalization takes place in fifth (output) layer, where the output

signal is computed.

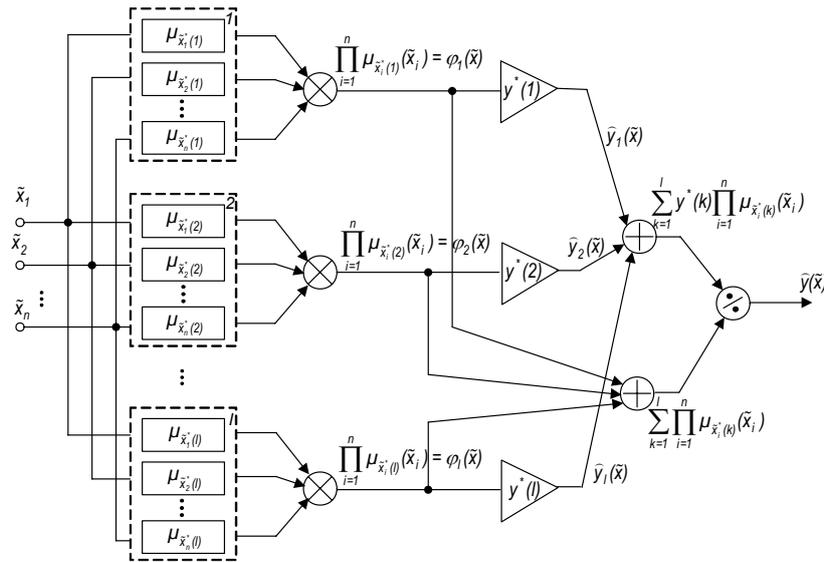


Fig.1 – General Regression Neuro-Fuzzy Network.

GENERAL REGRESSION NEURO-FUZZY NETWORK LEARNING

Since GRNFN belongs to memory-based networks, its learning is based on principle “neurons at data points” that makes it extremely easy and fast. Learning sample vectors $x^*(1), \dots, x^*(k), \dots, x^*(l)$ are normalized in advance on unit centered hypercube so, that $x_i^{*min} \leq x_i^*(k) \leq x_i^{*max}$, $-0,5 \leq \tilde{x}_i^*(k) \leq 0,5$, $i = 1, 2, \dots, n$. For each vector from the learning sample $\tilde{x}^*(k) = (\tilde{x}_1^*(k), \tilde{x}_2^*(k), \dots, \tilde{x}_n^*(k))^T$ in the first hidden layer own set of fuzzy-basis membership functions $\mu_{\tilde{x}_1^*(k)}, \mu_{\tilde{x}_2^*(k)}, \dots, \mu_{\tilde{x}_n^*(k)}$ is formed, so that centers of $\mu_{\tilde{x}_i^*(k)}$ coincide with $\tilde{x}_i^*(k)$, $k=1, 2, \dots, l$. The process of FBF formation is illustrated on Fig. 2. Note that GRNFN contains nl fuzzy-basis functions, that can't lead to the curse of dimensionality.

Theoretically, any kernel function with non-strictly local support can be used as FBF that allows to avoiding of appearance of “gaps” [4]. As such a function one can recommend generalized Gaussian

$$\mu_{\tilde{x}_i^*(k)}(\tilde{x}_i) = \left(1 + \left| \frac{\tilde{x}_i^*(k) - \tilde{x}_i}{\sigma_i(k)} \right|^{2b} \right)^{-1}, \quad b \geq 0,5, \quad (2)$$

that is the bell-shaped function, whose shape is defined by the scalar parameter b [6]. As for choosing of the width parameter $\sigma_i(k)$, standard recommendation leads to

the idea [11], that it must ensure small overlapping of FBFs neighboring. At the same time with FBFs forming in first hidden layer, the synaptic weights are formed in the third hidden layer and they are supposed to be equal to the signals of learning sample $y^*(k)$.

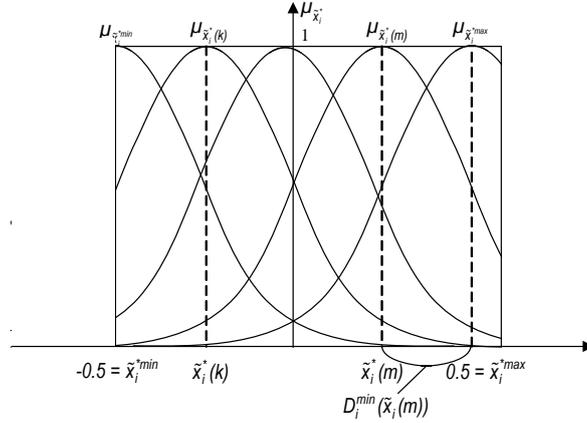


Fig.2 – Fuzzy-basis membership functions.

Thus, when arbitrary signal \tilde{x} is fed to the input of GRNFN in the first hidden layer membership levels $\mu_{\tilde{x}_i(k)}(\tilde{x}_i)$, $i=1,2,\dots,n$, $k=1,2,\dots,l$ are computed, in the second layer their aggregation is realized by forming multidimensional FBFs

$$\varphi_k(\tilde{x}) = \prod_{i=1}^n \left(1 + \left| \frac{\tilde{x}_i^*(k) - \tilde{x}_i}{\sigma_i(k)} \right|^{2b} \right)^{-1}, \quad k=1,2,\dots,l, \quad (3)$$

in the third layer products $\hat{y}(\tilde{x}) = y^*(k)\varphi_k(\tilde{x})$ are determined, fourth layer computes the values of signals $\sum_{k=1}^l y^*(k)\varphi_k(\tilde{x})$ and $\sum_{k=1}^l \varphi_k(\tilde{x})$, and, finally, in the output layer the estimate

$$\hat{y}(\tilde{x}) = \frac{\sum_{k=1}^l y^*(k)\varphi_k(\tilde{x})}{\sum_{k=1}^l \varphi_k(\tilde{x})} = \frac{\sum_{k=1}^l y^*(k) \prod_{i=1}^n \mu_{\tilde{x}_i(k)}(\tilde{x}_i)}{\sum_{k=1}^l \prod_{i=1}^n \mu_{\tilde{x}_i(k)}(\tilde{x}_i)}, \quad (4)$$

is forming, which coincides with (1) with the only difference, that instead of radial-basis functions multidimensional fuzzy-basis functions are used, that were formed of univariate FBF.

The scheme of fuzzy inference, which is realized by GRNFN can be presented as a logic equations system

$$\text{IF}(\tilde{x}_1.\text{IS}.A_1(1)).\text{AND}(\tilde{x}_2.\text{IS}.A_2(1)).\text{AND}.\dots.\text{AND}(\tilde{x}_n.\text{IS}.A_n(1)), \quad \text{THEN} \quad \hat{y}_1(\tilde{x}) = y^*(1)$$

$$\vdots$$

where the input to the plant $u(k)=\sin(2\pi k/25)+\sin(2\pi k/10)$ for $k=100$.

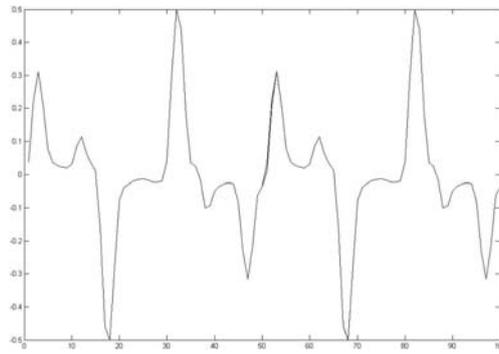


Fig.3 – Outputs of the original plant, GRNN and GRNFN for 50 signals in learning sample.

Numerical results of the experiment show that if learning sample consists of 50 signals, then both networks operates equally and have no mistakes. Fig.3 shows the plant and the outputs of GRNN and GRNFN and the differences between them are undistinguished. But if the number of signals which form the learning sample less than a half of all number of signals, GRNFN has the accuracy higher by 3-5% than GRNN.

CONCLUSIONS

General Regression Neuro-Fuzzy Network, that is generalization of conventional GRNN and adaptive fuzzy inference systems, is proposed in this work. This network is characterized by computational simplicity, interpretability of the results and ensures high accuracy in the nonlinear nonstationary processes identification and modeling problems.

References:

- [1] Specht D.E.: A general regression neural network. *IEEE Trans. on Neural Networks* 2 (1991) 568-576.
- [2] Nelles O.: *Nonlinear System Identification*. Springer, Berlin (2001).
- [3] Zahirniak D.R., Capman R., Rogers S.K., Suter B.W., Kabrisky M., Pyati V.: Pattern recognition using radial basis function network. *Proc. 6-th Ann. Aerospace Application of AI Conf.*, Dayton, OH (1990) 249-260.
- [4] Friedman J., Hastie T., Tibshirani R.: *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Springer, Berlin (2003).
- [5] Jang J.-S. R., Sun G.-T.: Neuro-fuzzy modeling and control. *Proc. IEEE* 83 (1995) 378-406.
- [6] Jang J.-S. R., Sun G.-T., Mizutani E.: *Neuro-Fuzzy and Soft Computing*. Prentice Hall, Upper Saddle River, NJ (1997).
- [7] Cios K.J., Pedrycz W. Neuro-Fuzzy algorithms// In: "Handbook on Neural Computation" – oxford University Press, 1997.- D1.3:1-D1.3:7.
- [8] Jang J.-S. R.: ANFIS: Adaptive-Network-based Fuzzy Inference Systems. *IEEE Trans. on Systems, Man, and Cybernetics* 23 (1993) 665-685.
- [9] Brown M., Harris C.J.: *Neural networks for modeling and control*. In: Eds. by C.J. Harris "Advances in Intellectual Control". Taylor and Francis, London (1994) 17-55.
- [10] Wang H., Liu G.P., Harris C.J., Brown M.: *Advanced Adaptive Control*. Pergamon, Oxford (1995).
- [11] Bishop C.M.: *Neural Networks for Pattern Recognition*. Clarendon Press Oxford (1995).
- [12] Narendra K.S., Parthasarathy K. Identification and control of dynamical systems using neural networks// *IEEE Trans. on Neural Networks* 1 (1990) 4-26.

Authors

Dr.-Ing. habil. Yevgeniy Bodyanskiy
Dr.-Ing. Irina Pliss
Researcher Nataliya Teslenko
Kharkiv National University of Radio Electronics
Lenin Av., 14
61166, Kharkiv, Ukraine
Phone: (+38) 057 7021890
E-mail: {bodya,pliss}@kture.kharkov.ua, ntntp@ukr.net

Dr.-Ing. habil. Peter Otto
Department of Informatics and Automation
Technical University of Ilmenau
PF 100565
98684 Ilmenau, Germany
E-mail: Peter.Otto@tu-ilmenau.de

Ye. Bodyanskiy / Ye. Gorshkov / V. Kolodyazhniy / P. Otto

Evolving Network Based on Double Neo-Fuzzy Neurons

INTRODUCTION

Hybrid computational intelligence systems, and mainly adaptive neuro-fuzzy systems, are widely used in the problems of analysis and processing of non-stationary signals of arbitrary nature under the uncertainty conditions. In most cases these systems have comparatively complex architecture (e.g. 5-layer ANFIS and the like) [1] which results in a complication and deceleration of the learning process in the problems that need to be solved in real-time.

To overcome these difficulties, a new approach called neo-fuzzy neuron (NFN) was proposed in [2–4]. The NFN architecture is quite similar to a conventional n -input formal neuron. However, instead of regular synaptic weights it contains nonlinear synapses NS_i , $i = 1, 2, \dots, n$, which are formed by a set of membership functions μ_{ji} , $j = 1, 2, \dots, h_i$ with tunable weight w_{ji} on each function.

The response of the NFN to the input signal vector $x(k) = (x_1(k), x_2(k), \dots, x_n(k))^T$ (here $k = 1, 2, \dots$ is the discrete time) is

$$y(k) = \sum_{i=1}^n f_i(x_i(k)) = \sum_{i=1}^n \sum_{j=1}^{h_i} \mu_{ji}(x_i(k)) w_{ji}(k), \quad (1)$$

where $w_{ji}(k)$ is the current value of the tunable weights at the time step k on a j -th membership function of i -th input signal component. As the error criterion for the learning of NFN a local quadratic error function is usually used

$$E(k) = \frac{1}{2} (d(k) - y(k))^2 = \frac{1}{2} e^2(k) = \frac{1}{2} \left(d(k) - \sum_{i=1}^n \sum_{j=1}^{h_i} \mu_{ji}(x_i(k)) w_{ji} \right)^2 \quad (2)$$

and the learning process itself is the minimization of this criterion using the gradient-based procedure

$$w_{ji}(k+1) = w_{ji}(k) + \eta e(k) \mu_{ji}(x_i(k)), \quad (3)$$

where $d(k)$ is the reference learning value, and η is the learning rate parameter which is usually chosen empirically and fully determines the speed of the learning process.

The membership functions of the NFN are formed as an array of triangular functions satisfying the following criterion

$$\sum_{j=1}^{h_i} \mu_{ji}(x_i(k)) = 1, \quad i = 1, 2, \dots, n \quad (4)$$

(the so called Ruspini partitioning). Thus, the network does not require a normalization layer.

The goal of this paper is to improve the approximating capabilities of the NFN by modification of its architecture and speed-up the learning procedure by a special selection of the learning rate parameter.

DOUBLE NEO-FUZZY NEURON

Consider the architecture of the double neo-fuzzy neuron (DNFN). Its architecture in a compact form is presented in Fig. 1.

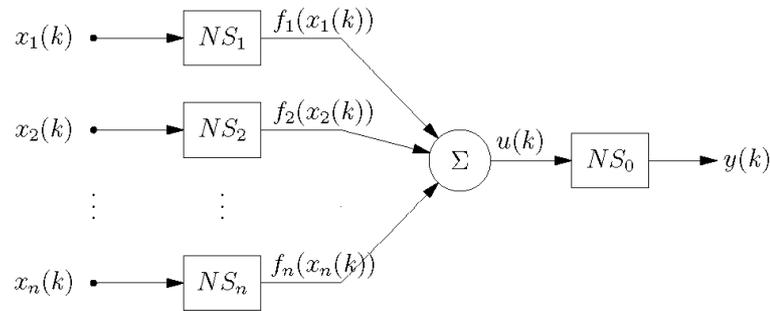


Fig. 1: Double neo-fuzzy neuron

Thus, the DNFN consists of two layers: the input layer of n nonlinear synapses NS_i with h_i membership functions and synaptic weights each, and the output layer formed by a nonlinear synapse NS_0 with h_0 membership functions μ_{l0} , $l = 1, 2, \dots, h_0$ and synaptic weights w_{l0} .

When the vector $x(k) = (x_1(k), x_2(k), \dots, x_n(k))^T$ is fed to the input of the DNFN, it produces the response in the form

$$y(k) = f_0(u(k)) = f_0\left(\sum_{i=1}^n f_i(x_i(k))\right) = \sum_{l=1}^{h_0} \mu_{l0}(u(k)) w_{l0} = \sum_{l=1}^{h_0} \mu_{l0}\left(\sum_{i=1}^n \sum_{j=1}^{h_i} \mu_{ji}(x_i(k)) w_{ji}\right) w_{l0}. \quad (5)$$

The output value of the DNFN is determined by the input vector and the values of $\sum_{i=1}^n h_i + h_0$ membership functions and corresponding tunable synaptic weights.

As we have already mentioned, the membership functions are of triangular form providing the Ruspini partitioning:

$$\mu_{ji}(x_i) = \begin{cases} (x_i - c_{j-1,i}) / (c_{ji} - c_{j-1,i}), & x_i \in [c_{j-1,i}, c_{ji}], \\ (c_{j+1,i} - x_i) / (c_{j+1,i} - c_{ji}), & x_i \in [c_{ji}, c_{j+1,i}], \\ 0 & \text{otherwise,} \end{cases} \quad (6)$$

$$\mu_{l0}(u) = \begin{cases} (u - c_{l-1,0}) / (c_{l0} - c_{l-1,0}), & u \in [c_{l-1,0}, c_{l0}], \\ (c_{l+1,0} - u) / (c_{l+1,0} - c_{l0}), & u \in [c_{l0}, c_{l+1,0}], \\ 0 & \text{otherwise,} \end{cases} \quad (7)$$

where c_{ji} , c_{l0} are the centers of the corresponding membership functions.

According to this partitioning, at each time step only two neighbouring membership functions of each nonlinear synapse are fired. Denote these functions by μ_{pi} and $\mu_{p+1,i}$ respectively. Then we can write

$$\begin{aligned} f_i(x_i(k)) &= \sum_{j=1}^{h_i} \mu_{ji}(x_i(k)) w_{ji} = \mu_{pi}(x_i(k)) w_{pi} + \mu_{p+1,i}(x_i(k)) w_{p+1,i} \\ &= \frac{c_{p+1,i} - x_i(k)}{c_{p+1,i} - c_{pi}} w_{pi} + \frac{x_i(k) - c_{pi}}{c_{p+1,i} - c_{pi}} w_{p+1,i} = a_i x_i(k) + b_i, \end{aligned} \quad (8)$$

where $a_i = \frac{w_{p+1,i} - w_{pi}}{c_{p+1,i} - c_{pi}}$, $b_i = \frac{c_{p+1,i} w_{pi} - c_{pi} w_{p+1,i}}{c_{p+1,i} - c_{pi}}$, and

$$u(k) = \sum_{i=1}^n a_i x_i(k) + b_i, \quad (9)$$

$$\begin{aligned} y(k) &= \sum_{l=1}^{h_0} \mu_{l0}(u(k)) w_{l0} = \mu_{l0}(u(k)) w_{p0} + \mu_{p+1,0}(u(k)) w_{p+1,0} \\ &= \frac{c_{p+1,0} - u(k)}{c_{p+1,0} - c_{p0}} w_{p0} + \frac{u(k) - c_{p0}}{c_{p+1,0} - c_{p0}} w_{p+1,0} = a_0 u(k) + b_0, \end{aligned} \quad (10)$$

where $a_0 = \frac{w_{p+1,0} - w_{p0}}{c_{p+1,0} - c_{p0}}$, $b_0 = \frac{c_{p+1,0} w_{p0} - c_{p0} w_{p+1,0}}{c_{p+1,0} - c_{p0}}$.

Hence, the DNFN provides a piecewise linear approximation of the unknown nonlinear function $d(k) = F(x(k))$ in the form

$$y(k) = a_0 \left(\sum_{i=1}^n a_i x_i(k) + b_i \right) + b_0. \quad (11)$$

The approximation is determined by the given set of membership functions and corresponding synaptic weights.

LEARNING ALGORITHM OF THE DNFN

To develop a learning algorithm for the DNFN parameters, consider the criterion (2) and

the gradient-based optimization procedure with a variable learning rate $\eta_l(k)$. Then we can write a simple algorithm for learning in the output synapse NS_0 :

$$\begin{cases} w_{l0}(k+1) = w_{l0}(k) + \eta_0(k)e(k)\mu_{l0}(u(k)), & l = p, p+1, \\ w_{l0}(k+1) = w_{l0}(k), & \forall l \neq p, l \neq p+1. \end{cases} \quad (12)$$

Thus, at each time step only two of the synaptic weights corresponding to the fired membership functions can be tuned.

In order to optimize the speed of the learning procedure, we propose using a one-step modification of the Levenberg-Marquardt algorithm with the Sherman-Morrison formula for inverse matrix computation as proposed in [5, 6]. This approach leads to the following procedure:

$$\begin{cases} w_{l0}(k+1) = w_{l0}(k) + r_0^{-1}(k)e(k)\mu_{l0}(u(k)), & l = p, p+1, \\ r_0(k+1) = \alpha r_0(k) + \mu_{p0}^2(u(k+1)) + \mu_{p+1,0}^2(u(k+1)), & 0 \leq \alpha \leq 1, \\ w_{l0}(k+1) = w_{l0}(k), & \forall l \neq p, l \neq p+1. \end{cases} \quad (13)$$

For a zero value of the forgetting factor α this procedure coincides with the optimal Kaczmarz-Widrow-Hoff algorithm, and with the Goodwin-Ramadge-Caines nonlinear identification algorithm possessing expressed smoothing properties for $\alpha = 1$. Varying of the parameter α provides tracking or filtering properties to the learning process.

It should be noted that if the learning data set $x(k)$, $d(k)$ is given *a priori*, the learning of the output nonlinear synapse NS_0 can be performed in a batch mode with the standard least squares method. In this case the learning process will be reduced to a single procedure. To learn the weights in the input layer nonlinear synapses, consider the criterion in the form:

$$E(k) = \frac{1}{2}(d(k) - f_0(u(k)))^2 = \frac{1}{2}(d(k) - f_0\left(\sum_{i=1}^n \sum_{j=1}^{h_i} \mu_{ji}(x_i(k))w_{ji}\right))^2. \quad (14)$$

Whence,

$$\frac{\partial E(k)}{\partial w_{ji}} = -e(k) \frac{\partial f_0(u(k))}{\partial u(k)} \frac{\partial u(k)}{\partial w_{ji}} = -e(k)a_0(k) \frac{\partial u(k)}{\partial w_{ji}}. \quad (15)$$

Taking into account (15) we can write to the following simple algorithm, which is a gradient-based optimization of the criterion (14):

$$\begin{cases} w_{ji}(k+1) = w_{ji}(k) + \eta_i(k)e(k)a_0(k)\mu_{ji}(x_i(k)), & j = p, p+1, i = 1, 2, \dots, n, \\ w_{ji}(k+1) = w_{ji}(k), & \forall j \neq p, j \neq p+1. \end{cases} \quad (16)$$

Denote

$$a_0(k)\mu_{ji}(x_i(k)) = \mu_{ji0}(x_i(k)). \quad (17)$$

Applying the technique described above, we can write the algorithm for learning of the input layer synaptic weights:

$$\begin{cases} w_{ji}(k+1) = w_{ji}(k) + r_i^{-1}(k)e(k)\mu_{ji0}(x_i(k)), & j = p, p+1, i = 1, 2, \dots, n, \\ r_i(k+1) = \alpha r_i(k) + \mu_{pi0}^2(x_i(k+1)) + \mu_{p+1,i0}^2(x_i(k+1)), & 0 \leq \alpha \leq 1, \\ w_{ji}(k+1) = w_{ji}(k), & \forall j \neq p, j \neq p+1, \end{cases} \quad (18)$$

which fully coincides with the procedure (13) by its structure. Thus, in fact, all the synaptic weights of the DNFN are learned using only a single algorithm.

NETWORK BASED ON DNFNs

The proposed DNFN is a basic building block of the evolving network shown in Fig. 2.

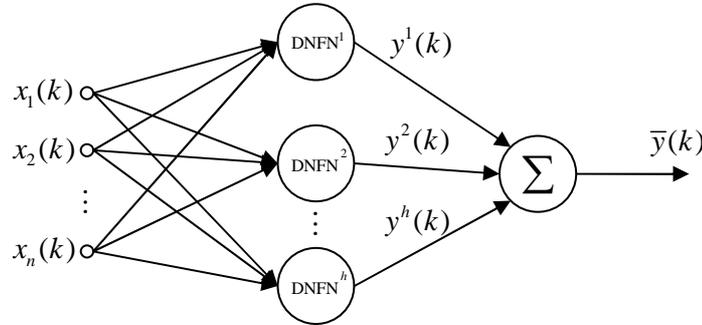


Fig. 2: Evolving network based on DNFNs

This network consists of a set of double neo-fuzzy neurons $DNFN^g$, $g = 1, 2, \dots, h$ combined into a layer. The network contains $(\sum_{i=1}^n h_i + h_0)h$ tunable weights and performs the following mapping:

$$\bar{y}(k) = \sum_{g=1}^h y^g(k) = \sum_{g=1}^h \sum_{l=1}^{h_0} \mu_{l0}^g \left(\sum_{i=1}^n \sum_{j=1}^{h_i} \mu_{ji}^g(x_i(k)) w_{ji}^g \right) w_{l0}^g \quad (19)$$

where $g = 1, 2, \dots, h$; $l = 1, 2, \dots, h_0$; $i = 1, 2, \dots, n$; $j = 1, 2, \dots, h_i$.

An essential feature of such network is the absence of the tunable weights in the output summation element synapses. This allows changing of the number of neurons by adding or removing neurons without the impact on the learning of the existing or newly added neurons. Each neuron learns independently from another according to the algorithm (18).

It can be easily seen that (19) corresponds, in fact, to the Kolmogorov's approximation scheme of a nonlinear function [7–9]. However, in contrast to the other neuro-fuzzy Kolmogorov's networks [5, 10, 11], the proposed approach possesses more flexibility, since it allows modification of the network structure directly during the learning process.

CONCLUSION

In the paper, an evolving network architecture based on double neo-fuzzy neurons is proposed. The proposed network performs Kolmogorov's approximation of an arbitrary nonlinear function, and possesses greater flexibility for network structure modification during the learning process in comparison to the other Kolmogorov's neuro-fuzzy systems. The learning algorithm is computationally simple and possesses filtering and tracking properties, which become significant in the processing of noisy non-stationary signals.

References:

- [1] J.-S. R. Jang, C.-T. Sun, and E. Mizutani. *Neuro-Fuzzy and Soft Computing — Computational Approach to Learning and Machine Intelligence*. Prentice Hall, Upper Saddle River, 1997.
- [2] Y. Yam, H. T. Nguyen, and V. Kreinovich. Multi-resolution techniques in the rules-based intelligent control systems: a universal approximation result. In *Proc. 14th IEEE Int. Symp. on Intelligent Control/Intelligent Systems and Semiotics ISIC/ISAS'99*, volume 3, pages 213–218, Cambridge, Massachusetts, September 15–17, 1999.
- [3] T. Yamakawa, E. Uchino, T. Miki, and H. Kusanagi. A neo fuzzy neuron and its applications to system identification and prediction of the system behavior. In *Proc. 2nd Int. Conf. on Fuzzy Logic and Neural Networks 'IIZUKA-92'*, pages 477–483, Iizuka, Japan, 1992.
- [4] T. Miki and T. Yamakawa. Analog implementation of neo-fuzzy neuron and its on-board learning. In N. E. Mastorakis, editor, *Computational Intelligence and Applications*, pages 144–149, WSES Press, Piraeus, 1999.
- [5] Ye. Bodyanskiy, Ye. Gorshkov, V. Kolodyazhniy, and P. Otto. Neuro-fuzzy Kolmogorov's network for time series prediction and classification. In *Lecture Notes in Artificial Intelligence*, volume 3698, pages 191–202, Berlin-Heidelberg-New York, Springer, 2005.
- [6] P. Otto, Ye. Bodyanskiy, and V. Kolodyazhniy. A new learning algorithm for a forecasting neuro-fuzzy network. *Integrated Computer-Aided Engineering*, volume 10, pages 399–409, 2003.
- [7] A. N. Kolmogorov. On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. In *Dokl. Akad. Nauk SSSR*, volume 114, pages 953–956, 1957.
- [8] V. Kůrková. Kolmogorov's theorem is relevant. In *Neural Computation*, volume 3, pages 617–622, 1991.
- [9] B. Igel'nik and N. Parikh. Kolmogorov's spline network. *IEEE Transactions on Neural Networks*, volume 14, pages 725–733, 2003.
- [10] V. Kolodyazhniy and Ye. Bodyanskiy. Fuzzy Kolmogorov's network. In *Lecture Notes in Computer Science*, volume 3214, pages 765–771, Heidelberg, Springer-Verlag, 2004.
- [11] Ye. Bodyanskiy, Ye. Gorshkov, V. Kolodyazhniy, and V. Poyedintseva. Neuro-fuzzy Kolmogorov's network. In *Lecture Notes in Computer Science*, volume 3697, pages 3–8, Berlin-Heidelberg, Springer-Verlag, 2005.

Authors:

Dr.-Ing. Habil. Yevgeniy Bodyanskiy

M.Sc. Yevgen Gorshkov

Control Systems Research Laboratory, Kharkiv National University of Radio Electronics, 14, Lenin Av., 61166, Kharkiv, Ukraine,

Phone: +38 057 702 1890

E-mail: bodya@kture.kharkov.ua, ye.gorshkov@gmail.com

Dr.-Ing. Vitaliy Kolodyazhniy

Institute for Psychology, University of Basel, Missionstrasse 60/62, CH-4055, Basel, Switzerland,

Phone: +41 61 267 0383

E-mail: v.kolodyazhniy@unibas.ch

Dr.-Ing. Habil. Peter Otto

Technical University of Ilmenau, Department of Informatics and Automation, PF 10 0565, 98684, Ilmenau, Germany

Phone: +49 0 3677/69-27 73

E-mail: peter.otto@tu-ilmenau.de

Ch. Wachten / Ch. Ament / C. Müller / H. Reinecke

Modeling of a Laser Tracker System with Galvanometer Scanner

INTRODUCTION

Laser trackers are systems that are capable of doing high accuracy 3D measurements. A laser beam is deflected by an actuator and follows the movements of a reflector in space. The position of the reflector can uniquely be defined by the distance measurement of an integrated interferometer and the measurement of the angle encoders of the actuator. Such systems offer the possibility of doing static as well as dynamic measurements. That is the reason why they are used in calibration tasks e.g. for robots or for coordinate measurement machines [1], [2].

In this paper we present a laser tracker system that is based on a galvanometer scanner. The accuracy of the system depends on the model as well as the calibration algorithm. Therefore a complete analytical model is developed that takes into account the internal behavior of the system and calculates the transformation of the measurements into 3D coordinates. An inverse kinematics is also developed to have the possibility to simulate the calibration process to reduce time and cost consuming experiments.

OPERATION PRINCIPLE OF THE LASER TRACKER SYSTEM

Fig. 1 shows the operation principle of the laser tracker system. A two-frequency HeNe laser emits two slightly different frequencies. The split frequency is about 2 MHz. An interferometer splits the laser beam into a reference beam (frequency f_2) and a measurement beam (frequency f_1). The measurement beam leaves the interferometer hits a non-polarizing beam splitter and is deflected by a magnetically driven actuator and is reflected by a retroreflector in space. On its returning path the reflected beam again hits the beam splitter and the beam position is analyzed by a quadrant diode. A part of the beam also interferes with the reference beam in the interferometer.

The quadrant diode generates analog output signals which are the inputs of a subsequent analog PI controller. This controller deflects the mirrors of the actuator in such a way that the laser beam follows the movements of the reflector. If the reflector is moved, the laser beam leaves its initial position – ideally the center position - on the quadrant

diode and the controller compensates the deviation. The controller has to be fast because the offset between the reference beam and the measurement beam must be below a certain threshold value to guarantee a stable function of the interferometer. With an analytical model the position of the reflector can be calculated. So, a direct kinematics is needed which transforms the angle and distance measurements into real coordinates.

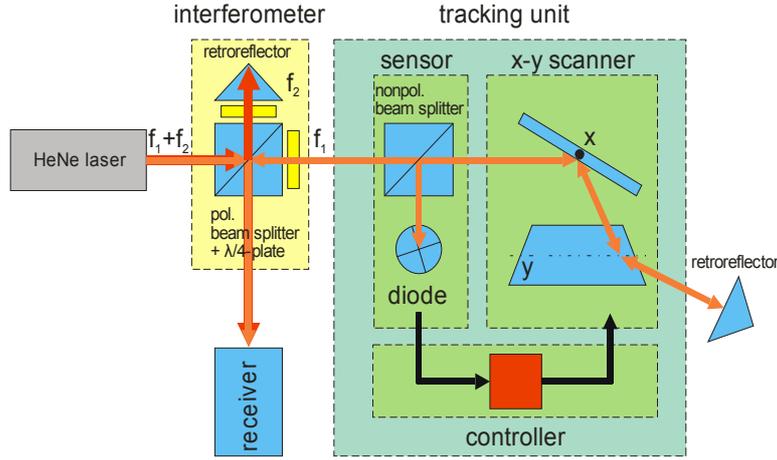


Fig. 1. Operation principle of the system

DEVELOPMENT OF THE DIRECT KINEMATICS

The analytical model of the system is an essential part because it is responsible for the accuracy of the system. It is important to consider all effects in contrast to [3] in the interferometer-scanner head to obtain a good model. First the interferometer function is neglected and the laser beam is propagated through the scanner head as proposed in [1]. So, the direction of the outgoing beam can be written as a line as shown in (1).

$$\vec{g}_{S1} = \vec{OC} + s \cdot \vec{r}_{S1} \quad (1)$$

The point C, the direction \vec{r}_{S1} and the parameter s have to be calculated (see Fig. 2). The direction is \vec{r}_{S1} obtained by a reflection of \vec{r}_L on the second mirror resulting in \vec{r}_{S21} and a reflection of \vec{r}_{S21} on the first mirror resulting in \vec{r}_{S1} . The normal vectors of the mirrors depend on the rotation of the corresponding axis. This is modeled by the Rodrigues formula given in (2) and (3).

$$\mathbf{R} = \begin{pmatrix} v_x^2 \cdot A + \cos \alpha & v_x \cdot v_y \cdot A - v_z \cdot \sin \alpha & v_x \cdot v_z \cdot A + v_y \cdot \sin \alpha \\ v_x \cdot v_y \cdot A + v_z \cdot \sin \alpha & v_y^2 \cdot A + \cos \alpha & v_y \cdot v_z \cdot A - v_x \cdot \sin \alpha \\ v_x \cdot v_z \cdot A - v_y \cdot \sin \alpha & v_y \cdot v_z \cdot A + v_x \cdot \sin \alpha & v_x^2 \cdot A + \cos \alpha \end{pmatrix} \quad (2)$$

$$\text{with } A = 1 - \cos \alpha \quad (3)$$

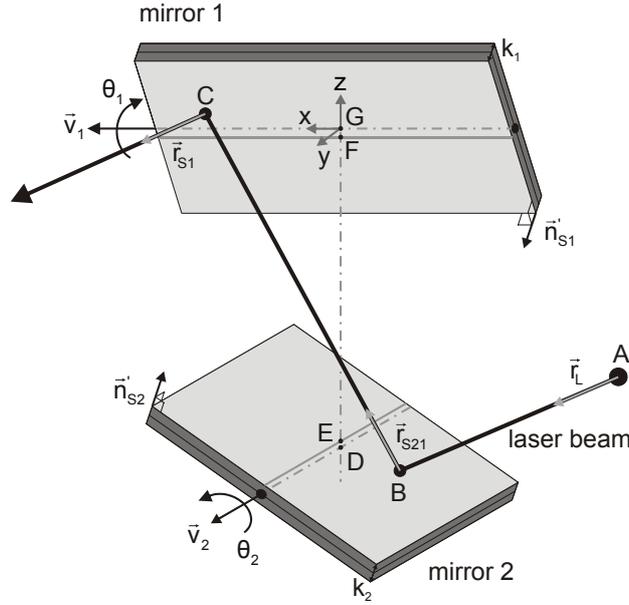


Fig. 2. Beam propagation through the scanner head

The vector $v_i = (v_x, v_y, v_z)$ in the matrix describes the direction of the rotation axis of the mirrors and α represents the rotation angle. With the updated normal vector in (4)

$$\vec{n}'_{Si} = \mathbf{R}(\vec{v}_i, \theta_i) \cdot \vec{n}_{Si} \quad (4)$$

and the law of reflection the direction \vec{r}_{S1} is calculated in (5).

$$\begin{aligned} \vec{r}_{S1} &= \vec{r}_{S21} - 2 \cdot (\vec{n}'_{S1} \cdot \vec{r}_{S21}) \cdot \vec{n}'_{S1} \\ &= (\vec{r}_L - 2 \cdot (\vec{n}'_{S2} \cdot \vec{r}_L) \cdot \vec{n}'_{S2}) - 2 \cdot (\vec{n}'_{S1} \cdot (\vec{r}_L - 2 \cdot (\vec{n}'_{S2} \cdot \vec{r}_L) \cdot \vec{n}'_{S2})) \cdot \vec{n}'_{S1} \end{aligned} \quad (5)$$

To obtain the point B, the line with its starting point A and the initial direction \vec{r}_L is cut with the mirror plane of the second mirror. Point B is calculated in (6).

$$\vec{OB} = \vec{OA} + \frac{(\vec{OD} + u_2 \cdot \vec{n}'_{S2} - \vec{OA}) \cdot \vec{n}'_{S2}}{\vec{r}_L \cdot \vec{n}'_{S2}} \cdot \vec{r}_L \quad (6)$$

To obtain point C the line with starting point B and direction \vec{r}_{S21} is cut with the mirror plane of the first mirror.

$$\vec{OC} = \vec{OB} + \frac{(\vec{OG} + u_1 \cdot \vec{n}'_{S1} - \vec{OB}) \cdot \vec{n}'_{S1}}{\vec{r}_{S21} \cdot \vec{n}'_{S1}} \cdot \vec{r}_{S21} \quad (7)$$

The parameter $u_i = k_i / 2 + a_i$ describes the distance between the model point of the rotation axis and the mirror plane. The parameter k_i represents the thickness of the mirrors and the parameter a_i is used to model a dumping of the mirror plane to the rotation axis. Now the interferometer function is considered. It is modeled with the parameter s of the

line in (1). The interferometer only measures relative distances to a given point. It is assumed that the retroreflector is put to a special point with known angles θ_{1start} and θ_{2start} and the initial length r_0 as shown in Fig. 3a). If the reflector is moved from point P_1 in Fig. 3b) to point P_2 then the interferometer detects a change in length that does not correspond to the length Δr because the beam path in the scanner head is neglected.

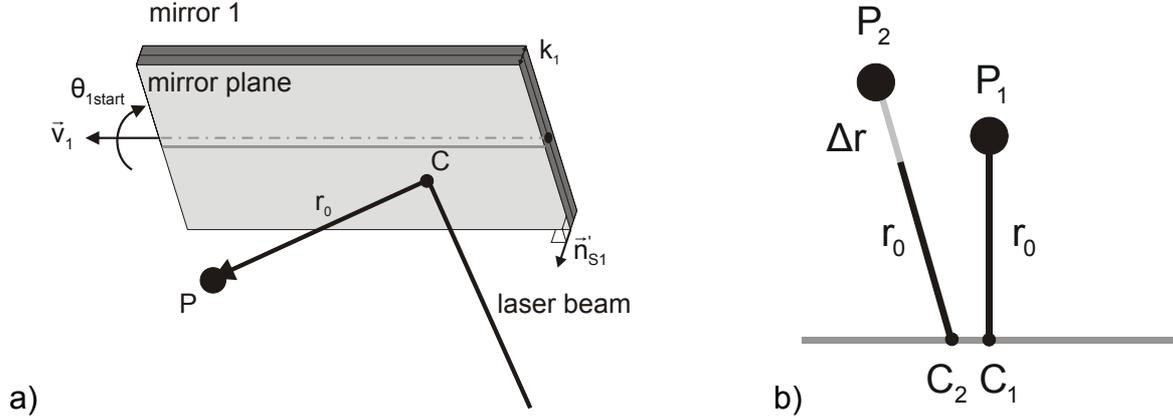


Fig. 3. Start condition a) and relative measurement of the interferometer b)

To overcome this problem the rotation angles of the mirrors have to be considered. So, it can be written for the parameter s :

$$s = r_0 + \Delta r_{gem} - \Delta AB - \Delta BC \quad (8)$$

with

$$\Delta AB = |\overline{OB}(\theta_2) - \overline{OA}| - |\overline{OB}(\theta_{2Start}) - \overline{OA}| \quad (9)$$

$$\Delta BC = |\overline{OC}(\theta_1, \theta_2) - \overline{OB}(\theta_2)| - |\overline{OC}(\theta_{1Start}, \theta_{2Start}) - \overline{OB}(\theta_{2Start})| \quad (10)$$

So, the outgoing laser beam is completely defined through its point C and its direction \vec{r}_{s1} and the parameter s . It is also of interest to calculate the inverse kinematics because a calibration algorithm can be tested without real experiments.

DEVELOPMENT OF THE INVERSE KINEMATICS

The inverse kinematics calculates to each reflector point in space the angles of the actuator and the distance of the interferometer from an arbitrary starting point. First the reflector is put to a point P in space as shown in Fig. 4. The laser beam is assumed to be deflected in space and does not hit the reflector. There is a distance a between point P and point F. The distance a is calculated in (11). The aim is to minimize this distance so that point P is coincident with point F as shown in (12).

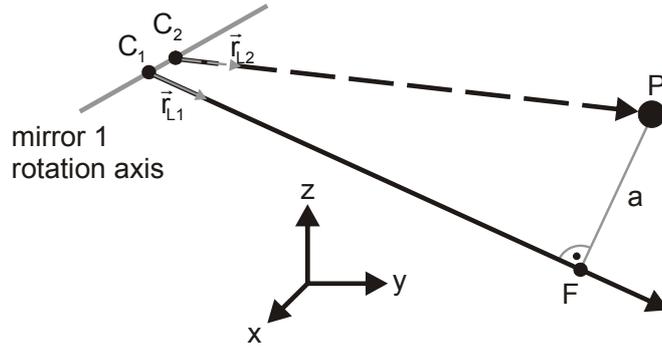


Fig. 4. Method to calculate the inverse kinematics

The angles of the mirrors are calculated with a numerical minimization algorithm implemented in Mathematica 5.2. This can only be done if the parameters of the laser tracker system are known.

$$a = |\overline{FP}| = |\overline{OP} - \overline{OC_1} - \left[(\overline{OP} - \overline{OC_1}) \cdot \frac{\vec{r}_{L1}}{r_{L1}^2} \right] \cdot \vec{r}_{L1}| \quad (11)$$

$$a \rightarrow 0 \Rightarrow \vec{r}_{L1} = \vec{r}_{L2} \wedge \overline{OC_1} = \overline{OC_2} \quad (12)$$

To complete the inverse kinematics the path change during the measurement is considered. Fig. 5 shows the beam path at the beginning of a measurement (black line) and during a measurement (gray line). The starting angles θ_{1start} and θ_{2start} are obtained by minimizing (8). So the points B_{start} and C_{start} are calculated. The relative path change is calculated by considering the system in sections.

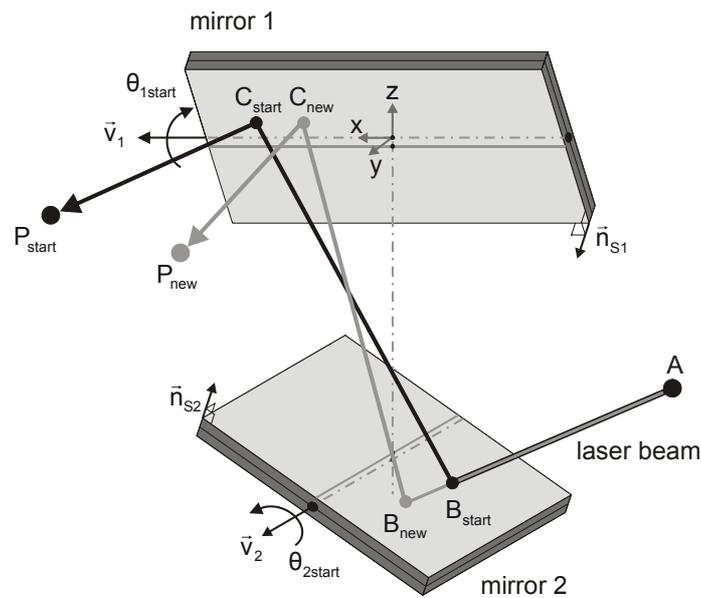


Fig. 5. Beam path at the beginning (start) and during (new) a measurement

The first section describes the path change between the reflector and the point C. The

second section describes the path change between the two mirrors and finally in the third section the path change between the incident laser beam and the second mirror is considered. This is resumed in (13).

$$\Delta r = \underbrace{|\overrightarrow{C_{new} P_{new}}| - |\overrightarrow{C_{start} P_{start}}|}_{\text{section 1}} + \underbrace{|\overrightarrow{B_{new} C_{new}}| - |\overrightarrow{B_{start} C_{start}}|}_{\text{section 2}} + \underbrace{|\overrightarrow{A_{new} B_{new}}| - |\overrightarrow{A_{start} B_{start}}|}_{\text{section 3}} \quad (13)$$

CONCLUSION

We have presented the direct and the inverse kinematics for a laser tracking system with galvanometer scanner. The analytical model covers nearly all effects in the scanner head. Furthermore it is possible with the inverse kinematics to simulate the calibration algorithm and to plan trajectories. This saves time and cost consuming experiments.

References:

- [1] J. R. R. Mayer and G. A. Parker, "A Portable Instrument for 3-D Dynamic Robot Measurements Using Triangulation and Laser Tracking", IEEE Transactions on Robotics and Automation, Vol. 10, No. 4, August 1994
- [2] H. Gander, M. Vinzce and J. Prenninger, "An external 6D sensor for Industrial Robots", IEEE/RSJ Intelligent Robots and Systems, Yokohama Japan, pp. 975-978, July 1993
- [3] J. Wang, Z. Hu, Y. Liu and J. Liang, "Experimental Research on Laser Tracking System with Galvanometer Scanner for Measuring Spatial Coordinates of Moving Target", Proceedings of SPIE - Optical Measurement and Nondestructive Testing: Techniques and Applications, Vol. 4221, pp. 10-14, October 2000

Authors:

Prof. Dr. Holger Reinecke,
Dr. Claas Müller,
Dipl.-Ing. Christian Wachten

Prof. Dr. Christoph Ament

University of Freiburg
Department of Microsystems Engineering
Laboratory for Process Technology
Georges-Köhler Allee 103
79110 Freiburg, Germany

University of Ilmenau
Institute for Automation and
Systems Engineering
Gustav-Kirchhoff Straße 1
98684 Ilmenau, Germany

Phone: +49 761 203 7304
Fax: +49 761 203 7352
E-mail: christian.wachten@imtek.uni-freiburg.de

Phone: +49 3677 69 2815
Fax: +49 3677 69 1434
E-mail: christoph.ament@tu-ilmenau.de

Kai Lüttkopf / Markus Abel / Bernd Eylert

Statistics of the truck activity on German Motorways

Introduction

In this contribution, we present results on data mining the data base of Toll Collect, the company operating the German toll system [1]. When using German motorways, a truck driver can either use the automatic system, or book his journey manually before entering the toll roads. Roughly, the automatic system consists of a central part, the computer centre and the OBUs as mobile components. About 540,000 trucks are equipped with this innovative device which determines its position by using Satellite technique (GNSS), and autonomously calculates the toll to be paid dependent on the distance traveled. The toll information is sent via mobile communication to the computer center of Toll Collect, and the logistics enterprise will be billed for it.

According to the German Data Protection Act (BDSG), Toll Collect stores the journey data for a certain time, after that time data will be deleted. One measure for the optimization of the operational procedures includes the statistics of the time an OBU spends on the German motorways, or the charged roads, respectively. This is important not only for operational questions. Furthermore it is a valuable basis to validate modern research on transport and logistics, not only for German roads. In this paper we present results only on the temporal behaviour, or activity, of the OBUs monitored in the Toll Collect data base. In a first step the data are anonymised in order to obey the data protection rules. In a second step we collect data pertaining to the OBUs. Thirdly, we identify journeys with and without breaks according to certain rules, specified below. Based on this classification we gather statistics on the start times and the duration of the OBU journeys with distinction of the weekdays. The statistics of the breaks between two subsequent journeys yields a clustering into two categories with clearly different activity. The direct impact on the operational conditions of Toll Collect lies in the optimization of the communication processes of the computer centre with the OBUs: we use the results as input for a simulation of the intelligent update process developed for the Toll Collect system [2].

Data Analysis

Storage of data in the Toll Collect data base is limited to a certain amount of time (120 days). The use of the data is restricted to toll-relevant events, as e.g., billing or system maintenance. In particular, individual-related data cannot be used for any further analysis. The analysis presented here serves as input to a simulation of the communication dynamics occurring when trucks receive new geo data. So, we had to use statistical procedures without access to personal information. We therefore worked with anonymized, and accumulated data to obtain a picture of the on-off behaviour of trucks on German motorways. Especially, we did not use any location-based information in the data.

You can try to figure out criteria to divide the trucks into clusters of different activity. One cluster could be day and night activity, another heavy or light trucks, or short-distance or long distance journeys, etc. Please note that the latter translate into short-time and long-time journeys in our analysis, since we do not access the spatial information. When running analyses in these directions, you quickly recognize that these typical categories are hard to apply. Let us consider day and night journeys: the trucks can depart in the afternoon and arrive late at night. When a break lies in between – you can ask whether it is a day journey, because of its early start, or is it a night journey, because the arrival is late. Or the short-time and long-time criterion: indeed, we find a continuum of durations for a journey, where it is virtually impossible to identify a particular scale on which you could speak of a “short” or a “long” journey. Further, to determine the duration of a journey, you are faced with the problem to identify breaks, which have to be taken at the latest after 4 hours of driving. If a truck stops on the motorway with the restrictions of the data we analyzed it is not possible to identify it as a resting truck.

Instead, we compare the duration of the journey with the typical duration of a similar journey, which is stored in a so-called distance matrix [3]. This matrix contains typical durations of truck journeys within the web of German motorways. If a truck takes much longer time for its travel, we assign a break to this journey. This is often found at night, when goods have to be delivered on schedule in the morning. Then trucks drive close to their destination and stay overnight on a nearby location on the motorway. Open questions are simply left out in our analysis and will be subject to more detailed investigations. Having corrected the journeys for the breaks we can proceed with the analysis. Basically, from the points made before we can determine three necessary identifiers of a journey: Start, duration and idle period between one journey and the next one. In addition, it is useful to divide the journeys further according to daily activity: If a truck drives once a day, most probably it has one driver who starts work, delivers and then takes a rest until the next day. If, however the On-Board-Unit is mounted in a truck where the drivers can change, there can be more journeys on one day. Then we have no means to distinguish the journeys, this results in journeys all around the day. An example for this are local delivery trucks which circle around without observable longer breaks, as e.g. for a company providing spare parts, or mail service in a populated region. If however, a single driver starts his truck twice a day, with a break of at least 8 hrs in between, two entries onto the motorways could be detected (in some rare cases even three – 3x8 hrs). In the following section we refer to these journeys as twice-a-day journeys.

Based on this classification, we define three categories: Once-a-day journeys, twice-a day journeys and all-around the clock profiles. The division into these sets is not unique: a truck can drive one day according to the once-a-day rule, and another day it is found in the twice-a-day category. In companies with many trucks, the vehicles are used on demand, with changing drivers, such that the switch between the groups is a usual process. A sketch of the classification is given in Fig. 1.

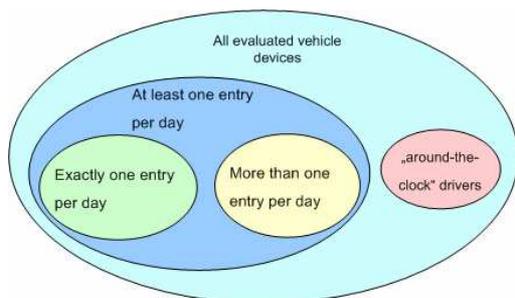


Figure 1: Composition of the categories.

Hereafter we consider a representative sample of 94,422 vehicles. For each category we analyze the start and duration of the journeys in statistical terms. The journeys have been corrected for the break duration as explained above. Let us consider the group with one, two or three journeys a day. All following analyses are done with SPSS [4]. We find the following histogram (all times in UTC).

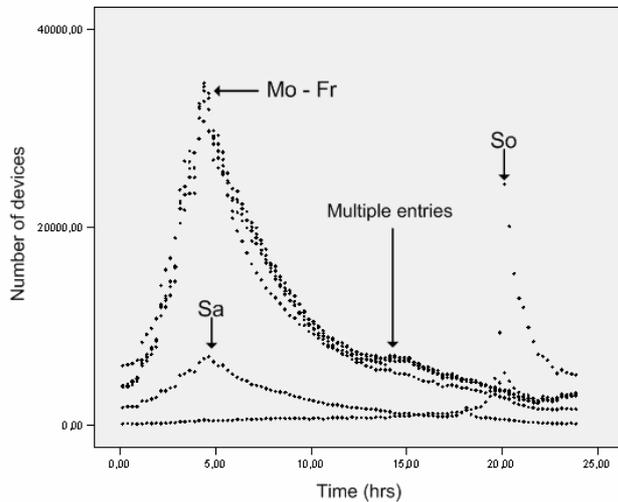


Figure 2: Distribution of start times of the journeys for each weekday. For working days, Monday to Friday, quite similar behavior are observed, on Saturday the activity is almost identical to the workdays, scaled by a factor 1:10. Sundays driving is banned before 20:00 UTC, this is seen as a sharp peak. Characteristic for the trucks with two journeys a day is the peak labeled “multiple entries”.

The major activity lies about 05:00 hrs UTC, then regular business hours start. In addition, trucks starting in the afternoon give rise to a small kink at 15:00 UTC. For Saturday, the same activity as for a working day is recovered, whereas the ban on driving on Sunday causes a completely different distribution with a peak at about 20:00 UTC, indicating the end of the ban. Of course, absolutely there are more trucks on the road during the night from Sunday to Monday.

One motorway entry per day

To get rid of the multiple daily journeys, we simply took out the respective trucks with 8 hrs break. As a result, the kink at 15:00 UTC disappears, as seen in Fig. 3. We conclude that this peak is due to multiple journeys.

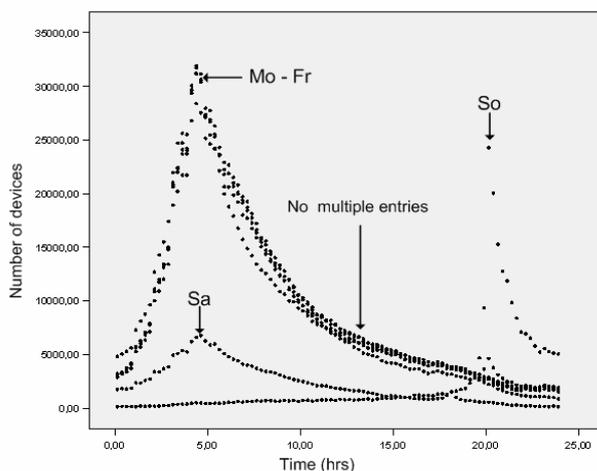


Figure 3: Same as Fig. 2, with multiple journeys per day subtracted.

Two motorway entries per day

Let us consider the multiple journeys alone, i.e., basically trucks driving twice a day with a break of at least 8 hrs in between. The distribution of entries over the day is displayed in Fig. 4. Clearly, there are two very distinct peaks at about 05:00 UTC and 15:00 UTC. Remarkably, you can notice another, peak at midnight. The dominant peak is at 05:00 UTC, and the other two approximately sum up to the height of the main peak. This means that the trucks typically enter in the morning and then wait either until afternoon or until midnight to start again.

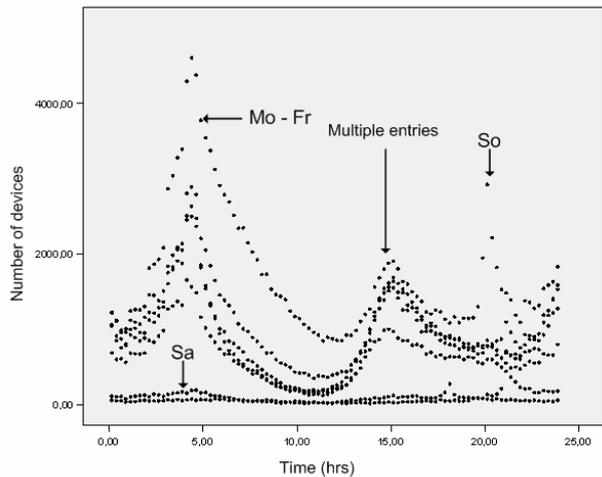


Figure 4: Twice-a-day entries. Trucks typically start in the morning and have a second journey starting the afternoon or at about midnight.

All-around the clock journeys

The last group, identified by us are all-around-the-clock journeys. They seem to drive locally or with two drivers in the way that no distinct break can be identified. The corresponding histogram is displayed in Fig. 5:

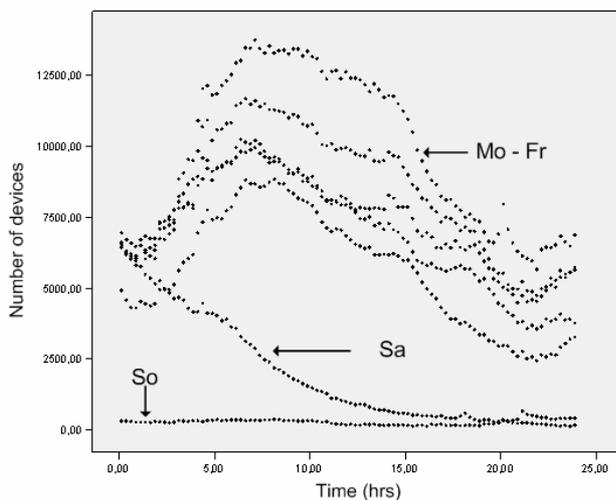


Figure 5: All-around-the-clock journeys show no peaked structure. Rather they are active during the normal working hours with slightly decreasing frequency. Late afternoon and night, the activity decreases to a smaller level, which arrives at a flat minimum at about 22:00 UTC.

Interestingly, the-all-around-the-clock journeys do not show the Sunday evening peak, neither do they show similarity on Saturday. This could indicate that the majority of the trucks in this category deliver locally, therefore no long-time journeys are necessary which would require a start at Sunday night.

Summary

In this paper, we present first results on data mining using the Toll Collect data base. In accordance with the data protection law, only accumulated, statistical analyses are possible on anonymized data without access to spatial information. Nevertheless we can obtain interesting results on the temporal activity of truck journeys for different weekdays. The statistics are used for the simulation of the communication of On-Board-Units for updating information in the devices. This is a regular task of Toll Collect which shall be controlled frequently. As a side product we get at a new division of truck activity suitable for the investigation of communication processes. We consider our work as important not only for toll collection. It rather shows that you can reach very interesting results on logistic processes with statistical evaluations.

References

- [1] www.toll-collect.de
- [2] Simulation of the Communication Behavior of the German Toll System, M. Abel and T. Lohfelder, this conferences proceedings.
- [3] D. Lohse, Verkehrsplanung. Grundlagen der Straßenverkehrstechnik und der Verkehrsplanung, Verlag Bauwesen, 1997.
- [4] www.spss.com

Authors:

Kai Lüttkopf
Toll Collect GmbH, Linkstr. 4,
10875 Berlin
Phone: +49 (0)30 74077 5747
Fax: +49 (0)30 74077 5554
E-mail: kai.luettkopf@toll-collect.de

Dr. habil. Markus Abel,
Toll Collect GmbH, Linkstr. 4,
10875 Berlin
Phone: +49 (0)30 74077 4913
Fax: +49 (0)30 74077 5554
E-mail: markus.abel@toll-collect.de

Prof. Dr. Bernd Eylert
TFH-Wildau , Bahnhofstr. 1
15745 Wildau
Phone: +49 (0) 171 523 0000
E-mail: beylert@igw.tfhwildau.de

K. Meissner / H. Hensel

A 3D process information display to visualize complex process conditions in the process industry

Human-machine Interface

1. Introduction

In the last few years, a clear increase in system complexity of production sites as a result of the technical developments in the process industry, and thus a not insignificant increase in the requirements of the operators of such systems, has been recorded. On the one hand, for cost reasons the degree of automation and the productivity of the systems have further increased, and on the other hand the number of system operators per system and shift have continued to be reduced. The information capacity to be supervised by an individual operator has thus been constantly rising [6]. The resulting mental load and the possible overloading of the operator described by Grams [4] result in an increasing error potential in the operator actions, particularly in emergency and stress situations. The solution methods already discussed in the technical literature in order to tackle this problem stretch on the one side from improved visualization, like the mass data display [3] or the 3D-visualization [8], which should unburden the user. On the other side, complex expert systems [1] were developed, which are supposed to actively support the operator during the processing by means of artificial intelligence methods, mathematical prognosis models and stored process knowledge [5].

In this paper, previous attempts at an improvement in visualization, in particular icon-based visualization [2] [7], are taken up and a new procedure with simple 3D icons is described. For this, the realizations of the cognitive science are to be used, in order to consider the abilities and the weaknesses of the human perception. It uses a cognitive scenic approach, which is to facilitate a recognizing and a classifying of critical process conditions or process tendencies of the operator. The second section presents the new icon and the imbedding into the 3D world. Here the cognitive scenic aspect of this

visualization method comes into being. The next section describes the process of the user evaluation and the conclusion forms a summary of the results as well as a look at further works.

2. Implementation

Based on the problems described above, a new type of icon based mass data display was developed. It contains a 3D view of several process values as figure 2 shows. Each value is represented by an icon made-up of a cylindrical symbol with an inner area, an outer ring, a trend line, a forecasting pointer and a label as shown in figure 1.

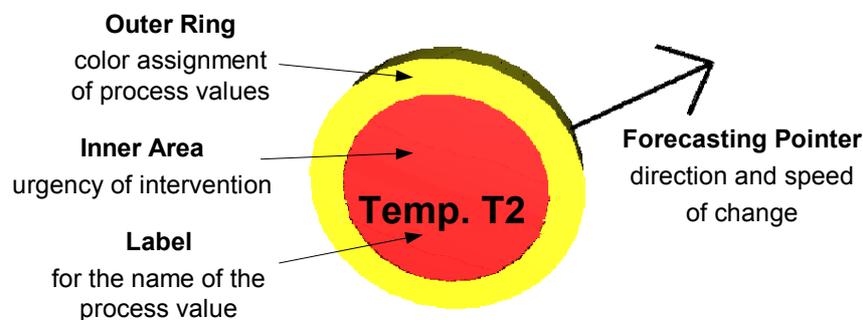


Figure 1: Design of the 3D visualization icon

The inner area shows the urgency of the operator intervention and changes his transparency. A clear area means that no intervention is necessary (figure 2a) and a red area represents a very high need of intervention by the operator (figure 2b). The colored outer ring represents the process value and shows its name in the legend. The forecasting pointer shows the operator the direction the process is running before and after an intervention. The trend line points up the history of the process value and the label will only be shown if the mouse pointer touches an icon. If a process value (e.g. Temp. T2) runs out of the limits, the inner area will change to red and the icon will enlarge by moving to the front. This scenic visualization indicates the importance of attention by the operator. Values close to the setpoint are smaller, with a transparent inner area and they are localized in the middle of the display. Therefore the operator can recognize the system state with one view only. If there are small icons in the middle of the display and outside the inner red area, the process is running within the specification [9].

In the test environment, the visualization is connected to the technical process via OPC. A self developed set of engineering tools gives the engineer the opportunity to generate the code for the visualization icons and to create the connection between the icons and the process values.

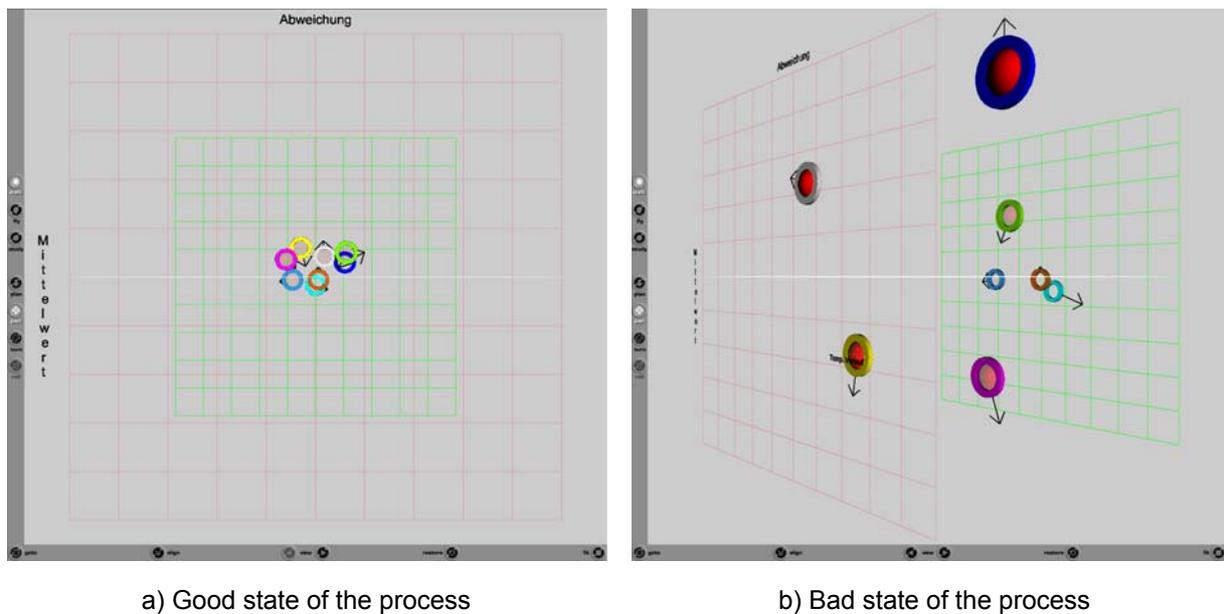


Figure 2: Views of the 3D process information display (3D-PID)

3. User Evaluation of the 3D process information display

Due to the range of the investigations and evaluations – altogether 92 questions in ten subject areas – only the most important results can be presented here. The entire results are the subject of a further publication, which will appear in due course. With the accomplished questioning, the possibilities of different representational forms for process data were compared with one another. The newly-developed 3D process information display was contrasted with the traditional representational forms of graphic displays, group pictures and curve representation. In the context of the evaluation, the implementation of the attributes shown in Section 2 and the following questions should be examined:

- How is the usefulness of the characteristics of the icon and the 3D-worlds evaluated?
- Is the navigation in the 3D-world reasonable for the user?

- How obstructive do the errors of the process topology seem?
- How is the overall acceptance of the new visualization system?

Except for the representation of the case history, whose use by the test persons was only partially recognized, all new characteristics of the 3D process information displays were highly evaluated, from positive to above average. This applies in particular to the representation of the urgency display. Navigation in the 3D world was regarded by the testers as unproblematic – this in particular due to the pre-defined observation points. In the view of the test persons, the recognition of abnormalities and the necessity of an intervention was greatly simplified and accelerated through this representation. Due to the connection to traditional process diagrams, the absence of the process topology in the 3D display is not evaluated negatively. The positive evaluation of the questions regarding the acceptance and the general impression of the 3D visualization system rounds off the results of the questioning. The system was consistently rated good to above average in these respects as well.

Parallel to the evaluation of the 3D visualization systems with the help of the questionnaire described above and testing at real systems, the visualization system was presented at various trade fairs, conferences and congresses. In the resulting discussions, in particular the two issues of process topology and coupling with the traditional visualization procedures were discussed. The problem of the strong abstraction of the visualization procedure and the resulting loss of the representation of the process topology was considered to be difficult. The general opinion was however that the gaining of additional information – like the urgency of an intervention, the speed and direction of the deviation and the quick detectability of process conditions – compensates for the loss of the process topology. The increased value of this representation method in contrast to the traditional visualization procedures was confirmed in the talks with experts regarding the function as a summary visualization. Through the coupling integrated in prototypes with the traditional visualization procedures, the requirement for a summary visualization is likewise fulfilled. Only the implementation of the 3D-world through VRML in prototypes was marked as problematic for reasons relating to performance.

4. Application of the 3-D Process Information Display

Having described the evaluation of the visualization system in the previous section, the process of the lime production, by means of which the 3D process information display was tested in practice, should be briefly considered at this point. The raw material for the production of quicklime is extracted through blasting operations in open-cast mining and transported by heavy-duty lorry or conveyor belt for processing. The lumps of calcium carbonate (CaCO_3) limestone are mixed with finer coal and filled into the top of a kiln. This lime-coal mixture slowly slides through the firing zones, individually fuelled by gas, oil or animal fat, in which the limestone is deacidified at a temperature of 1,000 °C to 1,300 °C. The quicklime is removed from the bottom of the kiln and transported to bunkers.

The process of lime production was chosen as many different process variables must be monitored in the kiln in order to ensure the quality of the lime. The process variables display varying high time constants, which complicate the monitoring and evaluation of the process state. In addition, the concentration of sensors for the monitoring of the process parameters within the processor is very high. This effect has up to now also been negative for the use of conventional display processes, as the process graphics were very overlaid and confusing. For this reason, it was suggested that the 3D process information display was deployed as a layered representation, and thus make the quality-relevant process data and information more quickly and simply available to the operator. Both the downtime of the entire plant as well as the reaction times during disruptions could be reduced through the use of the 3D process information display.

5. Summary and Perspective

Through the investigations described in Section 2 and the procedure presented in Section 3, it could be shown that an icon-based cognitive scenic visualization can provide the user, in a simple manner, with a summary representation of the information required to assess the process state. The evaluation in Section 3 shows a high acceptance of the abilities of the 3D-PID by operators and students. The operator is relieved from the influence of stress, allowing potential sources of errors to be minimized. Cognitive scenic visualization should not replace existing visualization forms

but rather supplement them, and be understood as abstract and superior visualization components. The number of markers recognizable by the operator as problem-free, the influence of three-dimensional visualization on the mental load of the operator and the possibilities of navigation in the three-dimensional information area are the subject of further investigations.

References:

- [1] Berheide, W.; Dörfel, G.; Döring, B.: Wissensbasierte Benutzerunterstützung in komplexen Mensch-Maschine-Systemen. In: *Verlässlichkeit von Mensch-Maschine-Systemen*; Hrsg. Willumeit, H.-P.; Kolrep, H.; Berlin: Technische Universität Berlin, 1995, pp. 61 – 72
- [2] Chernoff, H.: The Use of Faces to Represent Points in k-Dimensional Space Graphically. *Journal of American Statistical Association*, Vol. 68, (1973), pp. 361 – 368
- [3] Elzer, P. F.: Neuere Entwicklungen auf dem Gebiet der Mensch-Maschine-Schnittstellen: at - Automatisierungstechnik, 49 (2001), No. 4; München: Oldenbourg, pp. 178 – 186
- [4] Grams, T.: Bedienfehler und ihre Ursachen (Teil 1). In: *atp – Automatisierungstechnische Praxis*, Vol. 40 (1998), No. 3; München: Oldenbourg, pp. 53 – 56
- [5] Johannsen, G.; Architecture of Man-Machine Decision Making Systems. In: *Intelligent Decision Support in Process Environments*, Hrsg. Hollnagel, E.; Mancini, G.; Woods, D.D.; Berlin, New York: Springer, 1986, pp. 328 – 339
- [6] Komischke, T.; Informationsüberflutung in der Prozessführung. In: *atp – Automatisierungstechnische Praxis*, Vol. 45 (2003), No. 9; München: Oldenbourg, pp. 74 – 79
- [7] Pickett, R. M.; Grinstein, G.: Iconographics Displays for Visualization Multidimensional Data. *Proceedings IEEE Conference on Systems, Man and Cybernetics*, Peking, 1988, pp.514 – 519
- [8] Wittenberg, C.: Das Konzept der virtuellen Prozessvisualisierung für die industrielle Prozessführung. In: *at – Automatisierungstechnik*, Vol. 50 (2002), No. 4; München: Oldenbourg, pp. 178 – 186.
- [9] Meißner, K.; Hensel, H.: Visualisierung komplexer Prozesszustände durch den Einsatz von 3-D-Methoden. Hrsg. VDI, VDI-Berichte 1883 (2005), Düsseldorf, VDI-Verlag, pp. 91 – 98,

Authors:

Dipl.-Ing. (FH) Knut Meissner
Institute for Computer Science and Automation Systems
Dornbergsweg 2
38855 Wernigerode
Phone: ++49 3943 6259-745
Fax: ++49 3943 6259-747
Email: kmeissner@hs-harz.de

Prof. Dr.-Ing. Hartmut Hensel
University of Applied Studies and Research
Friedrichstraße 57 – 59
38855 Wernigerode
Phone: ++49 3943 659 313
Fax: ++49 3943 659 399
Email: hhensel@hs-harz.de

Recent Advances in the Estimation of Pointing Poses on Monocular Images for Human-Robot Interaction

1 Introduction and Motivation

In recent years, a lot of research work has been done to develop intelligent mobile robot systems, which can interact even with non-instructed users, making the robots suitable for applications in everyday life. Besides the verbal communication also the non-verbal communication plays a very important role in a dialog between humans. To the knowledge of the authors, only a few projects have already successfully integrated non-verbal communication parts in an interactive dialog on their mobile robots. In the work presented in this paper, we show how a basic non-verbal communication (more precisely: the problem, of instructing a mobile robot by the use of pointing gestures/poses) can be realized on a mobile robot system.

Some approaches already exist which focus on integrating gesture recognition into Man-Machine-Interfaces. In the works of Rogalla et al.[1], Paquin and Chohen [2] and Triesch and v.d. Malsburg [3] different approaches to detect and classify human gestures and poses are presented. However, most of this work concentrates on distinguishing different gestures, creating a command alphabet for robot control. A much more intuitive and smoother way to direct the robot is through pointing directly at the target position on the ground. In [4, 5] for the first time we presented an approach, which allows to direct a mobile robot to a certain position by means of such pointing poses. The system presented in [4, 5] was capable of estimating the target point of the pointing gesture on the floor with a low error, but could only operate in environments with unstructured background and ideal lighting conditions. Besides a computation time of 3-4 seconds was required for the estimation of a single target. These constraints conflict with the requirements for the usage of this approach in robotic real world applications. Therefore, in this paper we present several improvements on this approach making it possible to estimate the target point of a pointing pose in highly structured environments with variable lighting conditions with a computation time of only 80ms.

This paper is organized as follows: After this introduction, Section 2 describes the ground truth of the training data and the robot used for the experiments. Section 3 explains, how the pointing poses can be estimated and how our entire system is designed. In Section 4 the experiments and results will be presented. The paper ends with conclusions and a short outlook in Section 5.

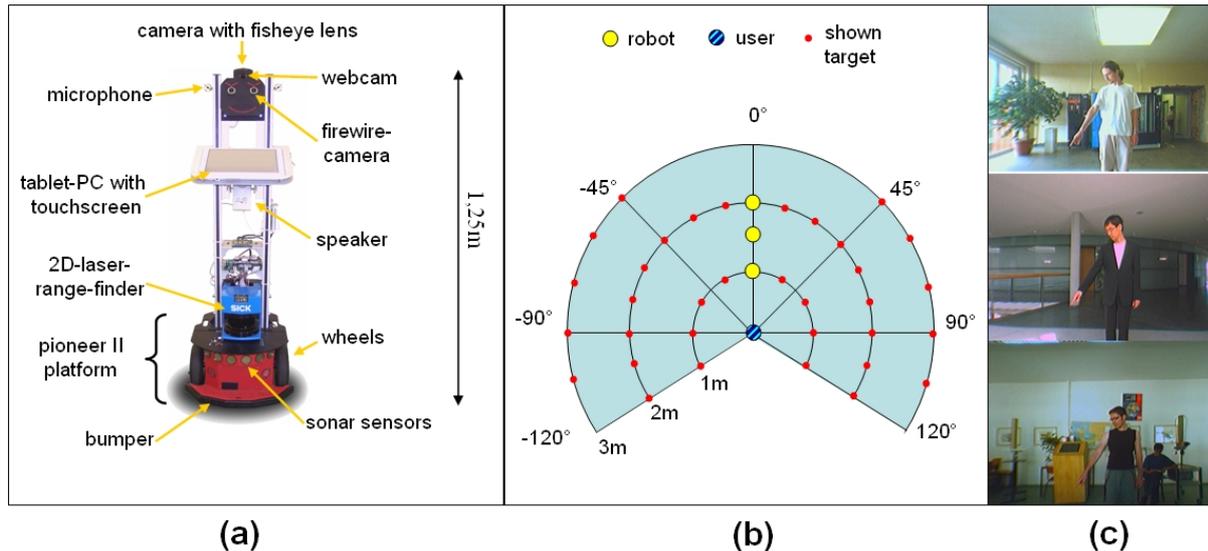


Figure 1: Image (a) displays our robot HOROS used for experimental investigation of the pointing pose estimation is displayed. The images for the estimation of the pointing target were taken with the front camera (located in the right eye). Image (b) displays the configuration used for recording the ground truth training and test data. The subject stood in front of the robot and pointed at one of the marked targets on the ground. The distance of the robot to the subject varied between 1 m and 2 m. Image (c) displays some examples of pointing poses recorded with the camera of the robot.

2 The Robot HOROS and the Ground Truth

The approach described in this paper was developed and tested on our mobile robot HOROS (HOMe ROBot System). HOROS' hardware platform is an extended Pioneer II based robot from ActivMedia (see Fig. 1(a)). Because one objective of our project is the development of a low-cost prototype of a mobile and interactive robot assistant, we are especially interested in vision technologies with a good price-performance ratio. Therefore, the two low-cost frontal cameras were utilized instead of a high-end stereovision system. We were interested if it would be possible to robustly estimate a target position at the floor from a pointing pose using only inexpensive hardware and monocular images.

A labeled set of images of subjects pointing to target points on the floor was required to train the system. We encoded the target points on the floor as (r, φ) coordinates in a subject-centered polar coordinate system (see Fig. 1) and placed the robot with the camera in front of the subjects. Moreover, we limited the valid area for targets to the half space in front of the robot with a value range for r from 1 to 3m and a value range for φ from -120° to $+120^\circ$. Figure 1 shows the configuration we chose for recording the training data. The subjects stood at distances of 1, 1.5 and 2m from the robot. Three concentric circles with

radii of 1, 2 and 3m are drawn around the subject, being marked every 15°. The subjects were asked to point to the markers on the circles in a defined order and an image was recorded each time. All captured images are labeled with distance, radius and angle, thus representing the ground truth used for training and for the comparing experiments with human viewers (see Section 4). This way, we collected a total of 2.340 images of 26 different interaction partners (90 different poses for each subject). This database was divided into a training subset and a validation subset containing two complete pointing series (i.e. two sample sets each containing all possible coordinates (r, φ) present in the training set). The latter was composed from 7 different persons and includes a total of 630 images. This leaves a training set of 19 persons including 1710 samples.

3 Estimation of Pointing Poses

Since the interaction partners standing in front of the camera can have different body height and distance, an algorithm had to be developed that can calculate a normalized region of interest, resulting in similar subimages for subsequent processing. We use an approach suggested by [4, 5] to determine the region of interest (ROI) by using a combination of face-detection (based on the Viola & Jones Detector cascade [9]) and some empirical factors. With the help of a multimodal tracker [4, 5] implemented on our robot, the direction and the distance of the robot to the interacting person can be estimated. The cropped ROI is scaled to 160*100 pixels for the body and the arm and 160*120 pixels for the head of the user. Additionally, a histogram equalization is applied to improve the feature detection under different lighting conditions. The preprocessing operations used to capture and normalize the image are shown in Fig. 2. To reduce the effects of different backgrounds, in the improved version we used a simple Background Subtraction algorithm and tested its influence on the pose estimation result in comparison with our approach in [4, 5] where no Background Subtraction was used. On the normalised image regions a feature extraction is used for the approximation of the target position the user is pointing to. In our work Gaborfilters of different orientations and frequencies, bundled in Gaborjets that are located on several fixed points in the selected ROIs, are used. The several steps of preprocessing and feature extraction used in our comparison are shown in Fig. 2.

In [4, 5] a cascade of several Multi-Layer Perceptrons (MLP) was used to estimate the target point from the extracted features. However other techniques are also often used for the estimation of certain human poses (but mostly not on mobile robots but under predefined observation conditions). Nölker and Ritter [10] used a Local Linear Map (LLM) and a Parametrized Self-Organizing Map (PSOM) to estimate the target of a pointing pose on a screen the user is pointing to. Krüger and Sommer [7] utilized Gaborfilters and a LLM to estimate the head pose, while Stiefelhagen [8] presented a system that works on edge-filtered images and uses a MLP for head pose estimation. Therefore, for this paper we implemented and compared several selected neural approaches, which all were trained

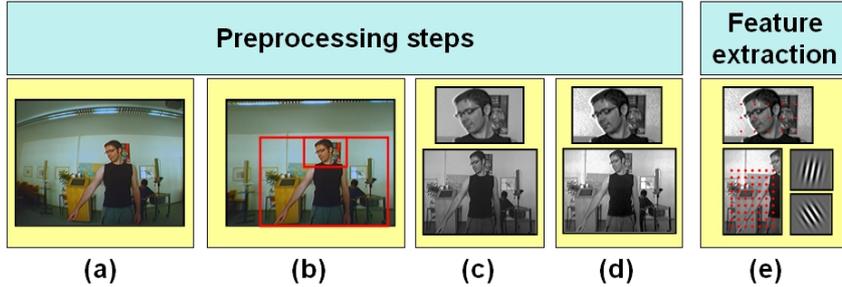


Figure 2: Steps of preprocessing and feature extraction: the raw distorted image of the lowcost camera in the robot’s eye (a) is transformed into an undistorted image and the face of the user is detected by means of [9] (b). Based on the height of the face in the picture and the distance of the user, two sections of the image are extracted and transformed into grayscale images (c). On these images a histogram equalization is used (d). Subsequently, features are extracted by Gaborfilters placed at pre-defined points of the image (marked as red dots in (e)). A Background Subtraction can optionally be used between steps (d) and (e).

and tested with the same sets of training and test data. This way we are able to give an overview of the suitability of the different approaches for the task of estimating a pointing pose on a monocular image. We compared a k-Nearest-Neighbour method (kNN), a Neural Gas network (NG, [11]), a Self-Organizing Map (SOM, [12]), a Local Linear Map (LLM, [13]) and Multi-Layer Perceptrons (MLP, [4, 5]).

4 Experiments and Results

To have a simple reference for the quality of the estimation, 10 human subjects were asked to estimate the target point of a pointing pose on the floor. At first, the subjects had to estimate the target on a computer screen where the images of the training data set were displayed. The subject had to click on the screen at the point where they estimated the target. Thus, the subjects were estimating the target on the images having the same conditions as the different estimation systems. Second, we determined the estimation result the subjects achieved under real world circumstances. Here, each subject had to point at a target on the ground and a second subject had to estimate the target. At first the recognizing person used both of their eyes to estimate the target, later we blindfolded one of the eyes and the person estimated the target again under monocular conditions. The results of the human based reference experiments are included in Fig. 3. The label *Human (screen)* refers to the experiments on the computer screen and the labels *Human (2 eyes)* and *Human (1 eye)* refer to the results under real world conditions.

The results of the several approaches for estimating the target position are shown in Fig. 3. As described in Sect. 2 the ground truth data is a tuple (r, φ) with the target radius r and the target angle φ . For the correct estimation of the target point, r as well as φ had to be estimated correctly. We defined the estimation result to be correct if r differed less

target point estimation (correct radius <i>and</i> correct angle)					
correct samples in %	k-NN	NG	SOM	LLM	MLP
Gaborfilters	11,12%	4,70%	6,70%	11,76%	29,16%
Gaborfilters and BG Subtraction (BGS)	22,28%	17,72%	15,34%	23,53%	44,90%
Gaborfilters and Discriminant Analysis	17,69%	9,38%	11,66%	16,04%	28,14%
Gaborfilters, BGS and Discriminant Analysis	34,72%	22,66%	23,66%	31,74%	50,63%

Human (2 eyes)
62,90 %

Human (1 eye)
40,75 %

Human (screen)
37,50 %

Figure 3: The results for the estimation of the target point of the pointing pose. The target point is determined by the radius r and the angle φ . For each method the percentage of the targets estimated correctly is determined. The results of the human viewers (on computer screen, and in reality (with both eyes "Human (2 eyes)" and with one eye blindfolded "Human (1 eye)")) are given for comparison. Methods that achieve a result equal to that of the human viewers are marked with a shaded background with different colors.

than 50cm from the ground truth radius and φ differed less than 10° from the ground truth angle. Figure 3 shows the results for a correct estimation of both values.

Every of the five selected approaches was trained and tested on the same training data set. For each system, we used four different feature extraction strategies: first only Gaborfilters were utilized, second we combined Gaborfilters with an additional Background Subtraction to reduce the effects of the different cluttered backgrounds in the images. Third, we used only those Gaborfilters that had a high discriminant value extracted by means of a Discriminant Analysis executed over all predefined Gaborfilter positions. Fourth, we combined Gaborfilter, Background Subtraction and utilized only the relevant features extracted by the Discriminant Analysis mentioned above.

The results demonstrate, that a cascade of several MLPs as proposed in [4, 5] is best suited to estimate the target position of a user's pointing pose on monocular images. A Background Subtraction and the information delivered by a Discriminant Analysis can be used to improve the results. The best system is capable of estimating r as good as humans with their binocular vision system in a real world environment and even better than humans estimating the target on 2D screens. The estimation of φ does not reach equally good values. The system is able to reach a result equally to humans on 2D screens or humans with one eye blindfolded, but it is not able to estimate the angle as good as humans in a real world setting using both eyes. This is because the estimation of the depth of a target in a monocular image is difficult for both, human and function approximators.

5 Conclusion and Outlook

In this paper we presented an extension to our approach introduced in [4, 5]. The major problems of the old approach (bad results in environment with structured background and a computation time which exceeds real-time requirements) could be solved. Extensive experiments with different neural function approximators have shown, that the MLP-based approximator leads to the best result. The realized approach is able to estimate a referred position on the ground based on monocular images with an accuracy nearly equal to humans and work in real-time (80ms). This enables the user to direct a mobile robot system into a target position based on pointing gestures only.

References

- [1] Rogalla, O. and Ehrenmann, M. and Zöllner, R. and Becher, R. and Dillmann, R.: *Using Gesture and Speech Control for Commanding a Robot Assistant*. In In Proc. of the 11th IEEE Int. Workshop on Robot and Human Interactive Communication, ROMAN. (2002) 454–459.
- [2] Paquin, V. and Cohen, P.: *A Vision-Based Gestural Guidance Interface for Mobile Robotic Platforms*. In Proc. of the Workshop on HCI, Computer Vision in Human-Computer Interaction, ECCV (2004) 39–47.
- [3] Triesch, J. and von der Malsburg, C.: *A System for Person-Independent Hand Posture Recognition against Complex Backgrounds* In IEEE Trans. Pattern Anal. Mach. Intell, Vol. 23, Number 12 (2001) 1449–1453.
- [4] Gross, H.-M. and Richarz, J. and Müller, S. and Scheidig, A. and Martin, C.: *Probabilistic Multi-modal People Tracker and Monocular Pointing Pose Estimator for Visual Instruction of Mobile Robot Assistants*. In: Proc. of the IEEE World Congress on Computational Intelligence, WCCI (2006) 8325–8333.
- [5] Richarz, J., Martin, C., Scheidig, A., Gross, H.-M.: *There You Go! - Estimation Pointing Gestures in Monocular Images for Mobile Robot Instruction*. In: RO-MAN 2006 - 15th IEEE Int. Symposium on Robot and Human Interactive Communication (2006) 546–551.
- [6] Takahashi, K., Tanigawa, T.: *Remarks on real-time human posture estimation from silhouette image using neural network*. In: Proceedings of the International Conference on Systems, Man and Cybernetics: The Hague (2004) 370–375.
- [7] Krüger, V., Sommer, G.: *Gabor wavelet networks for efficient head pose estimation*. In: Image and Vision Computing, Vol. 20, Number 9-10, August (2002) 665–672.
- [8] Stiefelhagen, R.: *Estimating Head Pose with Neural Networks - Results on the Pointing04 ICPR Workshop Evaluation*. In: Pointing 04 ICPR Workshop, Cambridge, UK (2004).
- [9] Viola, P. A., Jones, M. J.: *Rapid object detection using a boosted cascade of simple features*. In: Proc. of the Conf. on Computer Vision and Patter Recognition, Munich, Germany (2001) 511–518.
- [10] Nölker, C., Ritter, H.: *Illumination Independent Recognition of Deictic Arm Postures*. In: Proc. of the 24th Annual Conference of the IEEE Industrial Electronics Society, Aachen, Germany (1998) 2006–2011.
- [11] Martinetz, T., Schulten, K.: *A Neural-Gas Network Learns Topologies*. In: Proc. of the the ICANN 1991, Helsinki, Finland (1991) 397–402.
- [12] Kohonen, T.: *Self-organized formation of topologically correct feature maps*. In: Biological Cybernetics, Vol. 43 (1982) 59–69.
- [13] Ritter, H.: *Learning with the Self-Organizing Map, Eds.: T. Kohonen et al., Artificial Neural Networks*. (1991)

Author Information:

Dipl.-Inf. Frank-Florian Steege

Dipl.-Inf. Christian Martin

Prof. Dr.-Ing. Horst Michael Groß

Neuroinformatics and Cognitive Robotics Lab, Ilmenau Technical University

G-Kirchhof-Str. 2, 98693 Ilmenau, Germany

Tel: ++49 +3677 69 2858

Fax: ++49 +3677 69 1665

E-mail: frank-florian.steege@stud.tu-ilmenau.de, christian.martin@tu-ilmenau.de

A. Gonzalez / H. Fernlund / J. Ekblad

After Action Review by Comparison – an Approach to Automatically Evaluating Trainee Performance in Training Exercise

ABSTRACT

After-action Review (AAR) is an effective tool to evaluate and improve human performance in tactical training exercises. However, when the exercises grow in size, and possibly reside in several locations, providing feedback to the majority of the participants can be complicated. It requires extensive time and resources, and the review might be limited to the few most important tactical decisions made. To get the most out of AAR, it should be complemented with automated systems that help the instructor/operator (I/O) generate the appropriate feedback for each individual trainee. To improve the ability of the I/O to provide we investigated the development of intelligent tools to compose a *Smart After-Action Review* (SmartAAR) technology suite. This approach is based upon the concept of AAR-by-comparison. That is, we seek to build agents that represent expert human performance and then use them as benchmarks during execution of the tactical exercise, to which the trainee performance is compared continuously and possibly in real time. By pairing each trainee with his own 'personal' expert agent counterpart, individual feedback can be provided to each trainee. This paper presents a novel concept based on two dimensions: 1) comparing the spatio-temporal location of the trainee and 2) comparing the context in which the trainee finds himself. These could serve as a basis for automatic and self-instructing AAR.

Introduction

Evaluation of human performance against stated objectives is an important function in successful organizations. There are several different areas where evaluation of human performance is particularly important, such as sporting events (e.g., football),

rescue operations (e.g., fire fighting) or military operations. In such actions, a human must perform against other humans acting as adversaries, often in a life-or-death struggle. Especially challenging is evaluation of teamwork performances. We describe our research effort that has developed a novel approach for automatic support of human performance evaluation. The area of interest in this research was military exercises but the results are applicable to many different domains that employ simulator and/or live exercise training for its participants, such as sports training, etc. Particularly applicable are tasks that are tactical in nature, even if not adversarial, such as for example, driver training or flight training. However, in this paper, we refer only to the military training domain.

In military training, it is important that the trainee be provided with timely and individual-specific feedback in order to improve his performance in future missions. *After-Action Review* (AAR) is the process through which this feedback is traditionally provided. AAR is an important tool to evaluate the individual as well as collective task performances for trainees after the training session is completed. The *observer/controller* (OC) who normally provides the feedback must be aware of the actions executed by the trainee, and be able to determine their correctness. However, it is unrealistic to expect the OC to continuously monitor every single individual participant in the exercise [1]. This is especially true for large training exercises with many participants. The approach to deal with this problem is to conduct informal reviews by the leaders in the internal chain of commands, prior to the formal AAR. However, the leaders typically do not have a complete picture of all the events and the trainees' actions therein. The art of AAR is then to get each participant to perform accurate self-evaluations [2], in order to obtain a complete, high quality AAR experience. There is increasing interest in virtual simulations where the participants can be either real or virtual and in different training locations. Conducting constructive AAR in such exercises becomes even more difficult.

To get the most out of AAR, we believe it should be complemented with automated systems that generate the appropriate feedback for each individual trainee. To improve the ability of the OC to provide feedback, this research investigates the use of intelligent tools to compose a Smart After-Action Review (SmartAAR) technology suite. This approach is based upon the concept of AAR-by-comparison. That is, we build agents that represent expert human performance and then use them as benchmarks during execution of the tactical exercise, to which the

trainee performance is compared continuously and possibly in real time. By pairing each trainee with his own 'personal' expert agent, individual feedback can be managed for the benefit of the trainee.

The method used here to build the personal expert agents is a machine learning algorithm that builds the knowledge within the contexts, by observing human experts in action. It is called *Genetic Context Learning* or GenCL [3]. No matter how profoundly one might study a subject, it seems that actual experience is essential for perfecting a behavior. Experience increases the expertise level of a trainee beyond that provided by doctrines, manuals and regulations. If the agent could gain knowledge by observing experts with real experience performing the task to be taught to the trainees, the implicit knowledge might be effectively captured. We refer the reader to Fernlund et al. [3] for details on how to build these expert agents using GenCL. Nevertheless, our approach is designed to work regardless of how these expert agents are built, as long as they are built on a context-driven paradigm.

Today, there are many support systems for AAR in military exercises. Some of them record the actions of all actors during an exercise that could be re-played and viewed by the instructors and actors in an AAR session. These AAR aids are important to the individual participant to gain a more complete view of his actions during the exercise [2]. Extending such a support system for AAR with expert agents can then serve as the basis for a more detailed as well as personal evaluation as a result of AAR-by-comparison. If the expert agent receives the same inputs as its assigned trainee, its resulting action could be played in the simulated environment of the AAR support system and the discrepancies between the behavior of the trainee and of the expert agent could be identified, viewed and logged.

AAR by Comparison

Teaching guidelines and doctrines to military trainees has its drawback in that it is unrealistic to expose the trainee to all possible scenarios or actions that can happen in combat. The solution space is infinite in that sense. There is often no specific correct action to take for a given situation. More realistic would be to have models of the expertise at hand against which to compare the trainee's action.

Here we wish to establish a method whereby simulated expert agents experience the same situations, in a simulated environment, as does the human trainee in the simulated or real world military exercise. Drawing upon some of the basic tenets of Model-based Reasoning for equipment diagnosis, we can say that as

long as the actions of the trainee agree with those of the agent, the trainee is considered to be performing correctly. However, upon observation of a discrepancy from the benchmark expert agent, the discrepancy is noted and logged if the discrepancy is determined to be of enough importance. Such a system could be regarded as an evaluation support system. If the system juxtaposes the performance of the expert agent with the environmental data apparent to the agent, it will give the trainee an excellent platform for self-evaluation and learning.

Discrepancies between the Trainee and the Expert Agent

There can be different types of discrepancies in training exercises between the trainee and the expert agent, and with different severity. If the trainee and the agent for some reason chose different paths (physical or tactical) at a decision point, the discrepancy might become large. However, if neither encounters problems along the way, the discrepancy may be unimportant. Conversely, very small discrepancies in performance might have severe implications. The two entities could behave almost identically but, one might expose itself to the opponent's firing line of sight and be destroyed. Such a small discrepancy may have been the result of two completely different tactics applied to the same situation. It could be the difference between *seek cover* and *attack*. Hence, the investigation of any discrepancies between the expert agent and the trainee needs to be investigated with some intelligence. Therefore, we regard a discrepancy to be of two different types (not mutually exclusive): 1) the position, movement or firing action of the trainee is significantly different from the agent's; 2) the context of the human trainee is different from that of the agent. The first is rather easy to determine by merely overlaying the locations and actions of the trainee and of the expert agent. However, given the many possible moves and micro decisions, this type of discrepancy is likely to be only a very coarse filter that will result in many logged discrepancies. Many of these discrepancies will turn out to be of little tactical consequence.

The second type of discrepancy (Contextual Discrepancy) is more significant but more difficult to discover. First of all, the modeling paradigm of the expert agent must support contextual knowledge representation. To make a useful comparison, the SmartAAR system must also be able to infer the context in which the trainee is currently operating. Inferring a trainee's intentions and the set of skills being used at the time of the comparison can provide a very useful means of reviewing his performance. The problem, of course, is how to infer the context in which the human

is operating. One approach is to use a pattern matching technique that compares the trainee's action with that of the expert agent under various contexts simultaneously. The comparison that results in the closest match will indicate the context likely to be that of the trainee. This matching of patterns can be said to infer the context in which the trainee is operating. This is further described ahead.

It is our opinion that people in tactical situations behave in a context-based fashion. Several researchers in cognitive psychology promote models that are based on context-like structures, most notably Endsley [4] in her study of situational awareness, and Klein [5] in his recognition-primed decision making approach.

It is our assertion that the most important discrepancies between the expert agent and the trainee occur when they are in different contexts. While discrepancies in time and location may be common throughout an exercise, they may not always represent serious tactical misbehaviors. However, a discrepancy in the contexts of the expert agent and a trainee will nearly always be the result of inappropriate actions by the trainee and will also likely result in more future inappropriate actions. Hence, in order to facilitate this comparison, the modeling paradigm for the expert agent is context based.

When comparing the agent and the trainee, the expert agent executes in a simulated environment and acts upon the situation that the trainee encounters in the real world. Hence, the context model structure needs to be tailored for human behavior representation in simulated agents. This is fully in compliance with the way the expert agents are modeled according to the *Context-based Reasoning* (CxBR) behavior modeling paradigm. See Gonzalez and Ahlers 919980 for details on CxBR.

Summary and Conclusion of our Research

We propose an automatic self-evaluation approach called the SmartAAR technology suite that is applicable in military training, as well as in wide range of training and evaluation applications. The approach is able to detect both physical and contextual discrepancies between a trainee and an expert agent capable of acting as would an expert human performer. We refer to this approach as AAR-by-Comparison. Applying AAR-by-comparison could enhance the evaluation process and possibly be advantageous to more of the participants during an exercise. Giving each participant individualized feedback that focuses on their behavior by comparing it with an expert agent forms the basis for an automatic and self-instructing AAR. For training evaluation, the process of creating take-home packages or web portals can now be

automated. This would also ease conducting AAR in exercises with actors in different locations (live, virtual or mixed).

We assert that contextual discrepancies can be detected by comparison. This can be done by inferring the trainee's context and comparing it to the expert agent's context. If we can infer the context of a trainee, we can also say something about his intent. The comparison between the expert agent's active context and the context of the trainee then becomes a comparison of their intentions. The detection of contextual discrepancies is an important feature of the SmartAAR system. The prerequisite in doing such a comparison is the use of a modeling technique, such as CxBR, that models the context of the agent.

By detecting both contextual and physical discrepancies at the same time, SmartAAR provides full feedback to the trainee. Furthermore, by analyzing the contextual and physical discrepancies together, it could be possible to consolidate a number of discrepancies because it is likely that they correlate with each other. This correlation is left for future research, however.

We emphasize that the SmartAAR system is not a tutoring system that tells the trainee what to do or grades him in any way. It would be a risky to assume that the trainee and the expert agent at any moment would interpret the situation in the same manner (because of different inputs, view angles, assumptions, misinterpretations, stress, etc.). The SmartAAR system is a support system for self-evaluation that can help the trainee to make better evaluations of his behavior during the exercise. In this manner, the system fits, supports and enhances the way AAR is conducted today [2].

References:

- [1] "Training Circular 25-20, A leader's Guide to After-Action Reviews", Available at: http://www.au.af.mil/au/awc/awcgate/army/tc_25-20/table.htm, 1993
- [2] "Developing Adaptive Leaders", Pearson Custom Publishing, Pearson Merrill Prentice Hall, Boston, MA, Pp.112 – 123, 2005
- [3] Fernlund, H., Gonzalez, A. J., Georgiopoulos, M. and DeMara, R. F. (2006), "Learning Tactical Human Behavior Through Observation of Human Performance", IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, 36(1) pp.128–140
- [4] Endsley, M. (1995), "Towards a Theory of Situational Awareness in Dynamic Systems", Human Factors, 37(1), pp. 32-64
- [5] Klein, G. A. (1989), "Recognition Primed Decisions", in Advances in Man-Machine Research, W. Rouse (ed.), Greenwich, CT: JAI Press, pp 47-92

Authors:

Prof. Avelino Gonzalez
Dr. Hans Fernlund
Mr. Joachim Ekblad
School of Electrical Engineering & Computer Science
University of Central Florida
P.O. Box 162450
Orlando, FL U.S.A.
Phone: +407-823-5027
Fax: +407-823-5835
Email: gonzalez@mail.ucf.edu

R. Suzuki / N. Fujiki / Y. Taru / N. Kobayashi / E. P. Hofer

Internal Model Control for Assistive Devices in Rehabilitation Technology

INTRODUCTION

Control applications in rehabilitation technology are significant for the development for interaction between humans and machines. Assistive devices in rehabilitation technology are widely used by elderly or disabled people in order to support more life independence. For the interaction with humans, safety and reliability conditions have to be fulfilled and control design methods for appropriate assistive devices have to be robust. Moreover, in conventional situations, a lot of sensors are attached to assistive devices to detect ability of users or to estimate unknown disturbance added to systems. However, this complicates the mechanical structure and the controller scheme for such assistive devices.

To overcome these problems a control method based on the generalized internal model control (GIMC) is applied to assistive devices for rehabilitation, e.g. wheelchairs, assistive devices for standing-up.

Generalized Internal Model Control (GIMC) Scheme

The IMC design is widely used in process control and mechanical systems control. Nevertheless the procedure of controller synthesis is very simple, some satisfactory properties, e.g., a disturbance rejection and a trajectory tracking were shown. In the work of Suzuki (2002), the disturbance estimation property of the IMC structure based on LQ optimal control is also presented. We extend this structure to a general controller scheme.

The GIMC scheme has been firstly proposed by Zhou and Ren (2001). We discuss a new GIMC scheme from another point of views. Consider the linear time invariant system with disturbance as follows:

$$\begin{aligned} \dot{x} &= Ax + Bu + D\xi \\ y &= Cx \end{aligned} \tag{1}$$

where $x \in R^n$ is the state vector, $u \in R^m$ is the input vector, $y \in R^m$ is the output vector and $\xi \in R^q$ is the disturbance vector. For this system we propose the GIMC scheme shown in Fig. 1. The proposed method is based on a two-degree-of-freedom servo system using an observer-based stabilizing controller. Let us consider the following observer Σ_{ob}

$$\begin{aligned} \dot{\hat{x}} &= (A + L_\rho C)\hat{x} - L_\rho y + Bu \\ L_\rho &= \lim_{\rho \rightarrow 0} -P_\rho C^T \\ AP_\rho + P_\rho A^T - P_\rho C^T C P_\rho + DD^T \frac{1}{\rho^2} &= 0. \end{aligned} \quad (2)$$

Suppose the system (1) satisfies $B = D$, because input channels of mechanical systems are mostly affected by disturbance, and F is a feedback gain obtained by limiting feedback gain. $Q(s)$ is a stabilizing controller parameterized an approximate inverse system of Σ_{ob} . The parameters \hat{L} , H and G are obtained by (3) where S is a Hurwitz matrix. The associated relations are

$$\begin{aligned} F &= F_\varepsilon = \lim_{\varepsilon \rightarrow 0} -B^T M_\varepsilon, \quad A^T M_\varepsilon + M_\varepsilon A - M_\varepsilon B B^T M_\varepsilon + \frac{1}{\varepsilon^2} C^T C = 0 \\ \hat{L} &= C(A + B F_\varepsilon)^{-1} \\ H &= -(\hat{L} B)^{-1} \\ G &= -H S = (\hat{L} B)^{-1} S \end{aligned} \quad (3)$$

The GIMC structure shown in Fig. 1 contains several controller schemes.

- (1) The structure becomes an all stabilizing controller (Suzuki, et. al. 2001) if $G = 0$ holds. The all stabilizing controller has same disturbance decoupling property and trajectory tracking property as the IMC design structure by choosing an appropriate $Q(s)$.
- (2) From another point of view, this structure becomes an observer based servo controller with the 2DOF (Nakamoto 2003) if $Q(s) = 0$ holds.

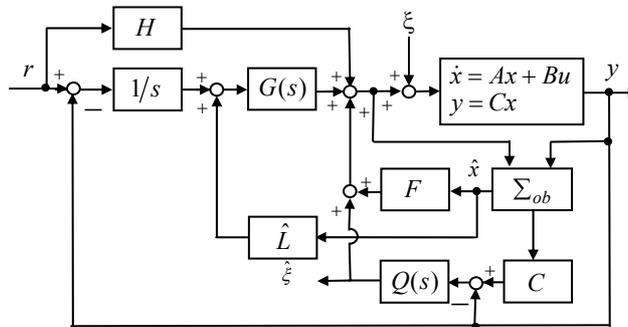


Figure 1 – GIMC structure with servo compensation

The error equation is obtained by

$$\Sigma_e(s) : \begin{cases} \dot{e} = (A + L_\rho C)e - B\xi \\ \mu = Ce = y_m - y \end{cases} \quad (4)$$

and $Q(s) = \Sigma_{e\tau}^{-1}(s)$ is calculated by

$$\Sigma_{e\tau}^{-1}(s)\Sigma_e(s) = \text{diag}((\tau s + 1)^{-d_1}, \dots, (\tau s + 1)^{-d_m}). \quad (5)$$

Let (A_Q, B_Q, C_Q, D_Q) be a minimal realization of the approximate inverse system of $\Sigma_{e\tau}^{-1}(s)$; τ is very small value and d_i is the integral index to obtain a proper approximate inverse system.

The augmented system is given by the following state equation

$$\dot{x}_e = \begin{pmatrix} A + BF_\varepsilon + BGL & BG & B(F_\varepsilon + GL + D_Q C) & BC_Q \\ -C & 0 & 0 & 0 \\ 0 & 0 & A + L_\rho C & 0 \\ 0 & 0 & B_Q C & A_Q \end{pmatrix} x_e + \begin{pmatrix} BH \\ I \\ 0 \\ 0 \end{pmatrix} r + \begin{pmatrix} B \\ 0 \\ -B \\ 0 \end{pmatrix} \xi \quad (6)$$

$$y = (C \ 0 \ 0 \ 0) x_e$$

where $x_e^T = (x^T \ w^T \ e^T \ \eta^T)^T$. By using the transformation matrix

$$T = \begin{pmatrix} I & 0 & 0 \\ \hat{L} & I & 0 \\ 0 & 0 & I \end{pmatrix} \quad (7)$$

the augmented system (6) is represented as follows

$$\dot{\tilde{x}}_e = \begin{pmatrix} A + BF_\varepsilon & BG & B(F_\varepsilon + GL + D_Q C) & BC_Q \\ \hat{L}(A + BF_\varepsilon) - C & \hat{L}BG & \hat{L}B(F_\varepsilon + GL + D_Q C) & \hat{L}BC_Q \\ 0 & 0 & A + L_\rho C & 0 \\ 0 & 0 & B_Q C & A_Q \end{pmatrix} \tilde{x}_e + \begin{pmatrix} BH \\ \hat{L}BH + I \\ 0 \\ 0 \end{pmatrix} r + \begin{pmatrix} B \\ \hat{L}B \\ -B \\ 0 \end{pmatrix} \xi \quad (8)$$

where $\tilde{x}_e = Tx_e$. For the augmented system (8), we obtain the following properties. The signals r and ξ are not limited to step input or step disturbance.

[Property 1]

Consider the augmented system (8), and suppose the system (A, B, C) has no unstable zeros. The closed loop system can immediately arrive at disturbance decoupled system by multiplicative effect of the disturbance rejection property of the IMC design structure and the disturbance rejection property of LQ control. That is, the following statement holds:

$$\lim_{\varepsilon \rightarrow 0, \tau \rightarrow 0, \rho \rightarrow 0} H_{y\xi}(s) \rightarrow 0 \quad (9)$$

where $H_{y\xi}$ is the transfer function from disturbance to output described as

$$H_{y\xi}(s) = C(sI - A - BF_\varepsilon)^{-1}B \{I + G(sI - \hat{L}BG)^{-1}\hat{L}B\} \{I - (D_Q + C_Q(sI - A_Q)^{-1}B_Q) C(sI - A - L_\rho C)^{-1}B\} \\ + C(sI - A - BF_\varepsilon)^{-1}B \{I + G(sI - \hat{L}BG)^{-1}\hat{L}B\} G\hat{L}(sI - A - L_\rho C)^{-1}B.$$

[Property 2]

Consider the augmented system (8), and suppose the system (A, B, C) has no unstable zeros. For the scheme shown in Fig. 1, if we parameterize $Q(s)$ as an approximate inverse system of $-\{ C(sI - A - L_\rho C)^{-1} B \}$, then the output signal of $Q(s)$ can be estimated unknown disturbance ξ as $\tau \rightarrow 0$. For the transfer function from disturbance to the estimated disturbance holds

$$\lim_{\tau \rightarrow 0} H_{\xi \hat{\xi}}(s) \rightarrow I, \quad (10)$$
$$H_{\xi \hat{\xi}}(s) = \{ D_Q + C_Q(sI - A_Q)^{-1} B_Q \} C(sI - A - L_\rho C)^{-1} (-B).$$

The approximate inverse system is obtained by (5). Moreover, the closed loop system maintains servo property.

Experimental Results

Firstly, the proposed controller scheme based on the GMC is applied to the assistive device for standing-up. By comparative experiments it is shown that the controller scheme can detect forces of the disabled person without additional sensors (see Fig. 2). By comparative experiments, we show that the disturbance estimation property can be applied to measure user's ability to stand up. To measure such kind of ability, we need additional sensors, e.g., force sensor, pressure sensor, or electromyography (EMG). However, by using the proposed controller scheme, we can estimate the user's ability to stand up.

Here dependency of a user to the assistive device is added to the disturbance channels. We design the controller scheme by tuning ε, ρ, τ . Figure 2 shows the experimental results of estimating human ability for standing up motion. The bottom solid line shows the result of standing with high dependence, and the middle solid line shows the result with low dependence. From this figure the user and helper can easily know ability for standing without additional sensors. Moreover, we can know whether users depend on the assistive device for standing or not. Note that the estimated value includes the data of human ability, weight of the experiment, influence of friction, and so on. However, we can extract the user's data easily by comparing the experimental data.

From the experimental result, we can see the effectiveness of disturbance estimation property of the proposed method. By using the proposed scheme we estimate

whether a user depends on the assistive devices. In addition, we can obtain instructing data for rehabilitation or training. The application is able to extend another assistive device which needs estimating ability for elderly or disabled persons.

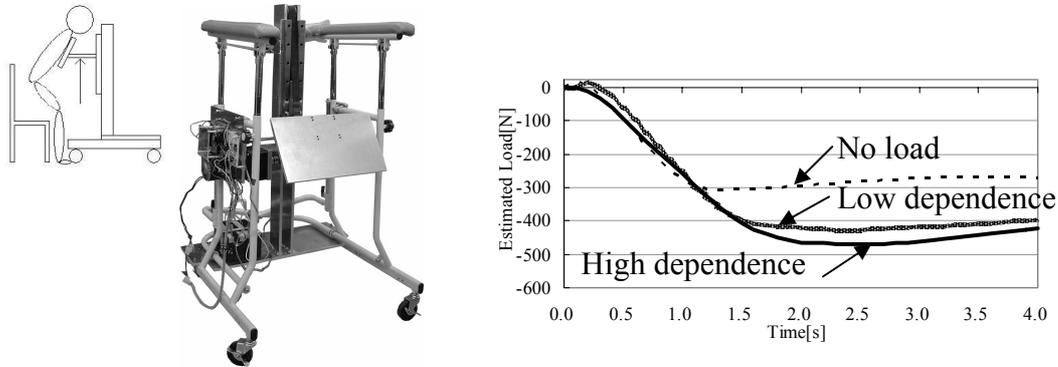


Fig. 2: Assistive device for standing and experimental result for detecting force

Secondly, the proposed scheme is applied to the single side driven wheelchairs. By using the proposed scheme the wheelchair is able to traverse on slant slopes without an inclinometer (see Fig. 3). As for wheelchairs, when it crosses on slant slopes, it is not possible to advance straight. Because center of gravity of wheelchairs changes on slopes, and then force given to two wheels would be changed. By using the estimation property of the GIMC, the controller scheme estimates this change as disturbance. As the result, the single side driven wheelchairs can traverse on slant slopes.

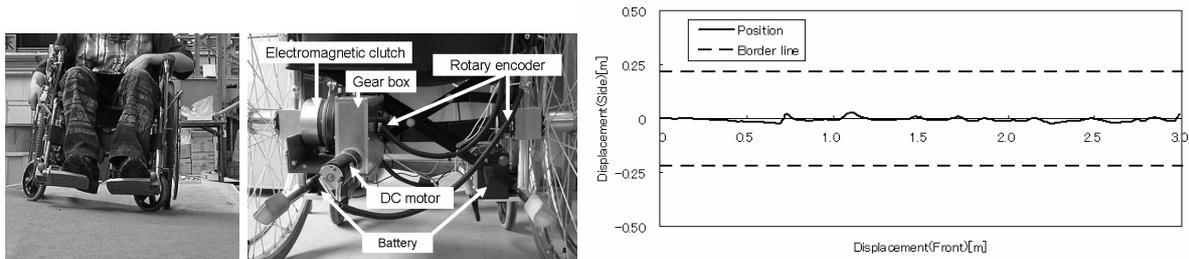


Fig. 3: Single side driven wheelchair and experimental result on slant slope

Conclusions

This paper showed a controller scheme for force detection based on the internal model control method in rehabilitation technology. The proposed method was applied for detecting force generated by humans. The proposed controller scheme based on the internal model control has the required robustness, especially, disturbance

estimation for parameter perturbations. The experimental results for an assistive device were also shown. By comparative experiments it was shown that the proposed controller scheme could detect forces of the disabled person without additional sensors. In addition, we can obtain instructing data as “bio-feedback” for rehabilitation or training. The controller scheme is able to extend another assistive device which needs estimating ability for humans.

The proposed research specifically emphasizes that from theoretical and experimental findings the disturbance estimation property of the proposed GIMC is useful for detecting unknown disturbance without additional sensors, e.g. inclinometers, force sensors, pressure sensors, or electromyography (EMG). The results are a further step in rehabilitation technology to contribute towards the development of equipment interacting with humans.

References:

- [1] M. Nakamoto, Two Degree of Freedom Servo Controller Incorporated with Reference and Disturbance Feedforward and an Application to a Steam Pressure and Flow Control Systems (in Japanese), Trans. on the Institute of Systems Control and Information Engineers, Vol. 16, No. 3, pp. 111-117, 2003.
- [2] R. Suzuki, T. Miyazaki and N. Kobayashi, Decoupling Control via All Stabilizing Controller with Limiting Properties of LQ Control and Its Application to A Mechanical System, Proceedings of IEEE International Conference on Control Applications, pp. 412-417, 2001.
- [3] R. Suzuki, M. Tani and N. Kobayashi, Design and Development of Single Side Driven Wheelchairs by Using Internal Model Control, Proceedings of IEEE International Conference on Control Applications, pp. 355-360, 2002.
- [4] R. Suzuki, et al., Disturbance Estimation by Generalized Internal Model Control and Its Application to Assistive Devices for Rehabilitation Technology, Proceedings of IEEE International Conference on Control Applications, pp. 921-926, 2006.
- [5] M. Tani, et al., Internal Model Control for Assisting Unit of Wheeled Walking Frames, Proceedings of IEEE International Conference on Control Applications, pp. 928-933, 2004.
- [6] K. Zhou and Z. Ren, A New Control Architecture for High Performance, Robust and Fault-Tolerant Control, IEEE Transactions on Automatic Control, Vol. 46, No. 10, pp. 1316-1618, 2001.

Authors:

Prof. Dr. Ryoichi Suzuki
Department of Robotics,
Kanazawa Institute of Technology,
7-1 Nonoichi, Ishikawa 921-8501, JAPAN
Phone: +81-76-248-9761
Fax: +81-76-294-6704
E-mail: r-suzuki@neptune.kanazawa-it.ac.jp

Dr. N. Fujiki, Y. Taru and Prof. Dr. N. Kobayashi
Department of Robotics,
Kanazawa Institute of Technology,
7-1 Nonoichi, Ishikawa 921-8501, JAPAN

Prof. Dr. Eberhard P. Hofer
Institute of Measurement, Control and Microtechnology,
University of Ulm,
Albert-Einstein-Allee 41, D-89081 Ulm Germany
Phone: +49 731 50-26300
Fax: +49 731 50-26301
E-mail: ep.hofer@uni-ulm.de

D. Sommer / M. Golz

Feature Reduction for Microsleep Detection

Abstract. In stages of extreme fatigue, e.g. while car driving, dangerous microsleep events may occur. Their detection in spontaneous biosignals still poses a challenge. For this purpose it is necessary to analyze signals with high temporal dynamics, like the brain and eye electric activity. From both types of measurements the Power Spectral Densities were estimated and were subsequently reduced by five different methods. Their performances were evaluated empirically in order to get lowest errors for the detection of microsleep events. The detection was realized by Computational Intelligence methods. It turned out that feature reduction performs best when averaging in many small spectral bands or in flexible spectral bands was utilized. Their free parameters were optimized by genetic algorithms. The Support Vector Machine with RBF kernel function established as best performing classification tool.

INTRODUCTION

One of the most important human factors causing accidents is operator fatigue and loss of attention. Their portion among all accidents is estimated as 15 to 20 % and is exceeding in this respect the importance of alcohol and drugs [1]. In stages of extreme fatigue dangerous microsleep events (MSE) may occur. MSE are defined as short intrusions of sleep under the demand of sustained attentiveness. Their detection in spontaneous biosignals still poses a challenge.

In contradiction to the evaluation of fatigue where a wide range of different signals are available, it is difficult to get suitable signals which immediately reflect ongoing MSE on a second by second basis. Mostly, measurements of brain electric and of eye movement activity are used, which are featured by high temporal resolution. Disadvantageously, they are non-contactless and are corrupted by large noise which is originated by other simultaneously ongoing processes. This leads to more or less extensive signal processing and pattern recognition.

It has been shown that the empirical error of recognition sensitively depends on several parameters of the preprocessing, the feature extraction and the classification stages [2]. Considerable improvements can be gained if different signal sources are fused on the feature level [3]. Furthermore, it turned out that feature extraction in the spectral domain is most successful compared to methods in the state space or time domain [3]. Therefore, we here report only of feature extraction and subsequent reduction in the spectral domain.

The common way of feature reduction during EEG analysis is band averaging. The Power Spectral Densities (PSD) are averaged in four or more frequency bands. Their definition is not fixed and varies to some extent between different authors. Typical values are 0.5 – 4.0 Hz (delta band), 4.0 – 8.0 Hz (theta band), 8.0 – 12.0 Hz (alpha band) and 12.0 – 30 Hz (beta band) [4]. The question arises if this or other choices are optimal for quantitative EEG analysis, especially when modern Computational Intelligence methods are applied. They do not suffer from the so-called 'curse of dimensionality' [5]. In this line, we have established five different cases:

- (1) No reduction; all available frequency bins up to the Nyquist frequency are included,
- (2) Feature reduction utilizing Principal Component Analysis (PCA),
- (3) Feature reduction by averaging in fixed band [4],
- (4) Feature reduction by averaging in equidistant bands,
- (5) Feature reduction by averaging in arbitrary bands utilizing genetic algorithms.

The outcome of all these methods were applied as input values to Computational Intelligence methods in order to establish a detector for microsleep [6]. The detection accuracy, estimated by the classification error of all feature vectors of the test set, is a useful empirical measure to compare all five cases. For this purpose the widely accepted multiple hold-out validation was used.

EXPERIMENTS

Experiments were conducted in our real car driving simulation lab (Figure 1). 23 young adults took part and each of them finished 7 driving sessions which are repeated every hour between 1 a.m. and 8 a.m. One driving session has a length of 35 minutes. It turned out that the likelihood of occurrence of MSE was gradually increasing due to monotony and fatigue. Fatigue has initially been induced by at least 16 hours without sleep prior to the experiment.

During driving the brain electric activity reflected by the EEG was recorded from seven different locations on the scalp. The eye and eyelid movements reflected by the Electro-oculogram (EOG) were recorded from two locations (vertical, horizontal): Both were sampled at a rate of 128 sec^{-1} .

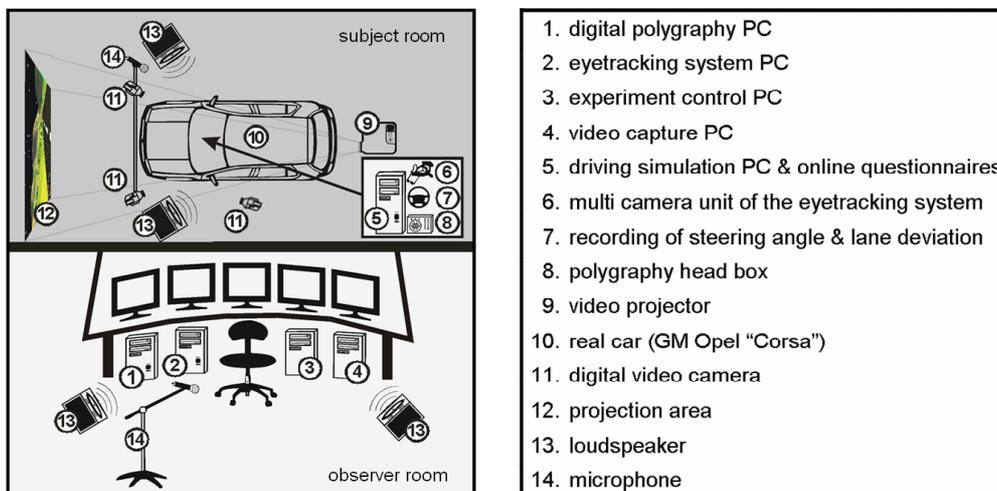


Figure 1: Our laboratory which is specialized for recording of overnight-driving simulations. A real small city car in conjunction with a 3D driving simulation software is utilized to give a monotonic lane tracking task to the subjects. Subject behaviour is video recorded. Also the driving performance and the biosignals of the subject are stored.

MSE are typically characterized by driving errors, prolonged eye lid closures or nodding-off. Towards automatic detection, two experts performed the initial MSE scoring, whereby three video cameras were utilized to record i) drivers portrait, ii) right eye region and iii) driving scene. For further processing, only clear-cut cases, where all the experts agreed on the MSE, were taken into account. Despite providing enough test data to tune our algorithms, the human experts could not detect some of the typical attention lapses, such as the one with open eyes and stare gaze. The number of MSE varied amongst the subjects and was increasing with time of day for all subjects. In all 3,573 MSE (per

subject: mean number 162 ± 91 , range 11 – 399) and 6,409 non-MSE (per subject: mean number 291 ± 89 , range 45 – 442) were scored. Non-MSEs are periods between MSE where the subject is drowsy but shows no clear or unclear MSE. This clearly highlights the need for an automated MSE detection system, which would not only detect the MSE also recognized by human experts, but would also offer a possibility to detect the critical MSE cases which are not recognizable by human experts.

METHODS

Segments of all 9 electrophysiological signals (seven EEG and two EOG channels) were extracted with respect to the observed temporal starting points of MSE / Non-MSE using two free parameters, the segment length and the offset between first sample of segment and starting point of an event. The trade-off between temporal and spectral resolution is adjusted by the segment length and the location of the region of interest on the time axis is controlled by the temporal offset. Both parameters are of high importance and are to be optimized [2]. The variation of the segment offset resulted in a relatively steep error function. An optimal offset value was found to be around -3 sec. In the same way an optimal segment length of 8 sec was found [2]. This means that classification is working best when 3 sec of EEG / EOG immediately before a MSE and 5 sec during ongoing MSE are processed.

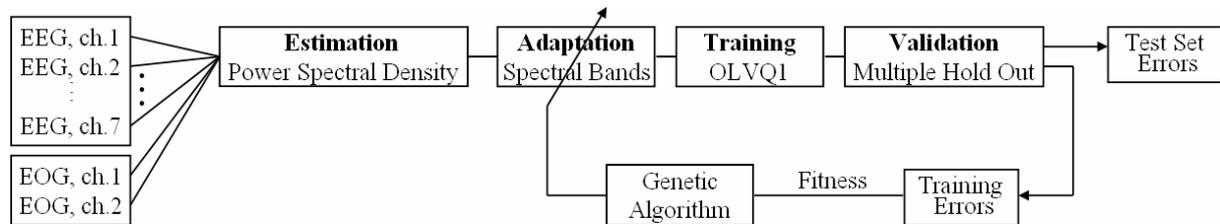


Figure 2: GA-OLVQ1 framework for empirical optimization of band averaging.

After linear trend removal the PSD values were estimated by the modified Periodogram method utilizing Hann windows. Subsequently, the PSD values were scaled logarithmically which has been shown to be important for error diminishing [3]. Now the above mentioned five cases of feature reduction were applied. Each feature vector consisted of 4617 components resulting of sampling in 9 channels at 128 sec^{-1} over a period of 8 sec (case 1). In order to reduce this large amount of components the PCA was utilized (case 2). As a criterion which subspace is optimal the eigenvalues of the covariance matrix were used. Case (3) is the well-known averaging of PSD values in fixed spectral bands. Their cut-off frequencies are mentioned above. This case came out by an extreme reduction from 4617 to 36 (4×9) components. In case (4) three free parameters of an averaging in equidistant bands were optimized: lower (f_L) and upper (f_U) cut-off frequency and the width (Δf) of each band. Optimal values were found to be $f_L = 0.5 \text{ Hz}$, $f_U = 23.0 \text{ Hz}$, and $\Delta f = 1.0 \text{ Hz}$, respectively. In the most flexible case (5) averaging was performed in spectral bands of arbitrary location and width. The training errors of Optimized Learning Vector Quantization (OLVQ1) were used as fitness function of genetic algorithms (GA) (Figure 2) [3]. Genetic representation was fixed to 10 spectral bands, each defined by lower and upper cut-off frequency for each EEG and EOG channel. This resulted in 180 real values ($10 \times 2 \times 9$), which were optimized by evolutionary strategy with Gaussian mutation and an averaging crossover method. Usually GA-OLVQ1 optimizations were finished after computation of 300 generations of 256

individuals per population. The variability of the resulting optimal gene expressions was estimated by repeated (100 times) application of all methods to the data set. All mentioned optimizations were done empirically. We employed mainly Optimized Learning Vector Quantization (OLVQ1) [7] as a robust, very adaptive and rapidly converging classification method. OLVQ1 has at least one further free parameter to be optimized, the number of prototype vectors. During parameter optimization the minimal test error was searched following the cross-validation paradigm of “multiple-hold-out”. Only when utilizing Support Vector Machines (SVM) the paradigm of “leave-one-out” was applied, which is an almost unbiased estimator of the true classification error [8]. In fact, this method is computationally much more expensive than “multiple-hold-out”, but only for the SVM classifier an efficient implementation exists [8]. The SVM applied on the given problem uses the RBF kernel function, because it matches best [6].

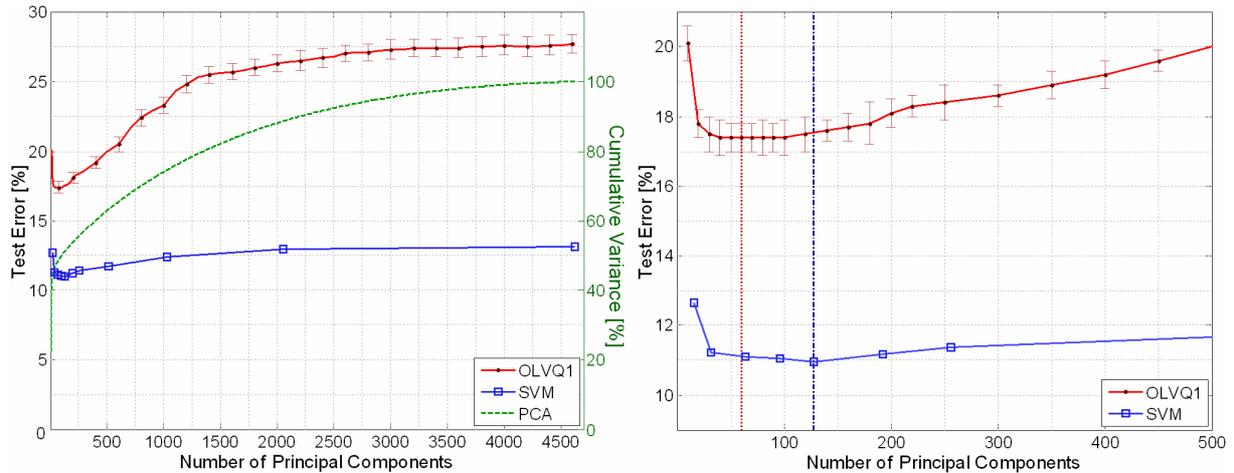


Figure 3: Test set errors of OLVQ1 (dots) and SVM (squares) and the cumulative variance (dotted) versus the number of ranked principal components. Minimal errors were achieved by processing the first 60 and first 128 principal components by OLVQ1 and SVM, respectively (see enlarged view, right). For each sample of the SVM plot separate hyper-parameter optimizations were performed.

Case	Number of Features		OLVQ1		SVM	
			E_{TRAIN} [%]	E_{TEST} [%]	E_{TRAIN} [%]	E_{TEST} [%]
(1) No feature reduction	4617	(513×9)	22.6 ±0.5	27.7 ±0.6	0.0 ±0	13.1 ±0
(2) PCA	60	128	10.4 ±0.2	17.4 ±0.4	1.4 ±0	10.9 ±0
(3) fixed band	36	(4×9)	11.6 ±0.2	17.5 ±0.4	4.9 ±0	13.2 ±0
(4) equidistant bands	216	(24×9)	9.3 ±0.1	15.7 ±0.4	0.1 ±0	9.9 ±0
(5) GA-OLVQ1 optimized	90	(10×9)	8.2 ±0.1	14.1 ±0.4	0.1 ±0	9.8 ±0

Table 1: Mean and standard deviation of training and test errors for 5 different cases of feature reduction. Errors were estimated by Multiple Hold-Out and by Leave-One-Out cross validation utilizing OLVQ1 and SVM, respectively.

RESULTS

If all 4617 PSD values are processed, then the classification methods have to adapt a relatively high dimensional decision function. OLVQ1 achieved test set error of 27.7 % in the mean, whereas SVM achieved 13.1 % (Figure 3, Table 1). These results can be outperformed by feature reduction. The Principal Component Analysis (case 2) [9] resulted with minimal errors when not more than the first 3 % of principal components

were utilized for classification (Figure 3). OLVQ1 came out by a tremendous decrease down to 17.4 % whereby only the first 60 principal components were needed. In contrast, SVM utilized twice more components (128) to achieve minimal test set errors down to 10.9 % (Figure 3).

Case (3) uses the common way of feature reduction in EEG analysis, namely the averaging of PSD values in fixed spectral bands [4]. Mostly, a relatively coarse split in four bands is applied, as explained above. This leads to 36 features (4 bands x 9 signals). OLVQ1 resulted in lower errors than in case (1) and in nearly same errors than in case (2). SVM came out with no improvements (Table 1). This is not the case with many small spectral bands (case 4). Empirical optimizations of the three free parameters resulted in $f_L = 0.5$ Hz, $f_U = 23.0$ Hz and $\Delta f = 1.0$ Hz, i.e. 24 band-averaged PSD values for each of the nine signals. Case (5) aimed at improving errors with a fixed number of spectral bands whereby the parameters for each band are optimized empirically. This method is most flexible and resulted best albeit only slightly better than case (4).

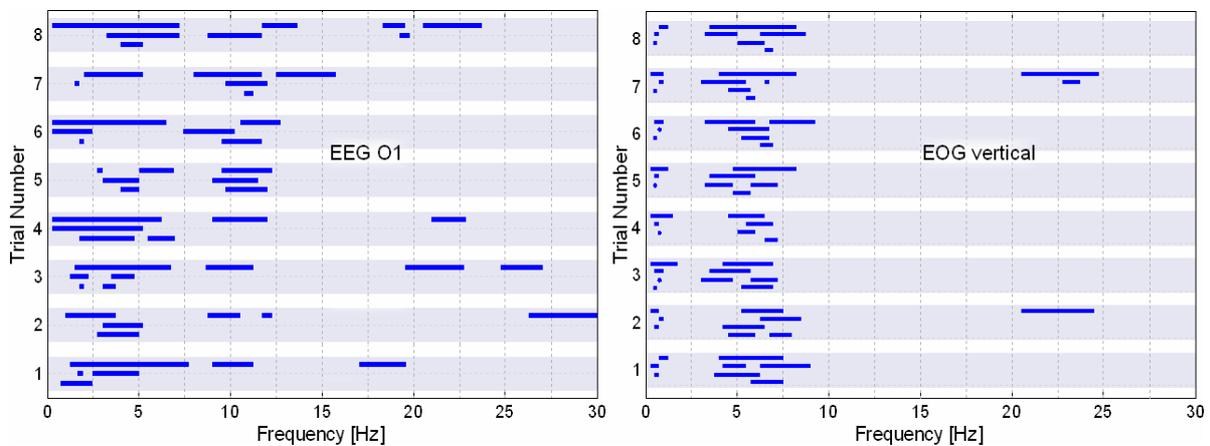


Figure 4: Eight randomly selected genetic representations (out of 100) found for EEG-O1 (left) and vertical EOG (right). Each bar represents both parameters (location and width) of the spectral bands. The plots are limited to 30 Hz while optimizations were performed up to 64 Hz.

The empirical optimizations of case (5) were performed individually for each spectral band, so that overlapping bands and different frequency intervals were found for each signal. These optimizations were repeated 100 times with random initializations of the genetic algorithm. It came out that the band parameters (location and width) varied slightly from run to run. Figure 4 shows 8 randomly selected genetic representations of spectral bands found for two signals (EEG-O1, vertical EOG). The other free parameters of the genetic algorithm, e.g. number of bands, population size and number of generations, were optimized empirically in prior steps.

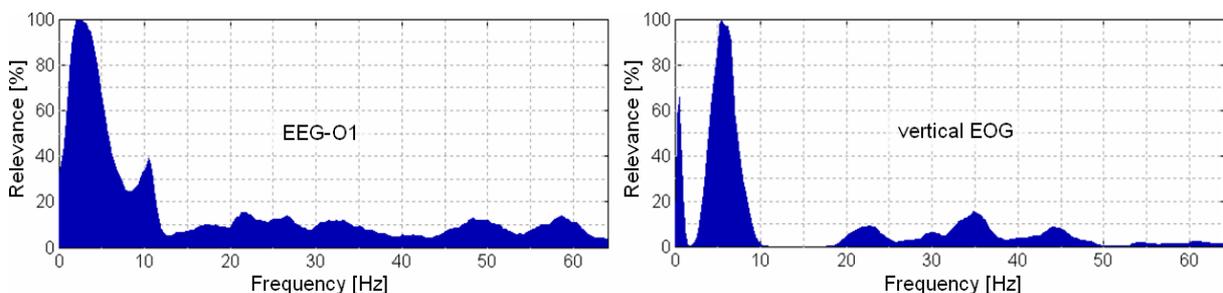


Figure 5: Relevance of frequency selection due to spectral band optimizations for the EEG-O1 (left) and the vertical EOG (right) signal. Results are the outcome of 100 GA-OLVQ1 runs.

In the EEG-O1 signal different bands resulted (Figure 4, left). Overlapping bands occurred mostly in the delta (0.5 – 4 Hz), theta (4 – 8 Hz) and alpha region (8 – 12 Hz), but not in the beta (12 – 30 Hz) and gamma region (30 – 64 Hz). As expected, the outcome for the vertical EOG was different (Figure 4, right). Optimizations resulted in higher probabilities to select a band at very low frequencies (up to 2 Hz) and at frequencies between 3 and 8 Hz, but scarcely at higher frequencies. The outcome of all calculations was summed and normalized (Figure 5) and is an estimate of how relevant are single frequencies to perform an optimal microsleep detection. This way, GA-OLVQ1 provides a way to extract knowledge from subsymbolic machine learning.

CONCLUSIONS

We have presented five different cases of feature reduction for the analysis of Microsleep events. Best results, with test errors down to 10 %, were obtained by dynamic adaptation of spectral bands for each signal utilizing genetic algorithms. Such an adaptation of frequency bands enables three facts. First, it reduces the number of features as well as the complexity of the problem which resulted in increased detection performance. Second, it lowers computational costs, and third, this methodology extracts knowledge on relevance or irrelevance of single PSD values in order to detect microsleep events.

Results showed common averaging of PSD in four bands (delta, theta, alpha, beta) to be outperformed by averaging in many spectral bands. SVM are less sensitive to high number of input features than OLVQ1. In the future OLVQ1 should be replaced by SVM in order to estimate training set errors. In the presented framework these estimates serve as fitness function of a genetic algorithm. In parallel, multiple hold-out validation should be replaced by leave-one-out validation which is an almost unbiased estimator of the true classification error. Furthermore, it should be considered to set the number of frequency bands per signal completely free. But, the basis for all future improvements is the expected rapid increase of computational power.

References:

- [1] T. Akerstedt (2000). Consensus Statement: Fatigue and Accidents in Transport Operations. *J Sleep Research* 9, 395
- [2] M. Golz, D. Sommer (2004). Automatic Recognition of Microsleep Events; *Biomedizinische Technik* 49, Suppl.2, 1, 332-333
- [3] D. Sommer, et al. (2005). Fusion of State Space and Frequency-Domain Features for Improved Microsleep Detection. *Int Conf Artificial Neural Networks (ICANN 2005)*; LNCS3697, 753-759, Springer, Berlin.
- [4] E. Niedermeyer, F. Lopes Da Silva (1999). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. 4th edition, Lippincott Williams & Wilkins, Baltimore.
- [5] Y. Bengio, O. Delalleau, N. Le Roux (2005). The Curse of Dimensionality for Local Kernel Machines. *Techn. Rep. 1258*, Université de Montréal; Canada.
- [6] M. Golz, et al. (2007). Feature Fusion for the Detection of Microsleep Events. *J VLSI Signal Proc Syst*, in press
- [7] Kohonen, T.; *Learning Vector Quantization*; in M. A. Arbib (eds.); *The Handbook of Brain Theory and Neural Networks*; MIT Press: Cambridge, MA, 1995, 537-540
- [8] Joachims, T.; *Learning to Classify Text Using Support Vector Machines*; Dissertation; Kluwer; Boston, 2002
- [9] B. Heisele, T. Serre, S. Mukherjee, and T. Poggio. Feature reduction and hierarchy of classifiers for fast object detection in video images. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii, 2001

Authors:

Dipl.-Inform.(FH) David Sommer
Prof. Dr. rer. nat. Martin Golz

Fachhochschule Schmalkalden – University of Applied Sciences
Am Schwimmbad, PF 10 04 52
98564 Schmalkalden

Phone: ++49 3683 688 4207
Fax: ++49 3683 688 4499
E-mail: d.sommer@fh-sm.de

F. Müller / A. Wenzel / J. Wernstedt

A new strategy for on-line Monitoring and Competence Assignment to Driver and Vehicle

ABSTRACT

This paper presents a new strategy for driver monitoring and competence assignment based mainly on standard sensory equipment. Within this a competence adapter assigns a competence level to each of both subsystems, driver and vehicle. In this manner short- and long-term reduced capabilities of the driver in regard to his driving task can be detected and may even be compensated under some circumstances. In this concept the driver still maintains his responsibility as long as he is conscious.

1. INTRODUCTION

Driving cars does not only mean mobility for many people but also is connected with freedom and individuality by them. In the next decades many western societies will undergo dramatic changes due to the demographic alteration. In addition to the fact, that drivers of vehicles will get older in the mean, there will be the need for them to have access to individual mobility even in their later phase of life. That's because of the turning away of traditional multi – generation households and increasing central supply facilities especially in urban areas. Driver assistance systems can help to safe the driving of older people. These often have face reduced capabilities notably in the perception of their environment and reaction time. Furthermore they are confronted with a higher risk for shock- or blackout syndroms resulting from specific diseases, like diabetes, represented more often at the group of elderly. The authors believe that the increasing possibilities of driver assistance systems are not only able to support and safe people in their late phase of life, but also can allow people with other limits in driving capabilities a form of individual mobility, which is safe for them and their environment in a most possible way. An integral part of this strategy is however the necessity for the continuous monitoring of the driver and vehicle as well as their competence assignment, which the rest of this paper deals with.

2. CYBERNETIC SYSTEM DRIVER – VEHICLE

Since the goal has to be that the vehicle and the driver do fulfill their driving task, the system driver – vehicle should be treated as one system and regarded as one. The cybernetic decision structure for the guidance of mobile systems/vehicles is based on the 3-level hierarchical structure illustrated in Figure 1, with level 1 as the autopilot layer for time-critical and safety-relevant control tasks as well as routine tasks, level 2 as the maneuver layer i.e. for track switching or overtaking and the Level 3 as mission layer for the planning, control and re-planning of a mission (route, course). The task distribution between the human and the vehicle within the levels is situation and task dependent and has changed drastically in the last decades.

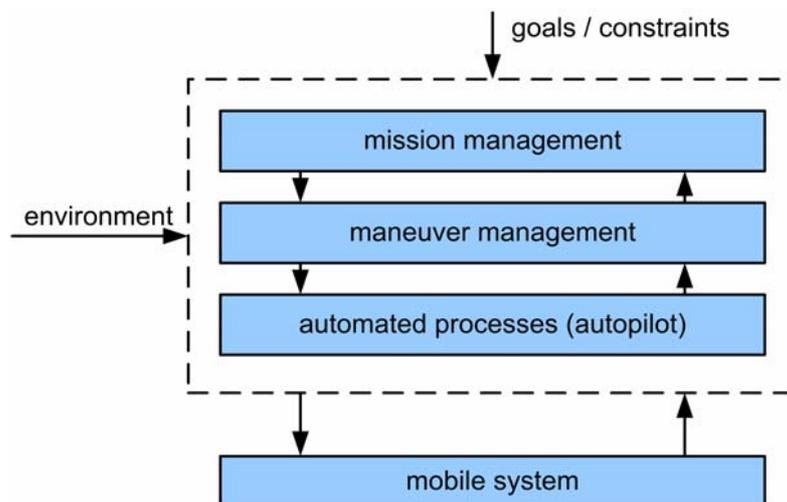


Figure 1: Decision structure for mobile systems

It is assumed that the development of automation and assistance systems will continue, but humans must and will further have the responsibility in the levels of mission and maneuver management.

3. SYSTEM CONCEPT

The concept of the proposed system is depicted in figure 2. The task of the competence adapter system (CAP) is to evaluate the monitoring signals from driver and vehicle, estimate the driver's condition as well as assist and safe the driver by appropriate measures. The system is based on the system description of the driver and the vehicle. Within this, the driver and the vehicle are modeled by their weighting functions $g_D(t)$ and $g_v(t)$ according to their laplacian transfer functions. The driver acts on the vehicle by the control vector $u_D(t)$, which consists mainly of accelerating,

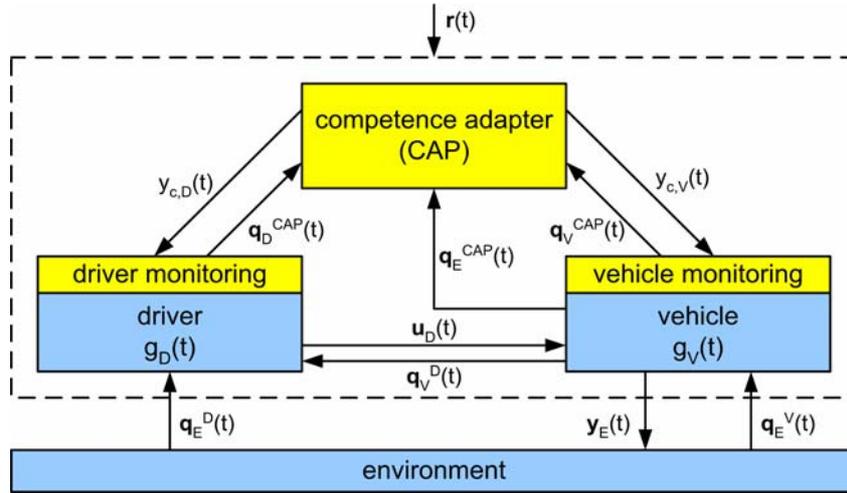


Figure 2: Overall system design

decelerating and steering the car, whereby the driver's inputs are the state vector of the environment $q_E^D(t)$ and the state vector of the vehicle $q_V^D(t)$ as seen by him. The changes in position and velocity as well as other output parameters of the vehicle with respect to itself and the environment are described by the vector $y_E(t)$. In addition to this basic structure, components of driver and vehicle monitoring as well as the competence adapter system are added to the system, which in result builds up the overall structure. In this regard vehicle monitoring is assumed to be relatively common in upcoming car generations and will provide a detailed survey about the vehicle's state [1]. The driver monitoring module and the CAP system are described more detailed in the subsequent parts of this paper. The vehicle monitoring system will then provide the state vector of the vehicle $q_V^{CAP}(t)$, while the driver monitoring system will provide the driver's state vector $q_D^{CAP}(t)$, both then as seen by the CAP system. In addition the vehicle sensory equipment generates the environment state vector $q_E^{CAP}(t)$ again as seen by the CAP system. The system's output are the parameters of competence assignment to driver and vehicle $y_{c,D}(t)$ and $y_{c,V}(t)$, as described later. The overall goal function is represented by $r(t)$. Note, that due to different perception the state vectors have to be distinguished according to the regarded system module. I.e. the state vectors of the environment as seen by the driver $q_E^D(t)$ and the CAP system $q_E^{CAP}(t)$ are different in general.

4. DRIVER MONITORING AND COMPETENCE ASSIGNMENT

Using the models for the driver, the vehicle [4] and the recording of substantial states of the environment (position, coordinates of the road, traffic conditions, etc.) the current condition of the driver $c_D(t)$ can be determined on the basis of a success-evaluated on-line action analysis using classification concepts [2]. It is suggested to make an on-line determination of a driver's condition $c_D(t)$ by sensor fusion according to the structure in figure 3. In this manner at present the behavior of the driver in regard to the distance

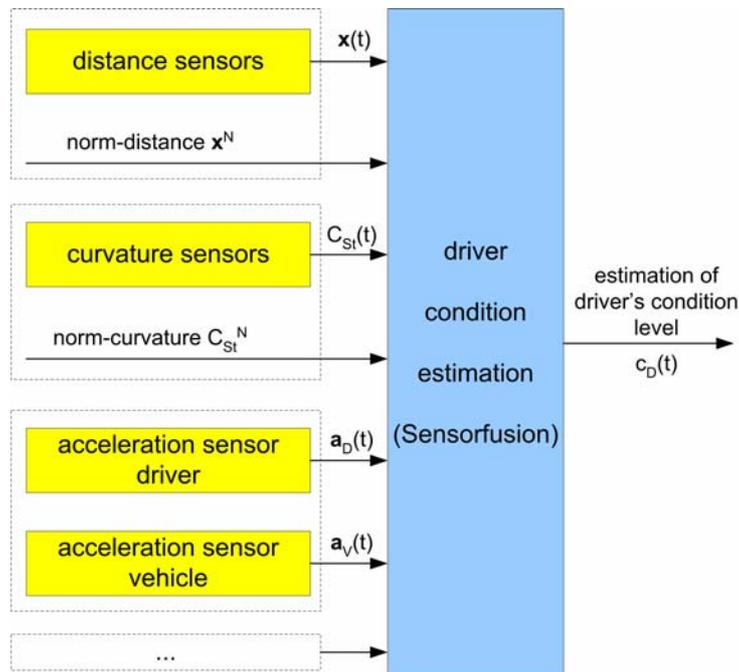


Figure 3: Determining of driver's capabilities with on-board sensors und measured physiological parameters with sensor fusion

to ahead driving vehicles and the driving of curved roads shall be examined. The novel concept addresses the expected standard equipments of vehicles in the coming years. Additionally selected physiological parameters of the driver are included in the monitoring. At first the norm distances x^N for e.g. the situations follow-up drivers, approaching, track switching, in-shearing and out-shearing have to be determined driver-specifically. These are compared situation-dependent with the actual recorded distances $x(t)$ of the driver. As a second input of the driver condition estimation module, the actual curvature driving behavior $C_{St}(t)$ is compared with the norm curvature C_{St}^N derived from maps. A prospective physiological parameter is the model-based evaluation of the driver's acceleration $a_D(t)$ compared to the vehicle's

acceleration $a_v(t)$. The driver is regarded as a transfer element with the vehicle's seat acceleration as input and the driver body acceleration as output. The resulting transfer function depends on the current driver state, i.e. in the manner of fatigue. So by monitoring the driving style as well as the driver's body acceleration changes in its driving condition can be detected. Investigations in the context of estimating the state of pilots are the basis of this strategy [10]. As a result, the estimation of the current driver condition can be expressed in the form:

$$c_D(t) = f\left(\|x(t) - x^N(t)\|, \|C_{St}(t) - C_{St}^N(t)\|, \|a_D(t) - a_v(t)\|\right) \quad (1)$$

As a next step mechanisms for the detection of reduced capabilities of the driver in relation to his driving task have to be developed. The goal is to give recommendations for safe driving though short or long term limitations. Therefore corresponding zones as functions of the environment vector are defined (see figure 4). The CAP's decision is then based on current driver condition $c_D(t)$ and current vehicle capability $c_V(t)$ as well as the current driving parameters like speed or total travel time and is outputted via the competence assignment to driver and vehicle $y_{c,D}(t)$ and $y_{c,V}(t)$. This competence assignment encompasses a range of measures from warnings to the driver or automatically running safety actions of the vehicle if the driver is losing his competence, i.e. by a shock.

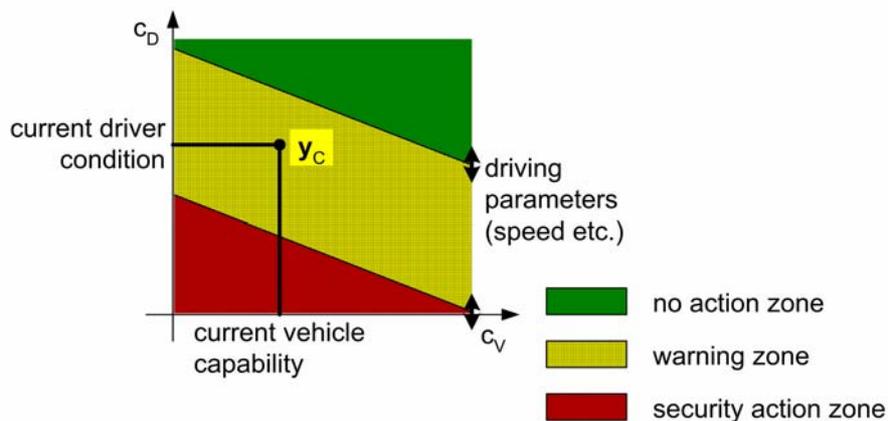


Figure 4: Determining competence assignment to driver and vehicle

5. IMPLEMENTATION ASPECTS

First experiences with an online monitoring of driver and vehicle were made in the project Superfour-in. In this project an outdoor wheelchair was equipped with a special developed surveillance technology. An onboard vehicle control center (VCC) collects technical data from the vehicle and medical data from the driver. This data is then

evaluated automatically. If an emergency or a technical fault is detected, a care service center (CSC) is informed automatically. Several kinds of technical faults are recognized. For instance the motor-currents are used to detect flat tires. Furthermore the state of the hybrid power-system is monitored to predict the remaining reach. In addition the state of the driver is monitored using an interface for bio-signals. A sensor for the measurement of the heart-rate and the blood oxygen saturation was integrated. This measurement allows detecting some kind of acute cardiovascular problems. Especially for the noninvasive measurement of relevant bio-signals in a driving vehicle there are only a few suitable solutions. In the mentioned project the evaluation of the vehicle state and the state of the driver are separate tasks. According to the approach of this paper this two tasks have to be joined to get more detailed information about the state of the driver.

6. CONCLUSION AND FURTHER WORK

The task of the competence adapter presented in this work is to identify the driver and vehicle condition and then allocate competence appropriately. If a durable and/or brief competence loss of the driver cannot be balanced by the vehicle, a mission cannot be realized. Therefore, a separate emergency system has to stop the vehicle in a controlled manner. At current the appropriate models for drivers are developed and driving behavior in the different age phases is being examined.

References:

- [1] Th. Engelhardt, Zustandsbezogenes Instandhaltungsmanagementsystem für mobile Systeme, Dissertation: TU-Ilmenau, 2006.
- [2] M. Koch and Th. Kuhn and J. Wernstedt, Fuzzy Control (Optimale Nachbildung und Entwurf optimaler Entscheidungen), R. Oldenburg Verlag : München, Wien, 1996.
- [3] J. R. Blau and A. Weicher and J. Wernstedt, Investigation of the Competence of Man as Component of Man- Machine Systems on the Base of Pilot Performance, 4th. IFAC / IFORS / JEA Conference on a Man-Machine Systems Analysis, Design and Evaluation : Xi' An, China, 1989.
- [4] Fabian Müller, Andreas Wenzel, Jürgen Wernstedt. Intelligent Monitoring and Adaptive Competence Assignment for Driver and Vehicle. Proceedings of Intelligent Vehicles Symposium, Istanbul, 2007.

Authors:

Fabian Müller

Dr. Andreas Wenzel

Prof. Dr.-Ing. habil. Jürgen Wernstedt

Fraunhofer Application Center System Technology (AST), Am Vogelherd 50

D-98693 Ilmenau, Germany

Phone: +49 3677 461-140

Fax: +49 3677 461-100

E-mail: {fabian.mueller, andreas.wenzel, juergen.wernstedt}@ast.iitb.fraunhofer.de

V. Borikov

Linear Parameter-Oriented Model of Microplasma Process in Electrolyte Solutions

PROCESS IDENTIFICATION AND MODELLING

Abstract: This paper presents the model of microplasma process in electrolyte solution based on linear electric circuit. Elements of the parameter-oriented model were defined. Its electrical parameters and example of the modeling are given. Modeling was carried out in MATLAB with Simulink.

Keywords: microplasma process, linear model, electrical parameters, equivalent circuit of microplasma process.

1. INTRODUCTION

The microplasma method of the metal surface processing is the most perspective method among traditional electrochemical process [1]. The method consists of processing of details in electrolytes by large density currents. During covering the local plasma sparks are observed that is the basic characteristic of process.

Modeling of coating process allows defining the most informative parameters for control and operating of the technological microplasma process of oxide, composite oxide-polymeric and oxide-metal coverings on aluminium, titan and their alloys of various structures.

Simulation is one kind of computer modeling. It is designing of system model and investigations of this model with the purpose either understand behavior of system or estimate (within the framework of restrictions by some criterion or its series) the various modes of the given system functioning [2]. For simulation, the structurally functional model is supplemented by parameters or data describing details of process functioning. Thus, the received model can be considered as algorithm of functioning of object, realized as the computer software complex.

The given researches are directed to construction the equivalent circuit, allowed to establish interrelation of received properties of oxide-ceramic coverings and registered electric parameters of system during microplasma oxidation process. As consequence, it is an opportunity of analysis of these properties at a stage of coverings formation. As a result, there is opportunity of control of microplasma oxidation process on the received cyclic volt-ampere curves with the purpose of creation of the covering with necessary quality.

Besides, the equivalent model will allow to observe change of process parameters without real experiments and will form a theoretical basis for creation of new microplasma technologies and definition of optimum modes, and structures of the process equipment.

2. PARAMETER-ORIENTED MODEL

As a model of the microplasma process, the parametrical model based on linear circuits has been offered.

The choice of linear circuits is caused by two reasons. The first reason is that electrochemical processes in electrolytes are submitted in the literature by the elementary equivalent circuits very frequently, representing combinations of capacity and resistor. In the general case the equivalent circuit for the metal-oxide electrode may be depicted by the parallel connected capacity C and the active resistance R . Full complex resistance of such circuit is:

$$Z = R/(1 + j\omega RC). \quad (1)$$

The separately active and reactive parts of the complete resistance are usually measured in experiment. Therefore, the equation is convenient as following:

$$Z = \frac{R}{1 + (\omega RC)^2} - j \frac{\omega R^2 C}{1 + (\omega RC)^2}. \quad (2)$$

The second reason is the proof of opportunity of the microplasma process control with the information of active and capacity parts of process' volt-ampere characteristics, recommendations for development of new methods of measurement of covering properties, and new methods of the technological processes control.

It was shown from experiment that during process time the active and capacity part of microplasma current was changed. It is connected with changing of active resistance and capacity of the equivalent circuit.

For electrochemical measurements the three-electrode electrochemical cell was used that representing a ceramic glass in diameter of 110 mm, height of 110 mm. The stainless steel electrode by thickness of 2 mm served as auxiliary electrode having the half ring form on internal glass diameter. The surface of the auxiliary electrode exceeded a surface of the working electrode. Standard platinum spherical electrode EPL-02 was chosen as comparison electrode.

In a microplasma mode the sample 40x40 mm size was exposed to processing made of the aluminium alloy 2021, preliminary smoothed out and skim. Experiment was

carried out in 4 componential electrolyte of the following structure, g/l: $\text{Na}_2\text{HPO}_4 \times 12 \text{H}_2\text{O}$ – 30; $\text{Na}_2\text{B}_4\text{O}_7 \times 10 \text{H}_2\text{O}$ – 30; H_3BO_3 – 20; NaF – 10.

The information-measuring complex (Fig. 1) was used for measurement of currents impacted and polarized voltage with the purpose of construction cyclic volt-ampere curves. The information-measuring complex includes the power supply, three-electrode electrochemical system and the measuring equipment for registration and processing of the information (computer measurement system).

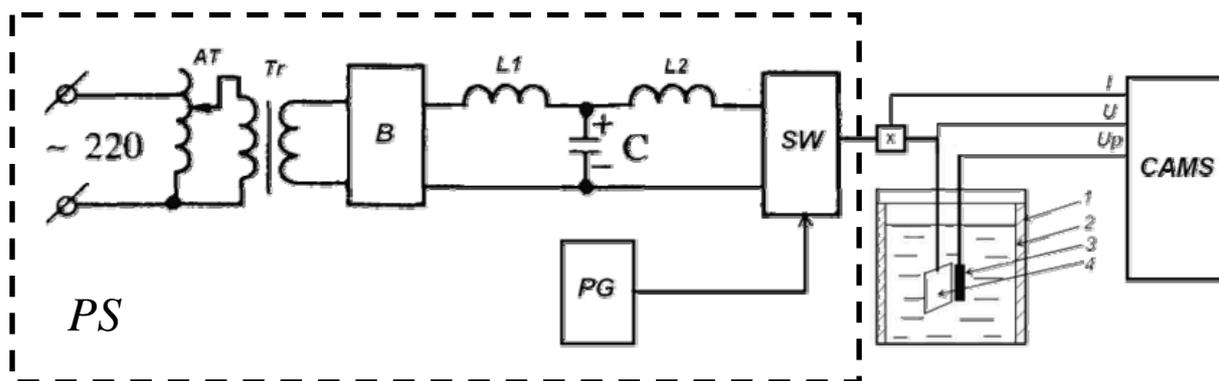


Fig. 1. The information-measuring complex: 1 is ceramic glass; 2 is auxiliary electrode; 3 is comparison electrode; 4 is working electrode, AT – autotransformer, SW – electric switch, B – electric bridge, PG – pulse generator, PS – power supply CAMS - Computer-Aided Measurement System

The pulse voltage (E) from the power supply to microplasma system is allocated between a working electrode (U) solution ($I * R_s$):

$$E = U + I * R_s \quad (3)$$

The computer measurement system allows to receive a volt-ampere dependences of microplasma processes in the pulse mode at voltage up to 300 V, voltage rise-time of 10^8V/sec , currents up to 100 A and to registration of the voltage and current signals with discrete 25 mV and 1 mA accordingly (Fig. 2).

As a result the time-coordinated data appropriate to input signal of voltage U , to voltage on the comparison electrode U_p and data of the current proceeding through sample I are formed in the computer. Measurement of all electric parameters (I, U, U_p) is carried out simultaneously during one pulse (200–250 usec). Representation of the information or as current and voltage diagrams or as tabulated values is carried out automatically.

Influence of microplasma processing time independent on alloy kind has the follow effect: cyclic volt-ampere curves are shifted in area of the large voltage and smaller values of current (Fig. 3).

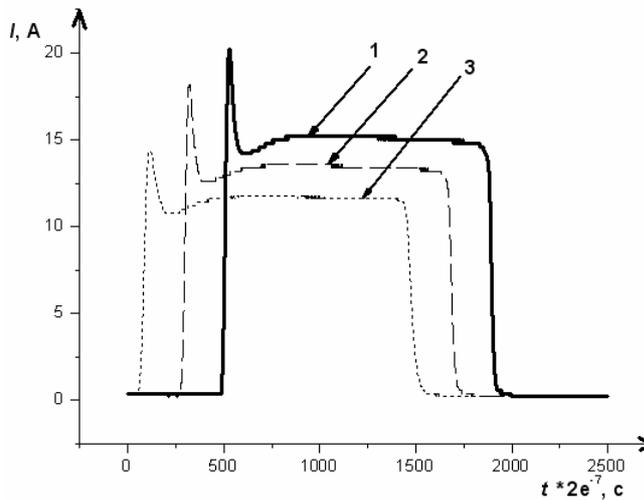


Fig. 2. The curves of current from oscilloscope: 1 - 3 min, 2 - 4 min, 3 - 5 min

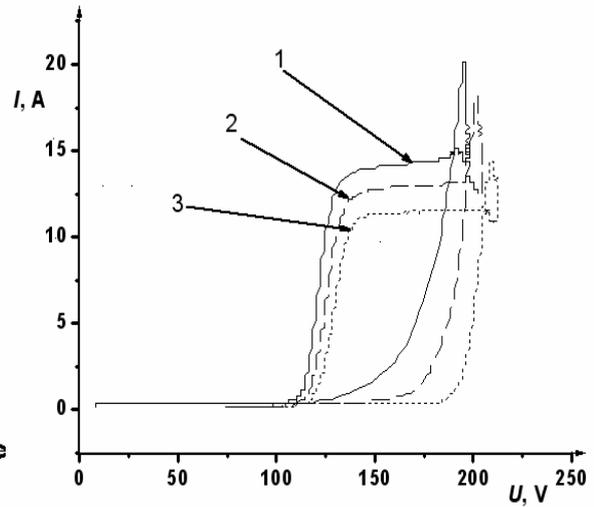


Fig. 3. Volt-ampere characteristics of aluminium alloy 2021 from time of covering process: 1 - 3 min; 2 - 4 min; 3 - 5 min

In this case the equivalent circuit of measurement is shown in Fig. 4, where R_w R_a is resistance of the spark stages on working and auxiliary electrodes, C_w and C_a is capacity of working and auxiliary electrodes, R_s is resistance of electrolyte solution, R_c and C_c is resistance and capacity of the comparison electrode (platinum electrode), R_g is resistance of the voltage source.

3. COMPUTER MODELING OF THE MICROPLASMA PROCESS

Modeling was carried out in MATLAB environment with Simulink toolbox.

The equivalent circuit of the microplasma system in Fig. 5 [3] was used for simulation.

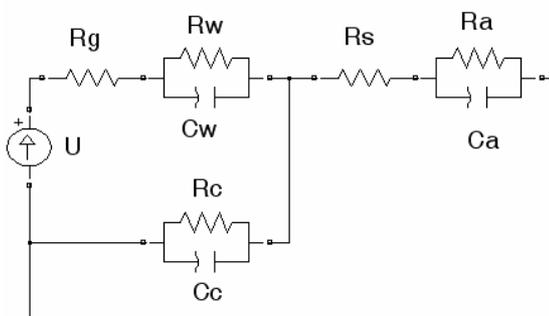


Fig. 4. Equivalent circuit for current from the three-electrode electro-chemistry cell

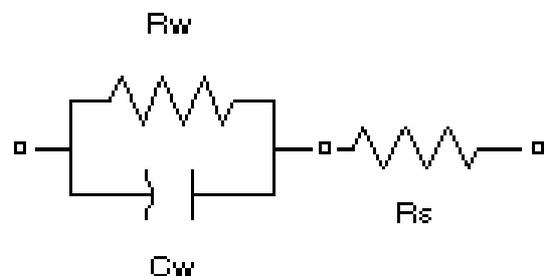


Fig. 5. The equivalent circuit of the microplasma system: C_w – the capacity part of electrode reactance, R_w – the active resistance of electrode, R_s – the solution resistance

In MATLAB environment the simulation model of coating system (Fig. 6) was designed. Elements of the circuit appropriate to experimental data for the current and voltage forms.

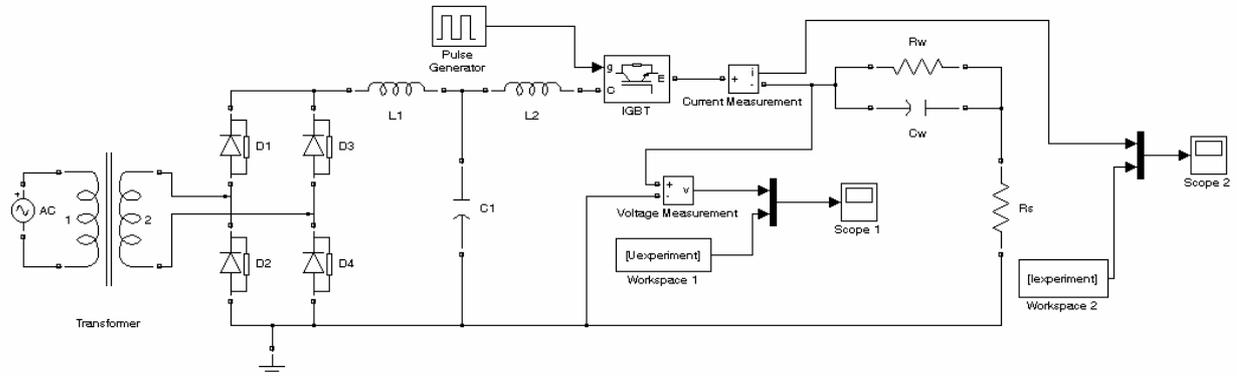


Fig. 6. The coating model in MATLAB

The data of current and voltage measured on experimental device [4] at the various points of time were used for modeling: a pulse voltage from the power source, the polarizing voltage (between the comparison electrode and the working electrode) and current in circuit.

Definition of resistance and capacities was made based on experimental data from the formula:

$$E_{PS} = U_c * \left(\frac{R_w}{R_s} + 1 \right) + L * C * \frac{\partial^2 U_c}{\partial t^2} + \left(\frac{L}{R_w} + C * R_s \right) * \frac{\partial U_c}{\partial t}, \quad (4)$$

recognizing that the RLC resonance is presented in system.

As a result of modeling the voltage pulse (Fig. 7) and the current pulse (Fig. 8) is received appropriate to form and size of the experimental data.

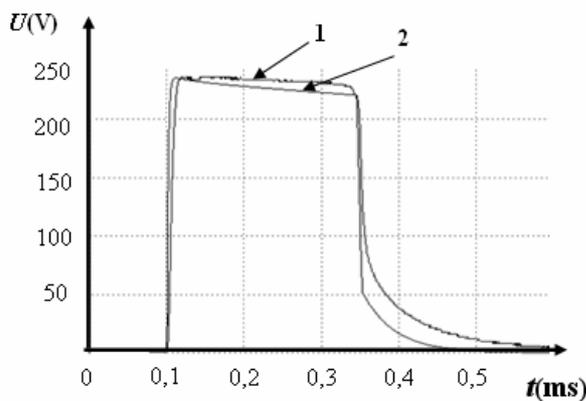


Fig. 7. The voltage pulse:
1 – from experiment, 2 – from model

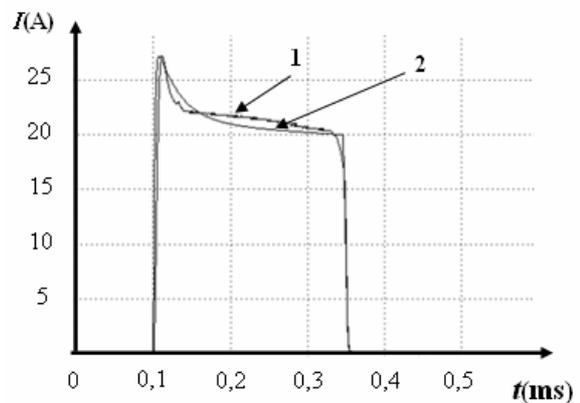


Fig. 8. The current pulse: 1 – from experiment, 2 – from model

4. CONCLUSION

As a result it is possible to present the microplasma process for pulse influences model with use of the certain restriction linear circuits. The given restriction is division of microplasma process into 3 phases: formation of a barrier layer, microplasma breakdown and relaxation of system.

The received results give initial approximation of the microplasma system model characteristics in two stages. Values of elements of system are received for the simulating model: $R_s=8.2$ Ohm, $R_w=2.8$ Ohm, $C_w=15$ uF for covering surface area $S=16$ cm², $t=20$ min. The third phase will differ of the oxide layer resistance which is not subject to breakdown, in the given phase resistance is $R_w=30$ Ohm.

The simulated signal retains approximately 95 % of the original signal energy. The estimation on energy mismatch was made on square-law norm:

$$\delta = \frac{\int_{t_0}^t [x(\tau) - y(\tau)]^2 d\tau}{\int_{t_0}^t x(\tau)^2 d\tau}, \quad (5)$$

where $x(\tau)$ is signal received from experiment, $y(\tau)$ – during simulation.

Models describing behavior of electrochemical system will allow to create power supplies for micro-arc oxidation equipment, plasma processing in electrolytes as well as to define parameters for coating control. Researches in the given field are carrying out. The modeling of the microplasma process with use of the nonlinear circuits, in particular, introduction of Zener diode is planned.

5. ACKNOWLEDGMENT

The study has been carried out with the support of analytical framework program “Development of scientific potential of the high school (2006-2008)” RNP 2.1.2.5253. Author thanks Professor A.I. Mamaev for his valuable support and useful discussions.

References:

- [1] Kurze P. et all. Micro Arc/ Spark Anodizing – was ist das? Micro Arc/ Spark Anodizing – what is that? // Galvanotechnik. № 8. – 2003. – P. 1850 – 1863.
- [2] Lennart Ljung, System identification: Theory for the User, Prentice-Hall, 1987
- [3] A.I. Mamaev, V.A. Mamaeva, “The High Current Processes in the Electrolyte Solutions”. Novosibirsk: Siberian Branch of the Russian Academy of Science, 2005.
- [4] A.I. Mamaev, V.N. Borikov, T.I. Dorofeeva, V.A. Mamaeva, “Computer Measurement System of Electric Parameters of the Microplasma Processes in Solutions of Anodizing”. Protection of Metals, 2003. No. 1, p. 23–26.

Authors:

Dr., Assoc. Professor Valery Borikov
Tomsk Polytechnic University, 30 Lenin Avenue
Tomsk, 634006, Russia
Phone: +7 3822 417527/ Fax: +7 3822 420449 / E-mail: borikov@camsam.tpu.ru

A. Avshalumov / G. Filaretov

Detection and Analysis of Impulse Point Sequences on Correlated Disturbance Phone

INTRODUCTION

The problem of low-powerful impulse point sequences analysis on phone of relatively intensive stochastic disturbance is considered. Subsequent consideration will be realized under next conditions:

a) The observable discrete process z_t can be represented as sum $z_t = x_t + \dot{t}_t$; $t = 1, 2, \dots, N$, where process x_t (disturbance) is the discrete correlated Gaussian process with variance σ_x^2 and standard autocorrelation function $\rho_{xx}(k)$, $k = 0, 1, 2, \dots$

b) Impulse component of observable process z_t is the Poisson impulse sequence; distribution function of intervals between impulses - $\varphi(t_u) = \lambda \cdot \exp(-\lambda t_u)$, where λ - unknown intensity of impulse point sequence, which necessary to find with help of observable realization.

в) Distribution function of impulses amplitudes $f(A_u) = f(A_u; \theta_1, \theta_2, \dots, \theta_k)$ is known up to parameters, which also have to estimate by experimental dates.

г) Impulses number M is too small in comparison with common discrete observations N (not more, than 10 – 15 %).

METHOD OF DETECTION AND ANALYSIS

Proposed method for detection and parameters estimation of Poisson impulse sequence includes two stages.

Stage 1 intended for detection and position localization of impulse sequence points. This detection can be effectively realized with the help of one of the algorithms proposed by authors [1]. The main purpose of this algorithm is to detect anomalous observations. Here, we understand anomalous observations to be those which essentially breaks the smoothness of observable sequence trajectory.

This algorithm is based on detection of statistically significant deviation observable value z_t from point z_t^* , which is found by linear interpolation in two neighbour points z_{t-1} and z_{t+1} .

Define stochastic value y_t :

$$y_t = z_t - z_t^* = z_t - \frac{z_{t+1} + z_{t-1}}{2} \quad (1)$$

This value, if anomalous observations are absent, has Gaussian distribution with zero mean and variance σ_y^2 . Then presence of anomalous observation in point z_t can be established by detection of significant deviation value y_t from zero, that is if

$$|y_t| > u_{1-P} \cdot \sigma_y \quad (2)$$

Here u_{1-P} - Gaussian distribution quantile, appropriated to confidence probability P .

Really for detection of anomalous observation it is more conveniently the other formula. For its finding we present y_t in another form:

$$y_t = z_t - \frac{z_{t+1} + z_{t-1}}{2} = \frac{(z_t - z_{t+1}) + (z_t - z_{t-1})}{2} = \frac{-\nabla_{t+1} + \nabla_t}{2} = \frac{-\nabla_{t+1}^{(2)}}{2}, \quad (3)$$

where $\nabla_{t+1}^{(2)}$ - the second order difference for time moment $t+1$. Then instead of (2) we can write the equivalent formula, using the second order differences:

$$|\nabla_{t+1}^{(2)}| > u_{1-P} \cdot \sigma_{\nabla^{(2)}} \quad (4)$$

In common form thus algorithm of anomalous observations detection includes next sequence of operations:

- Computing of the first order $\nabla_t = z_t - z_{t-1}$, $t = \overline{2, N}$ and the second order differences $\nabla_t^{(2)} = \nabla_t - \nabla_{t-1}$, $t = \overline{3, N}$ for time series z_1, z_2, \dots, z_N .
- Estimating of variance $\sigma_{\nabla^{(2)}}^2$ for time series $\nabla_t^{(2)}$.
- Choice of confidence probability P (usually $P = 0,90 - 0,99$).
- Fixing of anomalous observations: point z_t is related to the category of anomalous observations, if inequality (4) takes place.
- Elimination of fixed anomalous observations from time series z_1, z_2, \dots, z_N : all these points are replaced by new values, which are computed with help of linear interpolation in two neighbour points.

In practice we need to organize iterative regime of this algorithm work. This necessity is connected with distinguishes of statistical properties z_t and x_t , when there are anomalous points. In process of these points eliminating time series z_t , corrected by this way, will be approximated more and more to process x_t . The iterative process lasts until new anomalous points have not been found during the latest iteration. As a rule two iterations are enough for this. As a result the position of all found anomalous points on discrete time scale $t_1^*, t_2^*, \dots, t_{m+1}^*$ is fixed.

Stage 2 provides for realization of the next steps:

➤ For every found anomalous point the impulse amplitude is computed as deviation of observation in this point $z_{t_j^*}$ from value, which is found by linear interpolation in two neighbour points

$$A(t_j^*) = z_{t_j^*} - \frac{z_{t_{j+1}^*} + z_{t_{j-1}^*}}{2}; j = 1, 2, \dots, m. \quad (5)$$

After this the amplitudes distribution function $f(A_u)$ histogram with ordinates m_i ; $i = 1, 2, \dots$ is built.

➤ Extracting of subset I , which contains substantively warped values of histogram ordinates on account of limited sensitivity of the detection anomalous observations algorithm. This extraction can be made or visual with help of histogram, or by formal exception of histogram intervals from zone $\pm u_{1-p} \cdot \sigma_{\nabla(2)}$; value $\sigma_{\nabla(2)}$ is computed for corrected z_t after last iteration.

➤ Estimation of distribution function $f(A_u)$ parameters $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$ using values m_i , which not belong subset I .

➤ Determination of ordinates \hat{m}_i for histogram intervals, which belong subset I , using function $f(A_u; \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)$; building of corrected histogram.

➤ Computing of loss coefficient K_I in fixation of anomalous points, connected with limited sensitivity of the detection anomalous observations algorithm

$$K_I = \frac{m}{\sum_{i \notin I} \hat{m}_i + \sum_{i \in I} \hat{m}_i}. \text{ This coefficient defines the relativity decrease of impulses}$$

numbering comparison with its real value.

➤ Determination of time intervals between neighbor detection points $t_{uj}^* = t_{j+1}^* - t_j^*$; $j = 1, 2, \dots, m$. As impulses with amplitudes near zero cannot be detected, selected point process differs from real: it is more rare. However on account of random location of nondetected points, this process remains Poisson impulse process [2], but with another intensity $\lambda^* < \lambda$.

➤ Estimating of intensity parameter λ^* with help of maximum likelihood method:

$$\hat{\lambda}^* = m \cdot \left(\sum_{j=1}^m t_{uj}^* \right)^{-1}.$$

➤ Determination of corrected estimation for intensity parameter $\hat{\lambda} = \hat{\lambda}^* / K_I$.

Proposed algorithm works out the formulated above problem in full volume. It is necessary to underline that quality of end result depends on peculiarities of concrete applied task (what kind are characteristics of process x_t , impulse component, function $f(A_u)$ and so on).

MODEL EXAMPLE

Potential possibilities of proposed algorithm are illustrated by the next model example. Process x_t is formed by quadruple passing of discrete white noise through inertia element (constant time – 10 discrete time units). The observed realization length - $N= 20020$ discrete volumes. Poisson impulse point process includes 917 points. Model (empirical) volume of intensity $\hat{\lambda}_M = 0,046$. Distribution function of impulse amplitudes – exponential - $f(A_u; \theta_1) = \theta_1 \exp(-\theta_1 A)$; $\theta_1 = 250$. Histograms, characterizing model distributions of amplitudes and intervals between impulses, are represented on fig. 1a) and fig. 2a).

Stage 1: for the first iteration 565 points were detected (critical value $u_{1-P} = 1,28$); for the second iteration 158 points were detected (critical value $u_{1-P} = 3,09$); the third iteration was not fulfilled. In total 723 points from 917 were detected.

Stage 2: with the formula (5) amplitudes $A(t_j^*)$ are estimated and the appropriate distribution histogram is built (fig. 1b). As the subset I we choose points, belonging to the first histogram interval (number of these point is equal 164). Using

the all other histogram intervals with help of nonlinear estimation method we find approximation for dependence of histogram ordinates from interval centers and also unknown parameter of exponential (under condition) amplitude distribution:

$$m = b \cdot \exp(-\hat{\theta}_1 A). \quad (6)$$

The next estimations were got: $b = 461,87$; $\hat{\theta}_1 = 250,8$. With help of formula (6) we define the estimation for points number in the first grouping interval or another words – for subset I : $\hat{m}_1 = 359$. Corrected histogram with using of \hat{m}_1 practically coincides with histogram of fig. 1a). After this the loss coefficient K_I in fixation of anomalous points was computed: $K_I = 0,788$.

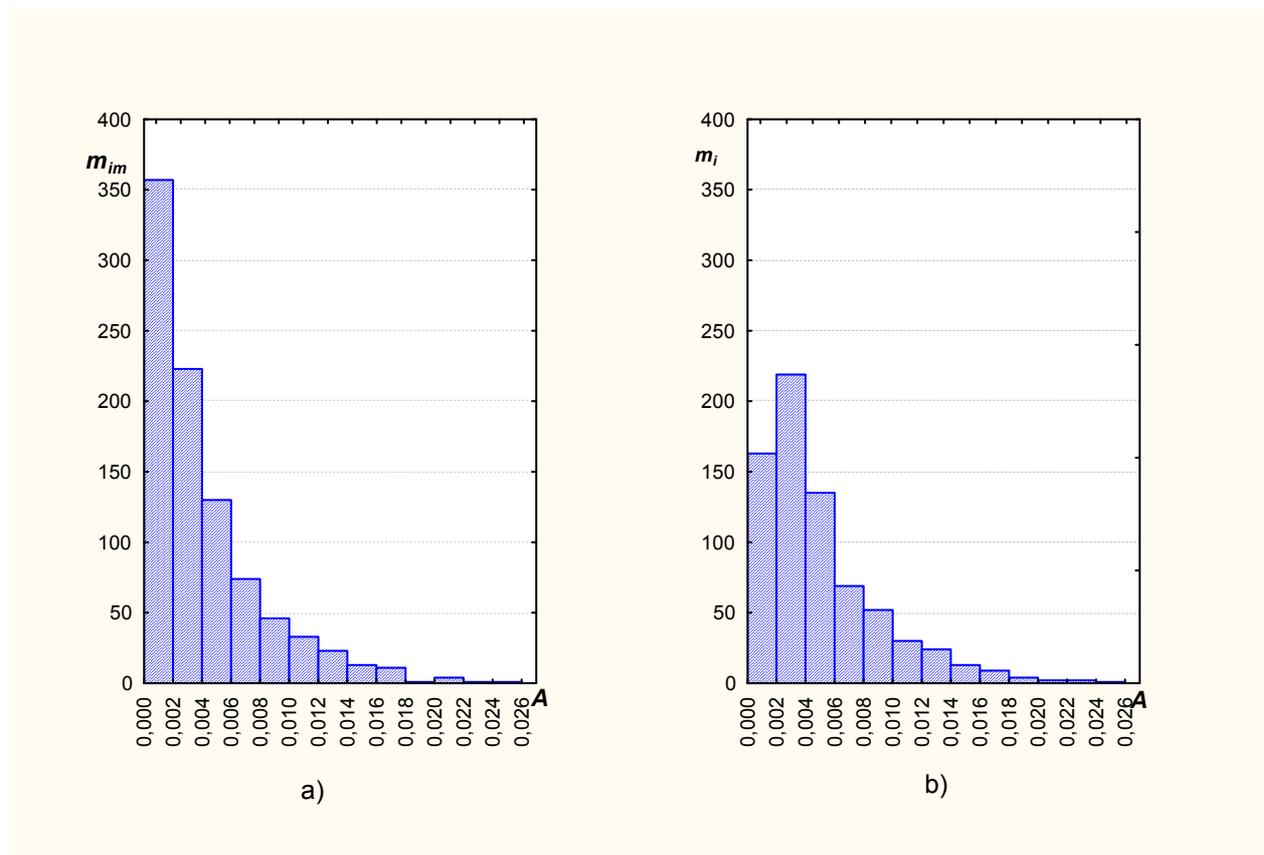


Fig. 1

Furthermore, time intervals between detected points were established. This intervals distribution histogram is shown on fig. 2b). How it was waited, we have distribution of exponential type. The intensity parameter estimation for this distribution was found: $\hat{\lambda}^* = 0,0361$. And, at last, corrected estimation of intensity parameter is computed: $\hat{\lambda} = \frac{0,0361}{0,788} = 0,0458$.

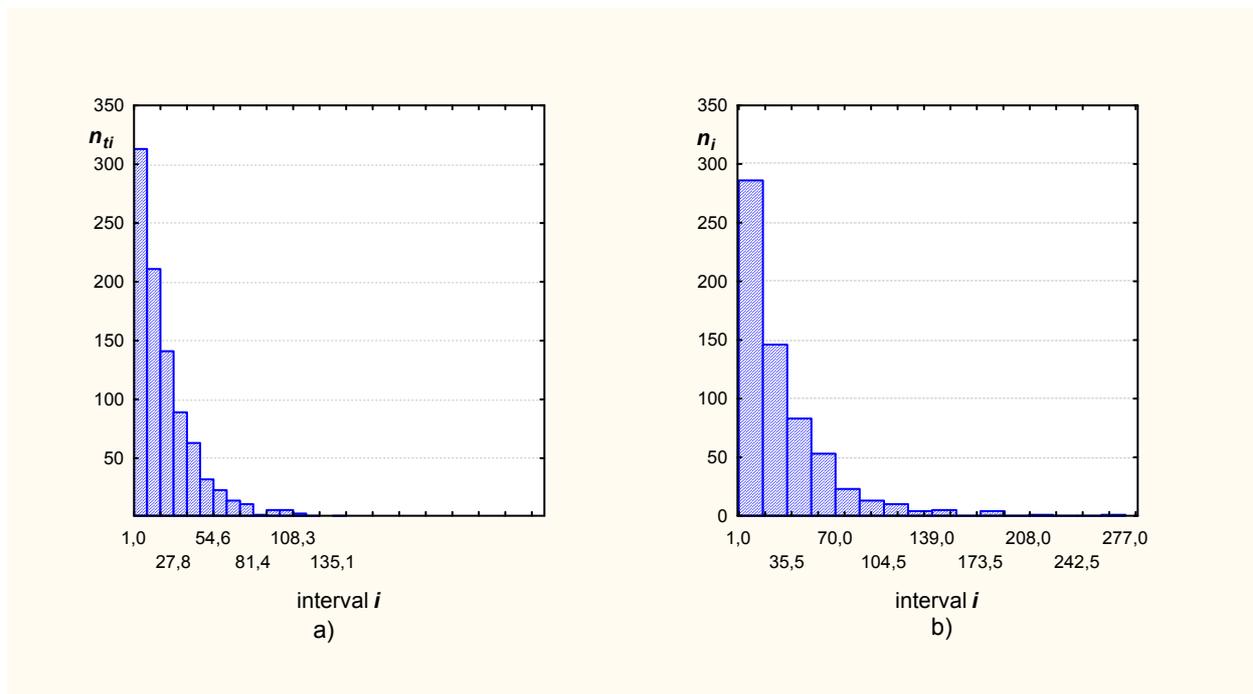


Fig. 2

Summing up, it is possible to say, that proposed method of detection and analysis is very effective. The relative error for amplitude mean is estimated to be less than 1%, and for the intensity parameter of Poisson impulse point process – about 2%. It is substantively that impulse point process capacity is less 0,1% from capacity of correlated stochastic process, on phone which this point process is detected and analyzed.

CONCLUSION

Areas of possible application proposed method can be various. In particular, it was used for aims of medical diagnostic as means of heart rhythm infringements.

References:

- [1] Avshalumov A.Sh., Filaretov G.F. Algorithm of anomalous observations detection in correlated stochastic sequences. // Vestnik MEI. – 2007, №3.
- [2] Cox D.R., Lewis P.A.V. The statistical analysis of series of events. New York: John Wiley, 1966.

Authors:

Dr. Gennadiy Filaretov
 Moscow Power Engineering Institute,
 Krasnokazarmennaya st., 14, 111250, Moscow,
 RUSSIA
 Phone: 007(095)517 63 19
 e-mail: gefefi@yandex.ru

Alexander Avshalumov
 Moscow Research Institute of Cybernetic
 Medicine,
 Mikluho-Maklaya st., 16/10, 117871, Moscow,
 RUSSIA
 Phone: 007(095)7632660
 e-mail: avs@cybermed.ru

Horst Salzwedel

COMPLEX SYSTEMS DESIGN AUTOMATION IN THE PRESENCE OF BOUNDED AND STATISTICAL UNCERTAINTIES

Abstract

Complex technical and organizational systems are designed by groups of people departments and/or companies. Because of product uncertainties during development, each of the individual engineers make individual assumptions in their employed design methodologies. Because of the complexity of the system they do not understand the impact of their decisions on the overall systems. Since current design approaches do not consider the impact of this uncertainty on the integrated system, the design specifications are error prone and not validated[1]. The likelihood that a system design does not have critical errors is less than 4%[2].

In this paper design methodologies of individual engineers are modeled and connected by a design process graph to make them executable. Bounded and statistical uncertainties of early design stages are included and bounds on integrated designs based on uncertainties at different design stages are determined using MLDesigner 3.0 [9] Simulation Set capabilities. The new design methodology permits treating uncertainties in early design stages, validate design specifications, and optimize the design flow for different criteria, like quality in design and design cost and verify systems in the presence of this uncertainty.

Introduction

Complex systems like networked systems in vehicles (aircraft, spacecraft, automobiles, ships, trains, autonomous systems), between vehicles, IT systems, communication or organizational systems can only be developed by groups of people, departments and/or companies. To meet time to market requirements, the system design is partitioned into subsystems and subsystem design is distributed to specialist teams before uncertainties about individual subsystems and interactions between subsystems has been resolved. The individual engineers make assumptions, many based on experience with less complex systems. It is not possible for these engineers to understand the impact of their decisions on these systems. Since

the models passed on for system integration do not include the model uncertainties, the uncertainty of the integrated system is not determined. The integrated system model or design specification is therefore not validated. The likelihood that this design approach will work without critical problems is less than 4% [2]. Very time consuming and expensive integration processes try to resolve conflicts and critical performance issues.

Figure 1 [3] depicts the probability of critical problems as a function of phases of the design process. The probability of critical problems is very high in early design stages and is low in late design stages. These failure rates often result in huge cost overruns for correcting mistakes in complex systems design. Several billion \$/€ projects were terminated due to these failures.

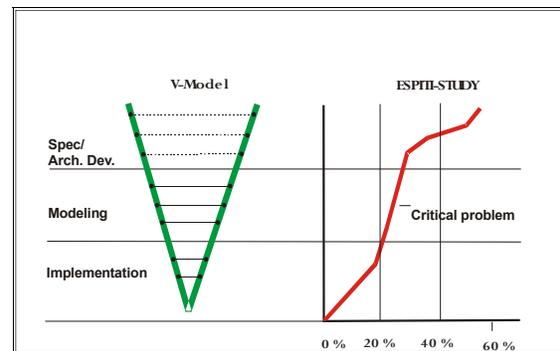


Figure 1: Critical problems in design of complex systems [3]

The analysis of designs shows that most causes for critical design flaws come from poor design specifications. The main reason is that most specifications are on paper only and therefore cannot be validated. The introduction of executable specifications [4] has improved the quality of specifications and significantly reduced the number of design iterations and the cost of design of electronics, software and organizational systems [5]. However, for complex systems, like large IT systems, complex systems design flows and aerospace systems the likelihood that the first design will be right is very low. In addition, in

silicon systems, cost for organizational processes for design and development is increasing exponentially and is already more than 30% for complex chip designs [6].

Development of executable development and organizational process models can catch critical problems in development processes [7] or result in significant improvements and cost savings in organizational processes [8]. However, in most projects these available technologies are not yet employed.

Even so uncertainty plays a major role in early design stages, it is only treated at a subcomponent level and not at the system level. The reason is that as the number of designers/design teams increases, the number of design alternatives increases. For example, 2 different design points of each designer in a 10-person design team, where each team member generates two alternatives, results in 2^{10} different designs. Current manual integration methods cannot cope with this size of design space exploration and verification.

In this paper a system design methodology is developed based on executable design flows. Bounded and statistical uncertainties of early design stages are included and bounds on integrated designs based on uncertainties at different design stages are determined. Examples will demonstrate the new design methodology.

The new design methodology not only permits to treat uncertainties in early design stages, but permit optimizing the design, as well as the design flow for different criteria like quality in design and design cost. Design changes can easily be implemented and design tracking automated.

Product Uncertainty

Information, knowledge and uncertainty about a product changes significantly during product life time. This knowledge about the product includes knowledge about the product itself, knowledge about the intended use, and knowledge about the environment in which the product is to be used. This uncertainty is bounded and may sometimes be described by bounded parameter sets, bounded statistical distributions, and/or bounded functional behavior described by e.g. H^∞ norms for continuous systems.

Figure 2 shows a typical behavior of the mean of a product uncertainty. At the start of a development project the product uncertainty,

U_0 , is a function on whether the development team has developed similar type of products before. During the development phase the knowledge about the product increases with the learning rate, l , reducing product uncertainty. This learning curve can be described by an exponential function,

$$U_D(t) = \begin{cases} U_0 e^{-lt} & \forall t < t_D \\ U_0 e^{-lt_D} & \forall t \geq t_D \end{cases},$$

where

$$l = \begin{cases} l_P & \text{for } 0 \leq t < t_P \quad \text{on paper} \\ l_M & \text{for } t_P \leq t < t_M \quad \text{model based} \end{cases}$$

for initial slow development on paper and rapid development with executable models, respectively.

At the end of the development phase, the product is delivered to customers, who will find additional insufficiencies that are reported to support and corrected by the development team. The learning effect due to customer testing and support can be described by,

$$L_S(t) = \begin{cases} 0 & \forall t \leq t_D \\ L_{S0}(1 - e^{-l(t-t_D)}) & \forall t > t_D \end{cases}$$

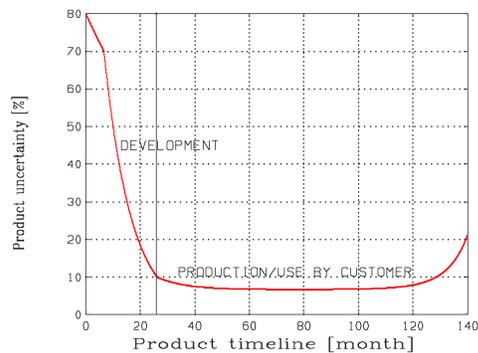


Figure 2: Product uncertainty

However information about a product is also lost during lifetime of a product. The reasons are that people leave development teams, are reassigned, or just not remembering all the decisions which went into the design of a product. Additionally, as time progresses, products are often used for applications that were not considered

during the development phase. These effects impose additional uncertainty on the product, which can be approximated by,

$$U_L = U_{L0} e^{L t^2}$$

As development of products with long life times like the Space Shuttle showed, people even have to be called back from retirement in order to overcome this loss of product information.

The uncertainty about a product can then be expressed by,

$$U_p(t) = U_D(t) + U_L(t) - L_S(t)$$

Comparing Figure 1 with Figure 2, we observe that observed occurrence of critical design errors and product uncertainty are highly correlated. We may conclude that critical design errors are primarily due to information uncertainty about a product and design processes that do not consider this uncertainty in the right way and do not validate and verify the system considering this uncertainty.

In the following sections it is shown how uncertainty in use cases, environment and design may be used to determine the system uncertainty. Making the design process itself executable enables its validation, harmonization, and may result in significant savings in development time and risk.

Complex System Design

Complex systems are always designed and developed by groups of people, departments and/or companies. The system is partitioned into subsystems and subsystems are assigned to different groups. Each group typically has unique knowledge in a particular field and will use their own methodology, models and software for the design of their assigned component. Some will use different software systems.

For design, validation and verification design groups will consider a range of possible parameters and use approximate models for analysis. Different designers/developers will exchange information about their components to other designers according to design or process graphs, Figure 3. The exchanged information typically will not include all assumptions and ranges of uncertainty considered in the design of components. The reason is that some of the uncertainties are very specific to the

field of knowledge used for a component and other designers will not understand this information nor will they be able to properly evaluate potential interactions between components in different subsystems. Other reasons include that the variability of a design may be too large to be considered for manual system integration.

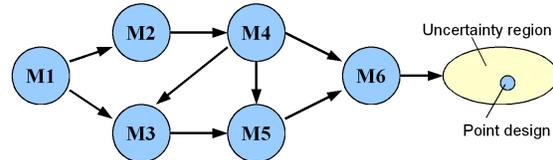


Figure 3: Design process graph

Ten designers doing a conceptual design and each is passing on a minimum and a maximum of one component specific parameter to the integration team, the integration team would require to integrate 2^{10} different system designs for analysis and verification. This would be too time consuming when done manually. It is therefore not done. For system integration, each team will pass on their “best” design to the integration team.

The integrated system is therefore not verified for variability due to uncertainties in subsystems and interactions between components located in different subsystems. As a result, the probability that this coupling is causing critical errors in the design is very high. Testing (expensive and time consuming) should eliminate most critical errors, but customers increasingly find unresolved critical errors after delivery of products. Additionally, different designs based on the same platform design will exhibit very different probability of failures.

Design Automation

To integrate, validate, and verify different designs (e.g. 2^{10}), and find the “best” solution between them is an impossible task if done manually. The design must be automated, in order to overcome this problem. 2^{10} iterations of a simulation are not very unusual, dependent on the complexity of the simulation. With distributed simulations and multi core processors even complex simulations can be run for, e.g., $\geq 2^{10}$ iterations.

For an automated design, a simulation must be developed that meets the following requirement,

- 1) modules, that are complete independent simulations of design methodologies
- 2) the simulations may be executed by the same or different simulation tools
- 3) a design process graph connects the design methodologies
- 4) Monte Carlo simulation capabilities, in order to be able to analyse sets or ranges of parameters
- 5) simulation model generation, in order to iterate over different architectures
- 6) optimization methodologies that can optimize a design with respect to system level objectives
- 7) distributed simulation to overcome potentially large processing requirements
- 8) the simulation must be connected to a data base, in order to keep track of the large volume of information
- 9) models and data must be stored in a standardized way, in order to ease comparison and analysis

Figure 4 depicts such an automated design process simulation. Each designer has to develop an executable model of her/his design methodology. Process engineers design the process design graph that connect the executable design methodologies.

The simulation generates a set of feasible designs and maps component uncertainties into system uncertainties of coupled designs and can determine which designs meet system requirements for all uncertainties in parameters, architecture, missions/use cases, and environment. If the system performance falls within the permissible performances of the design specifications, the design is verified.

The system uncertainty determined by the range of design variations also depicts the level of uncertainty at a given stage of development. As the different members of the design team get up the learning curve, Figure 2, this uncertainty is going to decrease. Unacceptable levels of system uncertainty can be analyzed, causes determined and eliminated.

The automated design process can be connected to a data base and automatically track design changes and their impact on overall system design criteria, including performance, cost, quality of design.

The design process itself can be modeled and optimized. In [7] the development process for a railway switching system was modeled in MLDesigner and optimized for team selection, team load. Critical components of the development process were identified and eliminated that would have doubled the development time. In [10] the design and development process for automotive electronic control units was optimized using genetic algorithms for finding the optimal combination of design methodologies (team selection), performance (timing, BER, cost and quality. Areas for significant improvements were identified.

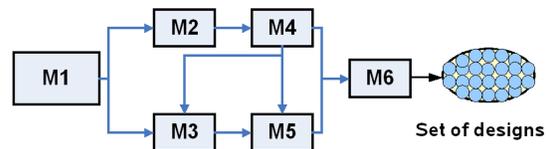


Figure 4: Executable design methodologies, connected by a design process graph

Implementation of Design Automation

The proposed design automation technology is being implemented and analyzed in the multi domain simulator MLDesigner [9], which already includes multiple execution domains, reducing the need of interfaces to other tools. When necessary, MLDesigner provides hooks that make it easy to interface to tools with similar execution domains. The Monte Carlo simulation capabilities of MLDesigner automatically distribute simulations to other computers that are registered to MLDesigner. MLDesigner 2.7 includes an SQL library for easy data base access.

All MLDesigner models are stored in XML, supporting mapping though XSLT scripts. Baumann [11] is developing network building blocks and XSLT based mapping functions to automatically map functional level models into architecture and combined architectural and functional models into implementations. Fischer

[12] is developing system level architecture optimization technology which generates the information about the optimal architecture, that can then be read by the mapping functions of Baumann. Rieher [13] is developing a standardized XML based data format for MLDesigner and Octave that is compatible with the Open Office Document Standard and can be read and generated by spreadsheets from Open Office and Microsoft Office. Genetic algorithm optimization methods have been developed to optimize MLDesigner models of design processes for performance and quality. These algorithms will be extended for optimization in automated design processes.

The next version of MLDesigner (3.0) will provide new simulation set capabilities which have been developed to make it possible to connect independent simulations, using UML activity diagram syntax that describes control flow, design flow, work flow and/or analysis flow connecting the elements of the simulation set. Parallel and sequential execution will support automated distributed simulation of independent simulations. The design process graph can be defined by the standard graphical model editor of MLDesigner, Figure 5.

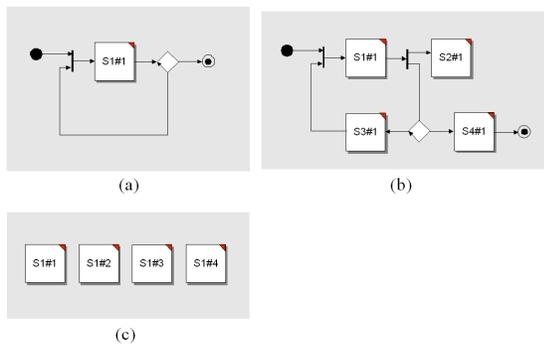


Figure 5: Sequential executions with termination conditions (a,b) and parallel executions (c) of MLDesigner simulation sets

Examples

This sections show two very simple examples using the simulation set capabilities of MLDesigner 3.0 and using two external tools --- SatLab and Octave --- for analysis. These low complexity examples can be verified directly in MLDesigner or by SatLab and Octave.

Example 1: The accuracy of a least squares signal processing algorithm is analyzed for the example of fitting a fourth order polynomial

$$y_i = p_1 + p_2 x_i + p_3 x_i^2 + p_4 x_i^3 + v_i, \quad v = N(0, R_v)$$

through a data set with normally distributed measurement errors. 30 different data sets are generated by MLDesigner and 30 to 1400 measurements are taken of these data for parameter estimation. The parameter estimation is performed by SatLab. Figure 6 shows how the parameter estimation error change as a function of number of measurements taken.

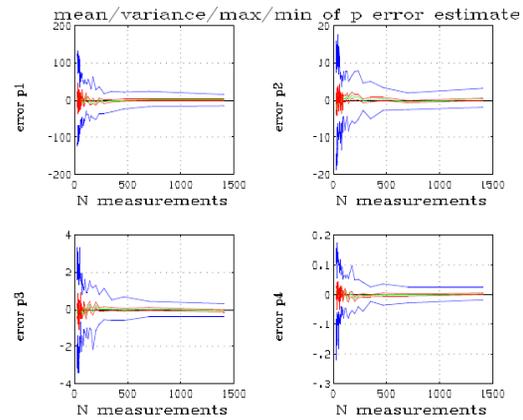


Figure 6: Analysis results for accuracy of least squares parameter estimation algorithms

Example 2: Chapter 10.3 of Reference 14 shows a design methodology for a design specification of a longitudinal controller for an aircraft (Boeing 747). Variations in loading and fuel consumption will affect aircraft dynamics parameters and flight behavior. These parameters changes are bounded by minimum and maximum weight and the permissible minimum and maximum center of mass location for the aircraft. Using maximum and minimum values instead of nominal values for seven longitudinal aircraft dynamics parameters, results in $2^7=128$ different dynamic models of the aircraft pitch behavior. Figure 7 shows minimum, mean and maximum behavior of an altitude change of the aircraft are affected by 40% changes in pitch parameters

$X_u, X_w, Z_u, Z_w, M_u, M_w, \text{ and } M_q$. The results show that the chosen closed loop eigenvalues for the autopilot design is very robust

against parameter changes and passengers will hardly notice the changes in behavior.

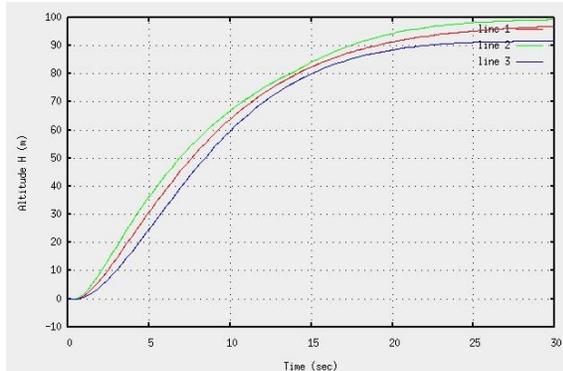


Figure 7: Aircraft altitude hold autopilot response verification for bounded aircraft parameter changes of $\pm 40\%$

Conclusions

Uncertainty of information in design of complex systems is the primary reason for critical problems and expensive corrective actions. A design automation methodology based on coupled models of design methodologies of participating design groups and developers is proposed that overcomes this critical design problem and enables large design space exploration and verification of uncertain systems. The developed methodology permits verification of system level functional and performance behavior in the presence of system uncertainty. Digital as well as analog systems may be verified by this methodology. A methodology for implementing this methodology by the simulation set capabilities of the development tool MLDesigner is shown. Two simple examples show how statistical and bounded system uncertainties may be mapped into bounded system behavior by automating design and analysis using an executable design flow.

REFERENCES

- [1] John Hines, We Don't Do Design Correctly!, Keynote speech at IEEE 9th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems, MASCOOTS 2001, Cincinnati, Ohio, Aug. 15-18, 2001.
- [2] Bruno Schienmann, *Kontinuierliches Anforderungsmanagement*, Addison-Wesley, 2002. – ISBN 3-8273-17-87-8

[3] Horst Salzwedel, Mission Level Design of Avionics, *AIAA-IEEE DASC 04 - The 23rd Digital Avionics Systems Conference 2004*, Salt Lake City, Utah, USA, October 24-28, 2004.

[4] Holger Rath, Gunar Schorch, Horst Salzwedel, Simulationsumgebung für Bordnetze - Bordnetz-Spezifikationen modellieren und validieren, *Hanser automotive*, 5-6.2006, S. 34-38, Carl Hanser Verlag München, ISSN 1860-5699.

[5] Horst Salzwedel, Frank Richter, Matthias Kühn. Standardized Modeling and Simulation of Hospital Processes - Optimization of Cancer Treatment Center, *International Conference on Health Sciences Simulation, HSS '07*, San Diego, California, January 14-18, 2007.

[6] Ralf Pferdmeiges, Presentation on current challenges in chip design, *EDACentrum Workshop: System Planning*, Hannover, Germany, November 30, 2006.

[7] Manuela Tröbs, *Analyse, Optimierung und Simulation des Simis IS Entwicklungsprozesses*, Thesis, Siemens/Technical University Ilmenau, June 2006.

[8] Matthias Kühn, Frank Richter, Horst Salzwedel, Process simulation for significant efficiency gains in clinical departments – a practical example of a cancer clinic, to be presented at 52th International Scientific Colloquium, Ilmenau, September 10-12, 2007.

[9] MLDesigner® Manual v3.0, <http://www.mldesigner.com>

[10] Tom Dengler, Horst Salzwedel, Optimierung von komplexen Entwicklungsprozessen mittels simulationsgestützter Prozessanalyse, to be presented at 52th International Scientific Colloquium, Ilmenau, September 10-12, 2007.

[11] Tommy Baumann, Horst Salzwedel, Mapping of Electronic System Level (ESL) Models Into Implementation, *DATE'07*, Acropolis, Nice, France, April 16-20, 2007.

[12] Nils Fischer, *Design of a Plug-and-Play Development Environment for Optimizing Avionics Systems Architectures*, Thesis, Airbus/Technical University Ilmenau, July 2007.

[13] Stefan Riehmer, *Standardizing Data Storage for MLDesigner and Octave using the Open Office Document Standard*, Thesis, Technical University Ilmenau, August 2007.

[14] Gene Franklin, David Powell, Abbas Emami-Naeini, *Feedback Control of Dynamic Systems*, Pearson Prentice Hall, New Jersey 2006.

AUTHOR:

Horst Salzwedel, MLDesign Technologies, Inc., horst@mldesigner.com

G. J. Nalepa / I. Wojnicki

Filling the Semantic Gaps in Systems Engineering

Abstract: The paper deals with some common problems present in knowledge and software engineering, related to the practical design, analysis, and implementation of systems. With different design methods used at subsequent design stages, the so-called semantic gaps appear, due to important differences in semantics between design methods. The paper discusses the semantic gaps present in software and knowledge engineering. In order to fill them it discusses a formal design method, based on the XTT knowledge representation.

Introduction

The paper deals with some common problems present in both knowledge and software engineering. These problems are related to the practical design, analysis, and implementation of systems in these domains. With different design methods used at subsequent design stages, the so-called *semantic gaps* appear. They are related to important differences in semantics between design methods. The best example is the problem with translating specification requirements, into a UML model, and then transforming it into object-oriented code. In order to cope with these problems some computer tools are used. However, these tools cannot solve these problems without proper formal foundations. This is why, the research in the field of formal methods in systems engineering is still active. In order to fill the semantic gaps while providing a bridge between these domains, the paper discusses a formal design method, based on the XTT knowledge representation. The method supports a hierarchical design, implementation, and on-line evaluation of systems.

Selected Issues of Software Engineering

Software engineering (SE) is a domain where a number of mature and well-proved design methods and approaches exist [1]. However, number of critical problems with efficient and integrated design and implementation of complex software persist. It will be argued, that sources of errors in software engineering are:

- *The Semantic Gap* between existing design methods, which are becoming more and more declarative, and implementation tools that remain sequential/procedural. This issue results in the problems mentioned below.
- *Evaluation problems* due to semantical differences of design methods and lack of formal knowledge model. They appear at many stages of the SE process, including the final software correctness, the validity of the design model, and the transformation from the model to the implementation.
- The so-called *Analysis Specification Gap*, which is the difficulty with proper formulation of requirements, and transformation of the requirements into an effective design, and then implementation.
- The so-called *Separation Problem*, which is the lack of separation between Core Software Logic, software interfaces and presentation layers.

The Software Engineering is derived as a set of paradigms, procedures,

specifications and tools from pure programming, which is coding. Historically, when the modeled systems became more complex, SE became more and more declarative, in order to model the system in a more comprehensive way. It made the design stage independent of programming languages which resulted in number of approaches. So, while programming itself remains mostly sequential, designing becomes more declarative. The introduction of object-oriented programming does not change the situation drastically.

In software engineering the software development process and life cycle is represented by several models [1]. In this process systems analysts try to model the structure of the real-world information system in the structure of computer software system. So the structure of the software corresponds to some respect to the structure of the real-world system. The task of the programmers is to encode and implement the model in some lower-level programming language.

UML approach identifies two distant domains of Software Engineering [2]. One of them is modeling software structure the other is modeling its behavior. There are two classes of diagrams then: Structure Diagrams and Behavior Diagrams containing different types of diagrams. *Structure Diagrams* model software structure, and comply with object-oriented software engineering. It seems that Structure Diagrams are the UML basis. They are fairly complete and allow for expressing software components and denoting relationship among them easily (i.e.: Class Diagram, Component Diagram etc.). *Behavior diagrams* model software logic. It is modeled at different abstraction levels. There is a big picture perspective: modeling what particular software should do, from the user point of view (i.e.: Use Case Diagram). There is also a detailed perspective: what particular software components defined by the Structure Diagrams should do (i.e.: State Machine Diagram, etc.).

A typical software design process based on UML consists of the following stages: general behavior modeling (use cases), structure modeling, behavior and interaction modeling. The general behavior modeling describes what the system should do in the most general terms. The second stage which is structure modeling tries to describe what the system will consist of, using class diagrams mostly. But the practice indicates, that the process is in fact in most cases the know-how of the users. The fact is that, UML is only a language suitable for software design but it does not offer a design process. The process is somehow hidden, and only the final result is visible. This can be partially fixed with the methodologies such as the MDA.

Since there is no direct bridge between declarative design and sequential implementation, a substantial work is needed in order to turn a design into a running application. This problem is often referred to as a *Semantic Gap* between a design and its implementation [3]. It is worth noting, that while the conceptual design can sometimes be partially formally analyzed and evaluated, the full formal analysis is impossible in most cases. However, there is no way to assure, that even fully formally correct model, would translate to a correct code in a programming language. What is even worse, if an application is automatically generated from a designed conceptual model, then any changes in the generated code have to be synchronized with the design.

There is also another gap in the specification-design-implementation process called *Analysis Specification Gap*. It regards the difficulty with the transition from the specification to the design. Formulating a specification which is clear, concise, complete and amenable to analysis turns out to be a very complex task, even in small scale projects.

Executable Design Concept

The *executable design* concept (ED) aims at solving the main problems outlined previously. The concept itself is not new, and can be considered one of the “holy grails” of systems engineering. The main goal of this concept is to avoid semantic gaps, mainly the gap between the design and the implementation [3]. In order to do so, the following elements should be developed: a rich and expressive design method, a high-level runtime environment, and an effective design process. A full ED method should eventually shorten the development time, improve software quality, provide a design-once-run-everywhere solution, transform the “implementation” into the runtime-integration.

The development of an ED has been approached on several fronts, namely: the implementation front, with the development of new, experimental languages; as well as on the design front, with new design approaches; with a lot of recent development in the area of advanced runtimes, including virtual machines.

From the ED perspective, in the domain of software design there are at least two interesting developments. The first one concerns the extension of UML into Executable UML (xUML) with action semantics, see [3] for more details. The principal idea is to fill in gaps present in UML, in order to offer a translation from an UML specification into an executable prototype. However, it must be pointed out the current state of the xUML is unclear, and its applications limited.

Another very important and influential concept concerns the so-called design patterns [4]. The idea is to identify certain patterns on the design level, and use them as the foundation for future design. The patterns are usually identified in the object-oriented paradigm. What is important, common patterns nowadays have practical implementations in the programming environments such as Java. So they are not only used to speedup and simplify the design, but also for providing a kind of ED.

There are a few assumptions and observations regarding ED. Since the software design process is declarative, its result, an application, is declarative as well (not counting interactions with existing non-declarative components, user interface, operating system etc.). This implies that execution of a declarative application must be provided through a declarative or at least partially declarative languages, including functional programming ones. Common choices are: Lisp, Prolog and Haskell. Moreover such an approach allows to formally analyze the designed application by the same runtime environment which runs it. It reduces number of software components implementing the runtime technology.

Finding a Bridge with Knowledge Engineering

What makes *knowledge-based systems* (KBS) distinctive is the separation of knowledge base from the knowledge processing facilities [5,6]. In order to store knowledge, KBS use various knowledge representation methods, which are *declarative* in nature. In case of rule-based systems (RBS) these are *rules*. Specific knowledge processing facilities, suitable for particular representation method being used, are selected then. In case of RBS these are logic-based inference engines.

What is important about the knowledge engineering process, is the fact that it should capture the expert knowledge and represent it in a way that is suitable for processing (this is the task for a knowledge engineer). The actual structure of a KBS does not need to be system specific - it should not “mimic” or model the structure of the real-world problem. However, the KBS should capture and contain knowledge regarding the real-world system. It should be pointed out, that in case of KBS there is no single universal engineering approach, or universal modeling method (such as

UML in SE). Different classes of KBS may require specific approaches.

It is worth considering how the standard SE language, UML, can be used to help build KBS. There are several possible approaches when it comes to practical UML application for knowledge engineering:

- Model system with a knowledge-based approach, that is use some classic knowledge representation method, such as decision trees, then design the software implementation using UML, and generate an object-oriented code.
- Model rule-based knowledge with UML diagrams, and then generate the corresponding OO code.
- Incorporate a complete rule-based logic core into an OO application, implementing I/O interfaces, including presentation layer, in an OO language.

The first solution is a “classic” and definitely the easiest one. It can be found in number of tools and approaches. In this case KE methods are used in the “design” stage, while SE methods provide “implementation” means (UML is somehow used to design the implementation previously designed with KE methods). But the fact is it can be considered the worst solution, since it exposes the semantic gap.

The second approach relies on either extending, or redefining the original semantics of UML. Some early beginning can be observed in OMG Production Rule Representation. However, a complete example of this approach may be found in the Unified Rule Modelling Language (URML) [7]. In this case existing UML diagrams are used to model different type of rules.

The last one is possibly the most complicated approach. It relies on the incorporation of the knowledge-based component into an OO application in a way that minimizes the semantic gap between SE and KE. This is the solution visible, to some extent, in the business rules approach. A similar, but more complete solution is being developed in the Hekate project, where a declarative, rule-based core is integrated into an OO application as a logical model (as in the MVC design pattern).

There are some general observations regarding the usability of UML. The syntax seems to be well defined; however, in some cases the semantics is not. One of the limitations of UML is its heavy dependability on the concept of object. This concept may be fundamental for OO languages, but it is of marginal importance for AI. The limitations of semantics are in some cases decreased with the use of UML profiles. However the problem is that in some cases profiles can totally redefine the original semantics, rendering its relation with the syntax nonexistent.

A problem is that two perspectives provided by UML (Structure and Behavior diagrams) do not mix well. While the detailed perspective corresponds to classes, the big picture one serves more as a guideline than a real modeling tool. What is worse some of the Behavior Diagrams share common functionality and judging which one to use is not clear quite often.

The semantic gap problem is the most important one from the software manufacturing perspective. Even if diagrams support the implementation process by describing software in a comprehensive way, it cannot be validated in reasonable time if the implementation matches the design. At some point there are structural diagrams which describe what the system consists of, and behavioral diagrams, describing how the system should work, and finally the implementation which consists of the designed structure and is believed to behave accordingly. It is worth noting, that while the behavior design can sometimes be partially formally analyzed and evaluated, a formal analysis of the implementation is impossible in most cases. There have been some substantial work on automating the transition from design to implementation, however none of these approaches solves the problem.

Hybrid Knowledge Engineering Methodology

The Hekate project aims at addressing the problems described previously. It is based on experiences with the Mirella project [8]. The main goal of that project was to fully develop and refine the integrated design process for RBS. The integrated design process proposed in Mirella can be considered a top-down hierarchical design methodology, based on the idea of meta-level approach to the design process. It includes three phases: conceptual, logical, and physical. It provides a clear separation of logical and physical (implementation) design phases. It offers equivalence of logical design specification and prototype implementation, and employs XTT, a hybrid knowledge representation.

Hekate [9] aims at extending Mirella's RBS perspective towards general SE. A principal idea in this approach is to model, represent, and store the logic behind the software (sometimes referred to as business logic) using advanced knowledge representation methods taken from KE. The logic is then encoded with use of a Prolog-based representation. The logical, Prolog-based core (the logic core) would be then embedded into a business application, or embedded control system. The remaining parts of the business or control applications, such as interfaces, or presentation aspects, would be developed with a classic object-oriented or procedural programming languages. The Hekate project should eventually provide a coherent runtime environment for running the combined Prolog and Java/C code.

The main idea behind XTT knowledge representation and design method aims at combining some of the existing approaches, namely decision trees and decision tables. It allows for a hierarchical visual representation of the decision tables linked into tree-like structure, according to the control specification provided. XTT, as a design and knowledge representation method, offers transparent, high density knowledge representation as well as a formally defined *logical*, Prolog-based interpretation, while preserving flexibility with respect to knowledge manipulation.

In Hekate, the knowledge base design process and knowledge visualization is derived from the XTT methodology. The XTT methodology is currently being extended towards covering not only forward and backward chaining RBS but also control applications, databases and general purpose software. From the implementation point of view Hekate is based on the idea of multiparadigm programming. The target application combines the logic core implemented in Prolog, with object-oriented interfaces in Java, or procedural in ANSI C. This is possible due to the existence of advanced interfaces between Prolog and other languages. Most of the contemporary Prolog implementations have well developed ANSI C interfaces. There is also a number of Object-Oriented interfaces and extensions in Prolog. The best example is LogTalk (www.logtalk.org).

In Hekate, the Semantic Gap problem is addressed by providing declarative design methods for the business logic. There is no translation from the formal, declarative design into the implementation language. The knowledge base is specified and encoded in the Prolog language. The logical design which specifies the knowledge base becomes an application executable by a runtime environment, combining an inference engine and classic language runtime (e.g. a JVM).

At the starting point for solving the Analysis Specification Gap problem the ARD design method is used. In Hekate Advanced Relationship Diagrams allow to specify components of the system and dependencies among them at different levels of detail. It allows to design software in a top-down fashion: starting from a very general idea what is to be designed, and going into more and more details about each single quantum of knowledge which refers to the system.

The executable design concept used in Hekate is based on ARD/XTT concept. ARD is used to describe dependencies in the knowledge base on different

abstraction levels, while XTT represents the actual knowledge. The design process starts with an ARD model at a very general level, which is developed to be more and more specific. The nature of knowledge dependencies, facts and rules, are encoded with XTT. An application model based on combined XTT and ARD, along with interfaces and views, becomes the Application executed by the Hekate (Hekate Run-Time), an inference engine supported with optional sequential (C/Java) runtime.

Two main approaches to provide an effective runtime environment for Hekate have been considered. The first one consists in generating native code in some classic object-oriented language such as Java. This solves both the practical implementation as well as runtime problem. This solution is used in products such as JBoss Rules (formerly Drools). However, it does have a major drawback: the object-oriented semantics is very distant from the declarative rule semantics of XTT. This instantly unveils a semantic gap which turns out to be a major limitation during the implementation and testing of the system.

The second approach is based on using a high-level Prolog representation of XTT. Prolog semantics includes all of the concepts present in XTT. Prolog has the advantages of flexible symbolic representation, as well as advanced meta-programming facilities. The Hekate-in-Prolog solution is based on the XTT implementation in Prolog. In this case a term-based representation is used, with an advanced meta interpreter engine provided.

Concluding Remarks

The paper discusses selected important issues present in systems engineering in the software engineering domains. These issues are commonly described as *semantic gaps* in the design process. They are responsible for making the design more complex and fragile to errors. In the paper a proposal for solving these problems is given. It is based on the idea of using knowledge engineering methods in software design. Custom representation method, the XTT, developed in the Mirella project, is the foundation of the Hekate project. Hekate aims at providing efficient AI tools for software engineering, basing on the extension of the executable design concept.

References:

- [1] Sommerville I., *Software Engineering*, Pearson Education, 2004.
- [2] Fowler M., *UML Distilled: A Brief Guide to the Standard Object Modelling Language*, Addison-Wesley, 2003.
- [3] Mellor S., Balcer M., *Executable UML: A Foundation for Model Driven Architecture*, Addison-Wesley, 2002.
- [4] Gamma E., et al., *Design Patterns*, Addison-Wesley, 1995.
- [5] Torsun I. S., *Foundations of Intelligent Knowledge-Based Systems*, Academic Press, 1995.
- [6] Ligeza A., *Logical Foundations for Rule-Based Systems*, Springer-Verlag, 2006.
- [7] Lukichev S., Wagner G., *Visual Rules Modeling*, 6th International Conference Perspectives of Systems Informatics, Springer LNCS, 2006.
- [8] Nalepa G. J., Ligeza A., *A Visual Edition Tool for Design and Verification of Knowledge in Rule-Based Systems*, *Systems Science*, vol. 31, num. 3, pp. 103-109., 2005.
- [9] Nalepa G. J., Wojnicki I., *A proposal of Hybrid Knowledge Engineering and Refinement Approach*, 20th FLAIRS Conference Key West Florida, AAAI Press 2007.

Authors:

Ph. D. Eng. Grzegorz J. Nalepa
Ph. D. Eng. Igor Wojnicki
AGH University of Science and Technology
Al. Mickiewicza 30, 30-059, Krakow
Phone: 48-12-617-3856
Fax: 48-12-634-15-68
E-mail: gjn@agh.edu.pl, wojnicki@agh.edu.pl

Acknowledgements

The paper is supported by the Hekate project funded from 2007-9 resources for science.

Compiling Experience into Knowledge

Abstract

TYPICAL application fields of Knowledge Based Systems are usually characterized by having human expertise as the only one source to specify their desired behavior. Therefore, their design, evaluation and refinement has to make effective use of this valuable source. After an introduction to the concept of collecting validation experience in a Validation Knowledge Base (VKB), the paper introduces an estimation of the significance of the cases collected in the VKB. A high significance signalizes that a VKB should not longer serve as a case-based source of external (outside the Knowledge Base) knowledge, but compiled into the Knowledge Base instead. Based on this significance estimation, a technology to compile well selected cases into the Knowledge Base of the system under evaluation is presented.

1 Introduction

The purpose of refinement approaches is to adjust a system according to new insights. Such insights may be uncovered by detecting invalidities when applying a validation technology, but also by gaining experience when considering cases. In particular, they aim at a knowledge base reconstruction so that the input–output behavior and, in some approaches, the complete rule trace is adapted to the new insights.

There is a history of attempts to face this problem for rule–based systems. A quite comprehensive digest about pros and cons of systems like TEIRESIAS [1], SEEK/SEEK2 [2, 4], the *Reduced Theory Learning System* RTLS [3] with a so–called relaxed retranslation that has been criticized in [13] and [16], and KRUSTWORKS [15], e.g., is provided in [5].

Besides particular individual drawbacks, these approaches share the property that they can't produce a rule base which is 100 % correct. Correctness, in this context, means correctness w.r.t. a set of test cases. Furthermore, they may cover some inherent anomalies and may not be interpretable by topical human experts.

AI systems' design and maintenance heavily depends on the quality of the human expertise and effectiveness of its involvement in the system's design, evaluation, and refinement. For their validation and refinement, the authors introduced a case based technology [6, 7].

To make validation results less dependent on the experts' opinions and to decrease the workload of the experts, the authors developed a concept to collect case oriented experience in a Validation Knowledge Base (VKB) [8, 10].

These concepts have been involved in a validation framework [7]. To estimate the usefulness of these concepts and to reveal their weaknesses, a prototype test was performed [9].

The VKB concept so far utilizes the external knowledge in a VKB as an additional source of knowledge for system validation. It is organized in a case-based manner. However, at some point a VKB should not longer serve as a case-based source of external (outside the Knowledge Base) knowledge, but compiled into the Knowledge Base instead. If a VKB content turns out to be accepted by topical experts, it should be utilized to refine the Knowledge Base to become a source of knowledge for the system itself, not just for its evaluation. Especially the cases, which gained a high agreement in the expert community

over a long period of time are worth to be included into the Knowledge Base. A technology to do so is introduced here.

The paper is organized as follows. Section two briefly introduces the VKB concept. In section three, the utilization of VKB for system refinement is outlined. Section four presents a technology to compile the cases of VKB into the rules of the Knowledge Base. The paper is finalized by a summary.

2 Collecting experience: The VKB approach so far

In spite of significant advances in recent years, validation of knowledge-based systems still requires significant involvement on the part of human validators. In contrast to verification, which seeks to assure compliance with specifications and the absence of specific errors without executing the system, validation typically involves rigorous and often extensive testing of the system.

The results of these tests are nearly always evaluated by expert validators who may not always agree among themselves. The size of the test case set, the frequency of the validation exercises and the number of expert validators required for each such exercise can combine to pose great burdens of time and effort on human expert validators. These expert validators are a scarce resource, have limited time, and are expensive to employ. These limitations have the potential to degrade a validation exercise.

Our concepts of a Validation Knowledge Base was originally proposed by TSURUTA et al. [14]. TSURUTA's work appears to be the first to specifically address the use of prior validation knowledge for improving the validation process. His work aimed at developing validation solutions for commercial applications, and he addresses this issue frequently [12].

Our work is implemented in the context of our previously described validation framework [7], which includes an expert validator review of test cases and results using a variation of a TURING Test for the validation of knowledge-based systems. In this step, humans play the role of expert validators as part of a validation panel. Their task is (1) to solve the test cases posed to the system under evaluation and (2) to review and provide their judgment on the correctness (the ratings) of all anonymous solutions (the system's as well as the panel's own).

To improve the validation process, the validation knowledge used in prior exercises, namely the set of test cases (the test inputs and the best rated solutions) along with their authors, must persist from one validation exercise to the next. This is effectively accomplished by the VKB.

The VKB and its historical validation knowledge can also significantly reduce the involvement of expert validators by eliminating their need to solve old test cases whose solutions are already found in the VKB. The expert validator panel needs only to solve new test cases created by the Validation Framework that are not already part of the VKB, because the VKB already have (formerly accepted) solutions available. However, they still have rate all solutions, because the former solution needs to be either confirmed or revised by the current expert panel to further "qualify" the VKB.

The VKB is a set of previous (historical) test cases and their best rated solutions, which can be described by 7-tuples $[t_j, E_{Kj}, E_{Ij}, sol_{Kj}^{opt}, r_{IjK}, c_{IjK}, \tau]$, in which

- t_j is a test case input,
- sol_{Kj}^{opt} is a solution associated to t_j , which gained the maximum experts' approval in a prior validation exercise,
- E_{Kj} is the list of experts who provided this particular solution,

- E_{I_j} is a list of experts who rated this solution,
- r_{I_jK} is the rating of this solution, which is provided by the experts in E_I ,
- c_{I_jK} is the certainty of this rating, and
- τ is a time stamp associated with the validation session in which the rating was provided.

Table 1 shows how the VKB would appear for a simple application. Here, e_1 , e_2 and e_3 are specific human expert validators. The outputs o_1, \dots, o_{25} are the test case outputs (solutions). The time stamps are denoted by natural numbers to indicate an unspecified time when the validation exercise was held in the right sequence.

The VKB is initially built as part of the first validation exercise. Here, each test case input used in the exercise, along with its optimal solution (as determined by the panel during that exercise), becomes a new entry. It is updated in subsequent validation exercises by adding all examined

Table 1: An example for *VKB*'s entries

t_j	E_K	E_I	sol_{Kj}^{opt}	r_{ijk}	c_{ijk}	τ_S
t_1	e_1, e_3	$[e_1, e_2, e_3]$	o_6	$[1, 0, 1]$	$[0, 1, 1]$	1
t_1	e_2	$[e_1, e_2, e_3]$	o_{17}	$[0, 1, 0]$	$[1, 1, 1]$	4
t_2	e_1, e_3	$[e_1, e_2, e_3]$	o_7	$[0, 0, 1]$	$[0, 0, 1]$	1

test cases of this session. New entries, however, do not supersede old entries. Instead, the 'updated information' is represented by the new entries with a more recent time stamp.

The VKB functions in the second step, the test case experimentation. In the original approach, the test case generation procedure consists of two steps (1) generating a quasi exhaustive set of test cases *QuEST* and (2) reducing it down to a reasonably sized set of test cases *ReST* [7]. A test case is a pair $[TestData, Solution]$. Both *QuEST* and *ReST* are sets of such pairs. Exactly between these two sub-steps is the entry-point of the external validation knowledge stored in a VKB that has been constructed in prior validation sessions.

Both *QuEST* and the historical cases in VKB are subjected to the criteria-based reduction procedure that aims to build a subset of test cases in *QuEST* or VKB. The cases in VKB are included in the reduction process to (1) ensure that they meet the requirements of the current application and (2) their number is small enough to be the subject of the time consuming and expensive test case experimentation.

The VKB, therefore is a database of test cases and their associated solutions that received an optimal rating in previous validation sessions. These solutions are considered an additional (external) source of expertise that did not explicitly appear in the solving session, but it is a subject of the rating session.

Regardless of their former ratings, the cases originated from the VKB have to be rated by the current expert panel in the current session.

The set of solutions $ExtSol \subseteq \Pi_2(ReST)$, which are contributed by the VKB and which are subject of the rating process, is¹

$$ExtSol := \{sol : \exists Entry : Entry \in VKB, \Pi_1(Entry) \in \Pi_1(ReST), sol = \Pi_4(Entry)\}$$

Because the criteria-based reduction process is controlled by a predetermined number m of cases that form the *ReST*, the **workload reduction factor** for the test case solving process for the expert validators can be quantified by the cardinality of *ExtSol* divided by the cardinality of *ReST*: *workload reduction factor* = $\frac{|ExtSol|}{|ReST|}$

¹ Π_i is the projection of the i -th element, i.e. $\Pi_i(T)$ with T being a tuple denotes the i -th element of the tuple and $\Pi_i(M)$ with M being a set of tuples denotes the set of i -th elements of the tuples in M .

The best rated solutions associated with the test cases in the VKB represent an additional (external) source of expertise. It is different from both the current expert validators' and the system's knowledge. The **expertise gain factor** introduced by the VKB is: $expertise\ gain\ factor = \frac{|ReST|}{(|ReST| - |ExtSol|)}$

The usefulness of the VKB approach could be proven by an experiment with human experts. Starting with an initial rule base, validation sessions as described in [6] and [7] have been performed and a VKB has been built. [11] and [9] could impressively show the VKB's contribution to the validation knowledge in this experimental case study.

3 Evaluating experience: Estimating its significance by a metrics

In case the VKB contains cases with (1) an optimal solution that is different from the system's solution and (2) a certain significance that the optimal solution in VKB is correct, a system refinement based on the VKB's content is indicated.

Depending on the particular application, a minimum number n^{min} of entries for a test case input t_j and a minimum significance level $0 \ll sig^{min} \leq 1$ needs to be determined.

Indications for the correctness of an optimal solution sol_{Kj}^{opt} to a case input t_j in VKB are high values of the following metrics

- approval rate $app = \frac{\# \text{ of positive ratings } r_{ijk} = 1}{\text{total } \# \text{ of ratings to } t_j}$
- persistence rate $per = \frac{\# \text{ of entries for } t_j \text{ with } sol_{Kj}^{opt}}{\text{total } \# \text{ of entries to } t_j}$
- agreement rate $agr = \frac{\# \text{ of experts providing } r_{ijk} = 1}{\text{total } \# \text{ of experts providing a } r_{ijk}}$

In case (1) the minimum number n^{min} of entries for a test case input t_j is reached and (2) all three of the above rates exceed the minimum significance level sig^{min} , the pair $[t_j, sol_{Kj}^{opt}]$ is worth to be compiled into the Knowledge Base as its input/output behavior for t_j . This is performed by a technology similar to the refinement technology introduced in [7] and described in the following section.

After compiling a case into the Knowledge Base, its usage as a validation test case becomes obsolete. Therefore, the related entries in VKB need to be removed.

4 Utilizing experience: Compiling it into knowledge

The refinement procedure looks for rule chains of cases in the VKB, which have a different solution in VKB than the rule chain ends up with. It starts with the last rule in this chain and analyses *all* VKB cases using this rule. It systematically constructs new rules as a substitute of it, which map the cases that have a different solution to this different solution and keeps the mapping of all other cases (at all, not only those in VKB) as it was before the refinement. So it is pretty "conservative", because it changes the I/O behavior of the rule as few as possible, i.e. exclusively for cases that are shown to be solved wrong by the rule base. The technique is applicable to rule bases as introduced in [7] and consists of the following steps.

Identifying "Guilty Rules" If the last rule r_l in the rule trace for a case input $t_j \in \Pi_{inp}(T)$ infers a solution different from $sol_{Kj}^{opt} \in \Pi_{outp}(T)$, this rule r_l is "guilty" and therefore, subject of the following refinement technology. Let $T_l \subseteq T$ be the set of cases that have r_l as their last rule in the rule traces for the cases in T .

Table 2: Reduction rules to construct substitutes for an invalid rule

R1	<p>$pos \in Pos$, s_{pos} has a value set with no \leq relation, $\{s_{pos}^1, \dots, s_{pos}^m\}$ are the values of s_{pos} occurring in T_l^s</p> <div style="border: 1px solid black; padding: 5px; margin: 5px 0;"> $[T_l^s, Pos, \{p_1, \dots, p_n\}]$ \hookrightarrow </div> <p>1. $[T_l^{s,1} \setminus \{[t_j, sol_s] \in T_l^s : s_{pos} \neq s_{pos}^1\}, Pos \setminus \{pos\}, \bigcup_{i=1}^m p_i \cup \{(s_{pos} = s_{pos}^1)\}]$ \dots</p> <p>m. $[T_l^{s,m} \setminus \{[t_j, sol_s] \in T_l^s : s_{pos} \neq s_{pos}^m\}, Pos \setminus \{pos\}, \bigcup_{i=1}^m p_i \cup \{(s_{pos} = s_{pos}^m)\}]$</p>
Continue with each $T_l^{s,i}$ ($1 \leq i \leq m$) separately.	
R2	<p>$pos \in Pos$, s_{pos} has a value set with a \leq-relation, $s_{pos}^{min} / s_{pos}^{max}$ are the smallest / largest value of s_{pos} within T_l^s</p> <div style="border: 1px solid black; padding: 5px; margin: 5px 0;"> $[T_l^s, Pos, \{p_1, \dots, p_n\}]$ \hookrightarrow $[T_l^s, Pos \setminus \{pos\}, \bigcup_{i=1}^m p_i \cup \{(s_{pos} \geq s_{pos}^{min}), (s_{pos} \leq s_{pos}^{max})\} \cup S_{excl}]$ </div> <p>S_{excl} is the set of excluded values for s_{pos}, which have to be mapped to a solution different from sol_s because of belonging to some other T_u^v with $v \neq s$: $S_{excl} = \{(s_{pos} \neq s_{pos}^j) : \exists [t_j, sol_s] \in T_l^s \exists [t_m, sol_v] \in T_u^v (v \neq s) \text{ with } \forall p \neq pos : s_p^j = s_p^m, s_{pos}^{min} < s_{pos}^m < s_{pos}^{max}\}$</p>

Simple Refinement by Conclusion Replacement If all cases $t_j \in T_l$ have the same solution sol_{Kj}^{opt} , in rule r_l the conclusion part is substituted by sol_{Kj}^{opt} .

Reconstructing the Remaining Guilty Rules The remaining guilty rules are used by a set of cases T_l , which have different optimal solutions. The subsets with the same optimal solution are considered separately:

1. T_l of the rule r_l is split into subsets T_l^s ($1 \leq s \leq n$) according to the n different solutions $sol_{Kj}^{opt,1}, \dots, sol_{Kj}^{opt,n}$ for the cases $t_j \in T_l$.

The if-part(s) of the new rule(s) that substitute r_l are expressions $e_i \in E$ of a set of p new alternative rules $\{r_l^1, r_l^2, \dots, r_l^p\}$ for each T_l^s and will be noted as a set of sets $P_l^s = \{\{e_1^1, \dots, e_{p_1}^1\}, \dots, \{e_1^p, \dots, e_{p_p}^p\}\}$. The corresponding rule set of P_l^s is

$$r_l^1 : \bigwedge_{i=1}^{p_1} e_i^1 \rightarrow sol_s \quad \dots \quad r_l^p : \bigwedge_{i=1}^{p_p} e_i^p \rightarrow sol_s$$

2. Pos is the set of Positions (dimensions of the input space), at which the input data $t_j \in \Pi_{inp}(T_l^s)$ of the test cases $t_j \in T_l^s$ are not identical.

The generation of the p different if-parts in P_l^s is managed by a formal *reduction system*, which is applied to triples $[T_l^s, Pos, P_l^s]$ until Pos becomes the empty set \emptyset .

3. The initial situation is $[T_l^s, Pos, P_l^s]$ with $P_l^s = \{\{(s_1 = s_1^{ident}), \dots, (s_q = s_q^{ident})\}\}$
 s_1, \dots, s_q are those positions, at which all test data $t_j \in \Pi_{inp}(T_l^s)$ have the same value s_i^{ident} . Initially, P_l^s stands for just one rule: $r_l^1 : \bigwedge_{i=1}^q (s_i = s_i^{ident}) \rightarrow sol_{Kj}^{opt,s}$
4. The reduction terminates with the situation $[T_l^s, \emptyset, P_l^s]$.

Table 2 shows the reduction rules applied to the triples. In [7] it is shown, that the reduction system is terminating, complete, and correct.

Recompiling the constructed rules The new rules generated so far are “one-shot-rules”, i.e. they infer directly from a system’s input to a system’s output. These rules might be difficult to read, because they may have very long *if*-parts, and difficult to interpret by subject matter experts. This problem can be defused by introducing the intermediate hypotheses into the computed new rules.

5 Summary

The formerly developed concept of a Validation Knowledge Base (VKB) was intended to model collective best experience of several human experts. The VKB is constructed and maintained across various validation exercises. If the knowledge gained in a VKB turns out to be well accepted by the expert community over a long period of time, this knowledge is worth to be compiled into the system’s Knowledge Base. This way, the knowledge dedicated to evaluate a system shifts to knowledge used to improve the system. Therefore, the paper introduced a technology to compile the case based knowledge of a VKB into the rule based knowledge of the system’s Knowledge Base.

References

- [1] Davis, R. and Lenat, D.B.: *Knowledge-Based Systems in Artificial Intelligence*. New York: McGraw Hill Int. Book Company, 1982.
- [2] Ginsberg, A.: *Automatic Refinement of Expert System Knowledge Bases*. London: Pitman Publishing, 1988.
- [3] Ginsberg, A.: Theory Reduction, Theory Revision, and Retranslation. *Proc. of the 9th National Conference on Artificial Intelligence (AAAI-90)*, Boston, MA, pp. 777–782, 1990.
- [4] Ginsberg, A., Weiss, S.M., and Politakis, P.: Automatic Knowledge Base Refinement for Classification Systems. *Artificial Intelligence*, 35(1988), pp. 197–226, 1988.
- [5] Kelbassa, H.-W. and Knauf, R.: A Process Approach to Rule Base Validation and Refinement. *Proc. of the Workshop on Knowledge Engineering and Software Engineering (KESE) at the 28th German Conf. on AI (KI 2005)*, University of Koblenz-Landau (2005), Fachberichte Informatik, ISSN 1860-4471, pp. 25–36, 2005.
- [6] Knauf, R.: *Validating rule-based systems – A complete methodology*. Habilitation Thesis (2000), Technical University Ilmenau, Germany. Aachen (Germany): Shaker, ISBN 3-8265-8293-4, 2000.
- [7] Knauf, R., Gonzalez, A.J., and Abel, T.: A Framework for Validation of Rule-Based Systems. *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, vol. 32, # 3, pp. 281–295, 2002.
- [8] Knauf, R., Tsuruta, S., Ihara, H., Gonzalez, A.J., and Kurbad, T. (2004): Improving AI Systems’ Dependability by Utilizing Historical Knowledge. *Proc. of 10th International Symposium Pacific Rim Dependable Computing (PRDC)*, Papeete, Tahiti, French Polynesia, pp. 343–352, IEEE Computer Society Press, 2004.
- [9] Knauf, R., Tsuruta, S., and Gonzalez, A.J. (2005): Towards Modeling Human Expertise: An Empirical Case Study. *Proc. of the 18th International Florida Artificial Intelligence Research Society Conference 2005 (FLAIRS-2005)*, Clearwater Beach, FL, USA, pp. 232–237, ISBN 1-57735-200-9, Menlo Park, CA: AAAI Press, 2005.
- [10] Knauf, R., Tsuruta, S., and Gonzalez, A.J. (2005): Overcoming Human Weaknesses in Validation of Knowledge-Based Systems. Jantke / Faehnrich / Wittig (eds.): *Marktplatz Internet: Von e-Learning bis e-Payment. 13th Leipzig Informatic Days 2005 (LIT 2005)*, Lecture Notes in Informatics (LNI) – Proceedings, Series of the Gesellschaft fuer Informatik (GI), vol. P-72, pp. 254–263, ISBN 3-88579-401-2, ISSN 1617-5468, 2005.
- [11] Knauf, R., Tsuruta, S., Uehara, K., Gonzalez, A.J. (2004) Validation knowledge bases and validation expert software agents : models of collective and individual human expertise. Technical Report # 01/04. Tokyo Denki University, School of Information Environment, Chiba, Japan, also available at <http://www.db-thueringen.de/servelets/DocumentServlet?id=4285>, 2004.
- [12] Onoyama, T. and Tsuruta, S.: Validation method for intelligent systems. *Journal of experimental and theoretical Artificial Intelligence*, vol. 12, # 4, pp. 461–472, 2000.
- [13] Liang, C.X.F. and Valtorta, M.: Refinement of Uncertain Rule Bases via Reduction. *International Journal of Approximate Reasoning*. 13(1995), pp. 95–126, 1995.
- [14] Tsuruta, S., Onoyama, T., Kubota, S., and Oyanagi, K.: Validation method for intelligent systems, *Proc. of the 13th International Florida Artificial Intelligence Research Society Conference (FLAIRS-2000)*, pp. 361–365, Menlo Park, CA, AAAI press, 2000.
- [15] Wiratunga, N., and Susan Craw, S.: Incorporating Backtracking in Knowledge Refinement. Vermesan / Coenen (eds): *Validation and Verification of Knowledge Based Systems – Theory, Tools and Practice. Proc. of the 5th Europ. Conf. on Validation and Verification (EUROVAV 99)*, pp. 193–205, Boston, MA Kluwer Academic Publishers, 1999.
- [16] Zlatareva, N.P., Preece, A.D.: State of the Art in Automated Validation of Knowledge-Based Systems. *Expert Systems with Applications*, 7(1994), pp. 151–167, 1994.

Author Information:

Priv.-doz. Dr.-Ing. habil. **Rainer Knauf**

Technical University of Ilmenau, Dept. of Computer Science and Automation, Chair of AI
P.O. Box 10 05 65, 98684 Ilmenau, Germany

Tel: +49 3677 69 1445, Fax: +49 3677 69 1665, E-mail: rainer.knauf@tu-ilmenau.de

R. Knauf / S. Tsuruta / Y. Sakurai

Toward Knowledge Engineering with Didactic Knowledge

Abstract

LEARNING systems suffer from a lack of an explicit and adaptable didactic design. A way to overcome such deficiencies is (semi-) formally representing the didactic design. A modeling approach, storyboarding, is outlined here. Storyboarding is setting the stage to apply Knowledge Engineering Technologies to verify, validate the didactics behind a learning process. As a vision, didactics can be refined according to revealed weaknesses and proven excellence. Furthermore, successful didactic patterns can be inductively inferred by analyzing the particular knowledge processing and its alleged contribution to learning success.

1 Introduction

Successful university instructors are often not those with the very best scientific background or outstanding research results. The most successful ones are typically those that successfully utilize didactical experiences as well as "soft skills" in dealing with other actors in the teaching process. Besides the students and colleagues, such actors include e-learning systems as well as the large amount of active (desirable and undesirable, conscious and unconscious) "content presenters" that include news, web sources and advertisements.

The design of learning activities in collegiate instruction is a very interdisciplinary process. Besides deep, topical knowledge in the subject being thought, an instructor needs knowledge and skills in many other subjects. This includes IT-related skills to use today's presentation equipment, didactic skills to effectively present the topical content, plus skills in fields like social sciences, psychology and ergonomics.

In particular, university instruction often suffers from a lack of didactic design. Since universities are also research institutions, their professors are usually hired based on their topical skills. Didactic skills are often underestimated in the recruiting process. We refrain from discussing reasons for that, but focus the issue of involving it a little more. Our approach to facing problems like these is a modeling concept for didactic knowledge called Storyboarding. A storyboard provides a roadmap for a course, including possible detours if certain concepts to be learned need reinforcement. Using modern media technology, a story-board also plays the role of a server that provides the appropriate content material when deemed required. Our suggestion to ensure a wide dissemination of this concept is to use a standard tool to develop and process this model, which is Microsoft Visio.

Section 2 is an introduction to the storyboard concept. It includes the present state of the current development. This is followed by a brief introduction to an exemplary storyboard in section 3. In section 4, we summarize the research undertaken so far and outline current work as well as research horizons.

2 Storyboarding

Former Storyboarding concepts to model information and learning processes have been introduced 1998/1999 [8]. The employment of storyboarding approaches for (unfortunately,

only) e-learning is characterized by misunderstandings. So-called storyboard concepts in use are mostly substitutes for software-technological documents of high-level design, but are not very much specific to the instructional design process [3][14]. Didactic concepts [9] are not made explicit and, thus, pondering about didactics is not sufficiently enforced. Again, also very recent approaches as introduced above (see also [13]) remain within the borders of IT systems.

There are contrasting approaches [10] that are conceptually very useful, but syntactically much too far from a workflow directed to technology enhanced learning implementations. The crux is that purely software-technologically driven concepts do not provide an opportunity to represent and discuss details of human learning [2] [4]. Learning is much more than memorizing: "Learning imposes new patterns of organization on the brain, and this phenomenon has been confirmed by electrophysiological recordings of the activity of nerve cells." ([2], p. 121).

Learning is reasonably understood as an interactive knowledge construction process. Illustrative case studies are discussed in [5]. This book's chapter "3B Organizing Shapes" reports process of conversation and co-operation between a teacher and his students in which a variety of media types, forms of interaction, and learners' activities are dovetailed. Didactic design means the anticipation of those communication processes [9], and storyboards may provide the expressive power suitable to the design and implementation of learning processes. This, however, needs to go beyond the limits of software systems specification – the crucial question for innovations in didactic design.

Our storyboard concept is built upon standard concepts which enjoy (1) clarity by providing a high-level modeling approach, (2) simplicity, which enables everybody to become a storyboard author, and (3) visual appearance as graphs. With respect to a better formal composition, processing, verification, validation and refinement the concept as introduced so far [7] [11] has been further developed. We adopt these modifications. Here, we define a storyboard as follows:

A storyboard is a nested hierarchy of directed graphs with annotated nodes and annotated edges. Nodes are scenes or episodes. Scenes denote leaves of the nesting hierarchy. Episodes denote a sub-graph. There is exactly one Start- and End- node to each (sub) graph. Edges specify transitions between nodes. They may be single-color or bi-color. Nodes and edges have (pre-defined) key attributes and may have free attributes.

The interpretations of these terms follow after presenting a small example.

The representation as a graph (instead of a linear sequence) reflects the fact that different readers trace the paper differently according to their particular interests, prerequisites, a current situation (like being under time pressure), and other circumstances. The story-board is the authors' de-

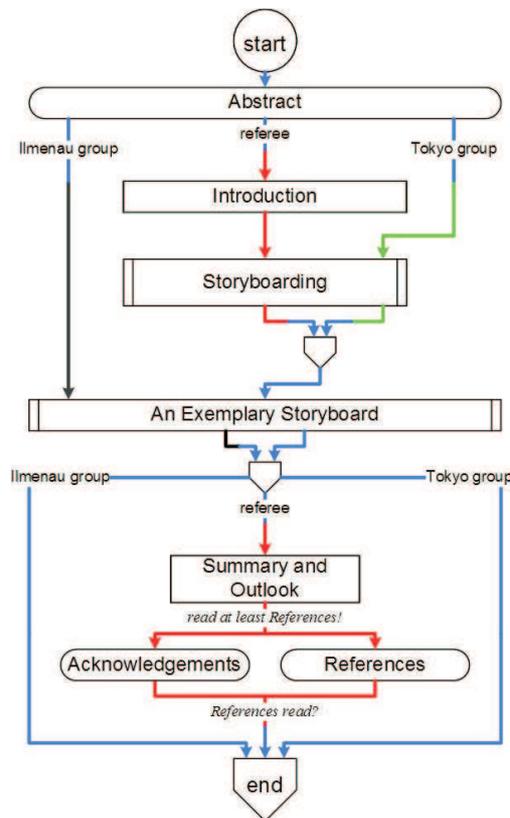


Figure 1: A storyboard on this paper

sign document representing expectations of human behavior. For exemplification, Figure 1 shows a top level storyboard on the present paper. Alternative paths may be driven by the reader's role:

- The Ilmenau research group may skip sections 1, 2 and 4, because they are familiar with it. Since the example application is new to them, they will study this example.
- The Tokyo research group may skip sections 1 and 4. Since they are interested in the recently developed refinements of the concept, they are interested in section 2. They like to know the way of storyboarding their study and therefore, study the section 3.
- Referees (hopefully) want to read all. After section 4, they can read Acknowledgements and References independently in any sequence. For their duty they have to check the References at least.

A storyboard can be traversed in different manners according to (1) users' interests, objectives, and desires, (2) didactic preferences¹, (3) the sequence of nodes (and other storyboards) visited before (i.e. according to the educational history), (4) available re-sources (like time, money, equipment to present material, and so on) and (5) other application driven circumstances. In fact, people may read the present paper in ways that are different from our assumptions modeled in Figure 1. However, for the ways we anticipate, we can ensure that they are coherent. A storyboard may be seen as a model of an anticipated reception process that is interpreted as follows:

- *Scenes* denote a non-decomposable learning activity that can be implemented in any way. It can be the presentation of a (media) document, opening a tool that supports learning (URL or e-learning system) or an informal activity description.
- *Episodes* are defined by their sub-graph.
- *Graphs* are interpreted by the paths, on which they can be *traversed*.
- A *Start Node* of a (sub-) graph defines the starting point of a legal graph traversing.
- An *End Node* of a (sub-) graph defines the final target point of a legal graph traversing.
- *Edges* denote transitions between nodes. There are rules to leave a node by an outgoing edge: (1) The outgoing edge must have the same color as the in-coming edge by which the node was reached. (2) If there is a condition specified as the edge's key attribute, this condition has to be met for leaving the node by this edge.
- *Key attributes of nodes* specify application driven information, which is necessary for all nodes of the same type, e.g. actors and locations.
- *Key attributes of edges* specify conditions, which have to be true for traversing on this edge.
- *Free attributes* specify whatever the storyboard author wants the user to know: didactic intentions, useful methods, necessary equipment, ...

Node and edge types, their visual appearance, their behavior on double click, and their behavior when following a hyperlink are as specified in tables 1 and 2. For both forking edges, there is a reverse fork at the end of the forked paths at the point of their merging. This reverse fork is marked with synchronization condition that needs to be satisfied before visiting the subsequent node.

What are peculiarities of the concept? At a first view, this purpose is similar to the purpose of traditional storyboards that are produced for shows, plays, theater games or movies, i.e. visual arts. The materials and tools of the storyboarded learning activities (e.g., text books, scripts, slides, hard- and software models, e-learning systems and others) are something comparable to the requisites of a show. Basic differences of our storyboards to those used to "specify" a show are:

¹In the authors' experience, some students understand better by presenting illustrations, others by providing a small example and others by providing formal descriptions.

Table 1: Node Types

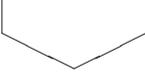
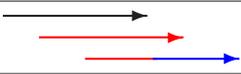
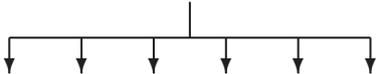
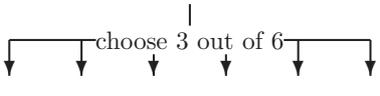
	Symbol	Behavior when	
		following a hyperlink	double clicked
Scene		<ul style="list-style-type: none"> opening a document (*.doc, *.pdf, *.wav, *.vsd, *.ppt, *.xls, ...) with the appropriate tool nothing, if just verbally described scene 	<ul style="list-style-type: none"> opening a document (*.doc, *.pdf, *.wav, *.vsd, *.ppt, *.xls, ...) with the appropriate tool visiting a website with the standard browser in case it is an URL opening the standard mail tool in case it is an e-mail address
Episode		opening a subgraph that specifies the episode	<ul style="list-style-type: none"> opening a document (*.doc, *.pdf, *.wav, *.vsd, *.ppt, *.xls, ...) with the appropriate tool visiting a website with the standard browser in case it is an URL opening the standard mail tool in case it is an e-mail address
Start Node		jumping to the start node of the related super-graph	not meaningful
End Node		jumping to the <i>Reference Node</i> that succeeds it's associated <i>Episode Node</i> in the related super-graph	not meaningful
Reference Node		jumping to the <i>End Node</i> of the subgraph that is associated to the preceded <i>Episode Node</i>	not meaningful

Table 2: Edge Types

	Symbol	Interpretation
Simple Edge		defines a unique successor node
Fork		defines several successor nodes, which have to be traversed independently from each other, i.e. in any sequence or parallel
Fork with conditions		defines several successor nodes, which have to be traversed independently from each other, i.e. in any sequence or parallel, according to the specified condition, e.g. take n out of m paths

- the primary purpose (learning vs. entertainment)²,
- the degree of formalization, and, as a consequence of being semi-formal,
- the obligation of everything above the level of scenes, which does (and should) not apply to storyboards in arts, in which the intendant has some freedom of individual interpretation and
- (thanks to formalization) the opportunity to formally represent, process, evaluate, and refine our storyboards, which does not apply at all to story-boards in visual arts.

Also, Storyboards have somewhat in common with classic AI knowledge representations like *Semantic Networks* and *Frames* as well as with process modeling languages like *State Diagrams* and *Petri Nets* (see e.g. [1] for use in learning processes), *Workflow Diagrams* (see e.g. [12] for use in learning processes) and *Float Charts* (see e.g. [15] for use in learning processes). Items that make this concept more expressive for didactic knowledge than representations as mentioned above are

- the potentially unlimited nesting of graphs,
- the opportunity to express "conditioned" edges by using the colors (bi-colored edges, e.g.) or respective key annotations to edges,
- the opportunity to use (two kinds of) fork-edges,
- the potential of nodes to carry many different teaching materials and tools as hyper-links³, and, most important, and
- the fact that a scene can be implemented in any way, i.e. is not restricted to something electronically available or even formally structured (like any knowledge representations and any material included in process models).

3 An exemplary storyboard

For *Tokyo Denki University* (TDU), we developed a storyboard on the undergraduate study of *Information Environment*. This project is of a special interest, because TDU introduced some dynamics into the study by a system called *Dynamic Learning Needs Reflection System* (DLNRS) [6]. For illustration, Figure 2 shows a sub-graph of the episode on *General Cultural Subjects* for the undergraduate study of *Information Environment* at TDU. To take the individuality and dynamics into account when composing a storyboard, issues like goals, pre-conditions, achieved *Grade Points* of the previous semester and other circumstances need to be formalized and associated to both the related episodes (as a key annotation) and the students. By programming, these annotations to subjects (like their number units, e.g.) and the annotations to students (like the achieved *Grade Points*, e.g.) can be analyzed. This way, storyboards can be individualized and regularly (after each semester) updated according to their new status.

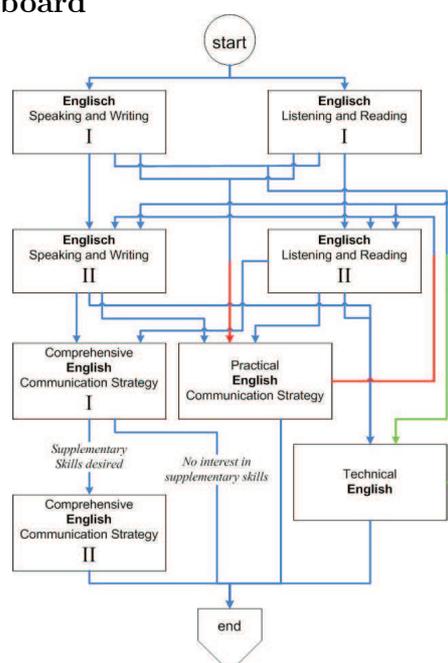


Figure 2: English subjects

²This is no ambivalence. To include entertainment into learning is a key of successful learning and an ultimate objective of storyboarding learning processes.

³The first author developed a storyboard for an AI course at an US university and included material of his own AI course in Germany. Now, this storyboard serves both universities and is also a common platform for internationally sharing teaching materials.

4 Summary and Outlook

Storyboarding makes didactic design explicit. Since the scenes are not limited to the presentation of electronic material and represent any learning activity, this concept goes far beyond the IT approaches to support learning so far. Representing knowledge at a high level by a modeling concept that is appropriate to be used by topical experts without the need of an IT- or even software technological background is very much AI-driven. To validate the usefulness in practice, we developed several storyboards on various subjects and for a complete university study.

One essential property is simplicity in terms of both the concept itself and the tool we used to implement it. Everybody, also university instructors of subjects that are far removed from information technology, are able to develop storyboards.

Current work is dedicated to following issues: (1) We develop various particular storyboards on many different subjects, that are not limited to university teaching to derive suggestions for the refinement of the concept. (2) We introduce automated storyboard verification by structure tests like (2a) the one-to-one mapping of episode nodes and related sub-graphs, (2b) the tree structure of the graph hierarchy, (2c) the reachability of each node from the start node of its (sub-) graph, (2d) consistency in edge coloring, and (2e) certain condition checks for fork edges and their synchronization when the forked paths merge. (3) We develop a set of basic operations to compose story-boards systematically, which guarantee consistency.

Our *short term objective* of upcoming work is promoting the use of storyboards. As a *medium term objective*, we plan to develop an evaluation concept for storyboards based on the learning results of the students as acquired from the final grade they achieve for the storyboarded courses as well as the students' specific comments in a questionnaire. Our dream and *long term objective* is to identify typical didactic patterns of successful storyboards. Since the learning result of a particular student is associated to a particular path through the storyboard, we will be able to identify successful storyboards in general, but also successful paths within storyboards in particular. By Machine Learning methods, we'll find out what these successful storyboard paths have in common and in which properties they differ from the less successful ones. Thus, we might be able to identify successful didactic patterns. The latter is the vision of knowledge discovery in didactics. By utilizing didactic insights for the upcoming storyboards, we intend to close the loop of the never ending storyboard development spiral.

References

- [1] Balaji, V., Arora, A., Jain, A., Goyal, G., Dubey, G., and Singh, S.: SCORM Learning Sequence Modeling with Petri Nets in Cooperative Learning. *Learning Technology*, 7(1), Special Issue on SCORM 2004 Sequencing & Navigation, ISSN 1438-0625, 2005.
- [2] Bransford, J.D., Brown, A.L., and Cocking, R.R.: *How People Learn: Brain, Mind, Experience, and School*. National Academic Press, 2000.
- [3] Briggs, R.M., Gagne, L.J., and Wager, W.W.: *Principles of Instructional Design*. Thomson Learning, 1992.
- [4] Damasio, A.: *The Feeling of What Happens: body and emotion in the making of consciousness*. Hartcourt, 1999.
- [5] Davis, B., Sumara, D., and Luce-Kapler, R.: *Learning and Teaching in a Complex World*. Lawrence Erlbaum Associates, 2000.
- [6] Dohi, S., Nakamura, Y., Sakurai, Y., Tsuruta, S., and Knauf, R.: Dynamic Learning Need Reflection System for academic education and its applicability to Intelligent Agents. *Proc. of the 6th IEEE International Conference on Advanced Learning Technologies (ICALT 2006)*, Kerkrade, the Netherlands, pp. 459-463, ISBN 0-7695-2632-2, Los Alamitos, CA: IEEE Computer Society, 2006.
- [7] Dohi, S., Sakurai, Y., Tsuruta, S., and Knauf, R.: Managing academic education through dynamic storyboarding. (accepted paper) *Proc. of the World Conference on e-Learning in Corporate, Government, Healthcare, and Higher Education 2006 (E-Learn 2006)*, October 2006, Honolulu, Hawaii, USA, 2006.
- [8] Feyer, T. and Thalheim, B.: E/R based scenario modelling for rapid prototyping of web information services. P.P.-S. Chen (ed): *Advances in Conceptual Modeling*, vol. 1727 of *Lecture Notes in Computer Science (LNCS)*, pp. 253-263, Berlin: Springer, 1999.

- [9] Flechsig, K.-H.: *Kleines Handbuch didaktischer Modelle*. (German) Neuland, 1996.
- [10] Forsha, H.I.: *The Complete Guide to Storyboarding and Problem Solving*. ASQ Quality Press, 1994.
- [11] Jantke, K.P. and Knauf, R.: Didactic design through storyboarding: Standard concepts for standard tools. *Proc. of the 4th Internat. Symposium on Information and Communication Technologies, Workshop on Dissemination of e-Learning Technologies and Applications*, Cape Town, South Africa, 2005, pp. 20-25, New York: ACM Press, ISBN 0-9544145-6-X, 2005.
- [12] Lin, J., Ho, C., Sadiq, W., Orlowska, W.E.: Using Workflow Technology to Manage Flexible e-Learning Services. *Educational Technology and Society*, 5(4), 2002.
- [13] Schewe, K.-D. and Thalheim, B.: The Co-Design Approach to WIS Development in E-Business and E-Learning Applications. *Proc. of FIPWIS 2004*, November 24, Brisbane, Australia, 2004.
- [14] Rothwell, W.J. and Kazanas, H.C.: *Mastering the Instructional Design Process: A Systematic Approach*. (Third Edition) Pfeiffer, 2004.
- [15] Sykes, E.R. and Franek, F.: A prototype for an intelligent tutoring system for students learning to program in Java. *Proc. of the IASTED International Conference on Computers and Advanced Technology in Education*, Rhodes, Greece, pp. 78-83, 2003.

Author Information:

Priv.-doz. Dr.-Ing. habil. **Rainer Knauf**

Technical University of Ilmenau, Dept. of Computer Science and Automation, Chair of AI
P.O. Box 10 05 65, 98684 Ilmenau, Germany

phone: +49 3677 69 1445, fax: +49 3677 69 1665, E-mail: rainer.knauf@tu-ilmenau.de

Prof. Dr.-Eng. **Setsuo Tsuruta**

Tokyo Denki University, School of Information Environment

2-1200 Musai-gakuendai, Inzai, Chiba, 270-1382, Japan

phone: +81 (476) 46 8491, fax: +81 (476) 45 8449, E-mail: tsuruta@sie.dendai.ac.jp

Prof. Dr.-Eng. **Yoshitaka Sakurai**

Tokyo Denki University, School of Information Environment

2-1200 Musai-gakuendai, Inzai, Chiba, 270-1382, Japan

phone: +81 (0) 476 46 9081, fax: +81 (476) 45 8449, E-mail: ysakurai@sie.dendai.ac.jp

U. Konigorski / A. López

Output Coupling by Dynamic Output Feedback

INTRODUCTION

In this paper, the synchronization of predefined control variables of a linear, time invariant multivariable system by means of dynamic output feedback is considered.

Many control applications require the synchronization or coupling of two or more control variables. Examples for this kind of problem are the speed synchronization of different electrical drives within a production line also called 'electronic gear' or the slip prevention between the different wheels of a car. The classical approach to solve this problem is the use of additional PI compensators to correct the divergences between the coupled control variables. This simple approach is often sufficient to accomplish the goals of stability and asymptotic synchronization of the closed-loop. However, this approach does not result in an exact dynamic synchronization and asymptotic synchronization is only achieved for piecewise constant reference inputs or disturbances.

Previous works [1,2] offer two different ways to tackle the problem of the synchronization of multiple outputs in state space. These methods solve the problem of synchronizing a given set of control variables by assigning a suitable eigenstructure to the closed-loop by means of state feedback. Since generally not all states are available for measurement, the implementation of these approaches relies on the use of observers to supply the missing degrees of freedom.

For that purpose also dynamic output feedback of appropriate order can be used where the states of the compensator supply the missing degrees of freedom for the design. Moreover, dynamic output feedback establishes the possibility to design PI-like compensators or to account for additional constraints as to the structure of the compensator. However, the problem of eigenvalue assignment by structurally constraint controllers generally has no analytic solution and therefore demands numerical methods to solve the underlying non-linear system of equations [3]. What's more, the problem of synchronization leads to some special restrictions in the eigenstructure of the closed-loop that have to be taken into account. In the following a new approach for the design of structurally constraint dynamic output feedback controllers is presented which not only allows arbitrary eigenvalue assignment but also assures the synchronization of some predefined output variables.

STATEMENT OF THE PROBLEM AND PRELIMINARY RESULTS

Consider a linear, time-invariant system of order n which is supposed to be completely controllable and observable.

$$\begin{aligned}\dot{x}_s(t) &= A_s \cdot x_s(t) + B_s \cdot u_s(t) \\ y_s(t) &= C_s \cdot x_s(t)\end{aligned}\quad (1)$$

The number of inputs and outputs is denoted by p and q , respectively. In what follows, for system (1) the dynamic output feedback

$$\begin{aligned}\dot{x}_d(t) &= A_d \cdot x_d(t) + B_d \cdot y_s(t) + F_1 \cdot w(t) \\ u_s(t) &= -C_d \cdot x_d(t) - D_d \cdot y_s(t) + F_2 \cdot w(t)\end{aligned}\quad (2)$$

of order r is used to place the $n+r$ poles of the resulting closed loop system

$$\begin{bmatrix} \dot{x}_s(t) \\ \dot{x}_d(t) \end{bmatrix} = \dot{x}(t) = \begin{bmatrix} A_s - B_s D_d C_s & -B_s C_d \\ B_d C_s & A_d \end{bmatrix} x(t) + \begin{bmatrix} B_s F_2 \\ F_1 \end{bmatrix} w(t)\quad (3)$$

at a predefined set $\Lambda = \{\lambda_1, \dots, \lambda_{n+r}\}$ of real or conjugate complex values. Simultaneously the solution $x_s(t)$ of (3) must be such, that $l < p$ linear coupling conditions

$$T_2^t \cdot y_s(t) = T_2^t \cdot C_s \cdot x_s(t) = 0\quad (4)$$

are met, where the superscript ‘ t ’ denotes the transpose of a matrix. Obviously, we can always assume that the l coupling conditions are linear independent and thus the $(l \times q)$ coupling matrix T_2^t has rank l .

Before proceeding further with the description of a new approach which numerically solves the aforementioned design problem, some useful results from the corresponding literature [3,4] are summarized shortly.

It is easy to verify that the closed loop system (3) can also be written in terms of a constant output feedback

$$\begin{aligned}\dot{x}(t) &= [A - BKC]x(t) + BF w(t) \\ y(t) &= Cx(t)\end{aligned}\quad (5)$$

with

$$A = \begin{bmatrix} A_s & 0 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} B_s & 0 \\ 0 & I \end{bmatrix}, \quad C = \begin{bmatrix} C_s & 0 \\ 0 & I \end{bmatrix}, \quad K = \begin{bmatrix} D_d & C_d \\ -B_d & -A_d \end{bmatrix}, \quad F = \begin{bmatrix} F_2 \\ F_1 \end{bmatrix}.\quad (6)$$

According to (6) the constant output feedback matrix K contains all free parameters of the original dynamic output feedback (2). Now from the eigenvalue/eigenvector equation of the closed loop system (5)

$$(\lambda_i I - A + BKC) \cdot v_i = 0$$

can be deduced the important relation

$$v_i = -(\lambda_i I - A)^{-1} B \cdot p_i\quad (7)$$

where v_i is the right eigenvector to the closed loop eigenvalue λ_i and

$$p_i = KC \cdot v_i \quad (8)$$

is a so called parameter vector (see [3,4]). In case of complete state feedback, e.g. $C = I$, this nonzero parameter vector $p_i \neq 0$ can almost arbitrarily be chosen. Otherwise it is subject to non-obvious constraints. In any case, substituting (7) into (8) yields

$$\left[I + KC(\lambda_i I - A)^{-1} B \right] \cdot p_i = [I + KG(\lambda_i)] \cdot p_i = H_i \cdot p_i = 0 \quad (9)$$

and for $p_i \neq 0$ this results in the condition

$$\det(H_i) = 0. \quad (10)$$

If (10) is solved by a suitable choice of K , λ_i is an eigenvalue of the closed loop system (5) and therefore minimizing the cost function

$$J(K) = \sum_{i=1}^{n+r} w_i \cdot \det(\bar{H}_i^t) \cdot \det(H_i) = \sum_{i=1}^{n+r} w_i \cdot \det(\bar{H}_i^t H_i), \quad w_i > 0 \quad (11)$$

with respect to K yields a solution K^* to the pole placement problem provided $J(K^*) = 0$. In (1) \bar{H}_i denotes the conjugate complex of H_i .

OUTPUT COUPLING BY CONSTANT OUTPUT FEEDBACK

First of all, the coupling condition (4) needs to be adapted to the new output equation in (5), where according to (6) the new output matrix C has r additional columns due to the compensator states. Since these additional outputs are not involved in the coupling, the matrix T_2^t needs to be expanded with a zero matrix of dimension $(l \times r)$

$$T_2^t \cdot y_s = \begin{bmatrix} T_2^t & 0 \end{bmatrix} \cdot y = \tilde{T}_2^t \cdot y = 0. \quad (12)$$

Now, substituting the Laplace transform of (5) in (12) leads to a new expression for the coupling conditions

$$\tilde{T}_2^t C (sI - A + BKC)^{-1} B \begin{bmatrix} \tilde{F}_1 & \tilde{F}_2 \end{bmatrix} w(s) = 0, \quad (13)$$

where $F = \begin{bmatrix} \tilde{F}_1 & \tilde{F}_2 \end{bmatrix}$. After applying a modal transformation to (13) (see [2]) it becomes

$$\sum_{i=1}^{n+r} \frac{\tilde{T}_2^t C v_i w_i^t B \begin{bmatrix} \tilde{F}_1 & \tilde{F}_2 \end{bmatrix}}{s - \lambda_i} \begin{bmatrix} w_1(s) \\ w_2(s) \end{bmatrix} = 0 \quad (14)$$

where v_i and w_i^t represent the right and left eigenvectors of the closed-loop system (5),

respectively. Obviously, with

$$w_2(s) = 0 \quad (15)$$

equation (14) can be split into two parts

$$\begin{aligned} \tilde{T}_2^t C \cdot v_i &= 0, \quad i = 1 \dots m \\ w_i^t B \cdot \tilde{F}_1 &= 0, \quad i = m + 1, \dots, n + r. \end{aligned} \quad (16)$$

The set of equations (16) represent the output- and input-coupling conditions for the constant output feedback system (5) and are formally equivalent to the coupling equations presented in [2] for the case of full state feedback, where it is shown that just the output-coupling conditions are relevant for the calculation of the controller K .

Therefore, substituting (7) into (16) the output-coupling condition reads

$$\tilde{T}_2^t C (\lambda_i I - A)^{-1} B \cdot p_i = \tilde{G}(\lambda_i) \cdot p_i = 0 \quad (17)$$

and thus the parameter vector p_i must be contained in the null space

$$\tilde{G}(\lambda_i) \cdot \tilde{N}_i = 0 \quad (18)$$

of $\tilde{G}(\lambda_i)$ to generate via (7) an eigenvector v_i that is compliant with the output-coupling condition (16). According to [5], the eigenvectors v_i constructed in this way span the (A, B) -invariant subspace in the kernel of $\tilde{T}_2^t C$ which in the sequel is assumed to have dimension m . So, any arbitrary $\tilde{p}_i \neq 0, i = 1, \dots, m$ results in an admissible parameter vector

$$p_i = \tilde{N}_i \tilde{p}_i \neq 0, \quad i = 1, \dots, m \quad (19)$$

and the remaining $n + r - m$ input-coupling conditions in (16) yield

$$[v_1, \dots, v_m, \quad B] \cdot \begin{bmatrix} M \\ \tilde{F}_1 \end{bmatrix} = 0 \quad (20)$$

from which a nonsingular prefilter \tilde{F}_1 can be calculated [2]. This \tilde{F}_1 makes the corresponding $n + r - m$ eigenvalues of the closed loop system (5) uncontrollable from the input w_1 . Therefore, what remains is the calculation of the constant output feedback matrix K in the feedback law

$$u = -Ky + \tilde{F}_1 w_1 \quad (21)$$

for the system (A, B, C) such that the closed loop has the $n + r$ predefined eigenvalues from the set Λ . This can be achieved by numerical minimization of (11) where for the first m eigenvalues the additional constraints (19) must be taken into consideration. With

respect to (9) this results in

$$J(K) = \sum_{i=1}^m w_i \cdot \det(\tilde{N}_i^t \bar{H}_i^t H_i \tilde{N}_i) + \sum_{i=m+1}^{n+r} w_i \cdot \det(\bar{H}_i^t H_i), \quad w_i > 0. \quad (22)$$

NUMERICAL EXAMPLE

The dynamics of a DC motor are described by the following equations

$$\begin{aligned} J \frac{d\ddot{\varphi}(t)}{dt} &= c_m i(t) - d\omega(t) \\ u(t) &= Ri(t) + c_m \omega(t) + L \frac{di(t)}{dt} \end{aligned} \quad (23)$$

where J represents the moment of inertia, c_m the motor constant, d the speed proportional damping, R the resistance, L the inductance, u the input voltage, i the motor current, φ the mechanical angle and $\omega = \frac{d\varphi}{dt}$ the mechanical angular speed.

Consider two independent motors "A" and "B" with $R_A = R_B = 1$, $L_A = L_B = 0,05$, $c_{mA} = c_{mB} = 1$, $d_A = d_B = 0$ but different moments of inertia $J_A = 0,1$ and $J_B = 0,025$ [1] (units are ignored). Their state equations can be summarized as follows

$$\dot{x}_s(t) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 40 & 0 & 0 & 0 \\ 0 & -20 & -20 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 10 \\ 0 & 0 & 0 & 0 & -20 & -20 \end{pmatrix} \cdot x_s(t) + \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 20 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 20 \end{pmatrix} \cdot u_s(t), \quad y_s(t) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \cdot x_s(t) \quad (24)$$

where $x_s = [\varphi_A \quad \omega_A \quad i_A \quad \varphi_B \quad \omega_B \quad i_B]^t$ is the n -dimensional state vector and the measured variables are $y_s = [\varphi_A \quad \omega_A \quad \varphi_B]^t$. The two uncoupled drives have their eigenvalues at $\Lambda_s = \{0 \quad 0 \quad -10 \pm 26,45i \quad -10 \pm 10i\}^t$ and they shall be shifted to

$$\Lambda = \{-4 \quad -20 \pm 25i \quad -15 \pm 6i \quad -8 \quad -10 \quad -12\}^t \quad (25)$$

by means of a dynamic output feedback of order 2. Simultaneously the angles φ_A and φ_B of the two motors must be synchronized. Thus the coupling matrix is set to $T_2^t = [1 \quad 0 \quad -1]$. Following the guidelines given in [5] for the calculation of the (A, B) -invariant subspace of $T_2^t C_s$ it is found that the dimension of this subspace is $m = 3$. Hence three eigenvectors can be found which are compliant with the output-coupling conditions (16) and these eigenvectors have been chosen to be $v_{\lambda=-4}$ and $v_{\lambda=-20 \pm 25i}$. Then from (18) the corresponding null spaces \tilde{N}_i can be easily calculated and after performing a minimization of the cost function (22) with the predefined set of closed-loop eigenvalues (25) the following constant output feedback matrix K is found

$$K = \begin{pmatrix} 2,525 & 0,513 & -0,513 & 0,390 & -0,167 \\ 2,660 & 3,428 & 0,599 & 1,368 & -0,452 \\ 156,35 & 14,41 & -2,772 & -6,753 & 11,637 \\ 343,65 & -158,17 & 147,64 & 119,42 & 70,753 \end{pmatrix} \quad (26)$$

from which the system matrices (A_d, B_d, C_d, D_d) of the dynamic controller (2) can be extracted according to (6). Finally, with the help of (20) the prefilter

$$\tilde{F}_1 = [-3,575 \quad -14,299 \quad -353,550 \quad -311,178]^t \cdot 10^{-3} \quad (27)$$

can be calculated and the transfer function of the closed-loop system then reads

$$G(s) = \begin{pmatrix} \frac{4100s^5 + 2,46 \cdot 10^5 s^4 + 5,974 \cdot 10^6 s^3 + 7,24 \cdot 10^7 s^2 + 4,348 \cdot 10^8 s + 1,027 \cdot 10^9}{s^8 + 104s^7 + 5282s^6 + 1,57 \cdot 10^5 s^5 + 2,86 \cdot 10^6 s^4 + 3,18 \cdot 10^7 s^3 + 2,09 \cdot 10^8 s^2 + 7,32 \cdot 10^8 s + 1,027 \cdot 10^9} \\ \frac{4100s^6 + 2,46 \cdot 10^5 s^5 + 5,974 \cdot 10^6 s^4 + 7,24 \cdot 10^7 s^3 + 4,348 \cdot 10^8 s^2 + 1,027 \cdot 10^9 s}{s^8 + 104s^7 + 5282s^6 + 1,57 \cdot 10^5 s^5 + 2,86 \cdot 10^6 s^4 + 3,18 \cdot 10^7 s^3 + 2,09 \cdot 10^8 s^2 + 7,32 \cdot 10^8 s + 1,027 \cdot 10^9} \\ \frac{4100s^5 + 2,46 \cdot 10^5 s^4 + 5,974 \cdot 10^6 s^3 + 7,24 \cdot 10^7 s^2 + 4,348 \cdot 10^8 s + 1,027 \cdot 10^9}{s^8 + 104s^7 + 5282s^6 + 1,57 \cdot 10^5 s^5 + 2,86 \cdot 10^6 s^4 + 3,18 \cdot 10^7 s^3 + 2,09 \cdot 10^8 s^2 + 7,32 \cdot 10^8 s + 1,027 \cdot 10^9} \end{pmatrix} \quad (28)$$

Obviously the outputs 1 and 3 or φ_A and φ_B share the same transfer function

$G_{\varphi_{A/B}}(s) = \frac{\varphi_{A/B}(s)}{w_1(s)}$ and thus are perfectly synchronized. Moreover, since $\omega_A = \frac{d\varphi_A}{dt}$ or

$\omega_A(s) = s \cdot \varphi_A(s)$, the second output $\omega_A(s)$ has the transfer function

$$G_{\omega_A}(s) = \frac{s \cdot \varphi_A(s)}{w_1(s)} = s \cdot G_{\varphi_A}(s).$$

References:

- [1] Reglerentwurf für Mehrgrößensysteme unter der Nebenbedingung vorgegebener Ausgangsgrößenverkopplungen, A. Jochheim VDI Verlag, Reihe 8 Nr. 457
- [2] Ausgangsgrößenverkopplung bei linearen Mehrgrößensystemen, U. Konigorski, at 4/99 P. 165
- [3] Ein direktes Verfahren zum Entwurf strukturbeschränkter Zustandsrückführungen durch Polvorgabe, U. Konigorski, VDI Verlag, Reihe 8, Nr. 156
- [4] Vollständige modale Synthese linearer Systeme und ihre Anwendung zum Entwurf strukturbeschränkter Zustandsrückführungen, G. Roppenecker, VDI Verlag, Reihe 8, Nr. 59
- [5] Linear Multivariable Control: A Geometric Approach, W. M. Wonham, Springer-Verlag, 2nd Edition

Authors:

Prof. Dr.-Ing. Ulrich Konigorski
 Dipl.-Ing. Alejandro López Pamplona
 Institute für Automatisierungstechnik, TU Darmstadt, Landgraf-Georg Straße 4
 64283, Darmstadt
 Phone: +49 615116 3014
 Fax: +49 615116 6114
 E-mail: ukonigorski@iat.tu-darmstadt.de

H. Toossian Shandiz / A. Hajipoor

Chaos in the Fractional Order Chua System and its Control

Abstract: In this paper, we study the chaotic behaviors in the fractional order Chua system. We found that chaos exists in the fractional order Chua system with order less than 3. The lowest order we found to have chaos in this system is 2.7. Linear feedback control of chaos in this system is also studied.

1. Introduction

Fractional calculus is one of the classical mathematical topics in recent years. According to [1,2], more attentions have been paid to the application of fractional calculus in physics, engineering systems and financial analysis.

The fractional-order dynamics of a system known to us include viscoelastic systems [3,4], dielectric polarization [5], electrode–electrolyte polarization [6], electromagnetic waves [7], quantitative finance [8], and quantum evolution of complex systems [9]. Moreover, the control of fractional-order dynamic systems is also performed by various researchers [10–15].

Zaslavsky [16] conducted a comprehensive review for the existing models of fractional kinetics and their connection to dynamical models, phase space topology, and other characteristics of chaos. Many researchers have found that the chaotic attractors indeed exist in fractional-order systems according to [17–24]. In 2004, Li and Chen [25] found that the hyper chaos in fractional order Rossler equations has an order as low as 3.8.

In this paper, we study the chaotic behaviors in the fractional order Chua system [24]. A linear feedback control is also presented for this fractional order system.

2. Approximation of Fractional Derivative

There are several definitions of fractional derivatives [1]. Perhaps the best known is the Riemann–Liouville definition, which is given by

$$\frac{d^\alpha f(t)}{dt^\alpha} = \frac{1}{\Gamma(n-\alpha)} \frac{d^n}{dt^n} \int_0^t \frac{f(\tau)}{(t-\tau)^{\alpha-n+1}} d\tau \quad (1)$$

where $\Gamma(\cdot)$ is the gamma function and $n-1 \leq \alpha < n$. This definition is significantly different from the classical definition of derivative.

Fortunately, the Laplace transform is still applicable and works as one would expect. Upon considering all the initial conditions to be zero, the Laplace transform of the Riemann–Liouville fractional derivative satisfies the following equation.

$$L\left\{\frac{d^\alpha f(t)}{dt^\alpha}\right\} = s^\alpha L\{f(t)\} \quad (2)$$

Thus, the fractional integral operator of order α can be represented by the transfer

function $F(s) = 1/s^\alpha$. The standard definition of fractional differential does not allow direct implementation of fractional operators in time-domain simulations. An efficient method to circumvent this problem is to approximate fractional operators by using standard integer order operators. In [26], an effective algorithm is developed to approximate fractional order transfer functions. Basically, the idea is to approximate the system behavior in the frequency domain. By utilizing frequency domain techniques based on Bode diagrams, one can obtain a linear approximation of the fractional order integrator, the order of which depends on the desired bandwidth and discrepancy between the actual and the approximate magnitude Bode diagrams. This approximation approach was adopted in [15], [18], [21–23]. In Table 1 of [15], approximations for $1/s^q$ with $q = 0.1–0.9$ in steps 0.1 are given, with errors of approximately 2 dB. We also use these approximations in the following simulations.

3. The Fractional Order Chua System

We consider the fractional order Chua system. The standard derivative [24] is replaced by fractional derivatives as follows:

$$\begin{aligned}\frac{d^q x}{dt^q} &= \alpha \left[y + \frac{x - 2x^3}{7} \right] \\ \frac{d^q y}{dt^q} &= x - y + z \\ \frac{d^q z}{dt^q} &= -\frac{100}{7} y\end{aligned}\tag{3}$$

where q is the fractional order. When $q = 1$, system (3) is the original integer order Chua system. Simulations are performed for $q = 0.9, q = 1.1$. The simulation results demonstrate that chaos indeed exist in the fractional order Chua system with order less than 3. When $q = 0.9, q = 1.1$ chaotic attractors are found and the phase portraits are shown in Figs. 1, 2 and 3, respectively. When $q = .8$ no chaotic behavior is found, which indicates that the lowest limit of the fractional order for this system to be chaotic is $q = 0.8–0.9$. Thus, the lowest order we found for this system to yield chaos is 2.7

4. Stability and Controller Design

In this section, stability of the fractional order Chua system is discussed. Then a controller is proposed to meet the stability criteria.

4.1. Stability Region of Fractional Order Systems

Stability of fractional systems has been thoroughly investigated. The necessary and sufficient conditions have been derived in [25]. It has been shown that the stability region of a linear set of fractional order equations with order q , is bounded by a cone, with vertex at the origin, and hat extends into the right half of the s-plane such that it encloses an angle of $\pm q\pi/2$ as shown in Fig. 1. For example, the stability region of the linearized part of equation (3) when $q = 0.5$ is the entire s-plane less the area enclosed by the cone making $\pm 45^\circ$. Thus, when $q = 1$, we get the all-familiar stability region of the integer order system, i.e. the left half-plane where the imaginary axis becomes the border of stability region. Hence, if the eigenvalues of the system Jacobian matrix are placed anywhere outside the cone in Fig. 1, the fractional order system will be stable. Moreover, a controller that stabilizes the integer order system stabilizes the fractional

order system. Therefore, a controller that places the characteristic roots in the left half-plane will stabilize both the integer order model as well as all of its fractional versions. However, from performance standpoint, it may be necessary to place the characteristic roots of the fractional system in the right half-plane but outside the stability cone. In this case, the fractional order system is stable whereas the integer order system is unstable.

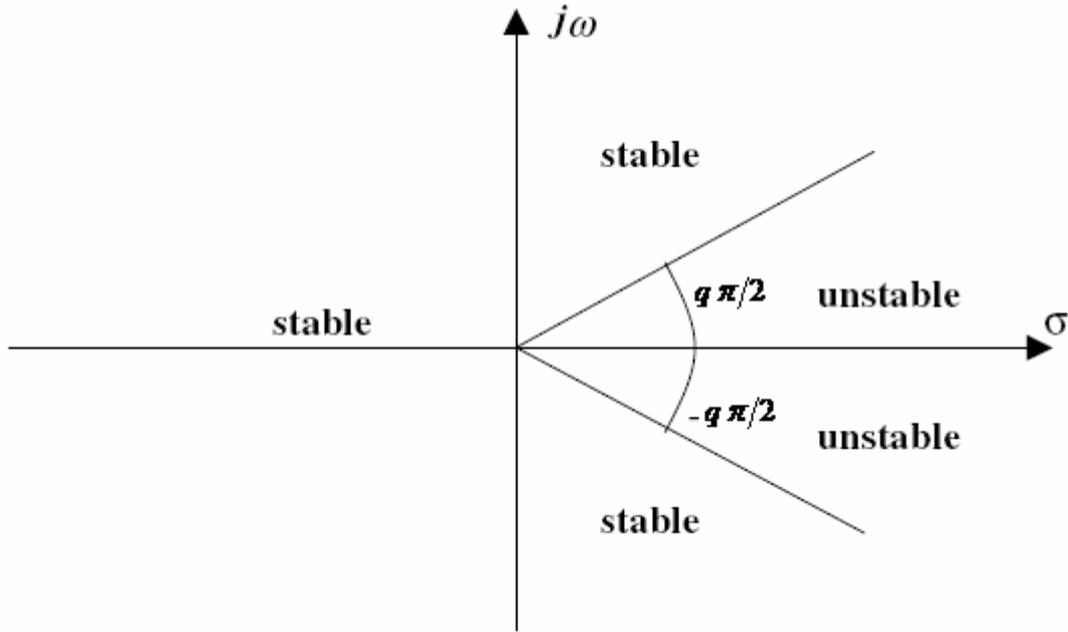


Figure 1. Stability region of fractional order system for order q

4.2. State feedback controller design

State feedback controllers will be proposed to stabilize the fractional chaotic system described by equation (3). The controller design may be based on placing the eigenvalues of the Jacobian system matrices for (3) when $q = 1$ in the left half of the s -plane. Alternatively, we will demonstrate how the fractional system can be stabilized by using static gains that place the eigenvalues in the right half-plane but outside the cone described in $\theta = \pm q\pi/2$. The composite fractional system models with a control law are described by:

$$\frac{d^q X}{dt^q} = AX + B_1 f(X) + B_2 u \quad (4)$$

Where $X^q = [x^q \ y^q \ z^q]$, $f(x) = \alpha x - 2x^3/7$, the matrix $B_1 = [0 \ 0 \ 1]^T$ for system (3), and where $[.]^T$ is the transpose of $[.]$. The input matrix B_2 is chosen so that the pair (A, B_2) for the corresponding system is controllable.

The static gain controller takes the form $u = -Kx$ where $K = [k_1, k_2, k_3]$. It can be seen that with $B_2 = [1 \ 1 \ 1]^T$, the Chua model is completely controllable as indicated by the controllability matrix $Q = [B \ AB \ A^2B]$. The dynamics of the controlled fractional chaotic Chua describe by:

$$\begin{bmatrix} \frac{d^q x}{dt^q} \\ \frac{d^q y}{dt^q} \\ \frac{d^q z}{dt^q} \end{bmatrix} = \begin{bmatrix} -k_1 & -k_2 + \alpha & -k_3 \\ -k_1 + 1 & -k_2 - 1 & -k_3 + 1 \\ -k_1 & -k_2 - \frac{100}{7} & -k_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} \alpha \\ 0 \\ 0 \end{bmatrix} \left(\frac{x - 2x^3}{7} \right) \quad (5)$$

The controller gains k_1, k_2 and k_3 are chosen such that the eigenvalues of $[A - B_2K]$ are placed outside the cone of angle $\theta = \pm q\pi/2$.

5 Conclusion

In this paper, we have studied the chaotic dynamics of the fractional order Chua system. We found that chaos exists in this system with order less than 2.7. A simple, but effective, linear feedback controller is also designed to stabilize the fractional order chaotic Chua system.

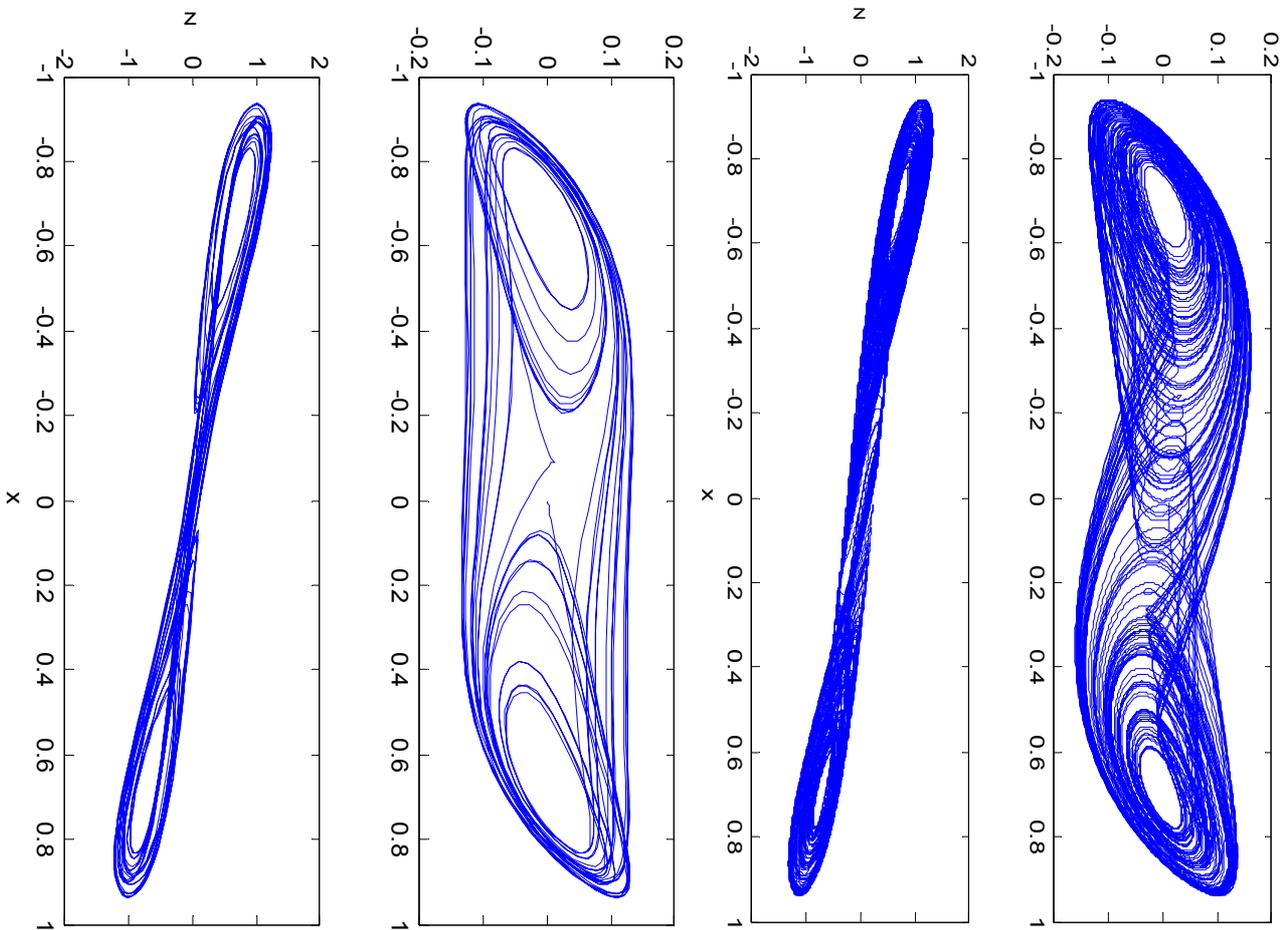


Figure 2. Chaotic attractor of the fractional order Chua system with order $q = .9$ $q=0.9$ and $\alpha = 12.75$

Figure 3. Chaotic attractor of the integer order Chua system with order $q = 1$ and $\alpha = 9.5$

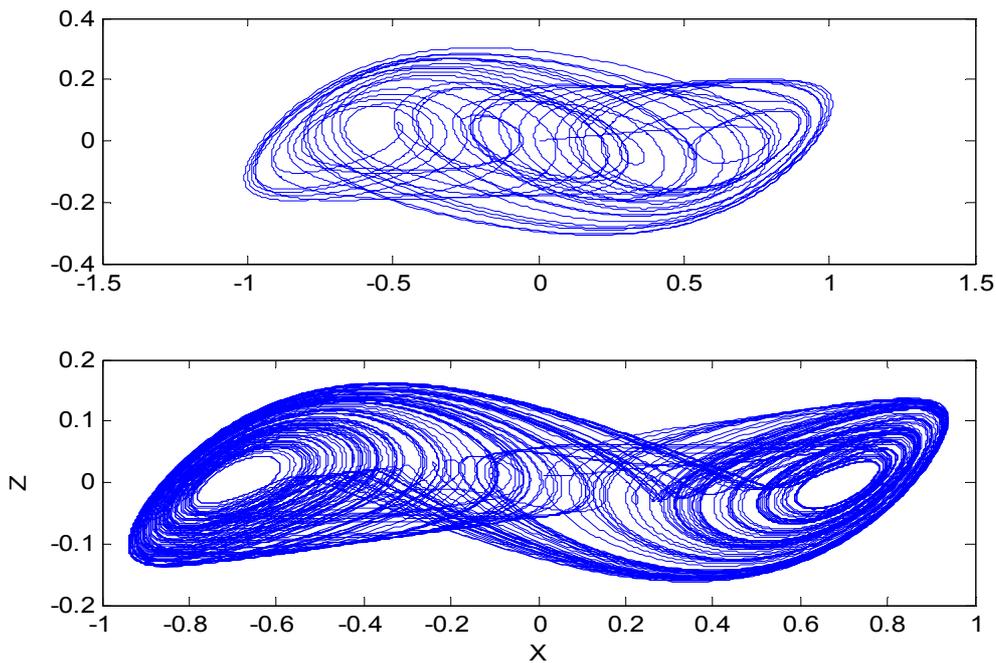


Figure 4. Chaotic attractor of the fractional order Chua system with order $q = 1.1$ and $\alpha = 7$

References:

- [1] Podlubny I. Fractional differential equations. New York: Academic Press; 1999.
- [2] Hilfer R, editor. Applications of fractional calculus in physics. New Jersey: World Scientific; 2001.
- [3] Bagley RL, Calico RA. Fractional order state equations for the control of viscoelastically damped structures. J Guid, Contr Dyn 1991;14:304–11.
- [4] Koeller RC. Application of fractional calculus to the theory of viscoelasticity. J Appl Mech 1984;51:199.
- [5] Sun HH, Abdelwahed AA, Onaral B. Linear approximation for transfer function with a pole of fractional order. IEEE Trans Auto Contr 1984;29:441–4.
- [6] Ichise M, Nagayanagi Y, Kojima T. An analog simulation of noninteger order transfer functions for analysis of electrode process. J Electroanal Chem 1971;33:253–65.
- [7] Heaviside O. Electromagnetic theory. New York: Chelsea; 1971.
- [8] Laskin N. Fractional market dynamics. Physica A 2000;287:482–92.
- [9] Kunsevov D, Bulagc A, Dang GD. Quantum levy processes and fractional kinetics. Phys Rev Lett 1999;82:1136–9.
- [10] Oustaloup A, Sabatier J, Lanusse P. From fractal robustness to CRONE control. Fract Calculus Appl Anal 1999;2:1–30.
- [11] Oustaloup A, Levron F, Nanot F, Mathieu B. Frequency band complex noninteger differentiator: characterization and synthesis. IEEE Trans CAS-I 2000;47:25–40.
- [12] Chen TQ, Moore K. Discretization schemes for fractional-order differentiators and integrators. IEEE Trans CAS-I 2002;79:363–7.
- [13] Hartley TT, Lorenzo CF. Dynamics and control of initialized fractional-order systems. Nonlinear Dyn 2002;29:201–33.
- [14] Hwang C, Leu JF, Tsay SY. A note on time-domain simulation of feedback fraction-order systems. IEEE Trans Auto Control 2002;47:625–31.
- [15] Podlubny I, Petras I, Vinagre BM, O_Leary P, Dorcak L. Analogue realizations of fractional-order

Nonlinear Dynamic controllers.2002;29:281–6.

[16] Zaslavsky GM. Chaos, fractional kinetics, and anomalous transport. Phys Rep 2002;371:461–580.

[17] Hartley TT, Lorenzo CF, Qammer HK. Chaos in a fractional order Chua_s system. IEEE Trans CAS-I 1995;42:485–90.

[18] Arena P, Caponetto R, Fortuna L, Porto D. Chaos in a fractional order Duffing system. In: Proc ECCTD, Budapest, 1997. p.1259–62.

[19] Ahmad W, El-Khazali R, El-Wakli A. Fractional-order Wien-bridge oscillator. Electr Lett 2001;37:1110–2.

[20] Ahmad WM, Sprott JC. Chaos in fractional-order autonomous nonlinear systems. Chaos, Solutions & Fractals 2003;16:339–51.

[21] Ahmad WM, Harb WM. On nonlinear control design for autonomous chaotic systems of integer and fractional orders. Chaos, Solitons & Fractals 2003;18:693–701.

[22] Grigorenko I, Grigirenko E. Chaotic dynamics of the fractional Lorenz system. Phys Rev Lett 2003;91:034101.

[23] Chen G, Dong X. From chaos to order: methodologies, perspectives and applications. Singapore: World Scientific; 1998.

[24]Hartley,T.T.:and Mossayebi,F.:Chontrol of Chua’s Circuit .J.of Circuits Syst., and Comput.,vol.3,no.1,March 1993,pp.173-194

[25] Matignon D. Stability results of fractional differential equations with applications to ontrol processing. In: IMACS, IEEE-SMC,. Lille, France, 1996. p. 963–8.

Authors:

Dr HEYDAR TOOSSIAN SHANDIZ

AHMAD HAJIPOOR

Shahrood University of Technology, Electrical Engineering
Faculty

7 Th Tir Square, P.o.Box 36155-316, Shahrood, IRAN

Zip code, city: 36155-316, Shahrood

Phone: 0098-9121733733

Fax: 0098-273-3334419

E-mail: htshandiz@shahroodut.ac.ir or
h_t_shandiz@hotmail.com

O. Katernoga / V. Popov / A. Potapovich / G. Davydau

Methods for Stability Analysis of Nonlinear Control Systems with Time Delay for Application in Automatic Devices.

SECTION HEADING 2

We investigated stability of nonlinear feedback systems, in which it is possible to separate a linear continuous part (LCP) from a nonlinear part (NP). Let's consider a formation method of mathematical models for discrete-time control systems (NDCS) on the following example. The characteristic $\varphi(\sigma)$ of random nonlinearity belongs to the sector $(0, k)$ and satisfies to the following conditions:

$$\varphi(0) = 0; 0 < \frac{\varphi(\sigma)}{\sigma} < k; \lim_{\sigma \rightarrow \infty} \varphi(\sigma) \neq 0, \quad (1)$$

where k is the constant given value, $k = y/\sigma$. In the Tsytkin's work [1] it is shown, that a sufficient condition of absolute stability of the position equilibrium for all $\bar{\omega}$ frequencies in the range $[0, \pi]$ is performance of the inequality [1]. It is corresponding to discrete system with NP (1) and a stable linear discrete part of the open-loop system (LDP). This inequality after updating by means of the advanced Z-transform according to the technique [1], i.e. replacement of expressions $\exp(j\bar{\omega})$ and the frequency transfer function $W^*(j\bar{\omega}, 0)$ of the open-loop system, (where: $z = \exp(j\bar{\omega})$, $\bar{\omega} = \omega T$, $\forall \bar{\omega} \in [0, \pi]$ and LDP with the transfer function $G^*(z, \varepsilon)$), we shall receive for values in the interval $-1 < z < 1$ and $\varepsilon = 0$ in the next form

$$\operatorname{Re} G^*(z, \varepsilon) + k^{-1} > 0. \quad (2)$$

Let's carry out geometrical research of stability analysis of the NDCS with time delay in complex plane $z = \delta^* + j\omega^*$, ($j^2 = -1$). In case of the stable LDP [1], the root locus of the effective frequency response $W^*(j\bar{\omega}, 0)$, is located completely on the right of the vertical straight line, which is passing through a point located on valid axis of a complex plane of frequencies. We investigated the discrete-time systems with time delay

$$T_d = nT + lT, \quad (5)$$

where T_d is time delay in a direct circuit or the plant (process); T is the sampling period; nT and lT are accordingly the whole and fractional part of delay, $n=0,1,2,\dots$. When $0 \leq l \leq 1$, we applied the Z-transform. In case of when value $T_d \ll T$, i.e. $l \ll 1$ we applied the advanced Z-transform.

Time delay (5) in a direct circuit we considered, included consistently with LCP, an element of pure delay with the transfer function which represent in the form of [4]

$$W_d(p) = k_d \cdot \exp(-nT p) \exp(-lT p). \quad (6)$$

For analysis of absolute stability NDCS with delay in case $T_d \ll T$ we used the transfer function LDP in following:

$$G^*(z) = D(z) W_h W_p(z) W_d(z), \quad (7)$$

where $D(z)$ is the transfer function of discrete filter ; $W_n W_p(z) = Z \{W_n(p) W_p(p)\}$, where Z - is a symbol Z-transform, $z = \exp(T p)$; $W_h(p)$ and $W_p(p)$ are accordingly the transfer functions of the zero-order hold and LCP without taking into account delay; $W_d(z) \approx z^{-(n+1)} [(1-l)z + l]$ is Z-transform (6), and $Z \{ \exp(-lT p) \} = \exp[-l(z-1)z^{-1}] \approx [(1-l)z + 1] \cdot z^{-1}$, i.e. is approximated by a linear member of the Tayloris series. In case $T_d < T$ of the transfer function LDP it is possible to present in the form

$$G^*(z, \varepsilon) = z^{-n} \cdot D(z) \cdot Z_\varepsilon \{W_n(p) W_p(p)\} \Big|_{\varepsilon=1-l},$$

where Z_ε is a symbol of the advanced Z-transform, $\varepsilon = 1-l$, $0 < \varepsilon \leq 1$ where $p = d/dt$. The transfer function LDP of system with the time delay T_d can be written in the form

$$G^*(z, \varepsilon) = G_0^*(z, \varepsilon) \cdot G_d^*(z, \varepsilon), \quad (3)$$

where $G_0^*(z, \varepsilon) = P_0^*(z, \varepsilon) / Q_0^*(z) = \sum_{j=0}^m b_j(\varepsilon) z^{m-j} / \sum_{i=0}^n a_i z^{n-i}$ (4)

is the transfer function LDP without taking into account time delay, $b_j(\varepsilon)$ and a_i are the real constant coefficients, $m \leq n$;

$$G_d^*(z, \varepsilon) = P_d^*(z, \varepsilon) / Q_d^*(z) = \sum_{j=0}^{m'} b'_j(\varepsilon) z^{m'-j} / \sum_{i=0}^{n'} a'_i z^{n'-i} \quad (5)$$

is the transfer function of time delay element, where $b'_j(\varepsilon)$ and a'_i are the real constant coefficients, $m' \leq n'$.

The transfer function LDP (3) can be presented in complex form [2] as

$$G^*(z, \varepsilon) = \frac{P^* + jR^*}{E^* + jF^*} \cdot \frac{A^* + jB^*}{C^* + jD^*} = U^*(\delta^*, \omega^*) + jV^*(\delta^*, \omega^*), \quad (6)$$

where polynomial functions from two independent variables δ^* and ω^* , P^* is $P^*(\delta^*, \omega^*)$, E^* , A^* , C^* - the even degrees, i.e. the real parts accordingly following polynomials $P_0(z, \varepsilon)$, $Q_0(z)$, $P_d(z, \varepsilon)$ and $Q_d(z)$, R^* , F^* , B^* and D^* - the odd degrees, i.e. the imaginary parts set forth above functions, which are identical on structure and, hence, are defined under the following formula for finding of polynomials of the even degrees, i.e. the real parts polynomials of degree \tilde{n} from z ,

$$\tilde{E}_{\tilde{n}}^*(\delta^*, \omega^*) = \sum_{i=0}^N \sum_{j=0}^{\tilde{n}-2i} (-1)^i C_{\tilde{n}-j}^{2i} \tilde{a}_j \omega^{*2i} \delta^{*(\tilde{n}-2i-j)} \quad (7)$$

and for finding of polynomials of the odd degrees, i.e. is the imaginary parts of polynomials of the degree \tilde{n} from z :

$$\tilde{F}_{\tilde{n}}^*(\delta^*, \omega^*) = \omega^* \sum_{i=0}^N \sum_{j=0}^{n-2i-1} (-1)^i C_{\tilde{n}-j}^{2i+1} \tilde{a}_j \omega^{*2i} \delta^{*(\tilde{n}-2i-j-1)}, \quad (8)$$

where C_v^λ - the binomial coefficients; \tilde{n} - the degree of polynomials from z , which can be equal n or m , (n' or m'); \tilde{a}_j - the real coefficients, i.e. a_i or $b_j(\varepsilon)$, (a'_i or $b'_j(\varepsilon)$);

$$N = \begin{cases} \tilde{n}/2 & \text{for even values } \tilde{n}; \\ (\tilde{n}-1)/2 & \text{for odd values } \tilde{n}. \end{cases}$$

We received after transformation (6) in view of (7) and (8)

$$U^*(\delta^*, \omega^*) = \frac{(E^* P^* + F^* R^*)(A^* C^* + B^* D^*) - (E^* R^* - F^* P^*)(B^* C^* - A^* D^*)}{(E^{*2} + F^{*2})(C^{*2} + D^{*2})}. \quad (9)$$

For geometric interpretation of absolute stability of NDCS for the Z-plane we equated the inequality (2)

$$U^*(\delta^*, \omega^*) + k^{-1} = 0. \quad (10)$$

Having substituted expressions (9) and (10), we received the mathematical model NDCS in view of nonlinearity (1) and time delay in the form of the analytical equation, which is the root-locus of vertical straight line (RGVSL) for research in the Z-plane of absolute stability of the position equilibrium:

$$k[(E^*P^* + F^*R^*)(A^*C^* + B^*D^*) - (E^*R^* - F^*P^*)(B^*C^* - A^*D^*)] + (E^{*2} + F^{*2})(C^{*2} + D^{*2}) = 0. \quad (11)$$

Let's considered properties of RGVSL for the NDCS in view of time delay by means of the equation (11) in the following cases:

1) $n > m$ and $n' > m'$. Let values $E^* = 0$, $F^* = 0$, $C^* = 0$, and $D^* = 0$. Then the left part of the equation (11) is identically equal to zero and, hence, the poles of the transfer function LDP located on the trajectory of roots. At values $P^* = 0$, $R^* = 0$, $A^* = 0$, and $B^* = 0$ the left part (11) is not equal to zero and, hence, the zeros $G^*(z, \varepsilon)$ do not located on the trajectory of roots. However, at value $k = \infty$ the zeros located on the trajectory of roots. 2) Values $n = m$ и $n' = m'$. This case is similar to the first case. 3) $m = 0$ и $m' = 0$. The $G^*(z, \varepsilon)$ does not contain zero. At values $E^* = 0$, $F^* = 0$, $C^* = 0$, and $D^* = 0$ the poles of functions LDP $G^*(z, \varepsilon)$ located on trajectory of roots. 4) $n = 0$ and $m' = 0$. The $G^*(z, \varepsilon)$ does not contain the poles. Zero of function $G^*(z, \varepsilon)$ at $P^* = 0$, $A^* = 0$ do not located on trajectory of roots, and at $k = \infty$ zeros of $G^*(z, \varepsilon)$ located on the trajectory of roots.

The geometrical analysis testify, that the branches non-degenerate RGVSL pass through the poles of $G^*(z, \varepsilon)$, and the zeros do not located on trajectories of roots at any value k , except for $k = \infty$ [2].

We have generated on the basis of the general theory [3] the root's condition for geometrical interpretation of absolute stability NDCS: that the position equilibrium in NDCS in view of time delay, and the stable linear discrete part and nonlinearity $\varphi(\sigma)$, satisfying to conditions (1), it was absolute stability in the sector $(0, k)$, there is enough, that RGVSL (11) completely was inside in the circumference of radius 1.0 with the center in the beginning of coordinates of the z - plane [2].

The mathematic models for some classes of continuous-time and discrete-time nonlinear control systems present analytical equations for construction trajectories of the concrete root locus on the complex plane P or Z , correspondingly. In calculations with computer-oriented methods of polynomial with two independent variables were used well-developed mathematical apparatus with the grafs tracing on the PC display. We have developed the method for analysis and the geometrical root criteria-tests for the absolute stability of nonlinear discrete-time control systems with time-delay [2,3].

These methods were applicated for controlling systems with time delay and may be realized in automatic devices [4].

Vibroacoustic protection system intended for voice data protection against leakages via acoustic channels and vibration channels from rooms far beyond the protection zone. The systems can be described as an automatically controlled source of acoustic noise and vibration, to mask voice, mountable into constructional elements of buildings and other possible acoustic channels of voice data leakage (depending on the

voice signal level in a room to be protected). The system generates masking signals, such as white noise, speech-like signals, and white noise + speech-like signals, by virtue of which the voice data leakage channels are shuttered. The speech-like signals are generated by a microprocessor randomly, and correspond with all the formal properties of voice (formant nature of signals, pitch frequency, equal to that of the voice pitch to be masked; pauses between words) and may be adapted to a certain person [7].

References:

- [1] Tsypkin J.Z., Popkov J.S. *Theory of nonlinear pulse systems*. Moscow: Nauka, 1973. 416 p. (in Russian).
- [2] Katernoga O.S., Skudnyakov J.A. *A method in designing digital systems with delay* // Modern vehicles for communication. Proceedings of 3th International Scientific and Technical Conference. Naroch, Belarus. Transactions of the Belarusian engineering academy, No 2 (6) / 1. pp. 23-26, 1998 (in Russian).
- [3] Rimsky G.V. *General theory of root trajectories in control systems*. Belarus, Minsk: Nauka i tehnika, 1972, 328 p. (in Russian).
- [4] Katernoga O.S. Skudnyakov J.A., Rimsky G.V. *Modification the root conditions absolute stability for discrete-time control systems with time delay*. - Modern vehicles for communication. Proceedings of 5th International Scientific and Technical Conference. Naroch, Belarus, Transactions of the Belarusian engineering academy, No 1 (9) / 1. pp. 25-26, 2000 (in Russian).
- [5] Katernoga O.S. Skudnyakov J.A. *A method for analysis absolute stability for sampled-data control systems with time delay*. - Modern vehicles for communication. Proceedings of 9th International Scientific and Technical Conference. Naroch, Belarus. Transactions of the Belarusian engineering academy, No 2 (18) / 1. pp. 33-35, 2004 (in Russian).
- [7] Davydau G.V., Popov V.A., Potapovich A.V. et al. *Protection device of voice data from leak by vibrational and acoustic channels*. Patent of Belarus Republic No 3053, (10.02.2006) (in Belarusian).

Authors:

Dipl.-Ing. Oleg Katernoga
Dipl.-Ing. Vasiliy Popov
Dipl.-Ing. Aleksandr Potapovich
Dr.-tech. Gennadiy Davydau

Belarusian State University of Informatics and Radioelectronics,
6, P.Browka Street,
BY-220013 Minsk,
Belarus
Phone/Fax: +375(17) 293-88-69
E-mail: andrew@nano.bsuir.edu.by

J. Zimmermann / O. Sawodny

Modelling and Control of a X-Y-Fine-Positioning Table

1. ABSTRACT

This contribution is about the modelling, simulation and control of an x-y-high precision positioning table and summarises results from a research work supported by the collaborative research centre SFB 622 'Nanopositionier- und Nanomessmaschinen' of the Technische Universität Ilmenau. The setup provides an operating range of 200x200mm² and is build up in a double-H configuration.

For modelling of the two dimensional system a Langrangian approach, with three mass approximation and couplings between the axes, was utilised. Additionally the friction of the four guides was implemented using the elasto-plastic friction model [3,4]. Since the normal force of the guides of the x-axis depends on the position of the y-axis, the friction model also must provide these properties. Therefore the model was adapted according to the results of an unequally stressed guide [8].

The control task is high precision path-tracking and fine positioning of multiple axes for coordinated movements. Therefore a PI state feedback controller with feedforward path is compared with a robust double-integrator control scheme [7] using the developed simulation model.

2. X-Y-POSITIONING TABLE

The setup which motivates this work was constructed by Steffen Hesse, member of the collaborative research centre SFB 622. It provides an operating range of 200x200mm² and is build up in a double-H configuration, which means there are two guides and drives for each axis. Both axes are coupled by a torsional stiffness of the two guides of each axis and a flexural stiffness between the two axes, which is modelled by spring-damper-elements.

Furthermore the accelerated mass of both guides of the x-axis depends on the position of the y-axis. This also involves the normal force and thereby the friction force of the guides, which is handled in section 3. Positioning of the tool-center-point with parallel feed of the axes is the main control task, which means the y-axis can be simplified to only one motor and guide for the Langrangian approach. Thus a three-mass model is developed, compare Fig. 1.

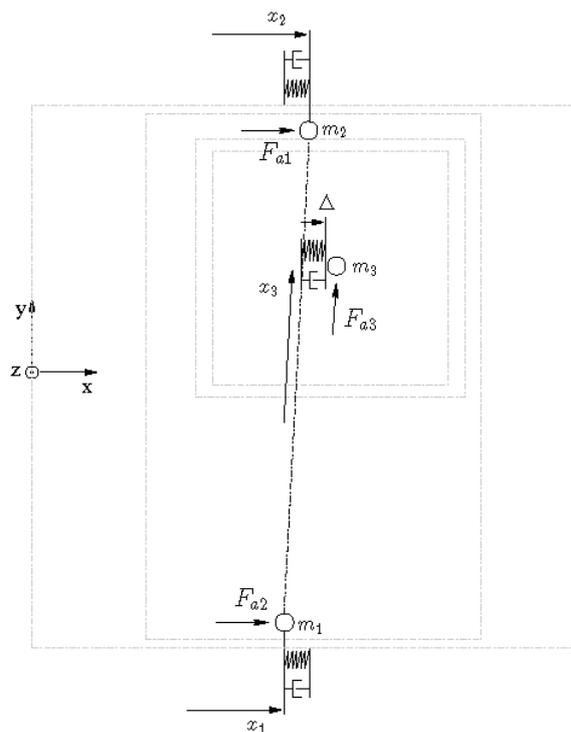


Fig. 1. x-y-table visualisation

Centres of gravity of the axes and locations of the corresponding motors and guides are arranged such that torsional movements are reduced, so there is no need to model the z-coordinates of the system. Additionally the following assumptions have been made:

- all masses are combined to three time dependent point masses
- friction is handled as an external force
- there is no torque causing torsional stress
- there is no bowing under load of the carrier

The resulting differential equations are derived from Mathematica® and implemented in MATLAB/Simulink® for simulation purposes. Together with the friction model from section 3 a state space representation is developed in section 4.

3. FRICTION MODEL

The Coulomb friction model (1) is well known and reflects the influence of the friction coefficient μ , which can depend on velocity, and the normal force between the two surfaces. When there is no movement the friction force is equal to the affecting one.

$$F_f = \text{sgn}(\dot{x}) \cdot F_c = \text{sgn}(\dot{x}) \cdot \mu \cdot F_N \quad \dot{x} \neq 0 \quad (1)$$

Since the model can not reflect pre-sliding effects, which are important in nanometer scale, the elasto-plastic model [4,5] based on the bristle idea [6] and similar to the LuGre model [2,3] is utilised. The basic model is given by equations (2) and (3); equations (4) and (5) define the transition from elastic to mixed elasto-plastic movement. Equation (6) reflects the influence of the velocity which serves as a good method to adapt the sliding friction, but can hardly be proven since the maximum displacement of the bristles can hardly be evaluated during movements.

The influence of the normal load on a high-precision linear guide, equipped with a laser interferometer, was observed in [8]. As a result equation (6) is expanded by position and load dependent friction forces. The friction coefficients σ_0 and σ_1 could also depend on z.

$$F_f = \sigma_0 \cdot z + \sigma_1 \cdot \dot{z} + \sigma_2 \cdot \dot{x} \quad (2) \quad \dot{z} = \dot{x} \cdot \left(1 - \alpha(z, \dot{x}) \frac{z}{z_{ss}(\dot{x})} \right) \quad (3)$$

$$\alpha(z, \dot{x}) = \begin{cases} 0 & |z| \leq z_{ba} \\ 0 < \alpha_m < 1 & z_{ba} < |z| < z_{ss}(\dot{x}), \text{sgn}(\dot{x}) = \text{sgn}(z) \\ 1 & |z| \geq z_{ss}(\dot{x}), \text{sgn}(\dot{x}) = \text{sgn}(z) \\ 0 & \text{sgn}(\dot{x}) \neq \text{sgn}(z) \end{cases} \quad (4) \quad \alpha_m(z) = \frac{1}{2} \sin \left(\pi \frac{z - \frac{z_{ss} + z_{ba}}{2}}{z_{ss} - z_{ba}} \right) + \frac{1}{2} \quad \text{with} \quad z_{ba} \leq |z| < z_{ss} \quad (5)$$

$$z_{ss}(x, \dot{x}) = \frac{\text{sgn}(\dot{x})}{\sigma_0} \left[F_C(x, F_N) + (F_S(x, F_N) - F_C(x, F_N)) e^{-\left(\frac{\dot{x}}{v_s}\right)^2} \right] \quad (6)$$

The pre-sliding domain can support fine positioning due to the large stiffness which decreases from its maximum in rest-position, but this also limits the dynamics due to the limited rate of force of the actuator. In the sliding domain there are small apparently random changes of the friction force during larger movements and velocities, which disturb exact path tracking especially with low velocities.

4. CONTROL

Control objectives are fine positioning, disturbance rejection, and exact path tracking as a basis for coordinated movements of multiple axes. Inputs of the system are references of actuator currents of the current controllers. The actuator force is proportional to the current, despite dynamics of the actuator and the influence of temperature and changes in the magnetic field. Measurements are only the x-y-positions and angles of the corner mirror located on the y-axis, so the other states must be reconstructed.

The path tracking properties of a PI state feedback controller with feed-forward path are compared with a robust double integrator controller according to [7]. All experiments are carried out with the simulation model according to sections 2 and 3 with respect to the real system inputs and outputs. The simulation model therefore contains also actuator and measurement dynamics including amplifier and measurement noise.

PI-State Feedback Regulator

Based on the mechanical model from section 2 and the linearised sliding friction model from section 3, a state space model is developed. Due to the fact that both axes are coupled not very strongly compared to the other dynamics, we here propose a decentralized control of both axes and present exemplary the x-control. The influence of the moving y-axis is taken into account as time-varying parameter, which is possible because the rate of change is very small compared to the other dynamics.

Equation (7) describes the influence of the y-position on the accelerated mass, while Equation (8) shows the overall system with the position of the tool-center-point x_{tcp} and the parallel position error δx as outputs of the system.

$$m_{E1}(x_3) = m_1 + \left(1 - \frac{x_3}{l}\right)m_3 \quad m_{E2}(x_3) = m_2 + \frac{x_3}{l}m_3 \quad (7)$$

$$\begin{bmatrix} \dot{x}_1 \\ \ddot{x}_1 \\ \dot{x}_2 \\ \ddot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -\frac{c_{12}}{m_{E1}} & -\frac{\sigma_{2,1}}{m_{E1}} & \frac{c_{12}}{m_{E1}} & 0 \\ 0 & 0 & 0 & 1 \\ \frac{c_{12}}{m_{E2}} & 0 & -\frac{c_{12}}{m_{E2}} & -\frac{\sigma_{2,2}}{m_{E2}} \end{bmatrix} \begin{bmatrix} x_1 \\ \dot{x}_1 \\ x_2 \\ \dot{x}_2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \frac{k_F}{m_{E1}} & 0 \\ 0 & 0 \\ 0 & \frac{k_F}{m_{E2}} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \quad (8)$$

$$\underline{y} = \begin{bmatrix} x_{tcp} \\ \delta x \end{bmatrix} = \begin{bmatrix} 1 - \frac{x_3}{l} & 0 & \frac{x_3}{l} & 0 \\ 1 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ \dot{x}_1 \\ x_2 \\ \dot{x}_2 \end{bmatrix}$$

Since the tool-center-point must be controlled, position and angle control around this point can be split into separate control tasks. Equations (9) state the calculation of the tool-center-point coordinates and the corresponding differential equation. The influence of the y-axis is handled as time-varying parameter, since the positioning distances are small compared to the operating range, and the system dynamics is fast compared to the rate of change of the x_3 position.

The state space model of the PI-state regulator is given by equations (10). Therefore we assume the integral angle controller to be affective, leading to perfect parallel feed.

$$\begin{aligned} x_{tcp} &= \left(1 - \frac{x_3}{l}\right)x_1 + \frac{x_3}{l}x_2 \\ \ddot{x}_{tcp} &= \left(1 - \frac{x_3}{l}\right)\ddot{x}_1 + \frac{x_3}{l}\ddot{x}_2 \end{aligned} \quad (9)$$

$$\begin{bmatrix} \dot{x}_{tcp} \\ \ddot{x}_{tcp} \\ \dot{e}_{tcp} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -\frac{\sigma_{2,tcp}}{m_{E1}+m_{E2}} & 0 \\ -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_{tcp} \\ \dot{x}_{tcp} \\ e_{tcp} \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{k_F}{m_{E1}+m_{E2}} \\ 0 \end{bmatrix} (u_1 + u_2) \quad (10)$$

$$y = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_{tcp} \\ \dot{x}_{tcp} \\ e_{tcp} \end{bmatrix}$$

Figure 2 shows the state space model together with the control system, which basically consists of a reference generator, a feed forward part based on the reference trajectory, a parallel feed controller with an integrative part only, and a PI state feedback regulator based on the simplified system. The reference trajectory is smooth with respect to input and state constraints. Additionally the internal variable z is planned based on initial states and the reference trajectory.

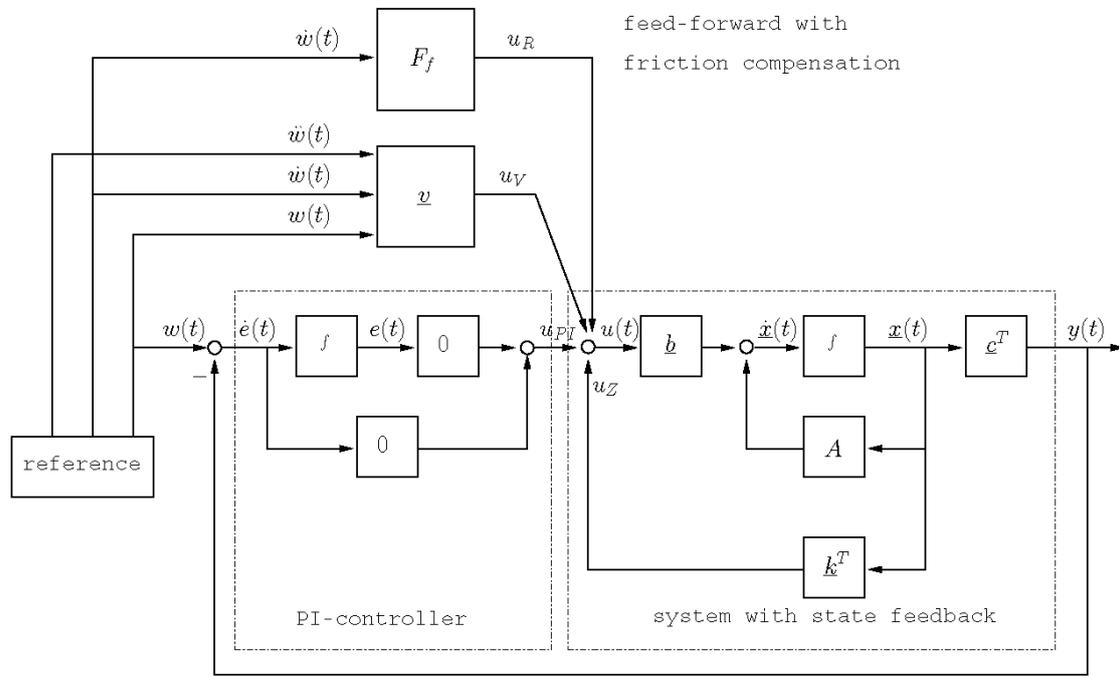


Fig. 2. State space model with PI-controller

Simulation results show that the PI-controller works well for fast direct movements due to the feedforward path and an integrative feedback. Especially when moving with small velocities the error dynamic is predictable well. For example a circle can be followed with a constant time delay leading to a synchronized movement with good accuracy. Figure 3 shows the positions, errors and feedback-forces for a circle with a diameter of $10\mu\text{m}$ and a movement time of around 750ms . The position errors are below $300\text{nm}/200\text{nm}$ and decrease to below $50\text{nm}/30\text{nm}$ outside direction-reversal domains.

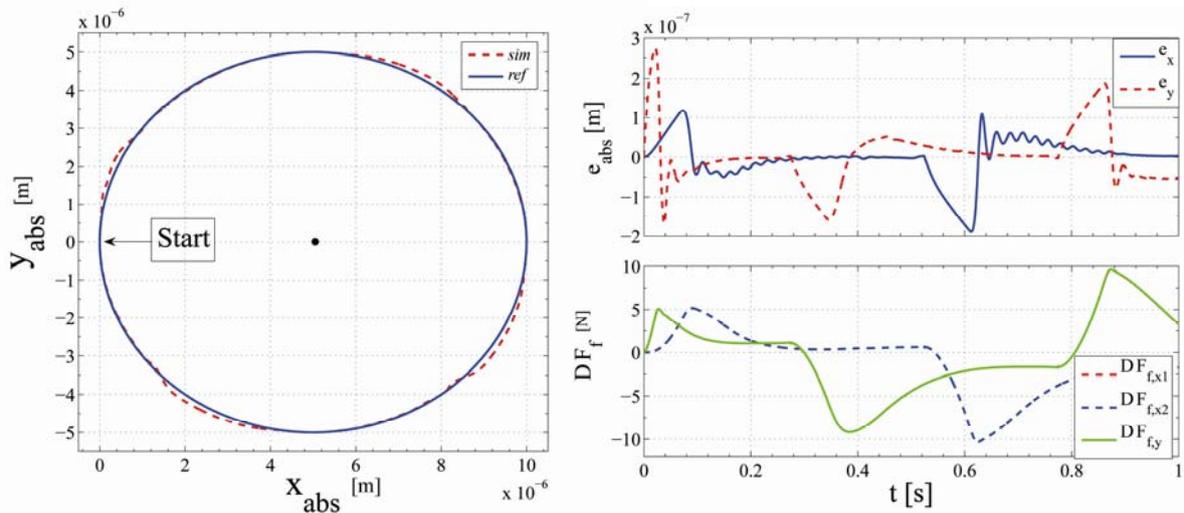


Fig. 3. Simulation results PI-state controller

Double Integrato Scheme

The double integrator controller was developed according to [7] and is shown in figure 4. Equation (11) denotes one simplified SISO axis with respect to time varying parameters and the total weight m of the axis. The calculated actuator force is divided into the two motor forces depending on the position of the y-axis by a decoupling feedforward.

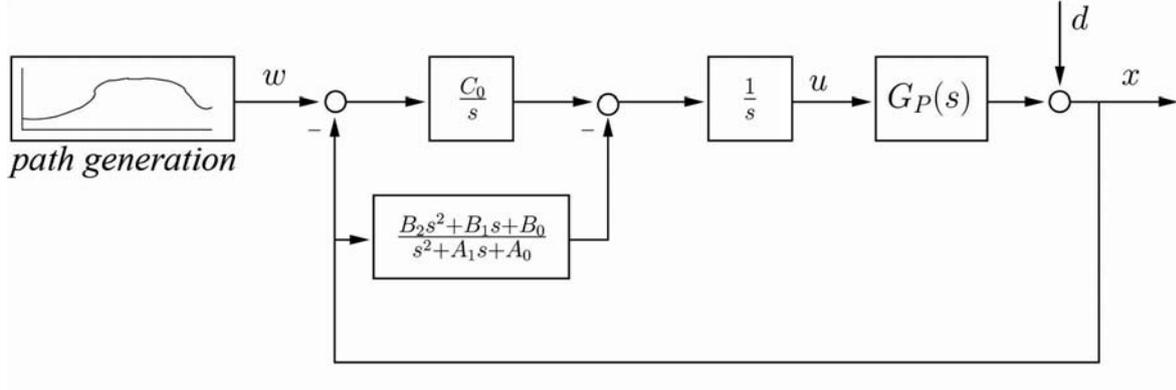


Fig. 4. Double integrator scheme [7]

$$G_p(s) = \frac{X_{icp}(s)}{U_1(s) + U_2(s)} = \frac{\frac{1}{m}}{s^2 + \frac{\sigma_z}{m}s} \quad (11)$$

$$G_{DI}(s) = \frac{\frac{1}{m}C_0(s^2 + A_1s + A_0)}{s^6 + \left(\frac{\sigma_z}{m} + A_1\right)s^5 + \left(A_0 + \frac{\sigma_z}{m}A_1\right)s^4 + \left(\frac{\sigma_z}{m}A_0 + \frac{1}{m}B_2\right)s^3 + \frac{1}{m}(B_1 + C_0)s^2 + \frac{1}{m}(A_1C_0 + B_0)s + \frac{1}{m}A_0C_0} \quad (12)$$

$$G_{DI,p}(s) = \frac{\frac{1}{m}C_0(s^2 + A_1s + A_0)}{(s + p)^6} \quad (13)$$

Equation (12) denotes the transfer function of the controlled system according to figure 4. A coefficient comparison with equation (13) results in equations (16) which are used to calculate the controller parameters online with the given poles p .

$$\begin{aligned} A_0 &= 15p^2 - 6\frac{\sigma_z}{m}p + \left(\frac{\sigma_z}{m}\right)^2 \\ A_1 &= 6p - \frac{\sigma_z}{m} \\ B_0 &= \frac{84p^7 - 35\frac{\sigma_z}{m}p^6 + 6\left(\frac{\sigma_z}{m}\right)^2p^5}{\frac{1}{m}\left(15p^2 - 6\frac{\sigma_z}{m}p + \left(\frac{\sigma_z}{m}\right)^2\right)} \\ B_1 &= \frac{224p^6 - 90\frac{\sigma_z}{m}p^5 + 15\left(\frac{\sigma_z}{m}\right)^2p^4}{\frac{1}{m}\left(15p^2 - 6\frac{\sigma_z}{m}p + \left(\frac{\sigma_z}{m}\right)^2\right)} \\ B_2 &= \frac{20p^3 - 15\frac{\sigma_z}{m}p^2 + 6\left(\frac{\sigma_z}{m}\right)^2p - \left(\frac{\sigma_z}{m}\right)^3}{\frac{1}{m}} \\ C_0 &= \frac{p^6}{\frac{1}{m}\left(15p^2 - 6\frac{\sigma_z}{m}p + \left(\frac{\sigma_z}{m}\right)^2\right)} \end{aligned} \quad (14)$$

Simulation results show that the double integrator controller works well for slow complex movements due to the high feedback gain for errors with low frequency. On the other hand measurement noise can be amplified easily and there is no feedforward path. Especially when moving with small velocities the tracking error is around the measurement noise. After a short initialisation phase of around 30ms nearly all smooth reference trajectories with respect to input constraints can be tracked with a small error. Figure 5 shows the positions, errors and friction forces for a circle with a diameter of $10\mu\text{m}$ and a movement time of around 750ms. The position errors are only below 500nm/300nm due to the high velocity, but decrease to below 10nm or even 5nm outside direction-reversal domains.

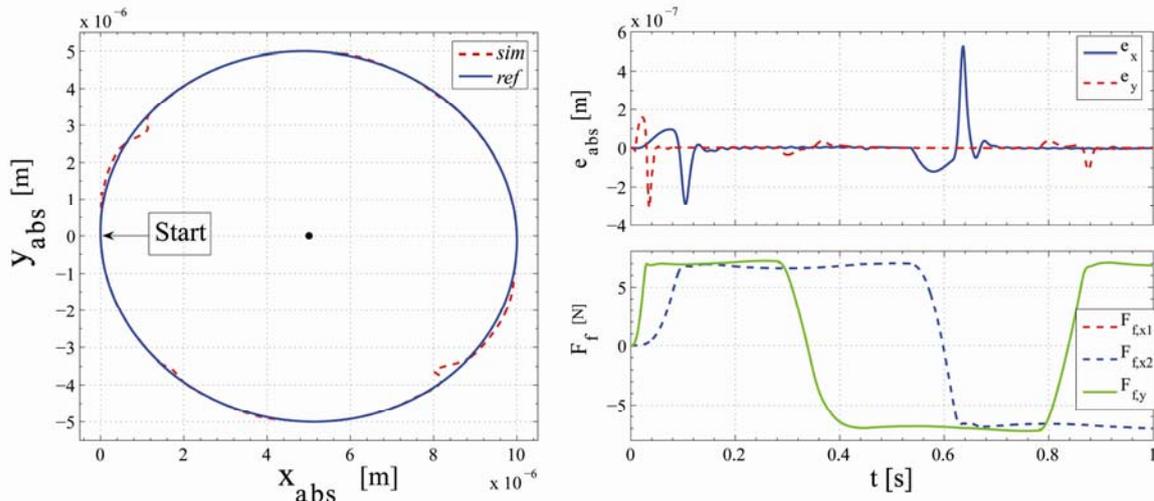


Fig. 5. Simulation results double integrator controller

4. CONCLUSIONS

In this paper a x-y-fine positioning table was modelled for simulation and control purposes. The friction model is based on former experiments with similar systems. In order to achieve exact path tracking of both axes the PI-state feedback regulator and the double integrator scheme were implemented. Finally both concepts are compared by using the developed simulation model. The PI controller is very fast and can be tuned well to achieve fast and exact fine positioning, which is ideal for positioning and step-height-measurements. The double integrator scheme is much more robust to low frequency disturbances and ideal for complex movements, which is ideal for scanning movements and three-dimensional tracking. In a next step the developed controllers will be tested with the experimental setup.

This work was supported by the DFG with the collaborative research center SFB622 'Nanopositionier- und Nanomessmaschinen' and with the SPP1159 'New strategies of measurement and testing technology for the production of micro-system and nano-structures'. The authors would like to thank all those colleagues who have contributed to the developments described.

References:

- [1] Hausotte, T. (2002). *Nanopositionier- und Nanomessmaschine*. Dissertation, Technische Universität Ilmenau
- [2] Åström, K. J.; Canudas de Wit, C.; Olsson, H.; Gäfvert, M.; Lischinsky, P. (1998). Friction Models and Friction Compensation. *European Journal of Control*, Vol. 29(4), pp. 176- 195
- [3] Canudas de Wit, C.; Olsson, H.; Åström, K. J.; Lischinsky, P. (1995). A new Model for Control of Systems with Friction. *IEEE Transactions on Automatic Control*, Vol. 40(3), pp. 419-425
- [4] Dupont, P.; Armstrong, B.; Hayward, V.; Altpeter, F. (2002a). Single State Elastoplastic Friction Models. *IEEE Transactions on Automatic Control*, Vol. 47(5), pp.787-792
- [5] Dupont, P.; Armstrong, B.; Hayward, V. (2002b). Single State Elastoplastic Friction Models for Friction Compensation. *IEEE Transactions on Automatic Control 2001*, Article TP99-255 Revision 2002
- [6] Haessig, D.; Friedland, B. (1990). On the modelling and simulation of friction. *Proceedings of the ACC San Diego 1990*, pp. 1256-1261
- [7] Mao, Junhong; Tachikawa, Hiroyuki et al. Double-Integrator control for precision positioning in the presence of friction, Tokio Institute of Technology, 2003
- [8] Zimmermann J., Sawodny O., Hausotte T., Jäger G.: *Friction Modelling for Control of a Linear High-Precision Actuator*, Mechatronics 2006 Heidelberg

Authors:

Jan Zimmermann
 Prof. Dr. Oliver Sawodny
 Universität Stuttgart, Institute for System Dynamics, Pfaffenwaldring 9
 70569 Stuttgart
 Phone: 0711-685-66295
 Fax: 0711-685-66371
 E-mail: jan.zimmermann@isys.uni-stuttgart.de, oliver.sawodny@isys.uni-stuttgart.de

A. Winkler / J. Suchý

Position Based Force Control of an Industrial Manipulator

INTRODUCTION

Force control of articulated robots is a research field quite often investigated during the last thirty years, [1, 2, 3, 4]. However, industrial applications of force controlled robots are very rare. Reasons for this may be the cost of a six component force/torque wrist sensor (F/T sensor) which is necessary in robot force control or the inadequate integration between the F/T sensor and the robot controller.

This paper shows that it is possible to implement high efficient force control algorithms into a commercial robot controller. For this purpose a very sensitive task has been chosen. Robot task is to draw a figure on a blackboard using a piece of chalk. However, the exact position of the blackboard and its inclination with respect to the ground are unknown. The improved force controllers developed for this compliant motion task are explained in this paper. The results, especially the drawing time and the drawing velocity, are compared with results attained by standard force control algorithms available using the commercial technology packet.

SYSTEM DESCRIPTION AND EXPERIMENTAL SETUP

The development of the improved approach to force control of an industrial manipulator was performed using the robot Kuka KR6/2. It is a six axes articulated robot with payload of 6 kg. The robot is controlled by the Kuka Robot Controller KRC2. A six component F/T sensor is mounted in the robot wrist. Because the force/torque measurement is based on PSD diodes the F/T sensor includes an appreciable compliance. The F/T sensor is connected with the robot controller via DeviceNet.

Using the robot in industrial standard applications like handling, the programming is performed with the Kuka Robot Language (KRL). However, using KRL real time robot control is not realizable. Therefore an additional module called Robot Sensor Interface (RSI) is necessary. With this module it is possible to generate and link RSI objects using

special KRL expressions. These objects provide sensor values, perform signal processing, and influence robots motion. A complex controller structure consisting of RSI objects can be built and used for sensor guided robot motion. It is executed in real time with the interpolation cycle of 12 ms. Robot motion can be influenced by RSI on the level of the position control loops. The functional scheme of RSI is shown in Fig. 1.

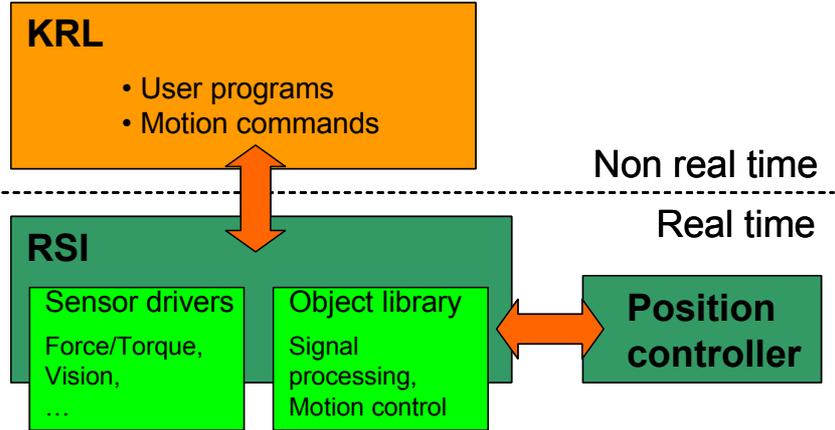


Fig. 1: Functional scheme of RSI.

To perform the drawing example task the robot is equipped with a two finger parallel gripper to hold a piece of chalk. The blackboard on which the figure has to be drawn is inclined with respect to the ground. Its exact position and the angles of inclination are unknown to the robot. Fig. 2 shows the experimental setup for the compliant motion task.

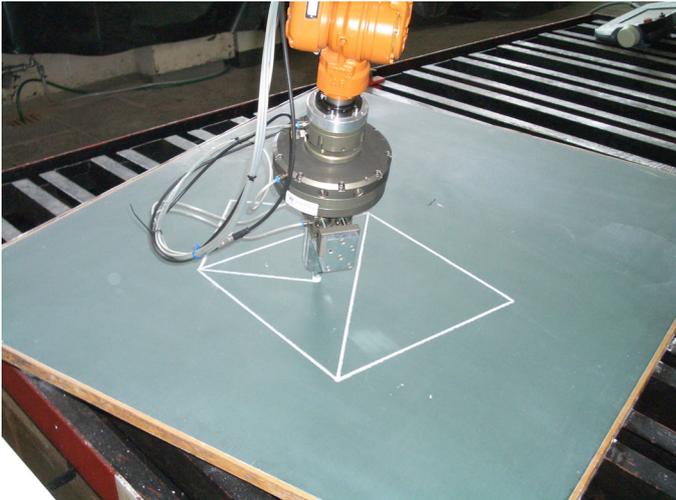


Fig. 2: Experimental setup.

The stiffness of the whole environment including the compliances of robot load limiter and F/T sensor is approximately 100 N/mm. To perform the drawing task two different

force controllers have been designed. The first controller is used to bring the tool (piece of chalk) in contact with the environment (blackboard). The second controller has to keep the stable and safe contact while drawing.

DESIGN OF IMPACT CONTROLLER

First, the end-effector has to get into contact with the environment. The exact position of the environment is unknown. An overshoot of the contact force should be avoided to beware of the damage of the tool (chalk). The controller is a variable structure controller based on a proportional plus integral controller. Its signal flow diagram is shown in Fig. 3.

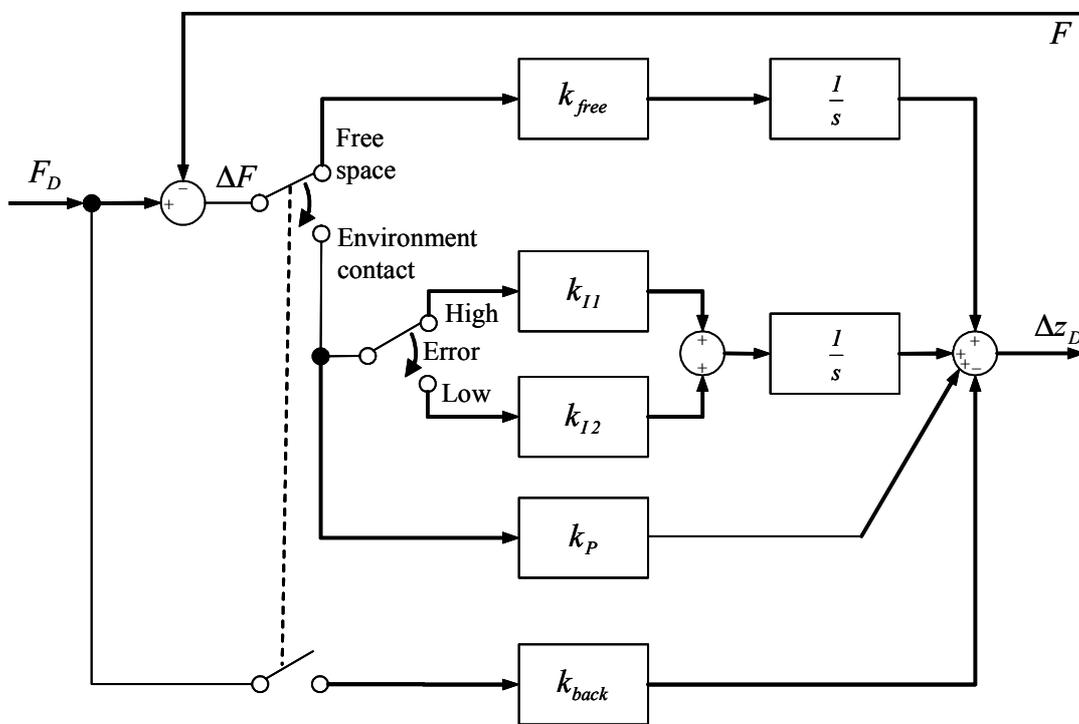


Fig. 3: Controller structure for impact control.

Controller inputs are the desired and the current contact force. The output signal is the desired end-effector position which is provided to the position control loops. If the contact force is zero, that means the end-effector is located in free space, only one integral branch is activated. Its controller gain is represented by factor k_{free} . During the environment contact the proportional part k_p starts functioning. The integral branch is switched between controller gains k_{I1} and k_{I2} for high and low force control error, respectively. This feature is necessary to obtain a fast and stable control behavior. To accelerate the contact detection process the factor k_{free} may be increased. To prevent

contact force overshoots the additional branch presented by controller gain k_{back} has to be integrated into the controller structure.

CONTROLLER DESIGN FOR CONTOUR FOLLOWING

After the impact process is finished the controller structure has to be changed to perform the contour following task (drawing). The signal flow diagram of the controller used for the drawing task is shown in Fig. 4.

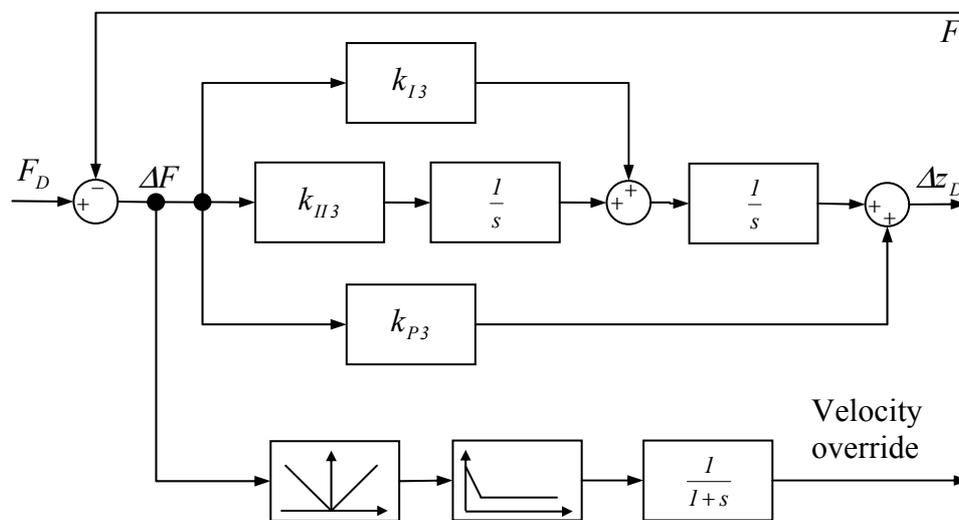


Fig. 4: Controller structure for the contour following task.

It consists of a proportional plus integral controller with controller gains k_{P3} and k_{I3} . Additional integral part is inserted in the control loop represented by factor k_{II3} , which results in the I^2 -behavior. With this functionality it is possible to adjust the controller to the particular environment. This adaptation may be seen as the estimation of the environment orientation.

To draw the figure on the blackboard the robot is commanded to move its end-effector in parallel to ground (x-y-plane). The contact force is controlled in z-direction by the controller shown in Fig. 4. The velocity of the motion command is controlled by the so called override. To accelerate the drawing process and to guaranty the safe and stable contact between blackboard and chalk an additional feature is integrated in the controller. It is the modulation of the velocity override in dependency on the force control error. In the case of a high control error the override has been reduced. After the inclination of the particular drawing direction is adapted by the I^2 -functionality and the

control error is decreased the velocity override can be increased.

RESULTS AND CONCLUSION

The two force controller structures described in the previous sections were implemented in the robot controller. Some supporting positions which represent the example figure to be drawn (see also Fig. 2) were taught in parallel to the ground without environment contact.

After starting the whole drawing program the robot moves its end-effector to supporting position No. 1. The controller structure for impact control is activated and the desired contact force F_d is set to the value of 10N. After the end-effector is in contact with the environment F_d is reduced to the value of 5N and the controller structure is change to contour following controller. Then the robot is commanded to move its end-effector to supporting position No. 2 using the path located in parallel to the ground. Because of the blackboard inclination, the end-effector coordinate in z-direction has to be adjusted to avoid the loss of contact or the damage of the chalk.

Reaching taught supporting location while drawing, the orientation of the end-effector has to be changed for the symmetrical wear of the chalk. The orientation change is performed in free space. It is followed by new contact detection.

The same compliant motion task was programmed using Kuka's Force Torque Control Technology Package (FTCtrl). It is also based on RSI programming. However, RSI is hidden in a comfortable Windows dialog and its functionality is restricted. The controller and its parameters were chosen to guarantee the stability during contact detection and contour following.

Fig. 5 shows the contact forces while drawing using the improved force controller presented in this paper and the standard force controller using FTCtrl. It can be seen that the whole drawing task was finished after 115s by the improved RSI force controller. Using FTCtrl the process took approximately 50% more time. It is finished after 175s. Because the orientation change and contact detection takes the same time in both controllers the time advantage while contour following (drawing) is approximately 100%. It is possible to calculate the average drawing velocity of 6 cm/s using the improved RSI force controller. For FTCtrl the average drawing velocity of 3cm/s can be determined. With the approach to an improved force controller which has been presented here it was able to shown that it is possible to realize such demanding tasks based on force

controlled commercial robot. The sensitive drawing task was chosen as an example because it has some relations to some industrial manufacturing tasks like polishing, grinding, or deburring.

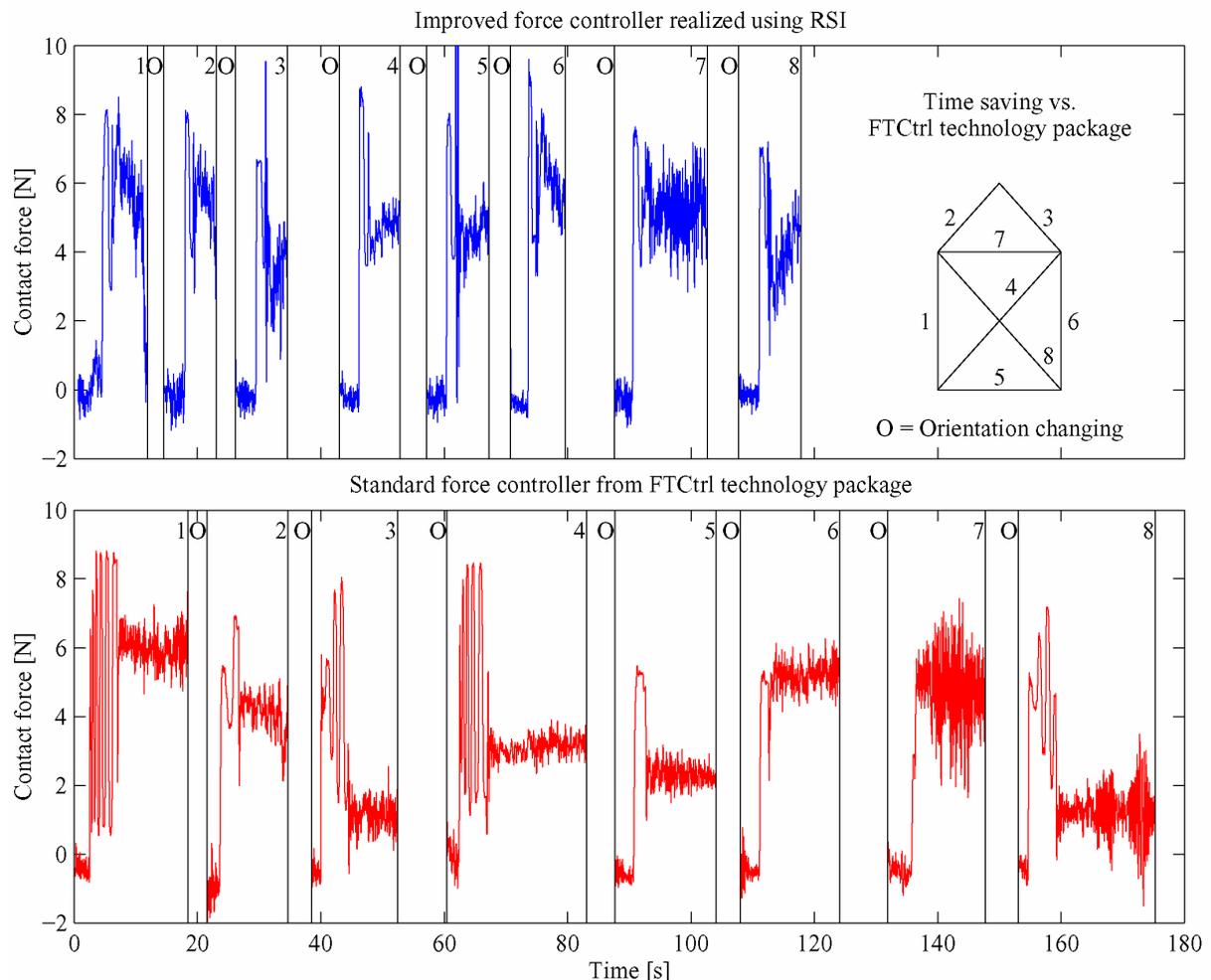


Fig. 5: Contact forces while drawing.

References:

- [1] N. Hogan: Impedance control, an approach to manipulation: Part i, ii, iii. Transactions of ASME Journal of Dynamic Systems, Measurement and Control, Vol. 107, 1-24, 1985.
- [2] M. T. Mason: Compliance and Force Control for Computer Controlled Manipulators. IEEE Transaction on Systems, Man. and Cybernetics, Vol. 11, No. 6, 418-432, 1981.
- [3] M. H. Raibert, J. J. Craig: Hybrid Position/Force Control of Manipulators. ASME Journal of Dynamic Systems, Measurement and Control, Vol. 102, No. 2, 126-133, 1981.
- [4] G. Zeng, A. Hemami: An overview of robot force control. Robotica, Vol. 15, 473-482, 1997.

Authors:

Dr.-Ing. Alexander Winkler
 Prof. Dr.-Ing. Jozef Suchý
 Chemnitz University of Technology
 09107 Chemnitz
 Phone: 0371 / 531-33495, -33427
 Fax: 0371 / 531-24139
 E-mail: alexander.winkler@e-technik.tu-chemnitz.de, jozef.suchy@etit.tu-chemnitz.de

E. Arnold / J. Neupert / O. Sawodny / K. Schneider

Trajectory Tracking for Boom Cranes Based on Nonlinear Control and Optimal Trajectory Generation

Abstract

The main objectives of crane automation are increasing the efficiency and safety of the transshipment processes. Therefore, advanced control strategies are applied for load sway reduction and trajectory tracking. The paper presents a nonlinear control strategy combined with a model-based optimal trajectory generation for the radial load movement of a boom crane. The results are validated by measurement results from a LIEBHERR harbor mobile crane.

1 Introduction

The paper addresses the problem of trajectory tracking and disturbance rejection for the transportation of crane loads. Advanced control systems are developed to meet the demands for fast, efficient, and safe transshipment of cargo in harbors, for example the CYCOP-TRONIC system (Sawodny et al. [1], [2], Arnold et al. [3], Neupert et al. [4]) provided by the company 'Liebherr Werk Nenzing' as anti-sway control for harbor mobile cranes (see Fig. 1). These boom cranes are characterized by a load capacity of up to 140t, a maximum outreach of 48m and a rope length of up to 80m.

Because of the dominant nonlinearities of the dynamic system, the accurate tracking of the crane load on the desired reference trajectory during luffing motion is a challenge for the control system. The nonlinear model of the crane dynamics for the radial movement excluding the influence of the centrifugal forces is derived and discussed in Neupert et al. [5]. Further on, the design of the flatness based controller is presented in detail. Another challenge is the generation of these reference trajectories that have to take into account the system dynamics as well as input and state constraints.

The theoretical foundation for the design of control structures for nonlinear systems and their analysis was introduced in numerous publications. Isidori et al. [6], [7] for example consider asymptotic output tracking of a certain class of nonlinear systems, where the reference or disturbance signals are generated by an exosystem. To calculate the feedforward trajectory partial differential equations are solved. Fliess et al. [8] discuss the differential flatness of nonlinear systems.



Figure 1: LIEBHERR harbor mobile crane (LHM).

They formulate the major property of differential flatness and propose the defect of a nonlinear system as a non-negative integer, which measures the distance from flatness.

Fliess et al. [9] study a two-dimensional overhead crane as an application example of the nonlinear control. The system is characterized as a differential flat system by deriving a linearizing output. Other publications from Piazzzi et al. [10] and Yanai et al. [11] are also focussed on the inversion based control of overhead cranes. This is why cranes are a typical example of an underactuated mechanical system with oscillatory behavior. Kiss et al. [12] show differential flatness for a class of cranes including overhead and rotary cranes.

Most of these contributions do not consider the actuator dynamics and kinematics. In case of a boom crane, which is driven by hydraulic actuators, the dynamics and kinematics of the hydraulic actuators are not negligible. Especially for the boom actuator (hydraulic cylinder) the kinematics have to be taken into account. The resulting nonlinear model for the luffing motion is derived in section 2. Based on the nonlinear model, the flatness based control approach is presented in section 3. It is shown, that a flat output can be found and a linearizing and stabilizing control law can be obtained.

The application of flatness based control methods requires sufficiently smooth reference trajectories that have to be feasible with respect to the input and state constraints of the system. For the tracking problem under consideration, the update of the reference trajectories requires the current state of the system thus forming an additional feedback loop.

An usual approach for on-line trajectory generation is the parameterization of the output and output derivatives profiles by stage-wise low-order polynomials. The coefficients of the polynomials are determined by the boundary values and bound constraints of the variables, see e. g. [13]. This can be interpreted as an approximate solution of a suitable optimal control problem. Because of the necessary increasing degree of the polynomials, this approach is limited to lower order derivatives of the output.

In this paper, a different approach is used. The flatness based controller linearizes and stabilizes the system. An optimal control problem is formulated and solved online to generate feasible reference trajectories for the linearized system including the state feedback. The reference trajectories take into account the current state of the system, therefore this outer feedback loop can be considered as a model predictive control (MPC) loop, see [14]. The formulation and the numerical solution of the optimal control problem is discussed in section 4.

The optimal trajectory generation and flatness based control for the luffing motion is applied to the LIEBHERR harbor mobile crane (LHM). The obtained measurement results are presented in section 5. In section 6 concluding remarks are given.

2 Nonlinear model of the crane

The performance of the crane's control is mainly measured by fast damping of load sway and exact tracking of the reference trajectory. To achieve these control objectives the dominant nonlinearities have to be considered in the dynamic model of the luffing motion.

The first part of this model is derived by the method of NEWTON-EULER considering the load as a point mass and neglecting the mass and elasticity of the rope as well as coriolis and centripetal terms. This results in the following differential equation which characterizes the radial load sway.

$$\ddot{\varphi}_{Sr} + \frac{g}{l_S} \sin \varphi_{Sr} = \frac{\cos \varphi_{Sr}}{l_S} \ddot{r}_A \quad (1)$$

As shown in Fig. 2, φ_{Sr} is the radial rope angle, $\ddot{\varphi}_{Sr}$ the radial angular acceleration, l_S the rope length, r_A the distance from the vertical axis to the end of the boom, \ddot{r}_A the radial acceleration of the end of the boom, m_L the mass of the load and g the gravitational constant.

The second part of the nonlinear model is obtained by taking the kinematics and dynamics of the actuator into account. This actuator is a hydraulic cylinder attached between tower and boom. Its dynamics can be approximated with a first order system

$$\ddot{z}_{zyl} = -\frac{1}{T_W}\dot{z}_{zyl} + \frac{K_{VW}}{T_W A_{zyl}}u_l \quad (2)$$

where \ddot{z}_{zyl} and \dot{z}_{zyl} are the cylinder acceleration and velocity respectively, T_W the time constant, A_{zyl} the cross-sectional area of the cylinder, u_l the input voltage of the servo valve and K_{VW} the proportional constant of flow rate to u_l . In order to combine equation (1) and (2) they have to be in the same coordinates. Therefore a transformation of equation (2) from cylinder variables (z_{zyl}) to outreach variables (r_A) with the kinematical equation

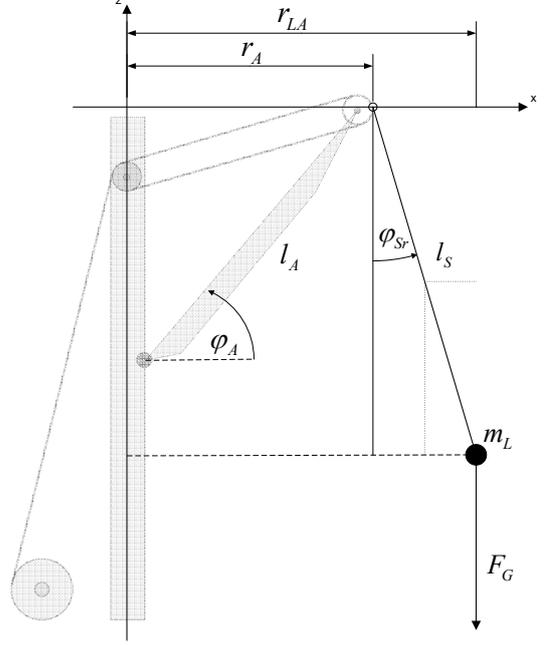


Figure 2: Radial movement of the load.

$$r_A(z_{zyl}) = l_A \cos \left(\alpha_{A0} - \arccos \frac{d_a^2 + d_b^2 - z_{zyl}^2}{2d_a d_b} \right) \quad (3)$$

and its derivatives

$$\begin{aligned} \dot{r}_A &= -l_A \sin \varphi_a K_{Wz1}(\varphi_A) \dot{z}_{zyl} \\ \ddot{r}_A &= -l_A \sin \varphi_a K_{Wz1}(\varphi_A) \ddot{z}_{zyl} - K_{Wz3} \dot{z}_{zyl}^2 \end{aligned} \quad (4)$$

is necessary. The dependency from the geometric constants d_a , d_b , α_1 , α_2 and the luffing angle φ_A is substituted by K_{Wz1} and K_{Wz3} . The geometric constants, the luffing angle and l_A , which is the length of the boom, are shown in Fig. 3.

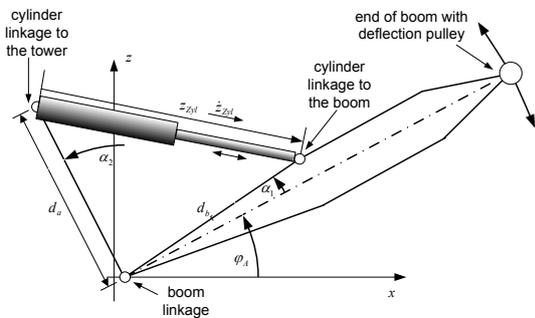


Figure 3: Mounting of the hydraulic cylinder.

As result of the transformation, equation (2) can be given in outreach coordinates.

$$\begin{aligned} \ddot{r}_A &= -\underbrace{\frac{K_{Wz3}}{l_A^2 \sin^2 \varphi_A K_{Wz1}^2}}_a \dot{r}_A^2 - \underbrace{\frac{1}{T_W}}_b \dot{r}_A \\ &\quad - \underbrace{\frac{K_{VW} l_A \sin \varphi_A K_{Wz1}}{T_W A_{zyl}}}_{m} u_l \end{aligned} \quad (5)$$

In order to obtain a nonlinear model in the input affine form

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x})u_l, \quad y = h(\mathbf{x}) \quad (6)$$

equations (1) and (5) are used. With the states defined as $\mathbf{x} = [r_A \ \dot{r}_a \ \varphi_{Sr} \ \dot{\varphi}_{Sr}]^T$ and the output $y = r_{LA}$ follow the vector fields \mathbf{f} and g and the function h , respectively.

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} x_2 \\ -ax_2^2 - bx_2 \\ x_4 \\ -\frac{g}{l_S} \sin x_3 + \frac{\cos x_3}{l_S} (ax_2^2 + bx_2) \end{bmatrix}, \quad g(\mathbf{x}) = \begin{bmatrix} 0 \\ -m \\ 0 \\ \frac{\cos x_3 m}{l_S} \end{bmatrix}, \quad (7)$$

$$h(\mathbf{x}) = x_1 + l_S \sin x_3 \quad (8)$$

3 Nonlinear control approach

The following considerations are made assuming that the right side of the differential equation for the load sway (1) can be linearized.

$$\ddot{\varphi}_{Sr} + \frac{g}{l_S} \sin \varphi_{Sr} = \frac{1}{l_S} \ddot{r}_A \quad (9)$$

In order to find a linearizing output for the simplified nonlinear system the relative degree has to be ascertained.

The relative degree concerning the output of the system is defined by the following conditions

$$L_{\mathbf{g}} L_{\mathbf{f}}^i h(\mathbf{x}) = 0 \quad \forall i = 0, \dots, r-2, \quad L_{\mathbf{g}} L_{\mathbf{f}}^{r-1} h(\mathbf{x}) \neq 0 \quad \forall \mathbf{x} \in \mathbb{R}^n \quad (10)$$

The operator $L_{\mathbf{f}}$ represents the LIE derivative along the vector field \mathbf{f} and $L_{\mathbf{g}}$ along the vector field \mathbf{g} , respectively.

A relative degree of $r = 2$ is obtained with the real output (8). Because the order of the simplified nonlinear model is 4, y is not a linearizing output. But with a new output

$$y^* = h^*(\mathbf{x}) = x_1 + l_S x_3 \quad (11)$$

a relative degree of $r = 4$ is obtained. Assuming that only small radial rope angles occur, the difference between the real output y and the flat output y^* can be neglected.

Because the simplified system representation is differentially flat an exact linearization can be done. Therefore a new input is defined as $v = \ddot{y}^*$ and the linearizing control signal $u_{l,lin}$ is calculated by

$$u_{l,lin} = \frac{-L_{\mathbf{f}}^r h^*(\mathbf{x}) + v}{L_{\mathbf{g}} L_{\mathbf{f}}^{r-1} h^*(\mathbf{x})} = \frac{g \sin x_3 x_4^2 - g \cos x_3 \left(-\frac{g}{l_S} \sin x_3 + \frac{a}{l_S} x_2^2 + \frac{b}{l_S} x_2 \right) + v}{\frac{gm}{l_S} \cos x_3} \quad (12)$$

In order to stabilize the resulting linearized system a feedback of the error between the reference trajectory and the derivatives of the output y^* is derived.

$$u_l = \frac{-L_{\mathbf{f}}^r h^*(\mathbf{x}) + v - \sum_{i=0}^{r-1} k_i \left(L_{\mathbf{f}}^i h^*(\mathbf{x}) - y_{ref}^{(i)*} \right)}{L_{\mathbf{g}} L_{\mathbf{f}}^{r-1} h^*(\mathbf{x})} \quad (13)$$

The feedback gains k_i are obtained by the pole placement technique.

4 Trajectory generation

The trajectory generation problem is formulated as a constrained, open-loop optimal control

problem for the linearized system including the state feedback. Because of the relevant calculating time for solution of the optimal control problem the model predictive trajectory generation operates with a non-negligible sample time. Likewise the numerical solution procedure itself introduces a discretization of the time axis, see below. But for the sake of simplicity, the open-loop optimal control problem is stated in continuous time in the following.

The model equations are given by

$$\begin{aligned}\dot{\mathbf{x}}_{lin} &= \mathbf{A}_{lin}\mathbf{x}_{lin} + \mathbf{b}_{lin}u_{lin}, & \mathbf{x}_{lin}(t_0) &= \mathbf{x}_{lin,0} \\ \mathbf{y}_{lin} &= \mathbf{C}_{lin}\mathbf{x}_{lin}\end{aligned}\tag{14}$$

The state variables \mathbf{x}_{lin} are the states of the integrator chain forming the linearized system as well as the state variables of the integrator chain for the output reference trajectory. Additional state variables are introduced to generate a smooth input v . The initial state $\mathbf{x}_{lin,0}$ is derived from the state of these integrators and the current system output and system output derivative measurements. The outputs \mathbf{y}_{lin} of the linear system (14) are the variables corresponding to the flat output y^* (eqn. (11)) and its first and second derivative that approximate the load position, velocity, and acceleration.

The objective functional

$$J_c = \frac{1}{2} \int_{t_0}^{t_f} ((\mathbf{y}_{lin} - \mathbf{w})^T \mathbf{Q} (\mathbf{y}_{lin} - \mathbf{w}) + r \dot{u}_{lin}^2) dt\tag{15}$$

is a standard form evaluating quadratically both the deviations of the predicted outputs \mathbf{y}_{lin} from their reference predictions $\mathbf{w}(t)$ and the rate of change of the input variables u_{lin} . The optimization horizon $t_f - t_0$ and the symmetric and positive semi-definite weighting matrix \mathbf{Q} and the weighting coefficient $r > 0$ are essential tuning parameters of the model predictive trajectory generator.

The optimization horizon $t_f - t_0$ should cover the essential dynamics of the process. These are defined by the period of the load sway (up to 18s for the crane under consideration). Practical experience shows that a horizon of 10s is sufficient.

Reference predictions $\mathbf{w}(t)$ are generated from the crane operator's hand lever signals (velocity targets) for the load position, velocity and acceleration. The prediction takes into account velocity reductions if the load approaches the radial boundaries.

The model predictive trajectory generation algorithm incorporates restrictions on the process variables as constraints of the open-loop optimal control problem.

$$u_{lin,\min} \leq u_{lin} \leq u_{lin,\max}, \quad \mathbf{y}_{lin,\min} \leq \mathbf{y}_{lin} \leq \mathbf{y}_{lin,\max}\tag{16}$$

Input rate constraints are applied to avoid high-frequency excitations of the system.

$$\dot{u}_{lin,\min} \leq \dot{u}_{lin} \leq \dot{u}_{lin,\max}\tag{17}$$

Therefore, the change rates \dot{u}_{lin} are to be considered as the control variables in the optimal control problem formulation.

The reference trajectory generation forms an outer control loop. Therefore stability results from model predictive control are applicable. Conditions for guaranteed stability of the closed loop system under nominal conditions usually require stabilizing constraints of the state variables at the end of the optimization horizon together with a suitable evaluation

of the final state [15]. This is approximated by a quadratic penalty term with symmetric, positive definite weighting matrix $\bar{\mathbf{Q}}$ which extends the original objective functional.

$$J = J_c + \frac{1}{2} (\mathbf{x}_{lin}(t_f) - \mathbf{x}_{lin,f})^T \bar{\mathbf{Q}} (\mathbf{x}_{lin}(t_f) - \mathbf{x}_{lin,f}) \quad (18)$$

The continuous-time constrained linear-quadratic optimal control problem (14)-(18) is discretized on the grid

$$\begin{aligned} t_0 = t^0 \leq t^1 \leq \dots \leq t^K = t_f \\ \mathbf{x}_{lin}^{k+1} &= \mathbf{A}^k \mathbf{x}_{lin}^k + \mathbf{b}^k u_{lin}^k, \quad k = 0, \dots, K-1 \\ \mathbf{x}_{lin}^0 &= \mathbf{x}_{lin,0} \\ \mathbf{y}_{lin}^k &= \mathbf{C}_{lin}^k \mathbf{x}_{lin}^k, \quad k = 0, \dots, K \end{aligned} \quad (19)$$

Here \mathbf{x}_{lin}^k , u^k , and \mathbf{y}_{lin}^k denote the values of the respective variables in the grid points t^k . The matrices and vectors \mathbf{A}^k , \mathbf{b}^k , and \mathbf{C}^k are obtained from \mathbf{A} , \mathbf{b} , and \mathbf{C} via solution of the transition equation in $[t^k, t^{k+1}]$.

The objective functional (18) and the constraints (16), (17) are discretized accordingly.

In this way the continuous-time optimal control problem is approximated by a quadratic programming problem (QP) in the control and state variables $[\mathbf{x}_{lin}^k, u_{lin}^k]$ of the discretized problem which can be solved by a standard interior point algorithm [16], [17]. Within this algorithm, the structure of the discrete-time model equations (19) is utilized in a RICCATI-like approach to obtain a solution of the NEWTON step equation with $\mathcal{O}(K(m^3 + n^3))$ operations, i. e. the computational effort grows linearly with the prediction horizon K and cubically with the number of control (m) and state (n) variables. For further details see [18], [19].

Non-uniform sample intervals $\Delta T^k = t^{k+1} - t^k$ within the prediction horizon of the MPC help to limit the dimension of the optimization problem. In this schema, the initial sample steps are determined by the trajectory generation loop and the length of the sample intervals increases linearly within the prediction horizon.

5 Measurement results

In this section, measurements of the boom crane LIEBHERR LHM 322 are presented. The rope length l_S is 57 m and the load mass is 3.5 t for all experiments. The flatness based nonlinear controller and the optimal trajectory generation is implemented on a rapid prototyping system dSPACE DS1103. The sample time of the model-based trajectory generation is 100ms.

Fig. 4 shows the load velocity given by the crane operator's hand lever and the optimized reference trajectory generated by the model-based trajectory generator. The upper bound for the load velocity depends on the radial load position. The acceleration constraint is $|\ddot{r}_{LA}| \leq 0.45 \text{ m/s}^2$.

Fig. 5 compares this reference trajectory with the load velocity measurement. The flatness based controller tracks the reference trajectory and the trajectory generation compensates model uncertainties by planning a model based reference trajectory. This results in a fast and damped motion of the load with nearly no overshoot.

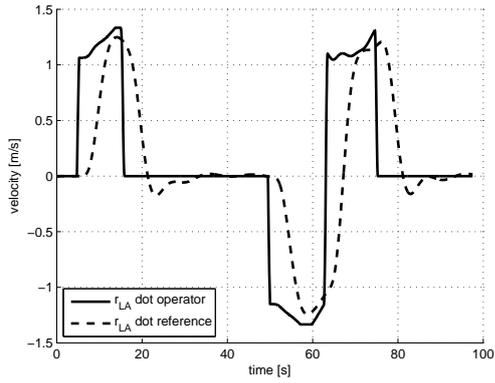


Figure 4: Load velocity reference trajectories.

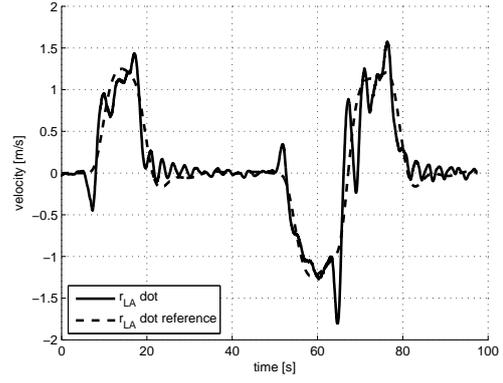


Figure 5: Load velocity: reference trajectory and measurement.

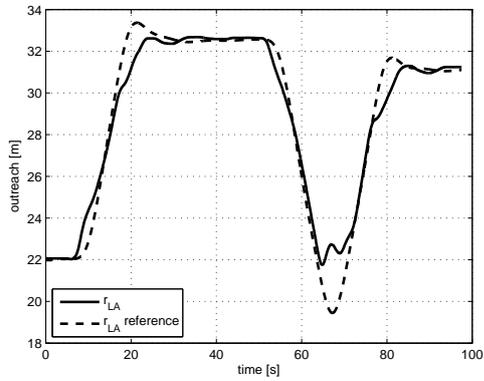


Figure 6: Load position: reference trajectory and measurement.

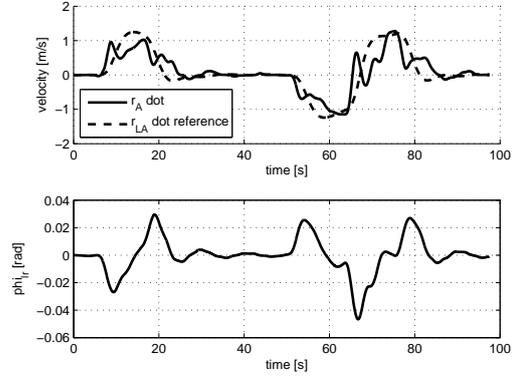


Figure 7: Radial rope angle φ_{Sr} .

The corresponding load position trajectories are shown in Fig. 6.

The control system damps the load sway by suitable compensating movements of the boom during and at the end of each maneuver. The resulting rope angle φ_{Sr} is less than 0.05 rad ($\approx 3^\circ$), see Fig. 7.

The actual calculation time for the online solution of the linear-quadratic optimal control problems is between 54ms and 66ms.

6 Conclusion

In this paper a nonlinear model for a rotary boom crane is developed utilizing the method of NEWTON-EULER. Dominant nonlinearities such as the kinematics of the hydraulic actuator (hydraulic cylinder) are considered. A nonlinear, flatness based controller is developed using a flat output that coincides with the load position for small rope angles. A model-based optimal trajectory generator provides feasible and sufficiently smooth reference trajectories. The optimal control problem to be solved online takes into account the linearized system including the state feedback as well as input and state constraints. The control system is implemented at the LIEBHERR harbor mobile crane to obtain measurement results. These results validate the exact tracking of the reference trajectory with reduced load sway.

References

- [1] O. Sawodny, H. Aschemann, J. Kümpel, C. Tarin, and K. Schneider, "Anti-sway control for boom cranes," in *American Control Conference*, Anchorage, USA, 2002, pp. 244–249.
- [2] O. Sawodny, A. Hildebrandt, and K. Schneider, "Control design for the rotation of crane loads for boom cranes," in *Proc. of ICRA03*, Taiwan, 2003, pp. 2182–2187.
- [3] E. Arnold, O. Sawodny, A. Hildebrandt, and K. Schneider, "Anti-sway system for boom cranes based on an optimal control approach," in *American Control Conference*, Denver, 2003.
- [4] J. Neupert, A. Hildebrandt, O. Sawodny, and S. Schneider, "A trajectory planning strategy for large serving robots," in *Proc. of the SICE Annual Conference*, Okayama, Japan, 2005, pp. 2180–2185.
- [5] —, "Trajectory tracking for boom cranes using a flatness based approach," in *Proc. of the SICE-ICCAS Conference*, Busan, Korea, 2006.
- [6] A. Isidori and C. I. Byrnes, "Output regulation of nonlinear systems," *IEEE Trans. Automat. Control*, vol. 35, no. 2, pp. 131–140, 1990.
- [7] A. Isidori, *Nonlinear control systems*, 3rd ed. Berlin: Springer, 1995.
- [8] M. Fliess, J. Lèvine, P. Martin, and P. Rouchon, "Flatness and defect of nonlinear systems: introductory theory and examples," *Internat. J. Control*, vol. 61, pp. 1327–1361, 1995.
- [9] M. Fliess, J. Lèvine, and P. Rouchon, "A simplified approach of crane control via a generalized state-space model," in *Proc. of the 90th Conference on Decision and Control*, Brighton, England, 1991, pp. 736–741.
- [10] A. Piazzzi and A. Visioli, "Optimal dynamic-inversion-based control of an overhead crane," *Proc.-Control Theory Applications*, vol. 149, no. 5, pp. 405–411, 2002.
- [11] N. Yanai, M. Yamamoto, and A. Mohri, "Feed-back control of crane based on inverse dynamics calculation," in *Proc. of the International Conference on Intelligent Robots and Systems*, Maui, USA, 2001, pp. 482–487.
- [12] B. Kiss, J. Lèvine, and P. Müllhaupt, "Modelling, flatness and simulation of a class of cranes," *Periodica Polytechnica*, vol. 43, no. 3, pp. 215–225, 1999.
- [13] I. M. S. Panahi and H. A. Toliyat, "An efficient processor-based online generation of time-optimal trajectories," *IEEE Trans. Contr. Syst. Technol.*, vol. 14, no. 5, pp. 966–973, 2006.
- [14] N. Zimmert, O. Sawodny, E. Arnold, and S. Oschmann, "An on-line trajectory generation based on a model predictive control approach," in *6th Asian Control Conference*, p. 946.
- [15] G. De Nicolao, L. Magni, and R. Scattolini, "Stability and robustness of nonlinear receding horizon control," in *Nonlinear Model Predictive Control*, ser. Progress in Systems and Control Theory, F. Allgöwer and A. Zheng, Eds., vol. 26. Basel: Birkhäuser, 2000, pp. 3–22.
- [16] S. J. Wright, *Primal-Dual Interior-Point Methods*. Philadelphia: SIAM, 1997.
- [17] S. Mehrotra, "On the implementation of a primal-dual interior point method," *SIAM J. Optim.*, vol. 2, no. 4, pp. 575–601, Nov. 1992.
- [18] C. V. Rao, S. J. Wright, and J. B. Rawlings, "On the application of interior point methods to model predictive control," *J. Optim. Theory Appl.*, vol. 99, no. 3, pp. 723–757, 1998.
- [19] R. Franke and E. Arnold, "The solver Omuses/HQP for structured large-scale constrained optimization: algorithm, implementation, and example application," in *Sixth SIAM Conference on Optimization*. Atlanta: SIAM, 1999.

Author Information:

Dr.-Ing. Eckhard Arnold
Dipl.-Ing. Jörg Neupert
Prof. Dr.-Ing. Oliver Sawodny
Universität Stuttgart
Institute for System Dynamics
Pfaffenwaldring 9
D-70569 Stuttgart, Germany
Tel: +49 711 685-65928
Fax: +49 711 685-66371
E-mail: [arnold|neupert|sawodny]@isys.uni-stuttgart.de

Dr.-Ing. Klaus Schneider
Liebherr-Werk Nenzing GmbH
Dr. Hans Liebherr Straße
A-6710 Nenzing, Austria
E-mail: k.schneider@liebherr.com

K. Shaposhnikov / V. Astakhov

The method of ortogonal projections in problems of the stationary magnetic field computation

ADVANCES IN CONTROL THEORY AND CONTROL ENGINEERING

We consider the problem of stationary plane magnetic field computation in the presence of infinitely long cylindrical ideal magnetics which have piecewise-smooth Lipschitz's boundaries Γ_k of cross-sections Ω_k . Let's introduce the following notation: Ω^-, Γ are the unions of the cross-sections and the boundaries of magnetizing bodies accordingly, $\mu = const$ is magnetic conductivity of the medium that is denoted as $\Omega^+, \Omega = \Omega^- \cup \Omega^+ \cup \Gamma$.

We have the following boundary problems for magnetic field computation:

$$\begin{aligned} \operatorname{rot} \mathbf{H} = \delta, \operatorname{div} \mathbf{B} = \mathbf{0}, \mathbf{B} = \mu \mathbf{H} \text{ in } \Omega^+, & \quad \Delta \varphi^* = 0 \text{ in } \Omega^+, \\ H_\tau = 0 \text{ on } \Gamma, & \quad \varphi^* = C_k - \varphi^0 \text{ on } \Gamma_k, k = \overline{1, N}, \\ \mathbf{H}(M) \xrightarrow{M \rightarrow \infty} 0. & \quad \varphi^*(M) \xrightarrow{M \rightarrow \infty} 0 \end{aligned}$$

where φ^* is scalar potential of intensity of reaction field \mathbf{H}^* , φ^0 is a potential of non-perturbed field intensity \mathbf{H}^0 , $\mathbf{H} = \mathbf{H}^* + \mathbf{H}^0$, $C_k = const$, $k = \overline{1, N}$.

We use a Hilbert space $H(\Omega)$ that is introduced and studied in [1]. The inner product and the norm in $H(\Omega)$ are the following:

$$(\psi_1, \psi_2)_H = \int_{\Omega} \nabla \psi_1 \nabla \psi_2 d\Omega, \quad \|\psi\|_H = \sqrt{(\psi, \psi)_H}.$$

According to [1], there is the orthogonal decomposition for this space: $H(\Omega) = H^*(\Omega) \oplus H_\sigma^0(\Omega)$. The subspace $H^*(\Omega)$ consists of the functions with constant value on Γ_k . $H_\sigma^0(\Omega)$ consists of the functions that are represented by the potentials of simple layer with the zero mean of densities on Γ_k . We have shown that if φ^0 is in the $H(\Omega)$ then φ^* belongs to $H_\sigma^0(\Omega)$ and the potential of the resulting field $\varphi = \varphi^* + \varphi^0$ is in $H^*(\Omega)$. Thus, the solution of the problem is reduced to φ^0 projection to subspace $H_\sigma^0(\Omega)$ that can be found by the calculation of φ^0 coordinates in some basis for $H_\sigma^0(\Omega)$. The coordinates is defined by the following SLAE solution:

$$\begin{bmatrix} (\varphi_1, \varphi_1)_H & \dots & (\varphi_1, \varphi_n)_H \\ \dots & \dots & \dots \\ (\varphi_n, \varphi_1)_H & \dots & (\varphi_n, \varphi_n)_H \end{bmatrix} \begin{bmatrix} c_1 \\ \dots \\ c_n \end{bmatrix} = - \begin{bmatrix} (\varphi^0, \varphi_1) \\ \dots \\ (\varphi^0, \varphi_n) \end{bmatrix}.$$

We denote $\Gamma_j = \Gamma'_j \cup \Gamma''_j$ where Γ'_j is cross-section of this body and Γ''_j is its

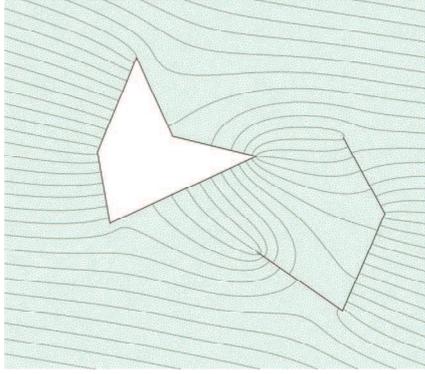


Fig. 1

closure to Γ_j in case if the cross-section of j -th magnetizing body is singular. We define that the density of φ^* is zero on Γ_j'' . We have shown that the potentials with such densities form a subspace in $H_o^0(\Omega)$ and all proofs are right with the change Γ_j to Γ_j' .

It is convenient to use the potentials of simple layer with finite step densities with zero mean on Γ_k as the coordinate functions. We have proved that such system of functions is complete in $H_o^0(\Omega)$ in sense of

Ritz's approximate sequence convergence. In this case the elements of SLAE are defined by the following formulas:

$$(\varphi_i, \varphi_k) = \frac{1}{2\pi} \int_{\Gamma} \sigma_i \int_{\Gamma} \sigma_k \ln \frac{1}{r} d\Gamma d\Gamma, \quad (\varphi^0, \varphi_i) = \int_{\Gamma} \sigma_i \varphi^0 d\Gamma.$$

The computing algorithm and its software realization are developed. This package computes magnetic field faster and more effectively than the famous software packages. The example of the developed software package usage is show on Fig. 1.

References:

[1] Astakhov V.I. Surface potentials and operators of potential's theory in Dirichlet's spaces // Electromechanics. 2000. #2. p.1-18. (in Russian language)

Authors:

Kirill Shaposhnikov
 prof. Vladimir Astakhov
 Southern Scientific Centre of Russian Academy of Science, Ordjonikidze Str., 122
 346504, Shakhty
 Phone: +7(918)5444579
 Fax: +7(86352)55374
 E-mail: _cyr_@mail.ru

J. Naumenko

The computing of sinusoidal magnetic fields in presence of the surface with bounded conductivity

ADVANCES IN CONTROL THEORY AND CONTROL ENGINEERING

There exist a number of technical problems in that necessary to provide a control of electro technical systems that include conducting plates and cases. The most common example of such systems is MAGLEV. The main aim of this work is the creation of the methods for computing of integral parameters of there systems such a force or electromagnetic field energy.

Let's consider the space $L_2(S)$ that consists of two-component complex square-integrable on S vector functions. We suppose that the multiconnected surface S and its boundary satisfy the Lipschitz's conditions. The space $L_2(S)$ can be decomposed to the sum: $L_2(S) = L_2^s(S) \oplus L_2^h(S) \oplus L_2^p(S)$. Here $L_2^p(S)$ consists of potential fields generalized by the Weyl [1], $L_2^s(S)$ consists of generalized solenoidal fields and $L_2^h(S)$ consists of generalized harmonic fields. Let's designate $L^s = L_2^s(S) \oplus L_2^h(S)$. We use the orthoprojector $P = P^L P^S$ below where P^S wanishes normal to S field component and P^L is orthoprojector $L_2(S) \rightarrow L^s$.

The computing of sinusoidal magnetic fields in presence of the surface with bounded conductivity can be reduced to the following operator equation for eddy currents density on S :

$$\delta = \lambda K \delta + \mathbf{f}.$$

Here $K = P\Gamma$, $\Gamma\xi = \frac{1}{4\pi} \iint_S \frac{\xi}{r} dS$, $\lambda = j\mu\gamma h\omega$ is some imaginary parameter, μ is the magnetic permeability of the medium, γ is the conductivity of the surface S , h is the thickness of the S , ω is the circle frequency of the exiting sources, $\mathbf{f} \in L^s$.

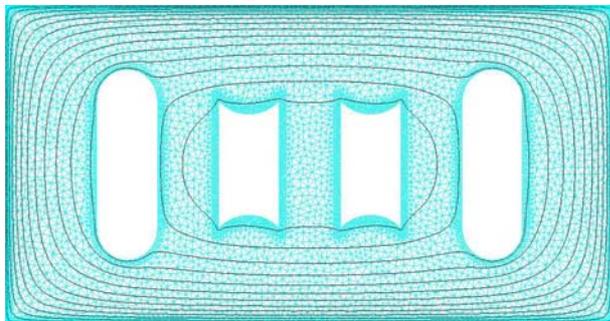


Fig. 1

We proof the existence, uniqueness and numerical stability of the described equation solution in the space L^s and also in the space $W_{2,L}^1 = L^s \cap W_2^1(S)$ by the theory of Riess – Fredholm. Here $W_2^1(S)$ is the two-component Sobolev space.

The software package was built on the basis of the described theory. Some practical problems of electrodynamics were computed by it. The example of package usage is shown on Fig.1. The exiting field is homogenous, the frequency is 500 Hz, the thickness is 1mm, the material of S is copper.

References:

[1] Weyl H. The method of orthogonal projection in potential theory // Duke Math. J. 1940. V. 7. P. 411 – 444.

Authors:

Dr. Jan Naumenko

Southern Science Centre of RAS, South-Russian State Technical Univ., Budyonovskaya st. 183-29
346421, Novocherkassk, Russia

Phone: +79054576301

Fax: +78635255374

E-mail: jan@novoch.ru

K. Bayramkulov / V. Astakhov

The method of the boundary equations in problems of computing static and stationary fields on the topological graph

ADVANCES IN CONTROL THEORY AND CONTROL ENGINEERING

There is a number of technical problems in which it is necessary to provide control of electrotechnical systems. The purpose of the given work is construction of methods of a numerical finding of parameters of such systems.

The problem of computing a plane stationary magnetic field in the presence of magnetizable bodies has been considered. The opportunity of application of the method of boundary equations on the graph of electric circuits in this example has been watched.

The initial boundary problem is following:

$$\begin{aligned} \operatorname{rot} \bar{H} &= \begin{cases} \bar{\delta} & \text{on } S_k, \\ 0 & \text{on } S \setminus S_k, \end{cases} \quad \bar{\delta} = \bar{e}_z \delta; \\ \operatorname{div} \bar{B} &= 0; \quad \bar{B} = \mu \bar{H}; \quad \mu = \mu^\pm \text{ on } S^\pm; \\ B_n^+ &= B_n^-, \quad H_\Gamma^+ = H_\Gamma^- \text{ на } \Gamma; \\ |\bar{B}(M)| &\xrightarrow{M \rightarrow \infty} 0 \end{aligned} \quad (1)$$

(S_k is the sections of the coil). In our statement the condition of solenoidalness an induction \bar{B} of the magnetic field including in (1), can be satisfied representation $\bar{B} = [\operatorname{grad} \varphi \bar{e}_z]$.

The problem (1) can be approximately transformed to the boundary problem for φ on the graph of Kirchhoff electric circuit (KEC). This graph may be construct by the method of power balance. The problem on graph KEC looks as:

$$\Delta \varphi(q) = \begin{cases} J, & q \in M_J, \\ 0, & q \notin M_J \cup M_\Gamma, \\ i_q(r-R), & q \in M_\Gamma, \end{cases} \quad \varphi(q) \xrightarrow{q \rightarrow \infty} 0 \quad (2)$$

M_q is the set of nodes connected by edges with M_q , $J = \frac{\mu^+ i}{W}$, i is the full current of the coil, $2W$ is the number of ineradicable nodes included in set M_J (belonging to section of the coil), M_Γ is the set of boundary nodes (in which converge edges with different resistance). The problem (2) has been reduced to the boundary matrix equation. The following representation has been used for this purpose

$$\varphi(p) = - \sum_M R(p, q) \Delta \varphi(q) \quad (3)$$

(M is the set of all nodes the graph) by the fundamental solution $R(p, q)$ of Laplace

equation on infinite graph (the finding $R(p, q)$ makes a separate problem).

After technically simple transformations the matrix equation gets a form

$$I = \lambda AI + B \quad (4)$$

$I = \{i_p\}_{p=1}^m$ is the boundary currents, $\lambda = \frac{r-R}{r+R}$, $B = \{b_p\}_{p=1}^m$, $A = \{a_{pq}\}_{p,q=1}^m$.

The research of the equation (4) in class of bounded functions with zero average values on M_Γ has shown its correctness. Let's note, that equality (3) will give the answer of a problem (2) if preliminary solve the equation (4) and, thus, define boundary currents. The software package for the numerical realization of the developed theory has been created. Examples of the calculations executed by the considered method are illustrated on fig. 1.

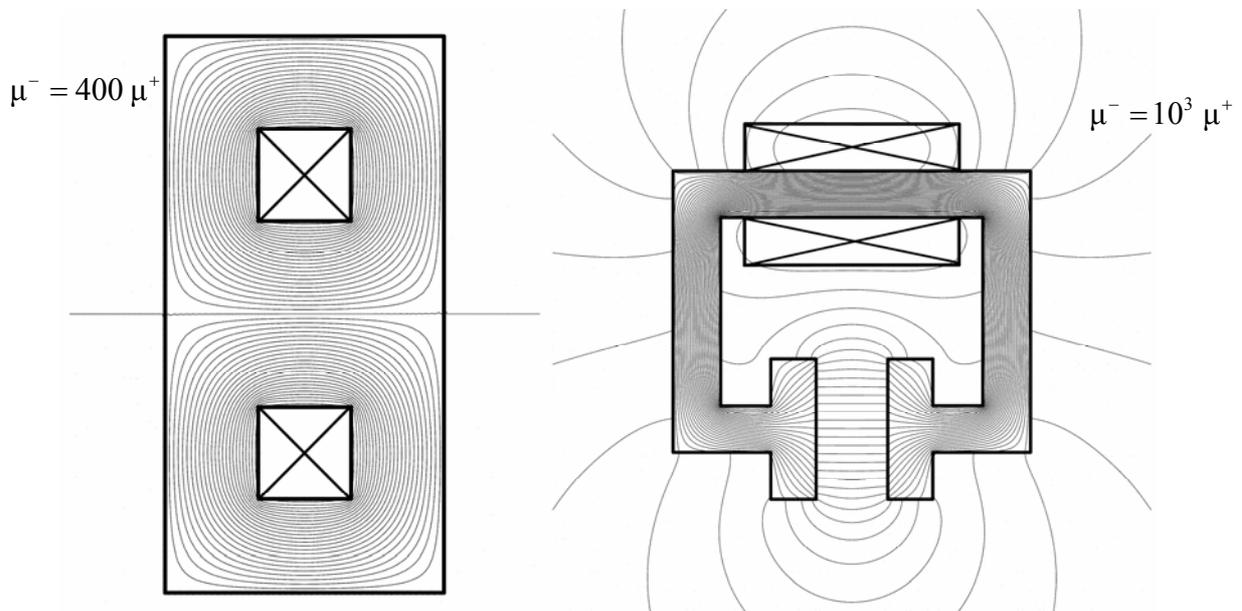


Fig.1

References:

[1] Bayramkulov K.N.-A. Astakhov V.I. Computing of a magnetic field by the method of the boundary equations on the graph of an electric circuit //Works of the Southern Scientific Centre of RAS. -- 2007. Vol.2 P. 72-79.

Authors:

Kazim Bayramkulov
 Prof. Vladimir Astakhov
 Company, street, P.O.B. Southern Scientific Centre of Russian Academy of Science, Pervomayskaya Str. 105-202
 Zip code, city: 346428, Novocherkassk Russia
 Phone: +7 9289558266
 Fax: +7 86352 55374
 E-mail: kazim82@mail.ru

T. Kochubey/ V. Astakhov

The computation of magnetic field in the presence of ideal conductors using the integral-differential equation of the first kind

ADVANCES IN CONTROL THEORY AND CONTROL ENGINEERING

Statements of many engineering problems include necessity of mathematical modelling and computation of stationary magnetic fields in the presence of ideal conductors.

The problem of stationary plane magnetic field computation in the presence of N infinitely long cylindrical ideal conductors has been considered. Cross-sections Ω_k^- of the ideal conductors have piecewise-smooth Lipschitz's boundaries Γ_k . The following notation is used here: Γ is the union of Γ_k , Ω^- is the union of Ω_k^- , Ω^+ is an exterior domain to Ω^- with magnetic conductivity $\mu^+ = const$.

The problem of magnetic field computation has been reduced to the following plane boundary problem which in vector and scalar statement looks like:

$$\begin{aligned} \operatorname{rot} \mathbf{H} = \boldsymbol{\delta}, \operatorname{div} \mathbf{B} = \mathbf{0}, \mathbf{B} = \mu^+ \mathbf{H} \text{ in } \Omega^+, & \quad \Delta \varphi^* = 0 \text{ in } \Omega^+, \\ B_n = 0 \text{ on } \Gamma, & \quad \frac{\partial \varphi^*}{\partial n} = B_n^0 \text{ on } \Gamma, \\ \mathbf{B}(M) \xrightarrow{M \rightarrow \infty} \mathbf{0}, & \quad \varphi^*(M) \xrightarrow{M \rightarrow \infty} 0, \\ \mathbf{B} = \mathbf{B}^* + \mathbf{B}^0, \mathbf{B}^* = -\operatorname{grad} \varphi^* \text{ in } \Omega^+ \cup \Omega^- & \end{aligned}$$

where \mathbf{B}^0 is an induction of non-perturbed magnetic field, \mathbf{B}^* is an induction of magnetic field of reaction, φ^* is a scalar potential of magnetic field of reaction. The balance of magnetic field sources is also taken into account.

The φ^* can be represented in the form of potential of double layer. The boundary problem has been reduced in this case to the integral-differential equation of the first kind for scalar density:

$$K_{\tau\sigma} \tau = f_\tau$$

where

$$K_{\tau\sigma} \tau = -\frac{1}{\pi} \frac{\partial}{\partial n} \int_{\Gamma} \tau \frac{\partial}{\partial n} \ln \frac{1}{r} d\Gamma, \quad f_\tau = -2B_n^0.$$

It is necessary to introduce Γ_k'' that is a closure to closed $\Gamma_k = \Gamma_k' \cup \Gamma_k''$ in case if k section is singular.

It was shown that the operator of the equation is linear, self-adjoint and positive in $L_2^0(\Gamma)$ that is Hilbert space of square-integrable functions with zero average value. It allowed to prove the availability, uniqueness and stability of the solution of the equation in the operator's energetic space [1] by applying the

variational principle and Riss's theorem.

The numerical solution of the equation is carried out by constructing the minimizing Ritz's sequence $\tau^{(1)}, \tau^{(2)}, \dots, \tau^{(n)}, \dots$ on the base of system of continuous piece-polynomial coordinate functions. This approach reduces the equation to the SLAE with the real positive-definite matrix.

The software package for the numerical realization of the developed theory has been created. The comparing of the results of the software package with the results of similar packages has shown that the developed software package has higher accuracy and shorter time of computation. Examples of the developed package usage are shown on Fig. 1.

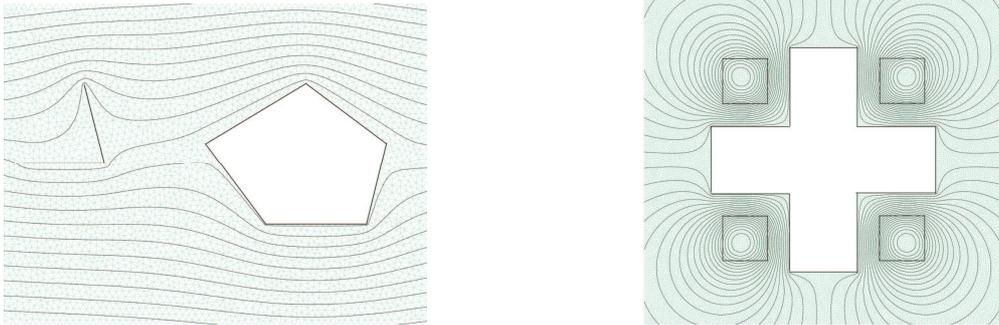


Fig. 1

References:

[1] Astakhov V. I. Surface potentials and operators of potential's theory in Dirichlet's spaces // Electromechanics. 2000. #2. p.1-18. (in Russian language)

Authors:

Tatiana Kochubey
Prof. Vladimir Astakhov
Southern Scientific Centre of Russian Academy of Science, Svoboda Str., 20a-17.
346405, Novochoerkassk, Russia
Phone: +79185970526
Fax: +78635255374
E-mail: tako84@yandex.ru

M. Schneider / U. Lehmann / J. Krone / P. Langbein / Ch. Ament / P. Otto / U. Stark /
J. Schrickel

Artificial neural network for product-accompanied analysis and control

1. Material behaviour of the side shell

In the context of the project in cooperation with Airbus the material behaviour of several components of a fuselage section is modelled and simulated by the Competence Platform “Computer Vision based on Computational Intelligence (KOPF CV&CI)” with artificial neural network (ANN) at the South Westphalia University of Applied Sciences. For this purpose fuselage components were measured in a series of experiments to model the interrelationship between movement of actors for positioning and corresponding change of component geometry [2]. Every actor can move independently of each other in x-, y-, z-direction. A very exact measuring system delivers the coordinates of the measuring points which are integrated as planned in the fuselage components. With these measuring points the geometry of the side shell and the position in space is determined. The data sets of every single measuring form the base to develop a simulation model showing the material behaviour of the right side shell in an artificial neural network (ANN). After the data sets had been processed and evaluated, they could be used for training. The training was not accomplished by absolute coordinates of the measuring points but by relative deviations from zero position of side shell.

Movements of controllable actors serve as input to artificial neural network and x-, y- and z-coordinates of every single measuring point serve as output. Due to a low number of data sets an artificial neural network was developed for every single measuring point. In addition, only the main moving direction (=y-direction) of the actors was taken into account to reduce the connections between the neurons. These artificial neural networks were put into a parallel order (figure 1). The predicted position of measuring points meets the high precision request of Airbus, i.e. artificial neural networks are suitable to predict this strong nonlinear material behaviour of fuselage components very well. With a larger number of data sets and taking into account of all possible actor movements it

would be better to train the material behaviour in one artificial neural network. So the result of the simulation might be even better. The interaction of measuring points with each other could be trained much better in a neural material model. Therefore the material behaviour of the side shell could be simulated more exactly. Due to the better relationship of connections between neurons to training data sets the neural model by 7 parallel artificial neural networks was favoured concerning the generalization performance.

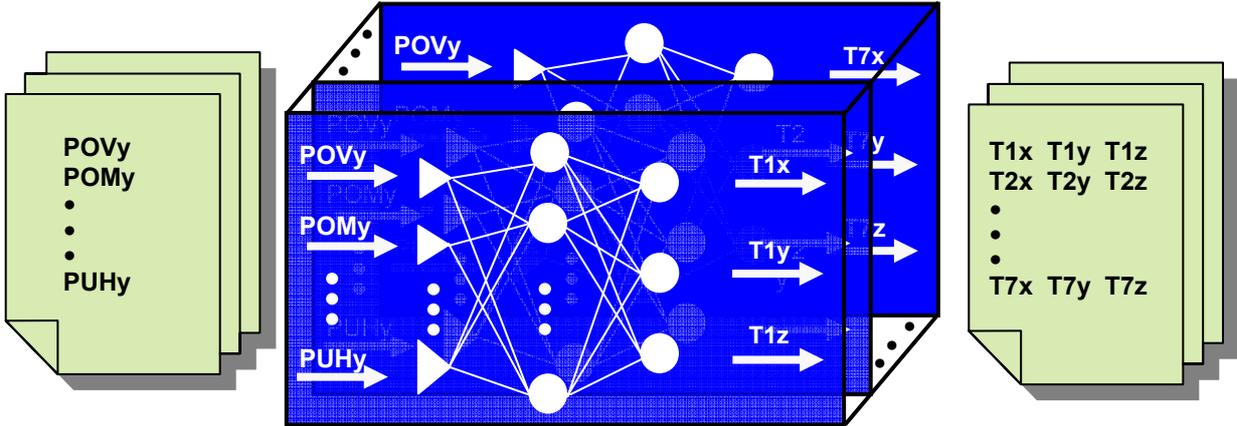


Figure 1: parallel order of the networks to simulate the material behaviour

In comparison to the finite elements method artificial neural networks are suitable very well to simulate this strongly nonlinear behaviour of components of a bent extensive fuselage section [1]. The data of process monitoring and prediction shall be used to simulate the process control.

2. Control concept

An aim of the project is to create a concept for production accompanying analysis of component geometry of a fuselage section by artificial neural networks [1]. “At present it is common for fuselage production, first to join extensive bent components to fuselage sections together [3]”. The individual components are riveted together at the component transitions. The fuselage section can stabilize now and is transported to another location for further assembly. This ready fuselage section shall correspond to the construction plan exactly. If the precision is not optimal, special approval procedures for the Aeronautics Federal Office or time- and cost-intensive improvement work are necessary to meet the demanded precision of the geometry of the fuselage section [5].

Various outer influences can have an effect on the components of the fuselage section and can influence the assembly. They shall be compensated automatically by a Computational Intelligence (CI) control concept. "It is not unusual, that the tool finally made produces parts, which one being outside the measure tolerances of the original design sheet ... In addition, deviations which are caused by temperature changes in the factory, can change ... the measurements of the final part [4]." In predefined intervals a visual measuring system (Computer Vision) delivers the coordinates of the measuring points. These data are led back in control system to ensure the component of a fuselage section corresponding to its specification. Controllable actors can specifically influence the geometry of the fuselage component in context of the permitted limits at measured deviations.

The control process is a neural model with 6 inputs (=actor's deviation) and 21 outputs (x-, y-, z-deviation of the 7 measuring points) as pointed in chapter 1. The applied controller needs 21 controlled process variable and 6 actuating variables. At the movement of one single actor all 7 measuring points are moved, e.g. no controlled process variable has to be assigned for exactly one actuating variable. With a classic multivariable control this feedback control problem is very difficult to solve because the controlled process variable and actuating variable are coupled ambiguously. A neural controller is able to meet these requirements. The neural model of the material behaviour cannot simply be inverted, because only one correction movement of the actors can return the measuring points into the desired position. A clear solution would not be received by 7 artificial neural networks in parallel. In spite of the low number of data sets only one network was trained. The data sets were partitioned the following way: 76% for training, 14% for validation and 9% for testing.

If one measuring point is out from zero position the neural controller displays the essential traverse path of the actors to move the measuring points back into zero position.

In change of outer boundary conditions and/or the system behaviour the artificial neural network delivers the traverse path of every single actor. Therefore the geometry of the fuselage section can be kept automatically constant by a CI-control of a close limit of a couple of one-hundredths of a millimetre.

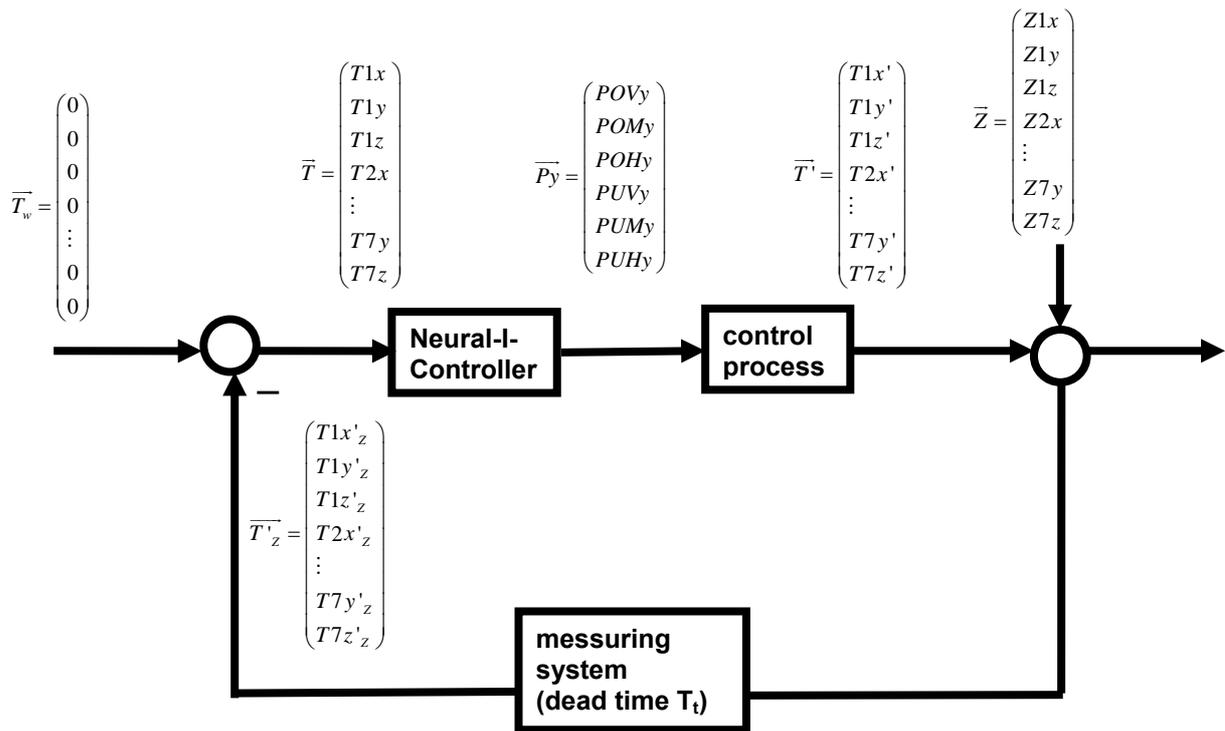


Figure 2: closed loop for simulation

The closed loop (figure 2) is dominated by the dead time. For continuous control an integral element was put in front of the neural controller. The setting rules of Ziegler and Nichols [6] were consulted to dimension the integration time T_i .

$$T_i \approx 3,33 T_t$$

For this simulation model was reached a good simulation result for: $T_i = 4 T_t$

Therefore oscillation of the closed loop could be prevented. The process model which is influenced by the disturbance input is controlled well.

3. Prospects

This model based control will be tested at a process model and can gradually be integrated into the real production expiry. Control accuracy of this control system by CI-controller can be improved permanently by an ongoing training in regular operation. The applicability of this Neural-I-Controller to side shells of other sections of the same aeroplane model and to side shells of other aeroplane models still has to be assayed. Furthermore the number of data sets should be extended and the simulation model should be tested on the real assembly. Advantages at the production costs could be realised by this CI-control concept.

4. Conclusion

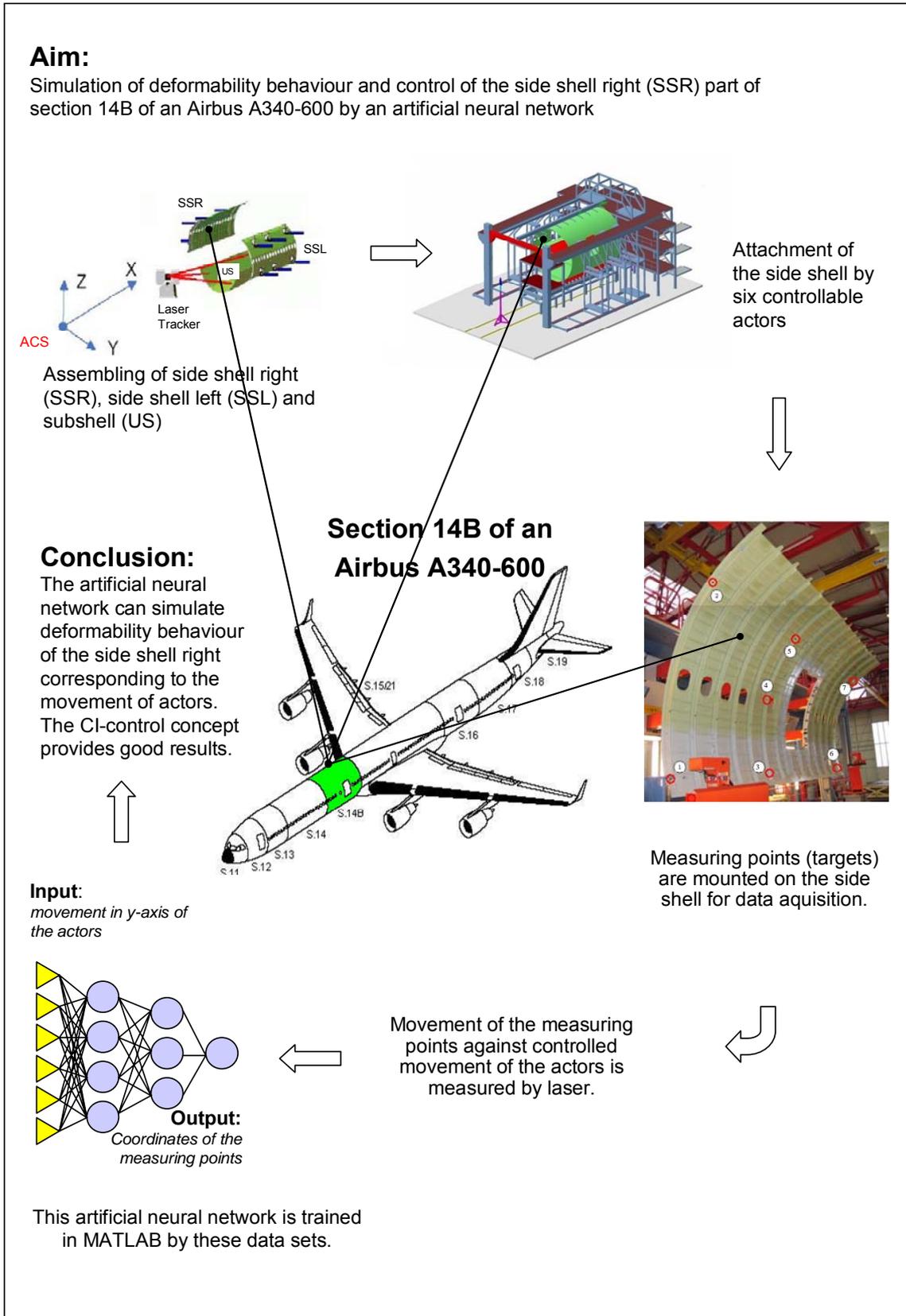


Figure 3: Artificial neural network for production accompanying analysis and control

5. Acknowledgement

The Competence Platform “Computer Vision based on and Computational Intelligence” (<http://www.fh-swf.de/cv-ci>) is funded by the Ministry of Innovation, Science, Research and Technology of the State of North Rhine-Westphalia.

The authors want to express their sincere thanks to Dipl.-Ing. Ulrich Stark and Dipl.-Ing. Jörg Schrickel from the Airbus Deutschland GmbH for supporting this research by providing real measuring data sets.

References:

- [1] R. Tillmann, M. Schneider, J. Krone, U. Lehmann, J. Schrickel, U. Stark, D. Welbhoff: „Airbus - KNN zur fertigungsbegleitenden Analyse und Prozessregelung“, Ausstellung “Lebendige Forschung an Fachhochschulen im Landtag NRW”, Düsseldorf, 24.01.-02.02.2007
- [2] Lehmann, U., Johannes, H., Hohage, S., Woestmann, S.: „Analyse eines Umformprozesses mit KNN und Data Mining“, GI/VDI-Computational-Intelligence-Kongress, Baden-Baden im Mai 2000
- [3] Patentschrift DE 198 34 703 C 1; Verfahren zur Herstellung, Ausrüstung und Ausstattung eines Flugzeugumpfes und Vorrichtung zur Durchführung des Verfahrens; Patentinhaber: DaimlerChrysler Aerospace Airbus GmbH, 21129 Hamburg, DE; Veröffentlichungstag der Patenterteilung 30.12.1999
- [4] Patentschrift DE 693 31 430 T 2; Platten und Rumpfmontage; Patentinhaber: The Boeing Co., Seattle, Wash., US; Veröffentlichungstag im Patentblatt 07.11.2002
- [5] Prof. Dr.-Ing. Klaus Rall, Dr.-Ing. habil. Jörg Wollnak, TU Hamburg-Harburg, „Genauigkeit muss nicht teuer sein“; SPEKTRUM: Special Luftfahrtforschung 2003
- [6] Prof. Dr.-Ing. Holger Wendt, Prof. Dr.-Ing. Wolfgang Wendt: „Taschenbuch der Regelungstechnik“, Verlag Harri Deutsch, 5. Auflage, 2003

Authors:

Dipl.-Ing. M. Schneider, Prof. U. Lehmann, Prof. Dr.-Ing. J. Krone, Prof. Dr.-Ing. habil. P. Langbein
Fachhochschule Südwestfalen
South Westphalia University of Applied Sciences
Frauenstuhlweg 31
D-58644 Iserlohn, Germany
Phone: +49/(0)2371/566-180
Fax: +49/(0)2371/566-209
E-mail: MSchneider@fh-swf.de, Lehmann@fh-swf.de, Krone@fh-swf.de, langbein@fh-swf.de

Prof. Dr.-Ing. habil. Ch. Ament, Priv.-Doz. Dr.-Ing. habil. P. Otto
Technische Universität Ilmenau
Faculty of Computer Sciences and Automation
System Analysis Group
Postfach 10 05 65
D-98684 Ilmenau, Germany
Phone: +49/(0)3677/69-2815, -2773
Fax: +49/(0)3677/69-1434
E-mail: christoph.ament@tu-ilmenau.de, peter.otto@tu-ilmenau.de

Dipl.-Ing. U. Stark, Dipl.-Ing. J. Schrickel
Airbus Deutschland GmbH
Kreetslag 10
D-21129 Hamburg, Germany
Phone: +49/(0)40/743-81268, -85167
Fax: +49/(0)40/743-74794
E-mail: ulrich.stark@airbus.com, joerg.schrickel@airbus.com

I. JAWISH

The Improvement of Traveling Responses of a Subway Train using Fuzzy Logic Techniques

* Advances in Control Theory and Control Engineering

Keywords : Metro Network , ATO- Automatic Train Operation , Fuzzy predictive controller(FPC) , PI - Like FC, FIS-Fuzzy Inference System , Traveling mode , Traveling response, Riding comfort.

Abstract

It's well known that the conventional PID controllers are fixed-gain feedback controllers, they can't compensate the parameter variations in the plant and can't adapt changes in the environment. In our case , the classical PID controllers of the automatic train operation system start or stop a train by reacting to markers that show how far the vehicle is from a station. Because the controllers are rigidly programmed, the ride may be jerky; the automated controller will apply the same brake pressure when a train is, say ,100 meters from a station, even if the train is going uphill or downhill. Moreover, depending on the difference between the actual and set point speeds, the drive motor is to be accelerated or retarded by increasing or decreasing the operating voltage not considering the actual train traveling mode (e.g., Slow, Moderate, Fast), i.e., not taking into account the kinetic energy of the entire moving parts of the train.

In this paper we are concerned with a specific aspect of Fuzzy-ATO , when the conversion of speed differences into a "Traveling response" is to take place in accordance with the actual train "Traveling mode". Using MATLAB 7.0.2 \ SIMULINK & Fuzzy TOOLBOX software and environment, a PI-like fuzzy controller FC for a speed control of dc traction motor model is designed. The designed FC has a Mamdani Model with two inputs: the linguistic variables "Speed deviation" , and 'Traveling mode'. The FC output is considered for the linguistic manipulated variable 'Traveling response'. Furthermore, 21 IF-THEN fuzzy rules is put up together, and the Max-Min inference mechanism is used in the rule evaluation process. The final step in the fuzzy logic controller design is to combine the fuzzy outputs into a crisp output , the Center of Gravity (COG) method is used in our application due to better results it gives, while the Integral action of the FC is taken separately. Finally, the influence of "Traveling mode" on the "Traveling response" at different "Speed deviation" has been investigated. The simulation results show has a sufficient performance of the train speed fuzzy control system using the designed PI-Like fuzzy controller.

I- Introduction

Electrical transportation means have become a part of modern life in developed countries. Subway trains are the famous one, since they offer an intensifier and fast access of passengers to the desired destination , contribute in saving and increasing efficient energy consumption , reducing traffic jam and air pollution.

In Syria a lot of joint projects have been carried out at domestic and international studies institutions. The common subscriber of such studies reveals the necessity to use large public transportation means, and to enhance the electrical ones , in particular, electrical trams and buses, and the underground tube. The intention to establish a Metro Network in a big Syrian cities (e.g., Damascus and Aleppo) are now strongly adopted by the Ministry of Transportation .

Metro Networks demand a considerable amount of electrical energy, which is consumed by: electric traction motors; escalators operation systems; control; measuring, monitoring (SCADA); lighting; conditioning and other needs. Meanwhile, the traction motors consume ~ 85% of the total supplied electric energy. Therefore, it's an important issue to improve the techno-economical performance indices of such distinguishable public transport means by implementing advanced control paradigms,

and introducing developed equipment .

An alternative approach to the control of any process is to investigate the control strategies employed by the human operator. One of the most successful example of implementation fuzzy logic technology was the application of a predictive fuzzy control to automatic subway train operation system in Metro Sendai-Japan, which outperformed both human operator and conventional automated PID controllers. The purpose of this system is to obtain automatic operation on a par with that of skilled operator; specifically this means multi-purpose operation that balances several requirements such as riding comfort, stopping accuracy, and energy conservation. This automatic train operation system is called Fuzzy-ATO. The speed over which the ATC (automatic train control) operates the emergency brake to maintain safety is called critical speed .The speed between stations is controlled so as not to exceed the speed limit, then at the next station the train is stopped accurately at the stop position through automatic application of driving control and a brake control command (notch) [1]. Such Fuzzy-ATO has beat an automated version on the riders'comfort , shortened riding times and even achieved a 10 percent reduction in the trains' energy consumption.

It's well known that the conventional PID controllers are fixed-gain feedback controllers,they can't compensate the parameter variations in the plant and can't adapt changes in the environment. In our case , the classical PID controllers of the automatic train operation system start or stop a train by reacting to markers that show how far the vehicle is from a station. Because the controllers are rigidly programmed, the ride may be jerky; the automated controller will apply the same brake pressure when a train is, say ,100 meters from a station,even if the train is going uphill or downhill. Moreover depending on the difference between the actual and set point speeds, the drive motor is to be accelerated or retarded by increasing or decreasing the operating voltage not considering the actual train traveling mode (e.g., Slow, Moderate,Fast), i.e., not taking into account the kinetic energy of the entire moving parts of the train.

In this paper we are concerned with this specific control issue , when the conversion of speed differences between the actual and set point speeds into a "Traveling response" is to be taken in accordance with the actual train "Traveling mode". Consequently, the drive traction motor should be accelerated or retarded with respect to a new different values of operating voltage then as in a conventional case.

II- Basic definitions ,and structures of the Fuzzy Logic Controllers

1- Fuzzy controllers (FC) are a class of knowledge based controllers using artificial intelligence techniques with origins in fuzzy logic. In the structure definition of a FC it is necessary to define the inputs and outputs (linguistic variables) of the control system , in terms of fuzzy sets, the number of linguistic labels(terms)and the respective membership functions for each , the design of fuzzy control rules ,the type of fuzzy inference system FIS, and finally the defuzzification methods .

Defuzzification method , i.e., transformation of the fuzzy control statement into specific control actions since the controlled process takes only crisp values as inputs , we have to use a defuzzifier to convert a fuzzy systems , set to a crisp value .In case of implementing Mamdani fuzzy models, there are two common means of doing this the Center of Gravity (COG) , and Mean of Maximum (MOM)[2],[3].

2- The Fuzzy Inference System (FIS)

FIS is a computational framework based on the concepts of *fuzzy sets*, *fuzzy if- then rules* ,and *fuzzy reasoning* .The basic structure of FIS consists of three conceptual components: a *rule-base* ,which contains a selection of *fuzzy rules* ;a *database* (or dictionary) ,which defines the *membership functions* used in the fuzzy rules ; and a *reasoning mechanism*, which performs the inference procedure upon the rules and given facts to derive a reasonable output or conclusion[4] .

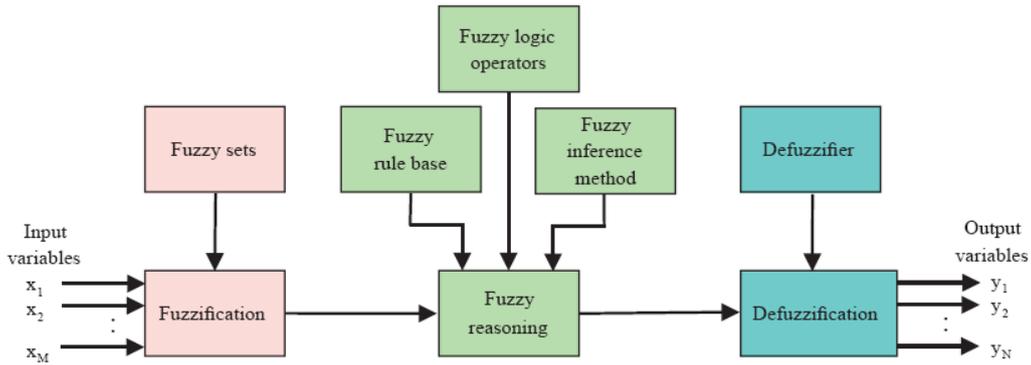


Figure.1 The Block Diagram of a typical Fuzzy Controller

3- Structures of a PID-Like Fuzzy logic Controllers

FLC can be derived from the original classical PID mathematical model [5],[6],[7],[8]:

$$u(t) = K_p e(t) + K_i \int e(t) dt + K_d \frac{de}{dt} \quad (1)$$

or in discrete form:
$$u(k) = K_p e(k) + K_i \sum e(m) + K_d \Delta e(k), \quad m=0, \dots, k \quad (2)$$

The increment of the output signal is :

$$\Delta u(k) = u(k) - u(k-1), \text{ then} \quad (3)$$

$$\Delta u(k) = K_p \Delta e(k) + K_i e(k) + K_d \Delta^2 e(k) \quad (4)$$

Depending on the choice made in the design phase we can have different types of FLCs : P, PI, PD, or PID .Usually, the universes of discourse of the controller variables are normalized in the [-1,0,+1] or in the [0,+1] domains. Associated to each linguistic variable is a scaling factor. Scaling factors of the FC enable the use of normalized universes of discourse and play a role similar to that of the gain coefficients in conventional controllers. The Block diagram of a typical Fuzzy Controller is shown in Figure.1

III-Design Steps of the subway train fuzzy speed control system

In this section we shall illustrate how a PI -like fuzzy controller for a speed control of dc traction motor is constructed . The recommended FC has a Mamdani Model with two inputs: the linguistic variables "Speed deviation" , and 'Traveling mode'.The FC output is considered for the linguistic manipulated variable 'Traveling response', while the Integral action of the FC has been taken separately. We now continue the design of the FC in details [9] :

1) Converting the control objectives Into Linguistic Variables and Fuzzy Sets

The control differences between the actual traveling rates corresponds directly to the set -point and actual speeds of the drive motor . A '**Speed deviation**' linguistic variable is thus required. To allow differentiated evaluation ,seven fuzzy sets are defined for the 'speed deviation' variable. They are 'very negative', 'negative', 'slightly negative' , 'zero' , 'slightly positive' , 'positive' , and 'very positive' (Fig.2) The second linguistic variable '**Traveling mode**' is represented as the second FC input .It encompasses the 'Slow' , 'Moderate' and 'Fast' fuzzy sets (Fig.3).

The linguistic variable required for the manipulated variable is termed '**Traveling response**' and it consists of seven fuzzy sets . They are : 'Strong braking' , 'braking' , 'slight braking' , 'constant speed' , 'slight acceleration' , 'acceleration' and 'strong acceleration' ,the shape of the Membership Functions is the same as of "Speed deviation" .

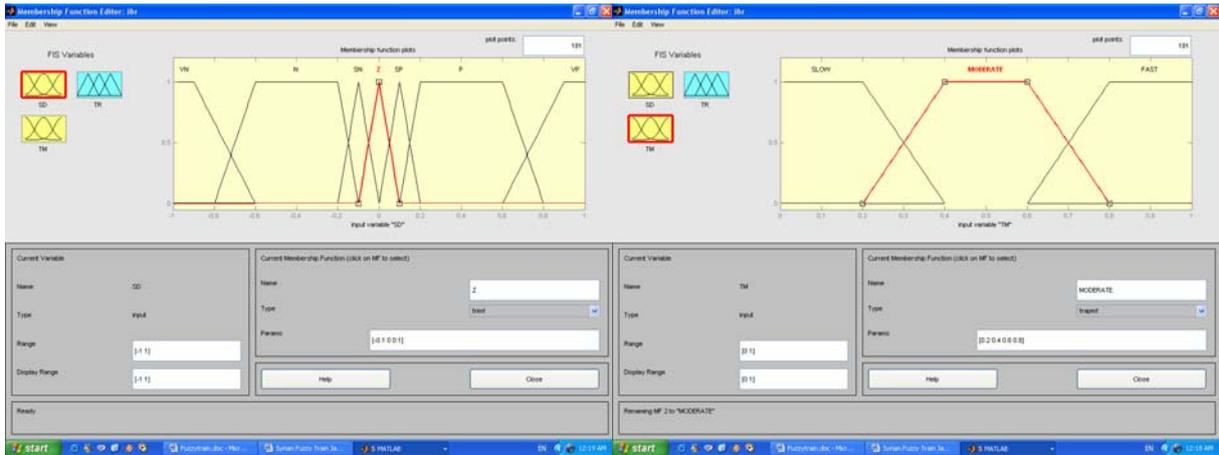


Fig.2 Membership Functions of Speed Deviation

Fig.3 Membership Functions of Traveling mode

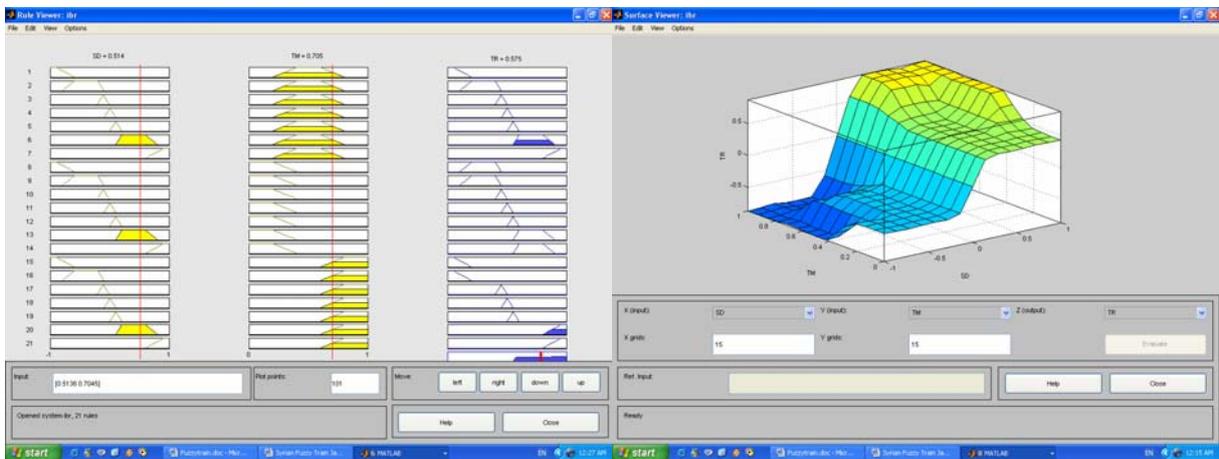


Fig.4 Th fuzzy rules viewer

Fig.5 The fuzzy controller Surface

2) Deriving the fuzzy inference Rules from the Control Objectives

The conversion of speed differences into traveling responses is to take place in accordance with the selected traveling mode .The following rules are associated with the one –to–one conversion in the ‘moderate’ traveling mode:

- Rule1: IF speed deviation very negative And traveling mode moderate THEN travelling response strong braking
- Rule2: IF speed deviation negative And traveling mode moderate THEN traveling response braking
- Rule3: IF speed deviation slightly negative And travelling mode moderate THEN travelling response slight braking
- Rule4: IF speed deviation zero And traveling mode moderate THEN traveling response constant speed
- Rule5: IF speed deviation slightly positive And traveling mode moderate THEN traveling response slight acceleration
- Rule6: IF speed deviation positive And traveling mode moderate THEN traveling response acceleration
- Rule7: IF speed deviation very positive And traveling mode moderate THEN traveling response strong acceleration

The production rules for "Slow" Traveling mode are the same as for ‘Moderate’ , the only two differences are that for the first rule the consequence should be ‘Braking ‘ instead of ‘Strong Braking" ,and for the seven rule: ‘Acceleration’ instead of ‘Strong Acceleration” . Furthermore,the production rules for"Fast" traveling mode is different to the ‘Moderate’ one, in that not only very negative or positive speed deviations , but also normal ones cause ‘Strong Braking’ or ‘Strong Acceleration’, i.e. to alter the consequenses of second and six rules of ‘Moderate’ travelling mode , respectively. Now we have 21 IF-THEN fuzzy rules which are to put up together,the AND operator is selected

as the aggregation operator , the Max-Min inference mechanism is used in the rule evaluation process , and OR as the accumulation operators. The final step in the fuzzy logic controller design is

to combine the fuzzy outputs into a crisp output , the Center of Gravity (COG) method is used in our application due to better results it gives . Finally ,the PI-like FC has been designed ,the fuzzy rules viewer,and thedesigned FC Surface are shown in Figs.4,and 5,respectively.

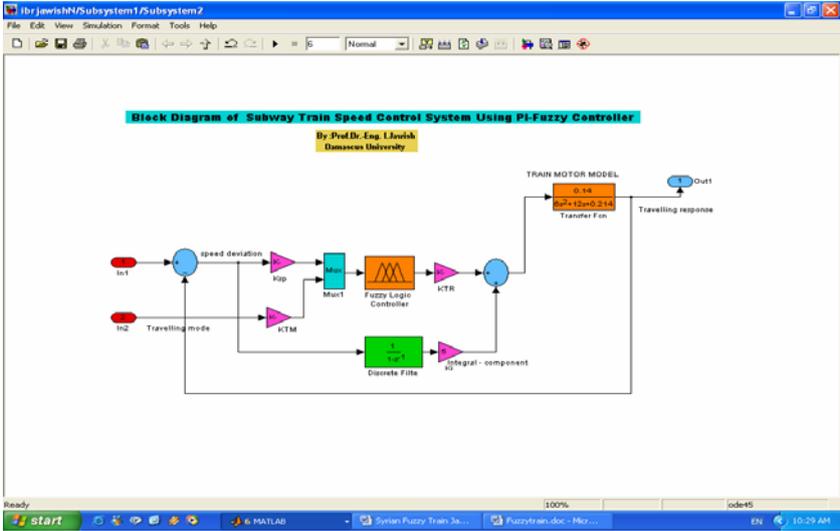


Fig.6 Block Diagram of the Train Speed fuzzy Control System



Figure 7. Speed Traveling Responses at different Travelling Modes
Slow = 10km/h,Moderate=50 km/h,Fast=90 km/h

IV-Computer Simulation Results

Using MATLAB 7.0.2 \ SIMULINK & Fuzzy TOOLBOX software and environment , we now use the computer simulation to demonstrate the behavior of the Subway Train Speed Control System using PI- like FC .

The mathematical model of of the traction dc traction motor for the purpose of simulation was adopted from [10]:

$$G_m(s) = \omega(s)/u(s) = 0.14/ 6s^2 + 3.4s+ 0.214 \quad (5)$$

The designed and investigated Subway Train Speed Control System using PI- Fuzzy is depicted in Figure.6 .

The speed traveling responses to a step-unit at different traveling modes is shown Figure 7. Analyzing the transient curves of the train fuzzy control system we can conclude that the designed PI-Like fuzzy controller has a sufficient performance

V- Conclusion

In this paper we have investigated the influence of the train traveling mode on the speed traveling response , which must be taken into consideration .In view of the fact that the kinetic energy is proportional to the square value of the actual train speed the amount of this energy might be more or less than the calculated nominal one(at moderate mode) , in such case applying the same value of strong acceleration or braking at slow mode is worthless, saving electrical energy ,and enhancing riding comfort . On the other side , at the fast traveling mode, even medium deviations in speed demands a strong acceleration or braking to overcome the increment of train inertia, saving travel time, and improving traceability .

References

- [1] Yen J.,Langari R.,Zadeh L.A., (Eds),Industrial Applications of Fuzzy Logic and Intelligent Systems , IEEE PRESS,1995 .
- [2] Zilouch , Jamshidi Mo.(ed) . Intelligent Control Systems Using Soft Computing Methodologies, CRC,2001.
- [3] Ying H., Fuzzy Control Modeling , Analytical Foundations and Applications ,IEEE PRESS, 2000.
- [4] Yager R.R., Dimitar P.F. Essentials of Fuzzy Modeling and Control .WILEY-INTERSCIENCE,1994.
- [5] Driankov D. Hellendroon H., Reinfrank M., An Introduction to Fuzzy Control (2ed), Springer –Verlag , 1996.
- [6] Tang K.S. An Optimal Fuzzy PID Controller, IEEE Transactions on Industrial Electronics, VOL. 48,NO 4, August 2001.
- [7] Tang W,Chen G. , Lu r. A modified fuzzy PI controller for a flexible-joint robot arm with uncertainties, Fuzzy Sets and Systems 118 (2001) 109-119.
- [8] Victor J, Adaptive Scaling Factors Algorithm for the Fuzzy Logic Controller. zevictor@dei.uc.pt.
- [9] Leybold Fuzzy Control Catalogue 2000.
- [10] MATLAB 7.0.2 /SIMULINK .

Authors:

Prof. Dr.-Eng., Ibrahim Jawish
Damascus University
+963-11,Damascus,Syria
Phone: +963-11-3121214
Fax: : +963-11-3320326
E-mail: ibrahim-jawish@fulbrightweb.org

Y. Gu / H. Su / J. Chu

An Approach for Transforming Nonlinear System Modeled by the Feedforward Neural Networks to Discrete Uncertain Linear System

ABSTRACT

In this paper, an approach for transforming a state-space or input-output representation nonlinear system modeled by feedforward NN to a discrete uncertain linear system is presented. First, The dynamics of the feedforward NN model are represented by linear difference inclusions (LDI's). Then, the LDI's are interpreted as discrete linear system with norm-bounded or polytopic time-varying uncertainties. Finally, we use some simulation results of a typical continuous stirred tank reactor (CSTR) to elaborate on the details of the above approach.

1. Introduction

Nonlinear system modeling and control is still an open problem now. Though nonlinear system modeling based on Neural Networks (NN) has made a great progress in recent years, the nonlinear controller synthesis for nonlinear system modeled by NN is still hard to be realized online due to the heavy computational burden and the difficulty in convergence. As a result, the most practical strategies recently are to linearize the involved NN model in the model based controller synthesis. In most cases, the uncertainties caused by linearization could not be neglected. Otherwise, the stability and performance characteristics of the nonlinear control system are not easy to be assessed. In the last few years, the research on robust stability and performance of linear systems with norm-bounded or polytopic uncertainties have achieved a lot^[1]. So we want to build a bridge between the NN model and the uncertain linear system to integrate the nonlinear system control with NN into robust control^{[2],[3]}. In this paper, a bridge between the feedforward NN and linear systems with norm-bounded or polytopic uncertainties is proposed.

2. Preliminaries

Usually, discrete nonlinear system could be described as state-space form or input-output form:

State-space form:

$$\begin{aligned} x(k) &= f[x(k-1), u(k-1)] \\ y(k) &= Cx(k) \end{aligned} \quad (2.1)$$

Input-output form:

$$y(k) = g[y(k-1), \dots, y(k-n), u(k-1), \dots, u(k-n)] \quad (2.2)$$

where f and g are nonlinear functions. f and g could be modeled by the Feedforward Neural Networks(FNN), which is sufficient to uniformly approximate arbitrary nonlinear mapping.

And, we describe some specific families of LDI's that we often encounter in robust control study:

Norm-bounded LDI's:

$$x(k+1) = (A_0 + \Delta A)x(k) + (B_0 + \Delta B)u(k) \quad (2.3)$$

$$y(k) = Cx(k)$$

$$[\Delta A \quad \Delta B] = DF(k)[E_1 \quad E_2] \quad (2.4)$$

where $\Delta A \in R^{n \times n}$ and $\Delta B \in R^{n \times m}$ represent time-varying parameter uncertainties in the system model. $D \in R^{n \times r}$, $E_1 \in R^{l \times n}$ and $E_2 \in R^{l \times m}$ are known real constant matrices. $F(k)$ is an unknown time varying matrix, satisfying

$$N = \{F(k) \in R^{r \times l} : F^T(k)F(k) \leq I\} \quad (2.5)$$

Polytopic LDI's:

$$x(k+1) = A(k)x(k) + B(k)u(k)$$

$$y(k) = Cx(k) \quad (2.6)$$

$$[A(k) \quad B(k)] \in \Omega$$

where Ω is polytope, $\Omega = Co\{[A_1 \quad B_1], [A_2 \quad B_2], \dots, [A_L \quad B_L]\}$ where Co devotes to the convex hull.

According to a simple extension from [1], there are two corollaries as the following:

Corollary 2.1 State-space form (2.1) could be described as the uncertain linear system as follows:

$$\begin{aligned} \bar{x}(k) &= A(k)\bar{x}(k-1) + B(k)\bar{u}(k-1) \\ y(k) &= Cx(k) \end{aligned} \quad (2.7)$$

$$[A(k) \quad B(k)] \in Co\left[\frac{\partial f}{\partial x(k-1)}, \frac{\partial f}{\partial u(k-1)}\right] \quad (2.8)$$

where, $\bar{x}(k) = x(k) - \tilde{x}(k)$, $\bar{u}(k) = u(k) - \tilde{u}(k)$. $[\tilde{x}(k) \quad \tilde{u}(k)]^T$ could be a known equilibrium trajectory.

Corollary 2.2 For input-output form (2.2), denote

$$p(k) = [y(k-1), \dots, y(k-n), u(k-1), \dots, u(k-n)]$$

then (2.2) could be described as the uncertain linear system as follows:

$$\bar{y}(k) = H(k)\bar{p}^T(k) \quad (2.9)$$

$$H(k) \in Co \left[\frac{\partial g}{\partial y(k-1)}, \dots, \frac{\partial g}{\partial y(k-n)}, \frac{\partial g}{\partial u(k-1)}, \dots, \frac{\partial g}{\partial u(k-n)} \right] \quad (2.10)$$

where $\bar{y}(k) = y(k) - \tilde{y}(k)$, $\bar{p}(k) = p(k) - \tilde{p}(k)$. $[\tilde{y}(k) \ \tilde{p}(k)]^T$ could be a known equilibrium trajectory.

Since most robust control theory focus on state-space model, we now consider the state-space observer realization of (2.9):

$$\begin{aligned} x(k+1) &= A(k)x(k) + B(k)\bar{u}(k) \\ \bar{y}(k) &= Cx(k) \end{aligned} \quad (2.11)$$

$$A(k) = \begin{bmatrix} 0 & \cdots & 0 & A_n(k) \\ I & \cdots & 0 & A_{n-1}(k) \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & I & A_1(k) \end{bmatrix} \quad B(k) = \begin{bmatrix} B_n(k) \\ B_{n-1}(k) \\ \vdots \\ B_1(k) \end{bmatrix} \quad (2.12)$$

$$C = [0 \ 0 \ \cdots \ I]$$

$$\text{where, } A_i(k) = \frac{\partial g}{\partial y(k-i)}, \quad B_i(k) = \frac{\partial g}{\partial u(k-i)}$$

3. Transforming FNN to Discrete Uncertain Linear System

In this section, we discuss the problem of the formulation for transforming weights of FNN to parameter matrices of norm-bounded uncertain linear system and vertexes of polytopic uncertain linear system. The active function in the neuron of FNN is Tansig function. The output value of Tansig function is bounded in (-1,1).

3.1 Norm-bounded Uncertainties

3.1.1 System(2.1) modeled by a single hidden layer FNN

Consider the nonlinear function f in system(2.1), which is modeled by a single hidden layer FNN. The Jacobian matrix of the FNN is as the following:

$$J = \begin{bmatrix} \frac{\partial f}{\partial x(k-1)} & \frac{\partial f}{\partial u(k-1)} \end{bmatrix} = W_2 \times \text{diag}(1 - a \times a) \times W_1 \quad (3.1)$$

where $a = f(W_1 \times p + B_1)$, is output vector of the hidden layer, \times represent the array multiplication. $\text{diag}(a_{s \times 1})$ denotes a square diagonal matrix with $a(1), \dots, a(s)$ on the main diagonal. W_1 is the matrix of connecting weights between the input-layer and the hidden-layer, W_2 is that between the hidden-layer and the output-layer.

According to Corollary 2.1, system (2.1) could be transformed to (2.7). Hence,

$$[A(k) \ B(k)] = W_2 \times \text{diag}(1 - a \times a) \times W_1 \quad (3.2)$$

Then system (2.1) could be further transformed to (2.3). The nominal model is the linearization model on an equilibrium.

$$[A_0 \quad B_0] = W_2 \times \text{diag}(1 - a_0 \times a_0) \times W_1 \quad (3.3)$$

The uncertainties are as the following:

$$[\Delta A(k) \quad \Delta B(k)] = W_2 \times \text{diag}(a_0 \times a_0 - a \times a) \times W_1 \quad (3.4)$$

$$D = W_2, \quad [E_1 \quad E_2] = W_1, \quad F(k) = \text{diag}(a_0 \times a_0 - a \times a) \quad (3.5)$$

And the value of each element in the diagonal matrix $F(k)$ is in $(-1, 1)$, so obviously the norm-bounded condition (2.5) is satisfied.

3.1.2 System(2.2) modeled by a single hidden layer FNN

Consider the nonlinear function g in system(2.2), which is modeled by a single hidden layer FNN with the input vector $p(k) = [y(k-n), \dots, y(k-1), u(k-n), \dots, u(k-1)]$.

According to Corollary 2.2, system (2.2) could be transformed to (2.9), and further transformed to state-space model (2.11), (2.12) . Then we have,

$$[A_n(k), \dots, A_1(k), B_n(k), \dots, B_1(k)] = W_2 \times \text{diag}(1 - a \times a) \times W_1 \quad (3.6)$$

We get W_1 partitioned with appropriate dimensions, and substituted the partitioned W_1 in (2.12) to get:

$$A(k) = \begin{bmatrix} 0 & \dots & 0 \\ I & \dots & 0 & W_{1A}^T \times \text{diag}(1 - a \times a) \times W_2^T \\ \vdots & \ddots & \vdots \\ 0 & \dots & I \end{bmatrix} \quad (3.7)$$

$$B(k) = W_{1B}^T \times \text{diag}(1 - a \times a) \times W_2^T \quad (3.8)$$

The nominal model could be set to the linearization model on one equilibrium.

Then (3.7),(3.8) could be transformed to(2.3),(2.4):

$$A_0 = \begin{bmatrix} 0 & \dots & 0 \\ I & \dots & 0 & W_{1A}^T \times \text{diag}(1 - a^0 \times a^0) \times W_2^T \\ \vdots & \ddots & \vdots \\ 0 & \dots & I \end{bmatrix} \quad (3.9)$$

$$B_0 = W_{1B}^T \times \text{diag}(1 - a^0 \times a^0) \times W_2^T \quad (3.10)$$

$$[\Delta A(k) \quad \Delta B(k)] = DF(k)[E_1 \quad E_2] \quad (3.11)$$

$$\text{where } F(k) = \begin{bmatrix} F_A(k) \\ F_B(k) \end{bmatrix}, \quad D = [D_A \quad D_B], \quad E_1 = \begin{bmatrix} E_A \\ 0 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 0 \\ E_B \end{bmatrix} \quad (3.12)$$

$$D_A = W_{1A}^T, \quad D_B = W_{1B}^T, \quad F_A(k) = F_B(k) = \text{diag}(a^0 \times a^0 - a \times a), \quad (3.13)$$

$$E_A = \begin{bmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix} W_2^T, E_B = W_2^T \quad (3.14)$$

Since the output value of Tansig function in hidden neuron is in $(-1,1)$, so obviously $F^T(k)F(k) < I$ is satisfied.

3.2 Polytopic Uncertainties

The above norm-bounded LDI's are all diagonal norm-bounded LDI with an uncertain diagonal matrix $F(k)$. And the diagonal norm-bounded uncertainties could be described as polytopic uncertainties^[1]. The vetexes are the linear nominal models when the elements of $F(k)$ take their extreme value. The number of the vetexes is obtained according to permutation and combination. There are two corollaries as the following:

Corollary 3.1 The nonlinear system modeled by a single hidden-layer feedforward NN could be represented by a polytopic uncertain system with 2^s vetexes at most, where s is the number of nodes of the hidden-layer.

Corollary 3.2 The nonlinear system modeled by a double hidden-layers feedforward NN could be represented by a polytopic uncertain system with $2^{s+t} - 2^s - 2^t + 2$ vetexes at most, where s is the number of nodes of the first hidden-layer, and t is that of the second one.

4. An Example

Let's consider a typical CSTR process^[4] now. The single hidden layer FNN for modelling the CSTR is first identified by the input-output data. The topology structure of the FNN is 3-4-2.

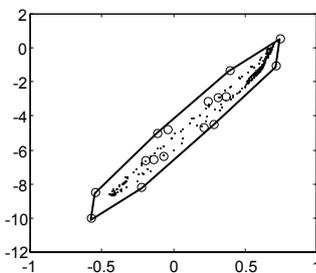


Fig. 4.1 a_{11} and a_{21} of $A(k)$

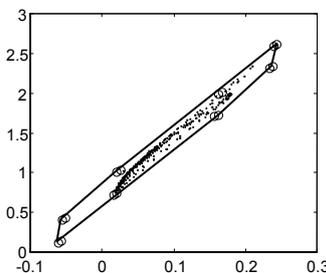


Fig.4.2 a_{12} and a_{22} of $A(k)$

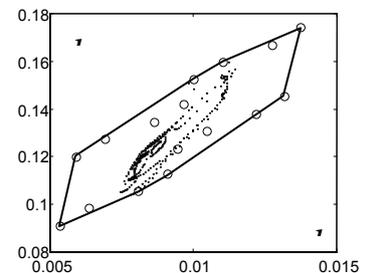


Fig. 4.3 b_1 and b_2 of $B(k)$

Then, the FNN could be transformed to an uncertain linear system (3.2)~(3.5). where $A(k) \in R^{2 \times 2}$, $B(k) \in R^{2 \times 1}$, and the equilibrium is selected as the middle unstable equilibrium(0.447, 2.751, 0). We group every two elements of the matrices $A(k), B(k)$

and project the uncertainties of the elements to the Fig. 4.1~4.3. The dots in the figures are obtained by the equations (3.3) on every input-output sampling data. According to Corollary 3.1, the polytopic uncertain system has $2^4 = 16$ vertices at most. The circles in the figures are the potential vertices.

5. Conclusion

Based on the above work, the robust control strategies could be applied to the nonlinear controller synthesis incorporating the feedforward NN model. And, the robust stability and robust performance of the corresponding nonlinear control system is convenient to be assessed. Then a systematic procedure for nonlinear process control could be provided with our contribution. We could firstly model the real process with the feedforward NN model just based on the input and output data. Secondly, the resulted NN would be transformed to uncertain linear system with the approach in this paper. Thirdly, many robust control strategies could be attempted to analyse and synthesis the uncertain linear system. Finally, the derived controller is applied to the real process with some anticipative stability and performance. It is noted that uncertain linear system is always conservative compared with the feedforward NN. The conservatism has direct impact on the successful implementation of this procedure. Then how to reduce the conservatism in this procedure will be further studied in the future.

References

- [1] Boyd, S. et al. 1994. Linear Matrix Inequalities in systems and control theory. SIAM.
- [2] Tanaka, K. 1996. An Approach to Stability Criteria of Neural Network Control Systems. IEEE Trans. on Neural Networks, 7(3): 629-642.
- [3] Limanond, S. and Si, J. 1998. Neural Network-Based Control Design: An LMI Approach. Proceeding of ACC, Philadelphia, Pennsylvania, June, 970-974.
- [4] Uppal, A., Ray, W.H. and Poore, A.B. 1974. On the Dynamic Behavior of Continuous Stirred Tank Reactors. Chem. Eng. Sci. Vol.29: 967-985.

Authors:

Associate Prof. Yong Gu
Prof. Hongye Su
Prof. Jian Chu
Inst. Of Advanced Process Control
Yuquan Campus, Zhejiang Univ.
Hangzhou, ZHEJIANG 310027
P.R.China
Phone: : +86 571 86667611
Fax: +86 571 86667616
E-mail: guyong@mail.hz.zj.cn

3 Optimisation and Management of Complex Systems and Networked Systems

R. Franke / J. Doppelhamer

Advanced model based control in the Industrial IT System 800xA

Abstract

Advanced model based control is a promising technology that can improve the productivity of industrial processes. In order to find its way into regular applications, advanced control must be integrated with the industrial control systems. Modern control systems, on the other hand, need to extend the reach of traditional automation systems – beyond control of the process – to also cover the increasing amount of information technology (IT) required to successfully operate industrial processes in today's business markets. The Industrial IT System 800xA from ABB provides a scalable solution that spans and integrates loop, unit, area, plant, and interplant controls.

This paper introduces the 800xA and the underlying Aspect Object technology. It is shown how model knowledge and optimization solver technology are integrated into the 800xA framework. This way, advanced model based control solutions can be set up in an open and modularly structured way. New model and solver aspects can be combined with available aspects covering standard functionality like process connectivity, management of process data, trend&history data and application data, as well as operator graphics.

A Nonlinear Model-based Predictive Controller (NMPC) for power plant start-up is treated as example. This paper discusses how NMPC can be integrated with a modern control system so that standard concepts are re-used for this advanced model based control concept.

Introduction

During the last decades, several advanced control technologies have been developed, including adaptive control, fuzzy control and neuro control. While each of these technologies offers advantages over classical control methods, PID controllers still dominate the vast majority of industrial applications.

One reason for the lack of mainstream use of advanced control technologies is seen in the fact that they require specialized engineering knowledge and tools. Normally, specialized experts are required to apply advanced control methods. A better integration of advanced control technologies with regular control systems is seen as a key factor for improved acceptance.

Nonlinear model based control (NMPC) has received much attention during the last years. The technology has several advantages from a control point of view: it accommodates nonlinear, multi-variable problems with state constraints. Important achievements have been made to treat the computationally challenging task of formulating and solving large-scale nonlinear optimization problems on-line [4, 2, 6]. Moreover, NMPC has the advantage that the technology is more open, compared to other advanced control methods. Models do represent the behavior of a plant and standard optimization algorithms are used to apply the models to control. This openness improves the acceptance of NMPC on the one hand side.

On the other side, still special purpose tools are required to implement model based control. This implies that concepts which are readily available in a standard control system need to be

specifically interfaced or even redeveloped for applications of model based control, including e.g. signal exchange with sensors, actuators and low level controls, operator graphics, trend&history display, signaling of alarms and events, as well as system maintenance. This is seen as an important burden for both: acceptance and cost of NMPC.

The Industrial IT System 800xA

The Industrial IT System 800xA seamlessly integrates traditionally isolated plant devices and systems, extending the reach of the automation system to all plant areas. The result is a simplified, software representation of the plant, from simple on/off-type switches and valves to smart field devices, dedicated control subsystems, and PC-based supervisory systems [1].

The framework for the 800xA system architecture is built upon ABB's Aspect Object technology. Aspect Objects relate plant data and functions – the aspects, to specific plant assets – the objects. Aspect objects represent real objects, such as process units, devices and controllers. Aspects are informational items, such as I/O definitions, engineering drawings, process graphics, reports and trends that are assigned to the objects in the system.

Aspect Objects are organized in hierarchical structures that represent different views of the plant. One object may be placed multiple times in different structures. Examples for different types of structures are:

Functional Structure: Shows the plant from the process point of view.

Location Structure: Shows the physical layout of what equipment is located where in the plant.

Control Structure: Shows the control network in terms of networks, nodes, fieldbuses, and stations.

The idea of placing the same object in multiple structures is based on the IEC standard 1346 [9, 3].

Integration of model based control

A new Model aspect has been developed so that mathematical model information can be added to an Aspect Object. The model has the form of a hybrid differential algebraic equation system (hybrid DAE)

$$\mathbf{0} = \mathbf{F}[\mathbf{x}(t), \dot{\mathbf{x}}(t), \mathbf{m}(t), \mathbf{u}(t), \mathbf{z}(t), \mathbf{y}(t), \mathbf{p}, t], \quad (1)$$

$$\mathbf{F} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_m} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_z} \times \mathbb{R}^{n_y} \times \mathbb{R}^{n_p} \times \mathbb{R}^1 \mapsto \mathbb{R}^{n_x}$$

$$\mathbf{m}(t) := \mathbf{G}[\mathbf{x}(t), \mathbf{m}(t), \mathbf{u}(t), \mathbf{z}(t), \mathbf{y}(t), \mathbf{p}, t], \quad (2)$$

$$\mathbf{G} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_m} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_z} \times \mathbb{R}^{n_y} \times \mathbb{R}^{n_p} \times \mathbb{R}^1 \mapsto \mathbb{R}^{n_m}.$$

Here \mathbf{x} denote continuous-time states, \mathbf{m} are discrete modes, \mathbf{u} and \mathbf{z} are controlled and not-controlled inputs, respectively, \mathbf{y} are outputs and \mathbf{p} are model parameters. Discrete modes are variables that change their values only at discrete time instants, so called event instants t_e . See [10] for more information on the treated hybrid DAE.

The Model aspect holds information related to the model, including

- Declaration of model variables in categories (Parameter, Input, Output, State, Generic),
- Values for model variables, e.g. for parameters,
- References to process signals, e.g. for inputs and outputs,
- Structural information for hierarchical sub-model structure,
- Reference to the implementation of the model.

The Model aspect does not provide any functionality nor does it deal with implementation details. Instead it references an external implementation. In this way available modeling tools can be applied and expensive re-implementation is avoided.

A model can be used to perform one or more model-based activities. A second aspect, the Dynamic Optimization aspect has been developed to interface a numerical solver, hold the solver configuration, and to exchange data between the solver and the control system. The exchanged data includes: configuration data, current process values (like sensor values and controller set-points), and history logs. Predictions are written back to the control system as history logs with future time stamps.

The integrated solver HQP is primarily intended for structured, large-scale nonlinear optimization [7]. It implements a Sequential Quadratic Programming algorithm that treats nonlinear optimization problems with a sequence of linear-quadratic sub-problems. The sub-problems are formed internally by simulating the model and by analyzing sensitivities. They are solved with an interior point method that is especially suited for a high number of inequality constraints, e.g. resulting from the discretization of path constraints. See [6], [8], and [7] for more details about the solver.

The treated model based activities include

- Initial value simulation for specified initial states $\mathbf{x}(t_0)$ and model inputs,
- Estimation of model parameters and initial states,
- Nonlinear optimal control with constraints on model inputs and outputs,
- Steady-state simulation, estimation and optimization at one time instant.

The object-oriented, physical modeling technology Modelica is used to build the models [10]. A physical plant model is built on available model libraries [5]. It is used by both: state estimator and optimizer. Moreover, specific preprocessor and the postprocessor models are formulated as computational algorithms in Modelica. The scheduler model is formulated as state graph [12].

Application example

A Nonlinear Model-based Predictive Controller (NMPC) for power plant start-up serves as example. The start-up problem is challenging as it is highly nonlinear in the covered large range of operation. Thermal stress occurring in thick walled components needs to be kept in given limits. Multiple manipulated variables must be coordinated. A long prediction horizon is required to fulfill the constraints during a start-up.

Figure 1 shows a process diagram of a power plant. Feed water goes through pre-heaters and the economizer into the evaporator, as seen in the lower left section of the diagram. Saturated steam leaving the evaporator is super-heated in several super-heater stages. The example uses

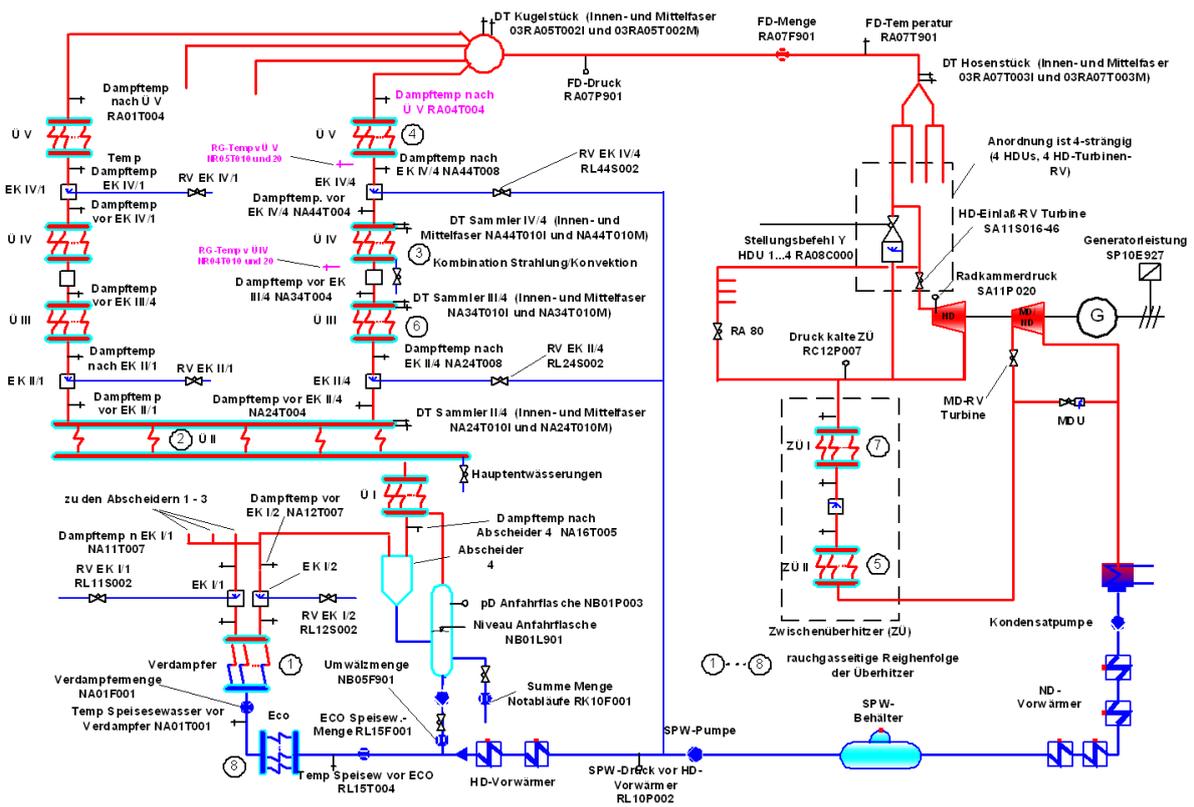


Figure 1: Simplified process diagram of a power plant.

five super-heater stages and four parallel streams, as seen in the upper left section of the diagram. The live steam leaving the boiler goes to the turbine. The example uses two turbine sections. In the turbine, thermal energy is transformed to mechanical energy, driving the generator. Afterwards the steam is condensed and water flows back to the feed water tank, as seen in the lower right section of the diagram.

A boiler model was built using the Modelica technology [5]. The model needs to be carefully designed so that it expresses the relationship between optimized control actions (fuel flow rate and valve positions) and constrained process values (pressures, temperatures and thermal stresses). In the example described here, a system of differential-algebraic equations (DAE) with 940 variables was built, using measurements of about 150 process values. The Dynamic Optimization aspect system was used off-line to identify model parameters based on data logs available for historical start-ups.

During a run of the NMPC, an optimization problem is solved on-line every minute. The model is adapted to the process based on 36 on-line signals. 18 values are communicated back to the process, including three controller set points and additional signals for switch conditions and operator displays. The time horizon for prediction and control is 90 minutes in the example. It gets divided into 90 sample periods. The optimized manipulated variables are parameterized piecewise linear. All other model variables are evaluated at the sample time points. This means that overall 85540 variables are present in the on-line optimization problem. The solution time is about five minutes for a cold start of the solver and about 40 seconds for a subsequent solver run. Please see [8] for details about the numerical formulation and solution of the optimization problem.

Figure 2 shows an operator display for boiler start-up optimization. The trend plot displays the

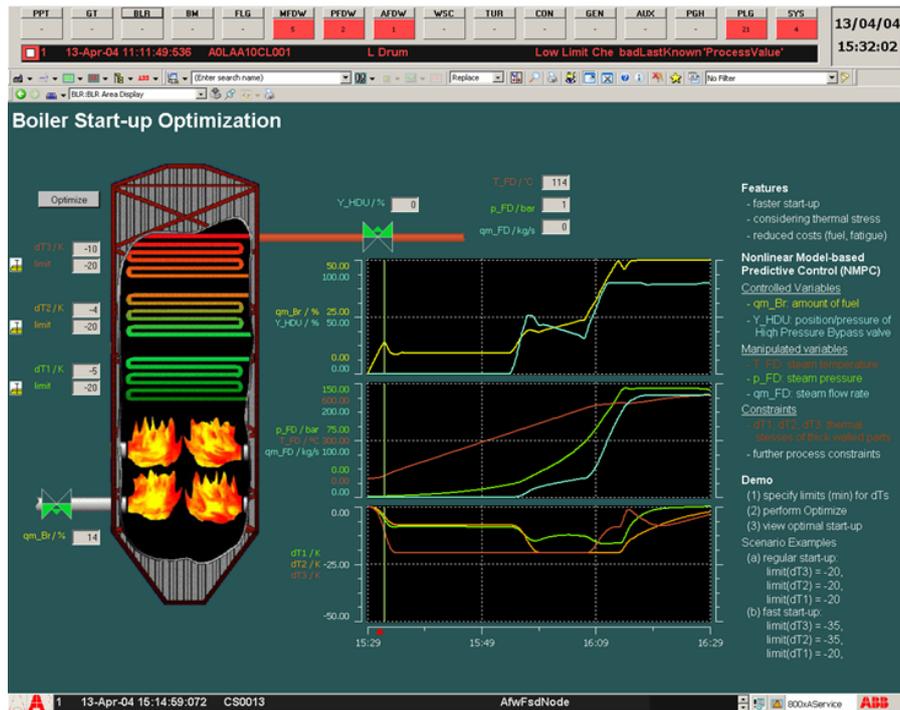


Figure 2: Operator display showing the optimal start-up predicted by the NMPC, in addition to current process values and history logs.

manipulated variables in the upper pane, the main process variables (live steam parameters) in the middle pane and constrained thermal stresses in the lower pane. As a result of the optimization, the process is driven along the allowed limits for thermal stresses.

Traditionally an operator display shows current process values and history logs. As a by-product of model predictive control, the operator can additionally see the prediction of the future behavior of the plant. As the NMPC runs integrated with the control system, this display can easily be configured.

Using the NMPC, the start-up time could be reduced by about 20 minutes and the start-up costs by 10% as compared to a well tuned classical control.

Conclusions

Nonlinear Model-based Predictive Control (NMPC) is a promising control technology. Due to advances in computational algorithms during recent years, it is now possible formulate and solve the underlying large-scale nonlinear optimization problems on-line under real-time conditions. The example discussed here was developed in detail in [8].

For a successful application of NMPC it is equally important to appropriately integrate the method with the control system. This paper discusses how this is done with the Industrial IT System 800xA by ABB. Based on international standards for control systems engineering and software, the System 800xA architecture and the Aspect Object technology allow a flexible integration of model knowledge and model based applications. Two new aspects have been developed in the Dynamic Optimization system extension. The new aspects can be combined with other available aspects, e.g. for controller connectivity, history logs and process graphics.

The NMPC runs on an application server that is integrated as additional node with the system. Installation and maintenance are identical to other nodes, like data servers and display clients.

This paper uses the start-up of a power plant as example. Batch processes are another promising application area, as described in [11].

References:

- [1] ABB Automation Technologies. Industrial IT System 800xA – System Architecture Overview. <http://www.abb.com>, Document Id: 3BUS092080R0101, 2005.
- [2] Lorenz T. Biegler. Efficient solution of dynamic optimization and NMPC problems. In Frank Allgöwer and Alex Zheng, editors, *Nonlinear Model Predictive Control*, pages 219–243. Birkhäuser, Basel, 2000.
- [3] L.G. Bratthall, R. van der Geest, H. Hoffmann, E. Jellum, Z. Korendo, R. Martinez, M. Orkisz, C. Zeidler, and J. S Andersson. Integrating hundred's of products through one architecture – the Industrial IT architecture. In *International Convergence on Software Engineering*. Orlando, Florida, USA, 2002.
- [4] M. Diehl, R. Findeisen, S. Schwarzkopf, I. Uslu, F. Allgöwer, H.G. Bock, and J.P. Schlöder. An efficient algorithm for optimization in nonlinear model predictive control of large-scale systems. *at – Automatisierungstechnik*, 50(12), 2002.
- [5] H. Elmqvist, H. Tummescheit, and M. Otter. Modeling of thermo-fluid systems – Modelica.Media and Modelica.Fluid. In *Proceedings of the 3rd International Modelica Conference*. Modelica Association, Linköping, Sweden, November 2003.
- [6] R. Franke and E. Arnold. Applying new numerical algorithms to the solution of discrete-time optimal control problems. In K. Warwick and M. Kárný, editors, *Computer-Intensive Methods in Control and Signal Processing: The Curse of Dimensionality*, pages 105–118. Birkhäuser Verlag, Basel, 1997.
- [7] R. Franke, E. Arnold, and H. Linke. HQP: a solver for nonlinearly constrained large-scale optimization. <http://hqp.sourceforge.net>.
- [8] R. Franke, K. Krüger, and M. Rode. Nonlinear model predictive control for optimized startup of steam boilers. In *GMA-Kongress 2003*. VDI-Verlag, Düsseldorf, 2003. VDI-Berichte Nr. 1756, ISBN 3-18-091756-3.
- [9] International Electrotechnical Commission. Industrial systems, installations and equipment and industrial products – structuring principles and reference designations. IEC Standard 61346, 1996.
- [10] Modelica Association. Modelica – A Unified Object-Oriented Language for Physical Systems Modeling, Version 2.2. <http://www.modelica.org>, 2005.
- [11] Z.K. Nagy, B. Mahn, R. Franke, and F. Allgöwer. Nonlinear model predictive control of batch processes: an industrial case study. In *16th IFAC World Congress*. Prague, Czech Republic, July 2005.
- [12] M. Otter, J. Årzén, and A. Schneider. StateGraph – a Modelica library for hierarchical state machines. In *Proceedings of the 4th International Modelica Conference*. Modelica Association, Hamburg-Harburg, Germany, March 2005.

Author information:

Rüdiger Franke
ABB AG, Power Technology Systems
Kallstadter Str. 1
68309 Mannheim, Germany
E-mail: Ruediger.Franke@de.abb.com

Jens Doppelhamer
ABB Corporate Research
Wallstadter Str. 59
68526 Ladenburg, Germany
E-mail: Jens.Doppelhamer@de.abb.com

H. Gerbracht/ P. Li/ W. Hong

An efficient optimization approach to optimal control of large-scale processes

INTRODUCTION

In this contribution, we consider the interior-point quasi-sequential approach as an efficient optimization method to handle optimal control problems of large-scale processes. This approach is an extension of the quasi-sequential approach [1]. In the original quasi-sequential approach an active-set strategy was applied. In this work an interior-point strategy for the efficient solution of large-scale dynamic optimization problems will be proposed. Numerical experiments showed that with an increasing number of active constraints the computational effort rises significantly (for detailed results see [1]). This high computational costs are a result of the active-set implemented in the recent quasi-sequential approach. Typically the finding of an active-set represents a massive combinatorial problem if the number of active constraints is high [3]. Therefore for this type of processes the active-set line-search strategy is not suitable. An alternative line-search strategy, the interior-point method, does not result in a comparable increase in computational cost in similar circumstances.

STATE-OF-THE-ART

Dynamic optimization approaches can be divided in two main branches, the direct and indirect methods. Current research concentrates on the direct methods. They can be distinguished in simultaneous, sequential and combined approaches. The combined approaches integrate the advantages of both, simultaneous and sequential, methods. Two well known combinations of those approaches exist: the multiple-shooting approach [4] and the quasi-sequential approach [1].

The fundamental advantages of the quasi-sequential approach are the following: On the one hand it is possible to maintain the path constraints of state variables. In the simultaneous approach the maintaining of this path constraints is realized by including

the state and control variables on discrete time points in the optimization. In the quasi-sequential approach, however, this is ensured by enforcing the equality and inequality constraints at the collocation points in the NLP solver. On the other hand only a small-scale NLP solver is needed as in sequential approaches and in terms of real-time optimization it is important that a feasible path strategy is used in the quasi-sequential framework.

QUASI-SEQUENTIAL INTERIOR-POINT METHOD

The quasi-sequential approach is divided in two layers, the simulation and optimization layer. In the simulation layer both control and state variables are discretized by orthogonal collocation, the equality constraints are solved by a Newton solver and the sensitivities and gradients are calculated. The elimination of the equality constraints by simulation simplifies the line-search problem in the optimization layer considerably and therefore larger steps can be taken towards the optimum. Only the control variables and inequalities are included in the optimization layer. Hence, if we want to modify the line-search method, which is a part of the optimization algorithm, the simulation layer can stay unchanged.

Until now the SQP based IMSL[®] routine DNCONG, which includes an active-set line-search strategy, has been used as the small-scale NLP solver in the optimization layer. This routine has to be replaced for the implementation of the interior-point method. In place of the DNCONG a primal-dual interior-point method (adapted from [2]) with a generalized augmented Lagrangian function as a merit function, which shows excellent convergence properties and computational performances, is employed. Due to the handling of inequality constraints with barrier terms the problem to be solved consists merely of an objective function. This leads to a considerable reduction of computational costs in situations where the number of active constraints is high. In the following section the developed algorithm will be explained in detail.

We consider a general NLP problem by approximating state profiles with orthogonal collocation on finite elements (time intervals) of the form

$$\begin{aligned}
 & \min_{u \in \mathbb{R}^{n-m}} f(z(u_{l,i}), u_{l,i}) & (1) \\
 s.t. \quad & g_{l,i}(z(u_{l,i}), u_{l,i}) = 0 \\
 & u_L \leq u_{l,i} \leq u_U \\
 & z_L \leq z(u_{l,i}) \leq z_U
 \end{aligned}$$

where $i = 1, \dots, NC$ and $l = 1, \dots, NL$ are the encounter for the collocation points and the time intervalls of the discretization, respectively. $u \in \mathbb{R}^{n-m}$ and $z(u) \in \mathbb{R}^m$ are the control (independent) and state (dependent) variables, $f(z(u_{l,i}), u_{l,i}) : \mathbb{R}^n \rightarrow \mathbb{R}$ is the objective function and $g(z(u_{l,i}), u_{l,i}) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ are the equality constraints. u_L, z_L and u_U, z_U are the lower and upper bounds of the control and state variables.

The equality constraints will be solved through simulation of the model equations, which provides the state variables $z(u_{l,i})$ for each collocation point. Additionally the inequality constraints will be replaced by logarithmic barrier terms that have been added to the objective function. The resulting optimization problem is

$$\begin{aligned} \min_{u \in \mathbb{R}^{n-m}} \varphi_\mu(z(u_{l,i}), u_{l,i}) = & f(z(u_{l,i}), u_{l,i}) - \mu \left\{ \sum_{j=1}^{n-m} \ln [u_{l,i}^{(j)} - u_L] + \sum_{j=1}^{n-m} \ln [u_U - u_{l,i}^{(j)}] \right. \\ & \left. + \sum_{j=1}^m \ln [z(u_{l,i})^{(j)} - z_L] + \sum_{j=1}^m \ln [z_U - z(u_{l,i})^{(j)}] \right\} \end{aligned} \quad (2)$$

with the barrier parameter $\mu > 0$ as well as $u_{l,i}^{(j)}$ and $z(u_{l,i})^{(j)}$ as the j th component of the vector $u_{l,i}$ and $z(u_{l,i})$. If we modify notation with $u_{l,i} = u$ and $z(u_{l,i}) = z(u)$ as the vectors of all discretized control and state variables, the following Lagrange function can be defined:

$$\begin{aligned} \mathcal{L}(z(u), u, \nu^L, \nu^U) = & f(z(u), u) - \sum_{j=1}^{n-m} \nu^{LU} [u^{(j)} - u_L] - \sum_{j=1}^{n-m} \nu^{UU} [u_U - u^{(j)}] \\ & - \sum_{j=1}^m \nu^{LZ} [z(u)^{(j)} - z(u)_L] - \sum_{j=1}^m \nu^{UZ} [z(u)^{(j)} - z(u)_U], \end{aligned} \quad (3)$$

where $(\nu^L)^T = [(\nu^{LZ})^T (\nu^{LU})^T]$ and $(\nu^U)^T = [(\nu^{UZ})^T (\nu^{UU})^T]$ are the Lagrange multipliers for the lower and upper bounds of the state and control variables as stated in (1). The related optimality conditions can be written as follows

$$\begin{aligned} \nabla_u f(z(u), u) - \nu^{LU} + \nu^{UU} - (\nu^{LZ})^T \frac{dz(u)}{du} + (\nu^{UZ})^T \frac{dz(u)}{du} &= 0 \\ \nabla_z f(z(u), u) - \nu^{LZ} + \nu^{UZ} &= 0 \\ [U - U_L] V^{LU} e - \mu e &= 0 \\ [U_U - U] V^{UU} e - \mu e &= 0 \\ [Z - Z_L] V^{LZ} e - \mu e &= 0 \\ [Z_U - Z] V^{UZ} e - \mu e &= 0 \end{aligned}$$

with $[U - U_L], [U_U - U], [Z - Z_L], [Z_U - Z]$ and V as diagonal matrices with the related entries on their diagonals and $e = [1, \dots, 1]^T$ in the adequate dimension. The optimality conditions are solved with a Newton method. The search directions

$(d_k^z, d_k^u, d_k^{LU}, d_k^{UU}, d_k^{LZ}, d_k^{UZ})$ at an iteration $(z(u)_k, u_k, \nu_k^{LU}, \nu_k^{UU}, \nu_k^{LZ}, \nu_k^{UZ})$ are obtained by linearization of the optimality conditions:

$$\begin{aligned}
& \begin{bmatrix} W_{UZ,k} & W_{UU,k} & -I & I & -\left(\frac{dz}{du}\right)_k^T & \left(\frac{dz}{du}\right)_k^T \\ W_{ZZ,k} & W_{ZU,k} & 0 & 0 & -I & I \\ 0 & V_k^{LU} & [U - U_L]_k & 0 & 0 & 0 \\ 0 & -V_k^{UU} & 0 & [U_U - U]_k & 0 & 0 \\ V_k^{LZ} & 0 & 0 & 0 & [Z - Z_L]_k & 0 \\ -V_k^{UZ} & 0 & 0 & 0 & 0 & [Z_U - Z]_k \end{bmatrix} \begin{bmatrix} d_k^z \\ d_k^u \\ d_k^{LU} \\ d_k^{UU} \\ d_k^{LZ} \\ d_k^{UZ} \end{bmatrix} \\
& = - \begin{bmatrix} \nabla_u f(z(u_k), u_k) - \nu_k^{LU} + \nu_k^{UU} - (\nu_k^{LZ})^T \frac{dz(u)}{du} + (\nu_k^{UZ})^T \frac{dz(u)}{du} \\ \nabla_u f(z(u_k), u_k) - \nu_k^{LZ} + \nu_k^{UZ} \\ [U - U_L]_k V_k^{LU} e - \mu e \\ [U_U - U]_k V_k^{UU} e - \mu e \\ [Z - Z_L]_k V_k^{LZ} e - \mu e \\ [Z_U - Z]_k V_k^{UZ} e - \mu e \end{bmatrix} \quad (4)
\end{aligned}$$

where $W_k = \nabla^2 \mathcal{L}(z(u)_k, u_k, \nu_k^{LU}, \nu_k^{UU}, \nu_k^{LZ}, \nu_k^{UZ})$ is the Hessian of the Lagrange function with respect to the mentioned indices. Eliminating $\nu_k^{LU}, \nu_k^{UU}, \nu_k^{LZ}$ and ν_k^{UZ} leads to the following system

$$\begin{bmatrix} W_{UZ,k} + \Sigma_{UZ,k} & W_{UU,k} + \Sigma_{UU,k} \\ W_{ZZ,k} + \Sigma_{ZZ,k} & W_{ZU,k} + \Sigma_{ZU,k} \end{bmatrix} \begin{bmatrix} d_k^z \\ d_k^u \end{bmatrix} = - \begin{bmatrix} \nabla_u \varphi_\mu(z(u_k), u_k) \\ \nabla_z \varphi_\mu(z(u_k), u_k) \end{bmatrix} \quad (5)$$

with

$$\begin{aligned}
\Sigma_{UZ,k} &= [Z - Z_L]_k^{-1} \left(\frac{dz}{du}\right)_k^T V_k^{LZ} + [Z_U - Z]_k^{-1} \left(\frac{dz}{du}\right)_k^T V_k^{UZ} \\
\Sigma_{UU,k} &= [U - U_L]_k^{-1} V_k^{LU} + [U_U - U]_k^{-1} V_k^{UU} \\
\Sigma_{ZZ,k} &= [Z - Z_L]_k^{-1} V_k^{LZ} + [Z_U - Z]_k^{-1} V_k^{UZ} \\
\Sigma_{ZU,k} &= 0
\end{aligned}$$

and

$$\begin{aligned}
d_k^{LU} &= -V_k^{LU} e + [U - U_L]^{-1} (\mu e - V_k^{LU} d_k^u) \\
d_k^{UU} &= -V_k^{UU} e + [U_U - U]^{-1} (\mu e + V_k^{UU} d_k^u) \\
d_k^{LZ} &= -V_k^{LZ} e + [Z - Z_L]^{-1} (\mu e - V_k^{LZ} d_k^z) \\
d_k^{UZ} &= -V_k^{UZ} e + [Z_U - Z]^{-1} (\mu e + V_k^{UZ} d_k^z)
\end{aligned}$$

as the search directions for the Lagrange multipliers. Additionally in comparison to [2] in Σ_k the sensitivities have been considered. d_k^z and d_k^u can be obtained by the solution of the following QP subproblem:

$$\min_{d \in \mathbb{R}^n} \nabla \varphi(x_k) d_k^x + \frac{1}{2} (d_k^x)^T (W_k + \Sigma_k) d_k^x \quad (6)$$

with $x^T = (z^T u^T)$ for better readability. Note that, in contrast to [2], this optimization problem contains no equality constraints. The solution can directly be calculated as

$$d_k^x = - (W_k + \Sigma_k)^{-1} \nabla \varphi(x_k) \quad (7)$$

In order to obtain the next iterate the step sizes $\alpha_k \in (0, \alpha_k^{max}]$, $\alpha_k^\nu \in (0, 1]$ must be determined. A first approximation for the step-length is given by the *fraction-to-the-boundary* rule to maintain the positivity of the variables:

$$\begin{aligned}\alpha_k^{max} &:= \max \{ \alpha \in (0, 1] : z(u_k) + \alpha d_k^z \geq (1 - \tau_j) z(u_k) \} \\ \alpha_k^{max} &:= \max \{ \alpha \in (0, 1] : u_k + \alpha d_k^u \geq (1 - \tau_j) u_k \} \\ \alpha_k^\nu &:= \max \{ \alpha \in (0, 1] : \nu_k^L + \alpha d_k^{\nu L} \geq (1 - \tau_j) \nu_k^L \} \\ \alpha_k^\nu &:= \max \{ \alpha \in (0, 1] : \nu_k^U + \alpha d_k^{\nu U} \geq (1 - \tau_j) \nu_k^U \}\end{aligned}\quad (8)$$

where $\tau_j = \max \{ \tau_{min}, 1 - \mu_j \} \in (0, 1)$ is the *fraction-to-the-boundary* parameter [2]. A sufficient decrease condition on a merit function is used to give an second approximation for the step-length by applying the *Armijo condition*:

$$M_{\mu_k}(x_{k+1}, \nu_{k+1}^L, \nu_{k+1}^U; \rho) \leq M_{\mu_k}(x_k, \nu_k^L, \nu_k^U; \rho) + c_k \alpha_k \beta \nabla \Phi_{\mu_k}(x_k, \nu_k^L, \nu_k^U; \rho) \Delta x_k \quad (9)$$

with $x^T = (z^T u^T)$, $\nu_L^T = (\nu_{LZ}^T \nu_{LU}^T)$, $\nu_U^T = (\nu_{UZ}^T \nu_{UU}^T)$, $c > 0$, $\beta \in (0, 1)$ and $\alpha_k = p^t \alpha_k^{max}$, where t is the smallest nonnegative integer value such that α_k satisfies the above condition [5]. For any $\mu > 0$ the merit function M_μ is defined by

$$M_\mu(x, \nu_L, \nu_U; \rho) = \mathcal{L}(x, \nu_L, \nu_U) + \rho \Phi_\mu(x, \nu_L, \nu_U) \quad (10)$$

where ρ is a nonnegative penalty parameter and Φ_μ is the penalty term

$$\begin{aligned}\Phi_\mu(x, \nu_L, \nu_U) &= \nu_L^T [x - x_L] + \nu_U^T [x_U - x] \\ &\quad - \mu \left\{ \sum_{j=1}^n \ln(\nu_{L,j} [x - x_L]_j) + \sum_{j=1}^n (\nu_{U,j} [x_U - x]_j) \right\}\end{aligned}\quad (11)$$

In Eq. (10) the penalty parameter $\rho = \tilde{\rho} + c$ is calculated by

$$\tilde{\rho} = \frac{\nabla_x \mathcal{L}^T \Delta x + \nabla_{\nu_L} \mathcal{L}^T \Delta \nu_L + \nabla_{\nu_U} \mathcal{L}^T \Delta \nu_U}{\nabla \Phi_\mu^T \Delta x} \quad (12)$$

with a given c . In the global line-search interior-point algorithm with given $\tilde{c} > 0$ and ρ the current penalty parameter to update ρ is

$$\rho_+ = \begin{cases} \tilde{\rho}_+ + \tilde{c}, & \text{if } \tilde{\rho}_+ + \tilde{c} > \rho, \\ \tilde{\rho}_+ + c, & \text{otherwise,} \end{cases} \quad (13)$$

where $\tilde{\rho}_+$ is given by (12), $c = \rho - \tilde{\rho}_+$ and $c \geq \tilde{c}$. After calculating the search direction the variables for the next iterate will be determined and the damped BFGS update for the matrix $W_k + \Sigma_k$ is taken from [6]. After that it will be checked if $(x_k, \nu_{L,k}, \nu_{U,k}) \notin \mathcal{N}_{\mu_k}(\gamma)$. If this condition is satisfied, then repeat the inner loop by going back to Eq. (6). $\mathcal{N}_{\mu_k}(\gamma)$ is defined as a neighborhood of a point of the quasicontral path [5] corresponding to μ by

$$\begin{aligned}\mathcal{N}_\mu(\gamma) &= \left\{ (x, \nu_L, \nu_U) \in \mathbb{R}^{3n} : x > 0, \nu_L > 0, \nu_U > 0, \right. \\ &\quad \left. \|w_L - \mu w_L^{-1}\|^2 + \|w_U - \mu w_U^{-1}\|^2 \leq \gamma \mu \right\}\end{aligned}\quad (14)$$

with $w_L = \{[X - X_L]V^L e\}^{\frac{1}{2}}$, $w_U = \{[X_U - X]V^U e\}^{\frac{1}{2}}$ and $(\mu, \gamma) > 0$. Otherwise μ is updated on the following way:

$$\mu_{k+1} = \sigma \left(\|w_{L,k} - \mu_k w_{L,k}^{-1}\|^2 + \|w_{U,k} - \mu_k w_{U,k}^{-1}\|^2 \right), \quad (15)$$

where $\sigma \in (0, 1)$ is chosen by the user. This loop will be repeated until the following stopping criteria is satisfied:

$$\epsilon_{tol} \geq \max \left\{ \left\| \nabla_u f(z(u), u) - \nu^{LU} + \nu^{UU} - (\nu^{LZ})^T \frac{dz}{du} + (\nu^{UZ})^T \frac{dz}{du} \right\|_{\infty}, \right. \\ \left. \left\| \nabla_z f(z(u), u) - \nu^{LZ} + \nu^{UZ} \right\|_{\infty}, \left\| [X - X_L]V^L e - \mu e \right\|_{\infty}, \right. \\ \left. \left\| [X_U - X]V^U e - \mu e \right\|_{\infty} \right\} \quad (16)$$

with $\epsilon_{tol} > 0$ as a user provided error tolerance. Additionally, if there arise numerical problems, it is possible to modify the stopping criteria by scaling factors [2].

The proposed approach has been applied to the dynamic optimization of a continuous stirred-tank reactor (CSTR) and a heat-integrated distillation column system [1].

CONCLUSION

In this work an efficient interior-point method for large-scale optimization problems has been proposed. This approach has been implemented in the quasi-sequential framework in order to substitute the recent active-set strategy, which has a low performance for large-scale problems with many active inequality constraints. Further research could be done by comparing some of those methods or modify them, e.g. introducing slack variables, to reach an additional decrease in computational costs. Another area of research could be the application of the quasi-sequential approach to instable processes.

References:

- [1] Hong, W.R., Wang, S., P. Li, Wozny, G., Biegler, L.T. (2006): A Quasi-Sequential Approach to Large-Scale Dynamic Optimization Problems. *AIChE Journal*, vol. 52, 255-268.
- [2] Wächter, A., Biegler, L.T. (2006): On the Implementation of an Interior-Point Filter Line-Search Algorithm for Large-Scale Nonlinear Programming. *Mathematical Programming*, vol. 106, 25-57.
- [3] Biegler, L.T., Cervantes, A.M., Wächter, A. (2002): Advances in Simultaneous Strategies for Dynamic Optimization. *Chem Eng Sci*, vol. 57, 575-593.
- [4] Bock, H.G., Diehl, M., Leineweber, D. and Schlöder, J.P. (2000): A Direct Multiple Shooting Method for Real-time Optimization of Nonlinear DAE Processes. In Allgöwer, F. and Zheng, A., editors: *Nonlinear Predictive Control*, vol.26 of *Progress in System Theory*, 246-267, Basel, Birkhäuser.
- [5] Argaez, M., Tapia, R.A. (2002): On the Global Convergence of a Modified Augmented Lagrangian Linesearch Interior-Point Newton Method for Nonlinear Programming. *Journal of Optimization Theory and Application*, vol.114, no.1, 1-25.
- [6] Nocedal, J., Wright, S.J. (1999): *Numerical Optimization*. New York, Springer.

Authors:

Dipl.-Ing. Heidi Gerbracht
 Prof. Dr.-Ing. Pu Li
 Prof. Dr. Weirong Hong
 TU Ilmenau, Postfach 10 05 65
 98684 Ilmenau
 Phone: +49 3677 69-1427
 Fax: +49 3677 69-1415
 E-mail: heidi.gerbracht@tu-ilmenau.de

Pham Thieu Nga / B. Wutke

Modifying Bellman's dynamic programming to the solution of discrete multi-criteria optimization problems under fuzziness in long – term planning

ABSTRACT

For optimal planning of the long-term development of large scale, complex systems we need suitable methods regarding the dynamic, multi-criteria and uncertain characters of optimization problems. In this paper we propose a method for solving multi-objective discrete optimization problems under fuzziness in long-term planning based on the Discrete Dynamic Programming (DDP) of Bellman. This modified method has been called **Fuzzy Discrete Dynamic Pareto-Programming (FDDPP)**.

INTRODUCTION

Long-term planning problems always have dynamic character. They require the optimization of a developing system over a finite period time horizon. Nevertheless, with respect to large scale, complex systems they usually have multiple criteria and their data are often uncertain. The planning processes can be represented by multistage models (graph models) and optimized by the method of dynamic programming of Bellman. But the multi-objective optimization problem under fuzziness requires modifications of this method.

In this paper, we present briefly the method of dynamic programming of Bellman first. Then, we discuss the modifications to apply the method of Bellman for multi-criteria optimization problems under fuzziness.

THE DYNAMIC PROGRAMMING OF BELLMAN

(1) Basic model of the multistage decision process

The whole process can be represented in a multistage model (*Fig. 1*). The order of the stages is determined at the beginning. The outgoing state of a stage t is the entry state of the next stage $t+1$. Thereby the system state can be controlled through the transformation of a state into another. Within a stage the system state is considered unchanged. The optimization problem of the multistage decision process can be described as follows:

t - Stage, $t=1(1)T$

- \mathbf{x}^t - State vector of the system in stage t , $\mathbf{x}^t \in X^t$
- \mathbf{u}^t - Decision/Control vector in stage t , $\mathbf{u}^t \in U^t$
- q^t - The cost for stage t , $q^t = q^t(\mathbf{x}^t, \mathbf{u}^t)$
- $\mathbf{x}^t = \mathbf{x}^t(\mathbf{x}^{t-1}, \mathbf{u}^t)$ - The transformation between stages.

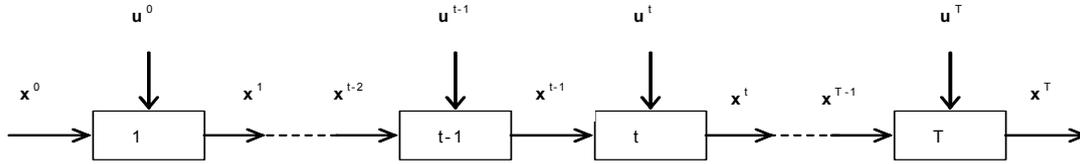


Figure 1. Multistage model

The problem is now to find a sequence of decisions/controls $(\mathbf{u}^{1*}, \mathbf{u}^{2*}, \dots, \mathbf{u}^{T*})$ that optimizes the criterion/return function $q = \sum_{t=1}^T q^t$ subject to the constraints $\mathbf{x}^t \in X$ and $\mathbf{u}^t \in U^t$. The sequence of controls is called the *optimal policy*.

(2) Bellman's principle of optimality

The Bellman's principle of optimality was formulated as follows [1]: "An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision". Using Bellman's principle of optimality, a general solution schema can be indicated through recursive substitution. Dependently on the direction in which the system will run during the recursion the forward solution (from the first to the last stage) and backward solution (from the last to the first stage) are to be distinguished. The dynamic programming in this paper bases on the backward solution principle. The conditions of the dynamic optimization procedure are as follows: Sequence of stages, separable criteria and restrictions, monotony of the return function.

(3) Application of the Discrete Dynamic Programming (DDP) to the planning task

The DDP method of Bellman is special appropriate for discrete dynamic optimization problems if the space and the time are modeled in graph models. The planning processes for large scale and continuous developing systems are represented by discrete graph models with a finite number of the discrete time stages, nodes and edges. Each node is a variant. In every stage several variants (nodes) can exist. The edges connect nodes in each stage. With this graph model, where the states are nodes and the controls are edges, the DDP method of Bellman can be applied effectively: Starting with the first node, the principle of the recursive optimization is carried out until the entry node is reached in the first stage.

The special feature of this planning task is that the optimal decision function, the stage transformation and the stage cost are given as discrete values. As far as that is concerned, the tabular calculation for the determination of the optimal functions is applied [6].

MODIFYING BELLMAN'S DYNAMIC PROGRAMMING TO THE SOLUTION OF THE MULTI-CRITERIA OPTIMIZATION PROBLEM

The DDP method of Bellman is usually used for optimizing the problems of single-criterion; however the main characteristics of the real-world planning problems have multiple criteria. For solving the multistage multi-objective discrete optimization problem the method of DDP must be expanded. The Pareto-principle can be applied in this case. If each edge of the graph model is a vector of goal quantities, the method is carried out also in stages at which all goal quantities of the edges outgoing from a node are compared with each other according to the half order principle during the selection of an optimal alternative. Thus, all pareto-optimal trajectories of the compromise set are found in the solution field. The modifying method is called as **Discrete Dynamic Pareto-Programming (DDPP)**. The following example demonstrates the procedure at the DDPP. Given is a graph model including 3 stages and 3 variants (*Fig. 2*). The value of every edge corresponds to the state of the node to which this edge directs. Each node is a state vector including 2 components (Q1, Q2), which correspond to 2 different objective functions.

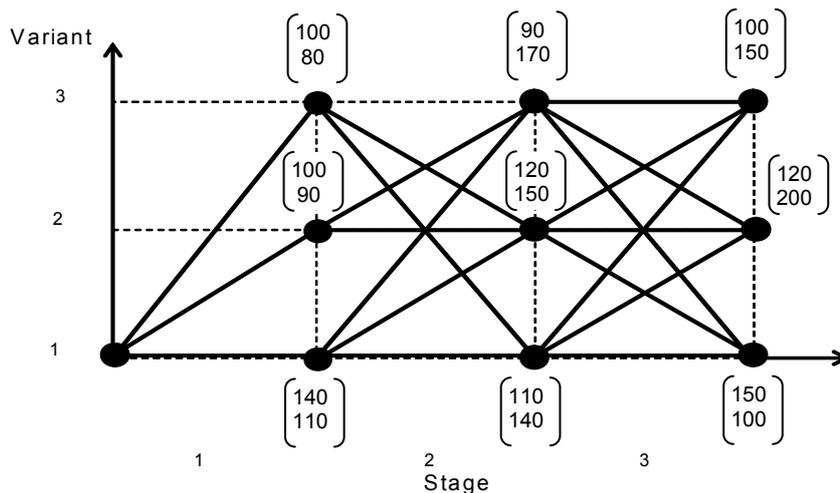


Figure 2. Example for Discrete Dynamic Pareto-Programming

Step 1: Starting at the last stage, the following edges are chosen:

- 1-1, 2-1 and 3-1 with $Q_1=150$ and $Q_2=100$
- 1-3, 2-3 and 3-3 with $Q_1=100$ and $Q_2=150$.

Step 2: In the next stage the following edges are chosen:

- 1-1-1 and 3-1-1 with $Q_1 = 110 + 150 = 260$ and $Q_2 = 140 + 100 = 240$
- 1-3-3 and 2-3-3 with $Q_1 = 90 + 100 = 190$ and $Q_2 = 170 + 150 = 320$
- 2-2-1 with $Q_1 = 120 + 150 = 270$ and $Q_2 = 150 + 100 = 250$
- 3-1-3 with $Q_1 = 110 + 100 = 210$ and $Q_2 = 140 + 150 = 290$

Step 3: In the first stage the following optimal trajectories are determined:

- 1-2-3-3 with $Q_1 = 100 + 190 = 290$ and $Q_2 = 90 + 320 = 410$
- 1-3-1-1 with $Q_1 = 100 + 260 = 360$ and $Q_2 = 80 + 240 = 320$

Unlike the DDP, as result the DDPP in general supplies **several** optimal trajectories.

MODIFYING BELLMAN'S DYNAMIC PROGRAMMING TO THE SOLUTION OF OPTIMIZATION PROBLEMS IN A FUZZY ENVIRONMENT

In the planning, the data will play an important role. Nevertheless, the data in the real world are often uncertain. In most cases, necessary data must be estimated in part or be approximated, or it is required to forecast future values. In order to give an optimal solution for the practical problems, this uncertainty of the data must be considered. Suppose the values of the states are not merely exact real numbers but fuzzy numbers, the use of the conventional Bellman's principle for the DDPP with fuzzy model requires a modification of the procedure.

(1) The stage transformation

For the stage transformation the extension principle of Zadeh [2] was applied. The extension principle may be used for the arithmetic operations, making it possible to handle fuzzy numbers. Let $A_i \in X_i$ be some fuzzy sets and $f : X_1 \times \dots \times X_n \rightarrow Y$ be some (nonfuzzy) function, $y = f(x_1, \dots, x_n)$. Then, due to the extension principle, the fuzzy set $B \in Y$ induced by the fuzzy set A_1, \dots, A_n through f is

$$\mu_B(y) = \begin{cases} \text{Sup}_{y=f(x_1, \dots, x_n)} \min(\mu_{A_1}(x_1), \dots, \mu_{A_n}(x_n)), & \text{if } \exists y = f(x_1, \dots, x_n) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

(2) Selection the better edge

For the definition of the "better" edge in the graph model the ranking procedure of Chen [3] was used.

The common concept of almost all ranking procedures consists in calculation a membership value for every alternative. It is valid for 2 alternatives a_i and $a_j \in A$:

$$a_i \succ a_j \Leftrightarrow \mu_C(i) > \mu_C(j) \quad (2)$$

The method of Chen is to find the total utility or ordering value of each fuzzy number A_i using the concept of maximizing set and minimizing set:

$$\mu_C^T(i) = \frac{1}{2} [\mu_C^M(i) + (1 - \mu_C^G(i))] \quad (3a)$$

where $\mu_C^M(i) = \text{Sup}_{u \in U} \min(\mu_i(u), \mu_M(u))$ $\mu_M(u) = \left[\frac{u - \text{Inf } U}{\text{Sup } U - \text{Inf } U} \right]^k$ (3b)

and $\mu_C^G(i) = \text{Sup}_{u \in U} \min(\mu_i(u), \mu_G(u))$ $\mu_G(u) = \left[\frac{\text{Sup } U - u}{\text{Sup } U - \text{Inf } U} \right]^k$ (3c)

The value of k can be varied to suit the application. For example: The case $k=2$ is risk-prone, the case $k=1/2$ is risk-averse.

Fuzzy numbers with triangular, trapezoidal shaped membership functions are often used in the practice. Suppose the values of states are fuzzy numbers A_i with triangular membership functions, each having three vertices, with coordinates $(c_i, 0)$, (a_i, h_i) , $(d_i, 0)$ as in *Figure 3*. In details, if the fuzzy number A_i has the membership function:

$$\mu_{A_i}(u) = \begin{cases} \frac{h_i(u - c_i)}{(a_i - c_i)}, & c_i \leq u \leq a_i, \\ \frac{h_i(u - d_i)}{(a_i - d_i)}, & a_i \leq u \leq d_i, \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

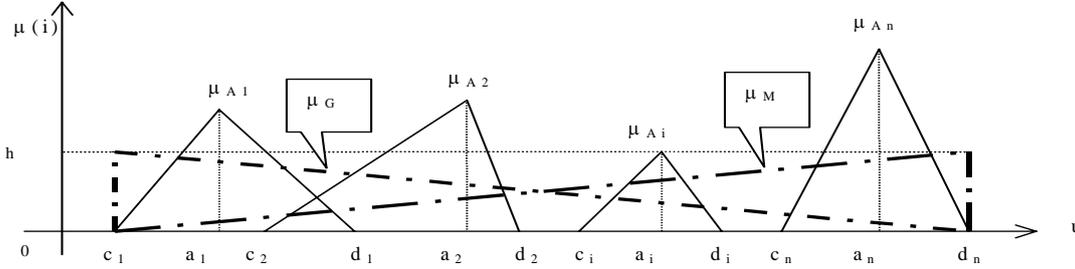


Figure 3. Ranking fuzzy numbers with triangular membership functions

When $k=1, i=1,2,\dots,n$, $h = \text{Inf } h_i$, $u_{\min} = \text{Inf } S$, $u_{\max} = \text{Sup } S$, $S = \bigcup_{i=1}^n S_i$, $S_i = \{u \mid \mu_{A_i}(u) > 0\}$

we obtain the total utility of each fuzzy number A_i as follows:

$$\mu_C^T(i) = \frac{h \cdot h_i}{2} \left[\frac{d_i - u_{\min}}{h_i(u_{\max} - u_{\min}) - h(a_i - d_i)} + \frac{1}{h_i} - \frac{u_{\max} - c_i}{h_i(u_{\max} - u_{\min}) + h(a_i - c_i)} \right] \quad (5)$$

If the values of states are fuzzy numbers A_i with trapezoidal membership functions, each having once vertex (b_i, h_i) more, $b_i > a_i$. When $k=1$, we obtain the total utility of each fuzzy number A_i as follows:

$$\mu_C^T(i) = \frac{h \cdot h_i}{2} \left[\frac{d_i - u_{\min}}{h_i(u_{\max} - u_{\min}) - h(b_i - d_i)} + \frac{1}{h_i} - \frac{u_{\max} - c_i}{h_i(u_{\max} - u_{\min}) + h(a_i - c_i)} \right] \quad (6)$$

$\mu_C^T(i)$ is a measure to compare the fuzzy numbers, which describe the values of states in graph model.

So the dynamic discrete optimization problem under fuzziness is solved mathematical cleanly. With this modification the DDPP has been expanded once more and is called as **Fuzzy Discrete Dynamic Pareto-Programming (FDDPP)**.

APPLICATION OF FDDPP - METHOD

The above-mentioned FDDPP-method was applied to the decision support system DSPES (Decision Support System for Planning of Energy Supply) [4]. DSPES is a system for long-term energy supply planning in the commune/region. The planning problem is modeled as multi-criteria, multistage evolution problems in decomposed systems with global resource constraints. The hierarchical decision-making system consists of 2 layers. The subordinate layer consisting of several levels is responsible for the optimal resource distribution. It is supposed, the large scale system here a city or a region is decomposed into (indirect) coupled subsystems through the restriction of the global resources, available for the overall system in each (time) stage. Since the system

is considered unchanged within a stage, the problem of the resource distribution is a static optimization problem under fuzziness due to the approximation and the forecast of data. The solution method of this problem [5] is not a subject of this paper. The results of the static optimization problem are fuzzy numbers with triangular membership functions. The above layer (the strategic layer) determines the optimal developing trajectory of the chronological evolution of the energy support for the overall system about the long-term planning horizon consisting of several stages. Each subsystem is represented by a discrete graph model consisting of nodes and edges at which the state of every node was supplied from the subordinate layer, (in form of a triangular fuzzy number). From these local graph models of the subsystems, a global graph model for the overall system is formed. Using the FDDPP-method the optimal trajectory in the global graph model is determined. Through a mapping of the optimal trajectory in the global graph model the optimal trajectories of the local graph models are determined.

DSPES was tested with 2 projects successfully [4]. The first project was for a city with 16 thousand of people, 60 technologies to supply energy for 9 different demands. The second project was for a city with 186 thousand of people, 301 technologies to supply energy for 22 different demands. Both these projects have time horizon of 20 years consisting of 4 stages, 4 global resource constraints and 3 criteria (the consumption of primary energy, total costs and emission of pollutants) to optimize.

CONCLUSION

In the long-term planning of real large scale and continuous developing systems the consideration regarding different criteria is very necessary. Simultaneously, with these complex systems the appearance of uncertain data is not avoidable. This paper proposed the method of Fuzzy Discrete Dynamic Pareto-Programming (FDDPP) by modifying the Discrete Dynamic Programming (DDP) of Bellman for solving the multi-criteria discrete optimization problem under fuzziness. The proposed method has several improvements and was applied to a decision support system successfully.

References:

- [1] Bellman, R.: "Dynamic Programming", Princeton, New Jersey, , 1957
- [2] Bothe H.-H.: "Fuzzy logic", Springer Verlag, Berlin 1995
- [3] Chen, S. H.: "Ranking fuzzy numbers with maximizing set and minimizing set", Fuzzy Set and System 17, 113-129, 1985.
- [4] Pham Thieu Nga: "Ein Entscheidungshilfesystem für komplexe Planungsprozesse der kommunalen-regionalen Energieversorgung", ISLE Verlag, Ilmenau, 2000.
- [5] Pham Thieu Nga: "Integration of uncertainties in a decision support system for the regional/communal energy supply planning", 49th IWK, Conference Proceedings, Volume 2, 67-72, Shaker Verlag, 2004.
- [6] Wutke, B.: "Beiträge zur Systemanalyse, optimierung und Koordination bei der hierarchischen Steuerung-Entscheidungsfindung in komplexen Systemen", Dissertation B, Kapitel 5, TH Ilmenau, 1982.

Authors:

Dr.-Ing. Pham Thieu Nga
 University of Civil Engineering
 Hanoi, Vietnam
 Phone: +84 - 4- 8212677
 Email:ptnga@hn.vnn.vn

Dr.-Ing. habil. B. Wutke
 Technical University of Ilmenau
 Faculty of Computer Science and Automation
 Institute for Automation and Systems Engineering
 98693 Ilmenau, Germany

S. Ritter / P. Bretschneider

Optimale Planung und Betriebsführung der Energieversorgung im liberalisierten Energiemarkt

Einleitung

Mit der Liberalisierung der Energiemärkte traten in den vergangenen Jahren Gesetze in Kraft, die Veränderungen bezüglich der Strukturen in den Bereichen des Energieeinkaufs, der Energiebereitstellung und des Energieverkaufs der Energieversorgungsunternehmen verlangten. Durch EnWG, EEG und KWK-Gesetz sind nicht nur Unbundling und andere regulatorische Maßnahmen an die Tagesordnung getreten, sondern es sind auch neue Bezugs- und Absatzmöglichkeiten geschaffen worden, die durch Art und Struktur die Freiheitsgrade einer optimalen Planung und Betriebsführung stark erhöhen und damit das manuelle Finden des kostenminimalen Optimums unüberschaubar und damit weitestgehend unmöglich machen.

Anhand eines Praxisbeispiels soll in diesem Beitrag gezeigt werden, dass mittels mathematischer Werkzeuge eine Abbildung der sehr komplexen Optimierungsprobleme erfolgen kann und welche Optimierungspotentiale sich im liberalisierten Energiemarkt für den Energieversorgungsprozess ergeben können. Im Weiteren werden die Ergebnisse verschiedener Szenarien verglichen und bewertet.

Der Energieversorgungsprozess im liberalisierten Energiemarkt

Im Bereich der Energiebeschaffung existieren wie auch schon vor der Liberalisierung Bezüge auf Basis vertraglicher Vereinbarungen mit dem Energie-Vorlieferanten. Diese sogenannten Bezugsverträge sind durch verschiedenste Parameter definiert und wurden in ihrer Struktur seit der Liberalisierung novelliert.

Eine zusätzliche Möglichkeit des Energiebezuges und auch der Energielieferung bieten Energiespotmärkte und -börsen. Hier kann Energie zu Marktpreisen gehandelt werden. Um beispielsweise die Energiebörse in die zukünftige Planung einzubeziehen gibt es die Möglichkeit, anhand von Preisforwards die geplanten Energiemengen zu bewerten. Somit können Börsenprodukte im Zuge der Optimierung vorgeschlagen werden, die bei

realisiertem Einkauf bzw. Verkauf das Portfolio im Sinne der Optimalität unterstützen. Des Weiteren kann Energie auch außerbörslich gehandelt werden. Dieser OTC-Handel zwischen Marktteilnehmern gilt für den Energieeinkauf wie auch –verkauf. Da die an den Börsen gehandelten standardisierten Produkte häufig nicht dem Wunsch der handelnden Partner entsprechen, ist der OTC-Handel zu einem wichtigen Instrument für Energiebeschaffung und den Energieabsatz geworden.

Auf Basis des EEG gibt es im Bereich des Energiebezuges eine Energieabnahmeverpflichtung. Dies ist eine Zwangsabnahme elektrischer Energie, die auf Grundlage von Berechnungen der Übertragungsnetzbetreiber in einer festgelegten Menge und zu einem festgelegten Preis abzunehmen ist. Die sogenannte EEG-Quote ist die aufgenommene Strommenge aus durch das EEG geförderten Anlagen bezogen auf die gesamte in Deutschland an Endverbraucher abgegebene Strommenge [1]. Im Jahr 2007 beträgt die EEG-Quote schätzungsweise 13,39 % [2].

Für die sichere und wirtschaftliche Deckung der Netzlast ist die Bereitstellung von Energie durch Eigenerzeugung unerlässlich. Hierzu gibt es verschiedenste Möglichkeiten. In dem zu behandelnden Beispiel geschieht dies mittels Kraft-Wärme-Kopplung, wodurch ein wesentlich höherer Wirkungsgrad bezogen auf die resultierende Nutzenergie erreicht wird als bei reinen Stromerzeugungsprozessen. Die anfallende Wärme wird zur Deckung der Fernwärmelast verwendet. In diesem Zusammenhang gibt es laut dem am 1. April 2002 in Kraft getretenen KWK-Gesetz eine Förderung auf aus KWK-Prozessen eingespeisten Strom. Es erfolgt eine Vergütung pro kWh elektrischer Energie, die von Größe und Modernisierungsgrad der Anlagen abhängig ist. Damit soll eine verstärkte Nutzung der KWK-Anlagen erreicht werden, um eine weitere Minderung der Kohlendioxid-Emissionen zu forcieren.

Um Versorgungsengpässe auszugleichen oder Lastspitzen nicht zu ungünstigen Zeiten in die Bereiche des Vorlieferanten wirken zu lassen, werden Speicher im Gas- und Wärmebereich eingesetzt.

Optimierung des Versorgungsprozesses

Durch die Liberalisierung der Energiemärkte sind die Energieversorgungsunternehmen zunehmend unter Druck, ihre Energie immer effizienter anzubieten. Ein enormes Optimierungspotential bietet die Querverbundoptimierung. Hierbei werden Verknüpfungen der einzelnen Sparten Strom, Gas und Fernwärme dazu benutzt, um eine ganzheitliche Betriebsplanung durchzuführen. In Folge dessen werden alle

Erzeugeranlagen und Bezugsverträge spartenübergreifend dazu verwendet, das wirtschaftlichste Ergebnis zu erhalten. Die Ergänzung des Querverbunds um Spotmarkthandel/Börse und Speichernutzung bietet im Moment das größte ökonomischste Optimierungspotential. Zusätzlich kommen durch Kraft-Wärme-Kopplung (KWK) und Speichernutzung ökologisch sinnvolle Konzepte zum Einsatz.

Die Aufgabe besteht nun darin, alle zur Verfügung stehenden Ressourcen so einzusetzen, dass unter Einhaltung aller technischen, ökonomischen und ökologischen Randbedingungen die Energieversorgung garantiert ist und dabei die geringsten Kosten entstehen.

Modellierung des Optimierungsproblems

Die enorme Komplexität des zu betrachtenden Gesamtsystems und die hohe Dynamik der beteiligten Prozesse erfordert die rechnergestützte Lösung einer Optimierungsaufgabe mit den Kosten als Zielfunktion. Bei dem im Folgenden vorgestellten Beispiel erfolgte die Modellierung mittels der Gemischt-Ganzzahligen-Linearen-Programmierung (GGLP). Für diese Methode existieren standardisierte Lösungsverfahren und es besteht eine hohe Flexibilität hinsichtlich der Modellzusammenstellung und damit des zu optimierenden Versorgungsprozesses [3].

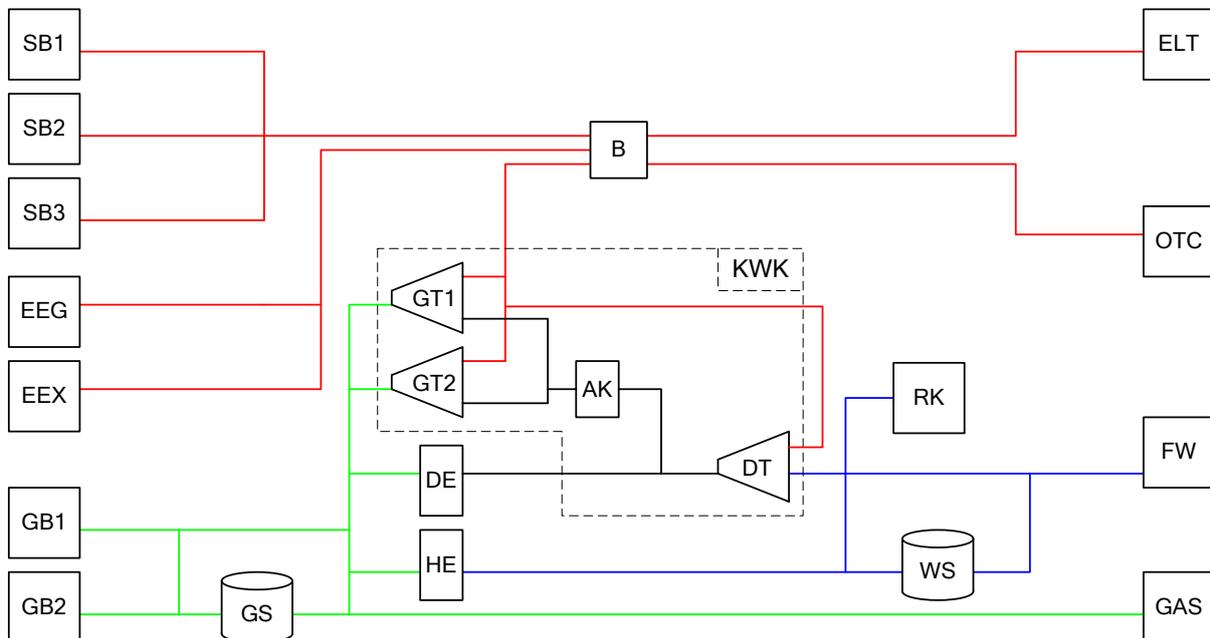


Abbildung 1: Struktur des Querverbund-Optimierungsmodells

Mittels GGLP gelingt eine gute Abbildung und Approximation der Prozesse und jegliche Randbedingungen können berücksichtigt werden [4].

Die Modellierung des Beispiels wurde in Gams umgesetzt und als Solver wurde CPLEX

verwendet. Die einzelnen Teile des Optimierungsmodells, wie z. B. Verträge oder Kraftwerksanlagen, wurden als separate Komponenten modelliert. Die Aus- und Eingänge der Komponenten wurden über Gleichungen verbunden. Somit ist es möglich, gezielt einzelne Teile des Optimierungsproblems bei Bedarf zu deaktivieren oder durch andere zu ersetzen. In Abbildung 1 ist die Struktur des zu betrachtenden Optimierungsmodells dargestellt. Wobei die Abkürzungen der Komponenten folgende Bedeutungen haben:

SB1, SB2, SB3	Strombezugsverträge
GB1, GB2	Gasbezugsverträge, GB1 ist ein offener Liefervertrag
EEG	EEG-Zwangsabnahme
EEX	EEX-Strombörse
GS	Gasspeicher
WS	Wärmespeicher
GT1, GT2	Gasturbinen
AK	Abhitzeessel
DE	Dampferzeuger
HE	Heißwassererzeuger
DT	Dampfturbine
B	Bilanz, hier kann Netznutzungsproblematik abgebildet werden
ELT	Bedarfsprognose Strom
OTC	Außerbörslicher Handel
FW	Bedarfsprognose Fernwärme
GAS	Bedarfsprognose Gas
RK	Rückkühler für Überschusswärme

Auswertung verschiedener Kostenszenarien

Zur Verdeutlichung der Kosteneinsparung, welche durch eine Querverbundoptimierung inkl. EEX-Börse und Speicherverwendung erzielt werden kann, wurden anhand des Beispiels 4 Kostenszenarien untersucht:

1. reiner Querverbund
2. Querverbund mit EEX-Börse
3. Querverbund mit EEX-Börse und Gasspeicher
4. Querverbund mit EEX-Börse, Gas- und Wärmespeicher

Die Ergebnisse der 4 Szenarien wurden für einen Optimierungshorizont von einer Woche ermittelt und sind in Tabelle 1 dargestellt. Anhand der ermittelten Kosten kann gezeigt werden, dass die Erweiterung des Querverbunds um EEX-Börsenbezug und Speichernutzung ein immenses Einsparpotential besitzt. Die größte Kosteneinsparung konnte im untersuchten Beispiel durch die Speichernutzung erreicht werden.

Tabelle 1: Kostenszenarien, KS – Kostenstelle, E – Energie in MWh, K – Kosten in €

KS	Szenario S1		Szenario S2		Szenario S3		Szenario S4	
	E	K	E	K	E	K	E	K
SB1	1282	39771	1508	46765	1500	46514	1502	46579
SB2	221	7120	525	16662	523	16598	526	16667
SB3	2114	69774	840	27722	870	28742	876	28931
GB1	3835	46024	4076	48917	312	3755	249	2988
GB2	41192	123651	41550	124723	45282	136435	45045	135238
EEX		0	304	12298	299	11842	297	11635
HE (Steuer)		11289		11286		11290		9903
RK	202		295		285		7	
Gesamt		297629		288373		255176		251941
ΔK zu S1		-		9256		42453		45688

Hier kann gezeigt werden dass z. B. durch Peak-Shaving Gasbezug aus teuren Bezugsverträgen vermieden werden kann. Des Weiteren sinkt die Verlustwärme des Rückkühlers, die ungenutzt an die Umgebungsluft abgegeben wird, auf ein Minimum. Bei einem Marktwert von durchschnittlich 20 € pro MWh KWK-Fernwärme ist das Einsparpotential enorm. Durch Strombezug von der Börse gelingt es ferner, zu günstigen Zeiten Bezug aus den Verträgen zu vermeiden und Kosten zu sparen. Der Gesamtenergiebezug konnte bei absoluter Lastdeckung um ca. 150 MWh verringert werden. Dies ist ein wichtiger Beitrag zum ökologisch sinnvollen Umgang mit Energie und trägt unmittelbar zur Verringerung der Treibhausgasemission bei. In Abbildung 2 sind der Energiebezug und die Kosten der untersuchten Szenarien grafisch dargestellt.

Zusammenfassung und Ausblick

Zusammenfassend kann festgestellt werden, dass sich durch die ganzheitliche Betrachtung des Versorgungsprozesses während der Optimierung immense Kosten- und Energieeinsparungen erreichen lassen.

In aktuellen Forschungsarbeiten wird untersucht, welche zusätzlichen Vorteile sich im praktischen Anwendungsfall durch die Berücksichtigung von Unsicherheiten während der Optimierung ergeben. Es wird zum Beispiel analysiert, welchen Einfluss die Unsicherheit der Preise/Preisforwards und anderer exogenen Größen auf das Optimierungsergebnis (Fahrpläne der einzelnen Komponenten) hat und welche Unterschiede sich dabei im Gegensatz zum determinierten Modell ergeben.

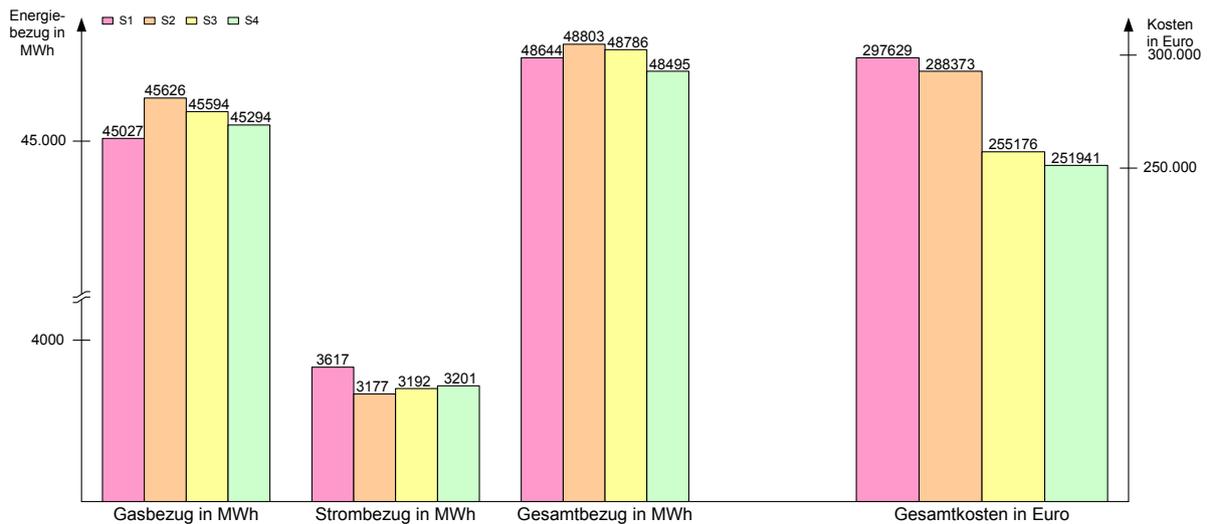


Abbildung 2: Energiebezug und Kosten der Szenarios 1 bis 4

Des Weiteren wird der praxistaugliche Einsatz der stochastischen Optimierung bezüglich der Performance untersucht. Anhand von Anforderungen von Industriepartnern wird festgestellt, welche Komplexität die stochastischen Modelle und welchen Umfang die zu berücksichtigenden Szenariobäume annehmen dürfen, damit ein gutes Aufwand-Nutzenverhältnis erreicht werden kann. Ziel ist es, anhand dieser Ergebnisse zu prüfen, ob ein praxistauglicher Workflow erarbeitet werden kann, um die stochastische Optimierung für die Energiewirtschaft nutzbar zu machen.

References:

- [1] Bundesverband Windenergie e.V., "<http://www.wind-energie.de/de/themen/windenergie-von-a-z/eeg-quote/>"
- [2] VDN e.V., "http://www.vdn-berlin.de/eeg_prognose_2007.asp"
- [3] Martin Hanselmann: Entwicklung eines Programmsystems zur Optimierung der Fahrweise von Kraft-Wärme-Kopplungsanlagen, Universität Stuttgart (1996)
- [4] Jonas Albiger: Integrierte Ressourcenplanung in der Energiewirtschaft mit Ansätzen aus der Kraftwerkseinsatzplanung, Universität Stuttgart (1997)

Authors:

Dipl.-Ing. Sebastian Ritter
 Dr.-Ing. Peter Bretschneider
 Fraunhofer Anwendungszentrum Systemtechnik, Am Vogelherd 50
 98693, Ilmenau
 Phone: 00493677 4610
 Fax: 00493677 461100
 E-mail: Sebastian.Ritter@ast.iitb.fraunhofer.de
 E-mail: Peter.Bretschneider@ast.iitb.fraunhofer.de

P. Bretschneider / D. Westermann

Intelligente Energiesysteme: Chancen und Potentiale von IuK-Technologien

Einleitung

Die europäische Energiewirtschaft ist in den letzten Jahren von einem tiefgreifenden Wandel geprägt. Treibende Kräfte sind sowohl die Liberalisierung der Energiemärkte als auch die Gestaltung einer nachhaltigen und umweltgerechten Energieversorgung. Zu nennen sind bspw. der Ausbau der dezentralen Energieerzeugung, die Entwicklung von neuen Technologien und Anlagen zur effizienten Bereitstellung und Nutzung von Energie sowie moderne Managementsysteme für die optimale Planung und Betriebsführung.

Weitere, noch ungenutzte Potentiale werden mit dem flächendeckenden Einsatz von IuK-Technologien für die Bereiche Erzeugung, Verteilung und Vertrieb erwartet. Hierzu werden bereits unterschiedliche Forschungs- und Entwicklungsthemen bearbeitet [4].

Die Technologieplattform „**SmartGrids**“ untersucht bspw. die Fragen der europäischen Netzintegration, Netzbetriebsführung, Speicher, MicroGrids, dezentralen Erzeugung, Demand-Side-Management und Kundenintegration. Dabei wird auch ein Umbau der Energieversorgungssysteme erwartet, welche sich künftig neben den bisherigen energetischen Strömen durch eine übergeordnete informationstechnische Vernetzung mit einer engen Verzahnung von technischen und geschäftlichen Prozessen auszeichnen [4].

Das BMWi prüft mit der Studie „**eEnergy**“ die Potenziale und Auswirkungen der IuK-Technologien für die Optimierung der Energieversorgung und Versorgungsprozesse sowie des Energieverbrauchs. Auch hier wird von einem erheblich veränderten Energieversorgungssystem ausgegangen, das sich neben den energetischen Strömen zukünftig auch durch umfangreiche Informationsflüsse auszeichnet [4].

Im Bereich „**Smart Metering**“ liegen ebenfalls sehr interessante Projekte vor. Der ver-

stärkte Einsatz von IuK-Technologien in diesem Sektor ergibt sich aus den Entwicklungen im Zähler- und Messwesen. Mittlerweile sind preiswerte elektronische Lastgangzähler sowohl für Sondervertrags- als auch Tarifikunden verfügbar. Sie ermöglichen erstmalig eine transparente Messung und Abrechnung von Energie über alle Kundensegmente hinweg [4].

Jüngstes Projekt in diesem Zusammenhang ist der BMWi-Förderschwerpunkt „E-Energy“ zum Aufbau eines IKT-basiertes Energiesystems der Zukunft. Die Ziele sind, die technischen und geschäftlichen Prozesse der Energieversorgung unter Beachtung der liberalisierten Marktanforderungen effizient zu verknüpfen, die Einbindung der dezentralen Energieerzeugung in die elektrischen Übertragungs- und Verteilnetze zu vereinfachen, die Durchlässigkeit der einzelnen energiewirtschaftlichen Prozesse zu erhöhen sowie passive Marktakteure zu aktiven Marktteilnehmern werden zu lassen.

Aspekte und Anforderungen des liberalisierten Energiemarktes

Die Liberalisierung des Energiemarktes basiert auf den EU-Richtlinien zum Elektrizitätsbinnenmarkt aus dem Jahr 1996 und zum Gasbinnenmarkt aus dem Jahr 1998. Die Umsetzung war mit tiefgreifenden strukturellen, organisatorischen, vertraglichen und prozessualen Veränderungen verbunden und erforderte entlang der Wertschöpfungskette die Entflechtung der Geschäftsbereiche Erzeugung, Übertragung, Transport, Verteilung und Vertrieb (vgl. Abbildung 1) [1], [5], [3]. Neue Marktregeln in Form von Gesetzen, Richtlinien und Vereinbarungen wurden zur Öffnung des Energiemarktes erlassen.

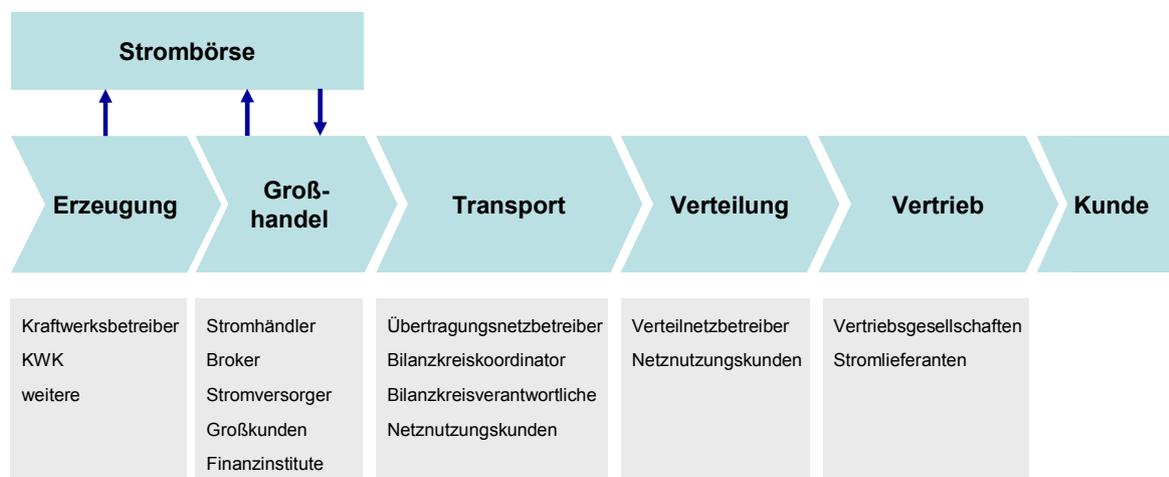


Abbildung 1: Wertschöpfungskette im liberalisierten Strommarkt

Sie führten einerseits zum Wegfall von „festen“ Versorgungsgebieten mit der freien Wahl des Stromlieferanten durch den Endkunden und andererseits zu neuen Marktplätzen und Marktteilnehmern wie z.B. Strombörse, Bilanzkreiskoordinator (BKO), Bilanzkreisverantwortliche (BKV), Lieferanten und Händler. Mit der organisatorischen Entflechtung von Netz und Vertrieb kommt es auch zur Trennung der bisher gemeinsam genutzten IuK-Technologien. Neue Geschäftsprozesse müssen definiert, in IT-Systemen umgesetzt und auf aktuelle und künftige Anforderungen ausgerichtet werden [1].

Eine wesentliche Rolle spielt dabei die Vernetzung der IT-Systeme wie z.B. Netzleitstelle, Zählerfernabfrage, Energiedatenmanagement, Energiemanagement und Energieabrechnung. Ziel ist es, möglichst durchgängige und systemübergreifende Verarbeitungsprozesse zu gestalten. Zu diesem Zweck müssen die betreffenden Softwaresysteme über offene, standardisierte Schnittstellen verfügen, um sowohl zeitgesteuert als auch transaktionsorientiert den Datenaustausch zwischen den Systemen zu automatisieren und zugleich konsistente unternehmensweite Datenbestände zu gewährleisten. Die Verarbeitungsprozesse sollen damit hochwertig, fehlerfrei und kostengünstig werden [2].

Intelligente Energieversorgungssysteme

Wie bereits im vorangegangenen Abschnitt geschildert, ergibt sich mit der Liberalisierung des Strommarktes und der zunehmenden Anzahl von dezentralen Einspeisungen eine Vielzahl von neuen Herausforderungen für die elektrischen Energieversorgungssysteme. Beispielsweise müssen die elektrischen Energieversorgungssysteme eine möglichst flexible Plattform für den freien Energiehandel für alle Marktteilnehmer darstellen, stark fluktuierende und nur schlecht prognostizierbare regenerative Einspeisungen aufnehmen und bei höchstmöglicher Belastung ein hohes Maß an Zuverlässigkeit aufweisen. Bei Ausfall von einzelnen Betriebsmitteln bis hin von Teilnetzen soll eine möglichst schnelle Wiederinbetriebnahme gewährleistet sein. Dies ist nur ein kleiner Anforderungsbereich, den intelligente elektrische Energieversorgungssysteme abdecken [6].

Durch den Einsatz moderner Informations- und Kommunikationstechnologien im Rahmen einer durchgängigen Applikationsarchitektur, verknüpft mit modernen schnell regelbaren Betriebsmittelkomponenten werden elektrische Energieversorgungssysteme „intelligent“. Diese Intelligenz umfasst alle Bereiche der elektrischen Energieversorgung.

Folgende Beispiele können hierfür aufgeführt werden:

- In den Privathaushalten halten mittlerweile die elektronischen Zähler mit Lastgangaufzeichnung Einzug. Über benutzerfreundliche Schnittstelle könnten aktuelle Verbrauchshistorie und verursachte Kosten visualisiert werden. Flexible Stromtarife können umgesetzt werden und nehmen Einfluss auf das Verbrauchsverhalten zur Minimierung der Energiekosten bei möglichst gleichzeitiger Maximierung der Netzeffizienz.
- Durch die informationstechnische Anbindung einzelner dezentraler Kleinsterzeuger können relevante Einspeisekapazitäten in Form eines virtuellen Kraftwerks realisiert werden. Die virtuellen Kraftwerke ermöglichen einen ökonomisch und ökologisch optimalen Betrieb parallel zu konventionellen Erzeugungseinheiten mit großen installierten Leistungen.
- Im elektrischen Übertragungs- und Verteilnetzen gestattet die informationstechnische Vernetzung eine koordinierte Betriebsführung von schnellen Netzreglern, um bestehende elektrische Anlagen optimal auszunutzen, Netzengpässe abzumildern und fluktuierende Einspeisungen effizient zu transportieren und zu verteilen.
- Die intelligente Vernetzung aller Betriebsmittel erlaubt im Falle von Störungen eine automatische Netzrekonfiguration, um einerseits eine geringere Beeinflussung des Gesamtsystems zu erreichen und andererseits das Risiko von Großstörungen zu minimieren [6].

Entwicklungs- und Testlabor für intelligente Energieversorgungssysteme

Im Entwicklungs- und Testlabor werden neue IuK-Technologien zur Führung, Überwachung und Monitoring von Energieversorgungssystemen sowie ihre Einbindung in liberalisierte Energiemärkte unter Labor- und Testfeldbedingungen entwickelt, erprobt und untersucht. Aufbauend auf den gewonnenen Erkenntnissen sollen neue Lösungen für die rationelle und nachhaltige Bereitstellung und Nutzung von Energie realisiert sowie den Technologietransfer in die Industrie unterstützt werden.

Das Entwicklungs- und Testlabor für Energieversorgungssysteme wird am Fraunhofer Anwendungszentrum Systemtechnik Ilmenau, Technische Universität Ilmenau und am

Fraunhofer Institut UMSICHT Oberhausen aufgebaut. Weitere Standorte sind ein Ökohaus in Japan, ein Geschäftshaus in Asien, Stadtwerke Erfurt Strom und Fernwärme GmbH und Solardorf Kettmannshausen (vgl. Abbildung 2).

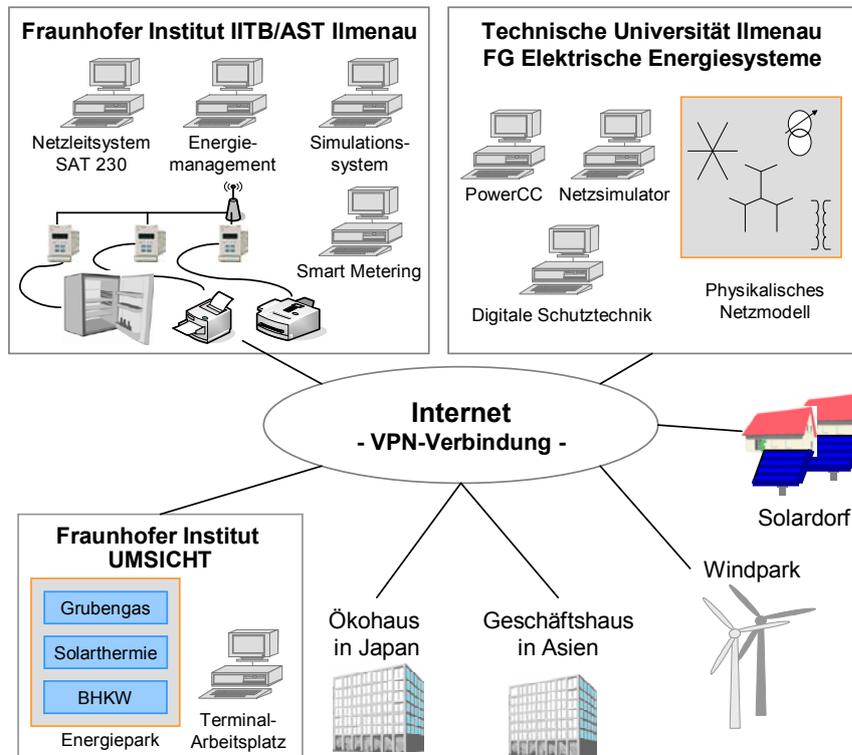


Abbildung 2: Entwicklungs- und Testlabors für zentrale und dezentrale Energiesysteme

Die Schwerpunkte des Entwicklungs- und Testlabors sind:

- Volldigitale Schutz- und Leittechnik;
- Untersuchung und Entwicklung von IuK-Technologien für die Umsetzung von Demand-Response und Demand-Side-Management;
- Koordinierte Betriebsführung dezentraler Einspeisungen;
- Methoden für ganzheitliches Energiemanagement (Strom, Gas, Wärme und Kälte);
- Echtzeitfähige Erfassung, Übertragung und Verwaltung von Massendaten;
- Untersuchung und Entwicklung von IuK-Technologien für den Einsatz in Privathaushalten wie z.B. Home-Portal-Interface zur Abrechnung und Visualisierung des Verbrauchs oder die energiebörsliche Anbindung;
- Testplattform für Leitsystemkomponenten industrieller Anbieter.

Zusammenfassung und Ausblick

Die in den vorangegangenen Abschnitten skizzierten Themen sind Gegenstand der aktuellen Forschungsarbeiten. Abschließende Ergebnisse liegen derzeit noch nicht vor. Die derzeitigen Arbeiten beinhalten ausschließlich den Aufbau und die Inbetriebnahme des Entwicklungs- und Testlabors. Die zukünftigen Themen konzentrieren sich auf das Demand-Side-Management, auf die koordinierte Betriebsführung dezentraler Einspeisungen sowie die echtzeitfähige Erfassung, Übertragung und Verwaltung von Massendaten.

Literatur

- [1] Bretschneider, P.: „Methoden und Lösungen für das Energiemanagement im liberalisierten Energiemarkt“, GMA-Fachausschuss, Bommernholz, 2006
- [2] Bunke, D.; Goetting, B.; Kück, H.: „Automatisierte Abwicklung des Lieferantenwechsels und der Fahrplananmeldung“, ew, Frankfurt, 6/2004
- [3] Brinker, Wemder: „VDEW-Kongress 2007“, VDEW-Kongress 2007, Berlin, 2007
- [4] Franz, O., Wissner, M., Büllingen, F., et.al.: Dr. Christin-Isabel Gries, D.: „Potenziale der Informations- und Kommunikationstechnologien zur Optimierung der Energieversorgung und des Energieverbrauchs (eEnergy)“, Studie, BMWi, 2006
- [5] Niessch, C.; Straßberger, F.: „Der europäische Strommarkt“, Studie, DG BANK, Frankfurt am Main, 11/2000
- [6] Westermann, D.: „Intelligente Elektrische Energieversorgungssysteme“, visIT, Karlsruhe, 01/2007

Autoren

Dr.-Ing. Peter Bretschneider
Fraunhofer Anwendungszentrum Systemtechnik Ilmenau, Vogelherd 50
98693 Ilmenau

Telefon: +49(0)3677 4610
Fax: +49(0)3677 46100
E-Mail: peter.bretschneider@ast.iitb.fraunhofer.de

Prof. Dr.-Ing. Dirk Westermann
Technische Universität Ilmenau
Fachgebiet Elektrische Energiesysteme
Telefon: +49(0)3677 692640
E-Mail: dirk.westermann@tu-ilmenau.de

Zhihui.Lu / Yiping.Zhong / Yu.Wu / Jie.Wu

WSReMS:A Novel WSDM-based System Resource Management Scheme

Abstract

Based on the OASIS-Web Services Distributed Management Standards- WSDM, we propose a novel system resource management scheme- WSReMS. Furthermore, we discuss MUSE-CIMOM based system management scheme implementation- WSReMS-Platform, including Manager Layer, WSDM-Gateway Layer, and Agent Layer, and then analyze the experiment results. At last, from our research experiences and related survey, we analyze the prospective research direction and challenges in this field.

Key words: Web Services, WSDM, System Management, Resource Management

1.Introduction and Background

IT system has been playing a more and more important role in the commercial world nowadays as well as in the daily life. Not only has the daily management of the company relied on the IT support, but also the process of critical business requirements have relied on the IT infrastructures and new IT technologies to assure the whole business continuity.

The total IT environment is becoming more and more complicated. Therefore, the difficulty of IT System management is also increasing. Because the system operators do not get all the information of the system operation, they do not know the bottleneck of current system and they do not know how many more resources should be added to fit for the new business requirements. They are often in the situation that they have put more resources but with little gains in productivity. And the IT environment is becoming more and more complex. All of this translates into additional costs and result in low ROI (return on investment). Enterprise Architecture believes that SOA(Service-Oriented Architecture) can help businesses respond more quickly and cost-effectively to the changing market conditions and the complexity of the IT environment. This style of architecture promotes reuse at the macro (service) level rather than micro levels (e.g. objects). It can also simplify interconnection to and

usage of existing IT (legacy) assets. Currently, Web Services is a series of feasible technology packages to implement SOA.

The OASIS Web Services Distributed Management TC is defining a set of Web Services Distributed Management (WSDM) specifications: The WSDM specifications define how to use Web services to expose manageable resources (MUWS, Management Using Web Services [1]), and in addition, define how to expose manageable Web service implementations (MOWS, Management Of Web Services[2]). In a word, the main idea of WSDM is to define a Web Service architecture to manage distributed resources. WSDM further exposes the Web Service Endpoint of the virtualized resource, as the manageability interfaces. And the manageability interfaces support common operations: (a) Control: start, stop, etc; (b) Monitoring: status & performance.

2.WSReMS: A Novel WSDM-based System Resource Management Scheme Design

Based on WSDM, we propose a novel system resource management scheme-WSReMS. Furthermore, we have developed and implemented a prototype system: WSReMS-Platform, WSReMS makes the physical IT resources virtualized by applying the international standards-WSDM. These resources are to be applied in our target scenarios to verify the effectiveness and the behavior of the WSReMS.

With our WSReMS-Platform, the system configuration manager can configure the properties of the managed IT resources. The operator can monitor the performance of the IT resources. The operator can subscribe to some management events. If an event happens, the operator shall get the notification. Then the operator can also execute some operations to react to the notification and to deal with the situation.

The WSReMS-Platform consists of three layers. They are the Manager Layer, the Gateway Layer and the Agent Layer as the following figure 1.

Manager Layer is a manageability consumer. It manipulates the manageable resource by invoking the management interface exposed by the Gateway component.

Gateway Layer takes the responsibility of mapping the physical resources that are manipulated by agent component to the WSDM resource. It should also provide WSDM-required management interface and the custom management interface which are mapped to the corresponding agent management interface. In a word, Gateway component takes charge of the property-mapping and management interface-

mapping job between the physical resources and corresponding WSDM resources.

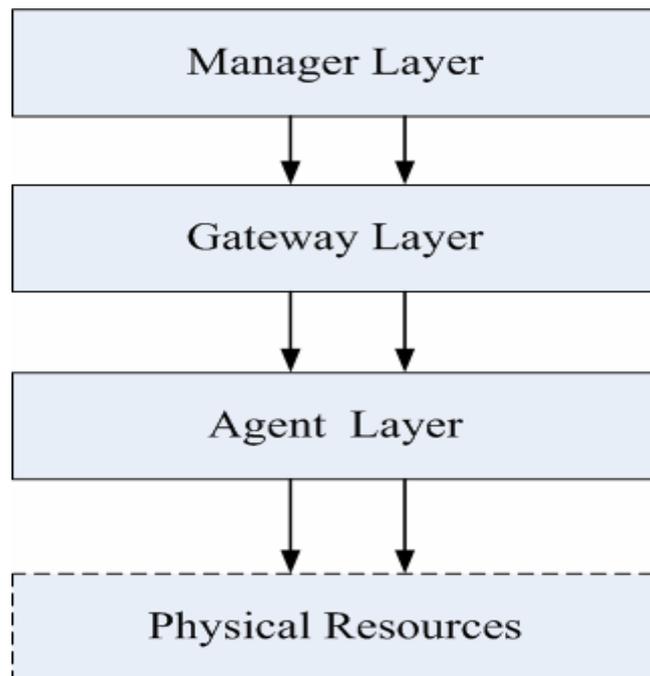


Figure1 General Description of WSRMS 3-Layer Architecture

Agent Layer uses its own management interface (such as JMX MBean, CIM interface, SNMP interface .etc.) to manipulate physical resources. Developers of software can use whatever technology they prefer to implement the manipulation on the ‘Real Resource’.

WSReMS Object Modeling Design

Object Modeling Design contains the Class design and the relationship of the classes of the Object in the WSRMS, including the attributes and operations.

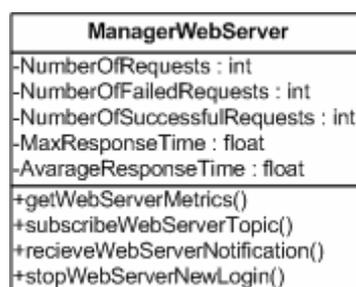


Figure2 Class Diagram of Manger Web Server

The Manager Web Server Class (Figure 2- Class Diagram of Manger Web Server) is the modeling of the service that is monitored in the Manager Component. The Manager Web Server Class has some operations that can be executed to react to some certain situation.

The instance image (Figure3 Instance Image of the Classes of Manager Layer) gives the general idea of the relationship of the classes we have designed above.

The relationship between the MangerService and the ManagerDatabaseServer (ManagerDBSvr), ManagerWebServer (ManagerWebSvr), ManagerLoadBalancer (ManagerLB) is aggregation.

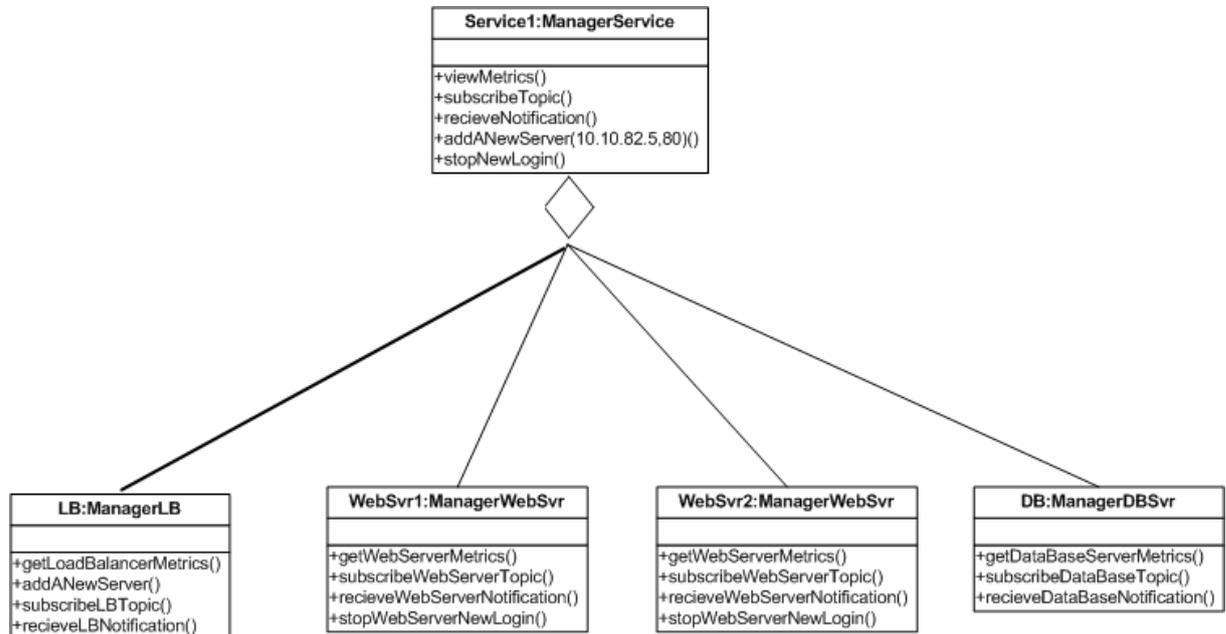


Figure3 Instance Image of the Classes of Manager Layer

3. MUSE and CIMOM Based Scheme Implementation and Experiment Analysis

In WSRMS -Platformv1.0, we have implemented the service monitoring and switching of Web server and DB server using CIMOM and MUSE Project development library [6].

In this system, there are three parts: Manager Layer, WSDM-Gateway Layer, and Agent Layer-the local resource service point. The WSDM-Gateway side is in charge of service status information collecting and publishing, while Manager Layer is in charge of retrieving status data from every resource service point, and managing and switching service.

WSReMS-Platform is developed using CIMOM-Common Information Model Object Manager[9]. It is the implementation of the core part of WBEM (DMTF - Web-Based Enterprise Management) protocols [7]. WSDM-Gateway Layer is developed based on the open source software 'MUSE 2.0', as its implementation of WSDM standards. MUSE [6] – is a Java-based implementation of the WS-Resource Framework (WSRF) 1.2, WS-Base Notification (WSN)1.3, and WS-Distributed Management (WSDM) 1.1 specifications.

In agent layer, we use JMX standard to encapsulate all the resource (hardware or

software) with java objects and expose them in a distributed environment at the resource service point side.

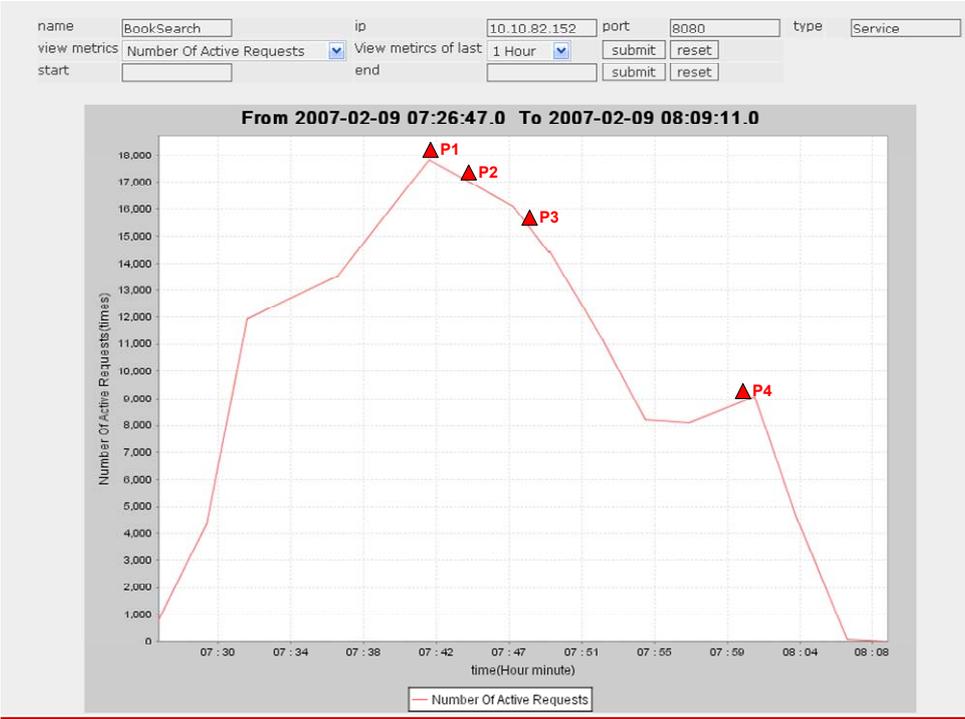


Figure4 Experiment Result Description

In our current WSRMS -Platform, we realize a case of a single service management. When overload scenario of the single service happens, the operator will get the notification. And then the new web server can be added to support more requests. Figure4 give a description of experement result. Manager first retrieves the numberOfCurrentRequests of the service and then it retrieves the numberOfCurrentRequests of the web1. Because of the time delay, P1 and P2 could be of different positions. And normally during our tests P2 is later than P1.

At P3, when web3 is added to the cluster, the numberOfCurrentRequests of the service drops down.

At P4, the increase of the value is not included in our test. The test ends at 08:00.

4.Conclusion and Futrue Works

In our current WSRMS -Platform, we just realize a management case of a single service. Actually, we don't use the complex composite service. In general, composing multiple Web services, rather than accessing a single service, is essential and provides more benefits to users. Composition primarily addresses the situation of a user's request that cannot be satisfied by any available service, whereas a composite service obtained by combining available services might be used [10].When the

complex Web Services composition is used in WSReMS -Platform, we should consider the following problem must be solved:

(1) The structure of the composite services, which is described by BPEL, is complex. If WSReMS -Platform would like to manage the composite service, the structure of the WSReMS -Platform, especially the Manager Layer might need to be re-designed. And the target service might need to re-build.

(2) We might need to consider the metrics of the composite service. What kind of metrics is useful should be evaluated carefully. The refinement of the metrics of the current system might be necessary.

WSDM-based System Resource Management Scheme is a new distributed system management scheme. The new scheme further being realized in global Internet will necessarily make this technology closer to business field and obtain more industry space. Before realizing WSDM-based System Resource Management service on a worldwide scale, researchers must overcome many key technologies and carry on many experiments in test bed with carefully controlled trials, continuously validating its scalability and evaluating its actual performance.

With the rapid development of Web Services, and WSDM technology, we believe this novel system monitoring and management solution will necessarily obtain powerful and persistent life force and industrialization development space.

References:

- [1] Web Services Distributed Management: Management Using Web Services (MUWS 1.0), OASIS Standard, Mar. 2005.
- [2] Web Services Distributed Management: Management Of Web Service (MOWS 1.0), OASIS Standard, Mar. 2005.
- [3] Modeling Statusful Resources using Web Services. [http://www-106.ibm.com/ developerworks/library/ws-resource/ws-modelingresources.pdf](http://www-106.ibm.com/developerworks/library/ws-resource/ws-modelingresources.pdf).
- [4] Management Using Web Service-A proposal Architecture and Roadmap, IBM, CA, HP.
- [5] Web Services for Management (WS-Management), Sun, Microsoft, Dell, Intel, Oct. 2004.
- [6] MUSE Project, <http://ws.apache.org/muse/>.
- [7] DMTF - Web-Based Enterprise Management (WBEM)-, <http://www.dmtf.org/standards/wbem/>.
- [8] Web Service Management for Adaptive Control, Sven Graupner, Tilo Nitzsche, HP Tech Con/04, Jun. 2004.
- [9] Proposal for a CIM Mapping to WSDM, CA, IBM, Cisco, CA, Jan. 2005.
- [10] D. Berardi, D. Calvanese, G. De Giacomo, M. Lenzerini, and M. Mecella, "A Foundational Vision for e-Services," Proc. Workshop Web Services, e-Business, and the Semantic Web (WES '03) 2003.

Authors:

Dr. Zhihui Lu

Professor Yiping Zhong

Ph.D. Candidate Yu Wu

Assistant Professor Jie Wu

Fudan University, Department of Computing & Information Technology, 220 Handan Road
200433, ShangHai, China

Phone: 86-21-6510-7607, 86-21-6564-3189

Fax: 86-21-6564-7894

E-mail: ypzhong@fudan.edu.cn, lzh@fudan.edu.cn, yuwu@fudan.edu.cn, jwu@fudan.edu.cn

M. Heit / E. Jenennchen/ V. Kruglyak/ D. Westermann

Simulation des Strommarktes unter Verwendung von Petrinetzen

3.6 Energy System Modelling and Simulation

Verbunden mit der zunehmenden Forderung nach Klimaschutz und Energieeinsparung, einem steigenden Energiebedarf in Europa, der Liberalisierung der Energiemärkte beginnend seit 1998 und der Osterweiterung der Europäischen Union werden die Energieversorgungssysteme immer komplexer. Die mathematische Beschreibung des Gesamtsystems, die so genannte Simulation, unter Berücksichtigung der transporttechnischen Infrastruktur gewinnt zunehmend an Bedeutung. Im Mittelpunkt steht dabei die Vorhersage der Preisentwicklung in Abhängigkeit von der jeweiligen Angebots- und Nachfragesituation und den vorhandenen Netzkapazitäten mit ihren Engpässen. Maßgeblich wird dabei der Strommarkt betrachtet, da die elektrische Energieversorgung die kritischste aller Infrastrukturen dar stellt. Der im Rahmen dieses Papers beschriebene Simulationsalgorithmus, der zukünftig als Softwaretool umgesetzt wird, soll aber auch für andere leitungsgebundene Energieträger, schwerpunktmäßig Erdgas, angewendet werden können.

Das vorliegende Paper gibt einen Überblick zur Problemstellung und erläutert, wie der Strommarkt mit Hilfe von Petrinetzen beschrieben werden kann. Es zeigt ein vereinfachtes Marktmodell, legt die Vorgehensweise dar und enthält erste mathematische Ansätze.

Der Energiemarkt kann durch ein vereinfachtes Knotenmodell [1] dargestellt, mit mathematischen Algorithmen [Formeln 1-5] abgebildet und unter Verwendung eines Petrinetzes [2] beschreiben werden. Dabei eignen sich Petrinetze besonders gut zur bildlichen Darstellung dynamischer Vorgänge mit komplexen internen Vorgängen [3]. Letzteres ist typisch für die Transportnetze elektrischer Energie.

Zur Vereinfachung des gesamten Verbundnetzmodells werden die ausgewählten Länder Deutschland, Tschechien und Polen repräsentativ betrachtet, da hierfür bereits erste

Untersuchungen durchgeführt wurden. Für diese drei Länder soll ein Marktsimulator erstellt werden, der die Wechselwirkung zwischen den Transportnetzen und den Handelsaktivitäten abbilden soll.

Der prinzipielle Aufbau eines Energiemarktes ist in Abbildung 1 dargestellt. Die Quellen des Petrinetzes (Position 2) sind die Summe der Einspeisung der Kraftwerke (die Energieerzeugung). Die Senken (Position 3) bilden die elektrische Last ab (Nachfrage an elektrischer Energie). Die Knoten der Energiemärkte (Position 1) sind die Einrichtungen zur Abwicklung des Energiehandels (Börse). Die Quellen beinhalten die Informationen der länderspezifischen Merit-Order-Curve. Diese Informationen werden an die dazugehörigen Strommarktknoten weiter gegeben. Die Senken jedes Landes tragen die Information über die Lastabnahme. Diese Daten werden an die Strommarktknoten in Form der Nachfragekurve weitergegeben. In den Strommarktknoten kann mit Hilfe dieser Daten der aktuelle Strompreis, d.h. der System Marginal Price (SMP) berechnet werden.

Einfaches Strommarktmodell für ein Land

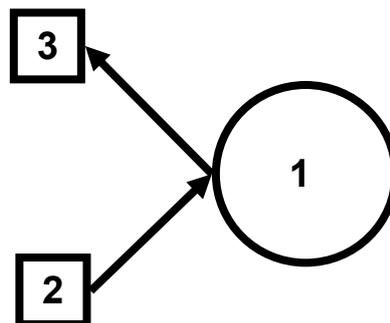


Abbildung 1 Einfaches Strommarktmodell für ein Land

Auf Grund der Liberalisierung des europäischen Strommarktes ist der Energiehandel auch zwischen den einzelnen Ländern möglich. Hierfür muss das Modell erweitert werden (Abbildung 2). Der Energietransport zwischen den Ländern erfolgt über die Kanten des Petrinetzes. Jeder Strommarktknoten ist mit den Strommarktknoten der anderen zwei Länder über Kanten in beiden Richtungen verbunden. Die Kanten besitzen Transitionen (Position 5), welche die leitungsgebundenen Übertragungskapazitäten an den Kuppelstellen abbilden. Um den Strompreisaufschlag an den Engpässen einbeziehen zu können, wird nach den Transitionen in Flussrichtung ein weiterer Knoten eingefügt (Position 4), der die fremdbezogene Leistung mit einem aufsummierten Preis für den Transport am Engpass an den Strommarktknoten weitergibt. Diese Daten werden bei der Berechnung des resultierenden Gesamtpreises eines Landes

wurde, unabhängig davon, über welche Engpässe die Übertragung erfolgt; welches der aktuellen Berechnungsformel entspricht. Für diesen Fall existiert keine maximal übertragbare Liefermenge zwischen zwei Ländern, sondern die Liefermenge wäre begrenzt durch die Kapazität der Kuppelstellen von allen Ländern. Wenn die Kapazität zwischen Land B und Land C voll ausgenutzt ist, kann Land B von Land C keinen Strom mehr an der Börse beziehen. Jedes einzelne Land kann nur über den Import, nicht aber über den Export entscheiden. Wenn also Land A aus Land B Strom beziehen will, kann Land B das nicht verhindern, auch wenn der Strompreis dadurch in Land B ansteigt. Die Einzige Beschränkung ist die Kapazität an den Länderengpässen. Wenn diese Einschränkung nicht vorhanden wäre, würde sich an allen drei Märkten der gleiche Gleichgewichtspreis einstellen. Es soll verhindert werden, dass ein drittes Land von den Importen profitiert. Dies wird realisiert, in dem nicht die Merit-Order-Curve eines Landes verändert wird, sondern die Nachfrage der einzelnen Länder variiert. Die Nachfrage berechnet sich wie folgt:

$$x_a = x_{aa} + x_{ab} + x_{ac} \quad \text{Formel 1}$$

$$x_{ab} = -x_{ba} \quad \text{Formel 2}$$

Mit: x_a – gesamte Nachfrage im Land A
 x_{aa} – eigene Nachfrage des Landes A
 x_{ab} – Liefermenge von Land A nach Land B

Also berechnet sich die gesamte Nachfrage des Landes A aus dem Eigenbedarf plus der Lieferungen an die restlichen Länder. Wenn Land A nicht Exporteur sondern Importeur ist würde die Lieferung nicht von A nach B erfolgen, sondern von B nach A. Somit würde nach Formel 2 bei der Berechnung der gesamten Nachfrage die Liefermenge nach B vom Eigenbedarf abgezogen. Im Falle des Stromexportes steigt die gesamte Nachfrage eines Landes und so auch unter Umständen der Strompreis an der Börse dieses Landes, auf Grund der höheren Nachfrage; im Falle des Importes sinkt die gesamte Nachfrage in einem Land und somit auch der Strompreis an der dazugehörigen Börse. Der Strompreis an der Börse berechnet sich nach dem Schnittpunkt der Merit-Order-Curve ($f(x)$) und der gesamten Nachfrage (x).Also:

$f_a(x_a)$ - der Börsenpreis des Landes A bei der gesamten Nachfrage im Land A.

Die Merit-Order-Curve (Abbildung 3) ist eine unstetige Stufenfunktion mit mehreren Stufen.

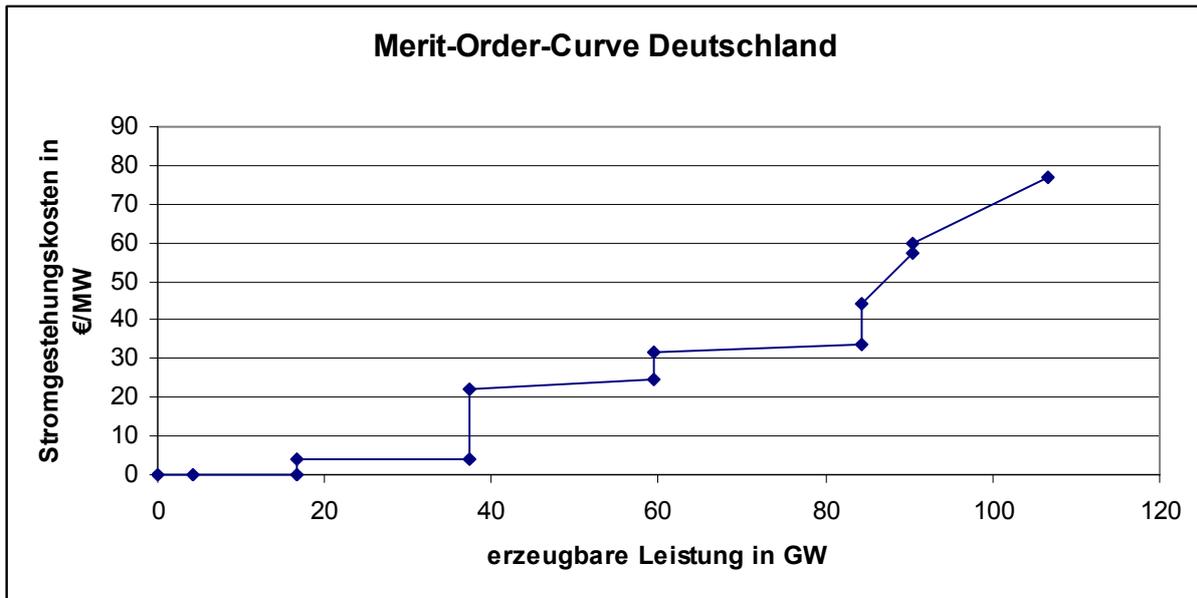


Abbildung 3 Merit-Order-Curve Deutschland

Variablendeklaration :

- D - Deutschland
- P - Polen
- T - Tschechien
- $f_a(x)$ - Merit-Order-Curve in Land A
- x_a - gesamte Nachfrage in Land A
- x_{aa} - eigene Nachfrage von Land A
- x_{ab} - Exportmenge von A nach B
- x_{abmax} - maximal übertragbare Liefermenge am Länderengpass zwischen A und B
- P_{ab} - Preis für Engpass zwischen A und B
- P_{ges_a} - resultierender Strompreis für Land A

Unter oben genannter Variablendeklaration werden die resultierenden Strompreise der Länder Deutschland, Polen, Tschechien wie folgt berechnet:

Deutschland:

$$P_{ges_D}(x_{DP}, x_{DT}) = \frac{f_D(x_D) \circ x_D + \begin{pmatrix} x_{PD} \circ f_P(x_P) + P_{PD} & \{\text{für } x_{PD} > 0\} \\ x_{PD} \circ f_D(x_D) & \{\text{für } x_{PD} < 0\} \\ 0 & \{\text{für } x_{PD} = 0\} \end{pmatrix} + \begin{pmatrix} x_{TD} \circ f_T(x_T) + P_{TD} & \{\text{für } x_{TD} > 0\} \\ x_{TD} \circ f_D(x_D) & \{\text{für } x_{TD} < 0\} \\ 0 & \{\text{für } x_{TD} = 0\} \end{pmatrix}}{x_{DD}}$$

Formel 3

Polen:

$$P_{ges_P}(x_{PD}, x_{PT}) = \frac{f_P(x_P) \circ x_P + \begin{pmatrix} x_{DP} \circ f_D(x_D) + P_{DP} & \{\text{für } x_{DP} > 0\} \\ x_{DP} \circ f_P(x_P) & \{\text{für } x_{DP} < 0\} \\ 0 & \{\text{für } x_{DP} = 0\} \end{pmatrix} + \begin{pmatrix} x_{TP} \circ f_T(x_T) + P_{TP} & \{\text{für } x_{TP} > 0\} \\ x_{TP} \circ f_P(x_P) & \{\text{für } x_{TP} < 0\} \\ 0 & \{\text{für } x_{TP} = 0\} \end{pmatrix}}{x_{PP}}$$

Formel 4

Tschechien:

$$P_{ges_T}(x_{TD}, x_{TP}) = \frac{f_T(x_T) \circ x_T + \begin{pmatrix} x_{DT} \circ f_D(x_D) + P_{DT} & \{\text{für } x_{DT} > 0\} \\ x_{DT} \circ f_T(x_T) & \{\text{für } x_{DT} < 0\} \\ 0 & \{\text{für } x_{DT} = 0\} \end{pmatrix} + \begin{pmatrix} x_{PT} \circ f_P(x_P) + P_{PT} & \{\text{für } x_{PT} > 0\} \\ x_{PT} \circ f_T(x_T) & \{\text{für } x_{PT} < 0\} \\ 0 & \{\text{für } x_{PT} = 0\} \end{pmatrix}}{x_{TT}}$$

Formel 5

Eine Optimierung über die variablen Im-/Exportgrößen x_{ab} , x_{ac} , x_{bc} ermöglicht, durch die Gleichungen 3-5, die Berechnung des jeweils niedrigsten resultierenden Strompreises der einzelnen Länder.

Ziel des Strommarktsimulators ist, die Strompreisentwicklung bei technischen Änderungen abzubilden. Solche technischen Änderungen können z. B. der Ausbau des Übertragungsnetzes an den Engpässen, oder auch der Ausstieg aus der Atomenergie sein. Diese Simulation soll als Grundlage bei der Wirtschaftlichkeitsbetrachtung von Investitionen in den Bereichen der Energieerzeugung und des Energietransportes dienen.

References:

- [1] Zolotarev, A. A.; Kryvenka, M.; Heit, M.; Westermann, D.: Mathematische Modellierung und Optimierung von Energieflüssen im deregulierten Umfeld, Tagungsband zum 51. Internationalen Wissenschaftlichen Kolloquium (IWK), Ilmenau, 2006
- [2] Rosenstengel, B.; Winand, U.: Petri-Netze – eine anwendungsorientierte Einführung, Vieweg-Verlag, Braunschweig, 1991, ISBN 3-528-33582-3
- [3] Schneeweiss, W. G.: Petri-Netz-Bilder-Buch (Eine elementare Einführung in die beste bildliche Darstellung zeitlicher Veränderungen), LiLoLe-Verlag, Hagen, 2002, ISBN 3-934447-05-8

Authors:

Univ.-Prof. Dr.-Ing. Dirk Westermann
Dr. Dipl.-Ing. Michael Heit
M. Sc. Valerija Krugljak
cand. ing. Eva Jenennchen
Technische Universität Ilmenau, Postfach 10 05 65
98684 Ilmenau
Phone: + 49 3677 69-2838
Fax: + 49 3677 69-1496
E-mail: dirk.westermann@tu-ilmenau.de

Olaf Sauer / Miriam Ebel

Engineering of production monitoring & control systems

Abstract

Adaptivity and interoperability of production equipment present extensive difficulties in business practice today, particularly if production systems are modified.

Production systems are adapted continually as products change, as capacities have to be re-adjusted owing to varying demands or as more efficient manufacturing technologies are applied. In practice modifications in production plants do not only result in equipment being shifted physically within a plant. Rather, they cause adjustments in the software of controlling machinery and equipment, for example programmable logic controller (PLC), as well as in the information technology (IT) which superposes the direct facility control, monitors automated plants or plans and controls the allocation of equipment. CIRP, the international organization of production scientists, considers this capability to undergo permanent change to be the strategic competitive advantage of plants and production systems in the future.

1 Introduction

At the same time, manufacturers of machinery and equipment have to re-adjust themselves continually to the individual 'company standards' of their customers who are the operators of the plants. One of the standards Fraunhofer IITB is facing in its projects is DaimlerChrysler's 'IntegraDCX' standard.

This standard comprises

- components such as PLCs, drives, etc.,
- automation functions such as ProfiNet,
- a methodology for equipment diagnostic such as plant visualization,
- distribution to production sites and supplier of equipment,
- support, e. g. 1st, 2nd and 3rd level and
- a staff training concept.

The standardization of components includes

- requirements for function blocks of PLCs and
- naming conventions for PLC variables.

Due to the benefit of such a standard for plant operators like DaimlerChrysler, other operating companies are likely to develop similar company standards in the near future – which is a big challenge for medium-sized machinery and equipment manufacturers. These ‘company standards’ including the above mentioned requirements have implications for machinery and equipment manufacturers with regard to

- naming conventions of PLC variables and function blocks,
- equipment components to be used,
- investment in engineering (CAE) and plant software,
- automation systems to be used,
- the proceeding of mechanical and electrical engineering and
- methods and tools of PLC programming.

Ultimately, this can result in equipment manufacturers losing their unique features and their own, cost-cutting standardization potentials.

Owing to the specific requirements of plant technology that results from the multitude of production tasks there is a nearly endless variety of machinery controls, software versions and manufacturing execution systems today. Whenever the equipment is changed, the software has to be adjusted as well, which results in high costs for the manufacturing industry.

Today 'digital factory' tools are mainly used to plan production systems, whereas operational processes are assisted by manufacturing execution systems which support the functional operation on the shop floor (figure 1).

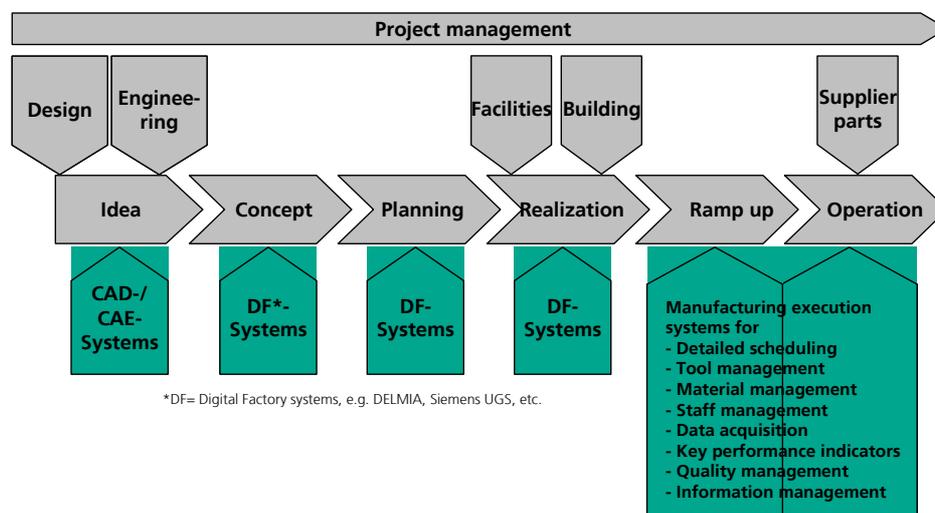


Figure 1: Cooperation between IT systems supporting planning and operation [Sr04]
Up to now, there has been no standard protocol between manufacturing plants, their

controls and the super ordinate IT ruling, for instance, the structure and contents of the capabilities and states of manufacturing equipment transmitted to the controlling or monitoring software. In contrast to this, there are a few standards for the integration of field devices into the industrial network, e. g. Industrial Ethernet, ProfiBus, ProfiNet, etc. However, most of these standards refer to the relevant network protocol or the supplier, i. e. only to Siemens PLCs or other brands.

In the future, information stored in 'digital factory' tools will be used to parameterize manufacturing plants and super ordinate IT systems as well as to start and operate them virtually. This aims at having the corresponding operational IT systems fully available when the modified or new production plants start up (figure 2).

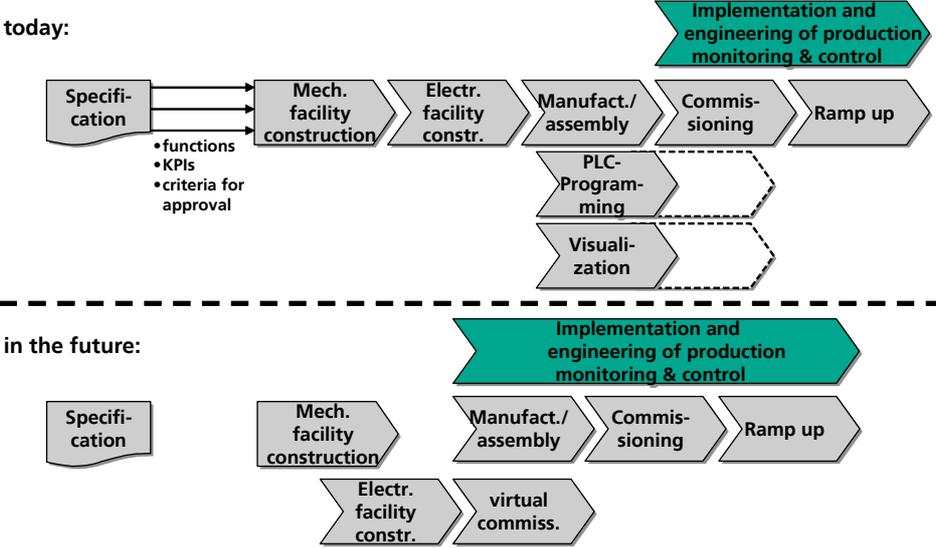


Figure 2: The early coupling of planning and operation aims at an early start up of shop-floor-related IT systems.

2 Trends in shop-floor-related IT systems

With reference to the enterprise-wide IT hierarchy, the efforts described below address the manufacturing control and – in part – the cell level (figure 3). Currently, it is expected that these systems will turn into factory information hubs in the years to come, both in discrete manufacturing and in the process industry. Nevertheless, their market share does not exceed 5 % to 10 % today, whereas the annual growth rate is projected to amount to approximately 11 % by 2010. Against this background, the efforts described in this article can be regarded as driving forces behind manufacturing execution systems because they contribute significantly to ensuring that shop-floor-related IT and plant controls are able to communicate in an automated and supplier-independent way.

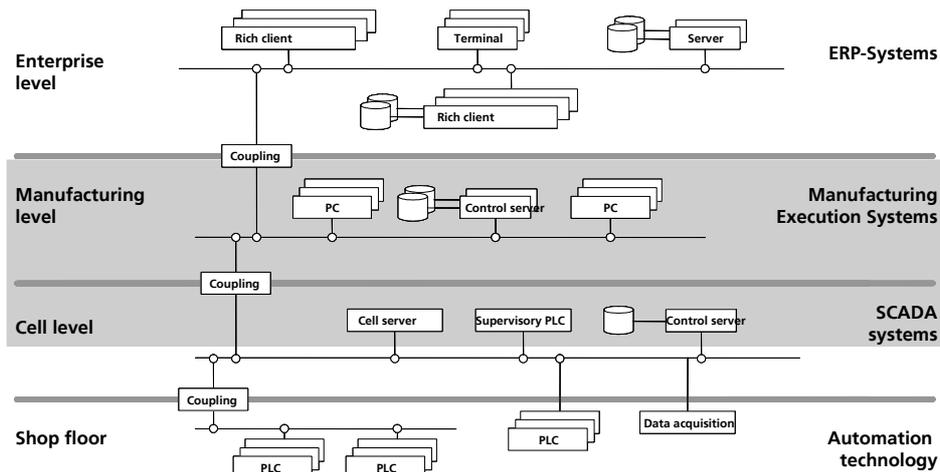


Figure 3: IT levels in manufacturing (Source: Betriebshütte, based on VDI guideline 5600, green paper)

The following section will deal with the trends that can be identified for tomorrow's MES systems. Two of these trends, namely trend 1 and 3, will be highlighted by means of examples because they are relevant for the closer integration of the 'digital factory' and shop-floor-related IT systems. In this context Simulation in the sense of concurrent realtime simulation is gaining more and more importance. It allows to forecast the effects of modifications to the process at any time, given the increased complexity of production and IT.

Tomorrow's manufacturing execution systems will be characterized by

- full integration with the 'digital factory', aiming at being ready to plan any time,
- simulation as a front-end in the sense of concurrent realtime simulation, allowing a quick response to unexpected events
- vertical integration with the manufacturing level, making use of standard plug-and-work mechanisms
- horizontal integration through service-oriented structure and consistent data management
- scalability including the support of decentral self-organizing production (RFID instead of factory data acquisition)
- 'human-centered' thanks to the task and role-specific provision of information to users

Trends 1 and 3 will be highlighted in this article.

3 Current projects to link the 'digital factory' and MES

3.1 Overview

The foreseeable integration of planning and operation calls for research and development as well as standardization services, e. g. the development of interfaces between system environments. The Fraunhofer Institute for Information and Data Processing (IITB) and its 'Monitoring & Control Systems' business unit focuses on making data of the 'digital factory' available for manufacturing execution systems. This implies, for instance, that data necessary for the engineering of MES systems are collected from 'digital factory' tools in a neutral exchange format such as XML and made available to the MES engineering. The 'digital factory' stores information on plant structure, plant parameters manufacturing processes and the arrangement of equipment. MES engineering also requires information about the topology and topography of production plants their parameters and manufacturing processes as well as PLC programs and variables.

The efforts of the Fraunhofer IITB are aimed at developing a kind of Universal Serial Bus (USB) for production. Yet unlike the well-known universal interface for personal computers, the conditions and requirements in the field of production are far more complex than for PCs.

The vision of plug and work between equipment and MES systems includes

- automated engineering of control systems and other MES systems resulting in earlier start-ups of software systems and fewer engineering errors
- provision of a technology defining an enterprise-wide namespace allowing machinery/equipment and IT infrastructure to communicate unambiguously about contents and meaning
- provision of mechanisms allowing new machinery/equipment to be identified automatically in a production system, including a description of the corresponding manufacturing capacity of the new machinery/equipment
- development of methods, procedures and software components enabling the automated integration of plants with the superordinate IT systems and mechanisms so that the information stored in the plant can be made available to these systems.

3.2 Status of the engineering of shop-floor-related IT systems

The current status of the engineering of new or modified production systems in superordinate IT systems is characterized by a large share of manual work.

Figure 4 shows a common example of the actual situation in today's enterprises, which is marked by system discontinuity and manual intervention when equipment and its controls are integrated in operational IT applications. The first step to enable communication is to publish the IP address of the control that has to be integrated on the network. Subsequently, the control-specific OPC server has to be announced to the client on which the operational application is running. The next step is to select all relevant variables from the set of variables provided by the OPC server and to link them with the process image created in advance. This is done manually using a browser (figure 4).

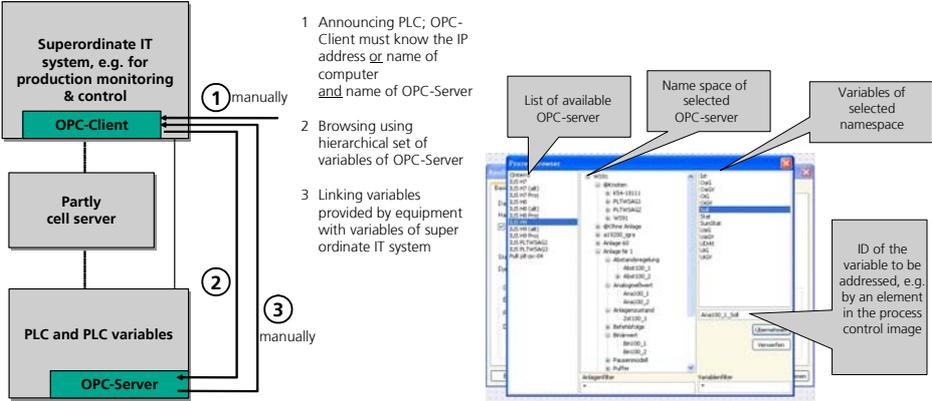


Figure 4: The example of PLCs illustrates the way equipment and its controls are integrated today.

In summary, the starting situation is characterized by a large amount of manual customizing as well as manual configurations of plant controls and super ordinate IT systems.

3.3 XML to support plug and work

IITB is specifically working on methods, software components and applications that can be standardized and that allow production equipment to be integrated into a production system simply, quickly and safely or modifications in equipment and its controls to be announced automatically to the production system and the super ordinate IT (figure 5). To this end, IITB's software engineers make use of existing standards, particularly CAEX (Computer Aided Engineering Exchange) to describe the static features of production plants and OPC-UA for dynamic components. CAEX is a process industry standard to describe the architecture and structure of process-engineered plants, whereas OPC-UA is used for control-relevant variables of which the values change dynamically during production.

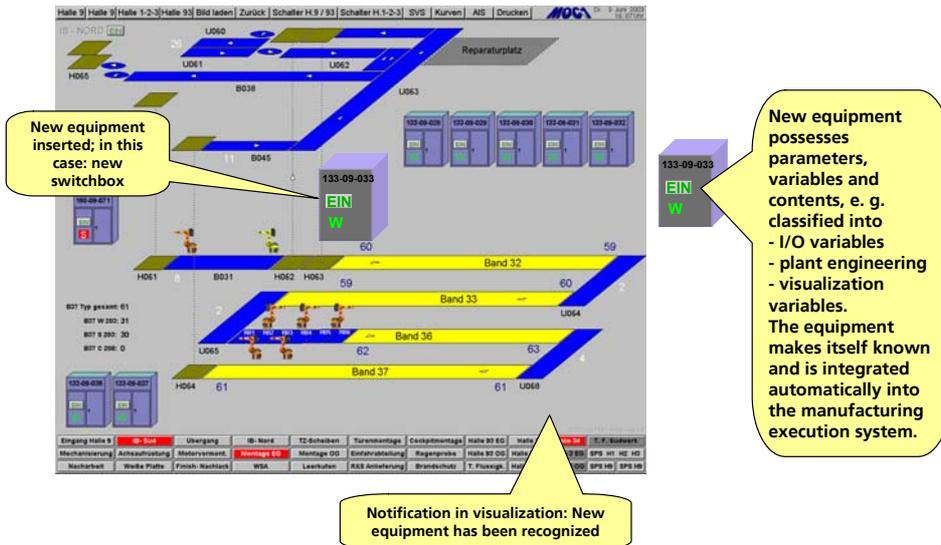
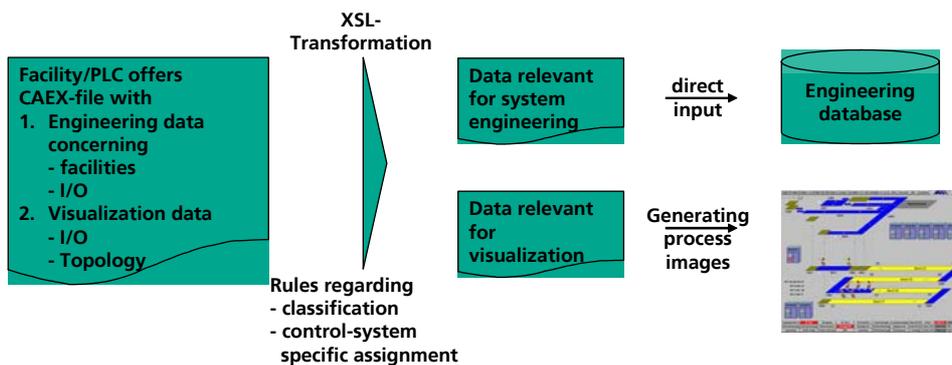


Figure 5: Plug and work of production equipment illustrated by the visualization of a plant monitoring system

In this process, equipment and its controls provide files describing their capabilities in a standardized format. The format chosen by IITB is CAEX, based on IEC-PAS-62424. By transformation, the files are classified into information relevant for system engineering and for visualization. This is used either to generate process control images or to transmit the information to a database, on the basis of which I/O and plant system engineering are generated for the process control image of the runtime system (figure 6).



Goal of development:
OPC-UA-Server

Figure 6: Application of XML standards for automated MES engineering

4 Outlook

The above mentioned work will be completed by the capability of using data from factory planning software tools. In this context, the most relevant data for automated engineering is topologic and topographic information such as the layout of factory halls,

allowing to arrange the elements of process control images that need to be visualized correctly. In addition, PLC programs in the sense of plant logics are of importance. IITB is cooperating with the well-known suppliers of 'digital factory' tools make this information available and utilizable for MES systems.

References

- [Dr04] Draht, R.; Fedai, M.: CAEX - ein neutrales Datenaustauschformat für Anlagendaten – parts 1 and 2. Automatisierungstechnische Praxis - atp volume 46 (2004) issue 2, pages 52-56 and issue 3, pages 20-27.
[Sr04] Sauer, O.: Einfluss der Digitalen Fabrik auf die Fabrikplanung. wt Werkstattstechnik online, issue 01/02, 2004, pages 31-34.
[Sr06] Sauer, O.; Sutschet, G.: ProVis.Agent: ein agentenorientiertes Leitsystem – erste Erfahrungen im industriellen Einsatz. VDE conference 2006, Aachen: Innovations for Europe. 23-25 October 2006, volume 2: pages 297-302.
[Sr07] Sauer, O.: Digitale Fabrik und MES. IT&Production: MES Wissen Kompakt 2007, pages 18-21.
[Sut01] Sutschet, G.: Störung im Griff. Ein Produktionsassistent für die Automobilfertigung. visIT 2 (2001), No. 2, pages 6-7.
[VDI4499] VDI guideline 4499, part 1
[VDI5600] Green paper of VDI guideline 5600

Authors:

Dr.-Ing. Olaf Sauer
Dipl.-Inform. Miriam Ebel
Fraunhofer IITB, Fraunhofer Straße 1
76131 Karlsruhe, Germany
Phone: +49 721 6091 477
Fax: +49 721 6091 413
E-mail: olaf.sauer@iitb.fraunhofer.de
miriam.ebel@iitb.fraunhofer.de

C. Behn / K. Zimmermann

Biologically inspired Locomotion Systems and Adaptive Control

GENERAL APPROACH TO WLLS

In this paper we discuss the problem of developing worm-like locomotion systems (WLLS), which have the earthworm as a living prototype. Non-pedal forms of locomotion show many advantages and are very interesting in current robotic research.

The following is taken as the basis of our theory:

(i) a worm is a terrestrial locomotion system of one dominant linear dimension with no active legs nor wheels; (ii) global displacement is achieved by (periodic) change of shape (such as local strain) and interaction with the environment; (iii) the model body of a worm is a 1-dimensional continuum that serves as the support of various fields. The continuum in (iii) is just an interval of a body-fixed coordinate. Most important fields are: *mass*, continuously distributed (with a density function) or in discrete distribution (chain of point masses), *actuators*, i.e., devices which produce internal displacements or forces thus mimicking muscles, *surface structure* causing the interaction with the environment.

Observing the locomotion of worms one recognizes first a surface contact with the ground. It is well known, that, if there is contact between two bodies (worm and ground), there is some kind of friction, which depends on the physical properties of the surfaces of the bodies. In particular, the friction may be anisotropic (orientation dependent of the relative displacement). This interaction (mentioned in (ii)) could emerge from a surface texture as asymmetric Coulomb friction or from a surface endowed with scales or bristles (we shall speak of *spikes* for short) preventing backward displacements. It is responsible for the conversion of (mostly periodic) internal and internally driven motions into a change of external position (undulatory locomotion [5]), see [6] and [9].

Summarizing, we consider a WLLS in form of a chain of point masses in a common straight line (**a discrete straight worm**), which are connected consecutively by linear visco-elastic elements, see [1], [2], [9] for instance and Fig. 1.

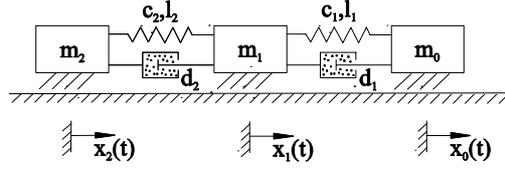


Fig. 1. Model of a WLLS - chain of point masses

In [6] and [7] the case is considered, where the point masses are endowed with scales, which could make the friction also orientation dependent (in sliding forward the frictional forces are minimal while in opposite direction the scales dig in and cause large friction). But, due to [1] and [2], we do not want to deal with reactive forces, we model this ground interaction as impressed forces - asymmetric (anisotropic) dry friction as a Coulomb sliding friction force.

WLLS AS A DYNAMICAL CONTROL SYSTEM

We model the ground interaction as an **asymmetric Coulomb dry friction force** F , which is taken to be different in the magnitude depending on the direction of each point mass motion:

$$\dot{x} \mapsto F(\dot{x}) = \begin{cases} -F^+ & , \dot{x} > 0 \\ F^0 & , \dot{x} = 0 \\ F^- & , \dot{x} < 0 \end{cases} \quad (1)$$

where $F^+, F^- > 0$ are fixed with $F^- \gg F^+$ and F^0 is arbitrary, $F^0 \in (F^-, -F^+)$ (neglecting stick-slip). For later simulation we restrict the number of point masses to $n = 2$, but we point out that our theory is valid for fixed but arbitrary $n \in \mathbb{N}$, see [3].

Mathematical model

Firstly, we derive the differential equations of motion of the WLLS by using Newton's second law:

$$\begin{aligned} m_0 \ddot{x}_0 &= -c_1(x_0 - x_1) - d_1(\dot{x}_0 - \dot{x}_1) + F_0(\dot{x}_0) + u_1(t) \\ m_1 \ddot{x}_1 &= \begin{cases} -c_1(x_1 - x_2) + c_2(x_0 - x_1) - d_1(\dot{x}_1 - \dot{x}_2) + d_2(\dot{x}_0 - \dot{x}_1) \\ + F_1(\dot{x}_1) + u_2(t) - u_1(t) \end{cases} \\ m_2 \ddot{x}_2 &= c_2(x_1 - x_2) + d_2(\dot{x}_1 - \dot{x}_2) + F_2(\dot{x}_2) - u_2(t) \end{aligned} \quad (2)$$

with $x_0(0) = x_{00}$, $x_1(0) = x_{10}$, $x_2(0) = x_{20}$, $\dot{x}_0(0) = x_{01}$, $\dot{x}_1(0) = x_{11}$, $\dot{x}_2(0) = x_{21}$ (all initial values are real numbers). Putting

$$u_1 := c_1 l_1 \text{ and } u_2 := c_2 l_2 \quad (3)$$

then u_{ij} is in fact a control of the original spring length. Therefore, we have *internal* inputs and no longer external force inputs, as in [1]. New outputs of this system could

be the actual distances of the point masses

$$y_1 := x_0 - x_1 \text{ and } y_2 := x_1 - x_2. \quad (4)$$

Therefore, this system (2), (4) is described by a mathematical model that falls into the category of quadratic, nonlinearly perturbed, minimum phase, multi-input $u(\cdot)$, multi-output $y(\cdot)$ systems with strict relative degree two.

Control objective

For the further analysis we suppose that the masses are all equal, but unknown, also the damping factors and spring stiffnesses, and the friction magnitudes as well (**uncertain systems**). The consideration of uncertain systems leads to the use of adaptive control. The aim is to design universal adaptive controllers, which learn from the behavior of the system, so automatically adjust their parameters and achieve a pre-specified control objective. Precisely, given an arbitrarily small $\lambda > 0$, a control strategy $y \mapsto u$ is sought which, when applied to the system, achieves λ -tracking for every reference signal $y_{ref}(\cdot)$ (belonging to a certain function class, for instance a given favoured kinematic gait presented), i.e., the following:

- every solution of the closed-loop system is defined and bounded on $R_{\geq 0}$, and
- the output $y(\cdot)$ tracks $y_{ref}(\cdot)$ with asymptotic accuracy quantified by $\lambda > 0$ in the sense that $\max\{0, \|y(t) - y_{ref}(t)\| - \lambda\} \rightarrow 0$ as $t \rightarrow +\infty$.

The last condition means that the error $e(t) := y(t) - y_{ref}(t)$ is forced, via the adaptive feedback mechanism (controllers (5) and (6)), towards a ball around zero radius $\lambda > 0$, see Fig. 2.

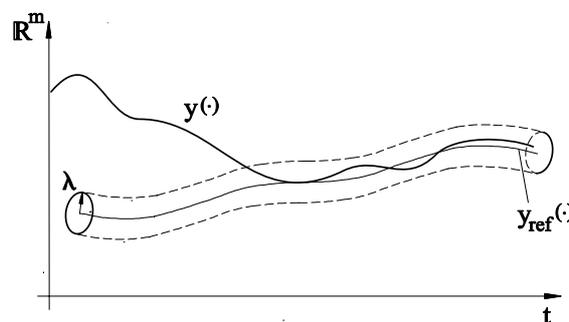


Fig. 2. The λ -radius around the reference signal

Controllers

Let us consider the following two λ -trackers, see also [1].

$$\left. \begin{aligned} e(t) &:= y(t) - y_{ref}(t) \\ u(t) &= -\left(k(t)e(t) + \frac{d}{dt}[k(t)e(t)]\right) \\ \dot{k}(t) &= (\max\{0, \|e(t)\| - \lambda\})^2 \end{aligned} \right\} \quad (5)$$

with $k(0) = k_0 \in R$, $\lambda > 0$, $y_{ref}(\cdot) \in W^{2,\infty}$ (a Sobolev-Space), $u(t), e(t) \in R^2$ and $k(t) \in R$.

The second one includes a dynamic compensator due to a controller of [4]. This controller allows us to avoid the drawback of using the derivative of the output:

$$\left. \begin{aligned} e(t) &:= y(t) - y_{ref}(t) \\ u(t) &= -\left(k(t)g(t) + \frac{d}{dt}[k(t)g(t)]\right) \\ \dot{g}(t) &= -k(t)^2 g(t) + k(t)^2 e(t) \\ \dot{k}(t) &= (\max\{0, \|e(t)\| - \lambda\})^2 \end{aligned} \right\} \quad (6)$$

with $\theta(0) = g_0$, $k(0) = k_0 > 0$, $\lambda > 0$, $y_{ref}(\cdot) \in W^{2,\infty}$, $u(t), e(t), g(t) \in R^2$ and $k(t) \in R$.

We stress, that the controller (6) does not invoke any derivatives. The structure of the feedback law and the simple adaptation law of the controllers in this subsection already exist in the literature, but they were only applied to systems with relative degree one. The considered WLLS has relative degree two. Therefore, the novelty is the application of the controller to systems with relative degree two. Only a few papers focus the adaptive λ -tracking problem for system with relative degree two, but the feedback law here is simpler than the introduced ones in [3], [8], [4]. These controllers achieve λ -tracking (for the proofs see [1]).

SIMULATIONS

We apply the presented simple adaptive λ -tracking control strategies to our WLLS in order to track a given reference signal: a kinematic gait developed in [7]. We try to track the “fast gait” in [7] in our dynamical theory, it is for $t \in [0,1]$:

$$t \mapsto y_{ref}(t) = \begin{cases} \left\{ \begin{array}{ll} l_0 [1 - \varepsilon(-1 + \cos(3\pi t))] & , t \in [0, 2/3) \\ l_0 & , t \in [2/3, 1) \end{array} \right. \\ \left\{ \begin{array}{ll} l_0 [1 - \varepsilon(1 - \cos(3\pi t))] & , t \in [0, 1/3) \\ l_0 [1 - 2\varepsilon] & , t \in [1/3, 2/3) \\ l_0 [1 - \varepsilon(1 + \cos(3\pi t))] & , t \in [2/3, 1) \end{array} \right. \end{cases}, \quad (15)$$

where l_0 is the original length (dimensionless chosen as 2 units) and $2\varepsilon = 0.3$ is the elongation. This gait is periodically repeated. Mind that one point mass is resting (active spike) at any time. In order to detect differences we present the simulation re-

sults with the λ -trackers (6) and (5), respectively, side by side.

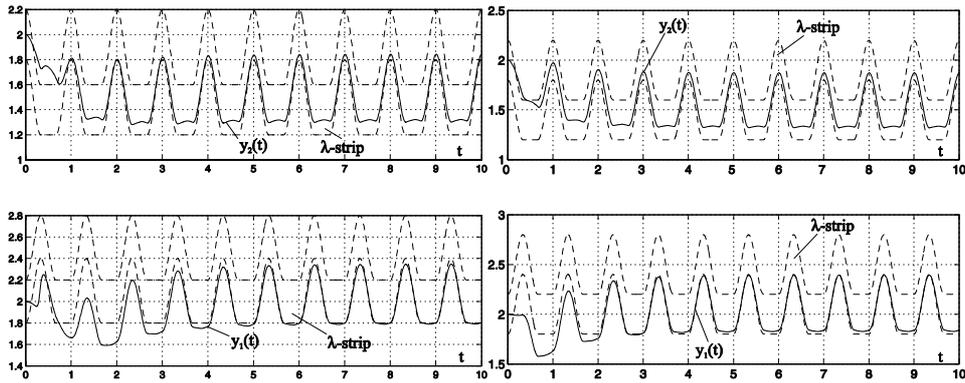


Fig. 3. Outputs and λ -strips – left: for (6), right: for (5)

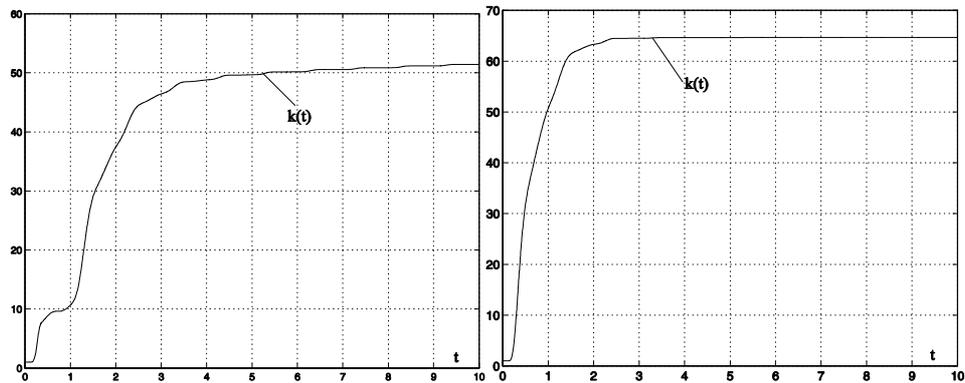


Fig. 4. The gain parameters - left: for (6), right: for (5)

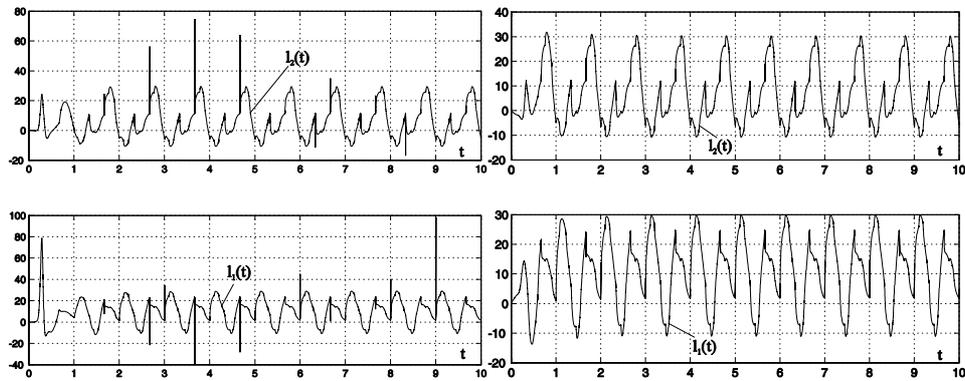


Fig. 5. The control inputs - left: for (6), right: for (5)

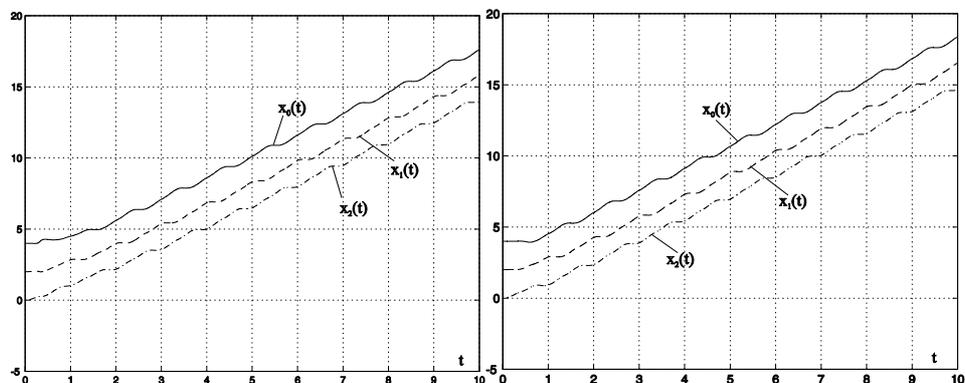


Fig. 6. The motions of the worm - left: with (6), right: with (5)

Fig. 3 shows the outputs of the systems and the according λ -strips. The reference

signal is tracked very quickly with controller (5) in comparison to controller (6). In Fig. 3, left, the outputs are not captured yet. The gain parameters, shown in Fig. 4, increase as long as the outputs are outside the λ -strips. Fig. 5 shows the necessary control inputs, and Fig. 6 the corresponding motions of the worm.

It can be seen that controller (5) works more effectively than controller (6) because we feed back more information about the output derivative than (6), which has to estimate the derivative. Hence, in the simulation with controller (6), the outputs are not captured on the considered time interval and the gain parameter is still increasing. Fig. 4, right, clearly shows the convergence of the gain parameter in the simulation with controller (5).

SUMMARY AND OUTLOOK

The motion of an earthworm was the inspiration for a (technical) solution of an artificial worm. In [7] a theory is developed for the peristaltic motion of such systems, which to a large extent allows to characterize these motions already on a kinematic level. Here, the advantage of adaptive control for the dynamical realization of these motions (tracking of kinematic gaits) is shown. The numerical simulations demonstrate and illustrate that the adaptive controllers work successfully and effectively. We point out, that the adaptive nature of the controllers is expressed by the arbitrary choice of the system parameters. It is obvious that for numerical simulation the system data have to be chosen fixed and known, but the controllers are able to adjust their gain parameter to each set of system data.

References:

- [1] Behn, C. (2005). Ein Beitrag zur adaptiven Regelung technischer Systeme nach biologischem Vorbild, Cuvillier, Göttingen
- [2] Behn, C. & Zimmermann, K. (2006). Adaptive lambda-tracking for locomotion systems, *Robotics and Autonomous Systems*, 54, pp. 529-545
- [3] Dragan, V. & Halanay, A. (1999). *Stabilization of Linear Systems*, Birkhäuser, Boston
- [4] Miller, D.E. & Davison, E.J. (1991). An adaptive controller which provides an arbitrary good transient and steady-state response, *IEEE Transaction on Automatic Control*, 36, pp. 68-81
- [5] Ostrowski, J.P., Burdick, J.W., Lewis, A.D. & Murray, R.M. (1995). The mechanics of undulatory locomotion: the mixed kinematic and dynamic case, *Proc. of IEEE Intern. Conf. on Robotics and Automation*, pp. 1945-1951, Nagoya, May, 1995, Japan
- [6] Steigenberger, J. (1999). On a class of biomorphic motion systems, *Preprint No. M12/99*, Faculty of Mathematics and Natural Sciences, TU Ilmenau, Germany
- [7] Steigenberger, J. (2004). Modelling artificial worms, *Preprint No. M02/04*, Faculty of Mathematics and Natural Sciences, TU Ilmenau, Germany
- [8] Ye, X. (1999). Universal lambda-tracking for nonlinearly-perturbed systems without restrictions on the relative degree, *Automatica*, 35, pp. 109-119
- [9] Zimmermann, K., Steigenberger, J. & Zeidis, I. (2003). On artificial worms as a chain of mass points, *Proceedings of 6th International Conference on Climbing and Walking Robots (CLAWAR)*, pp. 11-18, Catania, September, 2003, Italy

Authors:

Dr.-Ing. Dipl.-Math. Carsten Behn, Univ.-Prof. Dr.-Ing. habil. Klaus Zimmermann
TU Ilmenau, Faculty of Mechanical Engineering, Max-Planck-Ring 12 (Building F), 98693, Ilmenau
Phone: +49 3677 691813 / Fax: +49 3677 691823
E-mail: carsten.behn@tu-ilmenau.de

J. W. Vervoort / T. Kopfstedt

Mission Planning for UAV Swarms

Introduction

Unmanned Aerial Vehicles (UAVs) play an ever increasing role in a wide variety of scenarios in both the civilian and military sector, carrying out tasks like traffic surveillance, firefighting, or reconnaissance missions. Furthermore, the use of groups of vehicles, or swarms, has been shown to accomplish certain objectives more efficiently and more effectively than a single vehicle can, for example in terrain mapping or search missions. In all these scenarios, the autonomous vehicles need to fly on trajectories that match their flight envelope, are as short as possible, and most importantly, avoid collisions with obstacles and other UAVs at all cost.

In order to achieve this, Mixed-Integer Linear Programming (MILP) is used as the optimization principle. MILP extends regular linear programming to include variables that are constrained to integer or binary values. Thus, MILP offers the possibility to add logical and decision making constraints into the optimization, such as obstacle and collision avoidance.

However, finding long-range minimum-time trajectories in environments with many obstacles is a complex optimization problem. In this paper, a Model Predictive Control (MPC) approach is chosen to decrease computational complexity and limit computation time, therefore making the algorithm capable of real-time calculations as well as handling unknown or dynamically changing environments. The presented algorithm is capable of calculating near-optimal flight trajectories to ensure that the UAV swarm carries out its mission in the minimum time.

Problem Formulation

In general, a mission for a UAV swarm can be described as visiting a number of waypoints spread out on a map. All waypoints have to be visited once during the mission by one UAV. This so called Task Assignment constitutes the first part of the mission planning algorithm. During the Task Assignment phase, waypoints are assigned such that each UAV will have to travel only a minimum distance, therefore also bringing

the total mission time to a minimum. The problem is formulated as a multi-dimensional Traveling Salesman Problem (TSP), as implemented in [1].

In this paper, only planar motion is considered, meaning UAVs fly at constant altitude. The dynamics of the UAVs are modeled in the form of a simple point mass, bringing their discrete-time state space representation to:

$$\begin{bmatrix} x \\ y \\ v_x \\ v_y \end{bmatrix}_{k+1} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ v_x \\ v_y \end{bmatrix}_k + \begin{bmatrix} (\Delta t)^2 / 2 & 0 \\ 0 & (\Delta t)^2 / 2 \\ \Delta t & 0 \\ 0 & \Delta t \end{bmatrix} \cdot \begin{bmatrix} a_x \\ a_y \end{bmatrix}_k$$

UAV motion is constrained by a maximum velocity and turn rate, therefore giving us:

$$|v_x| \leq v_{\max}, |v_y| \leq v_{\max} \quad |a_x| \leq a_{\max}, |a_y| \leq a_{\max}$$

Calculating the length of a vector is a nonlinear operation. In order to adhere to the Linear Programming problem formulation, we approximate vector length by checking a number $n_{v_{\max}}$ of unit vectors onto which the velocity/acceleration vector is projected [5].

$$v^T \cdot i_k \leq v_{\max} \quad k=1, \dots, n_{v_{\max}} \quad a^T \cdot i_k \leq a_{\max} \quad k=1, \dots, n_{a_{\max}} \quad (1)$$

$$i_k = \begin{bmatrix} \cos\left(\frac{2\pi k}{n_{v_{\max}}}\right) \\ \sin\left(\frac{2\pi k}{n_{v_{\max}}}\right) \end{bmatrix} \quad i_k = \begin{bmatrix} \cos\left(\frac{2\pi k}{n_{a_{\max}}}\right) \\ \sin\left(\frac{2\pi k}{n_{a_{\max}}}\right) \end{bmatrix} \quad (2)$$

MILP also allows an efficient way to declare obstacles by using binary variables. Throughout this paper we limit ourselves to rectangular obstacles. Obstacles are described by their lower left and upper right corner $[x_{ll}, y_{ll}, x_{ur}, y_{ur}]^T$. The variable b_{object} is a binary, and the number M is an arbitrary large positive value.

$$\begin{aligned} x_{jk} &\leq (x_{ll}) + M \cdot b_{object_{ijk1}} \\ y_{jk} &\leq (x_{ll}) + M \cdot b_{object_{ijk2}} \\ x_{jk} &\geq (x_{ur}) - M \cdot b_{object_{ijk3}} \\ y_{jk} &\geq (x_{ur}) - M \cdot b_{object_{ijk4}} \end{aligned} \quad \begin{aligned} \sum_{l=1}^4 b_{object_{ijkl}} &\leq 3 \\ i &= 1 \dots N_O, \quad j = 1 \dots N_V \\ k &= 1 \dots T, \quad l = 1 \dots 4 \end{aligned} \quad (3)$$

$[x_{jk}, y_{jk}]^T$ is the position of UAV j at time k . Index i describes the number of obstacles, j lists the number of UAVs in the swarm, k the number of time steps in the planning horizon, and l enumerates the four inequalities. As long as $\sum b_{object_{ijkl}} \leq 3$ is fulfilled, the UAV is outside the obstacle.

Receding Horizon Controller

The constraints above can now be used to formulate a MILP optimization problem. Traditionally, trajectory optimization is done over a fixed horizon. This means that the complete trajectory is calculated from beginning to end point, forming a large and complicated optimization that does not take into account changing environments. When obstacles are added or subtracted, the precalculated optimal solution essentially becomes worthless and a recalculation has to be performed.

To solve this problem, a Model Predictive Control setup is chosen. [2] lists the properties of Model Predictive Control, also called Receding Horizon Control:

- At time i and initial state x_i the optimization is performed for only the next N_p time steps. N_p is the so called planning horizon.
- The first N_E values of the optimal solution are used as inputs to the system. N_E is the execution horizon. N_E is usually set to 1.
- In the new initial state x_{i+N_E} the optimization is performed again, repeating the process, thus taking into account possible changes in the environment.

Similar implementations of MPC can be found in [3] and [4].

To be able to use this approach, we need to overcome a problem. As seen in Fig. 1, the trajectory is only optimized within a small area around the current state, the planning horizon. But the total trajectory consists of the optimized piece plus the remaining pieces from the planning horizon to the goal point. However, that piece is unknown and the total distance cannot be calculated. Therefore, the remaining trajectory is approximated with straight line segments connecting the planning horizon to the goal point.

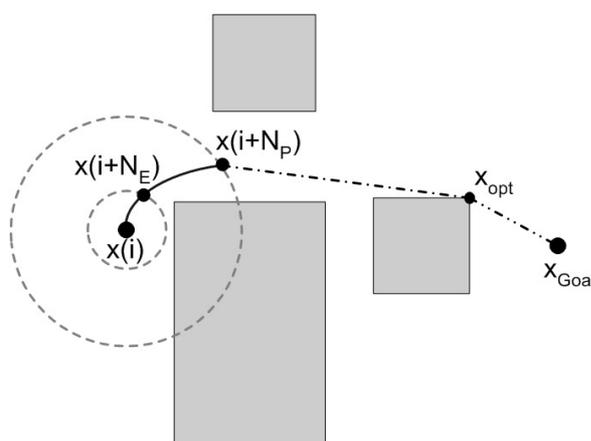


Fig. 1: Details of MPC

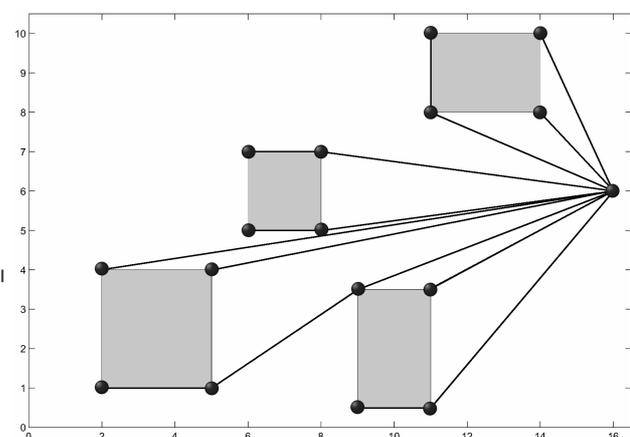


Fig. 2: Calculation of the cost map

In order to find the shortest distance to the goal point, a cost map of the complete environment is calculated, calculating and storing the distances of points to the goal point (Fig. 2). [6] shows that the shortest distance in an environment with convex obstacles is a combination of line segments between the points and the corners of the obstacles. Thus, the corners of the obstacles are cost points in our cost map. The visibility graph between all obstacles and the goal point is calculated and by using the Dijkstra algorithm, as outlined in [7], the shortest connections between all the points and the goal point are calculated and stored as the cost of each point.

The length of the trajectory outside the planning horizon can now be determined. As detailed in Fig. 1, it is:

- The distance from x_{i+N_p} , the edge of the planning horizon, to a known cost point x_{opt} that is visible from that point.
- The distance from that cost point to the goal point. This value has already been calculated and is stored in the cost map.

The visibility constraint between x_{i+N_p} and x_{opt} is very important. Because of it, local optimization within the planning horizon does indeed have a global influence on the whole trajectory, therefore also minimizing the total trajectory length.

Now the planning problem is expressed in MILP form. Each UAV can select only one cost point during each optimization:

$$\sum_{i=1}^{N_{CP}} \sum_{j=1}^{N_{goal}} b_{CP_{ijk}} = 1, \quad k = 1, \dots, N_V$$

with N_{CP} = number of cost points, N_V = number of UAVs, $N_{goal}=2$, the next two waypoints.

Equations (4) and (5) set the cost value and coordinates of the chosen cost point.

$$c_{opt_k} = \sum_{i=1}^{N_{CP}} \sum_{j=1}^{N_{goal}} c_i \cdot b_{CP_{ijk}}, \quad k = 1, \dots, N_V \quad (4)$$

$$\begin{bmatrix} x_{opt_k} \\ y_{opt_k} \end{bmatrix} = \sum_{i=1}^{N_{CP}} \sum_{j=1}^{N_{goal}} \begin{bmatrix} x_{CP_k} \\ y_{CP_k} \end{bmatrix} \cdot b_{CP_{ijk}}, \quad k = 1, \dots, N_V \quad (5)$$

The line connecting the edge of the planning horizon and the cost point is described by

$$\begin{bmatrix} x_{line_k} \\ y_{line_k} \end{bmatrix} = \begin{bmatrix} x_{opt_k} \\ y_{opt_k} \end{bmatrix} - \begin{bmatrix} (x_{i+N_p})_k \\ (y_{i+N_p})_k \end{bmatrix}$$

In order to test visibility, this connection is divided into N_{test} parts:

$$\begin{bmatrix} x_{test_{km}} \\ y_{test_{km}} \end{bmatrix} = \begin{bmatrix} x_{i+N_p} \\ y_{i+N_p} \end{bmatrix}_k + \frac{m}{N_{test}} \cdot \begin{bmatrix} x_{line_k} \\ y_{line_k} \end{bmatrix} \quad m = 1 \dots N_{test}$$

Each part is tested for interference with obstacles using a binary variable $b_{opt_{ikm}}$, very much like in (3).

Equations (6) and (7) handle which goal point to use in the optimization. If a goal point is reached in the last step of the planning horizon or not at all, then only this point will be part of the optimization. If a goal point is reached within the planning horizon, then goal points are switched and the optimization directs the trajectory to the next goal point.

$$\sum_{i=1}^{N_{CP}} b_{CP_{i1k}} = \sum_{i=T}^{T+1} b_{goal_{ij}} \quad (6)$$

$$\sum_{i=1}^{N_{CP}} b_{CP_{i2k}} = \sum_{i=1}^{T-1} b_{goal_{ij}} \quad (7)$$

Cost Function

The cost function to be minimized is the total trajectory lengths of the UAVs, consisting of three parts:

- The part within the planning horizon, from x_i to x_{i+N_p}
- The line between x_{i+N_p} and the selected cost point
- The distance between selected cost point and goal point.

The first part is described by the first part of Equation (8), which represents the number of time steps to the goal point times the distance traveled per time step.

$$Z = \sum_{j=1}^{N_V} \left(\left(v_{j_{max}} \cdot \Delta t \right) \cdot \sum_k^{T+1} k \cdot b_{goal_{kj}} + l_{line} + c_{opt_j} \right) \quad (8)$$

l_{line} is the distance between x_{i+N_p} and the selected cost point. Since it constitutes the length of a vector, it is calculated just as in Equations (1) and (2).

The third part is simply the stored value of the cost point, meaning the distance to the goal point.

Simulation

Two scenarios are presented that show the capabilities of the presented MCD algorithm, especially in changing environments. Fig. 3 shows a trajectory replanning after a UAV is lost during a mission.

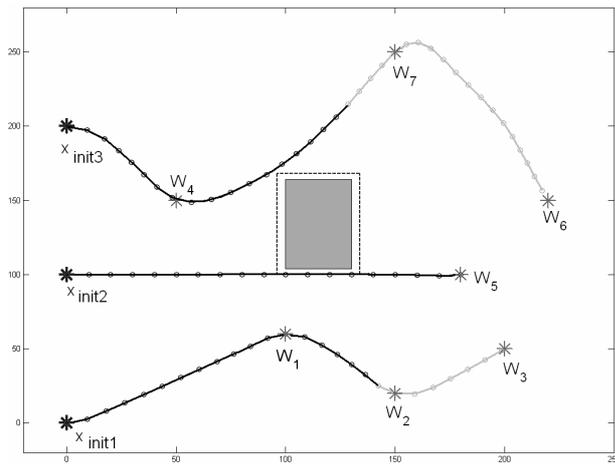


Fig. 3: Replanning after UAV loss

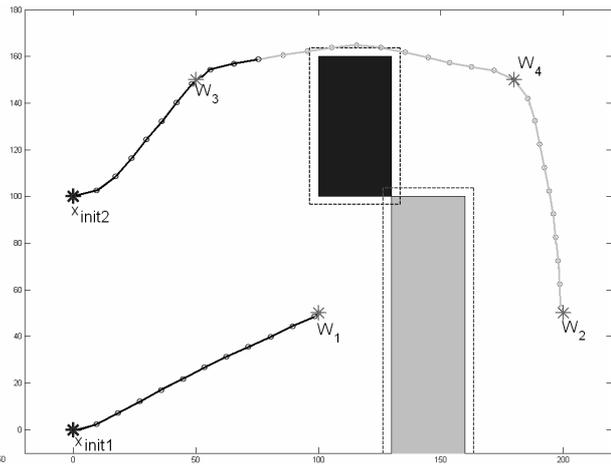


Fig. 4: Replanning due to unknown obstacles

The regularly planned trajectories are shown in black. UAV #2 is supposed to visit waypoints W_5 and W_6 . However, UAV #2 is lost at W_5 . Immediately, a replanning occurs and UAV #3 takes over W_6 . The newly planned trajectories are shown in gray.

In Fig. 4, UAV #1 is scheduled to visit W_1 and W_2 . However, it encounters an unknown obstacle, shown in gray, that was not part of the previous planning. After a replanning, it is determined that UAV #2 can reach W_2 faster and it takes over for #1.

Conclusion

The iterative MPC algorithm presented here has been shown to effectively handle various changes in the mission environment, decreasing computational complexity and still being able to calculate near-optimal trajectories for groups of cooperating UAVs.

References:

- [1] SCHUMACHER, Corey; CHANDLER, Phillip; PACTHER, Meir: *Constrained Optimization for UAV Task Assignment*. AIAA Guidance, Navigation, and Control Conference, Providence, RI, USA, 16-19 August 2004
- [2] MACIEJOWSKI, Jan M.: *Predictive Control with Constraints*. London: Prentice Hall, 2002
- [3] RICHARDS, Arthur; HOW, Jonathan P.: *Model Predictive Control of Vehicle Maneuvers with Guaranteed Completion Time and Robust Feasibility*. American Control Conference, Denver, CO, USA, 4-6 June 2003
- [4] FREW, Eric W.: *Receding Horizon Control Using Random Search for UAV Navigation with Passive, Non-cooperative Sensing*. AIAA Guidance, Navigation, and Control Conference, San Francisco, CA, USA, August 2005
- [5] RICHARDS, Arthur; HOW, Jonathan P.: *Mixed-integer Programming for Control*. American Control Conference, pp. 2676-2683, June 8-10, 2005
- [6] LATOMBE, J.-C.: *Robot Motion Planning*. Boston, MA, USA: Kluwer Academic Publishers, 1991, pp. 153-200
- [7] CORMEN, Thomas H.: *Introduction to Algorithms*. Cambridge, MA, USA: The MIT Press, 2001, pp. 595-601, 629-635

Authors:

Dipl.-Ing. Jan Willem Vervoort
 Dipl.-Ing. Thomas Kopfstedt
 Diehl BGT Defence GmbH & Co. KG, Alte Nußdorfer Str. 13
 D-88662 Überlingen, Germany
 Phone: +49 7551 89-2449
 Fax: +49 7551 89-2351
 E-mail: jan.vervoort@diehl-bgt-defence.de

M. Kaufmann / G. Bretthauer

Development and composition of control logic networks for distributed mechatronic systems in a heterogeneous architecture

INTRODUCTION

Information processing, plays an ever-increasing role in the functionality and quality of products in the field of mechanical engineering and becomes an inextricable part of the system function. However, program execution requires special hardware separated from the mechanical and electrical components. Typical hardware architectures in this domain are embedded systems (Micro controller, DSP board, micro processor, – with/without operating system), personal computer programmable logic controller (PLC) – and composed of several components: distributed systems with central supervisory control, distributed systems with distributed control, distributed autonomous cooperative systems. In many cases, a heterogeneous architecture consisting of diverse electronic units, software components, and communication interfaces is implemented (see Figure 1).

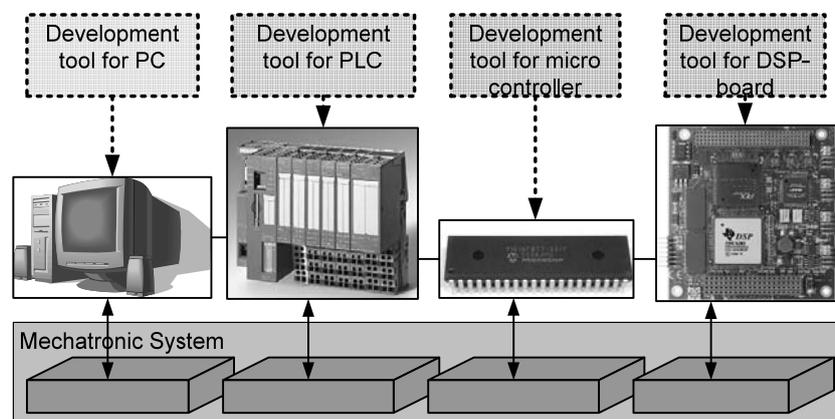


Figure 1: Mechatronic system with heterogeneous hardware and software architecture

Inside the different hardware units similar functional entities are executed – typically described by functional blocks. Among them are logical networks – consisting of logical operations, controllers – closed loop and open loop, various arithmetic operations, combinations of the aforementioned entities.

During the development of heterogeneous systems, different development tools using different standards, programming approaches and programming languages for basically similar functionalities are used. Developers make great efforts to verify functional consis-

tency of these systems, i.e. to test interoperation of entities developed by different tools. Especially for flexible systems of variable configuration procedures for test, verification and correction are passed through repeatedly. In this context functional consistency means: functional entities of the system cooperate in the desired manner and operate error-free. Depending on requirements, different aspects are considered and verified separately (see Table 1).

Table 1: Important aspects of functional consistency

Aspect	Remark
Data transfer between outputs and inputs of controllers and other function blocks,	Appropriate data connection, format and sequence
Hierarchical dependencies	Interoperation of several hierarchical layers
Failure management	Defined behavior in case of malfunction
Reliability, fault tolerance, FMEA [1]	Defined behavior to avoid damage in case of failure
Real time conditions, different cycle times	Compliance with real-time conditions
Concurrency, parallel operations	Appropriate sequence and interoperation of concurrent operations
Interlocking, enabling, priorities	Conditions and procedures for turn-on/turn-off
Different modes of operation	Switching between different operational states

Given the difficulties described above, an integration of all necessary tools into one comprehensive development suite would be desirable. Although comprehensive tools do exist for confined domains (see below), this approach is not practicable for heterogeneous systems in the foreseeable future. Hence, methods and tools for the assistance of design and verification of functional structures of information processing are necessary. The aim of this paper is to present an approach for a test-framework for functional consistency, applicable to a wide range of hardware and software, and to demonstrate the application to the development of an assembly machine [1].

TEST-FRAMEWORK FOR FUNCTIONAL CONSISTENCY

In computer science, the term data consistency model is used – among other meanings – to specify consistent access to data in a distributed system. In this regard, a number of possible data consistency models are known. A system supports a given consistency model if operations on memory follow a set of specific rules. The data consistency model specifies a contract between programmer and system, wherein if the programmer follows the rules, the system guarantees that memory will be consistent and the results of memory operations will be predictable. Accordingly, a functional consistency model (for a heterogeneous system) consists of a set of functional rules specifying conditions for connection and interoperation of functional entities of the system. If a functional structure – consisting of individual functional entities – follows the defined rules and is imple-

mented accordingly, the system is functioning failure-free and conflict-free.

In other domains where distributed hardware is used, comprehensive tools for integrated development have been implemented. For example, in process control engineering hardware and software architectures, functionalities, development tools etc. are standardized to a very large extent, see [2][3][4]. Open system standards are developed to further advance the flexibility and interchangeability of programs and hardware; see [5]. Due to this level of standardization, hardware manufacturers are able to provide integrated development tools covering the necessary range of verification. To cite another example, AUTOSAR (AUTomotive Open System Architecture) – developed by automobile manufacturers and suppliers – assists the development of networks of embedded systems in automobiles [6][7]. The AUTOSAR development framework supports the consistent distribution of functionality among several electronic control units.

The used consistency models in both domains – process control engineering and AUTOSAR – are tailored to the specific needs of supported hardware; implementation and programming are strongly supported. On the other hand, these consistency models are applicable to a narrow range of specialized hardware and software only and are mainly related to data transfer and interoperability of modules. An approach for a test-framework for functional consistency applicable to a wide range of hardware and software is presented in the following.

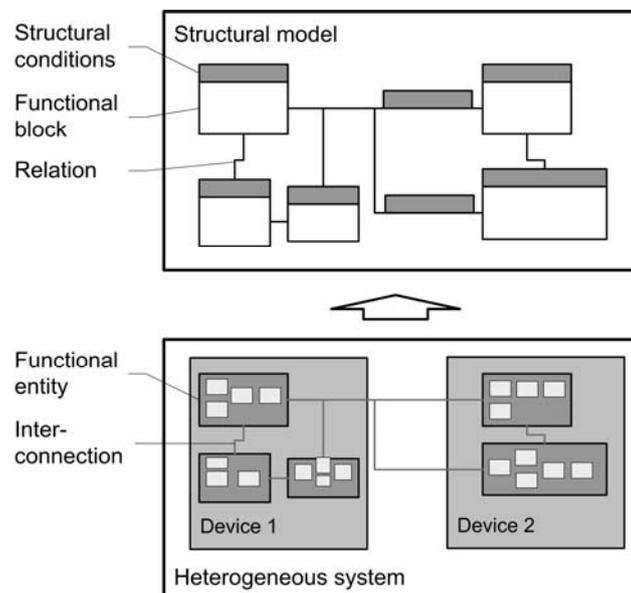


Figure 2: Structural Modeling

1. Structural Modeling: In the first step a structural model of the system is established for structural test purposes (see Figure 2). The structural model is composed of:

- Functional blocks representing the functional entities to be considered,

- Relations between functional blocks and
- Functional conditions.

Structural conditions are necessary conditions for interoperation between functional blocks to maintain functional consistency. Functional entities are – for example – controllers, pieces of control logic, arithmetic operations, data processing and supervisory control logic. The functional content of these system functions is *not* considered. Solely interfaces and relations between functional entities are taken into consideration – e.g. inputs and outputs, data transfer between entities, hierarchical relations, enabling and causal dependencies. Depending on the considered perspective, different information is incorporated. Functional blocks are connected by relations. Relations describe data transfer as well as the existence of functional dependencies between blocks. The type of dependency as well as interactions between functional blocks is described by functional conditions, e.g.: “B can be activated only if execution of A is completed”. A sample controller type block with hierarchical interconnections is given in Figure 3.

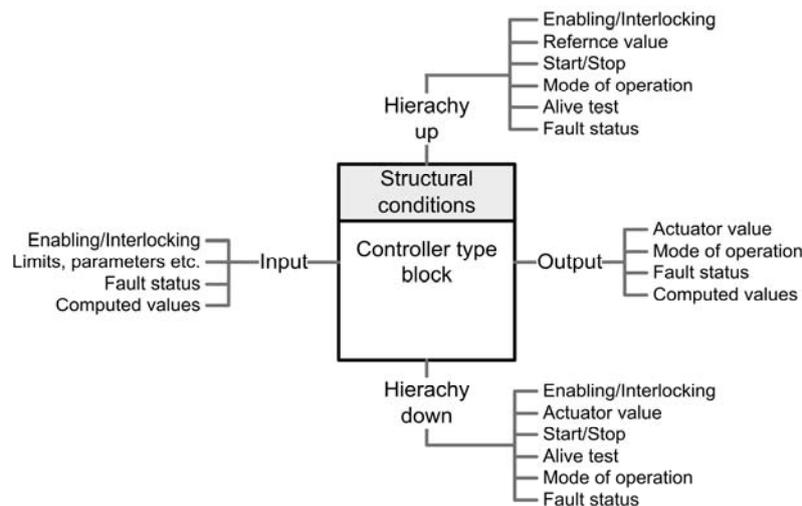


Figure 3: Sample controller type block with hierarchical interconnections

2. Consistency model and structural test: To define a consistency model, structural rule sets are defined for different aspects. Structural rules describe conditions and restrictions for consistency regarding a specific test aspect. For different test cases – e.g. failure management – several rule sets are arranged (see Figure 4). The structural rules are independent of the functional blocks, relations and functional conditions. The following rule exemplifies the possible content of structural rules: “A functional block A can be executed only if all functional blocks transferring inputs to A have been executed.”

A structural verification is accomplished by testing the structural model systematically for compliance with rules of one or more rule sets.

3. Application of results: As result of the structural verification inconsistencies are indi-

cated. In the third step the developer turns to system design again and eliminates the conflicts; results are incorporated in software coding. This procedure of test and re-design/re-programming is repeated until consistency is established.

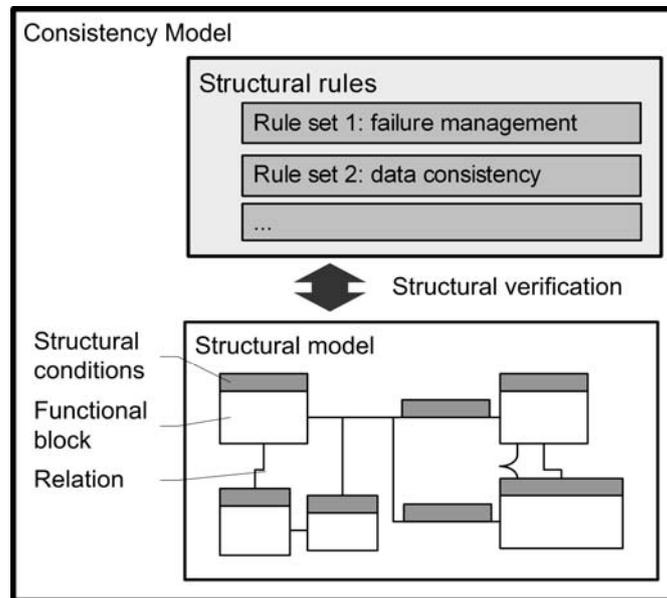


Figure 4: Consistency Model

SAMPLE APPLICATION

A section of the information processing of a servo driven assembly machine in symbolic form is given in Figure 5. It consists of equipment of heterogeneous architecture; functionality is finally coded using different tools. In the modeling of the functional structure, all functional units – e.g. controllers, interlocking logic, fault detection logic, data processing – are transformed to abstract standard blocks with specific interfaces, see Figure 6. In this process, only logical interconnections are considered, algorithms behind the blocks are not regarded. The resulting structure of connected standard blocks is used to analyze and optimize the interoperation of the blocks.

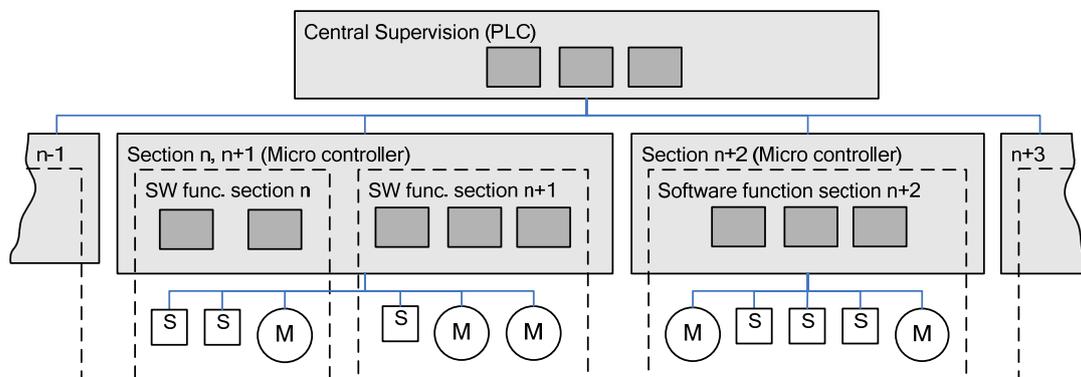


Figure 5: Servo driven assembly machine (M: Servo drive, S: Sensor)

This approach is used to eliminate fault conditions in the interaction between sections of the assembly machine [1]. The structural test based on the structural model reveals in-

consistencies – e.g. incorrect sequences, exceeded time limits. Functional entities can be changed accordingly. The automatic structural test eliminates the time-consuming, error-prone manual check of the structure. At present, a semi automatic test environment is established in order to develop modelling and test methods.

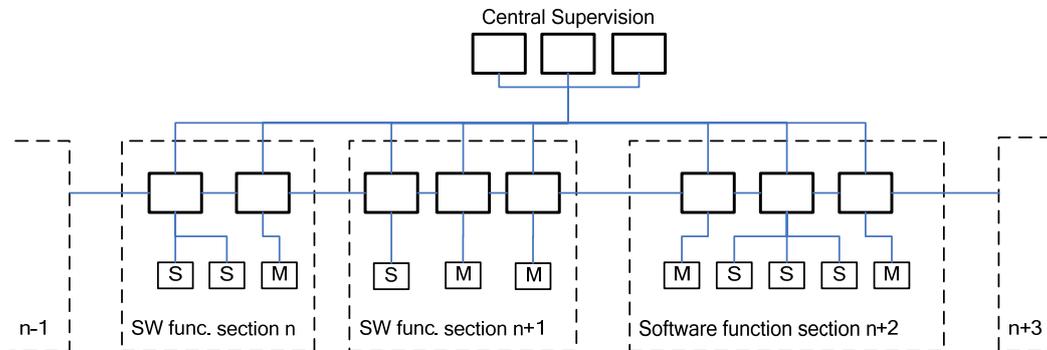


Figure 6: Model of the functional structure

CONCLUSION

Heterogeneous architectures of mechatronic systems require approaches for comprehensive structural tests during system design. Structural modeling of the system is used to create an automatic environment for device- and manufacturer-independent checking regarding different perspectives. This approach is used to check the failure management and start/stop-sequence/interlocking of an assembly machine composed of several assembly stations. Further work is necessary to extend the approach to more test perspectives and device combinations, to develop an uniform description method for functional blocks, relations, conditions and to develop a practicable software solution.

References:

- [1] Lorch, M.: Machbarkeitsanalyse eines neuen Antriebskonzeptes für einen Längstaktmontageautomaten mit hoher Taktzahl. Diplomarbeit, Universität Karlsruhe (TH), 2007.
- [2] DIN EN 61499-3. Funktionsbausteine für industrielle Leitsysteme, Teil 3: Programmiersprachen. Berlin: DIN Deutsches Institut für Normung e.V., 2006.
- [3] DIN EN 61131-1: Speicherprogrammierbare Steuerungen, Teil 1: Architektur. Berlin: DIN Deutsches Institut für Normung e.V., 2003.
- [4] DIN EN 61499-2. Funktionsbausteine für industrielle Leitsysteme, Teil 2: Anforderungen an Software-Werkzeuge. Berlin: DIN Deutsches Institut für Normung e.V., 2006.
- [5] PLCopen, Gorinchem, NL. <http://plcopen.org/>
- [6] Heinecke, H., Bielefeld, J., Schnelle, K.-P., Maldener, N., Fennel, H., Weis, O., Weber, Th., Ruh, J., Lundh, L., Sandén, T., Heitkämper, P., Rimkus, R., Leflour, J., Gilberg, A., Virnich, U., Voget, S., Nishikawa, K., Kajio, K., Scharnhorst, Th., Kunkel, B.: AUTOSAR – Current results and preparations for exploitation. 7th EUROFORUM Conference 'Software in the vehicle' 3-4 May 2006, Stuttgart, Germany.
- [7] AUTOSAR GbR, München, <http://www.autosar.org>

Authors:

Dr.-Ing. Michael Kaufmann
 Prof. Dr.-Ing. habil. Georg Bretthauer
 Universität Karlsruhe (TH), Institut für Angewandte Informatik/Automatisierungstechnik (AIA),
 Kaiserstr. 12, Bldg. 10.91, R. 203
 76131 Karlsruhe, Germany
 Phone: +49 (721) 608-7971
 Fax: +49 (721) 608-7972
 E-mail: [Michael.Kaufmann, Georg.Bretthauer]@aia.uni-karlsruhe.de

T. Kopfstedt / J. W. Vervoort

Formation Control for Groups of Mobile Robots Using a Hierarchical Controller Structure

1. INTRODUCTION

Nowadays most mobile robots are teleoperated or acting based on a preplanned mission plan as single individual robots. In the future more and more tasks will be done with robots and many of them require more autonomous functions of the robots and interaction between the robots. In future groups of robots will work together to fulfill missions such as humanitarian convoys, border patrol and search and rescue independently and reduce the risk for humans. In many of these missions a permanent communication between all members of the formation cannot be secured and is sometimes also not wanted, because intensive communication can be easily jammed or used for targeting on the formation in total or on single members of the formation of mobile robots. Accordingly a concept is presented which makes it possible for a group of mobile robots to fulfill a preplanned mission with limited inter-robot communication even in cases of unplanned changes of the trajectory for the group of mobile robots.

For such missions it is in most cases possible to generate a preplanned mission plan based on satellite maps, UAV data and digital maps. In this paper the mission plan for the group of mobile robots will be generated by Mixed Integer Programming as shown in [1] and [3]. This guarantees preplanned optimal trajectories for all members of the group of mobile robots. This information will be loaded into each robot a priori to the mission. During the mission the robots will follow this preplanned mission plan and, by the robots own onboard sensors, every robot will permanently measure the distance to the neighbors in the formation, to control and recalibrate the own position in the formation. The use of this concept enables a group of mobile robots to follow a preplanned optimal trajectory as a formation in a specified formation style without any communication if necessary.

In cases one robot of the formation of mobile robots detects an unknown obstacle it will activate its onboard fuzzy controller to avoid a possible collision with this obstacle. The other robots will detect the anomaly in the movement of this robot and automatically

react using their fuzzy controllers to try to keep in the predefined formation and, if this is impossible, to change the formation as long as necessary to surround the unknown obstacle and to return to the preplanned mission plan after that.

The following chapters will show the formulation of the MIP problem, the control structure, the design of the fuzzy controllers and simulation results for cases with a priori known and unknown obstacles.

2. HIERARCHICAL CONTROL STRUCTURE

The controller structure for the formation of mobile robots in this paper is divided into four control loops. These control loops represent different degrees of the level of control as shown in Fig. 1. In the Level 1 the complete mission is planned for each robot by several trajectory points which are connected by lines. Underneath this the fuzzy controller on each robot is implemented which is used during the movement of the robots along the trajectories to secure the formation stability and to avoid collisions with a priori unknown obstacles. The fuzzy controllers are also able to change the trajectories for the robots for passing around obstacles in a restricted area, but if greater changes should become necessary this must be done by the Level 1 using MIQP.

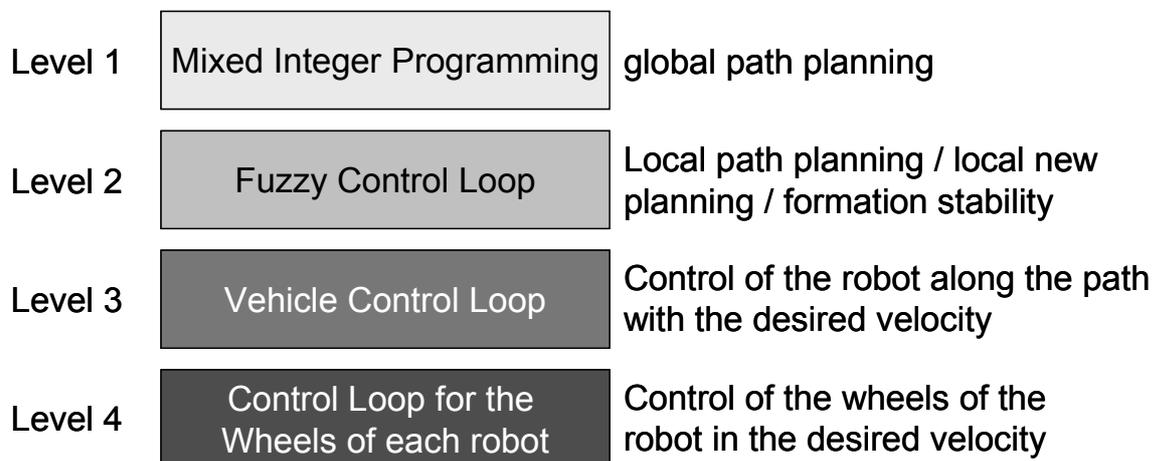


Figure 1: Hierarchical control structure

Beneath the fuzzy control loop each robot has a movement controller (Level 3) which generates the necessary command velocities for the wheels of the robot. As each wheel of the robot has is independently actuated the lowest level of the controller architecture is represented by the motor controller (Level 4) for each wheel of the robot.

3. MIXED INTEGER PROGRAMMING

The equations for the MIP problem which must be solved to generate the optimal trajectories for the formation of mobile robots can be described using [1], and [2]. The dynamic model for a mobile robot can be described according to [1] as

$$(1) \quad \frac{d}{dt} \begin{bmatrix} x \\ y \\ \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -gc_2 - \frac{\mu}{m} & 0 \\ 0 & 0 & 0 & -gc_2 - \frac{\mu}{m} \end{bmatrix} \begin{bmatrix} x \\ y \\ \dot{x} \\ \dot{y} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ \frac{1}{m} & 0 \\ 0 & \frac{1}{m} \end{bmatrix} \begin{bmatrix} f_x \\ f_y \end{bmatrix}$$

where m is the mass of the robot, f_x and f_y are the forces from the engine of the robot and $gc_2 + \mu$ are constants to describe the friction effect for rolling over ground and the mechanical parts around the engine in the robot as explained in [1]. The connection between the robots j and $j+1$ can be described according to [3] as

$$(2) \quad \begin{aligned} x_{t,j+1} - x_{t,j} &= x_{dist,j} \\ y_{t,j+1} - y_{t,j} &= y_{dist,j} \end{aligned}$$

in which $x_{dist,j}$ and $y_{dist,j}$ are static values for the description of the distance between two robots. This must be done for all connections between the robots in the formation to describe a complete formation. As shown in [1] alternative algorithms with more complex structures will also allow the description of variable formations and formation switching. In addition to the robot description and the description of the links between the robots also the obstacles must be described in a way as shown in [3] by using the system of equations:

$$(3) \quad \begin{aligned} x_{t,j} - S \cdot \varepsilon_1 &\leq x_{\min} \\ -x_{t,j} - S \cdot \varepsilon_2 &\leq -x_{\max} \\ y_{t,j} - S \cdot \varepsilon_3 &\leq y_{\min} \\ -y_{t,j} - S \cdot \varepsilon_4 &\leq -y_{\max} \\ \sum_{i=1}^4 \varepsilon_i &\leq 3 \end{aligned}$$

with S as a large positive number, the lower left edge x_{\min}, y_{\min} and the upper right edge x_{\max}, y_{\max} of the obstacle and $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$ as Boolean variables.

Further methods to describe of more complex obstacles which partially have already been tested with the algorithms shown in this paper, are described in [4] and [5].

The optimization problem itself can be described by the minimization criterion

$$(4) \quad \min_{\Delta s} \sum_{t=1}^{T-1} \sum_{j=1}^M \Delta s_{t,j}$$

Subject to

$$\left[\begin{array}{c} \text{MLD system representation of (1)} \\ (2) \\ (3) \\ s_{T,j} = s_{goal,j} \end{array} \right]$$

where $\Delta s_{t,j} = s_{t,j} - s_{t-1,j}$ is the difference in the position $s_{t,j} = [x_{t,j}, y_{t,j}]^T$ between two steps of a robot, and $s_{T,j} = s_{goal,j}$ is the reaching goal criteria to ensure that in the last step of the trajectory each robot reaches its own target point. This minimization problem can be solved by any MIQP solver and the result is the optimal trajectory through the scenario for the formation of mobile robots.

4. FUZZY CONTROL

The architecture of the fuzzy controller is an extension of the work from [4] and is based on [6] for the obstacle avoidance. As described in the hierarchical control structure part of this paper the fuzzy controller is permanently controlling the distances to the other robots to ensure that all robots keep in formation.

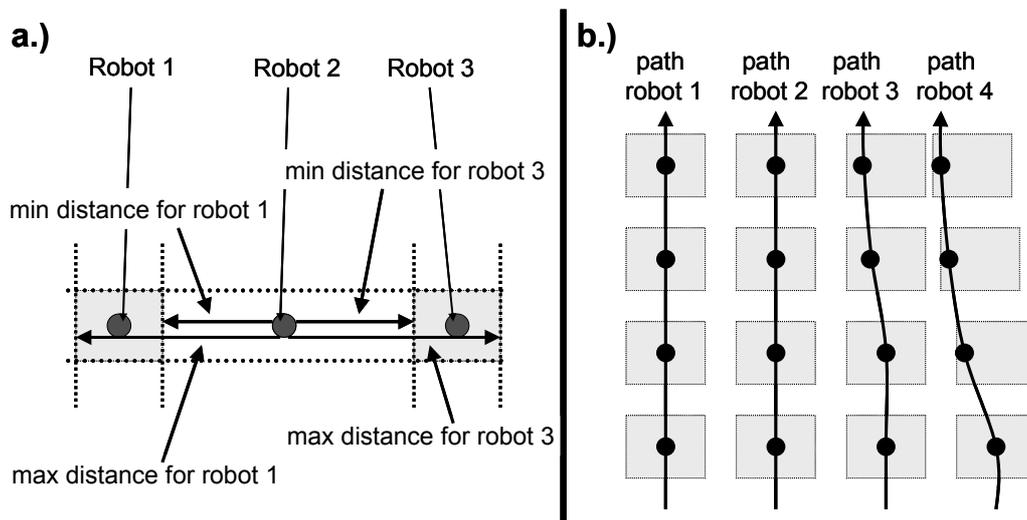


Figure 2: a) allowed robot positions b) collision avoidance

Therefore, as shown in Fig. 2a, each robot only controls the distances to its neighbors. If one of the robots has to change its position in a way that the formation stability is no longer guaranteed the neighboring robots will also change their positions as long as necessary to regain a stable formation. This is shown in Fig. 2b.

5. SIMULATION

Based on the information from Fig. 3a the trajectories in Fig. 3b are generated using MIQP for the formation of mobile robots. As all obstacles have been known during the planning phase of the mission (Level 1 in the hierarchical control structure) the solution for each robot is the optimal path through the scenario.

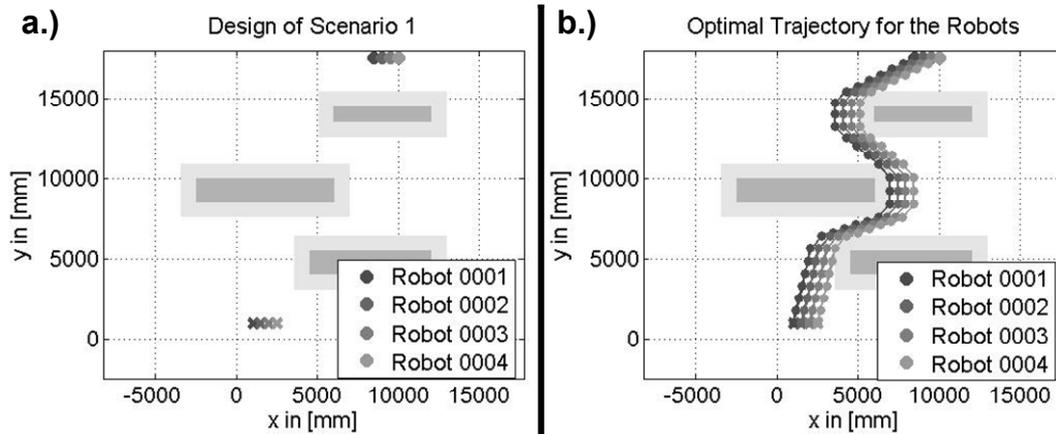


Figure 3: Scenario 1, all obstacles are known

In difference to the scenario shown in Fig. 3 the scenario in Fig. 4a is not known in total during the planning phase of the mission. The obstacle marked “unknown” is unknown during the planning phase of the mission. The result is that the trajectories from the MIQP are going through the unknown obstacle as shown in Fig. 4b. During the mission this obstacle will be detected by the sensors of one of the robots from the formation of mobile robots. At this moment the Fuzzy Controller of the robot No. four starts to change the trajectory for this robot. This is detected by robot No. three and this robot also changes its own trajectory a bit. The robots No. one and No. two are able to follow the preplanned trajectory.

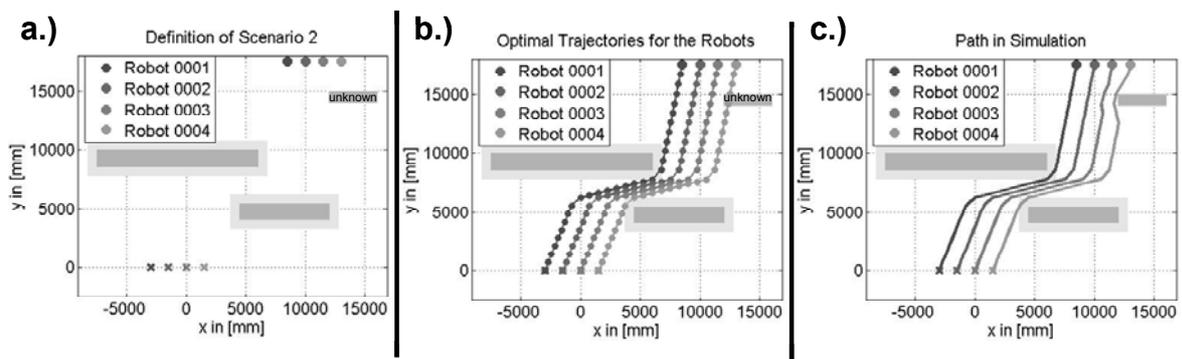


Figure 4: Scenario 2, working with unknown obstacles

Behind the unknown obstacle the robots No. three and No. four return to their preplanned trajectories and the mission continues using the preplanned trajectories for all robots. If the robots would have been unable to return to their trajectories or if all robots would have moved to a big distance from the preplanned trajectories a new planning of the trajectories for all robot using the MIQP would have taken place. If such new planning becomes necessary the robots have to communicate with each other to update their scenario information and resulting trajectories from the MIQP. In the other cases, as it is shown here in Fig. 4c, the robots do not need to communicate with each other during the mission as all necessary information about the other robots of the formation can be gathered by the onboard sensors of each robot.

6. CONCLUSION

As shown in Fig. 3 and Fig. 4 the hierarchical controller structure which is presented in this paper works with formations of four robots for scenarios in which all information is a priori known as well as for such scenarios in which some obstacles are unknown during the planning phase of the mission or in which the scenario changes during the mission. If all data are known before the mission starts the optimal path is found. In the other case only one possible path could be found due to the limited information available.

References:

- [1] T. Kopfstedt, M. Mukai, M. Fujita, O. Sawodny, "A Networked Formation Control for Groups of Mobile Robots using Mixed Integer Programming," IEEE Conference on Control Applications, pp. 579-584, Oct. 4-6, 2006
- [2] A. Richards, J.P. How, "Mixed-integer Programming for Control," American Control Conference, pp. 2676-2683, June 8-10, 2005
- [3] M. Mukai, T. Azuma, M. Fujita, "A Collision Avoidance Control for Multi-Vehicle Using PWA/MDL Hybrid System Representation," IEEE Conference on Control Applications, pp. 872-877, Sept. 2-4, 2004
- [4] Kopfstedt, M. Mukai, M. Fujita, O. Sawodny, "Formation Control for Mobile Robots in partially known Environments using Mixed Integer Programming and Fuzzy Systems," SICE-ICASE International Joint Conference, pp.1832-1837, Oct. 18-21, 2006
- [5] T. Schouwenaars, "Safe Trajectory Planning of Autonomous Vehicles," PhD Thesis at the Massachusetts Institute of Technology, Chapter 2.3, Feb. 2006
- [6] W. Li, "Fuzzy-Logic-Based Reactive Behaviour Control of an Autonomous Mobile System in Unknown Environments," Engng. Applic. Artif. Intell., Vol. 7, No. 5, pp. 521-531, 1994

Authors:

Dipl.-Ing. Thomas Kopfstedt
Dipl.-Ing. Jan Willem Vervoort
Diehl BGT Defence GmbH & Co. KG
Alte Nußdorfer Str. 13, 88662 Überlingen, Germany
Phone: +49 7551 89-2875
Fax: +49 7551 89-2351
E-mail: Thomas.Kopfstedt@diehl-bgt-defence.de

M. Abel / Th. Lohfelder

Simulation of the Communication Behavior of the German Toll System

Introduction

In Germany, toll is collected on all motorways and some selected national roads by the company *Toll Collect* [1] for vehicles heavier than 12t. In the following we will use “truck” for all toll-relevant vehicles. A truck driver can either use the automatic system, or book his journey manually before entering the toll-roads. Roughly, the automatic system consists of a central part, the computer centre and the OBUs as mobile components. About 540,000 trucks are equipped with this innovative device which determines its position by using Satellite technique (GNSS), and autonomously calculates the toll to be paid dependent on the distance travelled. The toll information is sent by mobile communication to the computer centre, and the logistics enterprise is billed for it. Taking into account the intricate data processing and spatio-temporal interdependencies of the sub-units of the computer centre, the system is among the most complex communication networks in the world. In this paper, we present simulations concerning the part of the automatic system occupied with remote device management: the update of the On-Board Units (OBU s) by mobile communication (GSM).

The key technology for the positioning is the Global Navigation Satellite System (GNSS), similar to the systems for personal use in cars. Of course, higher standards apply for toll collection with respect to availability, accuracy and reliability. Additionally, the OBU has to keep account about the distances driven, the tariffs in use and vehicle related information, like emission class, and number of axles. When an OBU drives on a toll-road it accounts for the toll and stores the amount to be paid encrypted on a smart card, similar to credit cards. When a certain limit is reached the amount is communicated by GSM services to the computer centre; from there the user receives a bill on a monthly basis.

Apart from money transfer, up-to-date information about new geo data (e.g. roads),

new tariff data and software (the “components”) has to be provided for and distributed to the OBUs. This distribution is highly nontrivial: the OBU is conceived in such a way that its many functions can be fulfilled autonomously, especially with respect to the process of updating information: it asks the centre, if any updated information is available. The centre communicates the state of the OBU data and the data present at the centre, this includes a timestamp after which the OBU information expires. Based on this procedure, the key requirements for a smooth update process are formulated: All OBUs should ask within the validity of the stored data for new information to ensure that no toll is lost due to missing geo or tariff data. Secondly, the OBUs should not ask all at the same time, otherwise mobile stations and central units would be unnecessarily overloaded. Thirdly, the distribution of the different components shall be independent from each other. To fulfill these requirements, Toll Collect has developed an intelligent distribution and monitoring system.

To obtain detailed predictions on the expected behavior of all registered OBUs, a simulation is of immense help. There are several advantages of the simulation approach: one can easily scale the system up or down in number, study worst-case scenarios, and finally one has the possibility to control efficiently the driving mechanism for the simulation, namely the activity of the trucks on the roads. This is a crucial factor, because a truck which is switched off most of the times obviously cannot communicate to the computer centre and thus needs special treatment for updates.

The modeling of the activities results from data mining; we use the unique data base of Toll Collect, where statistical analyses are run [2]. For our considerations, it is sufficient to model the temporal activity of the OBUs, i.e., on/off behavior, we do not need the spatial information on where the trucks drive.

Simulations have been carried out using MLDesigner [3], a very flexible next-generation system level design tool that combines the capabilities to model and analyze the architecture, function and performance of complex high-level systems. Different levels of abstraction can be accessed with most-common modelling domains like Discrete Event, Finite State Machine and Dataflow.

In the following section we describe some details of the update process, as used in the simulation. We explain our simulation strategy and show some graphs concerning programming and the final results. In the last section we conclude with a short discussion and outlook.

Simulation

The basic modules of the simulation are Driving, OBU, and Computer Centre, cf. Fig. 1. In these modules, the complex logic of the OBU and the computer centre is modeled and communication processes are worked out as detailed as needed. For the communication OBU-centre, the GSM service is used for a TCP/IP secure connection there are elaborate cryptographic algorithms in use which ensure that the data cannot be corrupted. The centre provides a certain number of ports which, of course, should be used in an optimal way. For toll collection all active OBUs must receive their update until validity expires, another task is the determination of the load on the ports. For the first, it is clear that a too short window for the distribution of updates yields a loss of the OBUS which have not been active in that window. On the other hand, geo-data must be up-to-date, such that distribution takes typically place within short time intervals. An obvious question for optimization concerns the number of ports necessary to communicate updated information to the OBUs reliably within a given time interval. Such questions are studied by means of our simulation.

.

System: Test System 5 [file:\$MLD_USER/Update/Test_System_5/Test_System_5.mml]

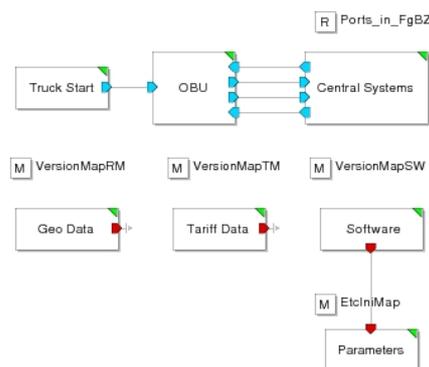


Figure 1: Schematic picture of the simulation with MLD Designer. The lower Modules provide the data relevant for the timing of the update process and other system parameters.

Results

To simulate the system, we create for each OBU a data structure containing the individual state of the OBU. The activity of the OBU is modeled by a hierarchical stochastic driver such that the important statistical properties, like number of daily journeys and recurrence rates are reproduced. Since our programming was optimized for storage (and speed of course), we can run easily up to 700,000 trucks on a computer with 1 GB RAM, to be compared with the current number of OBUs of ca. 600,000. A typical simulation yields the number of OBUs with updated information for each component. For this publication we simulated 50,000 OBUs with stochastic activity over 200,000 min, i.e., ca. 140 days. This is sufficient because traffic systems are scaling, i.e., apart from statistical outliers the main properties can be represented by a fraction of the real system [4]. The activity curve is shown in Fig. 2. One recognizes the weekly activity with gaps on Saturday and Sunday.

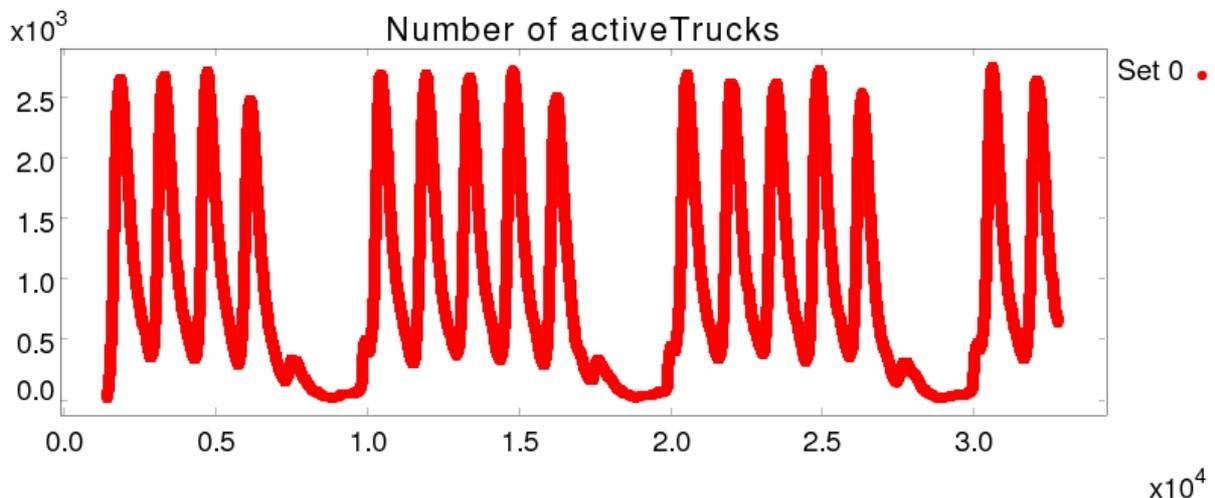


Figure 2: Activity output of our stochastic generator. The values correspond well to the activity observed by the Toll Collect data base. Saturday and Sunday evening activity (after the driving ban on German motorways) is seen by small peaks.

The time is given in minutes, the first 20,000 minutes (ca. 14 d) are transient. From 120,000 min to 167520 min (83 to 136 d) an update for geo and tariff data must be downloaded by the OBUs, software version 1 must be downloaded between 50,000

and 100,000 min, (35 to 69 d) another version from 120,000 to 180,000 min (83 to 125 d). In Figs. 3 and 4 the results are displayed. We see that the statistical distribution algorithm fulfills the requirements almost perfectly. In reality typical update intervals lie between 15 and 30 days.

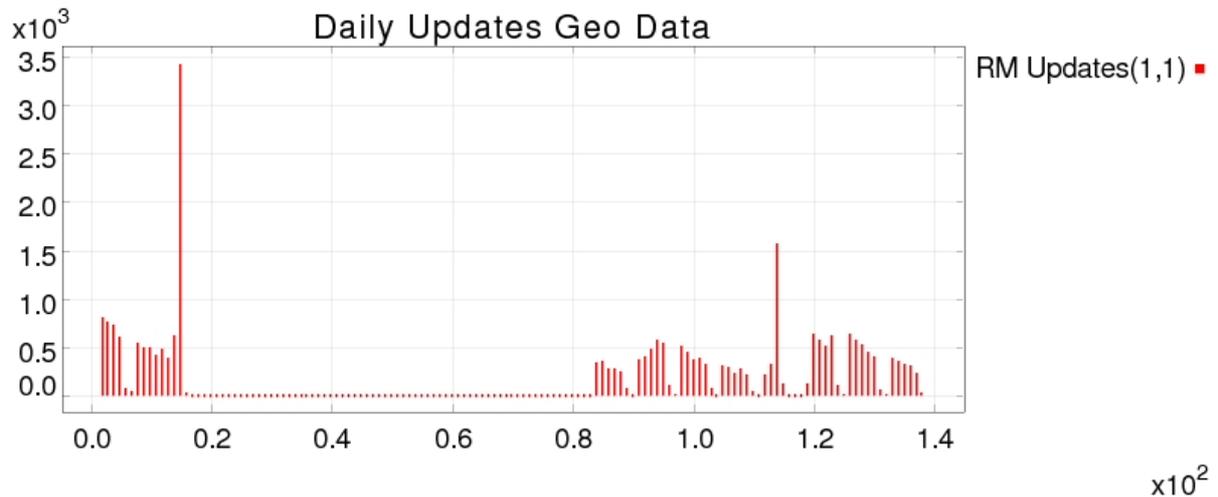


Figure 3: Update for the geo data, the x-axis unit is days. After a transient, the distribution of geo data is in the interval 83 to 136 d. Built-in, a security mechanism guarantees that a truck which otherwise might not obtain its update requests data briefly before validity of current data expires, this is seen as a peak. Fluctuations occur due to the irregular journeys of the trucks, as in reality. The tariff data are run with identical parameters, so the curve is almost identical and not shown here.

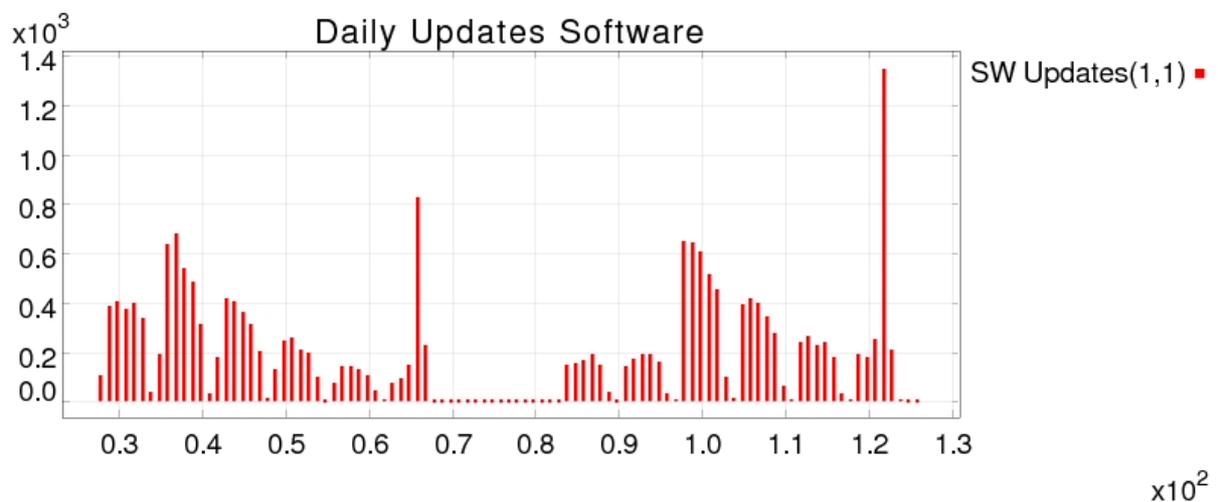


Figure 4: Update of software, x-axis units are in days. The distribution takes place over two large intervals from 35 to 69 days and from 83 to 125 days. Fluctuations occur due to the irregular activity of trucks.

Summary and Conclusions

We have demonstrated the great possibilities for the use of numerical simulation for telematic services in general and toll related tasks in particular. The tool – MLDesigner is a highest-level programming tool with strong object orientation and graphical support for modeling complex systems. The model displays the capabilities of remote device management in the Toll Collect fleet. Even though no detailed parameter studies are shown here to keep the paper short, the reader should have obtained an impression on the technological possibilities which result from the combination of high-level system design for communication and modern programming strategies.

The model as shown here is integrated in the routinely planning of update and information distribution, or the communication between central parts and mobile parts of the system, respectively. All procedures are highly optimized for toll collection. Of course, in full generality they are applicable to any system which requires remote device management. We think that the algorithms used and optimized by our work to be interesting for future applications involving communication between a central unit and mobile agents of arbitrary kind.

[1] www.toll-collect.de

[2] Statistics of the Truck Activity on German Motorways. K. Lüttkopf, M. Abel, and B. Eylert. This conference proceedings

[3] www.mldesigner.de

[4] M. Balmer, K. Nagel and B. Raney (2004) Large scale multi-agent simulations for transportation applications, Journal of Intelligent Transport Systems, 8 (4) 205-221.

Authors:

Dr. Markus Abel,
Toll Collect GmbH,
Linkstr. 4,
10875 Berlin
Phone: +49 (0)30 74077 4913
Fax: +49 (0)30 74077 5554
E-mail: markus.abel@toll-collect.de

Thomas Lohfelder,
Mission Level Design GmbH,
Ehrenbergstrasse 11,
D-98693 Ilmenau
Phone : +49 3677 4625-46
Fax : +49 3677 4625-11
Email: lohfelder@mldesigner.com

P. Hilgers / Ch. Ament

Control in Digital Sensor-Actuator-Networks

ABSTRACT

Due to recent advancements in technology powerful yet small embedded systems can be designed. The usage of wireless communication allows for the set-up of flexible and robust control systems. A model-based design method for distributed control systems without the need of central or fusion nodes is presented. As a means for an appropriate distribution of sensors and actuators and their communication the Gramian matrices can be used. Their properties lead to a data fusion which implements inherent system properties. The combination of these techniques results in a robust communication scheme which uses as little as possible energy for the wireless transceivers but still works within well-defined parameters.

Keywords: sensor-actuator-networks, event-discrete transmission, Gramian matrices

INTRODUCTION

It is generally accepted that the usage of sensor-actuator-networks for control purposes has several advantages if compared to classic centralised control. Many authors refer to this fact, see [1,2,3,4,5,6] to name just a few. This architecture is more flexible when a control system is set up and allows an easy reconfiguration as well as adding or removing network nodes. When using a wireless network the cost for the installation lies well below a wired solution. The recent trends in research and commercial products yield in several logical conclusions: Computing power is available at very little cost in very little space as the advancements in microelectronics go on and on. Small memory chips allow storing amounts of data which was impossible only a few years ago. In the mean time many different wireless communications standards have been developed and make it possible to choose a suitable solution for many applications. Thus, sensor-actuator-networks which are flexible, comparably cheap, and easy to maintain can be designed.

They incorporate the ideas of embedded systems and lately embedded microsystems, as well, which do not rely on electronics only but also the technologies of the MEMS area.

A sensor-actuator-network with embedded microsystems at its network-nodes shall be used to set up a decentralised and robust control system. First, we show how an event-discrete state transmission can be established with multiple state observers. Next, the observability and controllability of systems and their parts using Gramian matrices will be discussed. After that these two parts will be combined to set up an useful distributed estimation scheme which will be completed to a control system.

SYSTEM DESCRIPTION

Consider the following system

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k) + \mathbf{G}\mathbf{v}(k) \\ \mathbf{z}(k) &= \mathbf{C}\mathbf{x}(k) + \mathbf{w}(k) \end{aligned} \tag{1}$$

which is defined as usual and is assumed to be both observable and controllable.

This central structure shall now be distributed over the whole wireless network in the following manner:

- Each node includes a complete state model.
- Each *sensor* node has got an estimator.
- Each *actuator* node has got its according controller.

The wireless communication has several drawbacks which need to be compensated. Wireless communication is not always guaranteed, is delayed, and uses much more energy when compared to the power consumption of a microcontroller. That leads to the goal of controlling the system without the need to transmit data at every time step. The usage of state models is the main key of the proposed control method which will also counteract the discontinuous nature of this communications means. If the models are all synchronised to the real state of the system or are at least very close to the actual state each actuator is able to operate as needed since its controller is fed by the correct system state. The sensors, however, will be able to transmit their data only if the difference between the estimated state and the measured one is bigger than a certain threshold thus saving energy of unnecessary transmissions.

RESIDUAL-BASED EVENT-GENERATION

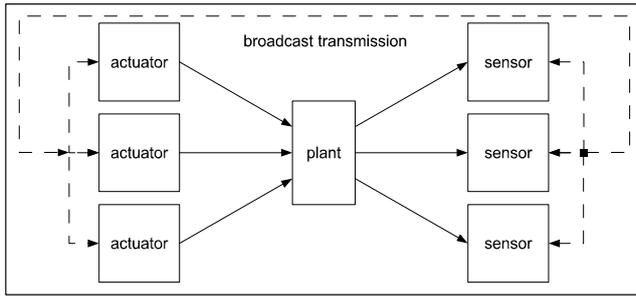


Fig. 1: Schematic description of information flow between sensor and actuator nodes in a distributed control network. Solid lines represent physical connections between actuators, plant, and sensors. Each node is able to transmit data via broadcast to all other nodes which is represented with dashed lines.

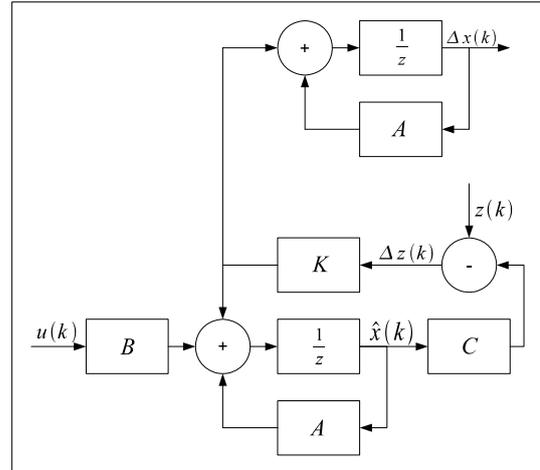


Fig. 2: Simplified Kalman filter with output for comparison with uncorrected model which runs in each sensor node.

Figure 1 shows in principle the connection scheme of a distributed control system. In the centralised case the physical sensors gather their measurements, which are then processed by a Kalman filter. Its state estimate is then used by a controller to give input signals to the actuators which influence the plant. In the simple way of a distributed control network this communication flow is established, as well. Several sensor nodes with their local Kalman filters, which inherently include a state model, gather information from the plant and process them. This is shown in figure 2. The lower part is essentially an uncorrected state model which runs equations (1). \mathbf{K} is the Kalman gain which is computed offline or online according to the standard formula (see [8]). The correction of the model is an additive term: $\mathbf{K}(\mathbf{z}(k) - \mathbf{C}\hat{\mathbf{x}}(k)) = \mathbf{K}\Delta\mathbf{z}(k)$. This yields

$$\hat{\mathbf{x}}(k+1) = \mathbf{A}\hat{\mathbf{x}}(k) + \mathbf{K}(\mathbf{z}(k) - \mathbf{C}\hat{\mathbf{x}}(k)) + \mathbf{B}\mathbf{u}(k) + \mathbf{G}\mathbf{v}(k) = \mathbf{A}\hat{\mathbf{x}}(k) + \mathbf{K}\Delta\mathbf{z}(k) + \mathbf{B}\mathbf{u}(k) + \mathbf{G}\mathbf{v}(k) \quad (4)$$

as state propagation. To account for the difference between the state vector in equations

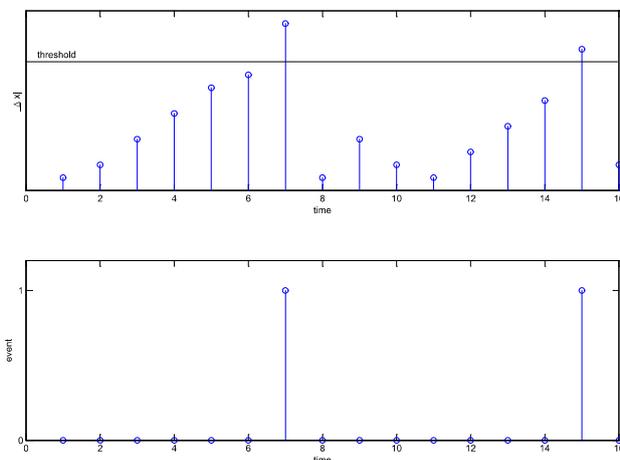


Fig. 3: Generation of transmission events.

(1) and (4) one has to propagate the correction with the state matrix. This means at each time the difference between an uncorrected model and the Kalman filter estimate is (upper part in figure figure 2)

$$\Delta\mathbf{x}(k+1) = \mathbf{A}\Delta\mathbf{x}(k) + \mathbf{K}\Delta\mathbf{z}(k) + \mathbf{G}\mathbf{v}(k) \quad (5)$$

At each time instant the biggest absolute value of the entries of the difference $\Delta\mathbf{x}$ is computed and compared with a threshold.

If the threshold is crossed $\Delta\mathbf{x}$ will be set

to zero and an event will be generated (see figure 3). This event starts the transmission of the state in order to correct all other node models. See [1,7] for further references.

GRAMIAN MATRICES

For a detailed description of the controllability and observability of distributed systems the Gramian matrices can be used. The controllability Gramian \mathbf{W}_c and the observability Gramian \mathbf{W}_o are the symmetric, non negative definite matrices which satisfy the following Lyapunov equations, respectively ([9]):

$$\begin{aligned} \mathbf{A}\mathbf{W}_c + \mathbf{W}_c\mathbf{A}^T + \mathbf{B}\mathbf{B}^T &= 0 \\ \mathbf{A}^T\mathbf{W}_o + \mathbf{W}_o\mathbf{A} + \mathbf{C}^T\mathbf{C} &= 0 \end{aligned} \tag{6}$$

A full rank n of \mathbf{W}_c or \mathbf{W}_o shows a complete controllability or observability, respectively. Otherwise it indicates the controllable or observable subspace. Furthermore, the entries in the matrices indicate how easy or hard it is to control or observe a state. In [9] several properties of the Gramians are discussed. In short these are: The Gramians depend on the state space realisation. However, their product $\mathbf{W}_c \cdot \mathbf{W}_o$ does not and can be used for further insight into the system. The Gramians can also be computed with reduced input or output matrices. That is, for single columns j of \mathbf{B} or lines i of \mathbf{C} a corresponding \mathbf{W}_{c_j} and \mathbf{W}_{o_i} can be computed. The sum over all these matrices results in the matrices of the full system.

EVENT-DISCRETE STATE TRANSMISSION AND DATA FUSION

The above mentioned technique of an event-discrete state transmission will now be combined with Gramian data fusion. The different observability properties of different sensor nodes will be taken to fuse all data correctly. When the local threshold of sensor i is crossed two things will be done: $\Delta \mathbf{x}_i(k)$ will be set to zero so that it can be computed correctly in the next time step and the event for state transmission will be generated. This event is transmitted to all actuators and triggers the data fusion.

Data fusion can be done in many different ways. In this work we propose the usage of Gramians as a means and show that this leads to better results than simply taking the average over all measurements. The advantage of this new approach is to exploit the properties of the Gramians. Each sensor has got its respective observability Gramian

W_{O_i} . The transmitted state is multiplied with it so that the data arriving at the actuators is $W_{O_i} \cdot \hat{x}$.

The fusing scheme is shown in figure 4. In the case of an event the trigger passes through the newly arrived data from the according sensor. The current model data is multiplied

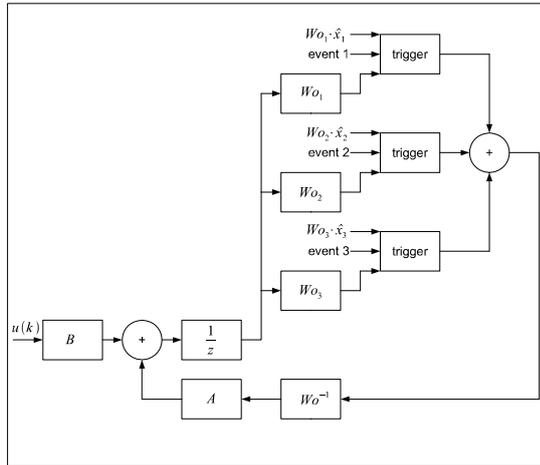


Fig. 4: Data fusion in actuator nodes.

multiplied with the remaining untriggered observability matrices and then summed up. Since all observability sub-matrices sum up to the full matrix a multiplication with the inverse Gramian ensures a proper transformation of these values back into state space. There are two advantages of this fusing scheme: First, at most times sensors will not always trigger an event at the same time. The here proposed fusion deals with this by taking the current model

value to substitute the missing sensors. Second, the multiplication with the observability matrices and its inverse make use of the inherent network and system structure. If a sensor can observe a state very well its new information will be given very much weight in the fusion. In the opposite case, if a sensor has no insight in a state during the fusion values close or equal to zero will be multiplied with it and therefore preventing its influence. The usage of these matrices is based on a profound method that relies solely on the model of the system and its precision.

RESULTS

The results of a test system are shown in figures 5 and 6. A small system of a simple plant, three sensors and actuators each was simulated. A classical PI-controller was designed and distributed onto the actuator nodes. Its goal is to bring the system into a desired state. The difference between the desired and actual state is summed up. At the same time the sum of all transmissions between sensor and actuator nodes is recorded. This scheme is repeated for rising threshold values. While the transmissions drop the total error is rising. To show the effectiveness of the Gramian data fusion a system with average value fusion was simulated as well. It can be seen that at a level auf 50% of maximal transmissions the sum of errors is 1.4 time higher when compared to the

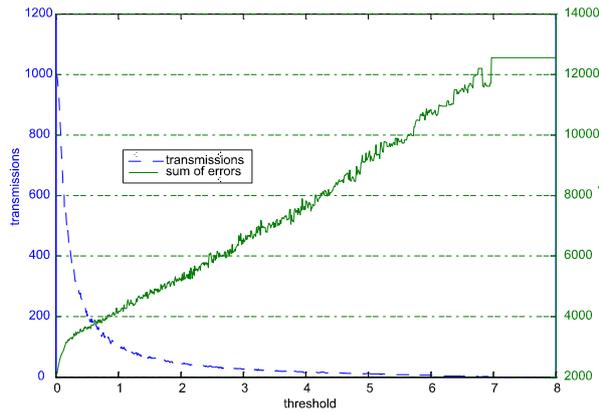


Fig. 5: Results for Gramian Data fusion.

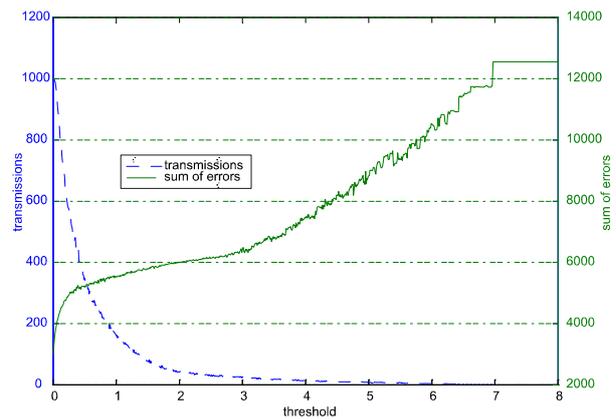


Fig. 6: Results for average data fusion.

Gramian data fusion. At 20% it is about 1.5 times higher. First, the influence of the threshold can be seen. Second, the effectiveness of the fusion method is obvious.

OUTLOOK

The ongoing work will use the properties of the Gramians to optimise the network topology by reducing it to a not fully-connected system. Next, the design of a controller, which was omitted here, will be analysed.

References:

- [1] João P. Hespanha, Payam Naghshtabrizi, and Yonggang Xu. Networked Control Systems: Analysis and design, 2006. "To appear in the *Proc. of IEEE*, Special Issue on Networked Control Systems".
- [2] Muis A. Montestruque and Panos J. Antsaklis. On the model-based control of networked systems. *Automatica*, 39(10):1837—1843, October 2003.
- [3] Reza Olfati-Saber. Distributed kalman filter with embedded consensus filters. In *Proc. of the 44th IEEE Conference on Decision and Control, and the European Control Conference 2005*, pages 8179—8184, Seville, Spain, December 2005. IEEE.
- [4] Sokwhoo Rhee, Deva Seetharam, and Sheng Liu. Techniques for minimizing power consumption in low data-rate wireless sensor networks. In *2004 IEEE Wireless Communications and Networking Conference*, volume 3, pages 1727—1731, Piscataway, NJ, USA, 2004. IEEE.
- [5] Andrew Stubbs and Geir E. Dullerud. Networked control of distributed systems: A testbed. In *Proc. of IMECE 200*, pages 1—5, New York, New York, November 2001. 2001 ASME International Mechanical Engineering Congress & Exposition.
- [6] Yodyium Tipsuwan and Mo-Yuen Chow. Control methodologies in networked control systems. *Control Engineering Practice*, 11:1099—1111, October 2003.
- [7] Yonggang Xu and João P. Hespanha. Estimation under uncontrolled and controlled communications in networked control systems. In *Proceedings of the 44th IEEE Conference on Decision and Control and the European Control Conference 2005*, pages 842—847, Seville, Spain, December 2005. IEEE.
- [8] Greg Welch and Gary Bishop. An introduction to the Kalman filter. <http://www.cs.unc.edu/~welch/kalman/>, 2001, ACM SIGGRAPH 2001.
- [9] Arthur Conley and Mario E. Salgado. Gramian based interaction measure. In *Proceedings of the 39th IEEE Conference on Decision and Control*, Sydney, Australia, December 2000. IEEE.

Authors:

Dipl.-Ing. Peter Hilgers
 University of Freiburg – IMTEK
 Department of Microsystems Engineering
 Georges-Köhler-Allee 078
 79110 Freiburg
 Phone: +49 761 2037305
 Fax: +49 761 2037439
 E-mail: Peter.Hilgers@imtek.de

Prof. Dr.-Ing. Christoph Ament
 TU Ilmenau
 Institute for Automation and Systems Engineering
 PF 100565
 98684 Ilmenau
 Phone: +49 3677 692814
 Fax: +49 3677 691434
 E-mail: Christoph.Ament@TU-Ilmenau.de

C. Saul / A. Mitschele-Thiel / A. Diab / M. Abd rabou Kalil

A Survey of MAC Protocols in Wireless Sensor Networks

ABSTRACT

Wireless sensor networks are one of the most important technologies of the 21st century. A sensor network is a network of smallest computers, which are equipped with sensors and collaborate to perform a common task. A network of these devices has a wide range of potential applications including environmental monitoring, intelligent facility surveillance, medical systems and robotic exploration.

Many Medium Access Control (MAC) protocols have been specifically designed for wireless sensor networks where energy awareness is an essential design issue. In this paper we present a survey of the current MAC protocols for wireless sensor networks. We highlight the state of the art and compare between these protocols with respect to the sources of energy waste.

I. INTRODUCTION

A sensor network consists of many densely arranged sensor nodes. In order to make them universally applicable, they need to be small, cheap, durable and able to hold a long lifetime. Because today available batteries have a limited capacity, the energy reserves have to be handled economically. When the sensor node has consumed its charge completely, it is inactive and does not have the ability to execute its task any more.

[1] identifies the major sources for energy waste, namely collision, overhearing, sending and receiving of control messages, idle-listening and overemitting. Thus, a well-designed MAC protocol should reduce the effects of these sources and consequently reduce the waste of energy. Many MAC protocols are developed to overcome these sources. Therefore, it is necessary to evaluate these protocols regarding the sources for energy waste, presented above. The rest of this paper is organized as follows: Section (II) focuses on state of the art. Following this, section

(III) presents a comparison between these protocols with respect to the above presented sources for waste of energy. After that, we conclude with the main results and the future work in section (IV).

II. REVIEW OF MAC PROTOCOLS

As in all shared-medium networks, MAC is an important technique that ensures successful operation of the network. MAC protocols control the mutual access of the shared medium. Current MAC design for wireless sensor networks can roughly be divided into contention-based and schedule-based protocols. Contention-based protocols contest admittance to the medium. Only one node can gain exclusive access. Contrary, schedule-based protocols underlie a predefined order. The schedule defines when and how long every node can access the medium.

a) Contention-Based Protocols

Power Aware Multi Access Protocol with Signalling Ad-Hoc Networks (PAMAS) [2] based on the Medium Access with Collision Avoidance (MACA) [3] protocol with the addition of a separate signaling channel. The nodes switch their power off if they are not active, so overhearing is prevented.

Sensor Medium Access Control (S-MAC) [1] attempts to prevent idle-listening, overhearing and collisions. The nodes stick to a fixed schedule. All nodes wake up at predefined cycles and stay awake for a while. During this time synchronisation, sending and receiving of data proceeds.

Timeout Medium Access Control (T-MAC) [4] is an improvement of the S-MAC protocol with less idle-listening. This is achieved through an adaptively determined wake-up cycle. All data are transmitted in bursts of variable length and nodes sleep in between. More protocol-overhead is caused however because of the additional control messages.

Berkeley Medium Access Control (B-MAC) [5] is based on the simple Carrier-Sense-Multiple-Access with Collision Avoidance CSMA/CA protocol. The underlying network can be diagrammed as a spanning-tree, where the root is the base station and the leaves are the nodes. B-MAC achieves a higher throughput and better energy efficiency than S-MAC and T-MAC. A disadvantage is the idle-listening, because B-MAC always eavesdrops on the medium.

Pattern Medium Access Control (P-MAC) [6] determines the sleep / wake-up schedules for a node adaptively, it relies on its own traffic and that of the neighbours. P-MAC achieves more power savings at low working loads and higher throughput at heavier traffic loads.

Zebra Medium Access Control (Z-MAC) [7] combines the strengths of Time-Division-Multiple-Access (TDMA) and CSMA while offsetting their weaknesses. Like CSMA, Z-MAC achieves a high channel utilisation and a low-latency at low contention and like TDMA it achieves a high channel utilisation at a high contention. Z-MAC is robust to dynamic topology changes and time synchronisation failures that occur in sensor networks and more energy efficient than B-MAC and S-MAC.

Carrier Sense Multiple Access with Preamble-Sampling (CSMA-PS) [8], [9] is a protocol that combines classical CSMA with a preamble-sampling. CSMA-PS prevents protocol overhead. However, it wastes energy through the sending of a large preamble.

Wise Medium Access Control (WiseMAC) [10] depends on the preamble sampling technique to minimise the idle listening. WiseMAC minimises the length of the wake-up preamble by taking advantage of knowing the sensor nodes sampling schedules. Therefore, WiseMAC achieves a better power efficiency.

b) Schedule-Based Protocols

Low Energy Adaptive Clustering Hierarchy (LEACH) [11] is a clustering-based protocol that minimises the global energy usage by distributing the load to all nodes at different points in time. This results in reducing the energy dissipation and enhancing the system lifetime.

Event Driven Time Division Multiple Access (ED-TDMA) [12] is an energy efficient TDMA protocol for event-driven applications. ED-TDMA improves channel utility by changing the length of a TDMA frame according to the number of source nodes and saves energy by using a bitmap-assisted TDMA schedule.

Traffic adaptive Energy Efficient MAC (TEEMAC) [13] is a cluster-based MAC protocol where each cluster is dynamically formed based on the cluster head. It prevents collisions in data transmission, idle-listening, overhearing and unnecessary control overhead.

Application Medium Access Control (AppMAC) [14] is an application-aware and event-oriented MAC protocol. AppMAC combines the advantages of the contention-

based and the schedule-based MAC protocols. AppMAC is able to support prioritised delivery of events, provide inter-event and intra-event fairness and improve channel utilisation while reducing energy consumption.

Self organizing Medium Access Control for Sensor-Networks (SMACS) [15] is a distributed MAC protocol. It enables a set of nodes to discover their neighbours and establish transmission / reception schedules for communicating with them.

Traffic Adaptive Medium Access (TRAMA) [16] is an energy-aware channel access protocol, which reduces energy consumption by ensuring that transmissions have no collisions and by allowing nodes to switch to a low-power mode whenever they are not transmitting or receiving data. TRAMA achieves high delivery guarantees and good energy efficiency.

Distributed Energy Aware Medium Access Control (DE-MAC) [17] is a TDMA based energy efficient MAC protocol. The protocol uses the TDMA technique together with periodic listening and sleeping to avoid the major sources of energy dissipation. DE-MAC achieves a significant gain in energy saving compared to other existing MAC protocols.

III. COMPARISON OF MAC PROTOCOLS

Contention-based protocols are preferable if the workload is not balanced. Busier nodes can get easier access to the medium. In the state of large workload schedule-based protocols are preferable, because they ensure an unproblematic transmission and a guaranteed throughput. The following tables present a comparative overview of different contention-based and schedule-based MAC protocols regarding the energy waste sources. A source of energy waste can be solved good, very good, bad or not solved in the discussed protocols.

Table1: Comparative overview of contention-based MAC protocols

	Idle-Listening	Collisions (Hidden-/ Exposed Terminals)	Overhearing	Protocol-overhead	Over-emitting
PAMAS	bad	very good/very good	very good	bad	very good
S-MAC	good	very good/very good	good	bad	very good
T-MAC	good	very good/very good	very good	bad	very good
B-MAC	bad	bad/bad	bad	very good	not solved
P-MAC	good	bad/bad	bad	bad	not solved
Z-MAC	bad	very good/very good	bad	bad	not solved
CSMA-PS	good	bad/bad	good	very good	not solved
WiseMAC	good	bad/bad	good	very good	not solved

Table 2: Comparative overview of schedule-based MAC protocols

	Idle-Listening	Collisions (Hidden-/ Exposed Terminals)	Overhearing	Protocol-overhead	Over-emitting
LEACH	very good	very good/very good	very good	bad	not solved
ED-TDMA	very good	very good/very good	very good	bad	not solved
TEEMAC	very good	very good/very good	very good	bad	not solved
AppMAC	good	good/good	very good	bad	not solved
SMACS	good	very good/very good	very good	bad	good
TRAMA	bad	bad/good	bad	bad	not solved
DE-MAC	bad	very good/very good	bad	bad	not solved

The table above shows that only LEACH, ED-TDMA and TEEMAC solve the idle-listening problem optimal. It's obvious that schedule-based protocols solve this problem better than contention-based protocols, because schedule-based protocols know when they are allowed to send and receive. In contrast contention-based protocols have to admittance to the medium. Collisions caused by hidden- and exposed stations are solved very well in all protocols expecting B-MAC, P-MAC, CSMA-PS, WiseMAC and TRAMA. Overhearing is more a problem of the contention-based than schedule-based protocols, because schedule-based protocols predefine who and when a node listens to the medium and when they shouldn't do. Regardless of those CSMA-PS and WiseMAC solve this problem in a good and PAMAS and T-MAC in a very good way respectively. With respect to protocol overhead they are completely different. Schedule-based protocols have more protocol overhead caused by organization and synchronization of the medium access while contention-based protocols don't need this and achieve a higher efficiency. The best protocols regarding protocol overhead are B-MAC, CSMA-PS and WiseMAC. With respect to the overemitting, only a few protocols attempt to solve this problem, namely PAMAS, S-MAC, T-MAC and SMACS.

IV. CONCLUSION

In this paper we have presented a survey of the well known MAC protocols and compared them to each other regarding the energy waste sources. Contention-based protocols are not underlying an explicit control. They react faster and better if the topology of the networks change often. The small or non-existent protocol overhead improves the scalability and efficiency. On the other hand, schedule-based protocols are subject to an explicit control because every node obtains a timeslot for sending

and receiving data. Thus, organisation and synchronisation are the imperative of schedule-based protocols. The disadvantage of schedule-based protocols is that their adaptability and scalability are lower compared to contention-based protocols. The additional protocol overhead in schedule-based protocols improves collision avoidance. It produces however more energy waste. In the case of large workload, schedule based protocols ensure an unproblematic transmission and a guaranteed throughput.

According to section (III) we don't have found a winner under the MAC protocols. It is recommended to gather firstly the requirements of the application. Based on this information decisions for an appropriate MAC protocol can met. We see that it is not likely that a unified or "standard" MAC protocol for sensor networks will emerge.

References:

- [1] W.Ye, J.Heidemann, D.Estrin: "An Energy-Efficient MAC protocol for Wireless Sensor Networks", Technical Report ISI-TR-543, 2001.
- [2] C.Raghavendra, S.Singh: "Pamas – power aware multi-access protocol with signalling for ad hoc networks", ACM Computer Communication Review, 1998.
- [3] P.Karn: "MACA - A new channel access method for packet radio", ARRL/CRRL Amateur Radio 9 Computer Networking Conf, 1990.
- [4] V.Dam, K.Langendoen: "An Adaptive Energy-Efficient MAC Protocol for Wireless Sensor Networks", Proceedings of Sensys'03, 2003.
- [5] A.Woo, D.Culler: "A Transmission Control Scheme for Media Access in Sensor Networks", Mobicom2001, 2001.
- [6] T.Zheng, S.Radhakrishnan, V.Sarangan: "PMAC: An adaptive energy-efficient MAC protocol for Wireless Sensor Networks", The 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS'05), 2005.
- [7] I.Rhee, et al: "Z-MAC: Hybrid MAC for Wireless Sensor Networks", ACM SenSys'05, 2005.
- [8] A.El-Hoiydi: "Aloha with Preamble Sampling for Sporadic Traffic in Ad Hoc Wireless Sensor Networks", IEEE ICC2002, 2002.
- [9] A.El-Hoiydi: "Spatial TDMA and CSMA with Preamble Sampling Aloha with Preamble for Low Power Ad Hoc Wireless Sensor Networks", IEEE ISCC2002, 2002.
- [10] A.El-Hoiydi, J.D.Decotignie: "WiseMAC: An Ultra Low Power MAC Protocol for the Downlink of Infrastructure Wireless Sensor Networks", 2004.
- [11] W.R.Heinzelman, A.Chandrakasanand, H.Balakrishnan: "Energy-Efficient communication Protocol for Wireless Micro sensor Networks", Proceedings of 33rd Hawaii International Conference on System Sciences, 2000.
- [12] H.Gong, et al: "An Energy Efficient TDMA Protocol for Event Driven Applications in Wireless Sensor Networks", Mobile Ad-hoc and Sensor Networks, Second International Conference, Hong Kong, China, 2006.
- [13] C.Sungrae, et al: "Traffic-Adaptive Energy Efficient Medium Access Control for Wireless Sensor Networks", Mobile Ad-hoc and Sensor Networks, First International Conference, Wuhan, China, 2005.
- [14] D.Junzhao, S.Weisong: "An Application-Aware Event-Oriented MAC Protocol in Multimodality Wireless Sensor Networks", Mobile Ad-hoc and Sensor Networks, Second International Conference, Hong Kong, China, 2006.
- [15] A.Sorabi, et al: "Protocols for Self-Organization of a Wireless Sensor Network", IEEE Pers. Commun, 2000.
- [16] V.Rajendran, K.Obraczka, J.J.Garcia-Luna-Aceves: "Energy-efficient collision-free medium access control for wireless sensor networks", Conference on Embedded Networked Sensor Systems, 2003.
- [17] S.Kalidindi, et al: "Distributed Energy Aware MAC Layer Protocol for Wireless Sensor Networks", International Conference on Wireless Networks, 2003.

Authors:

Christian Saul

Prof. Dr.-Ing. habil. Andreas Mitschele-Thiel

Dipl.-Ing Ali Diab

MSc.-Inf. Mohamed Abd rabou Kalil

TU-Ilmenau, Faculty for Informatic and Automation, Gustav-Kirchhoff-Str. 1, P.O.B. 10 0565

98693, Ilmenau

Phone: +49 3677 69 2819

Fax: +49 3677 69 1220

E-mail: christian.saul1@stud.tu-ilmenau.de / {mitsch|ali.diab|mohamed.abdrabou}@tu-ilmenau.de

T. Rossbach / M. Götze / A. Schreiber / M. Eifart / W. Kattanek

Wireless Sensor Networks at their Limits – Design Considerations and Prototype Experiments

ABSTRACT

Wireless sensor networks (WSN) consist of a number of autonomous nodes that are able to communicate within a limited range. In order to send a message to a destination which is not within reach of a given node, the message has to be relayed by intermediate nodes. This article details the design, implementation, and test of such a network for applications whose requirements exceed those of traditional WSNs.

I. INTRODUCTION

There is a plentitude of scientific publications dealing with the design and application of distributed sensor networks [1]. Often this concerns only a particular sub-area or a specific implementation for a given application scenario. This paper details the design of a wireless sensor network (WSN) according to specifications at the limit of what is technically feasible regarding data transfer rates and range of communications (fig. 1), aiming to reduce the impact of transmission errors and provide functionality for analysis.

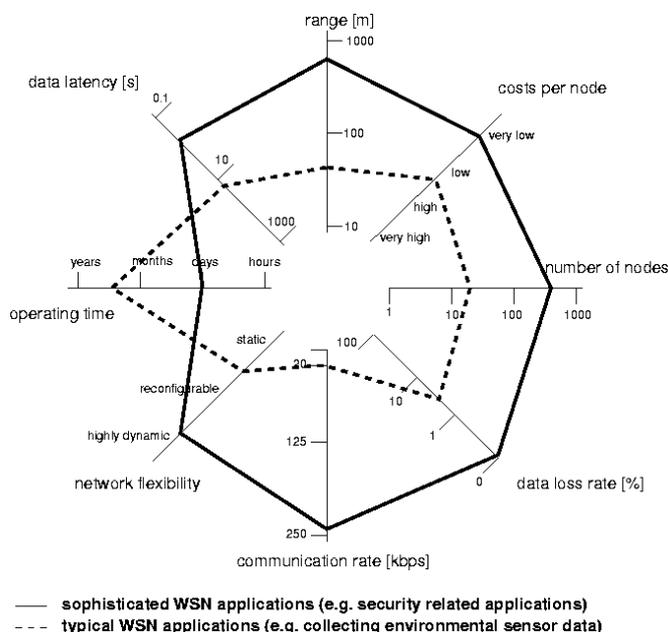


Figure 1: Different Classes of Application Requirements

A. Areas of Application

Besides military applications, there is a great variety of possible civilian applications of wireless sensor networks. These include monitoring a colony of ducks in a natural

preserve on an island [10], or the acquisition and transmission of data concerning bodily functions of athletes, such as pulse and blood pressure [11], or tasks in the domain of security and surveillance technology. It is also possible to think of applications in forests, such as the acquisition of environmental data or data concerning the behavior of wildlife.

B. Requirements

A traditional sensor network consists of a number of autonomous participants, so-called motes. Each mote monitors changes in one or more sensor values, logs these and forwards them to a destination via the network. This type of network often exhibits a high density of nodes, i.e., one mote has a large number of neighbors within immediate range of communications. Sensor values accumulate at a slow rate. Packet loss is in most cases uncritical. Missing values within series of measurements can be made up for by, e.g., interpolation. Often placement and environmental conditions are known in advance. This way, the network can be designed with fixed constraints, such as range and data transmission rate, in mind.

The implementation this paper is based on, however, aims at an application of WSN technology at its limits. Motes are spread across a large area, leading to a mote typically having merely two or three neighbors within reach. Looking at the network as a whole, it is designed to be comprised of a large number of participating nodes, in the range of several hundred. Packets are, in contrast to traditional networks, sent rather frequently. To make things even more difficult, data loss is intolerable for critical data. The future conditions of deployment and environmental conditions cannot be completely foreseen or restricted during the design phase. This means that the network might be deployed within a building, in the woods, or in a densely populated area of city. Although the usual destination of data packets will be a dedicated base station, it is still necessary to transmit commands and acknowledgments in the opposite direction, i.e., into the network. A low latency in between the transmission of a packet towards the base node and an acknowledgment or reaction to it was another criterion in the design.

Given these determining factors, the functional realm of traditional sensor networks is left behind. Still, the reason for employing WSN technology consisted in such aspects as physical size, power consumption and simple installation.

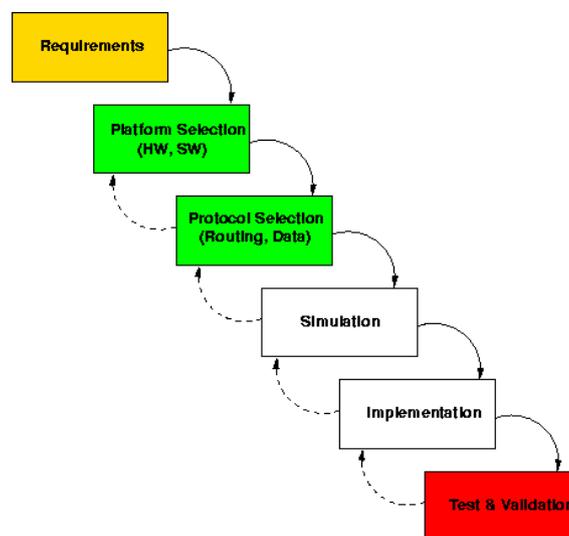


Figure 2: General Design Flow of WSN

II. DESIGN FLOW

The design of a network such as this starts with basic theoretical considerations (fig. 2). At this point, the following aspects are of primary concern:

1. The amount of data generated by each mote as well as its rate of transmission
2. The number of nodes in the network
3. A partition into critical (acknowledged) and non-critical data
4. The type and length of commands sent to a mote

When dealing with critical data, a safeguarding mechanism implemented via acknowledgments at the link and/or transport layer should be employed in order to ensure packets reach their destinations. Non-critical packets are those for which this is not a requirement. A loss of packets from this group is acceptable and is made up for by the receipt of the next packet of the same type.

A. Hardware and Operating System

One important point, particularly inasmuch as data transfer rates and communication range are concerned, is the choice of the hardware platform. Research publications tend to frequently cite the *Mica*, *MicaZ*, and *TelosB* modules by *Crossbow Technology, Inc.* An alternative to this for both research purposes and commercial deployment consists in the use of *TinyNode* [5,6] modules distributed by *Shockfish SA*. These modules employ *TinyOS* [7] — an embedded operating system tailored for WSN applications, major strengths of which consist in its modularity, portability and a reduced time-to-market. *TinyOS* is open-source, royalty-free software. Programming based on this OS is done in *NesC*, a C dialect extended for modularity, which speeds up development and facilitates reuse among projects. There already are a large number of ports and libraries for such purposes as routing, localization, and processing of measurement data.

B. Range

Under ideal conditions, radio waves spread according to the following formula:

$$P_E = P_S \left(\frac{\lambda}{4\pi \cdot r} \right)^2 \cdot G_S \cdot G_E$$

This formula describes how the power $P_S [W]$ emitted by the transmitter decreases with the square of the distance $r [m]$. Furthermore, the received power, $P_E [W]$, depends on the wavelength $\lambda [m]$ as well as the gains of the sending and receiving antennae, G_S and G_E , respectively.

Measurements under realistic conditions, however, are subject to further influences. For example, besides the transmission power and the sensitivity of the receiver, the maximum range also depends on the kind and structure of the environment. The height above ground and reflections of the signal by walls and obstacles also figure into this. The power of a signal received by a mote is represented by the so-called *Received Signal Strength Indicator (RSSI)*. Due to a multitude of factors, this value varies in the course of longer periods of time — even in a static network and under (seemingly) constant environmental conditions (fig. 3).

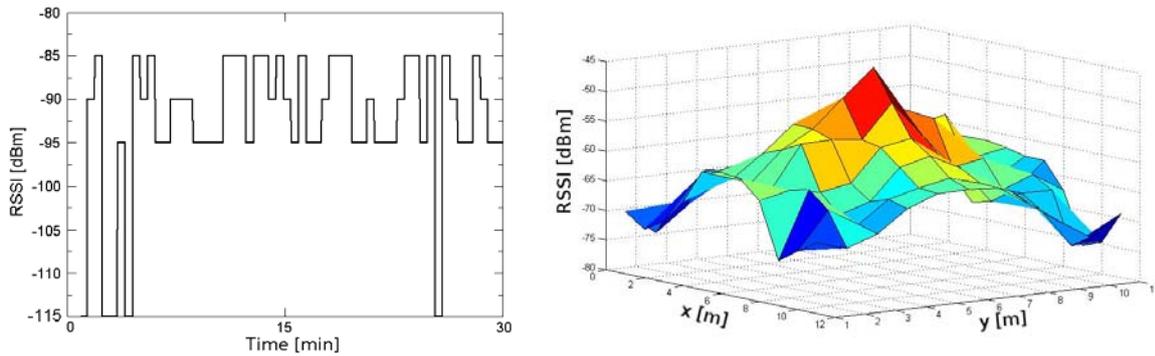


Figure 3: RSSI over the Course of 30 Minutes (left) and Radial Propagation Characteristics [2] (right)

Evaluating the radial propagation characteristics (fig. 3), it becomes obvious that those are not nearly ideal. RF links in between nodes operate accordingly.

The maximum range can be improved further through a careful design of antennae and PCBs and an optimization of transceiver parameters with respect to data transfer rate and filter bandwidth.

C. Routing Protocols

In contrast to traditional networks such as IP-based networks, new approaches and solutions had to be found for routing in mesh networks [4]. Such networks are based on links which are often unreliable and asymmetric (fig. 6) with a potentially high degree of mobility among its participants. Due to the dynamic nature of the network, the information about its topology is subject to continuous change. This also brings about the disadvantage that each node represents an unreliable component because it may fail at any time without advance warning.

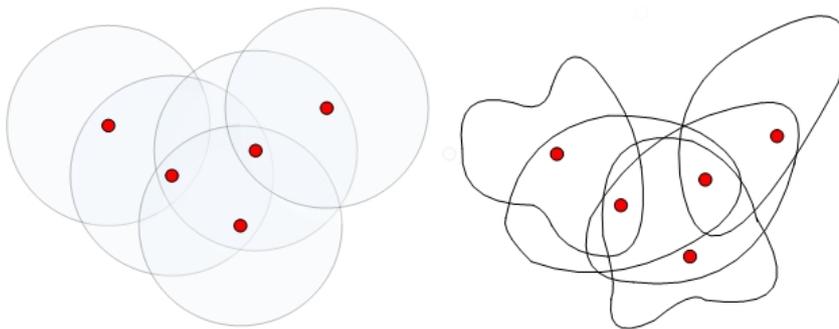


Figure 4: Ideal and Real Network Model

For TinyOS, a plentitude of routing protocols have been developed. There are geographical and topology-based, pro-active and reactive algorithms. They are often based on a link quality judgment. This may be obtained through the amount of packet loss among two nodes or based on the RSSI value. A routing algorithm chooses the optimal route according to a cost function based on the measurements.

D. Transmission Errors and Packet Loss

The link in between two nodes can be protected against packet loss by means of link-

layer acknowledgments. If an acknowledgment by the receiver fails to be received, the sender repeats the message and again waits for an acknowledgment. This mechanism by itself, however, represents merely a safeguard against lost packets — they may still have suffered from transmission errors. Checking a Cyclic Redundancy Checksum (CRC) value transmitted along with the message allows for conclusions regarding the occurrence of bit errors. Figure 7 shows the accumulated number of acknowledgment and CRC errors over a period of time. The missing link-layer acknowledgments and resulting packet loss could be avoided thanks to a transport-layer acknowledgment mechanism implemented in the network software. In subsequent tests, no loss of packets acknowledged this way occurred.

Packet loss can still never be completely avoided. If a node fails to receive an acknowledgment for an extended period of time and further messages accumulate in the meantime, messages will have to be dropped sooner or later (at the latest, once the queue’s capacity is exceeded). In conjunction with this, the network topology, the degree to which “bottleneck” links are utilized, and interference effects such as the Hidden Node Problem are important to consider. Acknowledgments on a secondary level (transport layer), i.e., end-to-end protection, can help reduce the impact of these effects when dealing with critical data, but they will never be completely avoidable.

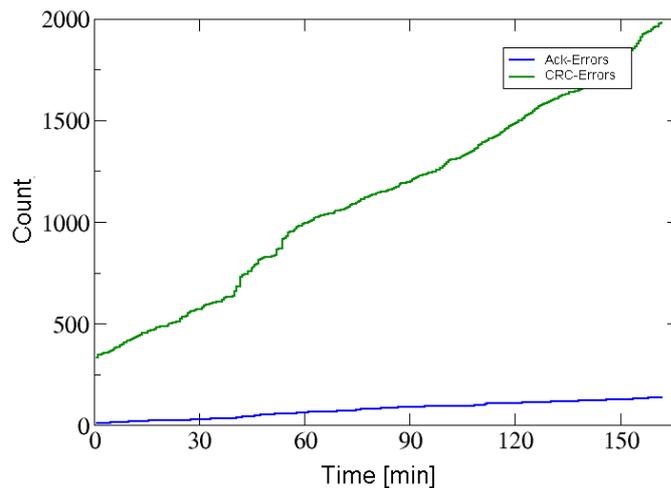


Figure 5: Acknowledgment and CRC Errors (Protocol: Drain) [3]

E. Simulation

Simulation can be used to analyze and optimize the behavior of a network and the selected routing protocol. There is a simulator already integrated with the TinyOS development environment. *TOSSim* [8] allows for simulating both lossless and lossy RF links. Unfortunately, it currently only includes a radio model for the Chipcon CC1000 radio transceiver. This makes a comparison with other platforms difficult. An alternative to *TOSSim* consists in using one of a variety of stand-alone simulators available, such as *OMNet++* [9]. However, in order to simulate a TinyOS-based environment, its code base first has to be transformed from NesC to regular C++ using a translator.

F. Implementation and Optimization

Profiting from the NesC concepts of software modules and components, a selected

routing protocol can be implemented in a modular fashion. For the application whose development this paper is based on, the protocols *Drain* (for any-to-one routing with a fixed base node) and *Drip* (for commands and acknowledgments) have been chosen. The existing TinyOS implementation offers a versatile programming interface which lent itself well to custom extensions, such as acknowledgments of messages on both the link and transport layer, improvements to the cost function employed by the Drain algorithm, and functions for processing measurement data. Added functionality includes:

- Route tracing to the base node
- Roundtrip time measurement as a basic latency indication
- Acknowledgment and CRC error monitoring
- Monitoring of the amount of data transmitted and received by individual nodes
- RSSI and battery voltage monitoring

IV. SUMMARY AND OUTLOOK

This paper has attempted to describe the process from the selection of a hardware platform through the required research into routing protocols and their simulation to a software implementation which is functional and can be deployed in practice. The system designed covers a broad spectrum of applications which go beyond what WSNs are typically used for. Besides the transmission of arbitrary packets of data, performance data concerning the network can be logged and analyzed.

The focus of ongoing work consists in improving upon and optimizing the routing protocol. Goals include a shorter response time to topology changes, error detection and correction, and further optimizations of the link cost calculation.

References:

- [1] A. Schreiber, M. Götze, E. Ulicna, W. Kattanek, „System design issues in advanced applications of wireless sensor networks,“ Internationales Wissenschaftliches Kolloquium, Ilmenau, 2006.
- [2] M. Eifart, „Entwicklung einer Lokalisierungslösung für mobile Knoten innerhalb drahtloser Sensornetzwerke für Freifeldanwendungen,“ Diploma Thesis, Institut für Mikroelektronik- und Mechatronik-Systeme gGmbH, Ilmenau, 2007.
- [3] T. Rossbach, „Aufbau eines Mesh-Netzwerkes und Optimierung in Bezug auf relevante Kennwerte,“ Diploma Thesis, Institut für Mikroelektronik- und Mechatronik-Systeme gGmbH, Ilmenau, 2007.
- [4] M. Abolhasan, T.A. Wysocki, and E. Dutkiewicz, "A Review of Routing Protocols for Mobile Ad hoc Networks", In: Elsevier Journal of Ad hoc Networks, 2 (2004), 1-22.
- [5] H. Dubois-Ferrière, R. Meier, L. Fabre, and P. Metrailler, "TinyNode: A Comprehensive Platform for Wireless Sensor Network Applications." In: Information Processing in Sensor Networks (IPSN 2006), 2006.
- [6] <<http://www.tinynode.com/>>
- [7] <<http://www.tinyos.net/>>
- [8] Philip Levis, Nelson Lee, Matt Welsh, David Culler, "TOSSIM: Accurate and Scalable Simulation of Entire TinyOS Applications." In: Proceedings of the First ACM Conference on Embedded Networked Sensor Systems (SenSys 2003).
- [9] < <http://www.omnetpp.org/>>
- [10] Intel Research Laboratory, "Great Duck Island, In-Situ Sensor Networks for Habitat Monitoring," <<http://www.coa.edu/html/motes.htm>>
- [11] "WHMS - Wearable Health Monitoring Systems," <<http://www.ece.uah.edu/~jovanov/whrms>>

Authors:

Dipl.-Ing. Tobias Rossbach, Dipl.-Inf. Marco Götze, Dr.-Ing. Axl Schreiber, Dipl.-Ing. Mario Eifart, Dipl.-Ing. Wolfram Kattanek
 Institut für Mikroelektronik- und Mechatronik-Systeme gGmbH
 Ehrenbergstr. 27
 D-98693 Ilmenau
 Phone: +49 (3677) 69 – 55 00
 Fax: +49 (3677) 69 – 55 15
 E-mail: {forename.surname}@imms.de

Yiping Zhong / Jianqing Ma

Ring Domain-Based Key Management in Wireless Sensor Network

ABSTRACT

Wireless sensor networks (WSNs) are Ad hoc networks that include sensor nodes with limited computation, memory, energy and communication capabilities. When WSNs are deployed in unprotected/hostile areas, Wireless sensor networks (WSNs) are known to be particularly vulnerable to all kind of attacks like eavesdropping communication, node capture attacks, etc. Hence WSNs require cryptographic protection of communications, sensor capture resistance, key distribution and key revocation. In this paper, we present a key management schemes based on location-aware and random key predistribution model. The schemes propose to generate the pairwise keys by union of diversity of random keys and spatial diversity, and therefore improve network resistance on compromised nodes. Comparing with these random key predistribution based models (e.g. q-composite schemes), our scheme can improve the performances both on secure link connection ratio and security for resistance on compromised node. Comparing with these location-aware key management schemes, the scheme has no assumption about preknowledge of deployment or aid of distributive key configuration servers, which the major location-based key management schemes usually need. The analysis indicate that this scheme have nice properties like high secure connection ratio, security, resource-saving, resilience of network, etc.

Key words: wireless sensor network; security; key management;

1. INTRODUCTION

In future, wireless sensor networks (WSN) are expected to have wide applications both in military and civil fields. These applications include battlefield surveillance, target detection and tracking by the military, microclimate control in buildings, nuclear, biological and chemical attack detection, home automation, environmental monitoring, etc. When WSNs are deployed in unattended/hostile environments such as building guard, battlefield, etc, the adversary may launch various attacks such as eavesdropping, falsifying legitimate nodes to disturb network's objective, etc. To prevent these attacks, communication should be encrypted and authenticated for the

sake of security. Secure communication is made possible through the use of keys which are themselves managed using the techniques of key management.

Unlike the traditional network, WSN have several factors (e.g. vagaries of wireless links; Ad hoc communication; vulnerability of nodes to physical capture; resource constraints such as limited memory, communication and computation capability, etc) which make the key management of WSN very different and difficult, especially on using public key schemes. Traditional techniques of key management like certification authority (CA) and key distribution center (KDC) cannot be applied in the WSN because of the following reasons: 1) single point of failure and incurring to denial of service (DOS) attack; 2) lowering service success ratio and prolonging service time because of the high bit error ratio in wireless ad hoc communication; 3) network congestion because of high communication overhead of node authentication. Symmetric key schemes have nice properties like simple and rapid encryption algorithms, short key bit, etc. Therefore, symmetric key schemes have been widely proposed to address the problem of key management in WSNs on account of the resource constraints associated with their features mentioned above[1][2].

Key management in WSNs require all nodes that need secure communication to share key for building secure links; support to securely distribute keys to additional nodes for secure communication with other nodes; support to renew older key, revoke compromised key effectively and efficiently, etc; Therefore, the main factors of key management scheme in WSNs include: 1) securely distributing keys to each sensor nodes; 2) securely discovering the shared key of sensor nodes and can prevent these attacks like eavesdrop, falsified nodes, etc ; 3) network resilience against node capture; 4) less influence on network performance, especially for connectivity, resource consumption and scalability 5)being convenient to renew or revoke the keys.

The sensor nodes of WSNs are collected, secure and trust each other before deployment, Therefore, the popular key management schemes using symmetric key in WSNs are random key predistribution schemes[4][5][6][7], in which a subset of a key pool are predistributed to each sensor node before deployment. These schemes keep a certain secure connectivity probability of any pairwise nodes because these nodes share at least one common key with probability. There are two main merits in these key predistribution schemes. 1) The major work of key management has been finished before deployment and the schemes only need run the key agreement after deployment, hence it can save communication overhead and energy consumption besides security risk for WSNs. 2) They trade off these network performances (e.g. resource consumption, network connectivity, network security, etc).

In this paper, we propose a novel location-based key management-RDBK by using random key predistribution. The RDBK scheme randomly distributes original key set to every sensor; then the sensor nodes generate derived key set by using their original keys and the broadcasted random number keys of base station after deployment; and last, RDBK builds the pairwise key of node to ensure secure link by hashing their common derived keys.

2. RDPK SCHEME

In this section, we will introduce details of RDBK scheme in three sequent phases: key predistribution before deployment, key distribution in initial phase of network deployment; building secure links before secure communication.

2.1 Key predistribution phase

In the phase of key predistribution, the base station store a key pool which composes of original keys K_o^i ($i=1, 2, \dots, P$), and generate random number keys- Rnd_j ($j=1,2\dots M$) by using random number Rnd_M and one-way hash function, where $Rnd_j=Hash(Rnd_{j+1})$. And then, The base station randomly predistributes each sensor node a key subset of key pool, which compose of different original keys K_o^i with number of $R(R \leq P)$, and a common one-way hash function $H()$.

2.2 Initial phase of network deployment

In the initial phase of network deployment, the base station broadcast the sequent random number keys- $Rnd_1, Rnd_2, \dots, Rnd_k$ ($K \leq M$) by increasing power level p (see formula 1). Thus, the sensor nodes in different ring domains receive different random number keys (See Fig1). Each sensor stores the first received $r+1$ random number keys (e.g. $Rnd_j, Rnd_{j+1}, \dots, Rnd_{j+r}$) and then verify them by $Rnd_j=H(Rnd_{j+1}), Rnd_{j+2}=H(Rnd_{j+1}), \dots, Rnd_{j+r}=H(Rnd_{j+r-1})$.

$$P_{out} = P_{max} / L * p \quad (1)$$

Where L is the number of power level which the base station can adjust to broadcast message.

$p=1,2,\dots,L$;

P_{max} is the maximum power of base station can launch.

After verified these keys, sensor nodes derive the keys $K_d^i, K_d^{i+1}, \dots, K_d^{i+r-1}$ by using their original keys (e.g. K_o^i), verified random number keys (e.g. Rnd_j, Rnd_{j+1}) and hash function $H()$. The derive key set can be generated by the formula like $K_d^i=H(Rnd_j, K_o^i), K_d^{i+1}=H(Rnd_{j+1}, K_o^i), \dots, K_d^{i+r-1}=H(Rnd_{j+r-1}, K_o^i)$ where the K_o^i is the original key and Rnd_j is the random number key (for example, see Figure 1). It can deduce that the number of derived keys is $r \cdot R$ where r is the number of stored random number keys and R is the number of stored original keys in each sensor node. It should point out that RDBK can increase the number of common derive keys and therefore improve the probability to build the pairwise keys comparing with q-composite scheme [3].

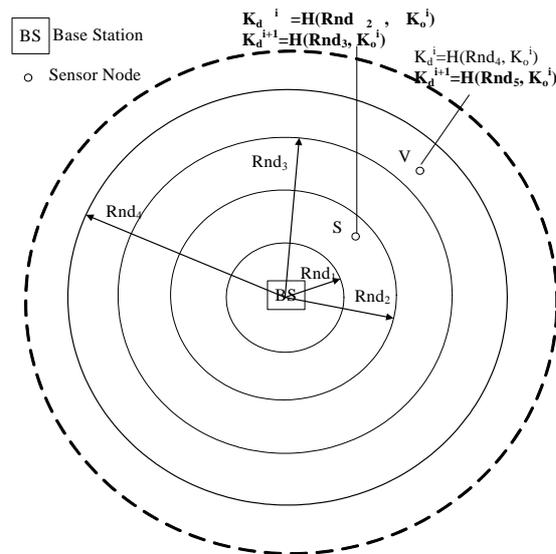


Fig1. Illustration of Derived Key Generation

2.3 Building secure links phase

In the phase of building secure links, the RDBK scheme is similar as q-composite scheme. When the number of common derived keys (denote as q') surpass the threshold number q , we build the pairwise key by function $H(K_d^{c1}, K_d^{c2}, \dots, K_d^{cq'})$, where H is the one-way hash function and K_d^{ci} is one of common derived keys between a pair of nodes. According to the approach mentioned above, we can deduce that the threshold of common original keys decreases to number q/r for successfully building secure link in RDBK. Therefore, RDBK have better performance both on security and connectivity, comparing with q-composite scheme [3].

3. ANALYSIS

In this section, we will compare RPBK scheme with q-composite scheme and other location-based key management scheme. Before comparison, we define the secure connectivity ratio and compromise ratio like [8]:

Secure connectivity ratio: For a given node, it is defined as the ratio of the number of neighbors of the node with which it can form secure links (since it shares keys with those neighbors) to the total number of neighbors of the node. The Secure connectivity ratio for the network is then the average of the connectivity values for each of the nodes in the network.

Compromise ratio: The compromise ratio is defined as the ratio of the number of secure links formed by the non-compromised nodes that have become vulnerable to the total number of secure links formed by non-compromised nodes in the network. The secure links become vulnerable on account of the leakage of keying material on the compromised nodes.

We use Matlab to simulate a network of sensors. In the simulations reported here, we assume that the sensors nodes with number of 600 are deployed randomly over an area of size 800x800 units. The default values for the transmission radius of sensor node is 30 units. The default threshold values of sharing key number is 3 ($q=3$). The default storing random key number of each node is 3 ($r=3$). Because the Hash function that derive the derived key is deleted after the initial phase of network

deployment, the adversary even capture all keys in pool, it still cannot construct the pairwise key. Therefore, we assume that the adversary have the ability to get the Hash function and therefore construct the pairwise key in experiment. Thus, when the ratio of compromised nodes is less than 30%, (the number of ring $L=10$ or 20 , the key number of pool $P=1000$ or 5000 , the number of original key $R=75$, the number of shared derived key $q=3$, the compromise ratio is less than 0.01% and the secure connectivity ratio are larger than 99%. At the same condition, the compromise ratio of q -composite is about 45% and its secure connectivity is 92%. In a word, the main performance of RPBK scheme is better than q -composite. Comparing with other location-based schemes, RPBK scheme have no assumption that the sensor node location can be predicted before deployment and therefore can deploy WSN more conveniently and expand the application scenario of WSN.

4. CONCLUSION

In this paper, we propose a key management scheme named RPBK, which is suitable to static WSN. The RPBK scheme unites the diversity of node deployment location with the diversity of random predistribution keys. Thus, on the one hand, the RPBK scheme can expand the size of key pool and reduce the probability of shared key of sensor nodes in different ring area to improve the performance of compromise resistance. On the other hand, RPBK scheme can improve the number of shared pairwise key of neighbor nodes and therefore improve the security connection ratio of network. Comparing with other location-based schemes, RPBK need not deployment preknowledge and therefore improve the convenience of WSN deployment and application scenario. In a word, RPBK scheme have nice performance like network security, connectivity and scalability.

References:

- [1] S. Basagni, K. Herrin, D. Bruschi, and E. Rosti. Secure pebblenet. Proceedings of the 2001 ACM
- [2] Laurent Eschenauer, Virgil D. Gligor, A Key-Management Scheme for Distributed Sensor, NetworksCCS'02, ACM, November 18–22, 2002
- [3] H. Chan, A. Perrig, and D. Song, "Random key predistribution schemes for sensor networks," Proceedings of 2003 Symposium on Security and Privacy. Los Alamitos, CA: IEEE Computer May 11–14 2003, pp. 197–215
- [4] Laurent Eschenauer, Virgil D. Gligor, A Key-Management Scheme for Distributed Sensor, NetworksCCS'02, ACM, November 18–22, 2002
- [5] H. Chan, A. Perrig, and D. Song, " Random key predistribution schemes for sensor networks, " Proceedings of 2003 Symposium on Security and Privacy. Los Alamitos, CA: IEEE Computer May 11 – 14 2003, pp. 197 – 215.
- [6] Wenliang Du, Jing Deng, Yunghsiang S. Han, and Pramod K. Varshney. A pairwise key pre-distribution scheme for wireless sensor networks. In ACM CCS 2003, pages 42–51, October 2003.
- [7] Wenliang Du, Jiang Deng. A key management scheme for wireless sensor networks using deployment knowledge[C]. In: Proceedings of the IEEE INFOCOM'04, 2004
- [8] Farooq Anjum, Location dependent key management using random key-predistribution In Sensor Networks. WiSe'06 September 29, 2006

Authors:

Prof. Yiping Zhong

Ph.D Jianqing Ma

Department of Computing and Information Technology, Fudan University,

No.220 Handan Rd, 200433, Shanghai, P.R.China

Phone: 86-21-65643189

Fax: 86-21-65647894

E-mail: ypzhong@fudan.edu.cn; jqma_edu@yahoo.com.cn

V. Nissen

Automatic Forecast Model Selection in SAP Business Information Warehouse under Noise Conditions

Introduction

Companies and markets are complex interacting systems. Despite inherent uncertainty it is a necessity for companies to plan ahead. Thus, forecasting based on IT-tools is today an integral part of management. Typical applications include sales and revenue forecasting as inputs for supply chain management and budgeting, respectively. One of the most important vendors for Enterprise Resource Planning (ERP) software in Europe and elsewhere in the world is the SAP AG with its IT-product portfolio centered around the R/3TM system.

Various SAP tools offer forecasting methods as part of their business functionality. Examples include SAP R/3TM, SAP Advanced Planner & Optimizer (APOTM), and SAP Business Information Warehouse (BWTM). The forecasting component examined here is the automatic forecast model selection in SAP BW Business Planning & Simulation (BPS). The automatic model selection fits a forecast model to the available historical data while minimizing some error measure, which is Akaike's Information Criterion in the case of SAP BW [1, 2].

It is a software feature routinely used by many companies in their forecasting, for instance in sales planning. At first sight, this feature seems to relieve the planner from choosing an adequate statistical forecast model for historical data himself, which can be a time-consuming and difficult task.

Because it is so often applied, one should aim for a good understanding of the strengths and weaknesses of automatic model selection in order to make a well-informed decision when to use it. The experiments outlined briefly in this paper are a contribution to this goal.

Method

Three basic functions were used in our forecasting experiments performed with SAP BW BPS release 3.5. The empirical tests are based on time series data showing trend, seasonal and trend-seasonal patterns.¹ The total horizon were 96 month of which 48 were treated as historical data and the other 48 were the planning horizon. Figure 1 gives an illustration of the time series patterns that were used.

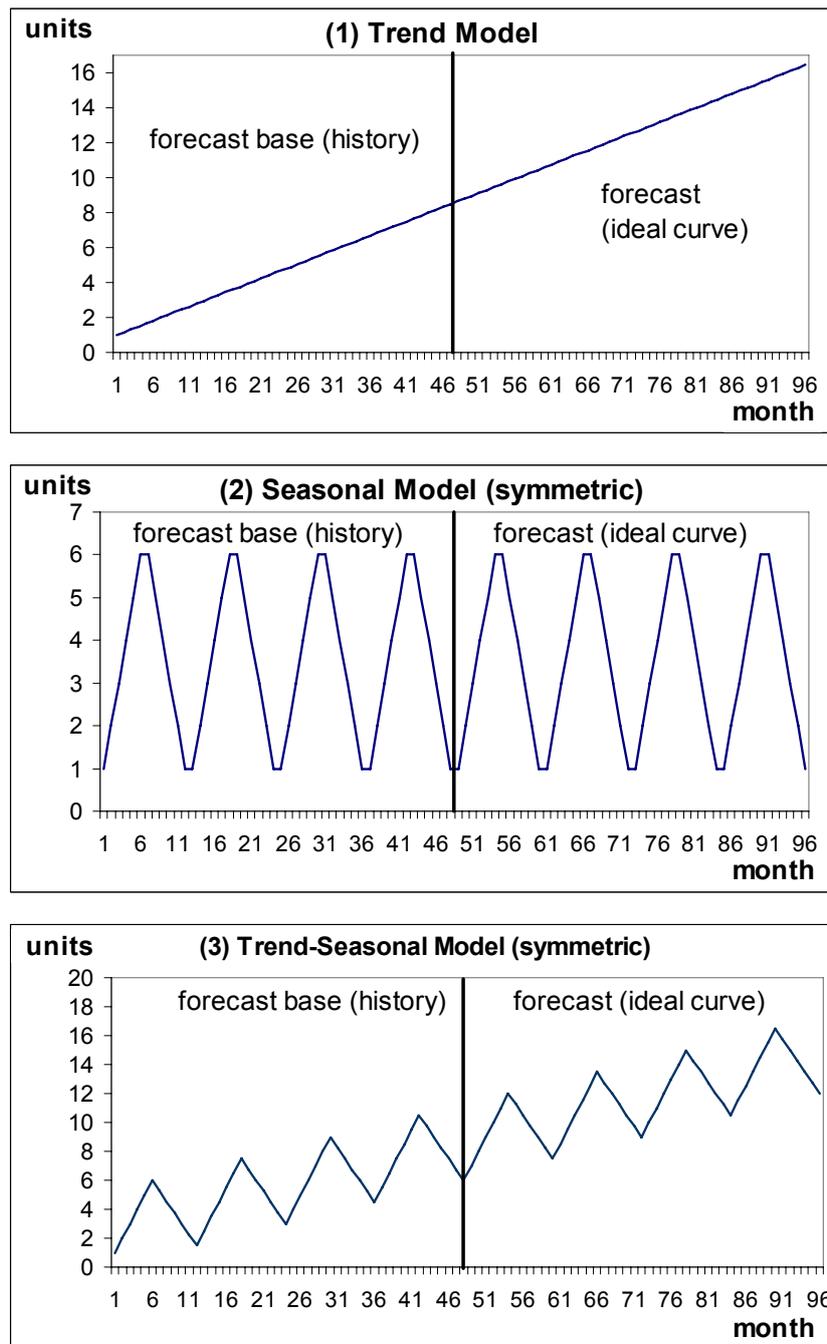


Fig. 1: The basic functions discussed here

¹ For a more detailed account of the results see [3].

The data was initially used in an ideal deterministic pattern which should provide a rather trivial task for automatic forecast model selection. Thereafter, varying amounts of normally-distributed noise are added to the historical data in order to see how this affects the quality of model selection and the forecasting result. The additive noise component mimics real-life data quality which can be characterized as a stochastic influence on trend, seasonal or other patterns. Practically, each historical data point x of the basic functions was modified in the following way:

$x' = x + a \cdot N(0,1)$, where $N(0,1)$ represents a standard-normally distributed random number and a is a noise factor that was varied according to $a = [0 ; 1 ; 2 ; 5]$. Thus, we conducted four sets of forecasting experiments for all functions. To remove arbitrary results, at each noise level and for each function 30 runs with different random number seeds for the time series modification were performed.

Results

Due to space limitations, results are presented here in the form of tables for each function and noise-level. The tables report on data averaged over all 30 runs for each experiment. The mean squared error (MSE) and the mean absolute percentage error (MAPE) are employed as standard forecast error measures. Furthermore, the percentage of correct models chosen by the automatic forecast model selection is given in the tables.

Set 1: Forecasting without noise (noise factor $a = 0$)

Function	MSE	MAPE (%)	correct model (%)
(1) trend	0,0	0,0	100,0
(2) seasonal	0,0	0,0	100,0
(3) trend-seasonal	0,0	0,0	100,0

Table 1: Forecast results with original historical data (noise level = 0), avg. from 30 runs

Without noise added, automatic model selection chose the correct model in every case.

Set 2: Forecasting with noise factor $a = 1$

Function	MSE (min, max)	MAPE (%) (min, max)	correct model (%)
(1) trend	3,1 (0,0 ; 24,0)	8,5 (0,5 ; 32,5)	80
(2) seasonal	0,4 (0,1 ; 2,3)	19,2 (8,9 ; 63,6)	90
(3) trend-seasonal	0,7 (0,2 ; 1,3)	5,7 (2,8 ; 12,6)	100

Table 2: Forecast results with modified data (noise level = 1), avg. from 30 runs

For the trend function, the system chose 24 times the trend model, 5 times a trend-seasonal and once a constant model. For the seasonal function, the system chose a seasonal model 27 times and a trend-seasonal model 3 times. For the trend-seasonal function, the trend-seasonal model was chosen in each case.

Set 3: Forecasting with noise factor $a = 2$

Function	MSE (min, max)	MAPE (%) (min, max)	correct model (%)
(1) trend	19,3 (0,0 ; 124,3)	25,2 (1,5 ; 73,8)	40
(2) seasonal	1,1 (0,5 ; 2,0)	33,4 (19,2 ; 49,9)	100
(3) trend-seasonal	5,5 (0,7 ; 31,9)	14,4 (5,5 ; 42,0)	80

Table 3: Forecast results with modified data (noise level = 2), avg. from 30 runs

For the trend function, the system chose 12 times a trend, 4 times a trend-seasonal and 14 times a constant model. For the seasonal function, the system chose a seasonal model in all cases. For the trend-seasonal function, the trend-seasonal model was chosen 26 times (with 24 times the correct additive model and two times a multiplicative model), while a seasonal and a constant model were both chosen in two cases each.

Set 4: Forecasting with noise factor $a = 5$

Function	MSE (min, max)	MAPE (%) (min, max)	correct model (%)
(1) trend	44,0 (2,5 ; 168,6)	44,0 (12,6 ; 90,3)	20
(2) seasonal	5,1 (2,9 ; 15,4)	78,2 (55,5 ; 142,4)	27
(3) trend-seasonal	26,3 (2,5 ; 86,3)	35,5 (11,9 ; 76,5)	7

Table 4: Forecast results with modified data (noise level = 5), avg. from 30 runs

For the trend function, the system chose 6 times a trend, once a trend-seasonal, once a seasonal and 22 times a constant model. For the seasonal function, the system chose a seasonal model 8 times and a constant model 22 times. For the trend-seasonal function, the trend-seasonal model was chosen in two cases, a seasonal model was chosen 4 times, a trend model 7 times, and a constant model 17 times.

Discussion

The performance of the automatic model selection is optimal when no noise is added to the underlying patterns. However, this situation never occurs in practical business forecasting as stochastic influence in historical data is abundant. On a general level, the correctness of the automatic model selection with respect to the underlying function pattern and the forecast quality deteriorates with the amount of noise – a result that can be expected. A closer look at the different noise levels and functions, though, reveals

some interesting details.

When the underlying pattern is a trend, the relative performance appears to be the worst. Even with the low noise factor $a = 1$, in 20% of the test cases the wrong model was chosen by the automatic model selection. Moreover, the error measures MSE and MAPE each display results within a wide interval. The maximum MAPE was as high as 32,5 % even at this comparatively low level of noise. This may be interpreted as a significant risk for a bad automatic forecast. The situation is aggravated at higher noise levels.

To a somewhat lesser degree the same applies to the trend-seasonal function, where we have a wide range of MAPE- and MSE-values at the noise level $a = 2$ and higher. Here, two MAPE-results higher than 40% are found, and in three out of 30 cases the MSE is higher than 25, thus again indicating a performance risk for a purely technical forecasting approach.

When analyzing the performance for the seasonal function, one should bare in mind that the MAPE-value depends on the underlying range of correct function values. As these values on average are smaller for the seasonal than for the trend and the trend-seasonal function, a high MAPE should not be overinterpreted in the seasonal case. The data indicates that the MAPE-error is particularly high at lower turning points of the curve. For the seasonal function, the MSE gives a more reliable account of forecast quality. Here, the performance appears reasonable up to a noise factor of $a = 2$.

However, it is worth mentioning that the performance in the seasonal (and the trend-seasonal) case very much depends on the customized seasonal length in the SAP BW forecast profile. We conducted some experiments with slightly wrong seasonal length factors (not documented here for space reasons) that led to very bad results. Even without noise the automatic model selection then generally does not find the correct forecast model. The necessity to enter the correct seasonal length in advance of the forecast limits the practical value of any forecasting tool. This is not an SAP-specific problem, though. One should also consider that in business practice the length of a season may vary over time, once more demonstrating the necessity to closely monitor and supervise automated planning.

For all patterns, the automatic model selection mostly fails at the noise factor $a = 5$, both in terms of the model chosen and the forecast quality achieved. Frequently, the automatic model selection chooses the constant model as no particular pattern can be identified by the system.

Large errors at all noise levels often occur when the automatic model selection chooses a model that does not correspond to the true underlying patterns in the historical data. When classifying this choice as “wrong” prior knowledge of the original function is required that the system does not have. From a purely statistical point of view, based on the available historical data in the individual run, the system’s choice is therefore perfectly justified. This demonstrates the limits of a technical approach to forecasting when really business knowledge is necessary to help deciding the right forecast model, tune it’s parameters or manually adapt the results.

Many companies routinely use automatic forecast model selection, for instance in sales planning, where the number of planning objects is very high. The argument is often that manual planning requires a high effort. A lack of statistical knowledge is another important reason why planners turn to automatic model selection instead of analyzing the data themselves. Our experiment reminds us that a blind trust in the results of a tool can lead to suboptimal business performance.

Conclusion

The results presented here may be interpreted as a warning for practical planners not to automate where their individual business knowledge can help to improve the forecast. Automatic model selection is useful where the variance of historical data around clear statistical patterns is relatively small and the importance of the forecast results is not exceptionally high. In all other cases, the planners knowledge and skills are important input in forecasting. This suggests that some company planners should reconsider their rather unreflective attitude towards automatic model selection in practical forecasting.

References:

- [1] Akaike, H.; Parzen, E. (1998) Selected Papers of Akaike Hirotugu. Springer, New York
- [2] SAP AG (2006) SAP Help Portal: SAP BW BPS Automatic Model Selection, <http://help.sap.com>
- [3] Nissen, V.; Matthesius, M. (2006) Automatic Forecast Model Selection in SAP Business Information Warehouse BPS – an Empirical Investigation, Ilmenauer Beiträge zur Wirtschaftsinformatik, Technical Report No. 2006-04

Author:

Prof. Dr. Volker Nissen
TU Ilmenau, FG Wirtschaftsinformatik für Dienstleistungen,
PF 10 05 65, D-98684 Ilmenau
Phone: +49 (0)3677 69-4047
Fax: +49 (0)3677 69-4219
E-mail: volker.nissen@tu-ilmenau.de

M. J. Kühn / F. Richter / H. Salzwedel

Process simulation for significant efficiency gains in clinical departments – a practical example of a cancer clinic

1 .Introduction

Cost for clinical treatments has been continuously increasing during the last several years. With 10.8% of GNP in 2003 it became the highest in Europe [1]. In order to overcome this problem, in 2004 reimbursements of hospital services based on disease patterns were introduced by the German government [2]. Under these new reimbursement rules hospitals are no longer reimbursed according to the number of days an inpatient is cared for, but on the basis of Diagnosis Related Groups (DRGs). This leads to financial restriction for inefficient hospitals that can only be compensated by better processes and cost reductions. Income and profit of hospitals may rise if treatment cost is reduced and the number of patients is increased. Treatment cost is strongly related to treatment time and number of hospital resources used during this time.

This article describes how treatment cost can be significantly reduced by model based process optimization. Standardized processes models [3] are used for the analysis. This showcase is derived from a highly successful project at a cancer clinic of a German university hospital [4].

2. Selection of simulation software

Processes in hospitals are dynamic over time. Therefore dynamic models are necessary to model these interacting and complex processes. To simulate those dynamic models, appropriate simulation software is required.

Some simulation software (e.g. CHESS [5], MedModel [6]) has been explicitly developed for the usage in hospitals [7]. An in-depth analysis of those simulation software turned out, that those systems can only be used for limited range of use. For simulating hospital processes, event-based simulation systems are required. Highly efficient and robust simulators have been developed to simulate complex communication networks and electronic systems and therefore perform much better than those developed for limited use for some aspects in hospitals. Additionally network simulators support distributed simulation, which may be required for real time hospital process optimization in hospital IT systems.

The chosen simulation system that was able to cover almost all requirements for optimizing hospital processes was MLDesigner [8]. MLDesigner belongs to class 5 simulation systems, which enables all users to rapidly develop a simulation model even without in-depth skills in programming [9].

3. Development and Validation of the simulation model

The methodology used for modeling and simulation in this study follows the steps described by Košturiak/Gregor [10]. The respective clinic covers medical examination and ambulant, chemotherapeutical treatment of cancer patients. Patients are differentiated by those, who gets only a medical examination without a therapy, and those who get examination as well as therapy (=therapy patients). The cancer clinic has an admission, one room for taking blood samples from patients, 3 examination rooms as well as a room for therapies with 24 chairs for patients. The staff includes 5 nurses and 4 medical doctors.

Patients are appointed in the timeframe between 7:30 a.m. and 8:30 a.m. for the current day. Appointments with patients get registered in the used standard software system SAP IS-H MED.

On the day of appointment a blood sample is taken from every patient right after admission. At this point, therapy patients get a venous access for further therapy. When blood analysis results are back from the laboratory, patients go to one of the physicians for medical examination, except patients who have to pass some more preliminary examinations, e.g. computer tomography. All medical findings/results as well as conducted analysis get documented and discarded in the patient documentation.

For therapy patients, the doctor decides, if the patient is able to get a chemotherapeutical therapy or not. In case of a “no go” decision, the patients get a new appointment for another day and are dismissed. If patients are in very bad condition, they will be taken inpatient. In all other cases, the therapy will be administered. Therefore the doctor calls the pharmacy for medication, which has been ordered the day before. Hereupon the pharmacy delivers the medication to the clinic’s facility. To prepare the patients, nursing staff administers the pre-medication. Upon the arrival of medication in the cancer clinic, for each patient a tray with the current medication and the therapy plan gets prepared. The doctor double-checks the tray. Afterwards, nursing staff administers the medication to the respective patient. Depending on the pharmaceutical form and type of medication (e.g. bolus, infusion, cytostatic drug or antibodies) the necessary time for administration is variable. At the end of treatment, every patient gets a discharge note from the medical doctor. If there are no intricacies, the last patient leaves the cancer clinic around 5 p.m. The detailed process was acquired on inductive way [11] and translated into a flow chart [12].

Modeling and simulation of the process with actual data resulted in the following observations: Based on the current process steps the chemotherapeutical therapy never starts before 10:00 o’clock. Till then the 24 chairs for administration of therapy are not used for any therapy purpose, although the first therapy patients arrive at the cancer clinic at 7:30 a.m.. Furthermore we determine that the mean waiting time is very high compared to other comparable clinical facilities. Hence there were a lot of complaints from patients. Waiting time is inefficient time and therefore it must be of interest for the cancer clinic to reduce it. In addition, because of intricacies during therapy or due to patients coming late, the regular closing time often is exceeded. This leads to inefficiency in resource allocation and additional cost.

The evaluation of existing data showed that within a period of three months [13] the clinic in average serves 34 patients a day. 69 percent of the patients got no therapy and 31 percent are therapy patients.

For further analysis, additional data were collected, modeled and validated through simulation. A structured template to capture data was developed. Data were collected

over a period of 55 working days. Afterwards, plausibility and completeness were checked. Based on those data, waiting times between all process steps were computed. Results for waiting time of therapy patients delivered values between 55 minutes and 5:05 hours. The average was 2:45 hours per patient. Patients stayed between 2:10 hours and 9:55 hours at the clinic. The average was 6:08 hours per patient. The ratio of waiting time to patient time was 45 percent.

System modeling has been done with MLDesigner by using the standardized building blocks, developed by Richter to speed up the modeling process [3].

The simulation model is a discrete, event based model. MLDesigner labels this type of system as “DE”. Figure 1 shows the top level of hierarchy (level 1) for this model and its interfaces to the environment. All further system elements are realized as modules (subsystem) or primitives (elements).

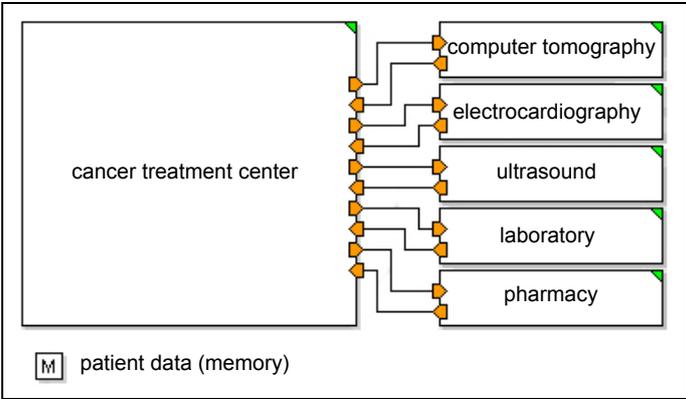


Fig. 1: Simulation model on level 1 – cancer clinic with related medical departments [14]

Relating to the master model designed by Richter [3], the processes of the clinic were organized and implemented with four main areas: admission, examination, therapy and discharge. Figure 2 shows the cancer clinic with the mentioned main areas.

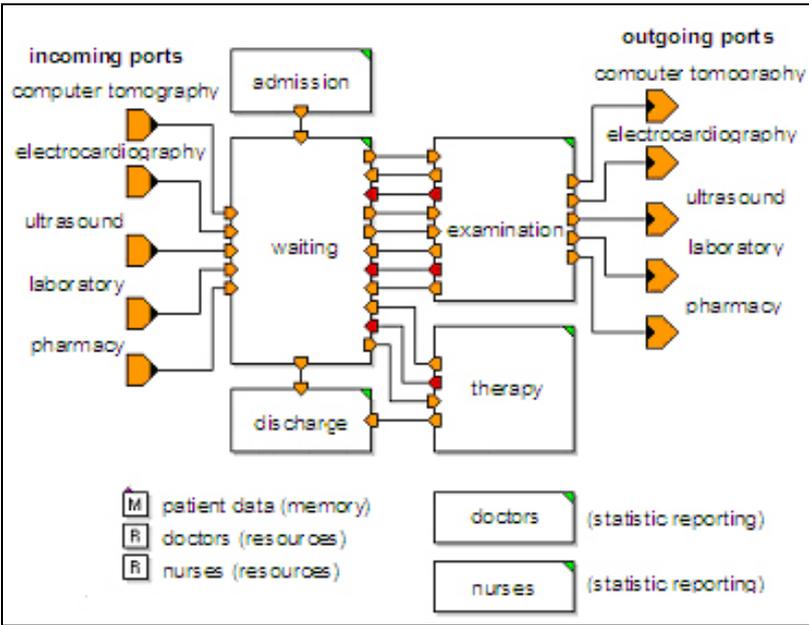


Fig. 2: Simulation model on level 2 – cancer clinic with main areas/units [14]

In the end output functions were implemented in the model to capture simulated data. MLDesigner offers a wide range of prepared functions, which also can be customized individually, e.g. statistics for the usage of resources during the simulation run.

Starting from a comprehensive catalogue of methods for validation and verification of computer models [15], selected methods were chosen to check the developed model. The deviation between model results and real data was less than 4% in average of waiting time all over, which is a good modeling result.

For validation and specially verification purposes, MLDesigner has some very useful functions. In addition to textual information about occurred failures the MLDesigner has a modeling check. Further more the simulation run can be checked with breakpoints and probes while it is running. MLDesigner permit visualization of simulation run in a standardized or user customized way.

4. Optimization of the simulation model

Analyzing simulation results, the model of existing process shows a high potential for reduction of patient time and expected treatment time. Further areas for improvements are integration with related departments like ultrasound, computer tomography or the laboratory. A third idea to optimize the cancer clinic is to check the sequence of process steps [4].

Optimization step 1: Elimination of a redundant process step:

In the past, patients who needed to get a computer tomography became a venous access (canula/needle) in the treatment center (for taking a blood sample) as well as in the radiology department (for application of radiopaque material). In the optimized simulation model, this group of patients gets only one venous access directly in the radiology department. This access will then also be used in the cancer center for taking blood samples. This saves time and increases the quality of treatment by only puncturing those patients one time instead of two times in a row.

In average 15 percent of the patients get a computer tomography. These are 5.10 patients a day based on 34 patients a day in average. The average time to apply a venous access was measured with 10 minutes. Average savings are 51 minutes a day.

Optimization step 2: Optimization of the planning sequence of scheduled patients

In the optimized model, patients were structured as groups according to their need for resources in the clinic. For example patients who will receive chemotherapy will be scheduled as the first group of patients. Patients who will only be examined by a doctor and will not get chemotherapy are scheduled as a second group.

The average reduction of waiting time in the optimized model compared to the as-is process model (after implementation of step 1 and 2) is 34 percent. This corresponds to 54 minutes in average per patient and day.

Optimization step 3: For selected patients, blood samples will be taken the day before their treatment with chemotherapy

For selected patients, blood samples will be taken the day before treatment. This enables the cancer center to receive results from the laboratory and order the chemotherapy from the hospital pharmacy the day before treatment. Those patients can then receive their chemotherapy right after admission to the cancer center.

The average reduction of waiting time in the optimized model compared to the as-is model (after implementation of step 1, 2 and 3) is 44 percent or 69 minutes in average

for each patient.

As result all key figures of the optimized model have resulted in significant improvements. Figure 3 shows the improvements on example of deployment of therapy chairs. Beginning and end of examination as well as beginning and end of therapies were significant earlier as in the existing process model. The distribution of patients over the day is more equal in optimized model compared to as-is model.

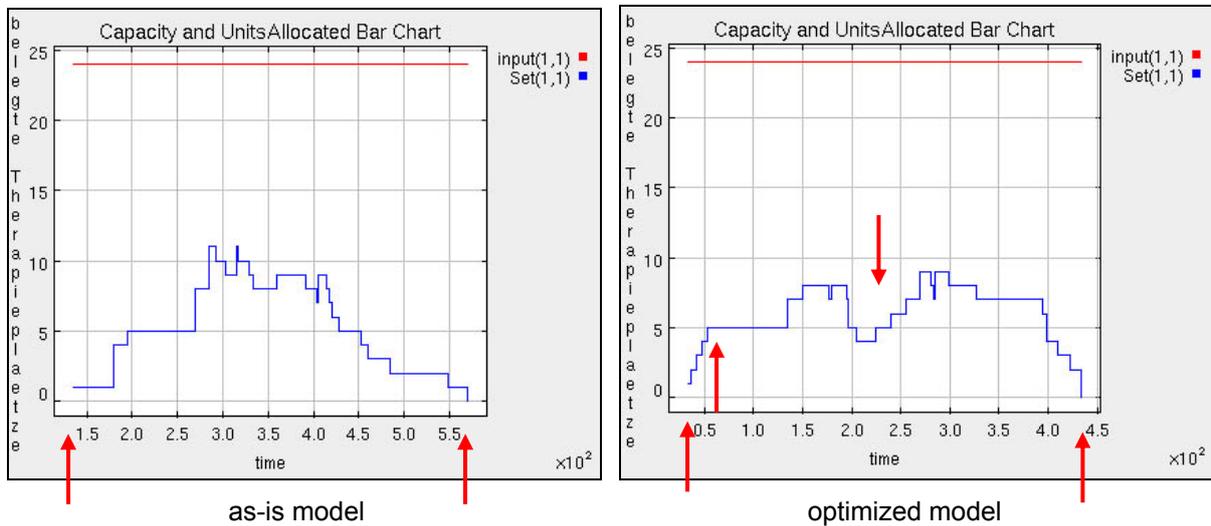


Fig. 3: Resource deployment of therapy chairs [14]

In average, the total waiting time of patient decreased 69 minutes. This equals an improvement of 44% for the optimized model compared to waiting time of as-is model with 158 minutes. Adjusted at calculated model deviation of 4% the total waiting time improves 66 minutes or 42% in result. The main reason for high cumulative waiting time in the existing clinical process was the result of unstructured random patient scheduling.

5. Conclusion

As the result, all project goals could be developed and demonstrated through the simulation model. All process changes have been implemented at the cancer clinic and delivered similar results. Resource utilization increased and opening hours could be reduced. The same number of patients can now be treated with reduced opening hours, less cost, and less effort.

References:

- [1] OECD Health Data 2006 <http://www.gbe-bund.de> - 30.01.2007
- [2] KHEntgG - Gesetz über die Entgelte für voll- und teilstationäre Krankenhausleistungen vom 23.04.2002
- [3] Richter, F.: Entwicklung standardisierter und wieder verwendbarer Komponenten zur effizienten Modellierung und Simulation klinischer Prozesse, Draft copy of dissertation at Technical University Ilmenau 13.06.2007, p. 137-157
- [4] Richter Management Consultants GmbH; internal project review, 2006
- [5] The Clinical Health Economics System Simulation (CHESS) – Department of Internal Medicine, University of Virginia Health System
- [6] MedModel is a product of ProModel Corporation, Utah, USA <http://www.promodel.com>
- [7] José A. Sepúlveda, et al, The Use of Simulation for Process Improvement in a Cancer Treatment Center, Proceedings of the 1999 Winter Simulation Conference, eds. P.A. Farrington, H.B. Nembhard, D.T. Sturrock, and G.W. Evans, p. 1541-1548.
- [8] MLDesigner is a product of MLDesign Technologies, Inc., USA <http://www.mldesigner.de>
- [9] Müller, J.-A.: Simulation ökonomischer Prozesse, Wien 1998, p.164-166
- [10] Košturiak, J., Gregor, M.: Simulation von Produktionssystemen, Wien, New York 1995, p.105
- [11] „Die Induktive Vorgehensweise geht von einer einzelproblemorientierten Formulierung von Prozessen „bottom-up“ aus.“ (vgl. Greiling, M., Hofstetter, J.: Patientenbehandlungspfade optimieren - Prozessmanagement im Krankenhaus, Kulmbach 2002)
- [12] Richter, F.: Entwicklung standardisierter und wieder verwendbarer Komponenten zur effizienten Modellierung und Simulation klinischer Prozesse, Entwurf Dissertation, Ilmenau 2007, p. 228
- [13] Quarter 04/05 – dataextract from SAP IS-H MED
- [14] Matthias Kühn, Simulation und Optimierung einer Tagesklinik – Simulationsstudie mit dem MLDesigner -, Diploma Thesis, Technical University Ilmenau, June 2006,
- [15] Page, B.: Diskrete Simulation: Eine Einführung in Modula-2, Berlin 1991, p.145 ff.

Authors:

Matthias Kühn, MJK Consulting, Westhausen, matthias.kuehn@mjk-cosulting.de
Frank Richter, Richter Management Consultants GmbH, Marloffstein, f.richter@richter-and-partner.com
Prof. Dr. Horst Salzwedel, TU Ilmenau, Fachgebiet SST, horst.salzwedel@tu-ilmenau.de

D. Westermann / M. Kratz / S. Kümmerling / P. Meyer

Architektur eines Simulators für Energie-, Informations- und Kommunikationstechnologien

Energy System Modeling and Simulation

IT basierte Lösungen für Anwendungen der elektrischen Energietechnik ersetzen veraltete Systeme der Schutz- und Leittechnik elektrischer Energiesysteme. Durch die damit einhergehende enge Verknüpfung von Energie, Informations- und Kommunikationstechnologien ergeben sich interdisziplinäre Fragestellungen an die Forschung und Entwicklung. Der Einfluss aktueller Datenübertragungssysteme auf die Prozessführung elektrischer Energieübertragungsnetze wird immer deutlicher, insbesondere dann, wenn Technologien wie „Demand Side Management“ und „Wide Area Measurement and Control“ eingesetzt werden sollen.

Die Untersuchung der auftretenden Phänomene soll über ein einheitliches Simulationssystem erfolgen. Die Anforderungen bestehen darin, dass unter Berücksichtigung der Eigendynamik des Prozesses der elektrischen Energieerzeugung und -übertragung untersucht werden kann, welchen Einfluss die Verarbeitung und Übertragung von Informationen auf das Prozessverhalten hat. Verfahren zur koordinierten Regelung von verteilten, dezentralen Energieeinspeisungen, zur Überwachung von Qualitätsparametern und Dämpfung von Weitbereichsschwingungen (Inter Area Oszillation) können mit einem solchen Simulator entwickelt und untersucht werden. Damit einher gehen Bestrebungen, Operatoren eine Simulationsumgebung zur Verfügung zu stellen, die eine nahezu identische Abbildung der Realität darstellt, um ihre Reaktionen und Entscheidungen in Sondersituationen zu schulen.

Die Liberalisierung der europäischen Energiemärkte hat einen entscheidenden Einfluss auf das Verhalten und die Prozessführung elektrischer Energiesysteme. Seit Ende der 90er Jahre sind die Kunden nicht länger an lokale Stromanbieter gebunden, sondern können ihren Energiebedarf auf dem freien Markt, z.B. über Energiebörsen, wie die EEX, oder den OTC-Markt (Over-The-Counter-Markt), decken [1]. Dadurch entsteht eine räumliche Differenzierung von Erzeugung und Verbrauch elektrischer Energie. Das bedeutet, die Energie, die ein Kunde vor Ort bezieht, wird oft an anderer Stelle im System eingespeist. Es resultieren Energieflüsse, die das Netz in Teilen oder als ganzes überlasten können [2]. Die Struktur der Erzeugung selbst unterliegt starken Veränderungen. Während früher in großen, zentralen Einheiten elektrische Energie erzeugt worden ist, werden heute vermehrt kleine, dezentrale Einheiten verwandt. Dies umfasst sowohl den autonomen, als auch den netzparallelen Betrieb [3]. Durch das Erneuerbare-Energien-Gesetz (EEG) beschleunigt, wird der Einsatz von stochastischen Einspeisungen (z.B. Windkraft, Photovoltaik) forciert. Dem gegenüber steht das Prinzip der bedarfsfolgenden Erzeugung, das mit sich erhöhenden Anteil stochastischer Einspeisungen nicht länger mit gleich bleibender Versorgungssicherheit eingehalten werden kann. Zudem entspricht die vorhandene Infrastruktur elektrischer Energiesysteme nicht den zukünftigen Anforderungen an ein System mit verteilten Einspeisungen nicht deterministischen Charakters [4].

Eine intelligente Netzregelung ist notwendig, um den Ausgleich von Erzeugung und Verbrauch elektrischer Energie gewährleisten zu können [3]. Dies bedingt den Aufbau und die Nutzung moderner Kommunikationssysteme zur Erfassung und Verteilung der für die Netzregelung relevanten Informationen. Die dafür notwendigen Technologien, wie z.B. Phasenwinkelmessgeräte (PMU - **Phasor Measurement Unit**) und Weitbereichsüberwachung (WAMS - **Wide Area Monitoring Systeme**), sind bereits entwickelt. Die in elektrischen Energiesystemen auftretenden Phänomene erstrecken sich über einen großen Zeitbereich von wenigen Millisekunden bis hin zu Stunden und Tagen. Große Entfernungen zwischen Sensoren und Aktoren führen zu übertragungsbedingten Zeitverzögerungen, die beim Entwurf und dem Einsatz von Regelungsverfahren berücksichtigt werden müssen. Insbesondere dann, wenn koordiniert geregelt werden soll, sind auftretende Zeitverzögerungen Störgrößen. Ein Anwendungsfall stellt die gezielte Regelung von Lasten dar (DSM - **Demand Side Management**). Am realen Prozess ist der Entwurf und Test neuer Verfahren im Sinne der Versorgungssicherheit nicht möglich. Deswegen muss es Ziel sein, mittels ge-

eigneter Simulationsplattformen den Prozess möglichst realitätsnah unter Berücksichtigung der Eigendynamik und Kommunikationssysteme nachzubilden. Mit Hilfe eines solchen Simulationssystems können neue Regelungsverfahren entworfen und getestet werden. Zusätzlich besteht die Möglichkeit Operatoren großer und kleiner Systeme in neuen Prozessführungsstrategien zu schulen.

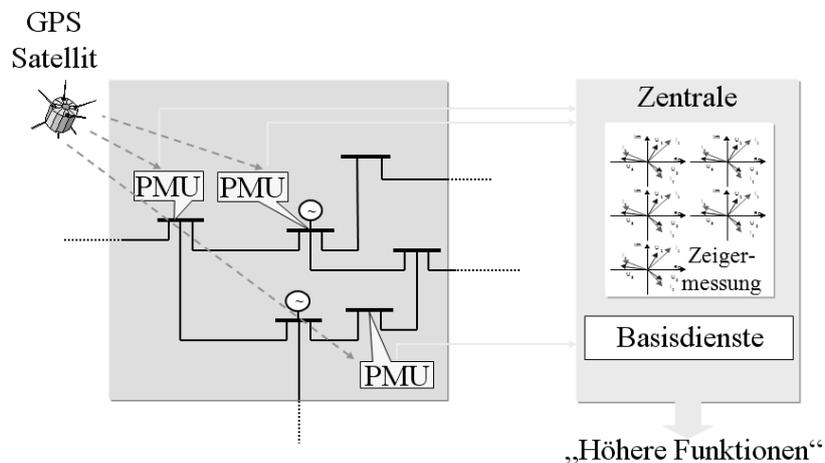


Abbildung 1: Anwendung von PMU für Wide Area Monitoring Systeme [5]

Die zukünftig vermehrt zum Einsatz kommenden Phasenwinkelmessgeräte (PMU) (siehe Abbildung 1), erlauben eine exakte Bestimmung von Betrag und Phase an jedem einzelnen Knoten. Gegenüber der ursprünglichen Erfassung des vollständigen Prozesszustandes mittels State Estimation, ergeben sich wesentlich schnellere und genauere Aktualisierungsraten im Sekundenbereich. Sofern schnelle bzw. zentral gesteuerte Regelungen angewandt werden, die diese Informationen nutzen, wird das Datenübertragungsmedium zur Schwachstelle. Ausfälle in der Erfassung und Übertragung der Prozessgrößen können zu fehlerhaften Regelverhalten führen. Das elektrische Energiesystem ist in diesem Fall von einem funktionierenden Kommunikationssystem abhängig. Diese Effekte sollen mittels der Simulationsumgebung untersucht werden.

Die Digitalisierung der Datenübertragung hat zu einer deutlichen Erhöhung der Anzahl am Markt verfügbarer Angebote von Dienstleistungen geführt. Die Übertragungskosten pro Informationseinheit unterliegen einem kontinuierlichen Rückgang. Damit können Energie bezogene Daten mit geringem Kostenaufwand bei ständiger Verfügbarkeit der Kommunikation übertragen werden. Das Internet als solches ist bereits heute eine umfassende Serviceplattform und übernimmt nach und nach die Rolle, die in der Vergangenheit das leitungsvermittelte Telefonnetz inne hatte [4]. Die

Flexibilität neue Dienste internetbasiert zu entwerfen und implementieren, ist neben den geringen Kosten ein Vorteil. Die Umstellung auf IPv6 erlaubt es künftig, dass jedes elektronische Gerät über eine Verbindung zum Internet verfügen kann. Die Anwendung dieser Technik ist fundamental, möchte man in Zukunft effizient im Energiesystem DSM anwenden, um auch Klein- bzw. Kleinstlasten steuern zu können. Umso mehr ist zu erforschen, inwiefern durch die Übertragung Energie bezogener Daten, die Einbindung der Endgeräte in die Regelung des Energiesystems, durch Ausfall der Kommunikation und Datenverfälschung bzw. -verlust negative Einflüsse auf das Gesamtverhalten zu erwarten sind. Dies gilt vor allem dann, wenn diese Techniken dazu dienen, das Energiesystem nicht nur zu optimieren, sondern auch das dynamische Systemverhalten zu stabilisieren. Ein Simulator, der sowohl Energie-, Leit-, und Kommunikationstechnik integriert, kann wie folgt aufgebaut werden:

- ein Simulationskern für die Nachbildung des elektrischen Energiesystems,
- ein Simulationskern für die Kommunikationsinfrastruktur,
- ein SCADA-System mit Visualisierung.

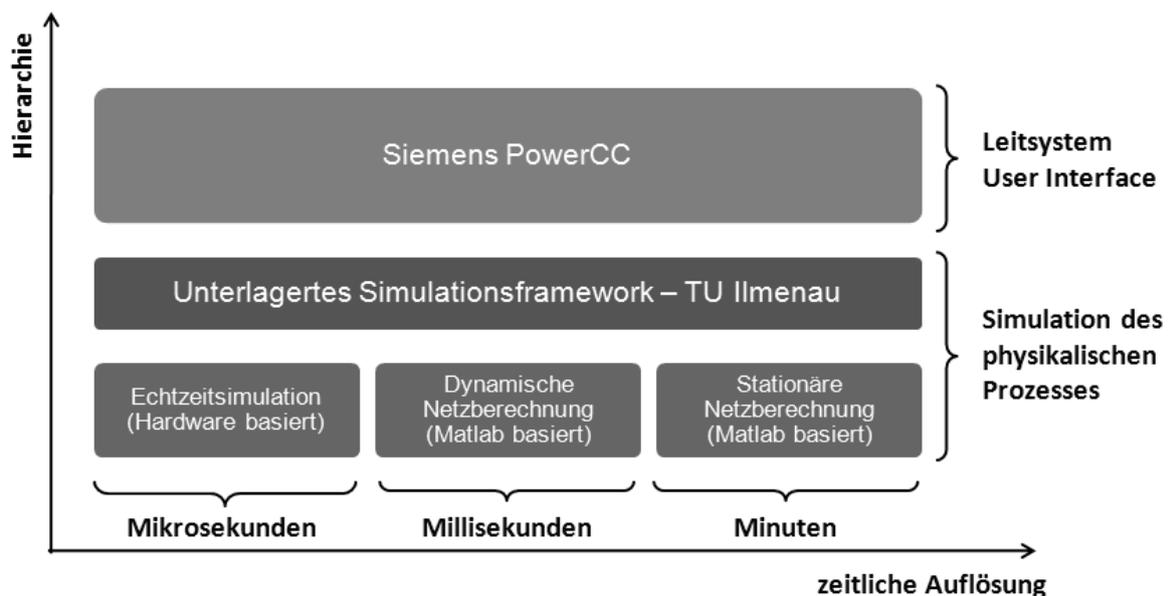


Abbildung 2: Übersicht der Module des Simulators

Je nach Komplexitätsgrad und Genauigkeit der Simulation sind die Simulationskomponenten mit unterschiedlichen Algorithmen zu realisieren. Für einfache Trainings von Operateuren ist eine stationäre Simulation des Prozesses ausreichend. Alle Einschwingvorgänge werden als abgeschlossen betrachtet. Es können z.B. einfache Simulationen durchgeführt werden, die das Verhalten des Operators in Sondersitua-

tionen trainieren, um Blackouts zu vermeiden. Dazu gehören Leitungsausfälle, wegfallende Einspeisungen oder Lasten, überlastete Betriebsmittel bzw. Optimierungsaufgaben.

Sollen Leistungs-Frequenzregelung und Ausgleichvorgänge vom Sekunden- bis in den Minutenbereich simuliert werden, so ist ein dynamischer Simulationskern zu verwenden. Dieser kann mit normalen Workstations auch für große Systeme realisiert werden. Für schnellere Vorgänge ist ein spezieller Simulationsrechner notwendig, der harte Echtzeitbedingungen erfüllt. Alternativ kann das dynamische Netzmodell der TU Ilmenau als reales Systemabbild verwandt werden. Dies ist insbesondere dann wichtig, wenn der Prozess in seiner Gesamtheit simuliert werden soll, um ein realitätsgetreues Abbild zu erhalten.

Die Simulation von Paketverlusten bzw. Ausfall einzelner Kommunikationskomponenten digitaler Datenübertragungssysteme soll der Untersuchung des Einflusses der immer stärkeren Vernetzung und den deutlich höheren Automatisierungsgrad, samt Risiken, dienen. Mittels eines in die Datenübertragungsstrecke eingebundenen Simulators sollen Datenpakete abgefangen, modifiziert oder verzögert werden. Typische TCP/IP basierte Protokolle für elektrische Energiesysteme sind IEC61850 für die Schutz und Leittechnik, das Fernwirkprotokoll IEC60870-5-104 bzw. OPC (OPC - **O**penness, **P**roductivity, **C**ollaboration).

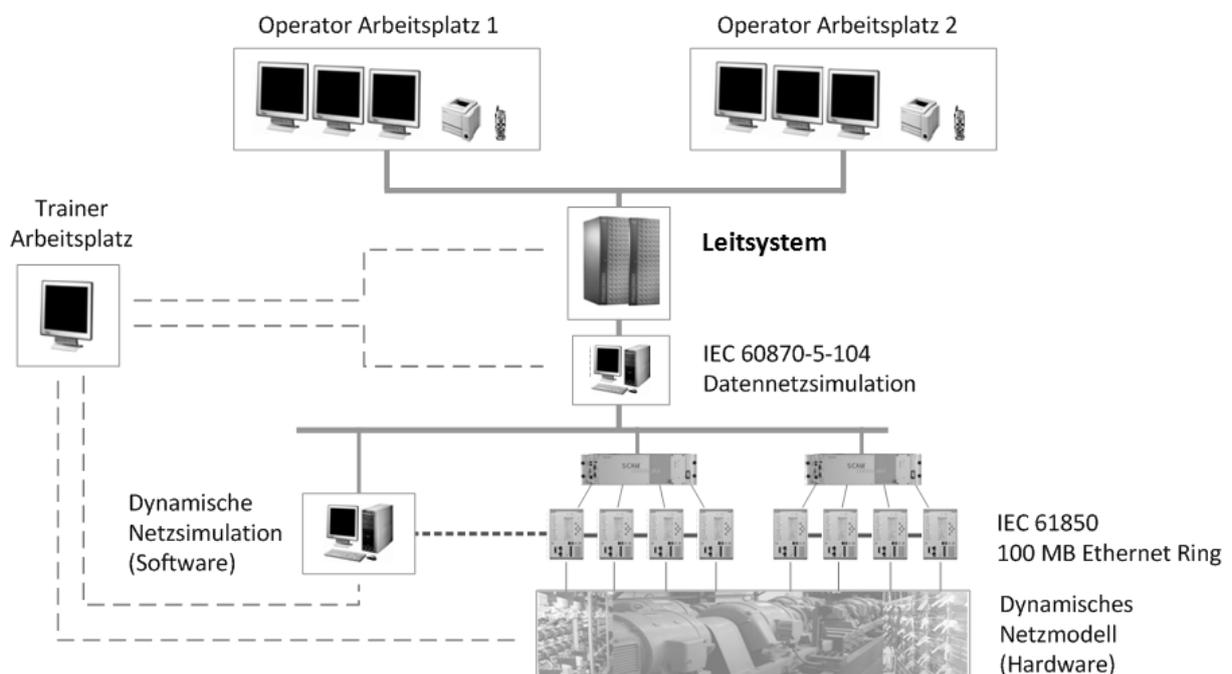


Abbildung 3: Infrastruktur des Simulationssystems mit Trainerarbeitsplatz für Operatortrainings

Die sich daraus ergebende Architektur (siehe Abbildung 3) berücksichtigt sowohl die speziellen Eigenschaften elektrischer Energiesysteme im Kurz- und Mittelzeitbereich, als auch den Einfluss digitaler Datenübertragungssysteme. Der gesamte Simulator ist sowohl für das Operatortraining, die Entwicklung neuer Steuerungen und Regelungen als auch der Untersuchung des Einflusses Internetprotokoll basierter Datenübertragung auf elektrische Energiesysteme geeignet.

Applikationen, wie „Wide Area Measurement and Control“ oder „Demand Side Management“ benötigen digitale Datenübertragungsdienste. Der zunehmende Anteil der informationsverarbeitenden Komponenten kann unter Laborbedingungen nachgestellt werden. Mit Hilfe des Gesamtsystems lassen sich Forschungsfelder erschließen, mit denen aktuelle und zukünftige Fragestellungen beantwortet werden können.

References:

- [1] Kahmann, M.; König, S.: „Wettbewerb im liberalisierten Strommarkt“, ISBN 3-540-66879-9, Springer Verlag Berlin Heidelberg New York, 2000.
- [2] VDN: „Blackouts in Italien, Skandinavien, USA und London“, <http://www.vdn-berlin.de/blackouts2003.asp>, 2003.
- [3] Preiss, O.; Wegmann, A.: „Towards a Composition Model Problem Based on IEC61850“, 4th ICSE Workshop on Component Based Software Engineering, 23th International Conference on Software Engineering, Toronto, May 2001.
- [4] Fritsch, T.: „Der Blackout in Europa – eine weitere unbequeme Wahrheit“, http://www.a-eberle.de/sonderdr/pdf/blackout_in_europa.pdf, 2007.
- [5] Westermann, D.; John A.: „Demand matching wind power generation with Wide Area Measurement and Demand Side Management“, 2006.
- [6] Franz, O., „Potentiale der Informations- und Kommunikations-Technologien zur Optimierung der Energieversorgung und des Energieverbrauchs (eEnergy)“, Oct. 2006.
- [7] Mackiewicz, R. E., „Overview of IEC61850 and benefits“, IEEE Power Engineering Society General Meeting, Jun. 2006.
- [8] Ericsson, G.N., „On Research topics in power system communication“, IEEE Power Engineering Review vol. 21, Oct. 2001.
- [9] Rehtanz, Ch.; Westermann, D., „Wide Area Measurement and Control for Increasing Transmission Capacity in Deregulated Markets, 14th Power System Computation Conference (PSCC), Sevilla, Spain, 2002
- [10] Stubbe, M.; Promel, F.: „Training Operators To Acute System Contingencies“, CIGRE, 2006
- [11] Shahidehpour, M.; Wang, Y.: „Communication and Control in Electric Power Systems“, ISBN 0-471-45325-0, IEEE Press, 2005.

Authors:

Univ.-Prof. Dr.-Ing. Dirk Westermann, Fachgebiet Elektrische Energieversorgung
Dipl.-Ing. Michael Kratz, Fachgebiet Elektrische Energieversorgung
Dipl.-Ing. Stefan Kümmerling, Fachgebiet Elektrische Energieversorgung
cand.-Ing. Patrick Meyer, Fachgebiet Elektrische Energieversorgung

Technische Universität Ilmenau
Gustav-Kirchhoff-Straße 1
98684 Ilmenau

Phone: +49 (3677) 69-1488
Fax: +49 (3677) 69-1496
E-mail: dirk.westermann@tu-ilmenau.de
michael.kratz@tu-ilmenau.de
stefan.kuemmerling@tu-ilmenau.de
patrick.meyer@stud.tu-ilmenau.de

P. Moreno/ D. Westermann/ P. Müller/ F. Büchner

Einsatzoptimierung von dezentralen netzgekoppelten Stromerzeugungsanlagen (DEA) in Verteilnetzen durch Erhöhung des Automatisierungsgrades

Einführung

Konventionelle Versorgungsstrukturen in der elektrischen Energieversorgung (EEV) basieren heute überwiegend auf dem historisch entstandenen Grundprinzip, dass der größte Teil der benötigten elektrischen Energie in zentralen Kraftwerken erzeugt und in Hochspannungsnetzen zur Verfügung gestellt wird. Auf Hoch- und Mittelspannungsebene erfolgt die Übertragung in die Versorgungsbezirke sowie auf Niederspannungsebene die Verteilung zum Endkunden.

Durch den vermehrten Einsatz neuer Energieerzeugungstechnologien auch für eine ökologisch nachhaltige Energieversorgung befindet sich die EEV-Struktur derzeit in einer Wende zu dezentralen Strukturen. In den letzten zwei Jahrzehnten sind dezentrale Energieerzeugungsanlagen (DEA) für den Einsatz im städtischen Versorgungsgebiet zur technologischen „Marktreife“ weiterentwickelt worden. Mit dem Erneuerbare-Energien-Gesetz (EEG), dem Gesetz zur Nutzung umweltfreundlicher Energiequellen, stellt sich die zukünftige Versorgungsstruktur als ein Netzwerk mit einem hohen Anteil an verteilter (dezentraler) Energieerzeugung und somit mit einem höheren Komplexitätsgrad dar. Durch die Ausschöpfung des Kraft-Wärme-Kopplungs-Potentials in öffentlichen Gebäuden und in Ein- und Mehrfamilienhäusern wird eine erhöhte Einspeisung im Niederspannungsnetz zu verzeichnen sein.

Die elektrische Energie von DEA wird hauptsächlich in Verbrauchernähe erzeugt. Im Verteilungsnetz werden derzeit viele Anlagen mit Leistungen von wenigen kW bis einigen 100 kW angeschlossen. Die Mehrzahl dieser Anlagen befindet sich im unteren Leistungsbereich. Derzeit sind vor allem Photovoltaiksysteme und Kraftwärmekopplungsanlagen relevant.

Aus diesem Grund ist in der nahen Zukunft mit einer zunehmenden Verbreitung von DEA im Hausbereich zu rechnen. Der Anteil von DEA, die in das Niederspannungsnetz eingebunden werden, steigt, wobei die maximal mögliche Energie, die als Primärenergie zur Wandlung zur Verfügung steht, ins Netz eingespeist wird. Die derzeit starren Vergütungsmodelle nach dem EEG und Kraft-Wärme-Kopplungs-Gesetz (KWKG) in Deutschland unterstützen diesen Prozess.

Zu welchem Zeitpunkt und in welcher Höhe derzeit die Einspeisung erfolgt, bleibt ähnlich wie beim Energieverbrauch letztlich dem Betreiber vorbehalten bzw. wird vom Wetter und der Tageszeit bestimmt. Die Netzbetreiber sind wegen fehlender Beobachtbar- und Steuerbarkeit der DEA „blind“ bezüglich der Einspeisung auf der Niederspannungsseite [1]. Dies führt zwangsläufig u.a. zu einer Verlagerung der Systemführung und -überwachung auf die unteren Netzebenen und erfordert eine „intelligente“, informationstechnologisch gestützte Einbindung und Steuerung dieser DEA im Rahmen von dezentralen Energieversorgungskonzepten, damit die DEA besser geplant und koordiniert zur Stromerzeugung eingesetzt werden können.

Durch die Netzanbindung von DEA an die Verteilnetze und mit zunehmender Anzahl installierter DEA kommt hier auch immer mehr die Frage auf, wie die vielen Anlagen besser in den Netzbetrieb integriert werden können, um weiterhin hohe Versorgungszuverlässigkeit gewährleisten zu können und einen optimierten Einsatz der verschiedenen DEA-Technologien zu ermöglichen.

Vor diesem Hintergrund sind zur systemkompatiblen Integration von DEA Lösungen in den Bereichen Automatisierung, Information und Kommunikation notwendig, um einen optimierten Einsatz von DEA in Verteilnetzen zu ermöglichen. Angepasste Betriebsführungskonzepte und Zuständigkeiten sind ebenfalls erforderlich, um zu durchgängig technologisch optimierten als auch wirtschaftlich vertretbaren Lösungen zu gelangen. Derzeit werden Energiemanagementkonzepte von Verteilnetzen mit DEA und Lasten entwickelt und der daraus resultierende Kommunikationsbedarf abgeleitet [1]. Diese Energiemanagementkonzepte umfassen im Allgemeinen die Planung, Koordination und die Kontrolle aller Aktivitäten der Verteilnetze mit DEA und haben das Ziel, die Stromerzeugung für den notwendigen Energieeinsatz zu optimieren [2].

Die Zielsetzung besteht darin, zukünftig DEA „intelligent“ mit einer Leitstelle anzuschließen, so dass integrierte, durchgängige Energie- und Informations- sowie Kommunikationsstrukturen einerseits die Optimierung von der Erzeugung über Übertragung und Verteilung bis zum Endverbraucher ermöglichen und andererseits den Betriebszustand der DEA liefern.

Eine Grundvoraussetzung für die Einsatzoptimierung einer großen Anzahl von DEA in Verteilnetzebenen ist die Einbeziehung eines effizienten Kommunikationssystems.

Es stellt sich die Frage, wie DEA in Verteilnetzen sich mit vorhandenen Informations- und Kommunikationstechnologien automatisieren lassen, um eine Einsatzoptimierung von verschiedenen DEA-Technologien zu ermöglichen.

Die vorliegende Veröffentlichung beschreibt die mit dieser Thematik verbundene Automatisierungsmöglichkeit von DEA in Verteilnetzen. Hierzu werden aktuelle Entwicklungen in der Technik verstärkt berücksichtigt.

Ausgangspunkt

Die DEA im Niederspannungsnetz werden derzeit in der Regel ohne Anbindung an ein übergeordnetes Leitsystem oder ein Energiemanagementsystem realisiert. Viele Anlagen haben inzwischen Schnittstellen, die für die Erfassung oder die Fernwartung von Betriebs- und Erzeugungsdaten benutzt werden können. Diese Schnittstellen sind nicht genormt und benutzen unterschiedliche Protokolle. Ein Problem gerade bei kleinen Anlagen ist, dass die Vergütung durch die Erzeugung/Einspeisung im Verhältnis zu den Investitions- und Betriebskosten der Kommunikation gering ausfällt. Praktisch wird bei vielen Anlagen nur die Regelung/Steuerung über die Netzgrößen Spannung und Frequenz vorgenommen, die als Abschalt- und Einschaltkriterium sowie als Synchronisationssignal genutzt werden [3].

Die informationstechnische Einbindung von DEA in eine übergeordnete Leitstelle ist eine Grundvoraussetzung für einen koordinierten, gemeinsamen Betrieb, aus dem ein wirtschaftlicher Nutzen für Anlagenbetreiber und Energieversorger entstehen kann. Hierzu werden grundsätzliche Varianten für die fernwirktechnische Ausrüstung von DEA vorgeschlagen. Im Stadtgebiet existierende DEA werden bisher meistens kommunikationstechnisch nicht in ein übergeordnetes Leitsystem eingebunden.

Es wird von einer Informationsverarbeitung/ -management auf übergeordneter Leitebene ausgegangen. Eine dezentrale Erzeuger-/ Lastoptimierung ist prinzipiell mit dieser Struktur ebenfalls möglich. Die „Intelligenz vor Ort“ entlastet die Kommunikationsverbindungen und die Ressourcen auf Leitebene. Entsprechende Lösungen befinden sich derzeit in Entwicklung, wie z.B: im DINAR-Projekt [4], [5].

Informationsumfang von DEA

Um den Umfang der Anlagenkategorien abzuschätzen, sind die benötigten technischen Informationen einer DEA zu ermitteln. Dabei lassen sich drei Datengruppen unterscheiden: Netzdaten, Anlagendaten und Randinformationen. Die Netzdaten beziehen sich auf den Netzzustand am Verknüpfungspunkt von DEA und werden als Meldungen zwecks Fernüberwachung in die übergeordnete Leitstelle übertragen. Die Anlagendaten beziehen sich auf den Betriebszustand von DEA und werden als Meldungen zwecks Fernüberwachung oder als Befehle zwecks Fernsteuerung in die bzw. aus der übergeordneten Leitstelle übertragen. Randinformationen dienen zur Überwachung der Umgebungsbedingungen von DEA und werden als Meldungen in das übergeordnete Leitsystem übertragen. In der Tabelle 1 sind wichtige Parameter für Netz-, DEA- und Randinformationen zusammengestellt, die entsprechend des Anlagentyps und Automatisierungsart (Fernsteuern/ Fernüberwachen) modifiziert werden können. Informationen der Umwelt können durch kompakte Wetterstationen am Anlagenort (DEA) erfasst werden, die Messwerte zu Temperatur, Windrichtung, Windgeschwindigkeit und Globalstrahlung als analoge Größen ausgeben.

Parameter			
Netzdaten	Anlagendaten		Randinformationen
<u>Meldungen</u>	<u>Meldungen</u>	- Betriebsstunden	<u>Meldungen</u>
- Ströme	- Anlagenkennung (Typ, Ort, Betreiber)	- Sammelstörung	- Temperatur
- Spannungen	- Hand/ Aus/ Automatik	<u>Befehle</u>	- Windgeschwindigkeit
- Frequenz	- Aus/ Standby/ An	- Betriebsartvorgabe	- Globalstrahlung
- Power	- Leistung	- Fahrplanvorgabe	- Status
Quality Werte	- Energiezähler	- Zeitsynchronisation	Wärmespeicher
	- Lastgang		- Sabotagemeldungen
	- Wartungsinformation		

Tabelle 1: Informationsumfang

Klassifizierung der Anlagentypen im Informationsverbund

Der gesamte Informationsumfang und die benötigte Aktualität der Daten ist vom Anlagentyp und dessen Beitrag zur Energieerzeugung abhängig. Dementsprechend kann wie in Tabelle 2 eine Klassifizierung von DEA mit Kommunikationsanforderungen vorgenommen werden.

Klassifizierung DEA mit Kommunikationsanforderung					
Kat.	Leistung [kW]	Funktionen	Parameter	Intervall	Volumen/ Monat
0	< 10	autark, installierte Leistung bekannt	-	-	-
1	10 – 50	Überwachungs-funktionen	Lastgang	monatlich	1 - 100 kB
			Servicemeldungen	(archiviert)	
2	50 – 250	Steuerungs-funktionen	Fahrplandaten	wöchentlich	0.1 – 1 MB
			Lastgangdaten	täglich	
			Störungsmeldungen	spontan	
3	250 – 1000	Regelungs-funktionen I	Soll-, Mess-, Zählwerte	¼ h-Werte	1 – 5 MB
			Stör- und Betriebsmeldungen	spontan	
4	> 1000	Regelungs-funktionen II	Soll-, Mess-, Zählwerte	minütlich	5 – 50 MB
			Stör- und Betriebsmeldungen	spontan	

Tabelle 2: Klassifizierung DEA mit Kommunikationsanforderung

Ausrüstungsgrad von DEA

Um einen Überblick über den Ausrüstungsgrad marktüblicher DEA mit Komponenten der Kommunikationstechnik zu erhalten, wurden Angaben von Herstellern zu Photovoltaik-Wechselrichtern (PV-WR) und Blockheizkraftwerksanlagen (BHKW-Anlagen) im Bezug auf folgenden Punkten analysiert:

- allgemeine Produktinformation (Leistungsklasse, Wirkungsgrad)
- herstellerseitige Möglichkeiten, Betriebsparameter elektronisch zu exportieren
- Schnittstellentypen und verwendete Protokolle (herstellerspezifisch/ genormt)
- Software zur Auswertung / Visualisierung verfügbar
- Konzepte zur Fernüberwachung

Die Eingrenzung auf diese zwei Anlagentypen erfolgte aus folgenden Gründen:

- in städtischen Versorgungsgebieten, wo sich die Mehrheit der Verbraucher befinden, sind derzeit vor allem PV- und BHKW-Anlagen relevant
- Durch die Serienfertigung dieser zwei Anlagentypen ist der Ausrüstungsumfang durch den Hersteller weitestgehend festgeschrieben, und im Einzelfall kaum flexibel zu gestalten. Dagegen sind z.B. Anlagen der Wind- oder Wasserkraft oder große BHKW durch die höheren Leistungsklassen in der Regel mit modular aufgebauten Steuerungen ausgestattet, die eine variable Kommunikationseinbindung ermöglichen.

Der Ausrüstungsumfang von DEA ist sehr unterschiedlich. Zusammenfassend lässt sich feststellen, dass mittlerweile alle BHKW-Anlagen mit einer Form der Fernüberwachung ausgestattet sind, oder diese optional erhältlich ist. Auf dem Markt der PV-WR sind die unterschiedlichsten Lösungen anzutreffen. Die meisten Geräte im unteren Leistungsbereich bieten intern oder extern realisierte Funktionen, die vorrangig zur Visualisierung der Anlage am PC dienen. Standardparameter sind die aktuelle Leistung und die Ausgangsspannung. Standardschnittstelle ist die RS-232, über die verschiedene Protokolle übertragen werden.

Bei PV- und BHKW-Anlagen werden die Meldungen auf unterschiedlichen Wegen ausgegeben. Die folgenden Punkte gelten für beide Bereiche:

- Datenübertragung durch teilweise offenes Herstellerprotokoll
- Visualisierung durch Software des Herstellers
- Visualisierung über Webportal auf Servern des Herstellers
- Fernübertragung über Wähl-/ Standleitung
- Meldungen über SMS / Fax / E-Mail
- Zugriff über Bussysteme (CANBUS, Profibus)
- Excel-kompatible Formatierung, meist erst händisch über Herstellersoftware zu exportieren
- Wertausgabe über ASCII-Files

Ein gemeinsamer, einheitlicher Anknüpfungspunkt für eine Datenfernübertragung besteht somit bei beiden Anlagenformen nicht. Bei Anlagen mit standardisierten Bussystemen, bzw. offen gelegten Herstellerprotokollen ist eine Fernwirkeinbindung möglich, setzt aber einen anlagenspezifischen Engineeringaufwand voraus. Dies ist vor allem bei großen PV-WR/ BHKW-Anlagen realisierbar, deren Einbindung in die Verteilnetze aufgrund ihrer Leistungsklasse als sehr effektiv angesehen werden kann.

Hardware-Ausrüstung

Das Hardwarekonzept basiert auf verfügbaren Systemkomponenten der Fernwirktechnik. Aufgrund der herstellerseitig unterschiedlichen Ausrüstungsumfänge und Ausbaugrade von DEA mit Kommunikationstechnik und Schnittstellen, ist ein einheitliches Fernwirksystem zu empfehlen.

Durch die Verwendung gleicher Komponenten entstehen folgende Vorteile:

- keine Eingriffe in Software der Kundenanlage notwendig
- Meldungsübernahme über Standardschnittstellen (Stromschleife/ Spannungspegel/ Binärkontakte)
- Nur wenige Ersatzmodule brauchen vorgehalten zu werden
- Funktionalität für alle Anlagen gleichermaßen bekannt
- Engineeringleistungen zur Ankopplung fremder Systeme sind gering
- Schnellere Fehlersuche durch einheitliches System
- Geringer Einweisungsumfang des Wartungspersonals

Entsprechend der dezentralen Erzeugerart werden zwei Ausrüstungsvarianten vorgeschlagen, die über unterschiedliche Übertragungswege eingebunden werden können.

Anlagenausrüstung für Überwachungsfunktion

Die Anlagen der Kategorie 1 im Kleinleistungsbereich, deren Einspeisung nur überwacht werden soll, können mit kleinen Fernwirkgeräten eingebunden werden. Messumformer können die Netzspannung messen und wandeln den Wert in ein Stromausgangssignal (z.B. 0 ... 20 mA) um.

Die Einspeise- oder Bezugsleistung kann durch Energieimpulse der Zähler erfasst werden. Die DEA muss dazu mit geeigneten Zählern ausgerüstet sein. Die Fernwirkeinbindung erfolgt z.B. durch ein externes Modem.

Anlagenausrüstung für Steuerung und Regelung

Für Anlagen der Kategorie 2 bis 4, die einen Ferneingriff gestatten, können mittels eines größeren Fernwirkgerätes eingebunden werden.

Messumformer können elektrische Kenngrößen auf Mittel- und Niederspannungsebene erfassen. Die Messgrößen können über eine Schnittstelle an das Fernwirkgerät mittels Telegramm übermittelt werden. Weiterhin können Messwerte als Analogsignal direkt ausgegeben werden und ein Binärausgang als Zähler oder Grenzwertmelder parametrisiert werden. Die Strom- und Spannungseingänge können bis 10 A bzw. 450 V ohne Wandler belegt werden. Die Frequenz kann z.B. am Spannungseingang L1 gemessen werden.

Knotenstation

Die Erfassung der genannten Anlagenparameter führt bei einer Vielzahl verteilter Erzeugeranlagen zu einem umfangreichen Datenvolumen. Aus diesem Grund ist aus derzeitiger Sicht eine Baumstruktur sinnvoll, in der die Informationen von DEA eines Gebietes in Knotenstationen zusammengefasst werden.

Die Funktionalität der Knotenstationen kann durch einen entsprechend ausgestatteten Industrierechner realisiert werden. Durch ein Windows Betriebssystem können auch herstellerspezifische Software zur Auswertung von Anlagendaten installiert werden.

Kommunikationsstruktur

Eine übergeordnete DEA-Leitstelle verwaltet die verschiedenen Knotenstationen und stellt relevante Betriebsdaten der Netzleitstelle zur Verfügung. In der Abbildung 1 sind die Ebenen mit bidirektionalen und unidirektionalen Kommunikationsbeziehungen dargestellt.

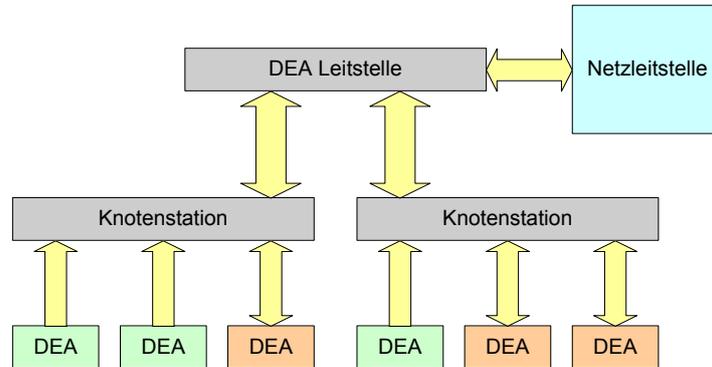


Abbildung 1: Kommunikationsstruktur verteilter Anlagen

Die Verarbeitung des Informationsumfanges beinhaltet die in Tabelle 3 beschriebenen Aufgaben, die auf den unterschiedlichen Ebenen bewältigt werden. Die geographische Anordnung der Knotenstationen ist von den verwendeten Datenfernübertragungsarten abhängig. Bei Verbindungen über öffentliche Übertragungsnetze können die Knotenstationen nahe der Leitstelle eingerichtet werden. Es kann auch eine Einbindung durch bereits fernwirktechnisch erschlossene Mittelspannungsstationen erfolgen, bzw. in diesem Zusammenhang eine Automatisierung von Umspannstationen in Betracht gezogen werden. Die gemeinsame Nutzung der Datenverbindung in die Leitstelle erhöht die Wirtschaftlichkeit der Verbindung.

Ebenen der Datenverarbeitung	
Aufgabe	Ebene
Korrekte Datenerfassung, Zwischenspeicherung, Zeitstempelung und Aufbereitung von Messwerten	Unterstation
Bewertung der einzelnen Meldungen und Einordnung in Prioritätsgruppen	Knotenstation
Zusammenfassung zu Gruppeninformationen, z.B. nach Erzeugertechnologie und Weitergabe an Netzleitstelle	DEA – Leitstelle
Ausgabe von Steuerbefehlen auf Grundlage eines Erzeuger-/Lastmanagements	DEA – Leitstelle

Tabelle 3: Ebenen der Datenverarbeitung

Datenübertragungsarten

Grundsätzlich gelten folgende Anforderungen an ein Kommunikationssystem zur Betriebsführung verteilter Energieerzeuger [6]:

- das System muss ausreichend leistungsfähig und zuverlässig sein
- Gewährleistung der Flexibilität, Offenheit aber auch Sicherheit der Information
- Standardisierte Strukturelemente
- Kostengünstig und zukunftssicher

Durch den Einsatz einheitlicher genormter Fernwirkprotokolle, wie z.B. IEC 61850 und 60870-5-101/104, wird die Einbindung in die Überwachung konventioneller Energieversorgungssysteme erleichtert. Die verschiedenen Möglichkeiten der Datenfernübertragung sollen hier kurz genannt werden:

- PSTN: Public Switched Telephone Network (analoge Wählverbindung)
- ISDN: Integrated Services Digital Network (digitale Datenübertragung)
- GSM: Global System for Mobile Communication
- GPRS: General Packet Radio Service (paketvermittelnder Dienst)
- Zeitschlzifunk (nicht öffentlicher Funkdienst)
- PLC: Powerline Communication (Datenübertragung über Stromnetz)
- Optischer Rundfunk (Informationsübertragung über modulierte Lichtsignale)
- Ethernet (kabelgebundene Datennetztechnologie für lokale Datennetze)

Prinzipiell kann durch moderne Technologien, bei entsprechendem technischen Aufwand, jede Information zu jeder Zeit überall zur Verfügung gestellt werden. Die Auswahl der geeigneten Technologie für die Fernwirkaufgaben bei verteilten

Energieerzeugern ist abhängig von folgenden Punkten: örtliche Gegebenheiten, Datenumfang und Aktualisierungszyklus.

Für die Übermittlung von Meldungen, Befehlen, Zählwerten, usw. sind Datenübertragungsraten bis 30 kbit/s, maximal 64 kbit/s ausreichend. Dies ist bereits mit herkömmlichen Modems bzw. einem ISDN-Kanal abzudecken.

Sollen dem Anwender In Zukunft gleichzeitig auch Mehrwertdienste wie Telefon, Internet oder Fernsehen angeboten werden, sind entsprechend breitbandige Datenkanäle vorzusehen, die prinzipiell technisch zur Verfügung stehen.

Hardwareausrüstung Kommunikation

Um die beschriebenen Datenübertragungswege beispielhaft darzustellen, können marktübliche Hardwarekomponenten (siehe Tabelle 4) ausgewählt werden.

Ausrüstung Kommunikation		
Verbindung	Hersteller / Typ	Eigenschaften
PSTN	Westermo TD-32B	Wähl- und Standleitungsmodem 33,6 kbit/s, RS-232
ISDN	Westermo ID-90	ISDN-Modem 128 kbit/s, vier Alarmeingänge mit SMS-Funktion, Rückruffunktion
GSM/ GPRS	Westermo GD-01	850/ 900/ 1800/ 1900 MHz; Datenrate 14,4 kbit/s; RS-232-Schnittstelle
Zeitschlitz- funk	Panasonic TRM710H	Zeitschlitzfunkmodem für nichtöffentlichen Datenfunk; Frequenzen und Zeitschlitz per DIP Schalter programmierbar; Funkuhr
PLC	Siemens DCS3000	Powerline Modem für 28,8 kbit/s mit kapazitiver Ankoppeleinheit
Optischer Rundfunk	Optel OPTICOMM	Reichweite bis 4.000 m möglich; Multimode-LWL-Anbindung bis 155 Mbit/s; für LAN oder Telekommunikation

Tabelle 4: Ausrüstung Kommunikation

Ein beispielhaftes Kommunikationsnetzwerk zeigt Abbildung 2:

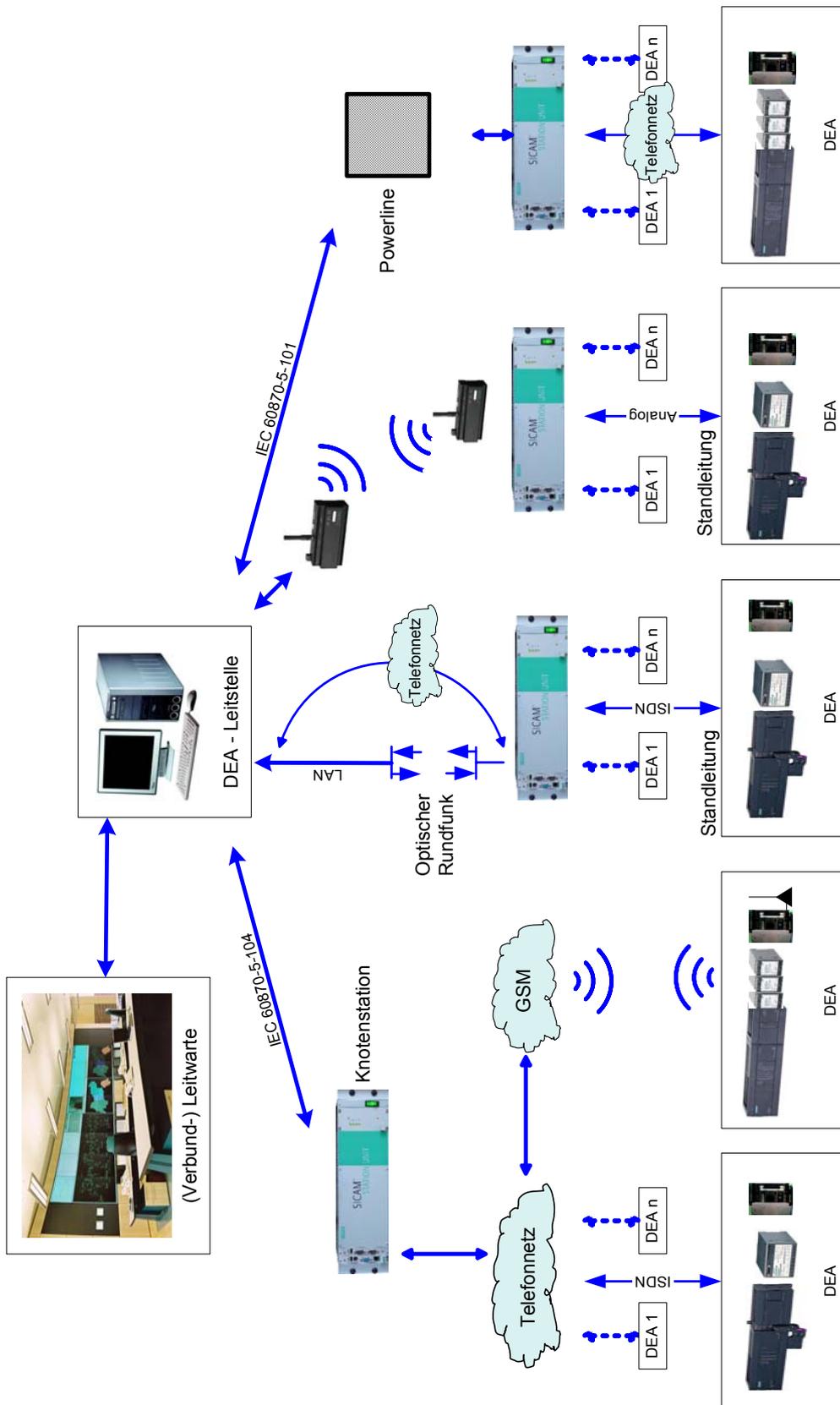


Abbildung 2: Beispiel-Datenübertragungsnetz

Zusammenfassend lässt sich feststellen, dass bereits heute und mit derzeitigen Informations- und Kommunikationstechnologien eine Automatisierung von Verteilnetzen mit DEA erfolgen kann. Dies ist die Grundlage, um den Einsatz von DEA in Energieversorgungsnetzen zu optimieren. Durch die Einbindung von DEA in einem umfassenden Kommunikationssystem eröffnen sich für alle Marktteilnehmer weitere Wertschöpfungspotentiale für zahlreiche Dienstleistungen.

References:

- [1] Bendel, C.; Nestle, D.: Bidirektionales dezentrales Energiemanagement im Niederspannungsnetz auf Basis zentraler und dezentraler Informationen, 11. Kasseler Symposium Energie-Systemtechnik 2006
- [2] Schlebusch, V.; Wolff, M.: Energiemanagement in Verteilnetzen mit hohem Anteil an dezentralen Erzeugungsanlagen, 11. Kasseler Symposium Energie-Systemtechnik 2006
- [3] Malcher, S.; Jahn, J.: Dezentrale Energieerzeugung, Weiterentwicklung von der reinen Einspeisung zum intelligenten System, 11. Kasseler Symposium Energie-Systemtechnik 2006
- [4] EUS GmbH, ISET: Dezentrale Energieversorgungsanlagen DINAR I, Dortmund Kassel 2004, Abschlussbericht
- [5] EUS GmbH: Dezentrale Energieversorgungsanlagen DINAR II, Dortmund 2005, Zwischenbericht, Forschungskennzeichen 0329900E
- [6] Englert, T.: IDS GmbH Ettlingen, Unterlagen Stadtwerke Unna
- [7] wik-Consult – FhG Verbund Energie: Potentiale der Informations- und Kommunikationstechnologien zur Optimierung der Energieversorgung und des Energieverbrauchs (eEnergy), Bad Honnef, Dezember 2006

Authors:

Univ.-Prof. Dr.-Ing. Dirk Westermann
M.-Ing. Pedro Moreno
Technische Universität Ilmenau, Gustav-Kirchhoff-Str. 1
98684, Ilmenau
Phone: +49 (03677) 69-2838
Fax: +49 (03677) 69-1496
E-mail: dirk.westermann@tu-ilmenau.de, pedro.moreno@web.de

Dr.-Ing. Peter Müller
Stadtwerke Leipzig GmbH, Arno-Nitzsche-Str. 35
04277, Leipzig
Phone: +49 (0341) 121 3900
Fax: +49 (0341) 121 3673
E-mail: peter.mueller@swl.de

Dr.-Ing. Frank Büchner
Siemens AG RD OST BLN PTD, Nonnendammallee 101
13629, Berlin
Phone: +49 (030) 386 35600
Fax: +49 (030) 386 35802
E-mail: frank.buechner@siemens.com

M. Heit / S. Rozhenko/ M. Kryvenka/ D. Westermann

Mathematische Bewertung von Engpass-Situationen in Transportnetzen elektrischer Energie mittels lastfluss-basierter Auktion

3.6 Energy System Modelling and Simulation

Vor ca. einen Jahrzehnt begann die Liberalisierung des Europäischen Energiemarktes. Durch die Osterweiterung der Europäischen Union vergrößert sich der transkontinentale Energiemarkt. Verbunden mit der politischen Forderung nach Klimaschutz und Energieeinsparung steigt die Anzahl dezentraler regenerativer Energiequellen mit meist nicht-deterministischer Erzeugungscharakteristik. Vor allem an Ländergrenzen und Zonen mit unterschiedlichen Preisstrukturen treten verstärkt wirtschaftliche und technischen Engpässen an den infrastrukturellen Transportnetzen auf. Besonders stark ist das im Bereich der elektrischen Energie zu erkennen, da hier der Liberalisierungsprozesse den größten Fortschritt zeigt. Verbunden mit dem sicherheitspolitischen Risiko eines elektrischen Blackouts werden hohe Anforderungen an die physikalische Stabilität der elektrischen Energiesysteme gestellt.

Deshalb ist es dringend erforderlich, diese Engpässe technisch zu überwachen und wirtschaftlich zu vermarkten. Das volkswirtschaftliche Potential der Engpässe ist zu nutzen und durch entsprechende Marktsignale ist ein Anreiz zum Ausbau der Transportnetze zu schaffen.

Bisher erfolgte das Engpassmanagement durch NTC (Net Transfer Capacity)-basierte explizite oder implizite Auktionsverfahren. Diese weisen einige Nachteile auf, die vor allem in einer mangelnden Markttransparenz zu sehen sind.

In diesem Paper wird ein lastflussbasiertes Allokationsmodell (LFBA) vorgestellt. Die Grundlage dafür bilden die BC (Border Capacity) und der PTDF (Power Transfer Distribution Factor). Hierbei werden die real physikalisch auftretenden Stromflüsse für die Engpassallokation zu Grunde gelegt.

Beide Verfahren werden hinsichtlich ihrer Vor- und Nachteile mit einer Beispielrechnung gegenüber gestellt. Es zeigt sich außerdem, dass für zukünftige Arbeiten der Energiemärkte verstärkt zu involvieren ist.

Im Jahre 1998 begann in Europa die Liberalisierung des Europäischen Strommarktes auf der Grundlage der EU-Richtlinie 2003/54/EG. Damit wurde die Grundlage für einen internationalen und transkontinentalen Stromhandel geschaffen, der sich seit dem zunehmend etabliert.

Davon ausgehend bildeten sich zahlreiche Binnenmärkte, die in der Regel an geografische bzw. transporttechnische Grenzen gebunden sind. Meistens liegen in den jeweiligen Binnenmärkten die Strompreise auf einem unterschiedlichen Niveau. Neben abweichenden Angebots- und Nachfragesituationen sind diese Preisdifferenzen eine wesentliche Motivation für den grenzüberschreitenden Stromhandel.

An den Grenzen der Binnenmärkte sind die Transportnetze für die elektrische Energie meistens nur schwach ausgebaut. Die Folge sind technische Engpässe an den Übergangspunkten. Um diese Engpässe wirtschaftlich zu vermarkten und aus technischer Sicht eine höhere Betriebssicherheit zu gewährleisten, wurde das Engpassmanagement [1] eingeführt, bei dem die entsprechenden Übertragungskapazitäten verauktioniert werden.

Bisher werden im Europäischen Strommarkt meistens ein explizites oder implizites Auktionsverfahren bzw. zukünftig auch Verfahren nach dem Prinzip des Market-Couplings eingesetzt [2].

Die Grundlage für die Auktion bildet immer eine vorher bestimmte zur Verfügung stehende Übertragungskapazität (NTC – Net Transfer Capacity) [3]. Dabei ist keine Übereinstimmung von physikalischen und finanziellen Flüssen gewährleistet.

Nachfolgend soll die Anwendung einer lastflussbasierten Allokation (LFBA) als eine Möglichkeit zum Engpassmanagement vorgestellt werden. Abbildung 1 zeigt die Einteilung der Verfahren. An einem Beispiel wird die Simulation der LFBA unter Anwendung geeigneter Methoden, an einem Testsystem dargestellt.

Wichtigste Ausgangsparameter zur Durchführung der LFBA sind die Border Capacity (BC) und der Power Transfer Distribution Factor (PTDF). Die BC gibt dabei die zur Verfügung stehende technisch zulässige Übertragungskapazität wieder und der PTDF repräsentiert den Einfluss einer Energietransaktion zwischen zwei Binnenmärkten. Als Binnenmärkte werden dabei die geografischen Regionen der Übertragungsnetzbetreiber (TSO) betrachtet. Die einzelnen Parameter werden als eine PTDF-Matrix dargestellt und sind den entsprechenden Kuppelleitungen zugeordnet.

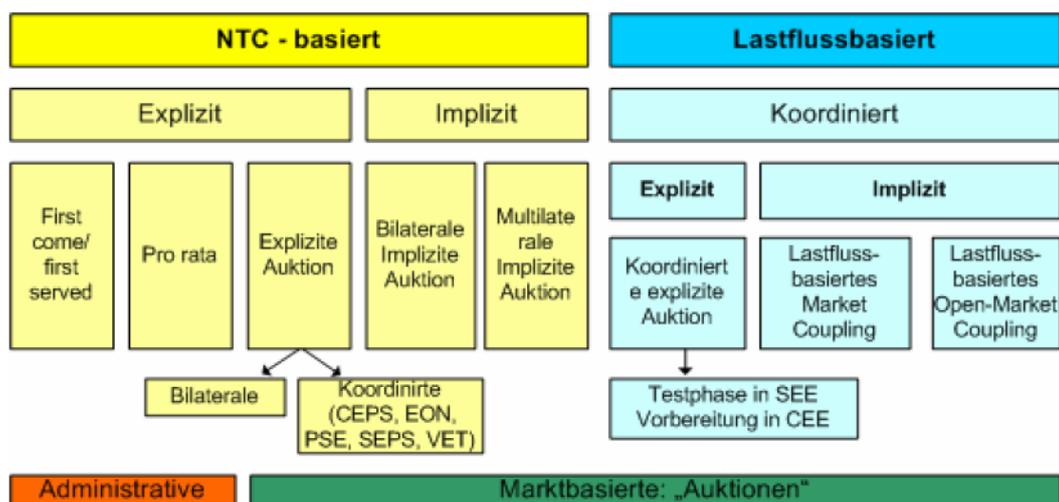


Abbildung 1 Einteilung der Verfahren zum Engpassmanagement [4]

Die Abbildung 2 verdeutlicht die Problematik der realen Lastverteilung beim Engpassmanagement. Die Skizze zeigt fünf angrenzende Versorgungsgebiete, die einer Verbundsituation in Europa entsprechen könnten.

Als Resultat eines Stromhandelsgeschäftes wird von einem Energietransport aus dem Gebiet C in das Gebiet E mit einer Höhe von 100 MW ausgegangen. Der Energiefluss wird hier als TTC (Total Transfer Capacity) dargestellt. Auf Grund der physikalischen Lastverteilung im elektrischen Netz nach dem Kirchhoffschen Gesetz fließt ein bemerkenswerter Teil der elektrischen Energie über die Nachbarländer.

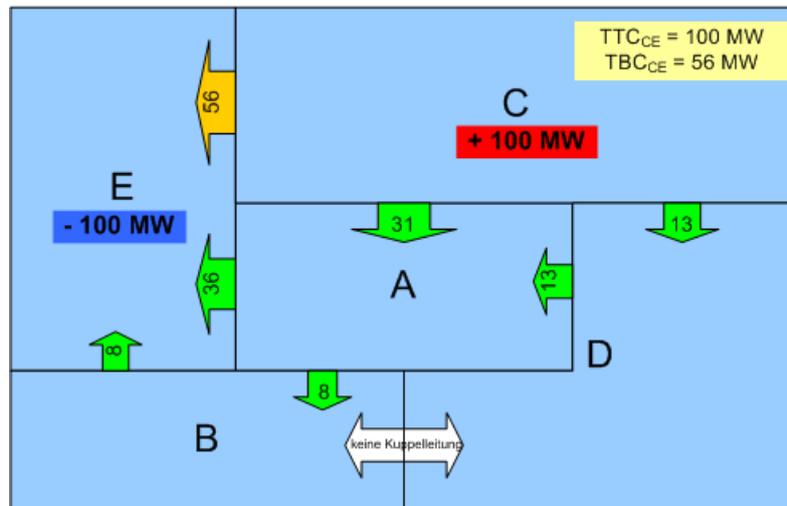


Abbildung 2 Finanzielle und physische Lastverteilung in einer Verbundsituation

Der reale Stromfluss, auch TBC (Total Border Capacity) genannt, beträgt nur noch 56 MW und es stellen sich nicht beachtete Leistungsflüsse zwischen 8 MW und 31 MW über die anderen Grenzen ein.

Wesentliche Konsequenzen durch die Anwendung des bisherigen Verfahrens sind:

- Jede Energietransaktion beeinflusst die Lastflüsse an allen Grenzen.
- Der direkte Lastfluss über eine Grenze stellt nicht den gesamten Energieaustausch zwischen zwei Versorgungsgebieten dar.
- Bei der Berechnung einer TC (Transfer Capacity) ist immer die gesamte Verbundsituation zu betrachten.
- Die TC werden für verschiedene Transportrichtungen getrennt berechnet. Die Anwendung eines Nettings (Saldierung von gegenläufigen Lastflüssen) wird nicht berücksichtigt.

Als Konsequenz werden die physikalisch gegebenen TC finanziell nicht vollständig ausgenutzt und es treten volkswirtschaftliche Verluste aus. Außerdem kann es zu Überlastungen im Verbundsystem kommen und die Netzstabilität ist nicht mehr gewährleistet.

Die lastfußbasierte Allokation hingegen betrachtet die reale Lastflussverteilung im Netz. Bei diesem Verfahren werden alle Grenzen gleichzeitig bewirtschaftet.

Das vorstehend genannte Verfahren wird zurzeit als internationales Versuchsprojekt zwischen fünf benachbarten TSO (Transmission System Operator) durchgeführt. Das Auktionsmodell wird als explizite koordinierte NTC-basierte Auktion umgesetzt.

Die NTC ergibt sich aus der TTC abzüglich einer Sicherheits- und Zuverlässigkeitsmarge, auch TRM (Transmission Reliability Margin) genannt. Die Grundlage ihrer Bestimmung ist die Einhaltung der (n-1)-Sicherheit und ist im Transmission Code [6] eines jeden Landes geregelt.

Die TTC eines jeden TSO wird mit Hilfe der Lastberechnung auf der Grundlage eines Netzmodells ermittelt. Die daraus ermittelten NTC der einzelnen Verbundleitungen werden immer zwischen zwei benachbarten TSO zu einer Grenzkapazität, auch BC (Border Capacity) genannt, zusammengefasst. Durch die BC sind auch die jeweiligen Netzengpässe definiert.

Der zweite Eingangsparameter ist der PTFD. Er gibt das Verhältnis zwischen dem jeweiligen physischen Lastfluss und dem gesamten kommerziellen Energiefluss zwischen zwei TSO wieder.

Die Ermittlung des PTDF erfolgt ebenfalls mit einem Netzmodell, wie vereinfacht in der Abbildung 3 dargestellt ist.

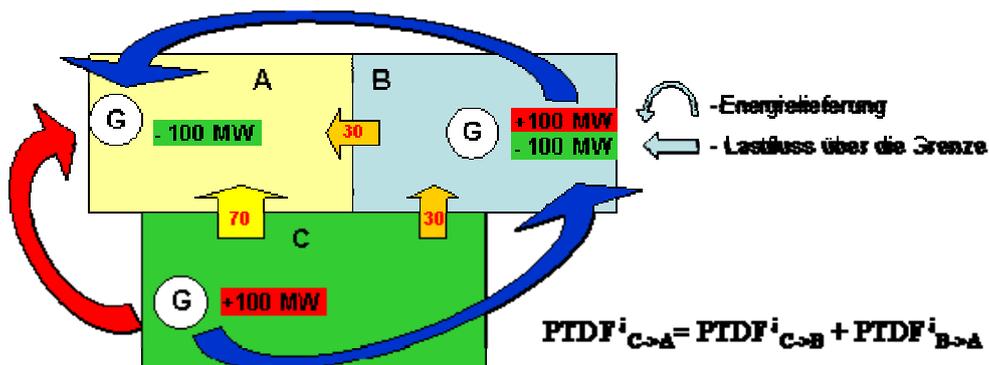


Abbildung 3 Vereinfachte Darstellung der Methode zur PTDF-Bestimmung

Dabei wird im Wesentlichen bei der Simulation die Kraftwerkseinspeisung in der Zone A um 100 MW reduziert und analog in der Zone B um den gleichen Betrag erhöht. Der PTDF-Faktor ergibt sich aus der Änderung des Energieflusses an der betrachteten Grenze in Bezug auf die Energietransaktion von A nach B.

Die ermittelten PTDF können zur besseren Übersicht in einer PTDF-Matrix analog Tabelle 1 dargestellt werden.

Tabelle 1 Beispiel einer PTDF-Matrix

Transaktion			Elektrische Grenze						
Nr.	von	nach	A_C	A_D	A_E	B_A	C_D	C_E	B_E
1	A	B
...	
12	C	E	-0,31	-0,13	0,36	-0,08	0,13	0,56	0,08
...	
19	E	C	0,31	0,13	-0,36	0,08	-0,13	-0,56	-0,08
20	E	D

Die Gebote werden durch die einzelnen Auktionsteilnehmer vorgegeben und es wird vorausgesetzt, dass die Gebote die entsprechenden Marktverhältnisse wieder geben und nur die Regeln für die Engpassbewirtschaftung geändert werden. Die eigentliche Simulation der Engpassvergabe wird mit einem Softwaretool durchgeführt, das im Rahmen der Kooperation mehrerer Energieversorger entstanden ist.

Abbildung 4 zeigt die vergebenen Kapazitäten und die zugehörigen Clearingpreise einer NTC-basierten Auktion, wie sie zurzeit im Versorgungsgebiet durchgeführt wird.

Diese Daten sind online verfügbar unter: www.e-trance.biz.

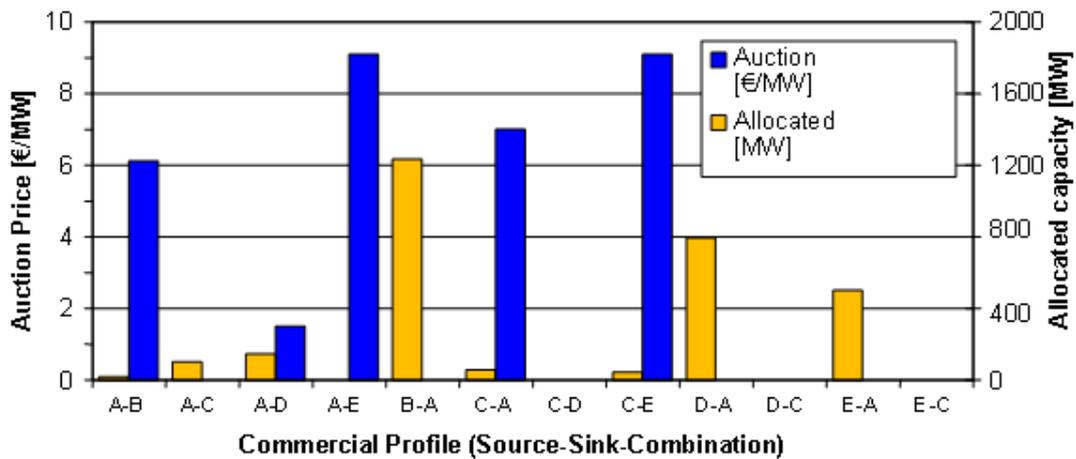


Abbildung 4 Vergebene Kapazitäten und Clearingpreise der NTC-basierten Auktion

Bei der Verwendung des konventionellen Verfahrens ist zu erkennen, dass an vier finanziellen Profilen keine Kapazitätsvergabe erfolgt. Es treten dem zu Folge Engpässe auf.

Im Vergleich dazu zeigt die Abbildung 5 die Resultate der Auktion bei der LFBA. Hierbei wird auch der Einfluss der Nettings betrachtet und es treten nur an drei Übergangsstellen Engpässe auf.

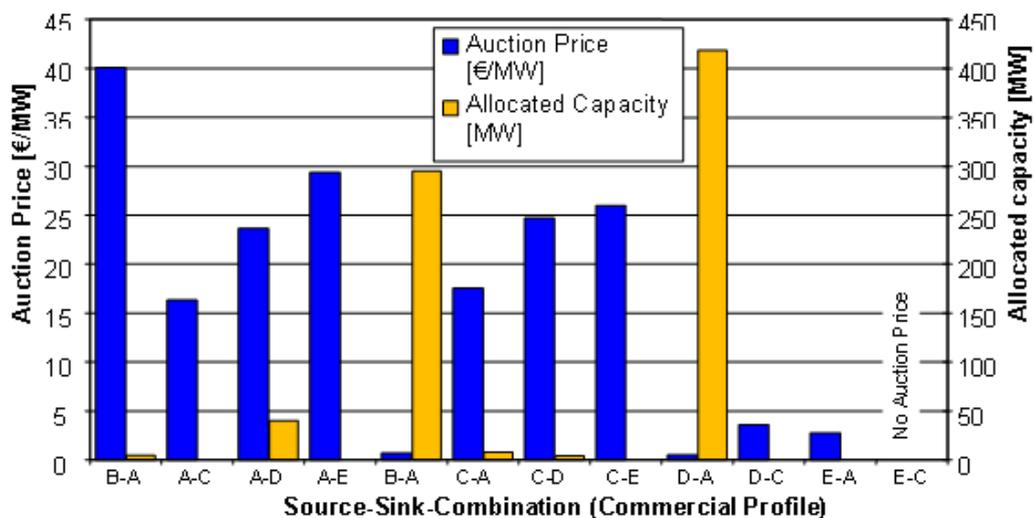


Abbildung 5 Vergebene Kapazität und Clearingpreise bei der LFBA

Nächstes Augenmerk ist die Verteilung der Auktionserlöse hinsichtlich der richtigen Marktsignale.

Bei den NTC-basierten Verfahren werden die Erlöse aus der Engpassauktion auf die benachbarten TSO nach der einfachen Regel 50:50 aufgeteilt. Bei der Anwendung der LFBA stehen die Erlöse als Ergebnis eines Clearingvorganges der gesamten Region zur Verfügung. Die Grundlage zur Erlösverteilung bilden dabei die Clearingpreise. Abbildung 6 zeigt eine einfache Gegenüberstellung beider Verfahren.

Die Verteilung der Einnahmen soll mit einem Verteilungsschlüssel erfolgen. Dieser muss gesetzeskonform sein und soll sinnvolle Marktsignale und Anreize zur Behebung von Engpässen für die TSOs hervorbringen. Er sollte akzeptabel für alle Marktteilnehmer und die Regulatoren sein.

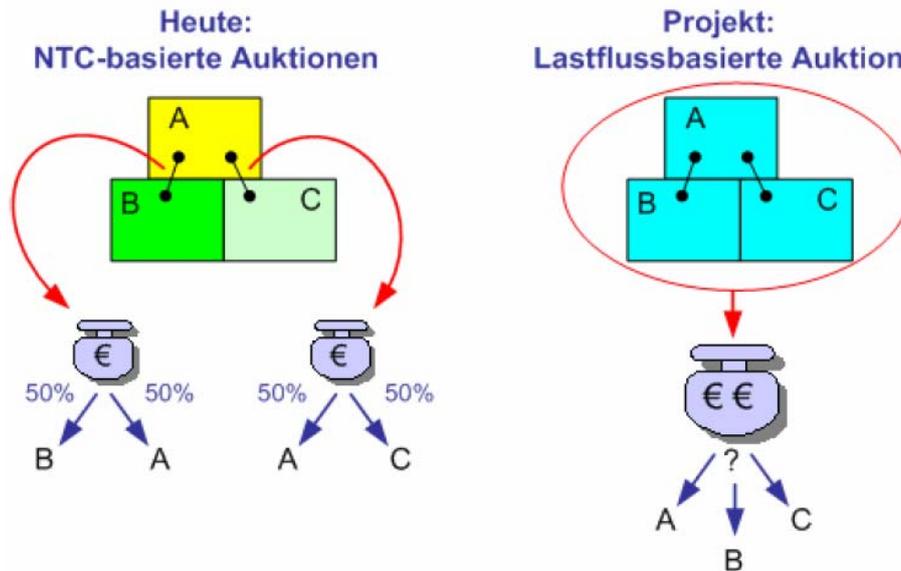


Abbildung 6 Gegenüberstellung von NTC- und lastflussbasierter Auktion [5]

Zum Abschluss ist in der Tabelle 2 ein qualitativer Vergleich beider Verfahren dargestellt.

Tabelle 2 Vergleich der Auktionsverfahren

Nr.	Kriterium	Auktionstyp	
		NTC-basiert	LFBA (Projekt)
1	Senkung des Risikos für TSOs	nein	ja
2	Aufwand bei Aufbereitung der Eingangsdaten	niedrig	sehr hoch
3	Richtige Marktsignale	nein	abhängig von Erlösverteilung
4	Vergütung der angebotenen Kapazität	nein	abhängig von Erlösverteilung
5	Bessere Nutzung der Kapazitäten	ja	nein
6	Stabilität der Ergebnisse	ja	nein

Es ist hervorzuheben, dass eine Senkung des Risikos für die TSO hinsichtlich einer Verletzung des (n-1)-Kriteriums eintritt. Je nach Verteilung der Erlöse können sich auch die marktbezogene Transparenz des Verfahrens und die Vergütung der angebotenen Kapazität erhöhen. Als Nachteil für die NTC-basierte Methode für die Aufbereitung der Daten.

Die vorliegenden Ergebnisse betrachten in erster Linie nur die Transportnetz bezogene Verauktionierung der gegebenen Übertragungskapazitäten und die Aufteilung der finanziellen Gewinne auf die einzelnen Beteiligten. Die Wechselwirkungen zum Energiemarkt und das Verhalten der Stromhändler und Engpasseigentümer bzw. –bewirtschafter ist noch nicht Gegenstand der Untersuchungen. Derartige Aufgaben werden erst bei weiteren Forschungsarbeiten mit verstärktem Einbezug von energiemarktwirtschaftlichen Fragen zunehmend an Beachtung gewinnen. Hierbei ist auf mathematische Algorithmen wie die Spieltheorie [7] oder Petrinetze [8] zurück zu greifen.

References:

- [1] CONSENTEC – Consulting GmbH: Analysis of Cross-Border Congestion Management Methods for the EU Internal Electricity Market, 2001
- [2] Hahn, R.: Wirtschaftliche Bewertung von Maßnahmen zum Engpassmanagement, Diplomarbeit an der TU Ilmenau, 2006
- [3] TransmissionCode 2003: Netz- und Systemregeln der deutschen Übertragungsnetzbetreiber, Verband der Netzbetreiber VDN e.V. beim VDEW, 2003
- [4] ETSO: Regional Flow-based allocations. State-of-play. Final paper, März, 2007
- [5] ETSO: Coordinated flow-based auctions in South-east Europe: Analysis of options for the distribution of congestion revenues among the participating TSOs, 2006
- [6] Verband der Netzbetreiber VDN e.V. beim VDEW: *Transmission Code 2003: Netz- und Systemregeln der deutschen Übertragungsnetzbetreiber*, 2003
- [7] Ventosa, M.; Baillo A.; Ramos A.; Rivier M.: Electricity market modeling trends, Energy Policy, vol. 33(7), 2005
- [8] Schneeweiss, W. G.: Petri-Netz-Bilder-Buch (Eine elementare Einführung in die beste bildliche Darstellung zeitlicher Veränderungen), LiLoLe-Verlag, Hagen, 2002, ISBN 3-934447-05-8

Authors:

Univ.-Prof. Dr.-Ing. Dirk Westermann
Dr. Dipl.-Ing. Michael Heit
Dipl.-Ök. Mariya Kryvenka
Sergey Rozhenko, M.Sc.
Technische Universität Ilmenau, Postfach 10 05 65
98684 Ilmenau
Phone: + 49 3677 69-2838
Fax: + 49 3677 69-1496
E-mail: dirk.westermann@tu-ilmenau.de

M. Lemmel / M. Schnatmeyer

RFID-Technology in Warehouse Logistics

Wireless technology becomes more important in logistics processes. Beside the usage of GPS or GSM technology for tracking and tracing of goods in the external logistics also internal logistics processes have a high demand on wireless technologies, which supports the quality, environmental and safety management. This paper resumes results from research projects, which have examined wireless technologies for positioning and identification of goods in warehouses.

State-of-the-Art in Wireless Technologies

Wireless technologies open the possibility for the seamless tracking and tracing of logistics process. For external logistics processes GPS and GSM technologies are in use for the position finding of logistics items. For internal logistics processes in a covered warehouse GPS is not available and GSM too imprecise for location goods on pallet level.

Alternative technologies for in-door processes are for example optical (Infrared), DECT, WLAN or RFID. WLAN utilizes spread-spectrum technology based on radio waves to enable communication between devices in a limited area, also known as the basic service set. This gives users the mobility to move around within a broad coverage area and still be connected to the network¹. Further the WLAN infrastructure enables determining the position of a user or item in this network.

Radio-frequency identification

Radio-frequency identification (RFID) is an automatic identification method such as barcode, relying on storing and remotely retrieving data using devices called RFID tags or transponders. An RFID tag is a device that can be attached to or incorporated into a product, animal, or person for the purpose of identification using radio waves. Chip-based RFID tags contain silicon chips and antennas. It operates at standardised frequencies between 134 kHz (LF), 13,56 MHz (HF) up to 868 / 915 MHz (UHF) [1].

¹ http://en.wikipedia.org/wiki/Wireless_LAN (3/22/2007)

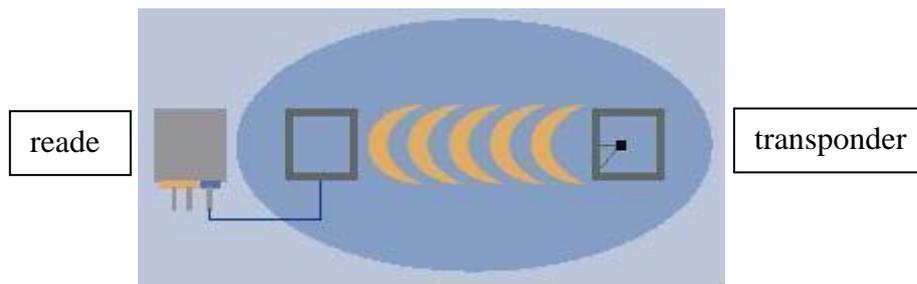


Figure 1: Principle of RFID

Passive tags require no internal power source². The antenna uses the radio field of an RFID reader for supplying the chip with energy. After the chip starts to operate, the transponder can exchange data with the reader (see Figure 1). Besides reading data, the reader can also write data on the transponder.

A typical application for passive RFID systems is the gate solution (see Figure 2).

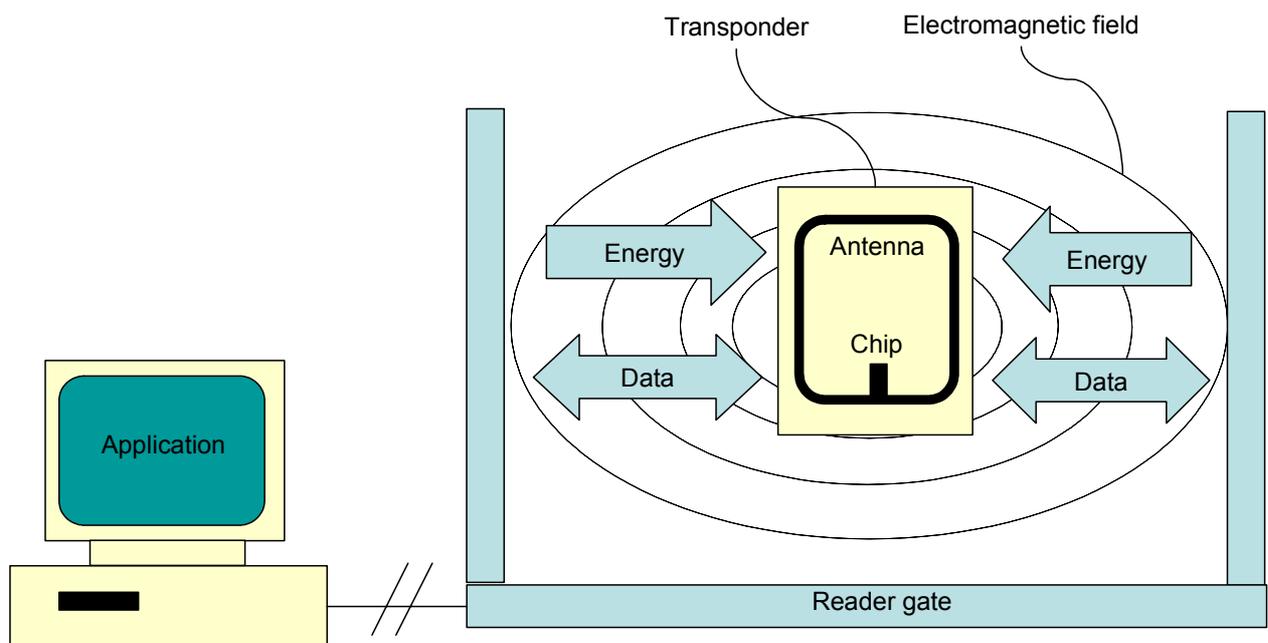


Figure 2: Principle of reader gate

In the centre of the gate it is possible to identify many tagged goods simultaneously. This type of configuration is often in use for retail and wholesale applications. Figure 3 shows a passive transponder which operates at 134 kHz.

² <http://en.wikipedia.org/wiki/Rfid> (3/22/2007)

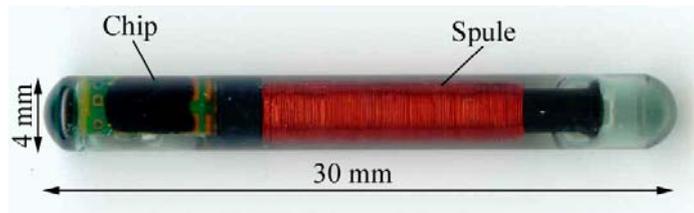


Figure 3: LF-Transponder (134 kHz)

Active transponders have onboard power supply. They have the opportunity to transmit data without being electromagnetically activated. This is reasonable in RFID-S (-sensor) for sensorial supervision of temperature or humidity etc.

For certain applications RFID has significant advantages against ordinary barcode. Some examples are:

- complicated optical detection, e. g. in refrigerated warehouse
- detection of piles, e. g. pallet full packed with different goods (see Figure 2)
- secondary data to be stored on tag, e. g. classification of hazardous goods
- sensorial supervision, e. g. temperature, humidity etc.
- high rate of circulation, e. g. tag integrated into deposit pallet

As the well known systems for positioning like GPS or GSM are just suitable for outdoor purposes, other systems have to be considered for in-house use. Nowadays many factories and warehouses facilities are already equipped with WLAN for data transmission. This existing infrastructure can also be used for position purposes. These in-house positioning systems, such as ekahau³ base on triangulation between a couple of access points and enables positioning on a cheap way. WLAN based positioning systems have also been evaluated within several research projects for in-house applications with the following disadvantages as results:

- The accuracy of ca. 3 m is insufficient for positioning goods in sizes of a euro-pellet or less.
- Due to the complex algorithms for processing the position the system is not suitable for real-time applications.
- The performance of the system is varying with the signal quality of the WLAN access, which can be alloyed by certain goods such as metal und fluids.

Therefore a new method for in-house positioning based on RFID has been implemented within a couple of actual research activities.

³ <http://www.ekahau.com>

RFID Positioning

Figure 4 describes a RFID positioning system, which was developed and used within the research project OPAK (Optimized PAcKing logistics in the life cycle economy) funded by the German ministry of education and research (BMBF). The project was focussing on the optimisation of the redistribution of plastic materials for recycling processes. Main goal was to improve logistic processes and optimise available technological infrastructures through RFID technology.⁴

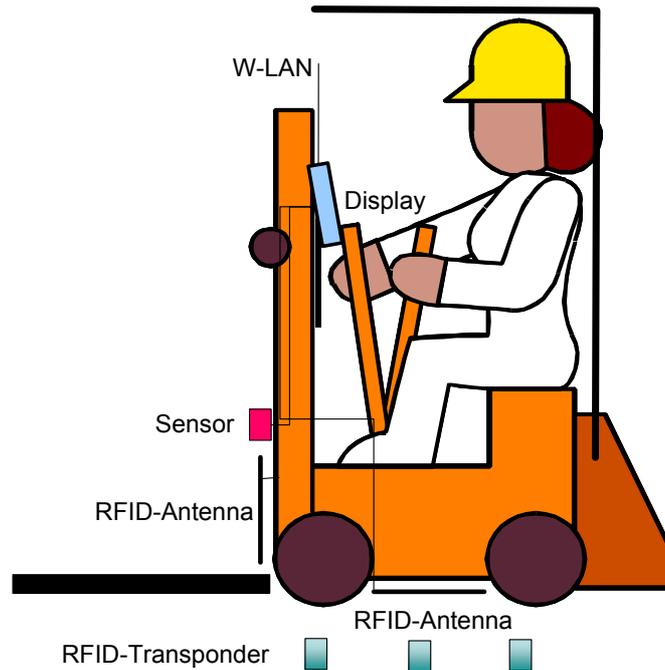


Figure 4: RFID positioning system [2]

For identification one RFID antenna is placed above the fork. This is responsible for reading information from the handled good or its pallet. A second RFID antenna is attached beneath the fork lift. This antenna reads the geographic data coming from the RFID transponder embedded into the warehouse ground.

Two additional ultrasonic sensors are installed at the front of the forklift: One sensor detects picked products on the fork. A second sensor measures the distance between fork and ground for calculating the storage level. This functionality is necessary for block and rack storage applications.

Combined with WLAN access, data collected from the cargo and the warehouse ground can be transmitted to a central WMS (Warehouse Management System) which traces the material movements and placement in the warehouse (or production plant) and provides the driver of the forklift with new transport orders after finishing the actual job.

⁴ http://www.biba.uni-bremen.de/projects/opak/Opak_homepage.htm

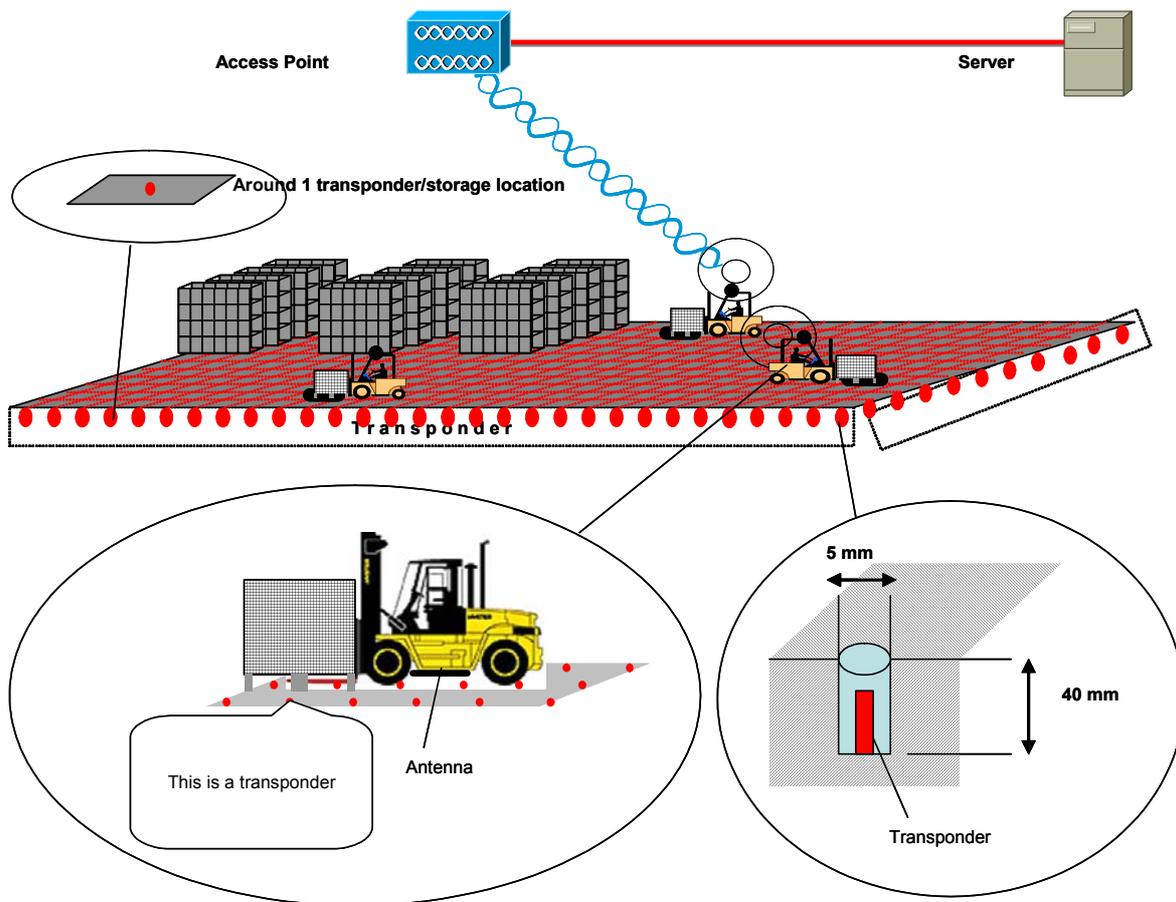


Figure 5: OPAK demonstrator overview

In addition this system can also be used to provide the fork lift driver with additional information, e.g. about dangerous materials, which have to be handled with care. As an example Figure 6 shows the user interface of the touch screen display. If the system identifies dangerous goods the traffic light will change from green to yellow (semi dangerous goods) or red (high dangerous goods).

This semi automated system is optimised for SMEs [3]. High investments in full automated systems seem to be inefficient at this type of industry because the logistic process can not be standardised.

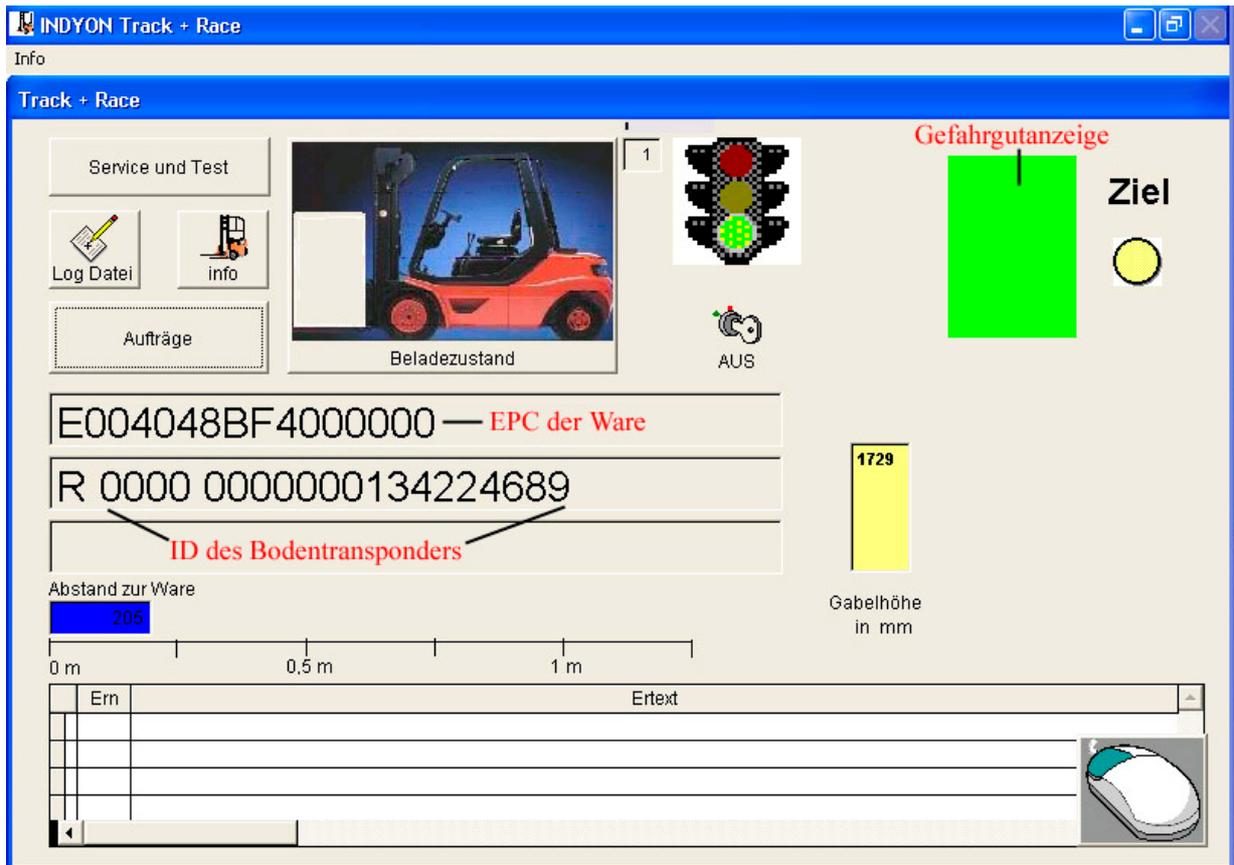


Figure 6: Display of the fork lift terminal [4]

Conclusion

The usage of RFID technology instead of barcode systems is depending on the specific application scenario. For in-house positioning the described RFID system is an attractive opportunity with sufficient accuracy at real time.

For identification purposes it is nowadays difficult to apply existing RFID systems to a whole process chain regarding to missing standards.

References:

- [1] Lemmel, M.; Plettner, A.; Schnatmeyer, M.: Optimierte Ein- und Auslagerung von Packstücken. Technische Umsetzung eines exemplarischen Realisierungskonzepts für Recyclingbetriebe. ZWF Zeitschrift für wirtschaftlichen Fabrikbetrieb, Band 101 (2006) Heft 3, Seite 109-113. München 2006.
- [2] Hans, C., Schnatmeyer, M., Schumacher, J., Thoben, K.-D.: Using Transponder Technology to Support the End-of-Life Phase in Product Life Cycle Management. Inter-national IMS Forum 2004 – Global Challenges in Manufacturing; Proceedings, Cernobbio, Italy, 17 – 19 May 2004; S. 1448 – 1455. (ISBN 88 901168 9 7)
- [3] Schnatmeyer, M.: MS Engineering & Consulting, Efficient Stock Management. <http://www.e.inos.de/3.html>.
- [4] Goch, G.; Thoben, K.-D.; Plettner, A.; Lemmel, M.; Schnatmeyer, M.; Christ, J.; Zhao, X.: Abschlussbericht zum Forschungsvorhaben Optimierte Verpackungslogistik in der Kreislaufwirtschaft (OPAK). Forschungsbericht. Hannover 2006 (TIB-Hannover, <http://edok01.tib.uni-hannover.de/edoks/e01fb06/515071951.pdf>).

Authors:

Dr. Marc C. Lemmel
 BIBA
 Hochschulring 20
 28359 Bremen
 Phone: +49 421 218 5627
 Fax: +49 421 218 5625
 E-mail: lem@biba.uni-bremen.de

Martin Schnatmeyer
 MS Engineering & Consulting
 Teerhof 34
 28199 Bremen
 Phone: + 49 421 59 751 80
 Fax: + 49 421 59 751 81
 E-mail: schnatmeyer@inos.de

V. Krugljak, M. Heit, D. Westermann

Approaches for modelling of power market. A comparison.

MOTIVATION

The electricity market is moving towards greater reliance on competition. Changing technology, new players in the generation market and a legislative mandate to provide access to the essential transmission facility have accelerated the process of competition that requires major changes in the institutions and operations of the electricity market. Nowadays, electricity markets are an evolving system of complex interactions between nature, physical structures, market rules and participants. They face risk and volatility as they pursue their goals, and make decisions based on limited information and their mental models of how they believe the system operates.

These crucial changes and challenges need to be analysed in order to make competent decisions. The necessity of modelling of power markets results from a number of economic, technical and educational reasons. Power market models enable the examination of new market design rules and restrictions. Furthermore they can serve as a platform for making reasonable investment decisions. The most challenging parts in power market modelling are the uniqueness of power markets and their products, as well as the complex interactions between market players and market institutions. Some more reasons refer to the necessity of generation pattern to be analysed, the input information for the physical network model to be provided, and model of the whole power system for power supply reliability to be established. Furthermore such a model could provide a powerful tool for training of future market operators and other market players.

For above mentioned reasons short-term models are needed to be investigated. Models for a long-term analysis are not considered here and can be found in [1], [2], [3], [5]. This paper concentrates the attention on modelling of imperfect competition and strategic bidding in short time horizons (spot-market), with special attention to investigation of market players' interactions within the framework of unified-pricing mechanism. The study is aimed at definition of main advantages and limitations of electricity market models.

From a structural viewpoint, the approaches to electricity market modelling reported in

the technical literature can be classified as showed in Figure 1:

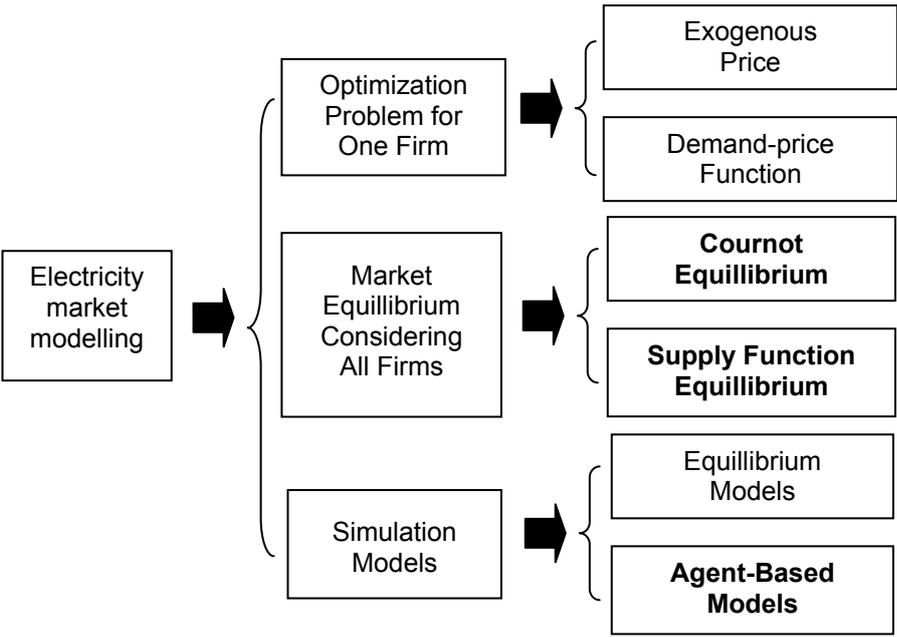


Figure 1: Main electricity market modelling trends [1]

COMPARISON OF POWER MARKET MODELS

Since the purpose of this paper is to analyse the competitive circumstances concerning an “all firms approach”, market equilibrium models are most suitable to be considered here. As an alternative approach the agent-based models are also included into analysis.

Cournot Equilibrium

In the Cournot model, each firm chooses an output quantity to maximize profit. Firms are assumed to produce homogeneous goods that are nonstorable. So all quantities produced are immediately sold. Market price in the model is determined through an auction process that equates industry supply with aggregate demand. The model also assumes that all firms in the industry can be identified at the start of the “game”, and that decision-making by firms occurs simultaneously.

According to this model each generation company uses its output decision as its decision variable, trying to define its volume targeting at maximizing its profit. However the individual profit maximization problem of each firm depends not merely on its own decision, but is strong dependent on the decisions of other market players.

Assuming the linearity of demand function and a quadratic form of the cost curve of each producer, each Generation Company (GENCO) now tries to determine its best response function, given the outputs decisions of other players. Solving the set of best response functions of all the market players present an equilibrium point, i.e. that point where no

individual can unilaterally increase its output (1) [4].

$$q_i = \frac{P_{\max} - b_i - K \sum_{\substack{j=1 \\ j \neq i}}^n \frac{b_i - b_j}{K + 2c_j}}{2(K + c_i) + K \sum_{\substack{j=1 \\ j \neq i}}^n \frac{K + 2c_i}{K + 2c_j}}, \quad \text{where} \quad (1)$$

q_i - profit maximizing output decision of i-th GENCO (MWh);

b_i, c_i bzw. b_j, c_j - coefficients of cost curves of i-th bzw j-th GENCOs ($i \neq j$) (€/MWh);

P_{\max} - maximum price (€/MWh);

$K = \frac{P_{\max}}{Q_{\max}}$ - the slope of the demand curve.

SFE (Supply Function Equilibrium)

In the SFE model, participants endowed with a cost curve find the equilibrium bid curve. I.e., a price-quantity offer that maximizes profit. An equilibrium is calculated on the basis of an interactive noncooperative game among the suppliers, where the supply function of each individual supplier is a function of production costs, capacity, the price elasticity of demand, and the extent to which the other suppliers will adjust their output to changes in market prices. This model is conceptually superior to the Cournot model in the electricity markets. This superiority consists in the “strategy space”—i.e., the range of choices over which suppliers can offer electricity products—which in the SFE modelling approach includes both price and quantity. In order to determine the optimal supply function (bid function) each GENCO, tries to find the profit maximizing point at each specified price level.

Among other papers [6] gives a good overview on a modelling approach. The persuasive convenience of its algorithm and its applicability to asymmetric environments allow to analyse the main features of this type of equilibrium simulation. Taking into consideration the lemmas for achieving the existence and uniqueness of SFE, capacity and price constraints the modelling algorithm can be sketched as follows.

Starting with capacity constraints of GENCOs, their supply functions are calculated using the alteration of the SF's slope:

$$q_i'(p) = \frac{1}{n-1} \sum_{j \in N_p} \frac{q_j(p)}{p - c_j(q_j(p))} - \frac{q_i(p)}{p - c_i(q_i(p))}, \quad \text{where} \quad (2)$$

$q_i'(p)$ - slope of supply function of the i-th producer;

p - market clearing price;

$c_j(q_j(p))$ - marginal costs;

N_p - subset of contrary bidders.

This algorithm is constructed downwards using the equilibrium system of differential equations targeted to individual profit maximization and is continued until the point $c_i(0)$ (marginal costs for zero output) is achieved.

Some numerical calculations were made to represent this approach of power market modelling. Following parameter values for two power producers were considered:

$$c_b = 0 \quad c_p = 20 \quad p^m = 100 \quad k_{1b} = 15 \quad k_{1p} = 15 \quad k_{2b} = 30 \quad k_{2p} = 30$$

where k_{it} is installed capacity of firm i in technology t ; c_t is the marginal cost of technology t ; p^m is the price cap. Additionally is assumed that quantities demanded are uniformly distributed on $[0, 100]$. Figure 2 shows the equilibrium supply functions (q_1^*, q_2^*) .

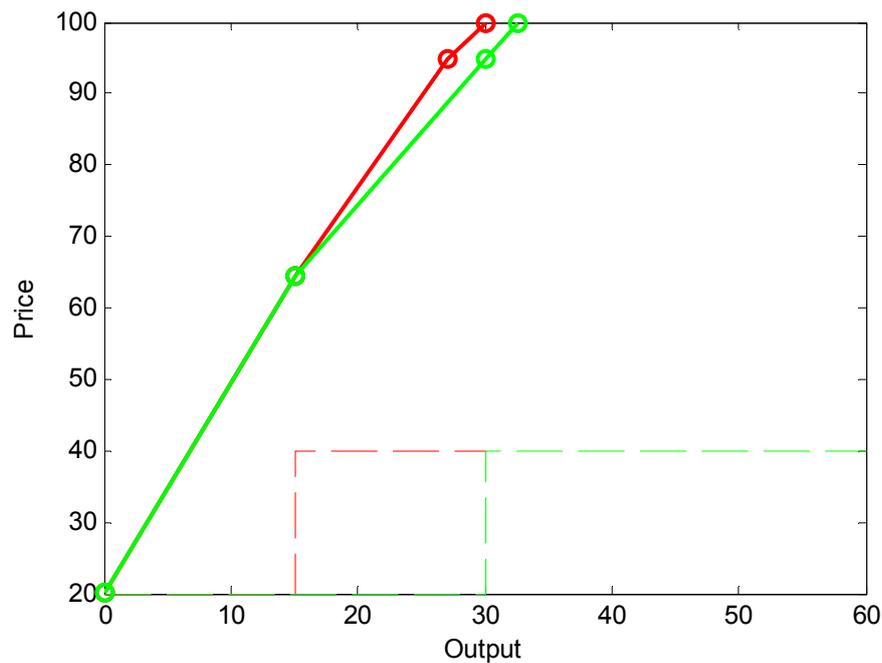


Figure 2 Equilibrium Supply Functions (q_1^* – red solid, q_2^* – green solid)

Both supply functions are identical up to $p = 64,5$. At that price the baseload capacity for firm 1 is binding; the slope of firm 1's supply function is larger than the slope of firm 2's supply function. At a price close to 94,7 firm 2 starts using peaker capacity. Starting at that price the gap in quantities offered decreases.

Agent-based models

Simulation provides a more flexible framework to explore the influence that the repetitive interaction of participants exerts on the evolution of wholesale electricity markets. Static models (e.g. equilibrium models) seem to neglect the fact that agents base their decisions on the historic information accumulated due to the daily operation of market mechanisms. In other words, agents learn from past experience, improve their decision-making and adapt to changes in the environment (e.g. competitors’ moves, demand variations or uncertain wind power infeed). This suggests that adaptive agent-based simulation techniques can shed light on properties of electricity markets that static models ignore.

One of the possible variants of power market model is presented in Figure 3 [7].

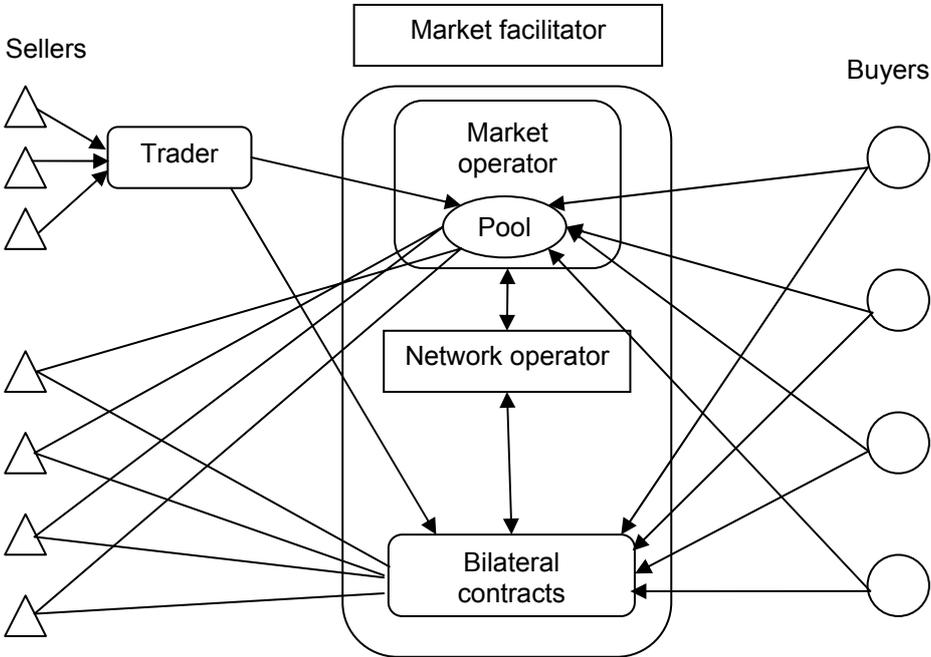


Figure 3 Multiple agents in power market modelling [7]

Superior to the equilibrium approaches, presented above, agent-based modelling avoid the restrictions to profit maximization in market players’ behaviour. Other important behavioral characteristics (e.g. risk joy/aversion, individual knowledge related to price cap, preferred prices, and available capacity, their willingness to cooperation) stand now in foreground and can be defined within this framework of modelling. One more advantage of a agent-based simulations lies in the flexibility it provides to implement the mechanisms of autonomous learning, allowing agents to alter their decisions in accordance with their individual characteristics and results of other agents’ actions. However, this freedom also requires that the assumptions embedded in the simulation

be more carefully (and empirically) justified. That is why it refers to the future research tasks for authors of this paper to identify and implement the appropriate behavioral models could be used to model the power market properly.

CONCLUSION

Within the framework of this paper the existing approaches of modelling the power markets were analysed. Approaches for simulation electricity markets have been analysed; in particularly two main streams of research in this domain: equilibrium and simulation models. Equilibrium modelling approaches have some weaknesses compared to agent-based modelling. They do not include tools to represent other firm's intentions beside profit maximization and do not allow to test different players' characteristics. Nevertheless they could be embedded into agent-based simulation in form of certain behavioral modules. In this sense supply function equilibrium should be used as more feasible and realistic approach. Table 1 contains main criteria of comparison.

Table 1 Main criteria of comparison of power market modelling approaches

Modelling possibilities		Modelling approaches		
		Cournot	SFE	Agent-based
Decisions are made on	price		+	+
	quantity	+	+	+
Possibility to represent	reactions to other players' actions	+	+	+
	learning capabilities		+	+
	attitude to risk			+
	other goals beside profit maximization			+
	different market designs			+

References:

- [1] Ventosa, M.; Baillo A.; Ramos A.; Rivier M.: *Electricity market modeling trends*, Energy Policy, vol. 33(7), 2005
- [2] Day C.J.; Hobbs B.F.; Pang J.S.: *Oligopolistic competition in power networks: a conjectured supply function approach*, IEEE Transactions on Power Systems, vol. 17, no. 3, 2002
- [3] Botterud A.; Koprass M.; Vogstad K.; Wangenstein I.: *A Dynamic Simulation Model for Long-Term Analysis of the power market*, in PROC. 14th Power Syst. Comput. Conf., Sevilla, Spain, 2002.
- [4] Kumar David A., Wen F.: *Market Power in Electricity Supply*, IEEE Transactions on Energy Conversion, vol. 16, no. 4, 2001
- [5] Krugljak V.: *Modelling of Power Markets. A comparison of existing approaches*. Technical Report - EV 0716
- [6] Aromi Daniel J.: *Asymmetric Supply Function Equilibrium with Applications to Investment Decisions in the Electricity Industry*. Job Market Paper, University of Maryland, 2006
- [7] Praca I.; Ramos C.; Vale Z.; Cordeiro M.: *MASCEM: A Multiagent System That Simulates Competitive Electricity Markets*, IEEE Intelligent Systems, vol. 18, no. 6, 2003

Authors:

M.Sc. Valerija Krugljak
 Dr. Michael Heit
 Univ.-Prof. Dr.-Ing. Dirk Westermann
 Ilmenau University of Technology, Power Systems Department, P.O.B. 10 05 65
 98684, Ilmenau, Germany
 Phone: +49(3677)69 2838
 Fax: +49(3677)69 1476
 E-mail: {valeriya.krugljak, michael.heit, dirk.westermann}@tu-ilmenau.de

S. Kümmerling / A. Friedemann / D. Westermann / N. Döring / M. Kratz

Demand-Side-Management in Privathaushalten – Der eBox-Ansatz

Alternative Energy Systems

Ein Ziel der 1998 begonnenen Liberalisierung des Strommarktes war es, elektrische Energie für jeden günstiger zu machen – auch für Privathaushalte. Während Kostensenkungen für Industrie und Handel größtenteils erreicht werden konnten, bleiben Privathaushalte zurzeit noch weitestgehend außen vor [1]. Dies wird sich mit dem innovativen eBox-Ansatz ändern.

Der Lösungsansatz umfasst die Installation eines fernparametrierbaren und fernauslesbaren Energieanschlusses (eBox) mit Schalt- und Zählerfunktion zwischen Energienetz und Endverbraucher. Über ein Kommunikationsnetz kann die eBox für verschiedene Energiehandelseinrichtungen, die als Energieagenturen bezeichnet werden, freigeschaltet werden. Damit ist die Grundlage für eine eCommerce-Anbindung der Privatkunden an den liberalisierten Energiemarkt geschaffen.

Die Energieagenturen sollen einzelne Lasten virtuell aggregieren, ein Beschaffungsportfolio erstellen und somit elektrische Energie auf deutlich niedrigerem Niveau beschaffen können. Zu den einzelnen Lasten, die mit Hilfe der eBox aggregiert werden sollen, zählen Geräte mit Speichercharakter sowie Geräte mit diskontinuierlichem Verbrauchsverhalten. Zur ersten Kategorie gehören zum Beispiel Kühlschränke oder Gefriergeräte, deren Versorgung einer Energiebedingung genügen muss, d.h. die Geräte müssen jeden Tag mit einem bestimmten Betrag an Elektroenergie versorgt werden, um die Kühlleistung aufrecht zu erhalten. Unter die zweite Kategorie fallen Geräte wie Waschmaschinen, Wäschetrockner oder Geschirrspüler, die lediglich in einem gewissen Zeitfenster, dann aber kontinuierlich betrieben werden müssen. Durch eine Freigabe all dieser Lasten an die Energieagentur kann die virtuelle Speicherkapazität zusammengefasst und am Energiemarkt vermarktet werden. Somit entstünde ein weiteres Potential, die Energiekosten für Endverbraucher

zu senken. [2]

Hinsichtlich des Betriebs großer Windkraftanlagen-Parks und des zu erwartenden, weiteren Ausbaus der Windkraft [3] wäre damit ein System geschaffen, das zur Speicherung bzw. zum Verbrauch der Windenergie die in den privaten Haushalten vorhandenen, schaltbaren Lasten nutzt. Konventionelle Kraftwerke für die Vorhaltung von Regelernergie sind dann nicht mehr bzw. in deutlich geringerem Maße notwendig. [1] [4]

Verfahren und Technologien zur aktiven Steuerung von Lasten sind seit mehr als drei Jahrzehnten bekannt und werden unter dem Fachbegriff „Demand Side Management (DSM)“ zusammengefasst [5]. Allerdings sind die bekannten technologischen Realisierungen zum DSM für den Einsatz im Haushaltsbereich weitestgehend unbrauchbar. Es sind andere, wesentliche Lösungskomponenten erforderlich [6]:

- eine bidirektionale Kommunikation zwischen Lasten und einer Betriebs-/ bzw. Abrechnungsstelle für Steuer- und Auslese Zwecke,
- eine informationstechnische Infrastruktur zur Vernetzung der Betriebsführungsinstanz mit den Lasten,
- eine Instanz zur Lastaggregation und Speichervermarktung,
- eine Betriebsführungsstrategie ohne Komfortverlust für die Endverbraucher bei gleichzeitiger persönlicher Akzeptanz des installierten Systems.

Die folgenden Beschreibungen präsentieren zwei unterschiedliche Ansätze zum DSM in Privathaushalten. Statt DSM wird die Bezeichnung Microload-Management verwendet. Diese erscheint insofern angemessener, als dass die zu schaltenden Haushaltsgeräte vergleichsweise kleine bzw. Kleinstlasten darstellen. Da die beiden Ansätze außerdem in zwei unterschiedlichen Zeitbereichen wirken, soll im Folgenden zwischen dem sogenannten High-Speed- und Low-Speed-Microload-Management unterschieden werden.

Während das High-Speed-Management im Bereich weniger Millisekunden bis einiger Sekunden arbeitet, liegt der Zeitraum des Low-Speed-Managements bei einigen Minuten bis Stunden und kann sogar über Tage und Wochen hinweg geplant werden.

Diese unterschiedlichen Zeitbereiche bedingen zusätzlich eine Unterscheidung bei den jeweils einsetzbaren Haushaltsgeräten: Für das Low-Speed-Management kommen vor allem diskontinuierliche Verbraucher in Frage, aber auch eine stundenweise Zu- bzw. Abschaltung von Geräten mit speicherndem Charakter wäre möglich. Für das High-Speed-Microload-Management hingegen eignen sich lediglich Geräte mit speicherndem Charakter. [2]

Beiden Verfahren gemein ist, dass zur Zu- und Abschaltung bzw. Steuerung der einzelnen Geräte eine bidirektionale Kommunikation notwendig ist. Anbieten würde sich eine Breitbandanbindung, z.B. via DSL oder Kabelanschluss. Vorteil dieser Anbindung ist unter anderem die Möglichkeit einer „Flatrate“, das bedeutet, es entstehen keine zusätzlichen Kosten für die Datenübertragung. Man kann relativ hohe Übertragungsgeschwindigkeiten und -bandbreiten erreichen, die Kommunikation ist über TCP/IP-Standards bereits gesichert, zusätzliche Maßnahmen zur Absicherung der Kommunikation gegen den Zugriff Dritter sind einfach zu realisieren [6]. Andere, denkbare Kommunikationswege wären Powerline oder Funkverbindungen z.B. via GSM oder Wimax, dürften aber aufgrund geringerer Bandbreiten und Geschwindigkeiten nur in wenigen Fällen Anwendung finden, wenn z.B. keine Breitbandversorgung möglich ist. Für die Kommunikation zwischen den einzelnen Geräten im Haushalt mit der Schnittstelle nach außen, der „eBox-Zentrale“, bietet sich die Powerline-Kommunikation an. Die Mess- und Steuersignale werden direkt über die im Haus befindlichen Stromleitungen übertragen. Es muss also kein zusätzliches Netzwerk installiert werden. Gegenüber funkbasierter Netzwerke (z.B. WLAN, Bluetooth) bietet Powerline zusätzlich eine Abhörsicherheit, da die eBox-Zentrale als Filter nach außen agiert [2].

Somit ergibt sich eine Kommunikationsstruktur, wie sie in Abbildung 1 dargestellt ist.

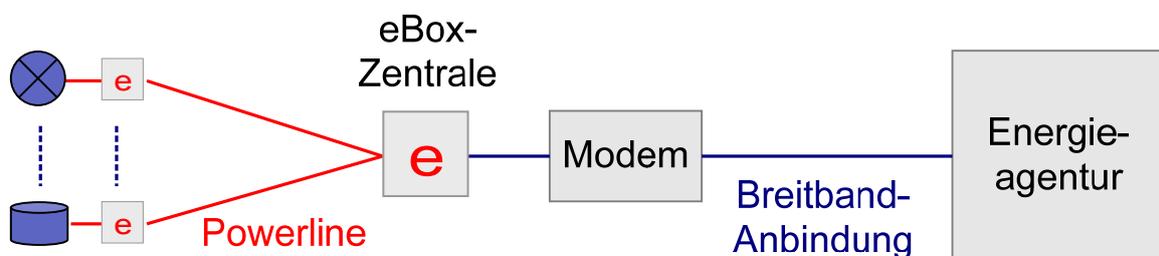


Abbildung 1: Die Kommunikationsstruktur der eBox [2]

Die Energieagenturen können einzelne Verbraucherlasten bzw. ganze Haushalte auf

diese Weise virtuell aggregieren (Lastpooling), ein Beschaffungsportfolio erstellen und an der Energiebörse oder im bilateralen Stromhandel (OTC-Markt: Over The Counter Markt) auf deutlich niedrigerem Preisniveau Elektroenergie beschaffen. Dieses Verfahren ist ebenfalls aus Industrie und Gewerbe seit einigen Jahren bekannt und kann für Privathaushalte umgesetzt werden. [6]

Eine Stärke des eBox-Projektes ist die von Beginn an stark interdisziplinäre Auslegung: Energietechniker und Sozialwissenschaftler arbeiten in jeder Phase des Projektes eng zusammen. So wird neben grundlegenden systemtechnischen Arbeiten, die die Umsetzung des eBox-Ansatzes erfordert, bereits in der Innovationsentwicklungsphase aufgrund der Involvierung privater Haushalte (User-Centred Design [7]) ein ebenso wichtiges Augenmerk auf die Akzeptanz der Endkunden gerichtet. Zudem fokussiert der zeitgemäße Universal Design-Ansatz [8] die besonderen Bedürfnisse einzelner Kundengruppen, wie zum Beispiel älterer Menschen.

Was ist die eBox für den Verbraucher? Welche Durchsetzungschancen hat die eBox-Technologie? Welche Akzeptanzhindernisse müssen beachtet werden und wie können diese abgebaut werden? Mit Rückgriff auf Theorien der Technikakzeptanz- [9] [10] und Nutzungsforschung [11] [12] sowie Ergebnisse qualitativer Leitfadenterviews (n = 19), Gruppendiskussionen und einer standardisierten Online-Befragung (n = 1500) zu zwei Nutzungsszenarien der eBox werden Antworten gegeben.

In einem Zwei-Szenarien-Modell werden zunächst die Wechselbereitschaft [13] sowie tatsächliche Wechselvorgänge auf dem Telefon- und Strommarkt verglichen, um in einem zweiten Schritt Steuerungspotentiale auf der Nachfrageseite (Demand Side Management) zu erfassen. Dabei zeigt sich sehr deutlich, dass einer hohen Wechselbewegung auf dem Telefonmarkt, die in allen Altersstufen und Einkommensklassen sowie Haushaltsgrößen gleichermaßen zu beobachten ist, eine hohe Anbieterbindung und geringe Wechselbereitschaft beim Stromnetz gegenübersteht. Die Befunde zeigen, dass sich unterschiedliche Verbrauchergruppen überraschend homogen in ihrer geringen Auseinandersetzung mit dem Problembereich Strom verhalten: mehr als 8 Jahre Liberalisierung wurden scheinbar kaum von den befragten Stromverbraucher/innen zur Kenntnis genommen [14]. Ihr Verhalten kann als konservativ und sicherheitsorientiert beschrieben werden.

Die ersten Reaktionen potentieller Nutzer auf die eBox als technische Innovation für den Energiemarkt können als skeptisch und überwiegend zurückhaltend beschrieben werden. Insbesondere die automatisierte Fernsteuerung von Geräten in Privathaushalten stieß bei den Teilnehmern der sieben Gruppendiskussionen auf Ablehnung. Wichtige Einflussfaktoren auf die Bewertung und mögliche Annahme der eBox sind dabei Gewohnheitsstrukturen, wahrgenommene Kosten, Steuerungs- und Kontrollmöglichkeiten sowie Sicherheitsbedürfnisse der Nutzer.

Um in Zukunft DSM-Lösungen im Haushaltsbereich einsetzen und Innovationen wie die eBox auf dem Strommarkt realisieren zu können, wird empfohlen, an diesen Faktoren anzusetzen. Dazu ist es notwendig in einem ersten Schritt ein Bewusstsein für Strom als Gut zu schaffen. Die Kontrolle über Preisanreizsysteme und/oder die Verfeinerung des Tarifsystems sind weitere wichtige Zwischenschritte auf dem Weg zu kompletten DSM-Lösungen.

References:

- [1] VDEW; Brinker, W.: (23.05.2007): *Innovationen unter Strom: Die Vision zur Zukunft der Energieversorgung*; Online-Dokument: http://www.strom.de/vdew.nsf/id/DE_20070523_Rede_Brinker [Zugriff am 05.06.2007]
- [2] Kümmerling, S.: *Technische Aspekte der eBox – Ausblick Modellregion E-Energy*, Vortrag im Rahmen des Workshop Modellregion E-Energy am 04.05.2007, Ilmenau
- [3] Welt der Physik: *Ausbau der Windkraft für die Energieversorgung*, 2006, <http://www.weltderphysik.de/de/4829.php>
- [4] Auer H.: *Modellierung von Kraftwerksbetrieb und Regelenergiebedarf bei verstärkter Einspeisung von Windenergie in verschiedene Energiesysteme unter Berücksichtigung des Lastmanagements*, Wien, 2005
- [5] Westermann D., John A.: *Demand matching wind power generation with Wide Area Measurement and Demand Side Management*, 2006
- [6] Kahmann, M.: *Wettbewerb im liberalisierten Strommarkt*, Springer-Verlag Berlin; Braunschweig, 2007
- [7] Vredenburg, K.; Isensee, S.; Righi, C. (2002): *User-centered design: an integrated approach*. New York: Prentice Hall PTR.
- [8] Grabowski, H. (Hrsg.) (1998): *Universal Design Theory*. Aachen: Shaker.
- [9] Dethloff, Claus (2004): *Akzeptanz und Nicht-Akzeptanz von technischen Produktinnovationen*. Lengerich [u. a.]: Pabst Science Publications.
- [10] Hüsing, B. [2002]: *Technikakzeptanz und Nachfragemuster als Standortvorteil*. Bundesministerium für Bildung und Forschung 2002. Online Dokument: www.bmbf.de/pub/Akzeptanz_Nachfrage_Standort.pdf [Zugriff: 03.03.2007]
- [11] Felser, G. (2001): *Werbe- und Konsumentenpsychologie: eine Einführung*. 2. Auflage. Stuttgart: Schäffer-Poeschel.
- [12] Ketola, A. & Matsson, P. (2001). „Help, I need somebody...“ – Consequences of a re-regulated competitive electricity market from the customer perspective. In European Council for an Energy Efficient Economy and the authors (Eds.). *Further than ever from Kyoto? Rethinking energy efficiency can get us there*. ECEEE Summer Study Proceedings, Vol. 1, p. 360-367.
- [13] Millett, S. M. (1998): *Futuring Consumer Products: An Illustrative Example of Scenario Analysis*. In: Fahey, L.; Randall, R. M. (Hrsg): *Learning From The Future*. New York, Chichester, Weinheim: John Wiley & Sons, p. 285-295.
- [14] *Ihm Wirkungen + Strategien* (Hrsg.) (1999). *Die widerständige Wechselbereitschaft – Psychologische Marktforschungsstudie zur Anfangsphase der Liberalisierung des Strommarktes*. Köln.

Authors:

Univ.-Prof. Dr.-Ing. Dirk Westermann, Fachgebiet Elektrische Energieversorgung
Prof. Dr. Nicola Döring, Fachgebiet Medienkonzeption/ Medienpsychologie
Dipl.-Ing. Stefan Kümmerling, Fachgebiet Elektrische Energieversorgung
Anne Friedemann, M.A., Fachgebiet Medienkonzeption/ Medienpsychologie
Dipl.-Ing. Michael Kratz, Fachgebiet Elektrische Energieversorgung
Technische Universität Ilmenau
Gustav-Kirchhoff-Straße 1
98684 Ilmenau

Phone: +49 (3677) 69-1488

Fax: +49 (3677) 69-1496

E-mail: dirk.westermann@tu-ilmenau.de
nicola.doering@tu-ilmenau.de
stefan.kuemmerling@tu-ilmenau.de
anne.friedemann@tu-ilmenau.de
michael.kratz@tu-ilmenau.de

A. Pedro Aguiar / R. Ghabcheloo / António Pascoal / C. Silvestre / Francesco Vanni

Coordinated Path Following of Multiple Marine Vehicles: Theoretical Issues and Practical Constraints

Abstract

We address the problem of making a group of marine vehicles follow pre-determined paths while keeping a desired spatial formation pattern (coordinated path following). We provide a brief summary of recent work in the area, leading to challenging theoretical issues that bear affinity with those that arise in Networked Control Systems. Practical constraints imposed by the underlying inter-vehicle acoustic communications network are discussed. The paper surveys some of the solutions developed so far and contains the results of simulations that illustrate the potential of the methodologies developed for coordinated path following.

Introduction

Spawned by the advent of small embedded processors and sensors, advanced communication systems, and the miniaturization of electro-mechanical devices, there is widespread interest in the design and deployment of groups of networked autonomous robotic vehicles operating in a number of challenging environments. Some of the potential applications include searching and surveying operations, as well as exploration and habitat mapping in hazardous environments.

A particular important scenario that motivates the cooperation of multiple autonomous vehicles and poses great challenges to systems engineers, both from a theoretical and practical standpoint, is *automatic ocean exploration/monitoring for scientific and commercial purposes*. In this scenario, one can immediately identify two main disadvantages of using a single, heavily equipped vehicle: lack of robustness to system failures and inefficiency due to the fact that the vehicle may need to wander significantly to collect data over a large spatial domain. A cooperative group of vehicles connected via a mobile communications network has the potential to overcome these limitations. It can also reconfigure the network in response to measurements of the environment in order to increase mission performance and optimize the strategies for detection and measurement of vector / scalar fields and features of particular interest. Furthermore, in a cooperative mission scenario each vehicle may only be required to carry a single sensor (per environmental variable of interest) making each of the vehicles in the formation less complex, thus increasing its reliability.

As an example, Figure 1 captures a conceptually simple mission scenario where an autonomous surface craft (ASC) and an autonomous underwater vehicle (AUV) maneuver in synchronism along two spatial paths, while aligning themselves along the same vertical line, so as to fully exploit the good properties of the acoustic communications channel under these conditions. This is in striking contrast to what happens when communications take place at slant range, for this reduces drastically the bandwidth of the channel, especially due to multipath effects in shallow water operations.

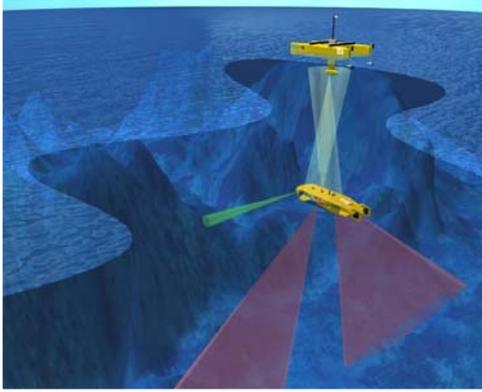


Fig. 1 Synchronization of two vehicles for data gathering at sea.

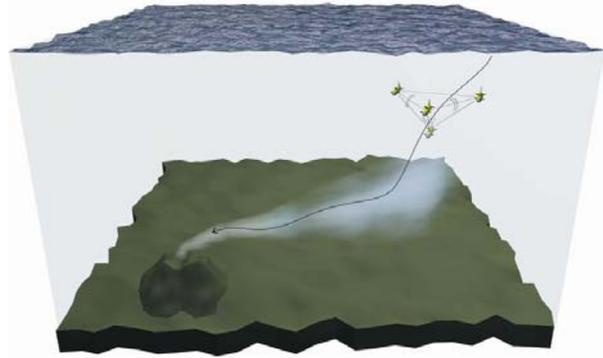


Fig. 2 AUV formation: the quest for hydrothermal vents

Figure 2 captures a different mission scenario involving a group of autonomous underwater vehicles (AUVs) in a “quest” for hydrothermal vents in the ocean floor. The mission is based on the knowledge that vents produce methane, which does not dissolve quickly in the water. This in turn allows for its detection and for the *measurement of the gradient of its concentration using methane sensors*. The vehicle baseline configuration is such that spatial estimates of the gradient of the methane concentration can be computed cooperatively. It is up to the fleet to maneuver so as to seek the region of higher concentration, and thus the localization of the vent. The scenario described requires multiple vehicle cooperation based on the type of information (methane concentration) that is acquired as the mission progresses. The mission poses formidable challenges to systems designers due to the need to develop a distributed, multi-vehicle cooperation scheme (requiring robust vehicle localization, navigation, and control) in the presence of severe underwater communication constraints. Other challenging scientific mission scenarios in the marine field can of course be envisioned (Cardigos et al., 2006).

Figure 3 shows the systems that are at the core of multiple vehicle cooperation. The scheme depicted is quite general and captures the basic trends in current research. Each vehicle is equipped with a *navigation and control system* that uses local information as well as information provided by a subset of the other vehicles over the communication network, so as to make the vehicle maneuver in cooperation with the whole formation.

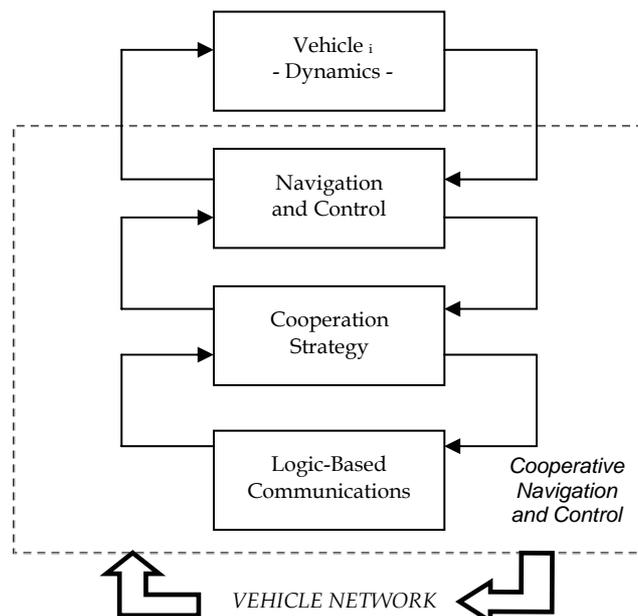


Fig. 3 General architecture for multiple vehicle cooperation

Navigation is in charge of computing the vehicle's state (e.g. position and velocity). Control accepts references for selected variables, together with the corresponding navigation data, and computes actuator commands so as to drive tracking errors to zero. The *cooperation strategy* block is responsible for implementing *cooperative navigation and control*. Its role is twofold: i) for *control* purposes, it issues high level synchronization commands to the local vehicle based on information available over the network (e.g. speed commands to achieve synchronization of a number of vehicles executing path following maneuvers). For *navigation* purposes, it merges local navigation data acquired with the vehicle itself as well as by a subset of the other vehicles. This is especially relevant in situations where only some of the vehicle can carry accurate navigation suites, whereas the others must rely on less precise sensor suites, complemented with information that is exchanged over the network. Finally, the system named *logic-based communications* is responsible for supervising the flow of information (to and from a subset of the other vehicles), which we assume is asynchronous, occurs on a discrete-time basis, has latency, and is subject to transmission failures. Central to the above scheme is the fact that each vehicle can only exchange information with a subset of the remaining group of vehicles. Furthermore, and because of the intrinsic nature of the underwater communications channel, communications should be parsimonious and take place at a very low data rate. This calls for the implementation of schemes to decide when and what minimum information should be transmitted from each of the vehicles to its neighbours. Interestingly enough, analogous constraints appear in the vibrant area of networked control systems, from which interesting and fruitful techniques can be borrowed.

Close inspection of the general architecture for multiple vehicle cooperation described above reveals the plethora of problems that must be addressed and solved:

- i) Cooperative Control (CC) (e.g. cooperative path following and cooperate trajectory tracking),
- ii) Cooperative Navigation (CN), and
- iii) CC and CN under strict communication constraints over a faulty, possibly switching network.

From a theoretical standpoint, much work remains to be done to derive analysis and synthesis tools aimed at assessing stability and performance of such a general scheme. In this respect, there are some very recent results that stand at the crossroads of control and information theory. These will probably shed some light into stability and performance limitations under constraints on the capacity of the communication channels involved. Finally, there is a need to bridge the gap between theory and practice by actually implementing selected sets of algorithms for cooperative navigation and control using prototypes of marine vehicles. Some of these issues are briefly discussed in the text below.

Cooperative control of multiple autonomous vehicles: state of the art and future challenging problems

The ever increasing sophistication of autonomous vehicles (AVs) is steadily paving the way for the execution of complex missions without direct supervision of human operators. A key enabling element for the execution of such missions is the availability of advanced systems for navigation and motion control of AVs. The past few decades have witnessed considerable interest in this area (Fossen, 1994; Leonard, 1995; Encarnação and Pascoal, 2000; Alonge et al., 2001; Jiang, 2002; Pettersen and Nijmeijer, 2003; Aguiar and Hespanha, 2004; Aguiar and Hespanha, 2007a; Aguiar and Pascoal, 2007b). The problems of motion control addressed in the literature can be roughly classified into three groups: point stabilization, trajectory tracking, and path following. For underactuated AVs, motion control is still an active

research topic (Aguiar and Hespanha, 2007a). The dynamics of these vehicles are nonlinear, which makes the control design task quite challenging. A common practice to deal with this issue is to simplify the dynamics using linearization-based techniques (Silvestre et al., 2002). The key assumption is that the range of operation is restricted to a small region for which the linear model remains valid. As a consequence, adequate control is only guaranteed in a neighborhood of the selected operating points and performance can suffer significantly when the required operating range is enlarged. Nonlinear Lyapunov-based designs can overcome some of the limitations mentioned above. See for example (Leonard, 1995; Encarnação and Pascoal, 2000; Alonge et al., 2001; Jiang, 2002; Pettersen and Nijmeijer, 2003)). Recently, in (Aguiar and Hespanha, 2007a) (see also (Aguiar et al., 2003) for experimental results conducted at Caltech) the authors have derived control algorithms for motion control of AVs (land and marine vehicles, in two and three-dimensional space). The important common feature that these designs share is the fact that they explicitly exploit the physical structure of the AVs instead of “fighting” it. In (Aguiar et al., 2007c), a robust control strategy called switched seesaw control is proposed that solves the challenging problem of point stabilization for a class of AVs in the presence of input disturbances and measurement noise.

Current research goes well beyond single vehicle control. In fact, recently there has been widespread interest in the problem of coordinated motion control of fleets of AVs. Applications include aircraft and spacecraft formation flying control (Beard et al., 2001; Giuletti et al., 2000), coordinated control of land robots (Desai et al., 2001), and control of multiple surface and underwater vehicles (Encarnação and Pascoal, 2001; Ögren et al., 2004; Skjetne et al., 2003; Skjetne et al., 2004, Pascoal et al., 2006; Ghabcheloo et al., 2006a; Ghabcheloo et al., 2006b; Aguiar and Pascoal, 2007d; Almeida et al., 2007). The concept of multiple AVs cooperatively performing a mission offers several advantages (over single vehicles working in a non-cooperative manner) such as increased efficiency, performance, reconfigurability, robustness, and the emergence of new capabilities.

The work reported in the literature is by now quite vast and addresses a large class of topics that include, among others, leader/follower formation flying, control of the “center of mass” and radius of dispersion of swarms of vehicles, and reaching a moving formation pattern. See for example (Rein et al., 2007) and the references therein. In the latter case, the goal is for the vehicles to achieve and maintain desired relative positions and orientations with respect to each other, while evolving at a desired formation speed. Central to the problems stated is the fact that each vehicle can only exchange information with a subset of the remaining group of vehicles. Similar constraints appear in the area of networked control (Hespanha et al., 2007).

The problem of coordinated motion control has several unique aspects that are at the root of new theoretical problems. As pointed out in (Fax and Murray, 2002) the following are worth stressing:

- i) except for some cases in the area of aircraft control, the motion of one vehicle does not directly affect the motion of the other vehicles, that is, the vehicles are dynamically decoupled; the only coupling arises naturally out of the specification of the tasks that they are required to accomplish as an ensemble.
- ii) there are strong practical limitations to the flow of information among vehicles, which may often be severely restricted due to the nature of the underlying communications network. In marine robotics, for example, underwater communications rely on the propagation of acoustic waves which travel at an approximate speed of $1500[\text{m s}^{-1}]$. As is well known, this fact sets tight limitations on the communication bandwidths that can be achieved and introduces unavoidable latencies that depend on the distance between the

emitter and the receiver (Pascoal et al., 2000). Thus, as a rule, no vehicle will be able to communicate with the entire formation. Furthermore, a reliable vehicle coordination scheme should exhibit some form of robustness against certain kinds of vehicle failures or temporary loss of inter-vehicle communications.

The coordination of AVs involves the design of distributed control laws with limited and disrupted communication, time-delays, model uncertainty, external disturbances, and possibly partial noisy state measurements. This is particularly significant in the case of underwater vehicles. It was only recently that these subjects have started to be formally tackled (see, e.g., (Beard et al., 2001; Giulletti et al., 2000; Ghabcheloo et al., 2006b; Borhaug et al., 2006)), and considerable research remains to be done to derive multiple vehicle control laws that can yield good performance in the presence of severe communication constraints. In (Ghabcheloo et al., 2006b), the concept of brief instabilities is exploited to model network failures and a distributed control law that ensures stability of a formation of autonomous marine vehicles is proposed. The results of a simulation with this control law are shown in Fig. 4. Further work is required to address the problems of robustness against communication delays and to develop strategies that can decide at each instant of time whether or not it is worth sending data through the network. Preliminary results in this direction can be found in (Aguiar and Pascoal, 2007d), which describes how a logic-based communication system that bears affinity with some of the ideas exposed in (Hespanha et al., 2007) can be incorporated in each of the vehicles. This system effectively decides when to transmit information to the neighbors by comparing its actual state with its estimate, “as perceived” by the neighboring system, and transmitting data when the “difference” between the two exceeds a certain level.

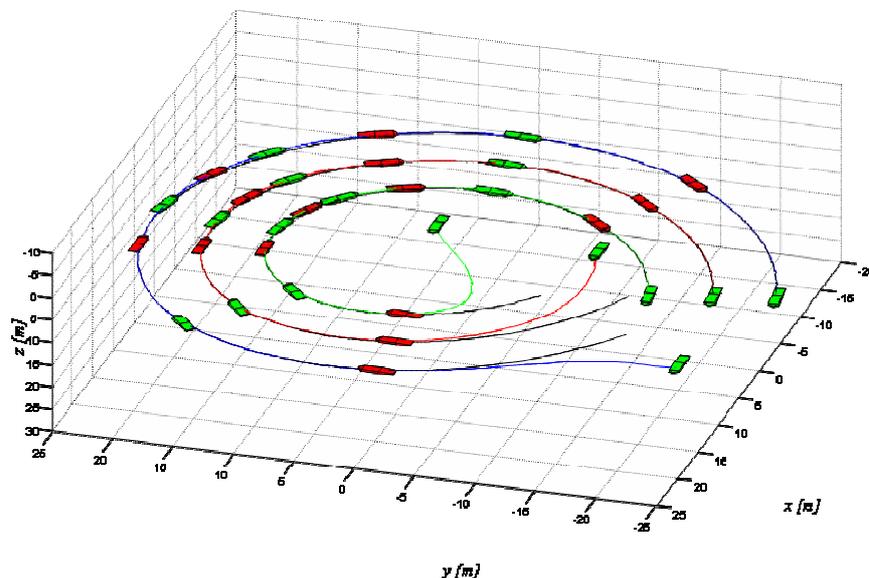


Fig. 4 Coordinated path-following of 3 AUVs under communication constraints

Acknowledgements

The authors would like to thank the European Community for the support of the research project GREX (FP6-IST-2006-035223) which is funded by the Sixth Framework Programme of the European Community.

References:

- [1] Aguiar, A. P. and J. P. Hespanha (2007a). Trajectory-tracking and path-following of underactuated autonomous vehicles with parametric modeling uncertainty. *IEEE Transactions on Automatic Control*. (In press).

- [2] Aguiar, A. P. and A. M. Pascoal (2007b). Dynamic positioning and way-point tracking of underactuated AUVs in the presence of ocean currents. *International Journal of Control*, (In press).
- [3] Aguiar, A. P., J. P. Hespanha, and A. M. Pascoal (2007c). Switched seesaw control for the stabilization of underactuated vehicles. *IFAC Journal Automatica* (In press).
- [4] Aguiar, A. Pedro and A. M. Pascoal (2007d). Coordinated path-following control for nonlinear systems with logic-based communication. Submitted for publication.
- [5] Aguiar, A. P. and J. P. Hespanha (2004). Logic-based switching control for trajectory-tracking and path-following of underactuated autonomous vehicles with parametric modeling uncertainty. In: *Proc. of the 2004 American Control Conf. Boston, MA, USA*.
- [6] Aguiar, A. P., L. Cremean, and J. P. Hespanha (2003). Position tracking for a nonlinear underactuated hovercraft: Controller design and experimental results. In: *Proc. Conf. Decision and Control (CDC), Maui, Hawaii*.
- [7] Almeida, João, Carlos Silvestre, António Pascoal (2007), Cooperative Path-Following of Multiples Surface Vessels with Parametric Model Uncertainty and in the Presence of Ocean Currents, IAV2007 6th IFAC Symposium on Intelligent Autonomous Vehicles, Toulouse France.
- [8] Alonge, F., F. D'Ippolito and F.M. Raimondi (2001). Trajectory tracking of underactuated underwater vehicles. In: *Proc. of the 40th Conf. on Decision and Control, Orlando, Florida, USA*.
- [9] Beard, R., J. Lawton, and F. Hadaegh (2001). A coordination architecture for spacecraft formation control. *IEEE Trans. on Control Systems Technology*, vol. 9, pp. 777–790.
- [10] Børhaug, E., A. Pavlov, R. Ghabcheloo, K. Pettersen, A. Pascoal, and C. Silvestre (2006). Formation Control of Underactuated Marine Vehicles with Communication Constraints. In: *Proc. 7th IFAC Conference on Manoeuvring and Control of Marine Craft (MCMC'2006), September 20-22, Lisbon, Portugal*.
- [11] Cardigos, F., A. Colaço, P.R. Dando, S. Ávila, P.-M. Sarradin, F. Tempera, P. Conceição, A. Pascoal, and R. Serrão Santos (2005). Shallow water hydrothermal vent field fluids and communities of the D. João de Castro Seamount (Azores). *Chemical Geology*, Elsevier, 224, pp. 153-168.
- [12] Desai, J., J. Otrowski, and V. Kumar (2001). Controlling formations of multiple robots. *IEEE Trans. on Control Systems Technology*, vol. 9, pp. 777–790.
- [13] Encarnação, P. and A. Pascoal (2001). Combined trajectory tracking and path following: an application to the coordinated control of marine craft. In: *Proc. IEEE Conf. Decision and Control (CDC)*.
- [14] Encarnação, P. and A. M. Pascoal (2000). 3D path following control of autonomous underwater vehicles. In: *Proc. of the 39th Conf. on Decision and Control, Sydney, Australia*.
- [15] Fax, A. and R. Murray (2002). Information flow and cooperative control of vehicle formations. In *Proc. 2002 IFAC World Congress, Barcelona, Spain*.
- [16] Fossen, T. I. (1994). *Guidance and Control of Ocean Vehicles*. John Wiley & Sons, England.
- [17] Ghabcheloo, R., A. Pascoal, C. Silvestre, and I. Kaminer (2006a). Nonlinear Coordinated Path Following Control with Bidirectional Communication Constraints. *Group Coordination and Cooperative Control*, Springer Series on Lecture Notes in Control and Information Sciences, Vol. 336, K. Pettersen, J. Gravdahl, and H. Nijmeijer (Eds.).
- [18] Ghabcheloo, R., A. Pedro Aguiar, Antonio Pascoal, Carlos Silvestre, Isaac Kaminer, and João Hespanha (2006b). Coordinated Path-Following Control of Multiple Underactuated Autonomous Vehicles in the Presence of Communication Failures. In: *Proc. 45th IEEE Conference on Decision and Control (CDC), San Diego, California, USA, Dec. 13-15, 2006*.
- [19] Giuletti, F., L. Pollini, and M. Innocenti (2000). Autonomous formation flight. *IEEE Control Systems Magazine*, vol. 20, pp. 33–34.
- [20] Hespanha, J. P., Payam Naghshtabrizi, Yonggang Xu (2007). A Survey of Recent Results in Networked Control Systems. *Proc. IEEE Special Issue on Technology of Networked Control Systems*, 95(1):138–162.
- [21] Jiang, Zhong-Ping (2002). Global tracking control of underactuated ships by Lyapunov's direct method. *IFAC Journal Automatica* 38, 301–309.
- [22] Leonard, N. E. (1995). Control synthesis and adaptation for an underactuated autonomous underwater vehicle. *IEEE Journal of Oceanic Engineering* 20(3), 211–220.
- [23] Ögren, P., E. Fiorelli, and N. Leonard (2004). Cooperative Control of Mobile Sensor Networks: Adaptive Gradient Climbing in a Distributed Environment. *IEEE Trans. on Automatic Control*, Vol. 49, No. 8, pp. 1292-1302.
- [24] Pascoal, A., C. Silvestre, P. Oliveira (2006). Vehicle and Mission Control of Single and Multiple Autonomous Marine Robots. *Advances in Unmanned Marine Vehicles*, IEE Control Engineering Series, G. Roberts and R. Sutton (Eds).
- [25] Ren, W., R. Beard, and E. Atkins (2007). Information Consensus in Multivehicle Cooperative Control. *IEEE Control Systems Magazine*, Vol. 27, No.2, pp. 71-82.
- [26] Pettersen, K. Y. and H. Nijmeijer (2003). Tracking control of an underactuated ship. *IEEE Trans. on Control. Systems Technology*. 11(1), 52–61.
- [27] Silvestre, C., A. Pascoal, and I. Kaminer (2002). On the Design of Gain-Scheduled Trajectory Tracking Controllers. *International Journal of Robust and Nonlinear Control*, 12:797-839.
- [28] Skjetne, R., T. Fossen, and P. Kokotovic (2004). Robust output maneuvering for a Class of nonlinear systems. *Automatica*, Vol. 40, pp. 373–383.
- [29] Skjetne, R., I. Flakstad and T. Fossen (2003). Formation control by synchronizing multiple maneuvering systems. *Proc. 6th IFAC Conference on Maneuvering and Control of Marine Craft, Girona, Spain*.
- [30] Skjetne, R., A. Teel and P. Kokotovic (2002). Stabilization of sets parametrized by a single variable: application to ship maneuvering. In: *Proc. 15th Int. Symposium on Mathematical Theory of Network Science, Notre Dame, IN, USA*.

Authors:

A. Pedro Aguiar / R. Ghabcheloo / António Pascoal / C. Silvestre / Francesco Vanni
 Institute for Systems and Robotics and Department of Electrical Engineering
 Instituto Superior Técnico
 Lisbon, Portugal
 Phone: + 351 21 841 8056
 Fax: + 351 21 841 8291
 E-mail: {pedro,reza,antonio,cjs,fvanni}@isr.ist.utl.pt

R. Engel/ J. Kalwa

Robust Relative Positioning of Multiple Underwater Vehicles for the European GREX Project

INTRODUCTION: THE GREX PROJECT

The main goal of the European GREX project is to achieve a first level of distributed “intelligence” through dependable underwater vehicles that are interconnected and cooperate towards the coordinated execution of tasks. Thus the project will witness the development of theoretical methods and practical tools for multiple vehicle cooperation, bridging the gap between concept and practice. The technology developed must be on one hand sufficiently generic in order to interface pre-existing heterogeneous systems. On the other hand, it must be sufficiently robust to cover problems caused by faulty communications. In this paper, we consider one interesting navigation problem arising in GREX.

NAVIGATION OF MULTIPLE AUVS

The positioning problem for multiple underwater vehicles, seen from the viewpoints of coordinated control and from underwater communication, shows some new aspects compared to the single vehicle positioning problem. In principle, one can define the following two subproblems of the positioning problem for multiple underwater vehicles.

Absolute positioning. The task of determining the position of each individual vehicle in an earth fixed frame using navigation data of the distributed AUV sensor network.

Relative positioning. The task of determining the mutual relative positions of the members of the group, using navigation data of the distributed AUV sensor network. Usually these subproblems are not subdivided in such a rigid way. The reason for this is that in most of existing projects dealing with multiple AUVs, the usage of an acoustic positioning system, such as LBL or USBL is proposed, see e.g. [2] or [3]. Since these

concepts allow the simultaneous tracking of multiple vehicles with bounded error, they can serve as a solution to both subproblems. However, the application of acoustic tracking systems impede large scale missions, call for large operational effort (in case of LBL) or make the presence of some surface vessel (in case of USBL) necessary.

We propose a method, by which relative positioning is possible by means of merely

- A dead reckoning navigation system on each vehicle of the group.
- A communication system which is able to distribute information in the AUV network (via acoustic modems).
- A device, to be implemented on each vehicle, which is capable of performing mutual range measurements (by the same modems).

Our general setting consists of a number N of underwater vehicles, each carrying a suite of sensors which enables the vehicle to operate by its own. The individual vehicles may be equipped with navigation systems of very diverse qualities. The state of the art for AUV navigation systems use an inertial navigation unit and several aiding sensors. By carefully fusing the sensor data, exhibiting partial redundancy of the navigation sensors, very accurate results can be achieved, even for the dead reckoning case without any absolute aiding of the position (see e.g. [6], [8]). The navigation accuracy achieved by AUVs range from some few meters per hour (for high performance INS/DVL combinations for site surveys) down to more than hundred meters per hour (for low cost compass/velocity sensor combinations) for site surveys.

OPERATIONAL SETUP

We assume that each individual navigation system is initialized independently, normally by means of a GPS fix after launch, while still remaining at the surface. Each member vehicle of the swarm will be equipped with an acoustic modem and a computer, on which the team functionalities are implemented. These functionalities consist in modules for team mission handling, coordinated control, communication management and team navigation, see fig. 1.

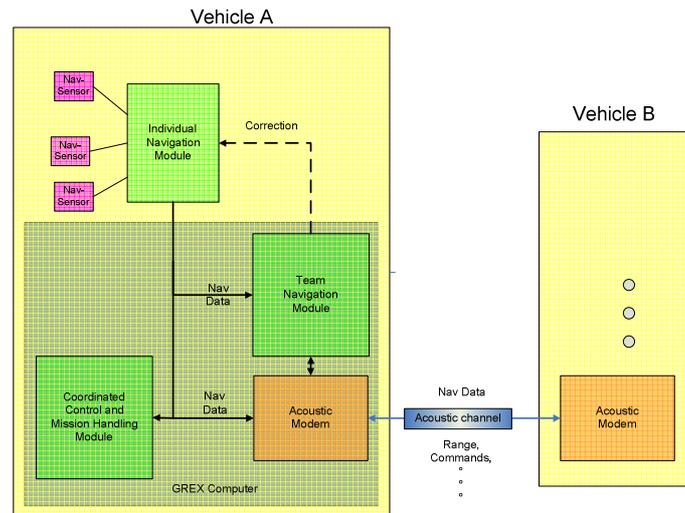


Figure 1: Vehicle specific hard- and software

The GREX navigation module receives input data from the proprietary navigation system of the respective vehicle on the one hand and from the communication module on the other hand, implementing the interface to other member vehicles. In particular, we make no special assumptions about the network topology, protocols, etc. We will merely assume that the GREX navigation module on vehicle A receives from time to time data from some other vehicle, say B. The origin of a data set will always be known to vehicle A via an ID contained in the header of the received data. The method for relative positioning, proposed here, only processes data originating from different vehicles independently. Thus, we can restrict ourselves to the special case of only two vehicles exchanging data. In the case of N vehicles, each navigation module runs N-1 instances of the algorithm described below, independently.

BASIC PRINCIPLE

Without an absolute position aiding, the individual position errors will show some drift behaviour, due to uncompensated errors in the dead reckoning navigation systems of the vehicle navigation systems. The respective position estimates can hence not be used directly for coordinated control purposes because of the growth of the relative position error over time. As a consequence, the information available to each vehicle has to be enhanced. According to the philosophy of the GREX project, any arbitrary AUV should be integrable (and, as well, removable) into (from) the group with comparably low extra cost. The least demanding type of additional measurements for relative position aiding, from a hardware point of view, consists in the determination of ranges between vehicles. Many acoustic modems on the market are capable to do

range measurements when establishing a connection between two stations. This is done by measuring round trip time of flight of a pulse signal, emitted by one and received by a second modem which in turn retransmits the pulse signal. The draw-back of this method is that one vehicle can exchange range data with not more than one other modem at the same time.

The basic concept for the solution of the relative positioning problem is illustrated in fig. 2 for the two dimensional case. It is based on the idea that a set of distances can be generated from the real measurements which geometrically determine the relative position of vehicle B to vehicle A uniquely by trilateration. In view of fig. 2, these distances are the range observations of three consecutive modem connections, r_1^{AB} , r_2^{AB} , r_3^{AB} , the position differences Δp_{12}^A , Δp_{12}^B for vehicle A and the position differences for vehicle B Δp_{12}^B , Δp_{23}^B , all of them taken between consecutive time instances of a connection between A and B. The crucial point here is that, although the individual position estimates of the the dead reckoning navigation systems of A and B are subject to drift, the difference quantities will be in general quite accurate, depending on the quality of the dead reckoning sensors and on the time span related to the position differences. The method is in some sense a generalization of the Synthetic Long Baseline concept, introduced in [9].

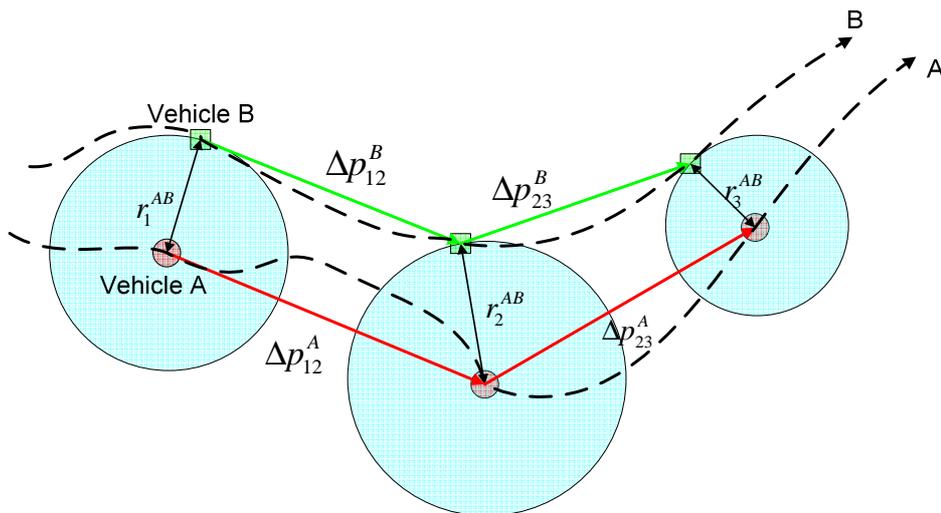


Figure 2: Consecutive range and position difference observations for two vehicles.

For a practical implementation, we will avoid solving the associated trilateration problem directly by a algebraic or numeric algorithm. The main reason is, that an “exact” solution might show large sensitivity with respect to the input data. A second reason is the need

to explicitly consider error characteristics of the measurements. Therefore, an extended Kalman filter solution is adopted here (see e.g. [5] for a standard reference). We will refer to this construction as a Relative Position Kalman Filter (RPKF) in the sequel.

SIMULATION RESULTS

For sake of clarity of presentation we will discuss a comparably simple case. Only two vehicles are considered, executing a coordinated turn manoeuvre, see fig. 3. These trajectories were occupied with a (quite large) random walk of $30\text{m}/\sqrt{\text{hour}}$, yielding the assumed trajectories as computed by the respective dead reckoning navigation system of the two vehicles. The data connection rate was chosen to be 60s, determining the rate of range observations. Position updates were assumed to be received by A from B every 15 s. The EKF time increment is chosen as $T_{tu} = 1\text{ sec}$.

Fig. 4 shows estimated positions of vehicle B. The green trajectory indicates the estimate of B's own dead reckoning navigation about four minutes after mission start. The red line represents the estimate for B's position, given from A's point of view, using the result of the RPKF running on A. For displaying the results, A's own navigation was supposed to be perfect. As can be deduced from fig. 4, the RPKF solution appears to converge to the real vehicle trajectory. Remember that this will not hold in reality for the absolute positions, displayed here.

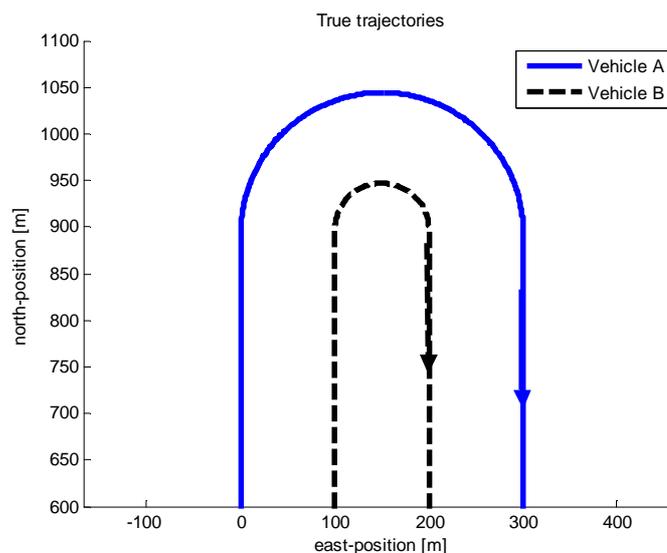


Figure 3: True trajectories of simulated vehicles

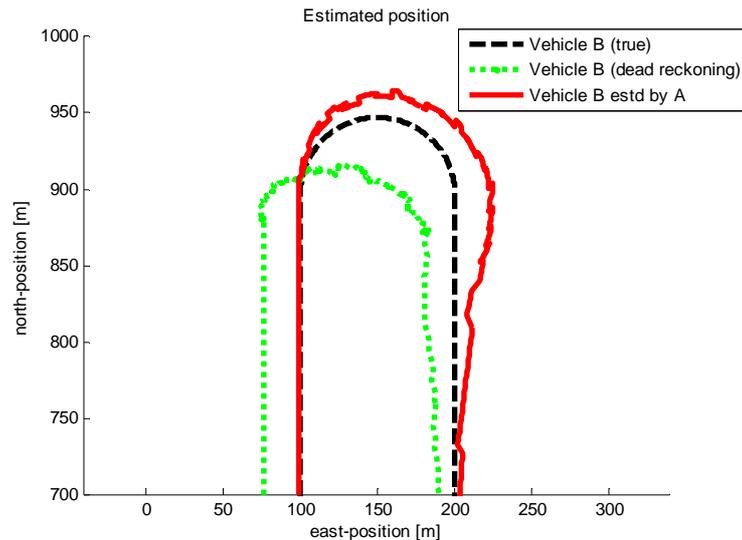


Figure 4: Relative position of B, estimated by own dead reckoning (green dotted) and by A's RPKF (red cont).

CONCLUSIONS

A method was presented, allowing the determination of relative positions of underwater vehicles in a coordinated group, using only range observations and the dead reckoning estimates of the individual vehicles. The method is based on the basic idea of a trilateration, intersecting certain spheres, which are determined by the relative motion of the vehicles.

References:

- [1] Bar-Shalom, Y, and Li, XR (1995). *Multitarget-Multisensor Tracking – Principles and Techniques*. Lecture Notes, Univ Connecticut.
- [2] Cruz, N, Matos N, Borges de Sousa, J, Pereira, FL, Silvay, J, Silvay, E, Coimbraz, J, and Brogueira Dias, J (2003). "Operations with Multiple Autonomous Underwater Vehicles: the PISCIS Project". *2nd ann. Symp. Autonomous Intelligent Networks and Systems*.
- [3] Cuff, TR, and Wall, RW, "Support Platform and Communications to manage Cooperative AUV Operations". www.ee.uidaho.edu/ee/digital/rwall/
- [4] Durrant-Whyte, H., and Bayley, T. (2006). "Simultaneous Localization and Mapping: Part I". *IEEE Robotics & Automation Magazine*, June 2006.
- [5] Gelb, A (1994). *Applied Optimal Estimation*. MIT Press.
- [6] Grenon, G, An, PE, Smith, SM, and Healey, AJ (2001). "Enhancement of the Inertial Navigation System for the Morpheus Autonomous Underwater Vehicles". *IEEE J Oceanic Eng*, Vol 26, NO 4, pp 548-560.
- [7] www.GREX-project.eu
- [8] Larsen, MB (2000a). "High Performance Doppler-Inertial Navigation – Experimental Results". *Proc IEEE OCEANS 2000*.
- [9] Larsen, MB (2000). "Synthetic Long Baseline Navigation of Underwater Vehicles", *Proc IEEE Oceans Conf*.

Authors:

Dr.-Ing. Robert Engel
 Dipl.-Ing. Jörg Kalwa
 ATLAS ELEKTRONIK GmbH, Sebaldsbrücker Heerstraße 235
 28309, Bremen
 Phone: +49 421 457 2853
 Fax: +49 421 457 2020
 E-mail: robert.engel@atlas-elektronik.com, joerg.kalwa@atlas-elektronik.com

M. Jacobi / T. Pfützenreuter / T. Glotzbach / M. Schneider

A 3D Simulation and Visualisation Environment for Unmanned Vehicles in Underwater Scenarios

ABSTRACT

In this paper we present the conceptual framework of a 3D simulation and visualisation tool that is particularly suitable for use with unmanned marine vehicles in underwater scenarios. We will describe the main ideas and the purpose of the software tool and discuss the principle constraints as well as the further development towards the finalised realisation. First results of the representation of underwater objects in realistic environments will be presented to document the effectivity and the lucidity of the approach.

INTRODUCTION

Marine influenced scenarios are a major application for the industrial use of unmanned vehicles. In the current research process, the level of autonomy of the unmanned vehicles is more and more raising. First team-based applications are on their way towards practical realisation.

Both scenarios demand possibilities for the simulation of different strategies to evaluate and visualise the mission execution. Especially, the last point must not be neglected because it is very hard to watch the vehicles performing their underwater missions in reality. For optimising the debugging process and visualization of results of unmanned vehicle missions to a wide audience, there is the need for a visualisation and simulation tool. In this work we introduce the concept and interfaces of a software tool that provides a realistic representation of unmanned marine vehicles in underwater scenarios. The three-dimensional area allows the user an independent movement. Furthermore, the software tool supports the simulation of the marine environment. For example it will be possible to simulate the sonar measurement of the seabed. The software will have clear defined interfaces to enable the operator to force the vehicle position using existing simulation tools that can be combined with the tool presented in this paper.

GENERAL CONDITIONS AND BASIC PARAMETERS

A main purpose within the development is the realisation of a visualisation tool that is as far as possible independent from the vehicle simulator. There are a lot of different possibilities in the modelling of marine vehicles (e.g. [1], [2]), differing both in proceeding and complexity. In the most simply case, a vehicle can be simulated by an integrator, translating velocity in position. It is also possible to consider all dynamic processes of a certain vehicle. Therefore many different simulators exist; each is adapted for the particular purpose. For this reason, the visualisation tool will not possess an own simulator for marine vehicles, but will have clear defined interfaces to enable a combination with different existing simulators, like the one described in [3].

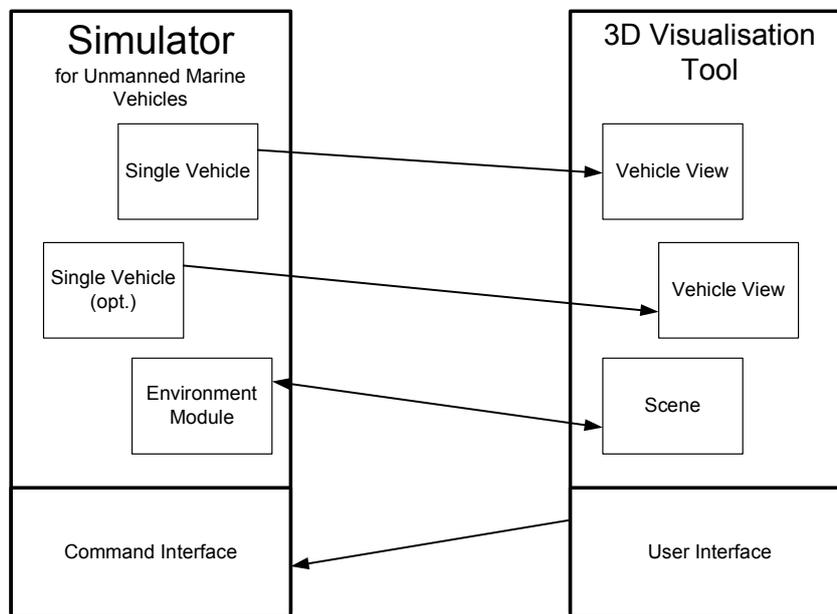


Figure 1: Connection of the visualisation tool to an existing simulator for unmanned marine vehicles

Like it is shown in Figure 1, the Vehicle View in the visualisation gets information about the position and the alignment of the vehicle(s). Information about the scene can be exchanged with the Environmental module of the simulation. The visualisation can get information about non-vehicle objects of the simulation, but it is also possible that the simulator gets different information, for example about collision of different objects. The visualisation tool will be able to supervise different objects for collision. Not only real objects can be considered; it is also possible to detect the contact of a virtual sonar beam with an object. The 3D representation of the visualisation tool is used as base for these calculations. Hence, it will also be possible to improve the calculation power of the simulation by support of the visualisation. With the user interface, it should also be

possible to send generalised commands to the simulator, like start / stop / pause simulation. The required design for the simulator is displayed in Figure 2: The vehicles are embedded in the environmental module that is responsible for collision control, communication, calculation of payload data etc. The vehicles can be controlled by KI or remote-controlled by the user.

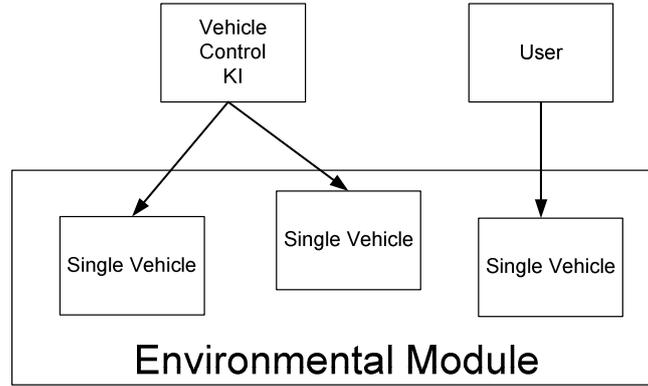


Figure 2: The required design of the simulator

SIMULATION SYSTEM

Coordinate Frames and Nomenclature

Objects moving freely in the 3D space (with 6 degrees of freedom, DOF) are often described by two coordinate frames [4]. An earth-fixed frame with an origin at the water surface is used as inertial reference system; the vehicle's position η_1 and orientation η_2 is described relative to this frame:

$$\boldsymbol{\eta} = [\boldsymbol{\eta}_1^T, \boldsymbol{\eta}_2^T]^T; \quad \boldsymbol{\eta}_1 = [x, y, z]^T; \quad \boldsymbol{\eta}_2 = [\phi, \theta, \varphi]^T. \quad (1)$$

The second frame, the body-fixed frame, has its origin in the centre of gravity of the vehicle. In this frame the velocities \mathbf{v} of the body are calculated:

$$\mathbf{v} = [\mathbf{v}_1^T, \mathbf{v}_2^T]^T; \quad \mathbf{v}_1 = [u, v, w]^T; \quad \mathbf{v}_2 = [p, q, r]^T, \quad (2)$$

where \mathbf{v}_1 denotes the linear and \mathbf{v}_2 the angular velocities. For the definition of force and momentum vectors see [4]. Using these definitions the transformation of velocity vector \mathbf{v} given in the body-fixed frame into the derivatives of the position and orientation vector $\boldsymbol{\eta}$ is described by the kinematic equation:

$$\dot{\boldsymbol{\eta}} = (\boldsymbol{\eta}) \mathbf{v}, \quad (3)$$

with $\mathbf{J}(\boldsymbol{\eta})$ as the orientation dependent transformation matrix.

The propeller rotation speed n is beside some propeller parameters an important measurement value for the calculation of the propeller thrust and torque. If a vehicle has more than one propeller, the vector \mathbf{n} denotes the rotation speeds of all propellers.

The current of the lake or sea is typically denoted in the earth-fixed frame (\mathbf{v}_C^E). It changes the vehicle's position according to all three vector components:

$$\dot{\boldsymbol{\eta}} = \mathbf{J}(\boldsymbol{\eta})\mathbf{v} + \mathbf{v}_C^E. \quad (4)$$

With (4) the position of a body with 6 DOF can be computed if the velocity vector \mathbf{v} of the vehicle is known. This calculation is, for instance, the task of the simulation model presented in the next section.

Vehicle simulation model

The vehicle simulators use the equations described above to compute the vehicle's position $\boldsymbol{\eta}$ out of propeller rotation speeds, flap angles and environmental conditions. For a flap-less underwater vehicle driven by four propellers and one vertical thruster the basic model can be described like follows.

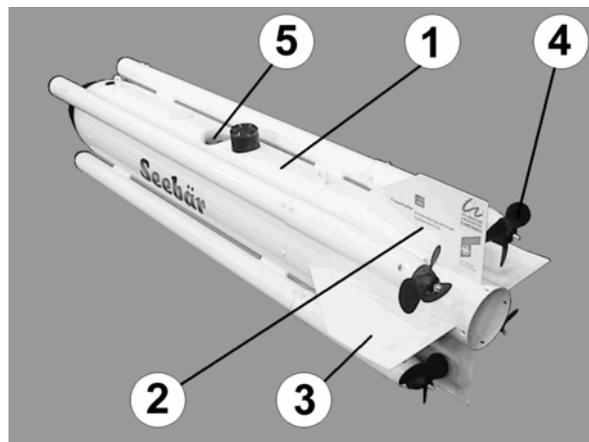


Figure 3: Components of the ROV "Seebär"

The behaviour of the underwater vehicle is described as a physical model. All forces and moments affecting the vehicle are computed separately according to origin and impact [5]. For this purpose models were created for the main components of the ROV as shown in Figure 3.

The model for the propulsion engines (4) calculates the propelling forces and moments using the rotation speed of the propellers, the vehicle's speed through water and the characteristics of both the electrical engines and the propellers. The vertical thruster (5),

which is responsible for stabilizing the position and depth changes at low velocities, is modelled the same way.

Forces and moments resulting from hydrodynamic drag and buoyancy are computed individually for the vehicle body (1), the vertical tails (2) and the horizontal tails (3). Together with other forces and moments due to the linear and angular movement of the vehicle the system model is formed by a force and a momentum equation.

This model is used to compute the velocity vector v in the body-fixed frame. For other vehicles models can be constructed in the same manner.

Environment module

The simulation of the vehicle's environment basically incorporates the sea's current (two- or three-dimensional), the elevation of the seabed (as height maps), constructions like quay walls, ships and other vehicles. If necessary, other information like temperature or salinity can be simulated if this is required by the vehicle's tasks to be solved.

Visualisation module

For direct user interaction and mission evaluation a 3D visualisation tool has been developed using open source 3D engine [6, 7]. The decision to use an open source engine depends on the ability to enhance the engine for own demands. An underwater sample scenario is shown in Figure 4.

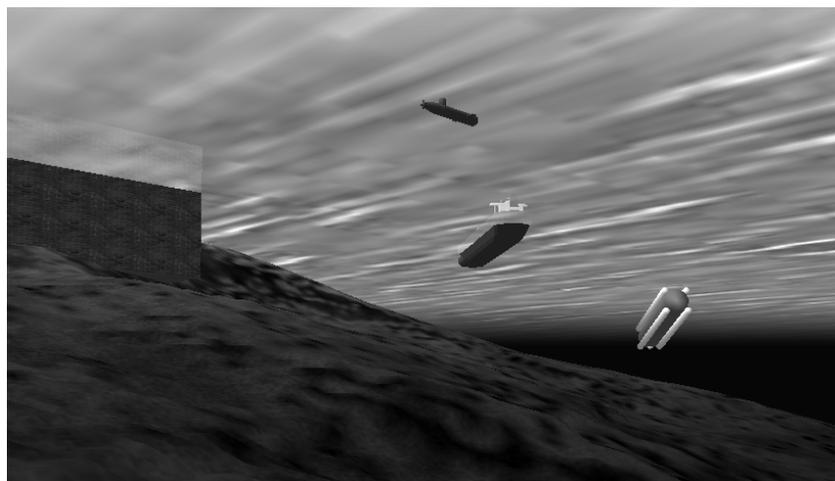


Figure 4: Underwater scenario

The visualization tool is also complete separated from the simulator as shown in Figure

1. The communication between simulator and visualisation tool is done by message exchange over a socket connection. When an object in the simulator is moving, a message is sent to the visualisation tool, and the object in the 3d environment moves also. To handle the communication gaps a simple motion model for each object is implemented.

The underwater environment and the vehicles have to be modelled for the 3D visualization using 3D modelling tools and data converter for the environment description used in the simulator. For this modelling issue, a “3D model production pipeline” had to be introduced [8]. This pipeline describes the process of using the different modelling tools in a specific way to create the 3D underwater environment.

CONCLUSION

The proposed visualisation tool for unmanned marine vehicles in underwater scenarios is an ideal extension for each existing software simulator. It allows a presentation in a reasonable design in a three-dimensional virtual reality and helps to demonstrate simulating results in an appealing manner. Furthermore, it is able to support the simulation software for all calculation concerning collision between graphical objects. Due to the clear defined interface it is guaranteed that the visualisation tool can easily be adapted to existing simulator software.

References:

- [1] T. Glotzbach, A. Picini, A. Zangrilli, M. Eichhorn, P. Otto, M. Schneider: “Evaluation of Different Control Strategies for Teams of Unmanned Marine Vehicles (UMVs) Comparing the Requirements of Underwater Communication and Navigation in MATLAB® Simulations,” 6th International Conference on Computer Applications and Information Technology in the Maritime Industries (COMPIT), Cortona, Italy, 2007, Conference Proceedings, pp. 331-344
- [2] A. Aguiar, A. Pascoal: “Dynamic Positioning of an Underactuated AUV in the Presence of a Constant Unknown Current Disturbance,” Proceedings of the 15th IFAC World Congress, Barcelona, Spain, 2002
- [3] M. Schneider, M. Eichhorn, T. Glotzbach, P. Otto: “A High-Level Simulator for heterogeneous marine vehicle teams under real constraints,” accepted for publication in the Proceedings of the 52nd IWK – Internationales Wissenschaftliches Kolloquium (International Scientific Colloquium), Ilmenau, Germany, 2007
- [4] T. I. Fossen, Guidance and Control of Ocean Vehicles, John Wiley & Sons, Chichester, West Sussex, 1994.
- [5] T. I. Fossen, “Underwater Vehicle Dynamics,” Proceedings of the Workshop on Future Research Directions in Underwater Robotics, 1994.
- [6] OGRE 3D Website, www.ogre3d.org, May 2007
- [7] MOGRE Website, <http://www.ogre3d.org/wiki/index.php/MOGRE>, May 2007
- [8] Assembling a production pipeline, http://www.ogre3d.org/wiki/index.php/Assembling_a_production_pipeline, May 2007

Authors:

Marco Jacobi
Thomas Glotzbach
Matthias Schneider
Technische Universität Ilmenau, P.O. Box 10 05 65,
98684 Ilmenau, Germany
Phone: +49 3677 69 14 21
Fax: +49 3677 69 14 34
E-mail: {marco.jacobi,thomas.glotzbach,
schneider.matthias}@tu-ilmenau.de

Torsten Pfütenreuter
Fraunhofer Center for Applied Systems
Technology (AST), Am Vogelherd 50
D-98693 Ilmenau, Germany
Phone: +49 3677 461-143
Fax: +49 3677 461-100
E-mail:
torsten.pfuetzenreuter@ast.iitb.fraunhofer.de

M. Schneider / M. Eichhorn / T. Glotzbach / P. Otto

A High-Level Simulator for heterogeneous marine vehicle teams under real constraints

ABSTRACT

In this paper, we introduce a high-level software simulator for teams of unmanned marine surface and underwater crafts. The work was performed in the framework of the European Research project GREX¹. This project has the goal to realise cooperation between heterogeneous marine vehicles by creating a conceptual framework and middleware system. The natural constraints of marine missions must be kept in mind. In underwater scenarios there is no global positioning available like GPS. The research in underwater communication is at present time at its very beginning. For the validation of different control strategies, there is the need for a high-level software simulator that is able to evaluate different approaches and to judge them according to their requirements in navigation and communication. Therefore, the above-mentioned constraints of reality must be realised within the simulator. All interfaces in the simulator need to be the same as in reality. So it is possible to use and to test the real control software modules in the simulator und make a first step towards the laboratory tests with real vehicles that need to be performed before the missions are executed in the real mission environments.

INTRODUCTION

One focus of the research project GREX [1] is on the developing of appropriate strategies for the realisation of the vehicle's coordination. It is a special challenge to force 'autonomous' vehicles to cooperate with each other. Therefore, there is the need for defined hierarchies of command and control between the vehicles. Different strategies can be imagined that will have an either more hierarchical or peripheral construction, as explained in [2].

A special challenge in marine scenarios is based on the lack of reliable global

¹ The research project GREX, FP6-IST-2006-035223 is funded by the Sixth Framework Programmed of the European Community (see [1])

positioning and communication. As navigation and communication are key technologies in the research area of autonomous mobile systems, experiences made with land and air vehicles can hardly be transferred in underwater situations. New strategies need to be developed and tested. Therefore, there is the need for a high-level software simulator that is able to evaluate different strategies according to their requirements for communication and the accuracy of the navigation. This simulator must fit to the described constraints and use the real interfaces so it can also be use to evaluate the real control software.

STRUCTURE OF THE SIMULATOR

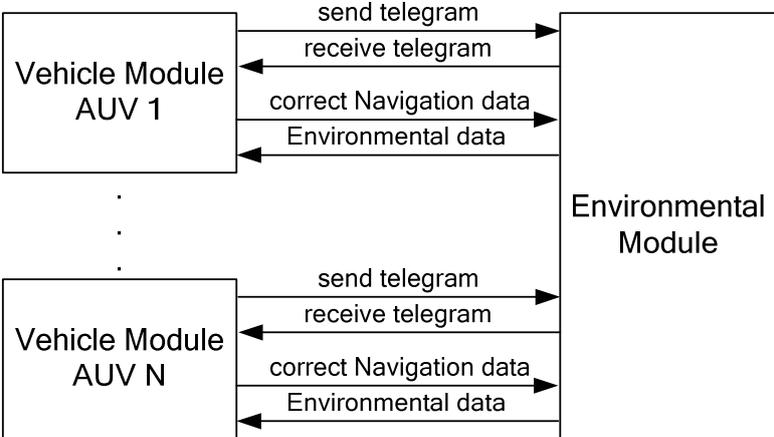


Figure 1: Principal structure of the simulator

To meet the described requirements, the simulator is made up of separated modules separating the environment as well as every single vehicle (see Figure 1). It is possible to employ the simulator under MATLAB®/Simulink® where separated blocks are used to model the mentioned modules. The blocks will be created and linked together in Simulink®. The MATLAB® based realisation enables the possibilities for an easy evaluation and presentation of the results as well as a visual evaluation with the Virtual Reality Toolbox. The use of Simulink® further guarantees an easy way to create a structure of test scenarios with the block construction method as well as the availability of all Simulink® libraries. Another possibility is the realisation as a stand-alone program whereas the vehicle modules are created and assigned to the environmental module at the program start.

To meet the already mentioned constraints, there is no data exchange directly between the vehicle modules. All data telegrams are exchanged between the particular vehicle

module and the environmental module. This module stores the incoming data telegrams in a queue together with the position and the attitude of the vehicle as well as the time of the transmission. The correct navigation data are stored in the environmental module (position, location). In each time step, it is checked whether a telegram in the queue must be delivered to a certain vehicle depending on the maximum transmitting range, the position and the attitude of the transmitting and the receiving vehicle. Moreover, the environmental data like sea current, altitude over ground and concentration will be transmitted to the separate vehicles in each time step depending on the position of the vehicle.

VEHICLE MODULE

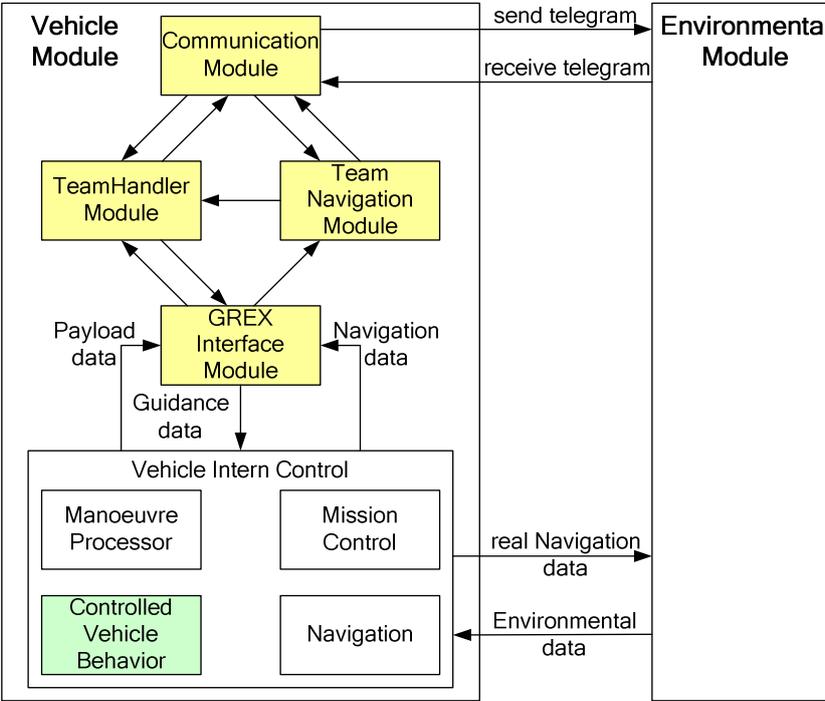


Figure 2: Components of the vehicle module

As shown in Figure 2, inside the vehicle module there are several sub-modules, which are responsible for specified, team relevant tasks. The four modules in the upper half represent the GREX-specific control software, while the Vehicle Intern Control in the lower box can be described as the single vehicle-typical control software. In the GREX-project, it can be assumed that this software already exists for every real vehicle in a different way. The GREX Interface Module describes the interface between the vehicle-own software and the new one for team coordination.

The internal control of the vehicle simulates the behaviour of the vehicle in dependence to the received guidance data like a manoeuvre list or a target position. In order to realise this actually as a sub module, the simulation of the controlled vehicle behaviour is integrated in addition to the error model for the navigation.

The modules TeamHandler, TeamNavigation, GREX-Interface Module and Communication Module will run as independent programs on the computer units of the vehicles, whereas a strict interface design of the separate modules exists and allows the consistent integration of the C++ code for the modules. The module TeamHandler, for instance, has several basic classes defining public interfaces (TeamHandler_C, InputProcessor_C, OutputProcessor_C). These classes have common references. To integrate the TeamHandler into the simulation only the classes of the data interfaces are deviated from their basic classes and their objects are implemented as s-function of the vehicle module block together with the sub-modules of the TeamHandler. During the translation, the separate libraries of the module TeamHandler are integrated into this process. This makes it possible to take over changes in the functionality locally within the TeamHandler code very easily just by a renewed translation of the Simulink®-program.

CONTROLLED VEHICLE BEHAVIOUR

To model the controlled behaviour of the different vehicle types, which are used in the project GREX, a simplistic kinematical model was designed. This model allows both a simple and a realistic simulation of the complex behaviour of the autopilot and the vehicle dynamics. In this case, only the control loop behaviour of the vehicle states *roll*, *pitch*, *heading* and *surge* will be reproduced using a time delay model. Figure 3 shows the hierarchical structure of model for the controlled vehicle behaviour.

Models for guidance tasks like the depth, track keeping and distance controllers work in combination with an algorithm to compensate for the influence of the sea current. This is possible by using known sea current and allows a simple control design for the guidance controllers (only P-controllers). The block *SeaCurrent Compensation* convert the desired set point values like course w_{course} or speed over ground w_{SoG} into the effective set points of the vehicle states w_{ψ} , w_{θ} and w_u . These set points can be determined by using the relationship of the sea current vector and the body- and earth fixed velocity vector of the vehicle as well as the intersection point between a line and a sphere as shown in Figure 4 according to the equations (1)-(3).

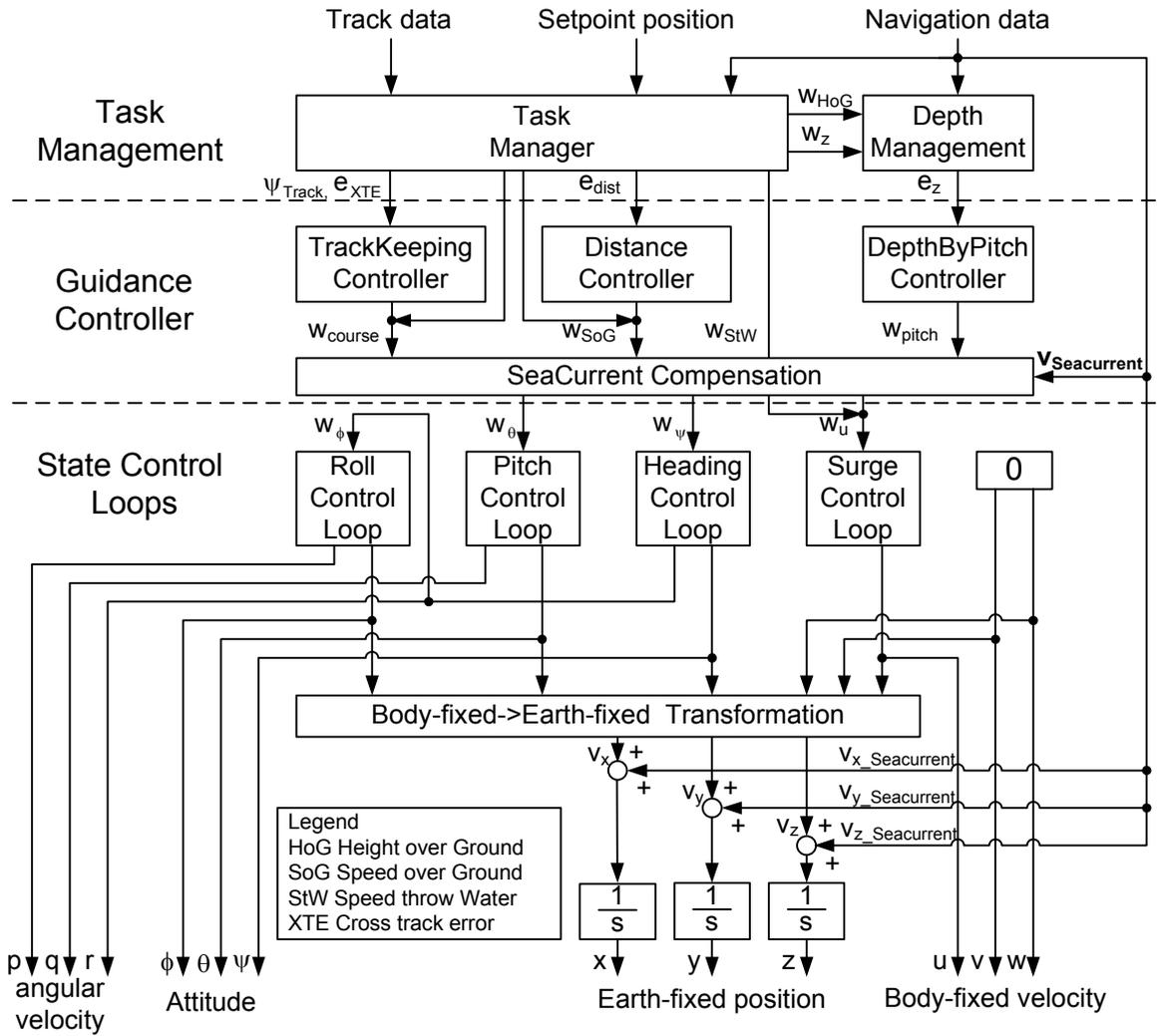


Figure 3: Structure of the controlled vehicle behaviour model

$$\text{line: } \mathbf{x}(w_{SoG}) = w_{SoG} \underbrace{\begin{bmatrix} \cos(w_{course}) \cos(w_{pitch}) & \sin(w_{course}) \cos(w_{pitch}) & -\sin(w_{pitch}) \end{bmatrix}^T}_{\mathbf{v}_{veh_ef}^0} \quad (1)$$

$$\text{sphere: } v_{veh_bf}^2 = \|\mathbf{x} - \mathbf{v}_{Seacurrent}\|^2$$

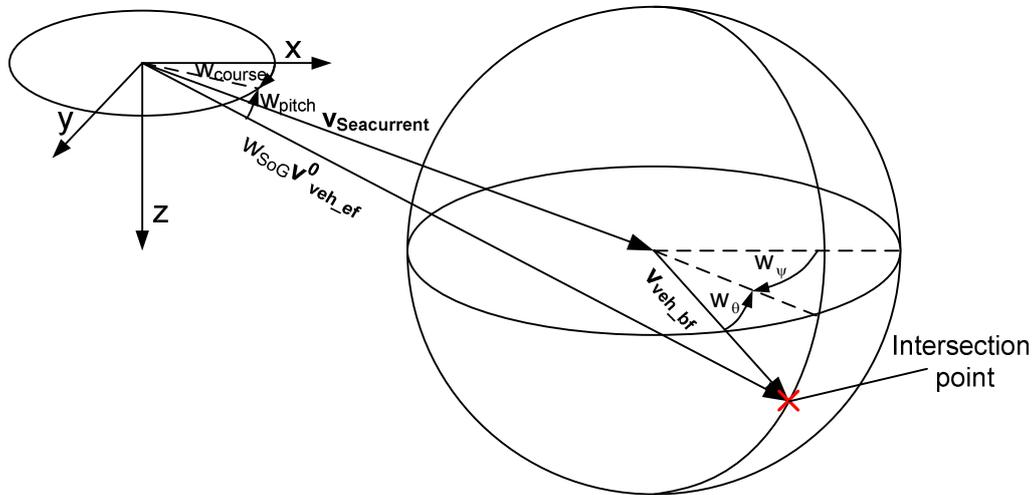


Figure 4: definitions of the velocities

$$disc = \left(\mathbf{v}_{veh_ef}^0 \mathbf{T} \cdot \mathbf{v}_{Seacurrent} \right)^2 + w_{u_max}^2 - \mathbf{v}_{Seacurrent} \mathbf{T} \cdot \mathbf{v}_{Seacurrent} \quad (2)$$

(disc < 0) impossible to hold the desired vector $\mathbf{v}_{veh_ef}^0$

$$(disc > 0): \begin{cases} w_{SoG_min} = \mathbf{v}_{veh_ef}^0 \mathbf{T} \cdot \mathbf{v}_{Seacurrent} \\ w_{SoG_max} = \mathbf{v}_{veh_ef}^0 \mathbf{T} \cdot \mathbf{v}_{Seacurrent} + \sqrt{disc} \end{cases}$$

$$\left(w_{SoG} > w_{SoG_max} \right) \rightarrow w_{SoG} = w_{SoG_max}, \quad \left(w_{SoG} < w_{SoG_min} \right) \rightarrow w_{SoG} = w_{SoG_min} \quad (3)$$

$$\mathbf{v}_{veh_bf} = w_{SoG} \mathbf{v}_{veh_ef}^0 - \mathbf{v}_{Seacurrent}$$

$$w_{\psi} = \text{atan2} \left(\frac{\mathbf{v}_{veh_bf}(y)}{\mathbf{v}_{veh_bf}(x)} \right), \quad w_{\theta} = -\text{atan2} \left(\frac{\mathbf{v}_{veh_bf}(z)}{\sqrt{\mathbf{v}_{veh_bf}(x)^2 + \mathbf{v}_{veh_bf}(y)^2}} \right), \quad w_u = \left| \mathbf{v}_{veh_bf} \right|$$

CONCLUSION

The described high-level software is able to meet the requirements that arise of the specific situation in marine environments. The natural constraints due to the problems of underwater communication and navigation are considered by the object-orientated approach. The interfaces are orientated on the real conditions, and thus the simulator cannot only be used to evaluate different strategies, but also to test the control software prior to its employment on the real vehicles. The environmental module contains considerable functionalities and allows the incorporation of many real aspects. The simulator will also be a good base for advanced research activities in the interesting field of Multiple Unmanned Marine Vehicles

ACKNOWLEDGEMENTS

The authors would like to thank the European Community for the support of the research project GREX (FP6-IST-2006-035223), which is funded by the Sixth Framework Program of the European Community.

References:

- [1] Atlas Elektronik GmbH, Ifremer, Innova S.p.A., MC Marketing Consulting, SeeByte Ltd., Technische Universität Ilmenau, Technical University of Lisbon – IST/ISR, IMAR-DOP/University of the Azores, Sciant (2005). "GREX: Coordination and control of cooperating heterogeneous unmanned systems in uncertain environments," Proposal for an EU Specific targeted research project, IST call 5. Homepage: www.grex-project.eu
- [2] T. Glotzbach and J. Wernstedt: "A Proposal for the Control of Teams of Autonomous Underwater Vehicles (AUVs) in Hierarchical and Peripheral Modes using the Concept of Adaptive Autonomy," 25th International Conference on Offshore Mechanics and Arctic Engineering (OMAЕ), Hamburg, Germany, 2006

Authors:

Matthias Schneider, Mike Eichhorn, Thomas Glotzbach, Peter Otto
 Technische Universität Ilmenau, P.O. Box 10 05 65,
 98684 Ilmenau, Germany
 Phone: +49 3677 461-145
 Fax: +49 3677 69 14 34
 E-mail: {schneider.matthias,mike.eichhorn,thomas.glotzbach,peter.otto}@tu-ilmenau.de

A. Zangrilli / A. Picini

Unmanned Marine Vehicles working in cooperation: market trends and technical requirements

ABSTRACT

This paper provides a preliminary overview of an on-going market survey on Multiple Unmanned Marine Vehicles (MUMVs), developed in the framework of the GREX project. GREX is a R&D initiative supported by the European Commission aimed at the development of middleware systems to coordinate groups of heterogeneous unmanned marine vehicles working in cooperation.

1. GREX PROJECT

Autonomous vehicles applications have been growing very fast over the last few years following the advent of new innovative technologies which provide high levels of autonomy and reliability. The past decade has witnessed the emergence of autonomous behaviours in single mobile systems, with applications to the safe operation of ground, air, and marine vehicles in the presence of changing and unknown environmental conditions. The experience which has been acquired is now steadily being brought to bear on the solutions of far more complex problems that arise when multiple systems must work together. This shift of attention was brought about by the introduction of the concept of multiple autonomous vehicles performing missions cooperatively as an attractive alternative to the traditional single vehicle paradigm.

In this context the GREX project [1] aims to create a conceptual framework and middleware system to coordinate a group of diverse, heterogeneous robotic vehicles working in cooperation to achieve a specific goal. A selected number of methodologies and systems developed under the GREX project will be validated through testing them in real situations with Unmanned Marine Vehicles including Autonomous Underwater Vehicles, Remotely Operated Vehicles and Unmanned Surface Vehicles. To summarize, the project will develop technological solutions that will allow for the execution of complex missions using Multiple Unmanned Marine Vehicles (MUMVs).

In order to develop a “market-driven” product, the project includes a survey to monitor and to investigate current and future market trends as well as the technical requirements

in the field of MUMVs. In the following sections the methodology used for carrying out the survey and the main results achieved until now are shown.

2. METHODOLOGY

The methodology followed to perform the survey consisted of four steps:

- i. Data collection to gather information on MUMVs state of the art technology and market trends through literature research;
- ii. Creating a questionnaire in order to collect information from experts worldwide belonging to companies and R&D organizations involved in the use and the development of Unmanned Marine Vehicles;
- iii. Direct research through expert interviews; 95 organizations were contacted from which 16 interviews were collected;
- iv. Final report drafting processing the data collected from the literature research and the results of the direct research. Although the survey is on-going, the information collected thus far allows us to provide a preliminary overview of the results.

3. MAIN RESULTS

The results shown below have been mainly extracted from interviews and are grouped in four parts:

- i. An overview of existing initiatives focusing on development of solutions for MUMVs management and coordination;
- ii. Benefits expected from the use of MUMVs;
- iii. MUMVs – User needs;
- iv. MUMVs - Market trends.

▪ Overview of existing MUMVs initiatives

The unmanned robots technology has gained a significant interest among a wide set of organisations operating in the marine domain due to its increasing reliability and cost effectiveness. Latest technology developments focuses on multiple vehicle approach potentially offering advance capabilities for applications including ocean sampling, mapping, surveillance and communication. So the use of multiple vehicles for defined tasks is currently an important topic within civil and military robotic research even if all attempts are still in their infancy [2]. Here a short description of some active initiatives follows:

- i) *Seaswarm Pty Ltd* commercializes autonomous underwater swarming vehicles for

highly efficient and rapid data collection and profiling of water and ocean bodies. The swarming concept is implemented through the use of small autonomous submarines designed to be deployed in a fleet [3].

ii) *The project “Decentralized Control of Multiple Autonomous Underwater Vehicles”* develops a comprehensive design procedure for communication among, and decentralized control of, a fleet of AUVs. Communication and distributed control concepts for a fleet of cooperating AUVs is investigated [4].

iii) *The project CADRE - Cooperative Autonomy for Distributed Reconnaissance and Exploration* develops a framework for the coordination of heterogeneous collections of unmanned vehicles for autonomous execution of goal-oriented missions [5].

iv) *The Multiple Cooperating AUVs (MCAUV) program* whose goal is to evaluate a Long Endurance Mobile Underwater Coastal Surveillance System through the development of innovative technologies focused on communication and cooperation multiple heterogeneous AUVs [6].

v) *The project ASAP - Adaptive sampling and prediction* whose goals are: learn how to deploy, direct and utilize autonomous vehicles (and other mobile sensing platforms) most efficiently to sample the ocean, assimilate the data into numerical models in real or near-real time, and predict future conditions with minimal error [7]. During the experiments led in the Monterey Bay 2006 field program the scientists have also studied how fleets of gliders can travel in different "formations" to cover hundreds of cubic kilometres of constantly evolving ocean [8].

▪ **Benefits expected from the use of MUMVs**

The MUMVs could have different advantages in a mission execution; interviewees were provided with a list of benefits in using MUMVs and they were asked to mark each one by level of importance. The following table shows the results.

BENEFITS EXPECTED	Score*
Mission time reduction	63%
New capabilities development	51%
Improvement of mission execution reliability	38%
Cost reduction for mission execution	25%
Optimisation of your existing vehicle “fleet”	14%

**Opinions were graded by a number from 1 to 6: 1= not relevant; 6= very important
Score is the ratio between: the number of answers with “5” or “6” divided by the number of total answers*

More than half of the interviewees cited the *mission time reduction* and the *development of new capabilities* (meant as new underwater tasks) as being the most **important** benefits in the use of MUMVs.

▪ **MUMVs – User needs**

To support the project technical partners in the development of a market-driven product, interviewees were asked to identify the **functionalities** and the **technical requirements** that a GREX-like system would be able to offer in a multiple vehicle scenario. Interviewees were provided with a list of features and they were asked to indicate the level of importance for each one.

The majority of functionalities listed in the table below were rated high priority by the respondents; in particular, the experts interviewed clearly indicated: (i) *mission re-planning*, (ii) *execution of downgraded mission* and (iii) *the monitoring of the state of the health of distributed systems* as the key functionalities to be implemented in a system for multi-vehicle applications.

SYSTEM FUNCTIONALITIES	Score*
Re-planning mission of the complete fleet in case of a new/modified objective	75%
Partially execution of the mission even in a downgraded situation	75%
Monitoring the “state of health” of the distributed communications and navigation system	75%
(Semi-) autonomous management capability for multiple vehicles mission execution	69%
Programming multiple-vehicle-missions using a single, central planning station	63%
Sharing sensor information inter-vehicles	44%
Common measure of mission execution status as the mission unfolds	31%

**Opinions were graded by a number from 1 to 6: 1= not relevant; 6= very important
Score is the ratio between: the number of answers with “5” or “6” divided by the number of total answers*

When considering the technical requirements three of them were clearly indicated as the most important (see table below): a system managing MUMVs is required to (i) *ensure the interoperability among the different acoustic modem systems*, (ii) *incorporate adaptive fleet formation capability in response to unforeseen events* and (iii) *ensure a full compliance with existing systems regarding user-interface*.

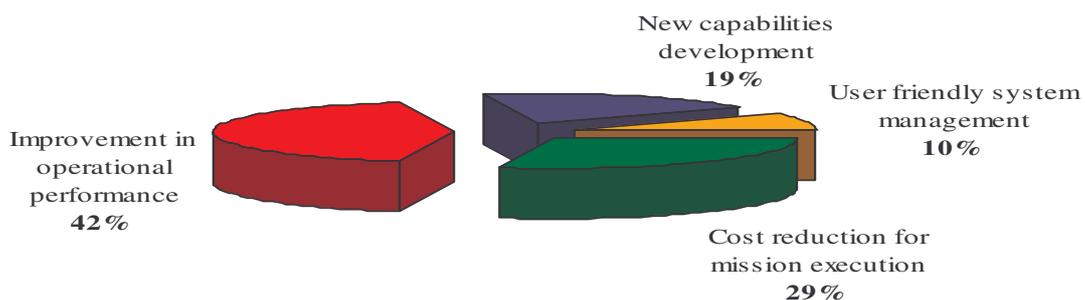
SYSTEM TECHNICAL REQUIREMENTS	Score*
<i>User-Interface</i>	
Compliance to link to existing programming systems for individual vehicles	50%
Using a single programming system for multiple heterogeneous vehicles	47%
Using graphical languages for programming	40%
<i>Middleware/hardware for inter-vehicle communication</i>	
Interoperability among different acoustic modem systems	71%
Seamless re-configuration of inter-vehicle communications network	43%
Constant monitoring of the status of the inter-vehicle communication network	43%
<i>System for coordinated control of multiple objects</i>	
Adaptive fleet formation capability in response to unforeseen events	67%
Tight fleet formation control as determined by a pre-defined strategy	29%

*Opinions were graded by a number from 1 to 6: 1= not relevant; 6= very important
Score is the ratio between: the number of answers with "5" or "6" divided by the number of total answers

▪ MUMVs - Market trends

Considering the market expectations, 37,5% of the interviewees predict that MUMVs applications will rise **significantly** over the **next three** years. The reasons for positive answers were: the ability to conduct increasingly complex missions, limited only by software and not by vehicle hardware, and the improved cost efficiency in offshore surveys. It is envisaged that such features will be required mainly by oil & gas operations, port protection and deepwater and military applications. However, the majority of respondents (62,5%) consider the market to still be in its developing stages. The main concerns were: MUMVs applications could be too expensive and there is the need for a lot of R&D (5-10 years) before MUMVs applications become really operational. These concerns are supported by the fact that, as shown in Figure 1, the majority of interviewees (42%) indicated "*improvement in operational performance*" the most important factor that will affect the MUMVs market development, followed by "*cost reduction through the use of cheaper MUMVs*" (29%).

Figure 1 The most important factors allowing for MUMVs market development.



Although a significant market growth in the short term is not foreseen, the majority of respondents (64%) thought that generic and modular (GREX-like) systems for MUMVs mission planning and execution could be adopted by existing Unmanned Marine Vehicles by both manufacturers and users. However, some interviewees pointed out that the market is not yet mature enough to adopt open solutions since *"manufactures are unlikely to welcome open solutions and customers currently place a premium on reliability best demonstrated by closed systems"*. In that respect, customers should *"create the incentive for manufacturers to adopt a generic system"*.

4. CONCLUSION

There is a growing interest in developing systems for multi-vehicle management and coordination in order to carry out missions with increasing complexity; that is witnessed by many R&D projects existing worldwide. GREX contributes to this trend by developing software solutions for multi-vehicle coordination starting from the real needs of the market in terms of technical and functional requirements. These are being collected from a direct survey involving a majority of robotic underwater experts worldwide.

From the initial results of the survey, it has emerged that the market for MUMVs applications is still immature and further R&D efforts are needed to prove that they will be really operational. However, many respondents provided positive feedback on market potential of the generic and modular (GREX-like) systems, confirming the value of the GREX project.

Acknowledgements

The authors would like to thank the European Community for the support of the research project GREX, in which this work was performed. The project GREX (FP6-IST-2006-035223) is funded by the Sixth Framework Programme of the European Community.

References:

- [1] Web site www.grex-project.eu
- [2] D. Spenneberg, C. Waldmann and R. Babb, *Exploration of underwater structures with cooperative heterogeneous robots*, Oceans 2005 Europe, Volume 2, pp.782-786
- [3] Web site <http://www.seaswarm.com.au/>
- [4] Web site <http://www.mrc.uidaho.edu/cisr/subs/index.htm>
- [5] S. Willcox, D. Goldberg, J. Vaganay and J. A. Curcio, *Multi-vehicle cooperative navigation and autonomy with the bluefin cadre system*. - 2006 IFAC
- [6] Web site <http://www.ausi.org/research/research.htm>
- [7] Web site <http://www.princeton.edu/~dcs/ asap/>
- [8] Web site <http://www.mbari.org/MB2006/ASAP/mb2006-asap-links.htm>

Authors:

Mr. Antonio Zangrilli
Mr. Andrea Picini
INNOVA, Via Giacomo Peroni, 386
00131 Roma
Phone: +39 06 40040358
Fax: +39 06 40040364
E-mail: a.zangrilli@innova-eu.net , a.picini@innova-eu.net

T. Glotzbach / M. Marinov / P. Otto / M. Schneider

A Concept for Team-Orientated Mission Planning and Formal Language Verification for Heterogenous Unmanned Vehicles

ABSTRACT

This paper proposes a concept for the team-orientated mission planning (TOMP) and Formal Language Verification (FLV) of several unmanned vehicles with a focus on marine surface and underwater crafts. The use of several heterogeneous unmanned marine vehicles is a goal of the European Research project GREX¹ in which framework this work is done. It is important to use an approach that prevents the operator from the necessity to perform the planning for all participating vehicles. We will describe the process of upgrading an existing software tool for marine vehicle mission planning to the ability of performing Team-Orientated Mission Planning, translating team mission plans to single vehicle mission plans and executing a Formal Language Verification to guarantee the feasibility of the team- and the single vehicle mission plans.

INTRODUCTION

It is important for the success of the GREX-project that the users have the possibility to perform the mission planning on a simple way, without the necessity to know all the technical details. Therefore there is the need for a meta-language to formulate team mission plans in an understandable way and also for a software editor that allows a simple editing. The realisation of a manageable mission editor is a special challenge. The editor has to accept two main tasks: The mission planning is performed in an offline mode before the mission starts. The same interface shall be used to analyse the past mission summary of the data collected from the swarm members. If an intervention of a human operator ought to be allowed during the mission, it will also be performed with this user interface.

¹ The research project GREX, FP6-IST-2006-035223 is funded by the Sixth Framework Programme of the European Community

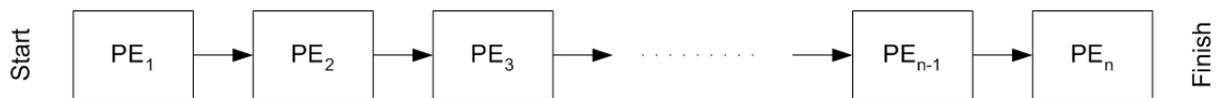


Figure 1: A sequential Mission Plan

According to the inputs of the operator, the editor software builds up a mission plan, which is composed of single plan elements. Every plan element describes a complex plot for the mobile systems. According to Figure 1, the mission plan can be stated as a linear sequence of plan elements (mission elements) that will further be referred as Vehicle Primitives. The advantage of sequential plans in opposite to parallel ones is that it is simpler for a human operator to create. For that reasons, many applications with autonomous systems (especially with Unmanned Air Vehicles, UAVs, and Autonomous Underwater Vehicles, AUVs) use sequential mission plans. [1] – [3]

The GUI for GREX Mission Planning will base on SeeByte’s SEETRACK [4], the de facto standard common operator interface for mission planning, monitoring and post mission analysis with sensorised remote assets and their data products. SeeTrack has at this stage the ability to plan, execute and analyse the missions of single autonomous marine vehicles and will be empowered during the project for the planning, executing and analysing of team missions. The concept of the software for cooperation between the vehicles is described in [5] and [6]

TEAM-ORIENTATED MISSION PLANNING (TOMP)

The operator needs to be released from the task to create the mission plans for every single vehicle. Every vehicle that will be used in GREX has already autonomous abilities, like reaching of a given position or the following of a certain track. These simple manoeuvres are called ‘Single Vehicle Primitives’ (SVPs) and are the ingredients of the vehicle’s mission plan. It is not planned to directly give commands to the autopilots of the involved vehicles from the Team Level. The SVPs shall be used as they are tested and work properly for each vehicle. Another advantage is that these SVPs – as they are used for Mission Planning – are usually available for all marine vehicles. So no further information about the vehicles and their technical realisations is needed. This eases the implementation of the vehicles into the team and may also ensure confidentiality about the functionality of the vehicles what may be important for vehicle providers.

This results in the needs of two Meta-languages for team mission planning. One is the

team-orientated Meta-language for mission planning which contains all actions that can be performed by a vehicle team, like moving on parallel tracks or following a plume. The elements of the team mission plan are called 'Multi Vehicle Primitives' (MVPs). They contain all necessary information and abilities for the predefined scenarios in the GREX project.

The other essential meta-language is a GREX-standardised language on vehicle level. It contains a couple of unitary SVPs to realise the missions defined by the MVPs. After the operator has planned the team mission plan, it will be translated to many single mission plans, one for each vehicle. This plans will be translated into the corresponding vehicle language by the 'GREX Interface Module' that in general is responsible for all interfaces between the existing hard- and software of the vehicles and the GREX-dedicated new ones. Figure 5 gives an overview on the described construction with the three language levels.

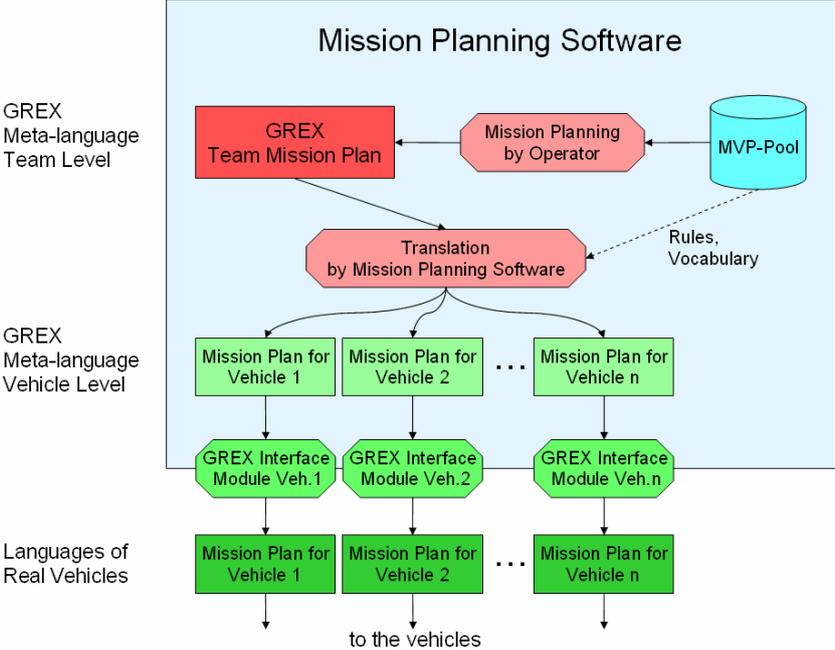


Figure 2: Concept of Team-Orientated Mission Planning

The proposed concept ensures the possibility to expand the mission- and vehicle-pool beyond the GREX-project. Every new mission simply needs to be described with MVPs. Possibly new MVPs need to be defined and translated into the meta-language on vehicle level. New vehicles can be introduced by developing an appropriate GREX Interface Module which in general is necessary not only for mission planning, but also for other tasks like navigation, communication, use of payload etc.

FORMAL LANGUAGE VERIFICATION (FLV)

The different mission plans must be subject to verification for syntax, semantic and logical execution of the mission. The goal of designing Formal Language Verification (FLV) is to present security layer in the GREX mission planning that can detect problems in execution of the mission plan and make the process of creating mission plan and its execution very reliable. To achieve this goal the Formal Language Verification must be accessible after the team mission plan is created of the MVPs, after the team mission plan is translated to the single vehicle mission plans, after the mission plan is uploaded on a specific vehicle and during execution of the mission plan on the vehicle. In every situation the FLV must be proceeded like shown on the Figure 3.

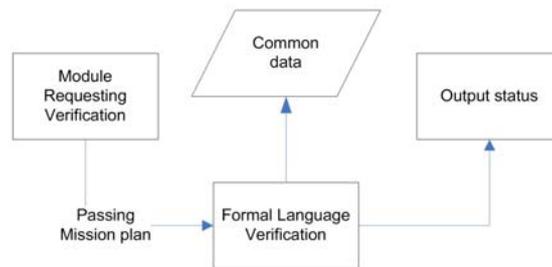


Figure 3: Formal Language Verification usage

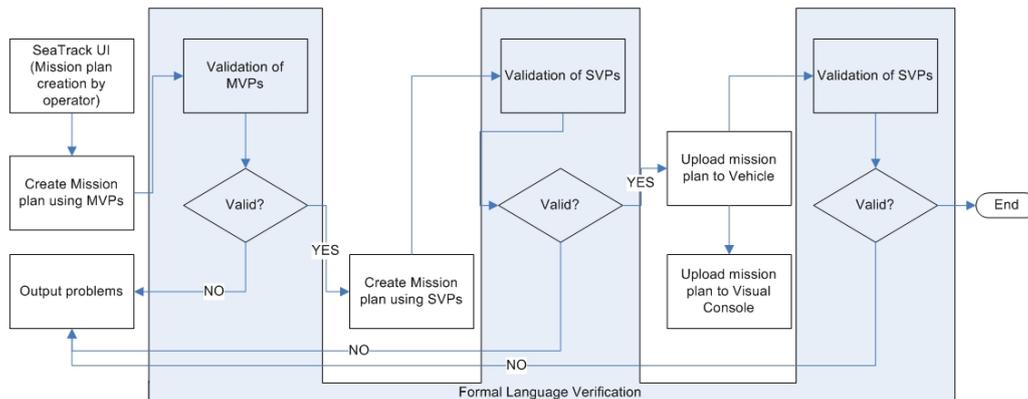


Figure 4: Mission plan creation and verification steps

The already defined steps for verification are integrated in the process of creating a working and verified mission plan that is ready for execution by a team of vehicles. Each mission plan must pass 3 levels of verification. During all of these three steps formal verification on the entire mission plan must be executed. As it can be seen from Figure 4 the first verification is made on MVPs but the other two steps on SVPs.

FLV is a process that must handle different verification aspects. Each of these aspects

must be accessible as stand alone verification or as a sub step of another aspect of verification. The different types of verification can be described like this:

- Syntax and semantic verification – This type of verification has the goal to verify the correctness of the language that defines the current mission plan.
- Vehicle/Payload/Communication constraints verification – The objective of the verification is to prove that all parameters included in the mission plan are correct in relation to the constraints defined for each vehicle. In the meaning of team-oriented mission planning this verification is not mandatory but it is a must for single vehicle-oriented mission planning. However, this step can be performed to the team-oriented mission plan to guarantee that the parameters are not out of range for all vehicles. This kind of verification on the team-oriented mission plan will prevent the translation from MVPs in SVPs if the values are not correct.
- Logical verification – the last and most important kind of verification must define if the mission plan as it is designed can be executed by the selected vehicles. This verification requires all other verifications to be completed with success because only logical algorithms are applied to the mission plan to prove that it is executable.

This kind of verification has the following major objectives:

- o Dependency verification.

One of the rules that must be valid so the mission plan can be executable is to have strong dependencies between every step. This requires that the execution of a specific step from the mission plan must result with parameters that are the same as the initial parameters for the next step.

- o Behaviour verification.

During mission execution it is common for every vehicle to have some predefined behaviours that can be executed but not actually to be described in detail as a part of the mission plan. The objective of such verifications is to minimize the amount of work and the risk when validation of the mission plan is performed.

- o Calculation verification.

The validity of the Mission plan is connected to calculations on every single step defined in the mission plan. The calculations are related with constraints defined by the vehicle, mission, communication, etc.

- o Path validation.

Using the objectives above the FLV must be able also to verify the paths that are defined in the mission plan. One of the major issues in this validation is to simulate the paths of

the vehicles and to verify that no collision or path intercept will occur during execution of the mission plan. If there are any interceptions of the vehicle's path then the role of this verification is to calculate the approximate time that every vehicle will pass the point of interception and to report the result to the operator.

CONCLUSION

The design of a Team-Orientated Mission Planning for unmanned vehicles has to deal with a lot of facts. We presented a concept with three different levels of planning languages that allow a modular and generic mission planning. The process can easily be expanded for new missions or new vehicles. A Formal Language Verification was presented that executes different verifications and allows the online changing of plan parameters.

ACKNOWLEDGEMENTS

The authors would like to thank the European Community for the support of the research project GREX (FP6-IST-2006-035223) which is funded by the Sixth Framework Programme of the European Community.

References:

- [1] D.E. Pritchard, "Dynamic Route Replanning and Retasking of Unmanned Aerial Reconnaissance Vehicles," Masterthesis, Air Force Institute of Technology, Wright-Patterson AFB Ohio, USA, 2000
- [2] B. Leonhardt, "Mission Planning and Mission Control Software for the Phoenix AUV," Masterthesis, Naval Postgraduate School, Monterey, USA, 1996
- [3] C. Barrouil and J. Lemaire, "Advanced Real-Time Mission Management for an AUV," NATO Symposium on Advanced Mission Management and System Integration Technologies for Improved Tactical Operations, Florence, Italy, 1999
- [4] A.L.R. Cormack, "Providing a Common Operator Interface for Port and Harbour security and Shallow Water Mine Warfare," Undersea Defence Technology Europe Conference and Exhibition (UDT Europe), Hamburg, Germany, 2006
- [5] T. Glotzbach and J. Wernstedt: "A Proposal for the Control of Teams of Autonomous Underwater Vehicles (AUVs) in Hierarchical and Peripheral Modes using the Concept of Adaptive Autonomy," 25th International Conference on Offshore Mechanics and Arctic Engineering (OMAE), Hamburg, Germany, 2006
- [6] T. Glotzbach, A. Picini, A. Zangrilli, M. Eichhorn, P. Otto, M. Schneider: "Evaluation of Different Control Strategies for Teams of Unmanned Marine Vehicles (UMVs) Comparing the Requirements of Underwater Communication and Navigation in MATLAB® Simulations," 6th International Conference on Computer Applications and Information Technology in the Maritime Industries (COMPIT), Cortona, Italy, 2007, Conference Proceedings, pp. 331-344

Authors:

Thomas Glotzbach
Peter Otto
Matthias Schneider
Technische Universität Ilmenau, P.O. Box 10 05 65,
98684 Ilmenau, Germany
Phone: +49 3677 461-145
Fax: +49 3677 69 14 34
E-mail: {thomas.glotzbach,peter.otto,schneider.matthias}@tu-ilmenau.de

Martin Marinov
SCIANT, 9 Chamkoria Street,
1504 Sofia, Bulgaria
Phone: +359 2 943 33 31
Fax: +359 2 943 33 37
E-mail: martin.marinov@sciانت.com

M. Arredondo / A. Cormack

SeeTrack: Situation Awareness Tool for Heterogeneous Vehicles

INTRODUCTION

SeeTrack is a mission planning, monitoring and post processing tool for rapid on-site analysis and data fusion of sensor data, including sidescan, imaging sonar, and video. It is a highly modular, equipment-agnostic system and is designed to perform on both notebook and desktop environments. It has been successfully deployed on numerous surveys, military exercises, and scientific experiments.

SeeTrack is platform and manufacturer independent, currently supporting a variety of underwater vehicles (e.g. Remus, OEX, Hugin, Morpheus, Gavia, Autosub, RAUVER) and sensor suites; including a number of sidescan and imaging sonars from different vendors (e.g. Marine Sonics, Klein, Didson), video, CTD, bathymetric and navigation sensors (INS/DVL, LBL, USBL, GPS). It runs on Windows compatible notebooks or desktop computers.

SeeTrack supports concurrent data visualization and processing of the various vehicle data products in an intuitive, easy to use fashion. Crucially, the system exhibits rapid operation, with data import and processing times which are orders of magnitude faster than other commercial systems, suiting in-stride operational needs.

In this paper we present SeeTrack's design concepts and some of its main features, which make it the ideal tool for situation awareness of heterogeneous vehicles.

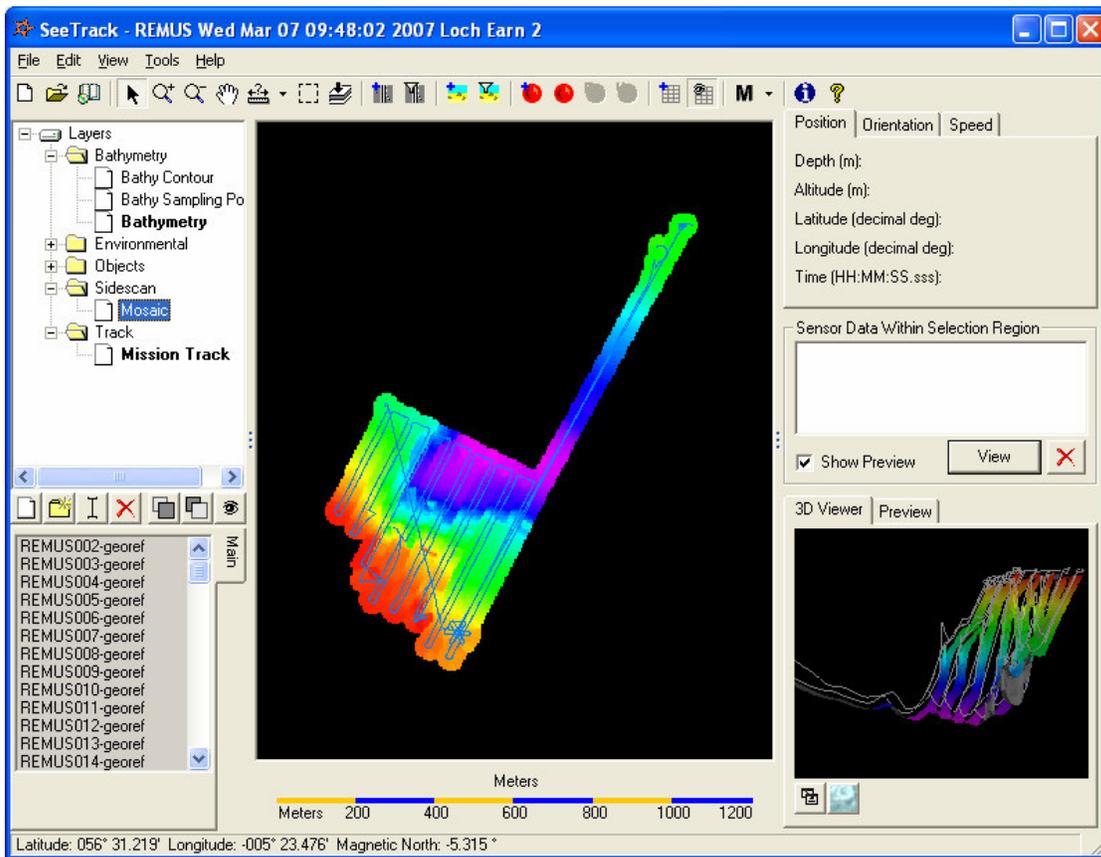


Figure 1 SeeTrack's main window.

DESIGN

SeeTrack has been designed around three main concepts:

1. SeeTrack uses a modular, hot-swap plug-in architecture. The free SeeTrack SDK allows for different vendors of vehicle platforms/sensors to create drivers that allow SeeTrack to import, visualize and analyze vehicle/sensor data. In the same way the SDK allows for the development of plug-ins for processing data (e.g. Seafloor classification, mine detection in sidescan sonar).
2. SeeTrack is designed as a geographical information system (GIS); that is, it allows the geo-location of vehicle and sensor data, which is crucial for mission planning, monitoring and analysis. Furthermore, working on a GIS allows users to import legacy data from other sources (e.g. charts, contacts, locations of interest, satellite images, etc.) For military personnel, working on a GIS allows them to identify potential targets for reacquisition and data fusion.
3. SeeTrack has been developed from the start as an operational tool. It has a very easy to use interface and displays the data in a friendly way. Users can access

data from any mission, process them and export the results to external systems (e.g. MEDAL and MINTACS) in only a few clicks. Finally, SeeTrack capabilities are well adapted to the needs of the user as it only requires a standard laptop to run.

CAPABILITIES

SeeTrack offers several key advantages compared to other software systems:

1. It integrates multiple sensors in one common geo-referenced environment, allowing users to run missions more efficiently. For example, sidescan sonar interpretation is enhanced as three dimensional navigation of the vehicle is available. Another example of the advantages of a common environment is the use of video data to confirm detections of mines made in the sidescan sonar.
2. Its open architecture allows to plug-in processing modules, which can be developed by SeeByte or by third party. SeeByte plug-ins like the Computer Aided Detection and Computer Aided Classification of mines (CAD/CAC) and the Seafloor Classification have proven very successful in different sonars.

SeeByte's CAD/CAC [1] uses a model-based approach, where both the sonar process and the physical parameters of the objects being searched for are considered during the analysis process. This provides several appealing properties. First, as AUV computational capabilities improve, the models can be continuously adapted to include more realistic information and simulations, resulting in improved performances. Second, contrary to the currently favored feature-based CAD/CAC systems which provide a 'black-box' solution, it is possible to determine why the model has provided a given solution. Last, the CAD/CAC is not sonar or platform specific and may be easily ported to access data from different sonars including emergent Synthetic Aperture Sonar (SAS) technologies.

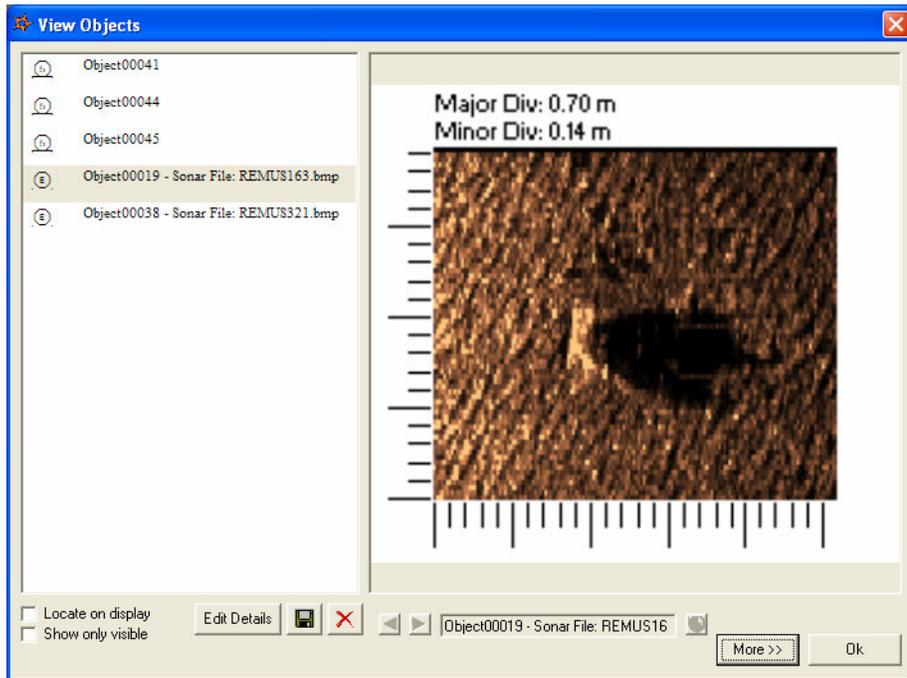


Figure 2 SeeByte's CAD/CAC has been successfully tested on different sidescan sonars.

In terms of navigation processing, SeeTrack can post-process navigation for error correction using an Extended Kalman Filtering with smoothing, ensuring better estimates of true locations of objects on the seabed and a high quality mosaicing of sonar and video data.

The latest addition to the repertoire of SeeTrack add-ins is the Performance Analysis and Training Tool (PATT) [2], which allows the evaluation of Mine Detection and Classification models. PATT uses an augmented reality approach: a mine simulator and sonar renderer model places ground truthed mine objects into real sonar data; a seafloor classification module is used in conjunction to classify the survey region into areas according to mine huntability. These two modules combined allow the evaluation of Automatic Target Recognition (ATR) algorithms in a large variety of scenarios. From the evaluation, Mine Counter Measure (MCM) missions may be planned to maximize the probability of mine clearance. Furthermore, 'what-if' scenarios (increasing sonar resolution, altering AUV trajectory) may also be considered and the results used to impact future MCM operations.

3. The system has the ability to perform fast and accurate mission analysis on a portable, low-cost platform enabling on-site mission assessment and decision

making. For example, a typical two hours REMUS mission with CTD data display and sidescan geo-referencing takes approximately 15 minutes.

SEETRACK AS GREX'S USER INTERFACE

The main object of the European Project GREX [3] is to create a conceptual framework and middleware systems to coordinate a swarm of diverse, heterogeneous physical objects (underwater vehicles) working in cooperation to achieve a well defined practical goal (e.g. search of hydrothermal vents) in an optimized manner.

SeeTrack provides the ideal software tool for the mission planning, monitoring and analysis of missions performed in the context of GREX. Planning of tasks for the heterogeneous vehicles can be performed using SeeTrack's friendly mission planning interface, which allows for simple point and click planning over charts or satellite images. During mission execution, live position and data from the vehicles can be displayed via Ocean-Shell messages [4]. Data can then be loaded and displayed in SeeTrack for post-mission analysis, so that success of the mission and quality of the data gathered can be evaluated.

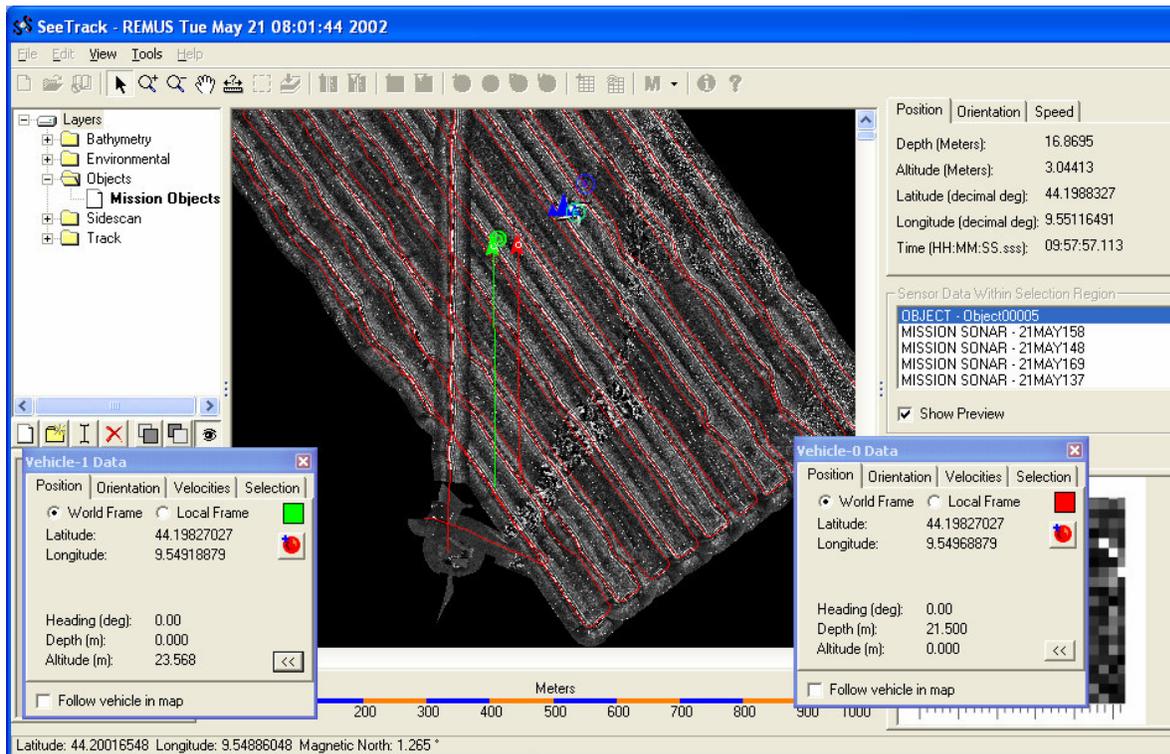


Figure 3 Live position and data from vehicles is displayed in SeeTrack using Ocean-Shell messages.

SUMMARY

The use of multiple sensors and vehicle platforms for AUV operations brings the need for a single interface to plan, monitor, analyze and report all the data products, regardless of the platform used to acquire the data. SeeTrack provides a system that enables this functionality and interoperability in a common application.

This paper has presented SeeTrack's design concepts and some of its main features, which make it the more suitable software tool for the situational awareness of heterogeneous underwater vehicles.

References:

- [1] S. Reed, Y. Petillot, J. Bell. Model-based approach to the detection and classification of mines in sidescan sonar. Applied Optics, Volume 43, Issue 2. Page 237-246. January 2004.
- [2] Y. Petillot, S. Reed, E. Coiras. An Augmented Reality Solution for Evaluating Underwater Sonar MCM Systems. 7th International Symposium on Technology and the Mine Problem. Naval Postgraduate School, Monterey, CA. May 2-5, 2006.
- [3] <http://www.grex-project.eu/>
- [4] 'Ocean-Shell: An embedded library for Distributed Applications and Communications', Ocean Systems Laboratory, Heriot-Watt University.

Authors:

Dr Miguel Arredondo
Mr Alastair Cormack
SeeByte Ltd, Level 7, Orchard Brae House
30 Queensferry Road. Edinburgh EH4 2HS
Phone: +44 (0) 131 447 4200
Fax: +44 (0) 131 447 4911
E-mail: miguel.arredondo@seebyte.com, alastair.cormack@seebyte.com

J. C. Ferreira / P. B. Maia / A. Lucia / A. I. Zapaniotis

Virtual Prototyping of an Innovative Urban Vehicle

ABSTRACT

The manuscript outlines the original solutions adopted to design and powering an original three-wheel lightweight hybrid electric vehicle (HEV) operating for medium and short distance drives in urban environments that has zero CO₂ emissions and has a silent performance. This vehicle is legislated as a tricycle and it can carry three people, i.e., two passengers and one driver. The electrical HEV motorization is carry out by an axial-flux permanent magnet machine (AFPM) with a single toroidal stator being placed between two permanent-magnet (PM) rotor discs. The electrical supply is realized by a fuel-cell-battery core.

INTRODUCTION

Considerable efforts have been expended to develop hybrid electric vehicles (HEVs) as replacements for high-emission cars, buses, and trucks powered by conventional gasoline or diesel engines [1]. The main objective of this work is to describe a virtual prototype of a HEV by the use of a suitable simulation model. This is an important step in the development of the HEVs due to the following two reasons:

- (i) a good virtual prototype allows for proof testing before hardware is assembled, which means likely reduction in the manufacturing cost and time, and
- (ii) new design possibilities can be explored; e.g., study of tradeoffs between sizes of components in the HEV is feasible.

A virtual prototype of a hybrid electric vehicle (HEV) is created within the virtual test bed (VTB) environment, which has been developed for modeling, simulation, analysis and virtual prototyping of large-scale multi-technical dynamic systems.

Some attention is also committed on the electric system, which is composed of:

- (i) a fuel cell system as a prime power source,
- (ii) battery and super capacitor banks as energy storage devices for high and

- intense power demands,
- (iii) DC-to-DC power converters to control the flow of power,
 - (iv) a three-phase inverter-fed permanent magnet synchronous motor as a drive,
 - (v) and a common DC bus.

VIRTUAL PROTOTYPING OF THE URBAN VEHICLE

In the last two years a research project was carried out to model and construct the prototype of three-wheel HEV. This is a lightweight vehicle intended for use in urban mobility with mission tasks such as 50 km/h cruising speed and 80 km range of autonomy. The propulsion system is arranged with a 10.0 kW prototype of slotless AFPM being totally enclosed in the twin rear wheels of the vehicle and fed from a fuel-cell-battery. The HEV urban vehicle is legislated as a tricycle and it can carry three persons, two passengers and one driver. The energy is generated from an electric motor, alimented from a fuel-cell-battery core.

The casing body dimensions of the HEV urban vehicle to accommodate the three persons and the electrical motorization are shown in Figure 1.

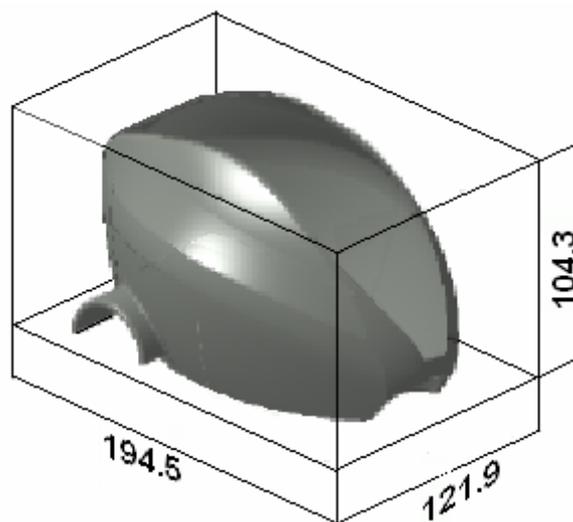


Figure 1: Dimensions of the casing body of the HEV urban vehicle.

The objective of geometric models is to show the form and the design of the product. In these cases the prototype is either designed in great detail comprising a manipulation model, either it is roughly designed representing a conceptual form of the product under design.

The casing body of the three-wheel lightweight electric vehicle was been designed in CAD Solid-Works™ and studied through Virtual Prototyping (VP). The results are exposed in Figure 2.

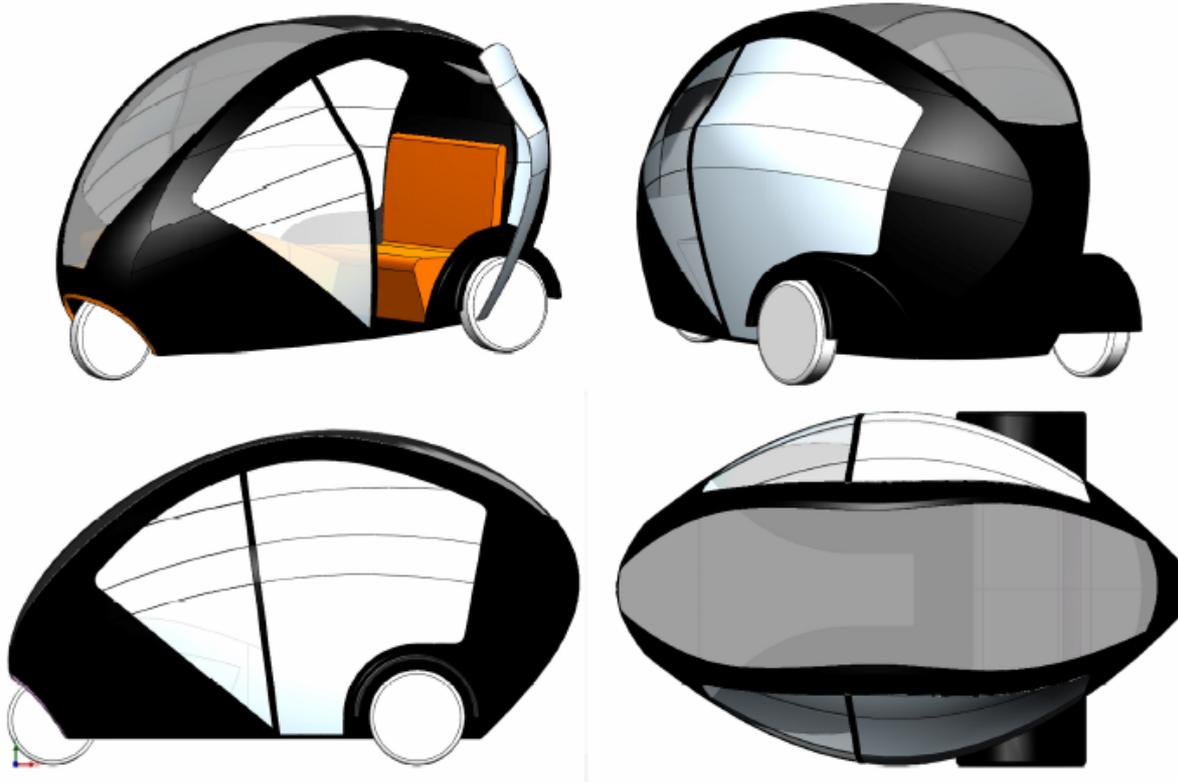


Figure 2: VP of the three-wheel lightweight electric vehicle draft in Solid-Works™.

In this work, the virtual test bed (VTB) was utilized for virtual prototyping of a HEV. The VTB has two important features [2,3]:

- (i) it has the capability of integrating models that have been created in a variety of languages into a single simulation environment; and
- (ii) it provides advanced visualization of simulation results, including full-motion animation of mechanical components, and imaginative mappings of computed results onto the system topology.

The first feature of the VTB allows each component of a large-scale multi-technical system to be described in the most appropriate language. On the other hand, the second feature enhances user's comprehension of the simulation results significantly.

FUELL CELL SYSTEM

A simplified block diagram for the electrical part of the HEV [4] is schematized in Figure 1. The components classifications of the fuel cell system are: battery bank, super capacitor bank, boost converter, DC Cuk converter, PMSM, and the PWM inverter.

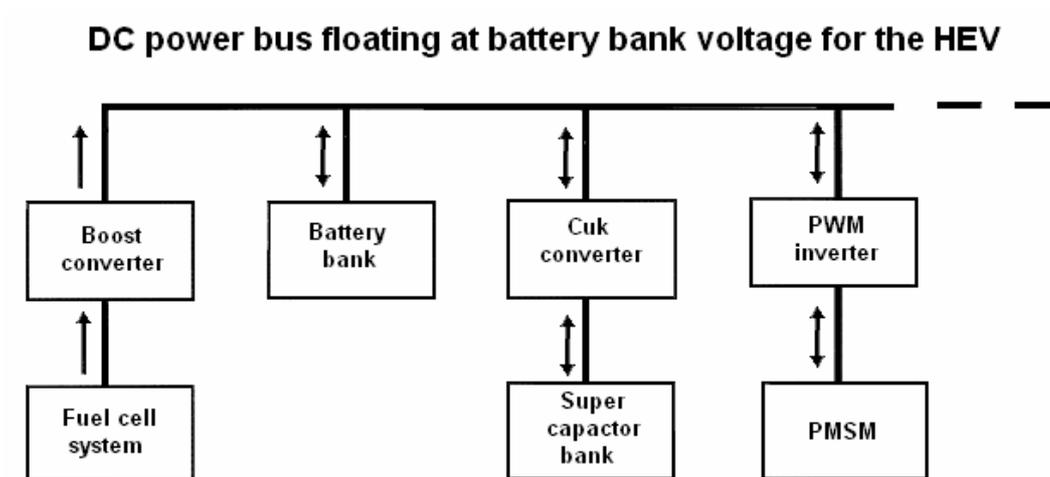


Figure 2: Block diagram for the electrical part of the HEV.

The voltage-current characteristic of a single proton exchange membrane (PEM) hydrogen fuel cell is illustrated in Figure 2. In this figure, V_{fc} (vertical axis) is the voltage at the terminals of the fuel cell, and I_{fc} (horizontal axis) is the current flowing out of the fuel cell. It is seen that there are basically three operation regions. These are:

- (i) the low current region in which the voltage decreases exponentially as the current increases,
- (ii) the linear region that covers a large portion of the characteristic, and
- (iii) the high current region in which there is a sharp drop of the voltage to near-zero [5,6].

Note: the units for V_{fc} and I_{fc} are millivolts and milliamperes, respectively.

For the values of I_{fc} which remain in the low current and linear regions, V_{fc} versus I_{fc} may be expressed by the graph in Figure 3 [7].

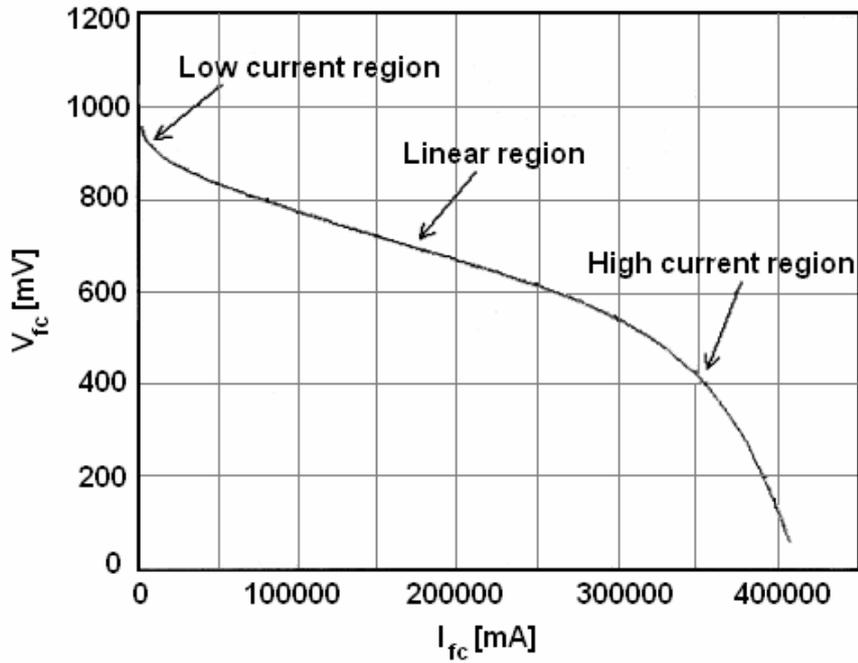


Figure 3: The fuel cell voltage in mV versus fuel cell current in mA.

ELETRICAL MOTORIZATION

The electrical HEV motorization is carry out by axial-flux permanent magnet machine (AFPM) with a single toroidal stator being placed between two permanent-magnet (PM) rotor discs. These systems prove to be the best candidates for such a low-speed high-torque drive application, as they can be designed to achieve the required high torque density without loss of efficiency. In addition, their disc shape is very well suited to housing the motor in a wheel rim as the double PM rotors could be mounted on the wheel side walls and the stator could be mounted centrally on the wheel axle, as shown in Figure 4.

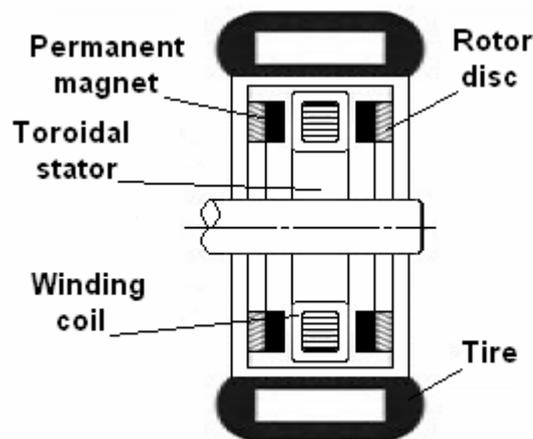


Figure 4: Cross-sectional view of a AFPM being used as wheel direct-drive motor.

In the AFPMs the machine stator, being mounted centrally on the wheel axle, consists of a coils arranged to form a three-phase winding wound in a toroidal fashion. The rotor comprises two mild-steel discs, one on each side of the stator, being mounted on the wheel side walls and carrying axially-polarized magnets.

CONCLUSIONS

A virtual prototype for a three-wheel HEV was developed devoted to urban mobility and numerically verified by simulation results within the urban environment.

One of the unique features of the virtual prototype is that it includes all possible energy devices (fuel cell system, battery bank, and super capacitor bank) for the next generation HEVs.

Further, to be consistent with the real world applications, the nonlinear dynamics, ohmic losses, and voltage/current limits of the components are taken into account.

Acknowledgment:

The authors would like to express its gratitude to Fundação para a Ciência e Tecnologia the support of project POCI/2010.

References:

- [1] Riezenman MJ. Engineering the EV future. IEEE Spectrum 1998;35(11):18–20.
- [2] Dougal RA, Brice CW, Pettus RO, Cokkinides GJ, Meliopoulos APS. Virtual prototyping of PCIM systems – the virtual test bed. In: Proceedings of the 37th International PCIM '98 Power Electronics Conference, November 7–13, 1998, Santa Clara, CA. p. 226–34.
- [3] Brice CW, Gökdere LU, Dougal RA. The virtual test bed: an environment for virtual prototyping. In: Proceedings of the International Conference on Electric Ship (ElecShip '98), September 1, 1998, Istanbul, Turkey. p. 27–31.
- [4] Gökdere LU, Benlyazid K, Santi E, Brice CW, Dougal RA. Hybrid electric vehicle with permanent magnet traction motor: a simulation model. In: Proceedings of IEEE International Electric Machines and Drives Conference (IEMDC '99), May 9–12, 1999. Seattle, Washington. p. 502–4.
- [5] Gökdere LU, Brice CW, Dougal RA. An advanced modeling and simulation tool for analysis of mechatronic systems. In: Proceedings of the Seventh Mechatronics Forum International Conference (Mechatronics 2000), September 6–8, 2000, Atlanta, GA.
- [6] Blomen LJMJ, Mugerwa MN. Fuel cell systems. New York: Plenum Press, 1993.
- [7] Ogburn MJ, Nelson DJ, Wipke K, Markel T. Modeling and Validation of a Fuel Cell Hybrid Vehicle, Future Car Congress 2000, Washington, DC.

Author:

Prof. José Carvalho Ferreira
Eng. P.B. Maia, Eng. A. Lúcia and Eng. A.I. Zapaniotis
IDMEC & In+, Instituto Superior Técnico, TULISBON
Av. Rovisco Pais / 1094-001 Lisboa / Portugal
Phone: +351 (0) 21 8417644
Fax: + 351 (0) 21 841 9058
E-mail: joseferreira@ist.utl.pt

A. Wenzel / A. Gehr / T. Glotzbach / F. Müller

Superfour-in: An all-terrain wheelchair with monitoring possibilities to enhance the life quality of people with walking disability

ABSTRACT

This paper presents an overview of a project in which a concept of safe mobility for walking impaired people in an outdoor environment was developed and implemented. A prototype of an innovative commercial outdoor wheelchair was equipped with a special surveillance technology which monitors the state of the vehicle and the driver and contacts a control center if an emergency or a fault occurs. The system was successfully tested in rough terrain. This project results are the basis for the further product development, like a rental system for such vehicles in touristic regions.

1. INTRODUCTION

More than 4 million people with walking disabilities live in Germany. Approximately 350000 of them have to use a wheelchair. According to the demographic trend the number of walking disabled people will grow in the next decade. People with such disabilities are handicapped in multiple areas of life. In the framework of the research project TAS-2 a consortium of the Fraunhofer Application Center System Technology Ilmenau and the Technical University of Ilmenau explored possibilities to enhance the quality of life for disabled people during their holidays in a model region in the Thuringian Forest. The main focus of this paper is put on the support for walking disabled people in outdoor activities. This paper primarily describes the work which was done to develop an intelligent outdoor wheelchair system. Recent work on this area mainly was concentrated on the tasks of adding autonomous driving functions to wheelchairs [1,4]. Further strategies for driver assistance like obstacle avoidance, intelligent path planning and convoy functions have been developed [2,3]. Most of this work has been done in the context of indoor driving. In [4] the need of developing driver assistance systems and not of autonomous driving is accentuated. According to the aim of TAS-2 to support people

with walking disabilities in touristic outdoor activities, an outdoor wheelchair was equipped with a surveillance technology which allows the monitoring of the vehicle and the driver state due to the following reasons: Individuals with a walking disability are usually not able to save themselves in the case of technical malfunctions or accidents. Moreover, this group of people often suffers from several medical limitations. Especially in the case of the poorly populated model region in the Thuringian forest an automatically autonomous surveillance technology is absolutely necessary.

2. SYSTEM DESIGN

2.1 Platform

Most of the commercial available electric powered wheelchairs are not able to drive over rough terrain or even dirt roads for relevant distances. This leads to the fact that most of the relevant touristic areas of the Thuringian Forest are unreachable. During the mentioned project several wheelchairs were evaluated. A prototype of the outdoor wheelchair SuperFour© of the company Otto Bock Healthcare GmbH did apply for the described problems. The new created system of SuperFour and the developed monitoring technology was called 'Superfour-in'. A number of innovative features of the base platform SuperFour were adopted by the Superfour-in system. The wheelchair is powered by a four wheel drive with two steerable axles. A hybrid system of accumulators and a combustion engine does provide an operation range of more than 100 km. Furthermore the wheelchair is able to overcome pitch and bank slopes up to 40 percent. The seat which is in a central position has an automatic balancing system and can be driven out of the wheelchair for an easy transfer from or to a normal wheelchair. These features of the platform provide an excellent all-terrain ability of the system.

2.2 Components

In figure 1 the general concept of Superfour-in is shown. In a so called "Care Service Center" (CSC) several of the special vehicles are supervised. This CSC can be located at a rental station for instance where it is possible for people with walking disabilities to rent such vehicles. Each vehicle has an onboard "Vehicle Control Center" (VCC). This component described later collects technical data of the vehicle and medical data of the driver. These data is automatically interpreted by the developed software of the VCC. The user interface of the VCC gives information about navigation and the vehicle state to

the driver. The VCC is connected via General Packet Radio Service (GPRS) to one CSC. In case of an emergency the CSC alerts a supervisor. Because of the permanent position transmission from VCC to the CSC rescue activities can then be started very effectively.

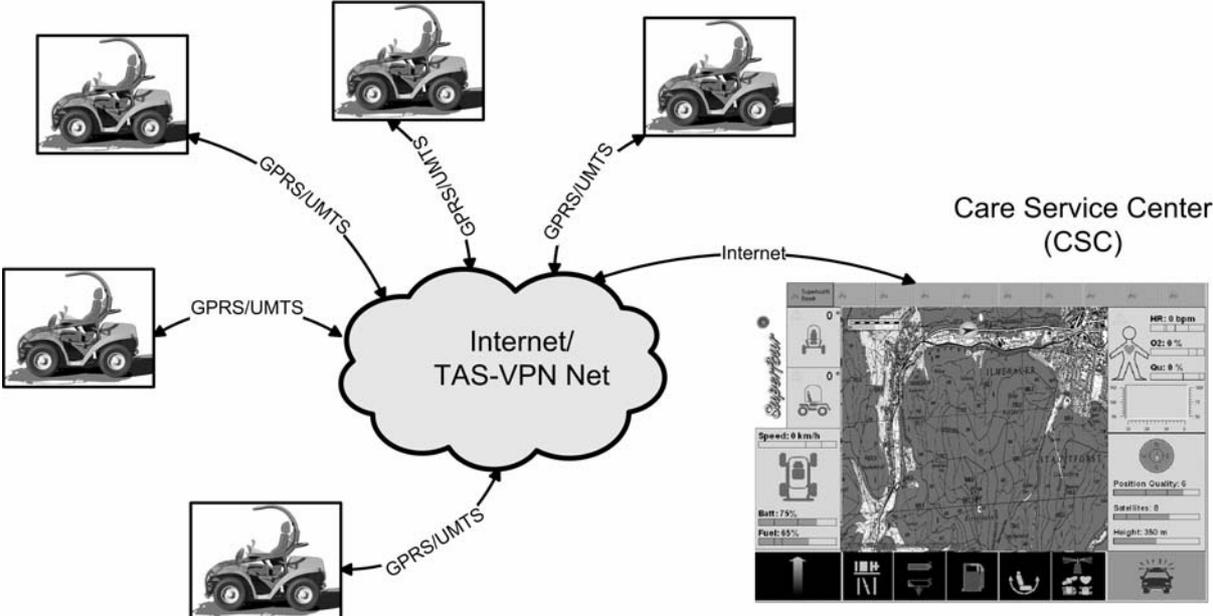


Figure 1: General concept of the system Superfour-in

2.3 Vehicle Control Center (VCC)

The structure of the developed vehicle control center is shown in figure 2. The main component is a waterproof tablet PC that runs the VCC application. This software collects the vehicle data and the medical data from the driver. The screen of the tablet PC is mounted in view of the driver so he can see the navigation information on the presented map and the state of the vehicle. The essential handling operations of the VCC are done by using the buttons on the tablet PC. According to figure 2, the vehicle information is collected in two different ways. The information about the state of the electrical drive is transmitted via a CAN-bus network so that the VCC Software can read this essential information. Some of the evaluated information is for instance the currents and the rotation speed of the four electrical engines and the state of the batteries. This information is used to detect or predict malfunctions like a flat tire, defect engine or empty battery. For the extraction of other important technical data a special embedded system was developed. This microcontroller application collects several additional sensor values. Anymore this circuit communicates via a special three wire serial bus with the control of the combustion engine of the vehicle. Values like the fuel and engine state

are acquired on this way. An important information in this context is the current longitudinal and transversal angle of the vehicle. By the use of this information the VCC warns the driver prior to a rollover or it is able to detect that a rollover has happened. The mentioned embedded system collects all these data and sends it cyclically via an USB-interface to the tablet PC where the information is interpreted by the VCC-Software. In addition the current position of the vehicle is determined by a GPS receiver and is sent to the VCC. Not only technical measurements are evaluated by the VCC. The developed software provides an interface for bio-sensors. In this way biosensors with cyclical or event-driven messages can easily be integrated into the system. For the application on a vehicle, only noninvasive robust measurements are considered. Furthermore these measurements should be easily applicable to the driver. Measurements of cardiovascular parameters are relevant for many diseases. Therefore as an example application a photoplethysmography sensor was integrated. The evaluating processor unit of this sensor provides information about current heart rate and the arterial oxygen saturation of the driver. Additionally a parameter which represents the current signal quality is available. Some kind of medical problems, like tachycardia can be automatically detected in this way.

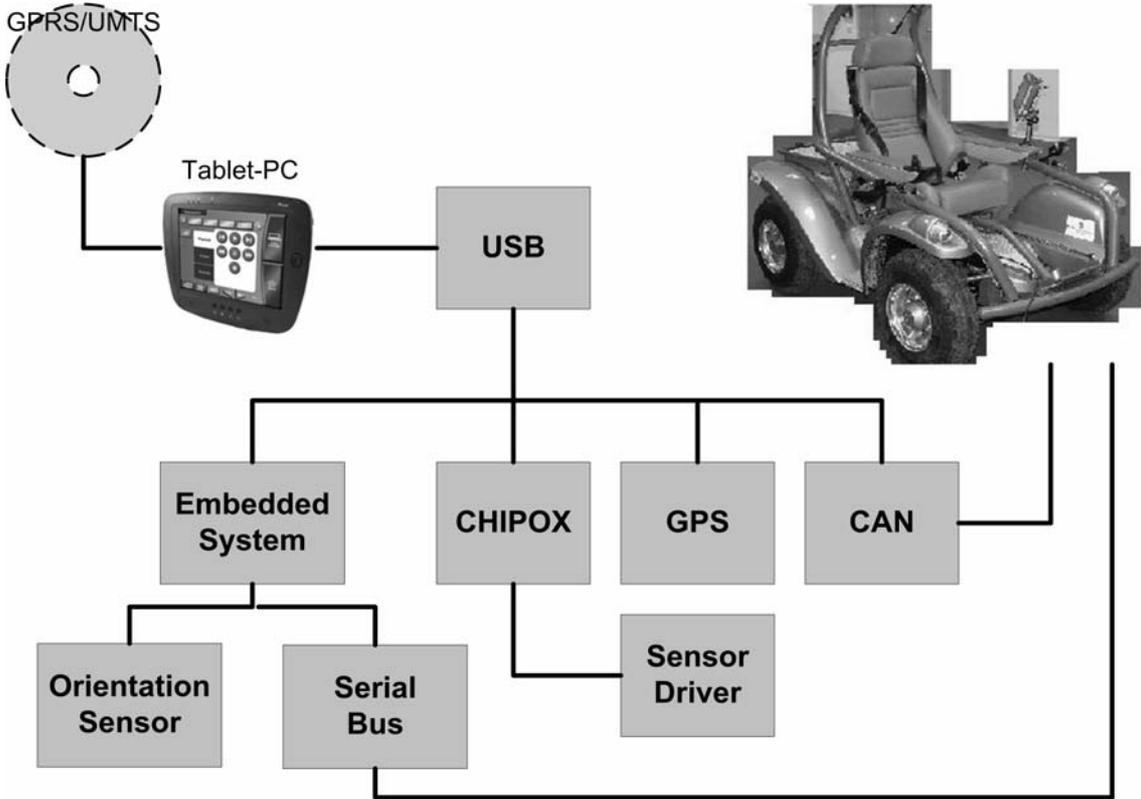


Figure 2: Structure of the Vehicle Control Center (VCC)

2.4 Care Service Center (CSC)

The developed CSC consists of a PC or Laptop which needs an internet connection and a special CSC-Software. These systems can be located for example at a rental station for Superfour-in systems or at a hotel reception desk. When the software is started it waits for incoming connections of VCC's. Once the communication is established the position of the relevant vehicles is shown on a map. By selecting a special vehicle, the current state of the vehicle and the medical data of the driver can be observed. If some of the connected VCC's report an automatically detected dangerous state or if one of the drivers sends an emergency call the system warns the supervisor acoustically and marks the relevant system in the map. Then the supervisor can try to contact the driver or if some kind of emergency has happened, he can initiate the recovery actions.

3. TEST RESULTS



Figure 3: The system Superfour-in outdoor tests

The developed system Superfour-in was sufficiently tested by different people with and without walking disabilities in different situations in the above mentioned region. The tests were partially carried out under snowy conditions. Especially the all terrain abilities and the weatherproofness met the demanded requirements. The system showed a good behaviour in both, road traffic and off road situations (Fig. 3). There are only a few available solutions for a mobile measurement of relevant medical data at the moment. This area of medical equipment technology in our opinion needs further research and development. It can also be stated that localisation with standard GPS technology offers good suited solution for the traffic on mapped public roads, but is often not exact enough regarding to the localisation of smaller objects on off road terrain where only rough

mapping is available. Therefore after some testing more precise GPS-receivers with dead reckoning had to be implemented. The area-wide availability of GPRS has to be guaranteed in every region where the system is intended to be used. During the tests in some areas there was only a weak GPRS-network available. In certain situations, it may be helpful to negotiate with the network providers or to take into account the usage of backup systems for critical areas based on narrow band radio data transmission. Such systems are commercial available and have been successfully evaluated.

4. CONCLUSIONS

The presented system Superfour-in was developed by equipping a prototype of a commercial hybrid powered outdoor wheelchair with a special surveillance technology. The system consists of an onboard vehicle control center which monitors the state of the vehicle and the driver. Furthermore the vehicle control center supports the user by navigation and with other relevant information. The developed care service center software allows to track and to support several vehicles. The whole system was successfully tested by impaired and non-impaired people under rough outdoor conditions. The results of this work are the fundament for the development of new touristic services especially for people with walking disability.

5. ACKNOWLEDGMENT

This work was funded by the German Federal Ministry of Education and Research (Promotional reference: 03i2818B)

References

- [1] Mazo M.: An Integral System for Assisted Mobility. In IEEE Robotics & Automation Magazine; 2001, March; p. 46-56
- [2] Schilling K.; Roth H.: Convoy Driving and Obstacle Avoidance Systems for Electrical Wheelchairs. In Wang P. P. (ed.): Proceedings of the Fourth Joint Conference on Information Systems; 1998; p 353-356
- [3] Yanco, H; Wheellesley A.: A Robotic Wheelchair System: Indoor Navigation and User Interface. In Mittal et. al. (eds.): Assistive Technology and AI. In LNAI 1458, Springer-Verlag; 1998
- [4] Lankenau A., Röfer T.: Smart Wheelchairs – State of the Art in an Emerging Market. KI, 4100; 2000, p. 37-39
- [5] Nisbet P. D.: Who's intelligent ? Wheelchair, driver or both ? In Proceedings of the 2002 IEEE International Conference on Control Applications, 2002; p. 760-765

Authors:

Dr. Andreas Wenzel

Achim Gehr

Thomas Glotzbach

Fabian Müller

Fraunhofer Application Center System Technology (AST), Am Vogelherd 50

D-98693 Ilmenau, Germany

Phone: +49 3677 461-144

Fax: +49 3677 461-100

E-mail: {andreas.wenzel,achim.gehr,thomas.glotzbach,fabian.mueller}@ast.iitb.fraunhofer.de

Th. Krause

Verteiltes, dynamisches Antriebssystem zur Steuerung eines Luftschiffes

Einleitung

Luftschiffe sind im heutigen Luftfahrtbild eher Exoten. Es gibt jedoch immer mehr Anwendungsbereiche, in denen die Vorteile eines mit Traggas gefüllten, „schwebenden“ Fluggerätes die Nachteile, wie hohe Windanfälligkeit aufwiegen. Bei Messaufgaben ist ein Luftschiff als Messplattform ideal, da es keine Luftverwirbelungen verursacht und so Messergebnisse nicht verfälscht. Für die Überwachung und Suche eignet es sich besonders, weil es sehr leise fliegt und in der Nähe von Menschenmassen operieren kann, ohne diese zu gefährden. Durch seinen geringen Energieverbrauch ist es ideal für längerfristige Überwachungen [1]. Ein sehr interessanter Anwendungsbereich ist das teilautonome und autonome Fliegen, der den Einsatz auch für Sicherheitskräfte wie Technisches Hilfswerk, Feuerwehr oder Polizei für die Überwachung und das Sammeln von Informationen interessant macht. Dies setzt jedoch eine gute Regelung der Fluglage und der Position des Luftschiffes im Raum voraus.

In dieser Veröffentlichung wird bereits beim Design des Antriebskonzeptes der regelungstechnische Aspekt betrachtet. Präsentiert wird ein auf diesen Überlegungen basierendes neues Antriebskonzept für Luftschiffe, welches eine hohe Manövrierbarkeit und ein direktes Reagieren auf Störgrößen ermöglicht. Dabei werden in einem weiteren Schritt Ausblicke auf die dann umzusetzende Regelung betrachtet. Erste Erprobungen einer solchen Regelung wurden bereits in Lissabon an einem kleineren Modell durchgeführt [2]. Das hier vorgestellte System wird auf einem 9 Meter Luftschiff erprobt.

Das Antriebskonzept

Die Regelung von physikalischen Größen vereinfacht sich, wenn man mit Stellgrößen direkt oder indirekt über einfache Modelle auf diese Größen Einfluss nehmen kann. Bei der Steuerung eines Luftschiffes sollen idealer Weise alle sechs Freiheitsgrade geregelt werden. Die Störgrößen sind hier Kräfte, die durch Luftbewegungen und Gravitation entstehen. Von außen wirken Luftbewegungen auf den Ballon. Dadurch wird das Luftschiff aus seiner Lage gebracht. Thermik und Temperaturschwankungen des Traggases durch Sonneneinstrahlung verändern den Auftrieb des Luftschiffes. Bisherige Lösungen weisen zwei am Bauch des Luftschiffes befestigte Antriebe auf, die sich um die Querachse drehen lassen, so dass ein Kräftevektor in Flugrichtung und vertikal aufgebaut werden kann. Die beiden Antriebe werden dabei oft synchron angesteuert [3]. Mit einer solchen Anordnung lassen sich nur die zwei translatorischen Bewegungen Vorwärts und die Höhe regeln. Seitenkräfte und Lageänderungen können damit nicht korrigiert werden. Hinzu kommt, dass die Aktoren an einer anderen Stelle wirken als die Störgrößen, wodurch bei Korrekturen Drehmomente entstehen.

An ein einfach zu regelndes Antriebssystem, welches eine hohe Manövrierbarkeit realisiert, werden daher folgende Anforderungen gestellt:

- Der resultierende Kräftevektor soll dort wirken, wo die Störkräfte wirken
- Der Kräftevektor soll sich in der Richtung so verändern lassen, dass er den Störkräften direkt entgegenwirkt

Wenn man diese Forderungen erfüllt, kann jede Größe einzeln für sich geregelt und das sehr komplexe System einfach entkoppelt werden.

Ordnet man mindestens drei Antriebe unabhängig ansteuerbar um den Ballon an und befestigt diese so, dass der Schubvektor parallel zur Ballonoberfläche wirkt und um die orthogonale Achse zur Ballonoberfläche gedreht werden kann, erreicht man die oben genannten Forderungen (siehe Abbildung 1). Damit lassen sich die Kräftevektoren einzeln drehen und in ihrem Betrag variieren. Dadurch kann der resultierende Kraftvektor in der Fläche, die die Antriebe aufspannen, positioniert und ausgerichtet werden. Gleichzeitig können auch die erforderlichen Drehmomente erzeugt werden um Lageänderungen, etwa durch Trägheiten, auszuregeln. Eine entsprechende Anmeldung finden Sie in der Patentschrift [5].



Abbildung 1: Beispielanordnung für die Antriebe

Durch Vektoraddition und unter Berücksichtigung der Geometrie des Antriebssystems können Kräfte in allen 6 Freiheitsgraden erzeugt werden. Diese orthogonalen Kräfte (3 translatorische Kräfte und 3 Drehmomente) werden im Folgenden als Grundkräfte im Raum beschrieben. Umgekehrt kann man für jede Grundkraft die dazu notwendigen Propellerstellungen und Stärken berechnen. Überlagert man dann diese einzeln ausgerechneten Propellerstellungen und Stärken erhält man die Ansteuerwerte für die Antriebe für die gewünschte Bewegung des Luftschiffes. Das heißt, man entkoppelt zunächst jeden Freiheitsgrad und entwickelt für jeden Freiheitsgrad einen eigenen Regelkreis. Daraus resultieren für jeden Antrieb sechs Kräfte mit sechs Winkeln. Dabei treten als Winkel jeweils nur 0 Grad oder 90 Grad auf. Abbildung 2 zeigt drei Beispiele für die Stellung der Antriebe für die Einzelbewegungen.

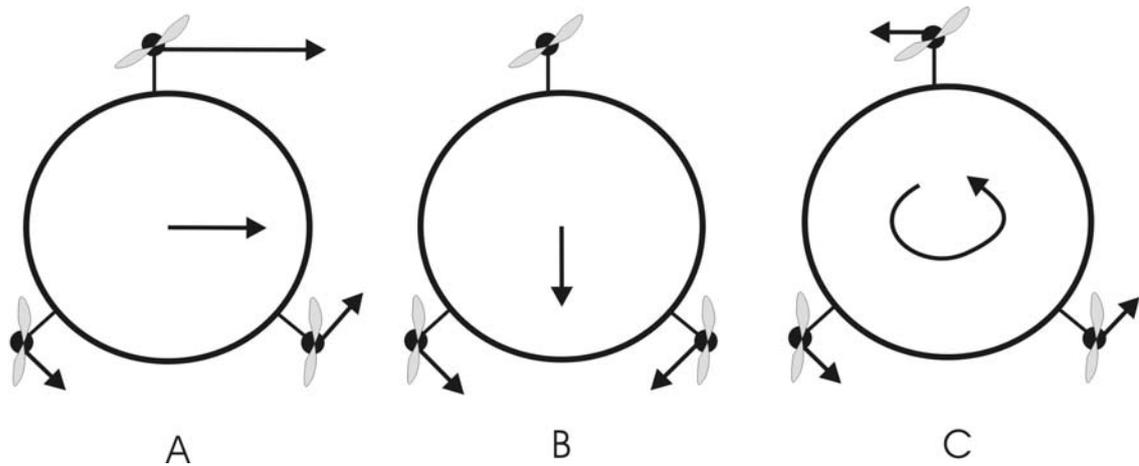


Abbildung 2: Drei Beispielbewegungen und ihre Antriebskonfiguration. A) Bewegung zur Seite B) Bewegung nach unten C) Drehung um die Längsachse

Unter dieser Voraussetzung lassen sich alle sechs Vektoren addieren, wodurch sich ein resultierender Vektor ergibt, mit dem der Antrieb angesteuert werden muss. Die Schwierigkeit liegt hier jedoch in der Endlichkeit der Hardware. Man kann die Stellgrößen für jede einzelne Regelstrecke begrenzen, so dass die Ergebnisse innerhalb der physikalischen Grenzen der Antriebe liegen. Nach der Addition der sechs Ergebnisse kann das Resultat jedoch trotzdem außerhalb dieser Grenzen liegen. Hier ist eine geschickte Synchronisation mit Rückwirkung auf die Führungsgrößen erforderlich. Für spätere autonome Aufgaben ist es vorteilhaft, wenn sich das Luftschiff so genau wie möglich in die vorgegebene Richtung oder auf dem vorgegebenen Pfad bewegt. Die Geschwindigkeit spielt dabei eine untergeordnete Rolle. Die Synchronisation wird also so durchgeführt, dass das Verhältnis der Kräfte der Antriebe zueinander gleich bleibt. Es wird lediglich der Betrag so weit reduziert, dass alle Antriebe innerhalb ihrer Grenzen bleiben. Rückführend müssen dann die Führungsgrößen reduziert werden, um die Regler wieder in einen Regelbereich zu bringen, in dem die Stellgrößen nicht mehr beschnitten werden und die Führungsgrößen eingehalten werden können.

Ein vereinfachtes Beispiel:

Das Luftschiff soll sich mit einer bestimmten Geschwindigkeit v und einer Drehrate r bewegen. Die resultierende Bewegung wäre also eine Kurve mit dem Radius x . Bei dieser Bewegung erhöht sich auf einmal der Wind von vorn und der Regelkreis für die Vorwärtsbewegung erhöht die Ansteuerwerte der Antriebe. Nach der Addition der Vektoren stellt sich dabei heraus, dass ein Antrieb die geforderte Kraft nicht bringen kann. Die Synchronisation reduziert die Motorleistung aller Motoren, so dass deren Verhältnis der Kräfte gleich bleibt. Gleichzeitig reduziert die Synchronisation die Führungsgrößen für Drehbewegung und Vorwärtsbewegung. Dabei bleibt das Verhältnis der beiden Führungsgrößen ebenfalls erhalten. Als Ergebnis fliegt das Luftschiff immer noch die Kurve mit dem Radius x aber mit einer geringeren Geschwindigkeit.

Würde man diese Anpassung nicht vornehmen, würde das Luftschiff in diesem Fall eine engere Kurve fliegen.

Das Luftschiff

Der Lehrstuhl Prozessautomatisierung der TU-Chemnitz verfügt über ein 9 Meter langes Luftschiff auf dem eine leicht abgeänderte Antriebskonfiguration aufgebaut wurde. Abbildung 3 zeigt den Versuchsaufbau während des ersten Testfluges.



Abbildung 3: Das Luftschiff mit dem neuen Antriebssystem beim ersten Testflug

Die Antriebe einzeln um den Ballon zu befestigen ist bei unserem Prallluftschiff technisch sehr aufwendig. Deshalb wurde eine etwas geänderte Anordnung gewählt. Am Ballon befindet sich jeweils oben und unten eine Gondel an der links und rechts an einer drehbaren Achse ein Antrieb befestigt ist. Alle vier Achsen und Antriebe lassen sich unabhängig voneinander ansteuern. Abbildung 4 zeigt links die einfache Mechanik zur Aufnahme der Achsen und rechts die Anordnung dieser beiden Mechaniken.

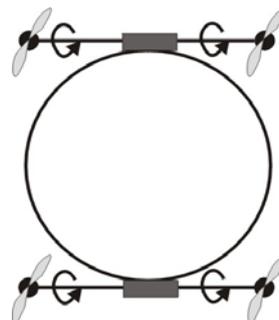
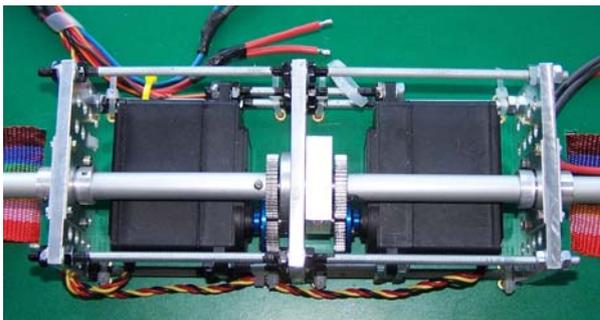


Abbildung 4: links: Mechanik zur Aufnahme und Steuerung der Antriebsachsen; rechts Anordnung der zwei Mechaniken mit den Antrieben.

Durch diese Anordnung werden die oben genannten Forderungen erfüllt. Lediglich eine Bewegung zur Seite kann nicht realisiert werden. Hierfür muss das Luftschiff gedreht werden.

Mit dieser Konstruktion kann die Regelung der Bewegung des Luftschiffes stark vereinfacht werden. So können kaskadierte Regler mit einer modellbasierten Steuerstrecke die Einzelbewegungen (vorwärts, hoch, drehen, neigen, rollen) mit klassischen Reglerstrukturen übernehmen. Die komplexe Bewegung des Luftschiffes

wird dann in einer Ebene darüber über eine Vektorregelung vorgenommen. Dieser Regelungsansatz wurde bereits in Lissabon an einem kleineren Luftschiff erfolgreich getestet [2][4]. Abbildung 5 zeigt die Reglerstruktur für die Einzelbewegungen. Die Bewegungsmodelle können für jede Einzelbewegung individuell angepasst werden.

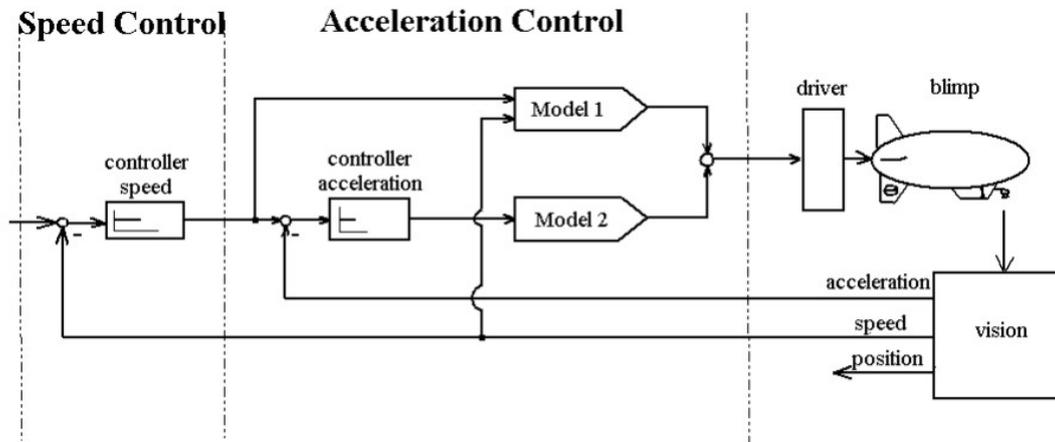


Abbildung 5: Kaskadierte Reglerstruktur zur Regelung der Einzelbewegungen

Abbildung 6 zeigt die globale Reglerstruktur zur Regelung der Lage und Flugbewegung. Da die Bewegungen auf der unteren Ebene bereits geregelt werden und auf dieser Ebene lediglich die Führungsgrößen für die Bewegungen berechnet werden, sind keine Modelle des Systems notwendig. Diese ergeben sich aus den Regelungskreisen der unteren Ebene. Die Level 1 Regler sind auf diesem Schiff 5 Regler. Gezeigt werden aus Platzgründen nur drei Strecken.

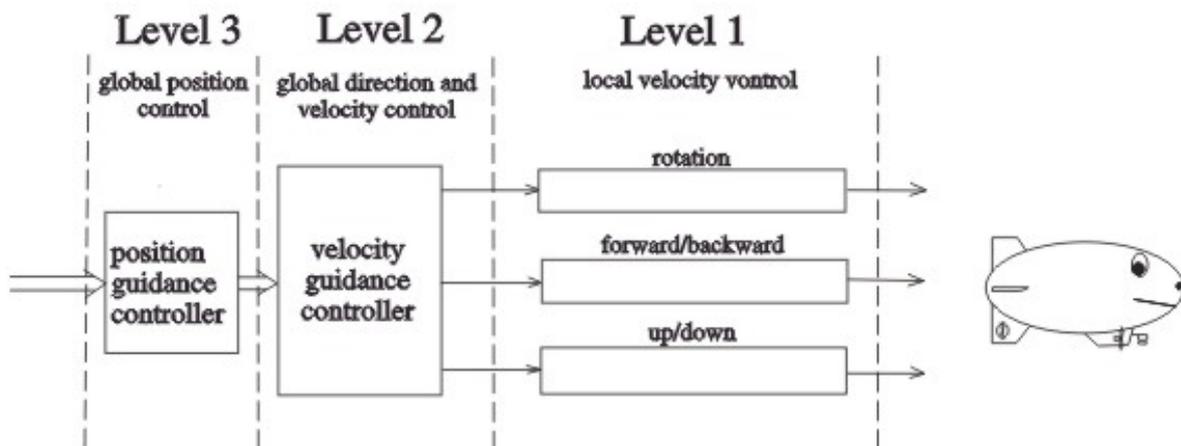


Abbildung 6: Globale Reglerstruktur zur Steuerung der Bewegung des Luftschiffes

Weitere Arbeiten

Nachdem beim ersten Testflug die Funktionalität des neuen Antriebssystems getestet wurde, werden nun die Übertragungsfunktionen der Antriebe zur Linearisierung in der Software integriert. Zum Testflug wurden nur drei Bewegungen (Höhe, Vorwärts und Drehung um die Hochachse) gesteuert. In der nächsten Stufe soll jetzt das Antriebssystem auch in der Software voll ausgebaut werden, so dass alle 5 Bewegungen gesteuert werden können. Zusätzlich zum Mikrokontroller, der die

Ansteuerung der Aktoren übernimmt und die Akkuspannung überwacht, bekommt das Luftschiff ein PC-System mit GPS und Beschleunigungssensoren sowie einem Stereokamerasystem. Auf dem PC wird dann die Regelung implementiert und die Auswertung der Sensoren zur Positionsbestimmung. In der ersten Stufe soll das Luftschiff mit Hilfe der Regelung in der Lage sein, vorgegebene Navigationspunkte abzufliegen. Abbildung 7 zeigt die weiteren Komponenten für das Luftschiff.

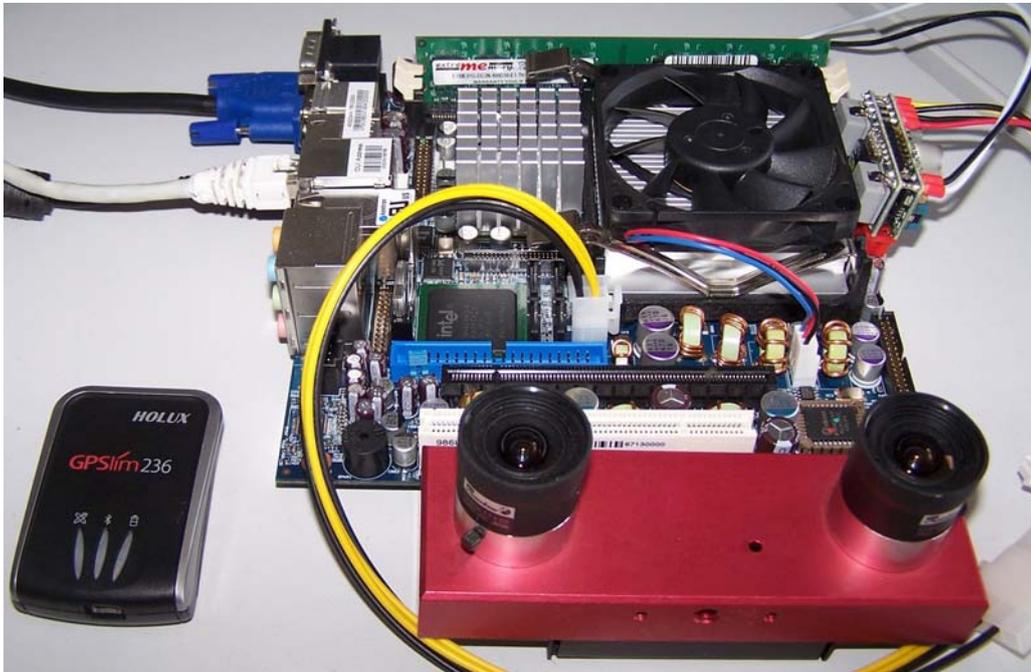


Abbildung 7: GPS Empfänger, Stereokamera und PC-System für das Luftschiff

Technische Daten:

- Länge: 9,50 m
- Breite: 2,20 m
- Antriebsleistung: 6 kg Schub
- Akkuleistung: 22,2 V, 12000 Ah
- Flugdauer: ca. 1 Stunde
- Maximale Geschwindigkeit: 25 km/h
- Maximale Windgeschwindigkeit: 15 km/h
- Gewicht des Luftschiffes inkl. Ausrüstung: ca. 20 kg

References:

- [1] Technikfolgenabschätzung – Leichter-als-Luft-Technologie – Innovations- und Anwendungspotentiale, März 2005
- [2] Krause, Protzel, Flugregler für ein autonomes Luftschiff, Autonome Mobile Systeme 2003, Karlsruhe, Germany
- [3] Gebrauchsmusterschrift, DE20107956 U1
- [4] Lima, Custódio, Ribeiro, The Rescue Projekt – Cooperative navigation for rescue robots, Aser 2003, Italy
- [5] Patentschrift, DE102007013147.1

Authors:

Dipl.-Ing. Thomas Krause
Prof. Dr. Peter Protzel
Technische Universität Chemnitz, Reichenhainer Straße 70
09126 Chemnitz, Germany
Phone: +49 371 531 33357
Fax: +49 371 531 833357
E-mail: thomas.krause@etit.tu-chemnitz.de
peter.protzel@etit.tu-chemnitz.de

T. Behrmann / M. Lemmel

Vehicle with pure electric hybrid energy storage system

Abstract

Hybrid systems to power vehicles are developed to compensate certain disadvantages of the main power supply, like low power dynamics in respect to the environmental factors or low power density relative to mass and costs. The advantages of hybrid systems are bought by higher level of system complexity and the need to handle these additional degrees of freedom. The quality of energy management systems is essential for the performance of the whole hybrid system.

The basis of this presentation is the German pilot project EFRB¹ - funded by the BMWi. The main focus of this project was the development, implementation and proving of a hybrid energy storage system consisting of different storage components. Well-known German automotive and energy storage manufacturer like DaimlerChrysler, Varta, Epcos etc. were partners within a consortium that was coordinated by BIBA/MAQ².

Within this research project a flexible automotive concept has been developed, which enables flexible adaptation of the operation mode appropriate to power requirements of the hybrid vehicle application. The key idea is a composition of different electric energy storages into one energy storage system. The objective is to achieve both: large mileage and high acceleration in the same vehicle. The results can be transferred on other usual hybrid combinations of fuel cells, combustion engines, batteries and UltraCaps. The system components developed in this project could lead to an optimised possibility for building hybrid vehicles.

The system complexity of a hybrid vehicle leads to new degrees of freedom which require an operating control combined with an optimised energy management. It has the objective to set the power flow of each storage device appropriate to drive situation and the optimised working point of the combined system. Basic and advanced methods of control theory, modelling, rapid control prototyping, optimised control and model

¹ Mobile electric storage system for vehicles with great mileage and high acceleration

² Bremen Institute of Industrial Technology and Applied Work Science at the University of Bremen
Division Metrology, Automation and Quality Science

predictive control are the tools used for the development process.

1. Basic idea and concept

The comparison of energy density and power density shows that energy storages can either be optimized for a maximum energy density with considerable reduced power density or alternatively for a maximum power density thus in a decreased energy density. Electric energy storages with different properties of energy and power density have been chosen purposely for application in vehicles (see Fig. 1). The target is to compensate these divergent design objectives by combining different energy storages.

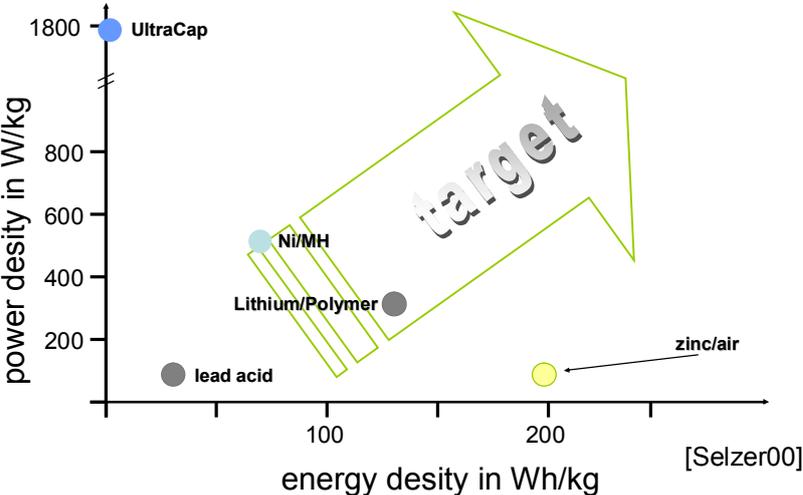


Fig 1 Ragone diagram of typical battery systems according to [Selz00]

Great mileage is covered by a long-term energy supply. Electric Fuel Company provides a zinc/air-battery with a high energy density. It has a relative low power density. In practice the fuel cells and small internal combination engines with generator have similar technical features, so that the results can be transferred on this issue.

The supercapacitor UltraCap from Epcos has a high power density and is able to supply and sink high currents. In recuperation mode it stores energy from breaking manoeuvres. This component is an ideal for covering short term power demand. This brings high acceleration abilities into electric vehicles while energy balance increases. A Ni/MH battery from Varta provides mid-term energy and power supply. It supports in case of sustaining acceleration processes the zinc/air basic supply and serves as recuperation storage for long down hill routes.

To handle the hybrid storage system energy management is essential. It identifies ac-

tive drive situation (drive, roll, break, stop, acceleration, stop and go etc.) and computes the state-of-charges (SOC) of the storage units. Depending on this an assessment the power flows will be controlled according to the expected drive state [Stan03]. The state-of-charges will instead be controlled for strategic planning depending on the expected drive state [Qu04]. E.g. an empty UltraCap is useful in case of high probability of breaking manoeuvres like in city traffic. It influences the overall efficiency.

The schematic image in Fig. 2 shows the drive train and the energy storage system with all vehicle components. On the left hand side the logical in- and output units for the operating system are shown. The Energy management for the control of the power converters and reaction to the drivers' decisions is situated at the centre.

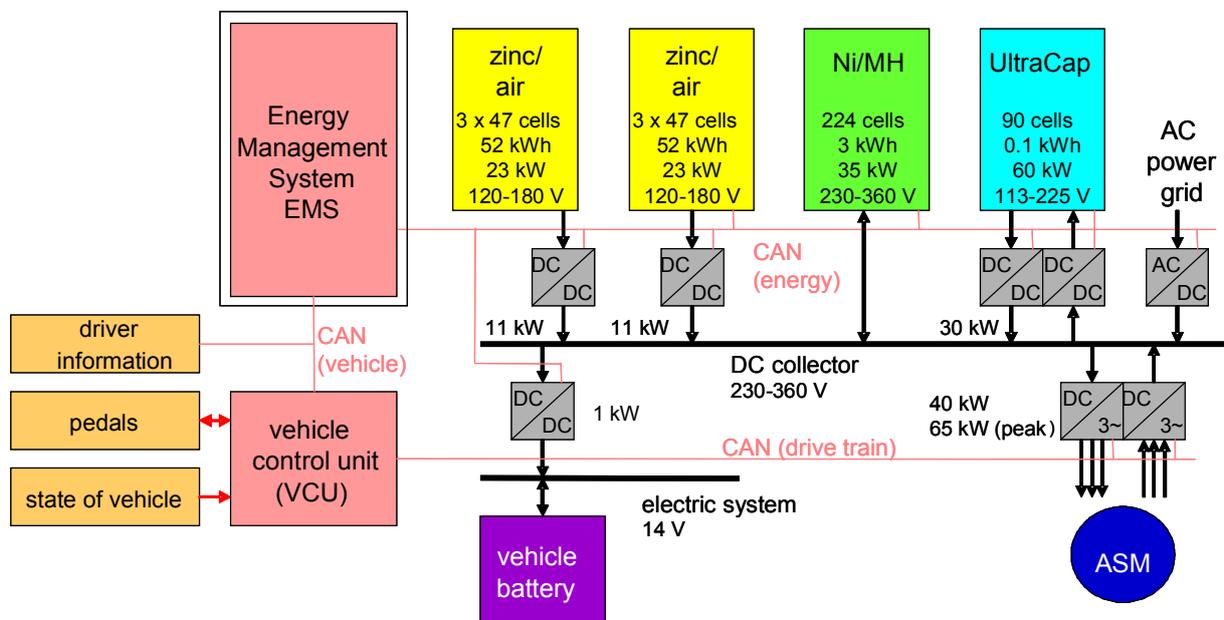


Fig 2 Schematic picture of the EFRB hybrid energy storage concept

The energy storages have different voltage operating ranges, so DC/DC converters are necessary for the connection over a main power bus. Direct connection of the Ni/MH battery keeps the bus voltage almost constant and has a robust behaviour in emergency situations. The zinc/air battery is connected by a unidirectional DC/DC converter with continuous power profile, while the UltraCap has to be operated bidirectional with high peak power characteristic.

The real-time communication of the active components of the storage system is implemented by a storage systems CAN bus.

2. The management of the traction energy

The two main components of the vehicle power supply are the energy management system (EMS) and the operating system (BFS). They ensure the operation of the three storage components and take care for optimal power flow. See Fig 3.

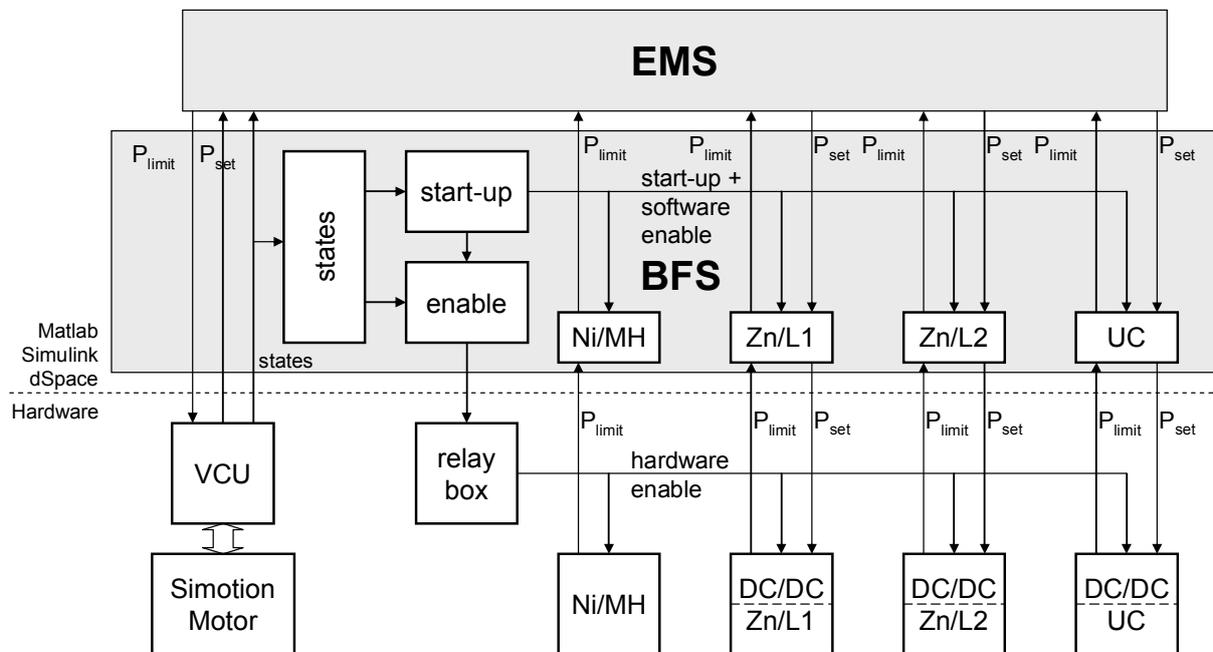


Fig 3 Scheme for the management of the supply of traction energy

The operating system (BFS) coordinates the boot process by enabling the single components in correct order. It supervises thresholds and global safety parameter and activates emergency modes if necessary. It interfaces the access of the available components to the energy management system (EMS).

The energy management contains several levels of decision and processing, which are illustrated schematically in Fig. 4 from left to right. Input data feeds are marked with arrows from the bottom.

The strategic planning components on the left hand side of the block diagram in Fig. 4 lead to the controller of the state of charge (SOC) of the energy storages. This controls the fraction of power supplied by each component (P-Mix-Planning). The planning components are implemented in rule based fuzzy controllers [Sorg04].

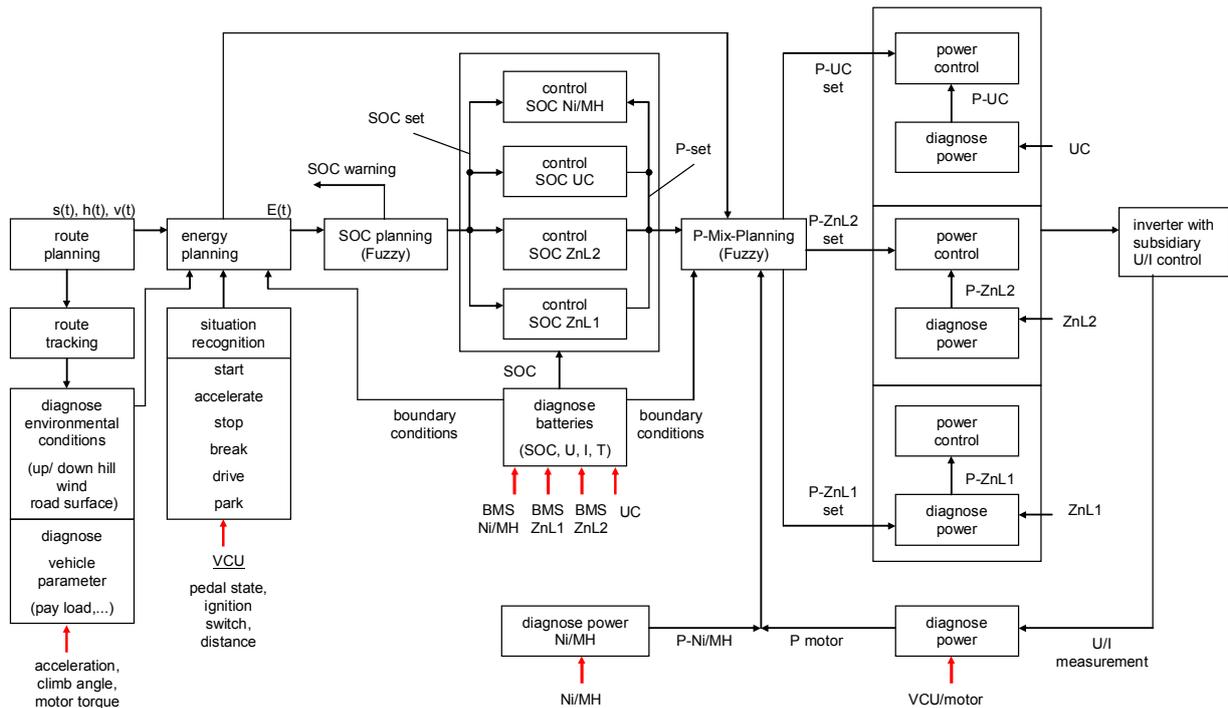


Fig. 4 Schematic overview of the energy management system (EMS) [Sorg04]

3. Conclusion

It has been shown, that the option to use a low power energy supply like zinc/air batteries or fuel cells can be supported by additional mid- and short-term components like Ni/MH batteries and UltraCaps. This situation dependant energy management leads to an up-to-date drive appeal with high efficiency. The results can be transferred to other hybrid constellations.

References:

- [Selz00] H. Selzer, Dr. Ch. Ament, M. Sorg: Electric storage system for great mileage and high acceleration; ISATA; Dublin; 2000
- [Sorg04] M. Sorg: Batterie-Tray mit modularem Batteriemanagementsystem; 26. Kolloquium der Automatisierungstechnik; Salzhhausen, 05.11.-06.11.2004
- [Qu04] Qu. Yu, M. Sorg, R. Stanislawski, Ch. Ament, H. Selzer: Fuzzy Prediction Control Strategy of EMS with Energy Hybridization of High Energy and High Power; Asian Electric Vehicle Conference(AEVC-3), Jeju Island, South Korea; 24.-29.03.2004
- [Stan03] R. Stanislawski, G. Goch, Ch. Ament, C. Bredemeier: Energy Management System for Combined Electric Storage System for High Energy and High Power; EVS, CA/USA; 2003

Authors:

Dipl.-Ing. Thomas Behrmann
 BIBA/MAQ
 Hochschulring 20
 28359 Bremen
 Phone: +49 421 218 5627
 Fax: +49 421 218 5625
 E-mail: RT@biba.uni-bremen.de

Dr. Marc C. Lemmel
 BIBA/MAQ
 Hochschulring 20
 28359 Bremen
 Phone: +49 421 218 5627
 Fax: +49 421 218 5625
 E-mail: lem@biba.uni-bremen.de

C. Schröter, M. Höchemer, H.-M. Gross

A Particle Filter for the Dynamic Window Approach to Mobile Robot Control

Abstract

In this paper we present an anticipative local navigation algorithm for an autonomous mobile robot. The purpose of local navigation is to move the robot according to a specified goal, like a planned path to a target, and avoiding collisions with obstacles during operation. The robot is perceiving its immediate surroundings by laser and sonar range scanners and by a stereo camera. All the sensor information is represented in a local map. In order to choose the best action, a number of possible trajectories are then evaluated. The trajectories are modelled as clothoid curves, a parametric curve which is well suited for moving vehicles. A fitness function that takes into account the likelihood of collisions, the compliance with the navigation goal and the speed that can be achieved selects the best trajectory, which is then translated into motion commands for the drive system.

1 Introduction

Autonomous navigation is one of the key features of a mobile robot. Typically, an autonomous system is possessing a map of the environment, which is either given by the designer or built from sensory data during operation. Furthermore, the robot has some means of determining and tracking its own position in the environment. In order to reach a certain position within that map a 2 step process is conducted: First, the robot has to plan a series of actions to move from its current position A to the desired target position B. Then, it has to translate the actions into concrete motion commands to be executed by the drive system. In most cases it is not possible to plan the entire sequence of motion steps that would move the robot from A to B because of the occurrence of positional errors during execution as well as discrepancies between the map and the actual environment, caused e.g. by dynamic obstacles such as people appearing in the area or unpredictable motion of obstacles. Therefore, typically the planner just generates a shortest path from A to B which is a sequence of positions, while a controller/local navigator is trying to sequentially generate motion commands that make the robot follow that path, with respect to the current position and the updated state of the environment as perceived by the robot's external sensors. In this paper we will consider a new implementation of the local navigator only. The global map, the planning of a shortest path from the robot's position to its target as well as the self-localization of the robot within the environment are taken for granted and not discussed here. The local navigation approach we propose is an anticipatory behaviour because it not only tries to avoid obstacle collisions but does so by evaluating the results of a number of possible actions using a limited foresight into the future. To this purpose,

actions are coded as parametrical trajectory curves and managed by a scheme similar to the way particle filters are used for state estimation problems.

2 Local Navigation - Related Work

Existing obstacle avoidance approaches can be roughly divided into two classes: reactive and anticipatory. A very simple reactive method is the potential field approach, which works by assigning virtual repelling forces to obstacles close to the robot, and attracting forces to navigation goals. However, this method often fails to pass narrow passages like doors. An improvement is the Vector Field Histogram proposed by Borenstein et.al. and its various enhancements [1], [2]. In this group of algorithms, the robot explicitly distinguishes between free and blocked directions and chooses a free direction that is closest to the navigation goal. In contrast to these reactive methods, anticipative approaches explicitly evaluate the consequences of certain actions and try to choose the one yielding the highest return or lowest cost. In the Dynamic Window Approach [3], a number of circular trajectories are tested for the distance they keep to the obstacles around the robot. An enhancement to this is the Global Dynamic Window Approach, which additionally incorporates global navigation goals in the cost function. Our approach here basically is a modification of the Global Dynamic Window Approach, where we use clothoid instead of circular trajectories. Furthermore, instead of re-generating and evaluating all the possible trajectories from scratch in each time step, in analogy to a particle filter, the hypotheses are sampled from the best trajectory of the last time step, imposing an implicit smoothness constraint.

3 Clothoids

In order to choose the best action in the current situation, given the current state of the robot and the local environment as well as the overall target, we need to generate and evaluate a number of possible local motion trajectories. For the representation of the trajectories a form of parametrized curves called clothoids are used. The definition of a clothoid is a curve with linearly changing bending

$$c(l) = c_0 + c_1 * l \tag{1}$$

where the bending c is the inverse of the curve radius r . Clothoids are used in road construction because the linearly changing curvature in turn means a linear change of lateral force, avoiding a jump in the force imposed on vehicles following the road. Obviously, for the same reason they are a good model for robot motion.

Independently of the actual drive system, the robot's motion is usually seen as a superposition of translational and rotational velocity, denoted as v and w respectively. For constant

translation velocity v and rotation velocity w , the robot will move on a circle with radius

$$r = \frac{v}{w} \quad (2)$$

Therefore, if the robot is currently moving at a velocity (v, w) , the initial curvature c_0 is fixed

$$c_0 = \frac{w}{v} \quad (3)$$

Furthermore, there are limitations on the change rate of the curvature c_1 , which reflect the physical properties of the robot such as mass and motor power. The actual sequence of positions described by the clothoid trajectories is then given by

$$x(l) = x_0 + \int_0^l \cos(\phi(l))dl \quad (4)$$

$$y(l) = y_0 + \int_0^l \sin(\phi(l))dl \quad (5)$$

$$\phi(l) = \phi_0 + \int_0^l c(l)dl \quad (6)$$

where (x_0, y_0, ϕ_0) is the current pose of the robot, containing position and orientation.

4 Trajectory Evaluation

When the robots navigator module receives a new target, it plans a path to the target position using the Dijkstra algorithm on the global map. During the path calculation, a potential field is generated which holds for each position of the global map the distance to the target, assuming a shortest path motion. This target distance will be used, together with other costs, in evaluating possible trajectories.

While the robot is moving, the external sensor measurements are continuously integrated into a local 2D map. This local map holds information about traversable and blocked space in a local vicinity of the robot. Due to the reliable perception distance of the sensors, the local map has a radius of about 3 meters around the current robot position.

The navigator is using the local map to generate motion commands for the drive system in intervals of 100 ms (at a maximum speed of 1m/s, this corresponds to a maximum driven distance of 0.1m). In order to determine the best local trajectory, a number of candidate clothoid trajectories are generated. Each clothoid is described by parameters c_0 and c_1 . As explained in section 3, c_0 is equal for all possible clothoids, determined by the current translational and rotational robot speed, which is reported by the drive system. However, c_1 is sampled from a random distribution. When no best trajectory was selected in the

previous loop run, e.g. at the very beginning of autonomous motion, the distribution is just a Gaussian with mean 0 and a fixed variance. When a previous best trajectory is already known, the new candidates are sampled with c_1^{old} as mean value. Together with $cost_{change}$ (see below), a behaviour of permanent alternating is suppressed in situations where 2 possible trajectories are approximately equally good (e.g. an obstacle in the center of a hallway that could be passed to the left or right), imposing an implicit smoothness constraint.

Each trajectory is assigned a cost that is a weighted sum of a number of costs, each one representing a certain objective:

- $cost_{closest_obstacle}$: Along the trajectory, normal vectors are calculated in regular intervals. Along each normal line, the closest obstacle (blocked cell) is searched. If an obstacle is found, the cost is $(1.0 - d'_{traj}) * (1.0 - d'_{norm})$, where d_{traj} is the distance along the trajectory, d_{norm} is the distance from the trajectory along the normal line. d' denominates normalization by dividing by the maximum trajectory/normal line length respectively. The maximum cost for a single found obstacle determines $cost_{closest_obstacle}$
- $cost_{sum_obstacles}$: The summed obstacle cost sums the values over all normal lines. In contrast to $cost_{closest_obstacle}$ it does not only consider the most extreme obstacle approach, but the overall distance keeping to obstacles along the entire trajectory.
- $cost_{bending}$: In order to enforce straight motion of the robot when possible, a high bending of the trajectory is punished with high cost. $cost_{bending}$ is directly proportional to the trajectory parameter c_1
- $cost_{target}$: While the robot must avoid collisions, it is still expected to follow a path that will take it to the target position. This is reflected by $cost_{target}$. The cost is proportional to the decrease of the target distance (when following the optimal path) between the current position and the trajectory end point..
- $cost_{change}$: This cost is proportional to the difference between the current trajectories parameter vector (c_0, c_1) and the previously selected trajectories parameter vector $(c_0, c_1)^{old}$ (see above).

Bending cost as well as change cost also depend on the current robot speed: at low speeds, a strong bending and a faster change of bending are less punishable than when driving at maximum speed. The overall cost is then given by

$$cost = \alpha * cost_{closest_obstacle} + \beta * cost_{sum_obstacles} + \gamma * cost_{bending} + \delta * cost_{target} + \epsilon * cost_{change}$$

with $\alpha = 10.0$, $\beta = 50.0$, $\gamma = 0.5$, $\delta = 8.0$ and $\epsilon = 0.2$. Obviously, when obstacles are near, they determine the cost mainly. Only in free space situations the bending and change cost have any significant influence.

Finally, the trajectory with the lowest overall cost determines the motion command. By choosing a certain c_1 , the desired change of the curve bending and, with a fixed update cycle, a bending c to be reached till the next step is given. From eq. (3) follows that only the relation between v and w is determined by the bending c . Therefore, in order to find specific values for v and w , additional rules are needed. One possibility would be to always keep a constant translational velocity $v = v_0$. However, for safety reasons we prefer to slow down if we get closer to obstacles, therefore v depends on the trajectory obstacle cost too.

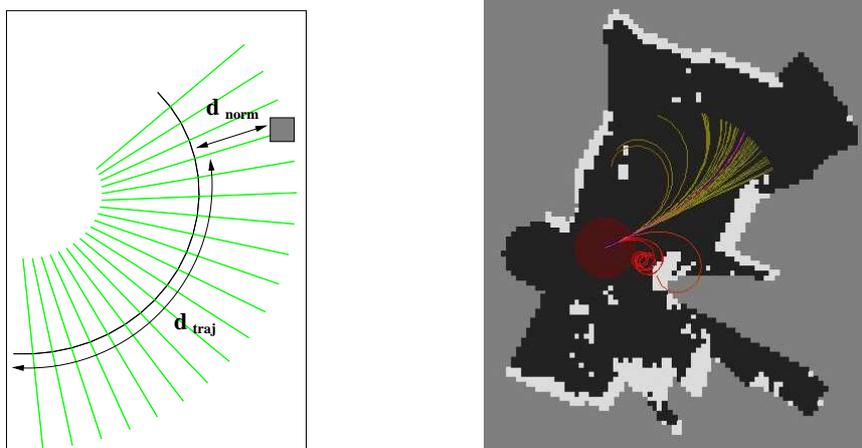


Figure 1: Left: The obstacle cost for a trajectory depends on the distance of the obstacle to the robot (along the trajectory) and the distance to the trajectory itself.

Right: The figure shows the robot and the local environment, which is perceived using a laser scanner and a stereo camera. White areas are obstacles, while black areas are free space. Grey indicates areas which have not been seen by the robot. A number of trajectories are shown, where the color shows the cost associated with each trajectory. Green colors mean low costs, while red colors show high costs. The preferred trajectory, which determines the motion command, is marked magenta.

5 Results

To compare the new navigation algorithm presented here to an implementation of the Vector Field Histogram (this is actually an enhanced version of VFH that has been our standard local navigation approach for years), we show results of a test run where the robot's task was to go down a hallway and turn into an adjacent room, crossing a very narrow door (only a few cm space to either side of the robot). In both cases the maximum robot speed was limited to 0.5 m/s. The plots show that the robot moves significantly faster and smoother using the new algorithm for local navigation (Fig. 2). While with VFH the robot took 36 seconds for the path (average velocity 0.25 m/s), it arrived 40the new algorithm.

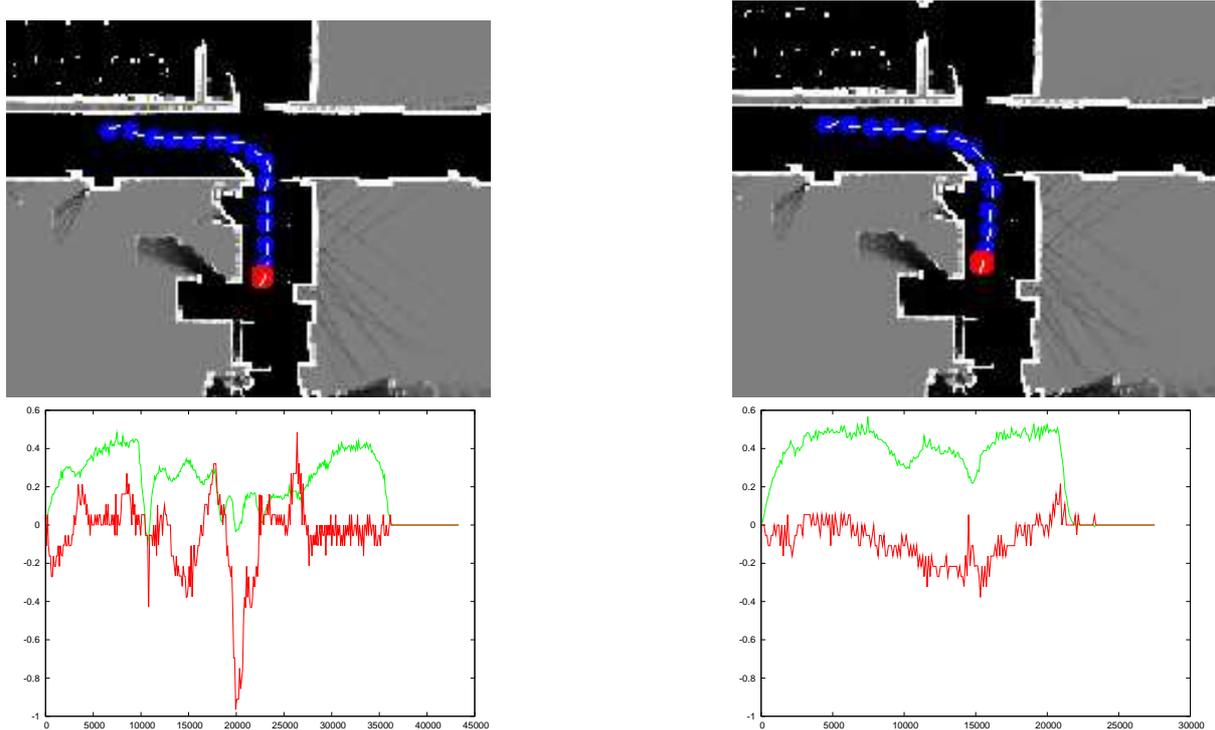


Figure 2: Test run results - left: results using an implementation of the Vector Field Histogram, right: using the trajectory particle filter. The upper row shows the robot path for both algorithms respectively. The path is slightly smoother using the new algorithm. Results are more obvious from the velocity plots (2nd row). Here, translational (green) and rotational (red) velocity are drawn along the path.

References

- [1] J. Borenstein and Y. Koran, "The Vector Field Histogram - fast obstacle avoidance for mobile robots", *IEEE Journal of Robotics and Automation*, Vol. 7, Num. 3, pp. 278-288, 1991.
- [2] I. Ulrich, J. Borenstein, "VFH+: Reliable Obstacle Avoidance for Fast Mobile Robots", *Proc. 1998 IEEE Intl. Conf. on Robotics and Automation (ICRA98)*, pp. 1572 - 1577, 1998
- [3] D. Fox and W. Burgard and S. Thrun, "The Dynamic Window Approach to Collision Avoidance", Technical Report, University of Bonn, IAI-TR-95-13, 1995.
- [4] O. Brock and O. Khatib, "High-speed navigation using the global dynamic window approach", *Proc. 1999 IEEE Intl. Conf. on Robotics and Automation (ICRA99)*, 1999

Author Information:

Christof Schröter, Matthias Höchemer, Horst-Michael Gross
 Department of Neuroinformatics and Cognitive Robotics, Faculty of Computer Science and Automation,
 Ilmenau Technical University, PO Box 10 05 65, 98684 Ilmenau
 Tel: +49 3677 69 1306
 Fax: +49 3677 69 1665
 E-mail: christof.schroeter@tu-ilmenau.de

M. Schenderlein / K. Debes / A. Koenig / H.-M. Gross

Appearance-based Visual Localisation in Outdoor Environments with an Omnidirectional Camera

1 Introduction

After achieving many goals in indoor robot navigation the focus of research is shifting to outdoor applications. This poses a new challenge since outdoor environments do not only lack the presence of mainly straight and close in range walls as can be found in indoor office environments but also come up with more dynamic lighting conditions. Thus localising the robot in its environment becomes a greater challenge. On the other hand robots can have the advantage of an additional sensory input - the Global Positioning System (GPS) receiver - which measures a global position of the robot and is not applicable indoors. Though it has a bounded error, the GPS position estimate can not be exclusively used for robot localisation since it is prone to jumps in precision and might include an offset over longer periods of time because of multipath reflections from buildings or trees. Therefore we use a visual localisation system which was made robust to changes in illumination and sensor occlusion. Our robot platform is a three-wheeled vehicle equipped with an omnidirectional camera as well as a low-cost GPS receiver, a laser range finder, sonar sensors and an inertial odometry measurement unit (see Figure 2(a)).

2 Related Work

This section provides a brief review of the literature in visual outdoor localisation. The proposed approaches can be divided into ones dealing with relative localization and ones dealing with absolute localisation. Relative localisation means deriving the current pose of the robot from the previous one by measuring or estimating the relative pose change. Optical flow techniques on monocular images have the drawback of a flat world assumption in order to get reliable movement estimations [5]. This is overcome by 3D pose estimation from tracking features (e.g. corners) in an image stream [14]. Since such a calculation needs a lot of computational power and moreover requires the camera platform to be in motion, the current focus of research in those so-called visual odometry techniques is the tracking of 3D points derived from stereo camera images. Candidate points can be found by corners [14] or SIFT features [2]. The other group of approaches deals with absolute localisation which means deriving the robots current pose from a topological visual map. Such a map consists of reference locations at which low dimensional features from belonging reference images are stored. Colour histograms are used as feature vectors describing the

colour distribution in the image and are combined with nearest neighbour searches to derive the possible location of the query image [3]. The image pixel intensity gradients pose a feature which is robust to illumination changes [4]. Local image structure is also examined by integral invariant features [16] or SIFT features [17], [16]. Features that go towards object recognition are region based object classification [12] and edges of buildings [7]. Simultaneous Localisation and Mapping (SLAM) in the visual domain in large scale outdoor environments is yet tackled by only a few approaches. Pose estimation by means of a stereo camera and loop closing detection with the help of gradient features is done in [11] to solve the SLAM problem. Separate loop closing detection using SIFT features is done in [13]. Another SLAM approach is the application of a relaxation algorithm on the topological map based on image feature and odometry relations [1]. The visual odometry techniques for relative robot localisation have the advantage to be independent from a map and can therefore localise the robot immediately. Their drawback is the fact that they have no global reference and thus accumulate the estimation error. By contrast using a visual map provides this global framework. Such a map has to be build beforehand which is not always possible though. SLAM approaches are meant to solve this drawback and are therefore a promising future technique for visual outdoor localisation.

3 Approach

We propose the use of appearance-based features from panoramic images in combination with a probabilistic framework as a solution to the localisation problem. The full localisation process consists of two phases, a mapping phase and a localisation phase, which are lined out below.

First, in the mapping phase, a topological map of visual representations of the environment is built by manually acquiring the image data of the environment in conjunction with the associated pose $\underline{x} = (x, y, \varphi)$ of the robot. This kind of map stores the information in a few distinct reference places, which is especially useful in large outdoor environments. The uncertainty in position of the robot caused by erroneous odometry information is integrated out by means of a Bayesian filter [6] combining the odometry information and the GPS measurements. For further details see [15]. The mapping process presented in this work still utilises the GPS sensor despite its mentioned shortcomings. This is due to ongoing research in this field. The accuracy of acquired maps therefore has to be judged by a human inspector. Inappropriate maps possibly have to be rejected and acquired again.

For fast and noise-reduced computation the image data is stored as low dimensional representations. At this point the use of the omnidirectional camera has the advantage that only one image per reference position is needed to gather visual information from all view directions. Several test runs showed that structure-based image features are the best choice

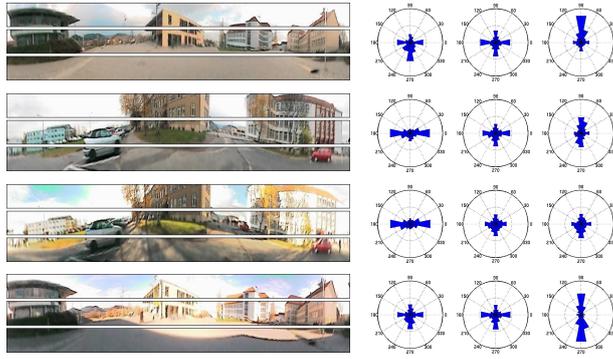


Figure 1: Two omnidirectional scenes of the campus divided into three horizontal segments and the corresponding 16-bin gradient orientation histograms left to right for the top to bottom segments.

considering invariance to different illumination settings and general capability to distinguish adjacent as well as relocate mapped places. Hence we decided to use the gradient orientation histogram as image feature for our purpose. To calculate the feature vector first the

omnidirectional image is converted into a panoramic image. Of that panoramic image the approximate first order gradients G_x, G_y are calculated from the pixel intensities by means of Sobel edge filtering. Each pixel contributes to the histogram by its gradient orientation. We found a number of 16 orientation classes to give the best results. To overcome the need of a threshold for very low gradients they are weighted by their magnitude $M = \sqrt{G_x^2 + G_y^2}$ as in [10]. For the feature vector \underline{h} to be more distinctive several histograms are computed from horizontal image segments (see Figure 1) and concatenated. We found a number of six segments to give the best results. Two feature vectors m and n are compared by calculating the statistical χ^2 distances

$$d_{\chi^2}^{(i)}(\underline{m}^{(i)}, \underline{n}^{(i)}) = \sum_{k=1}^b \frac{(m_k^{(i)} - n_k^{(i)})^2}{(m_k^{(i)} + n_k^{(i)})} \quad (1)$$

for all segment histograms indexed by i with b histogram bins. In order to be robust to partial image occlusions by e.g. persons, only the three smallest distances $d_{\chi^2}^{(i)}$ are summed up to the overall distance d_{χ^2} .

In the localisation phase the robot needs to estimate its pose \underline{x} by comparing the current visual sensory input with the acquired map. The visual localisation algorithm applies a Bayesian filter variant, namely a standard Monte-Carlo-Localisation (MCL) method [6], for estimating the robots pose via the current sensor input of the camera. This is accomplished by approximating the probability distribution of the pose by a number of discrete samples. Each of those samples is moved by a motion model derived from the robot vehicle dynamics given the last movement of the robot. To adapt the distribution according to the sensor input, for every sample the closest reference feature vector \underline{h}_r at the samples position in the map is compared to the feature vector \underline{h}_c calculated of the current sensor input. In the resampling step samples whose distance $d_{\chi^2}(\underline{h}_r, \underline{h}_c)$ is small are more likely to be kept and

duplicated as opposed to the remaining ones which are likely to be erased. Eventually the pose estimation is calculated as the mean of the pose distribution of all samples.

4 Results

To examine the performance of our localisation system we measured the precision at measuring points arranged as markers at known positions on the ground. Their relative positions were determined by tape measure and best fitted to the GPS measurements. Two places on the campus of the TU Ilmenau with an area of 60 m×18 m and 90 m×35 m (see Fig. 2(b)) were investigated. First a topological visual map was build at each of those two places on a sunny as well as a cloudy day. We found a minimum distance of the maps reference points of 2 m to be sufficient. Also in each of the four cases a test run of about 1.5 km length was performed which included several loops. In those test runs we did not stick to the paths used in the mapping stage, but arbitrarily passed the measuring points. At those points we were able to calculate the localisation error. Since our robot is not equipped with a compass we used the GPS as best available ground truth for the robots heading direction. The robots pose was estimated for all pairs of test runs and maps at each place. The results are lined out in table 1.

Test run	Map	Position				Heading			
		Place A		Place B		Place A		Place B	
		μ	σ	μ	σ	μ	σ	μ	σ
sunny	sunny	2.77	1.62	3.14	1.73	18.07	14.29	10.30	7.55
cloudy	cloudy	3.48	1.77	3.62	2.45	13.79	10.86	12.40	9.57
sunny	cloudy	3.86	2.77	3.95	2.29	20.60	15.28	12.11	9.11
cloudy	sunny	3.65	3.24	4.43	2.56	14.92	13.04	12.35	12.47

Table 1: Mean difference and standard deviation of the position estimation in meters and the heading direction estimation in degrees.



Figure 2: (a) The outdoor robot MILVA. (b) Places A and B on the campus of the TU Ilmenau where the performance examinations were carried out (magenta regions).

The results show that there is no big difference in estimation precision between the test runs performed on maps with equal illumination conditions and those with different conditions.

This means the localisation system is robust to changing environment illumination caused by different day times and weather situations. The gradient orientation histogram is robust up to 50% of image occlusion. The GPS/odometry integration achieved a mean position estimation error of 2.54 m and a mean heading direction estimation error of 10.89 degrees. Figure 3 shows estimated trajectories of the visual MCL and the GPS/odometry integration.

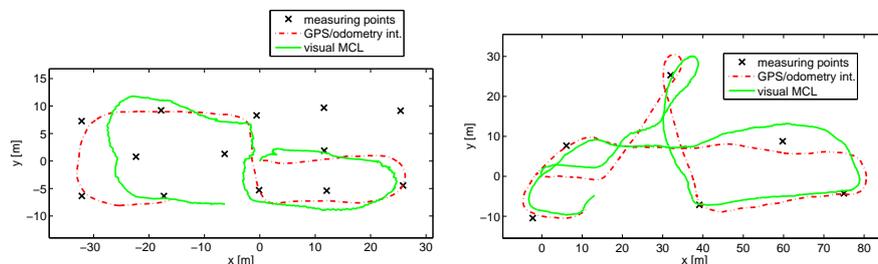


Figure 3: Parts of the estimated trajectories of the visual MCL and the GPS/odometry integration.

5 Conclusion

We presented a visual localisation framework able to localise our robot on the campus of the TU Ilmenau. It is robust to changes in environment illumination and image occlusions. At the moment its localisation precision is lower than the one of the GPS system used for map building. Considering the precision of the visual system is only as good as the map it relies on, the results are promising though. Moreover has the visual localisation the advantage that the map serves as a stable observation reference, with extreme image distortions causing only a temporal uncertainty of all particles which would shift the belief to the odometry. By contrast jumps in the GPS data which are possible for a longer period of time cause the pose estimation to drift to the erroneous measurements. This especially bears problems in the context of robot navigation. As mentioned above for map building this way of pose estimation is still useful since maps with extreme pose outliers can be identified manually after the mapping process and reacquired. Future investigations will have to tackle a gain in heading direction estimation precision which is likely to improve the overall performance as well.

References

- [1] H. Andreasson, T. Duckett and A. Lilienthal, “Mini-SLAM: Minimalistic Visual SLAM in Large-Scale Environments Based on a New Interpretation of Image Similarity”, *Proc. of the 2007 IEEE International Conference on Robotics and Automation (ICRA 2007)*, pp. 4096-4101, 2007
- [2] T. D. Barfoot, “Online Visual Motion Estimation using FastSLAM with SIFT Features”, *Proc. of the International Conference on Robotics and Intelligent Systems (IROIS 2005)*, pp. 579-585, 2005

- [3] P. Blaer and P. K. Allen, "Topological mobile robot localization using fast vision techniques", *Proc. of the IEEE International Conference on Robotics and Automation (ICRA 2002)*, Vol. 1, pp. 1031-1036, 2002
- [4] D. C. Bradley, R. Patel, N. Vandapel and S. M. Thayer, "Real-Time Image-Based Topological Localization in Large Outdoor Environments", *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005)*, pp. 3062-3069, 2005
- [5] P. Corke, D. Strelow and S. Singh, "Omnidirectional visual odometry for a planetary rover", *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004)*, Vol. 4, pp. 4007-4012, 2004
- [6] F. Dellaert, D. Fox, W. Burgard and S. Thrun, "Monte Carlo localization for mobile robots", *Proc. of the IEEE International Conference on Robotics and Automation (ICRA 1999)*, Vol. 2, pp. 1322-1328, 1999
- [7] A. Georgiev and P. K. Allen, "Localization Methods for a Mobile Robot in Urban Environments", *IEEE Transactions on Robotics*, Vol. 20(5), pp. 851-864, 2004
- [8] H.-M. Gross, A. Koenig, Ch. Schroeter and H.-J. Boehme, "Omnivision-based Probabilistic Self-localization for a Mobile Shopping Assistant Continued", *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)*, pp. 1505-1511, 2003
- [9] H.-M. Gross and A. Koenig, "Robust Omniview-based Probabilistic Self-Localization for Mobile Robots in Large Maze-like Environments", *Proc. of the 17th Int. Conference on Pattern Recognition (ICPR 2004)*, Vol. 3, pp. 266-269, 2004
- [10] J. Kosecka and F. Li, "Vision based topological markov localization", *Proc. of the IEEE International Conference on Robotics and Automation (ICRA 2004)*, Vol. 2, pp. 1481-1486, 2004
- [11] T. Lemaire and S. Lacroix, "SLAM with Panoramic Vision", *Journal of Field Robotics (Special Issue on SLAM in the Field)*, Vol. 24(1-2), pp. 91-111, 2007
- [12] H. Morita, M. Hild, J. Miura and Y. Shirai, "Panoramic View-Based Navigation in Outdoor Environments Based on Support Vector Learning", *Proc. of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2006)*, pp. 2302-2307, 2006
- [13] P. Newman, D. Cole and K. Ho, "Outdoor SLAM using visual appearance and laser ranging", *Proc. of the IEEE International Conference on Robotics and Automation (ICRA 2006)*, pp. 1180-1187, 2007
- [14] D. Nister, O. Naroditsky and J. Bergen, "Visual Odometry for Ground Vehicle Applications", *Journal of Field Robotics*, Vol. 23(1), pp. 3-20, 2006
- [15] M. Schenderlein, "Outdoor-Roboternavigation und Lokalisation in interaktiv erstellten topologischen Graphen mit Omnivision-Attributierung unter Einsatz probabilistischer Zustandsschaetzer", *Diploma Thesis*, 2007
- [16] C. Weiss, A. Masselli, H. Tamimi and A. Zell, "Fast Outdoor Robot Localization Using Integral Invariants", *Proc. of the 5th International Conference on Computer Vision Systems (ICVS 2007)*, 2007
- [17] W. Zhang And J. Kosecka, "Image Based Localization in Urban Environments", *Third International Symposium on 3D Data Processing, Visualization and Transmission*, pp. 33-40, 2006

Author Information:

M. Schenderlein

Dr. K. Debes, Dipl.-Inf. A. Koenig, Prof. Dr. H.-M. Gross

Department of Neuroinformatics and Cognitive Robotics

Technische Universitaet Ilmenau

Fakultaet fuer Informatik und Automatisierung

G.-Kirchhoff-Str. 2 (Bionikgebaeude)

98693 Ilmenau

Tel: +49-(0)3677-69-2770 (K. Debes), Fax: +49-(0)3677-69-1665

E-mail: klaus.debes@tu-ilmenau.de

G. Al Zeer / A. Nabout / B. Tibken

Hindernisvermeidung für Mobile Roboter mittels Ausweichecken

1. EINLEITUNG

Der Einsatz von intelligenten fahrerlosen Transportsystemen gewinnt u. a. in modernen Industrieanlagen immer mehr an Bedeutung. Solche Systeme können beispielsweise in modernen Lager- und Produktionshallen für eine automatisierte Lagerhaltung eingesetzt werden, um den Materialfluss zu optimieren und so die Herstellungskosten zu reduzieren.

Bahnplanung ist ein zentrales Problem im Bereich der mobilen Robotik und wesentlicher Bestandteil solcher Transportsysteme. In dieser Anwendung ist zwischen statischer und dynamischer Bahnplanung zu unterscheiden [1]. Die statische Bahnplanung berechnet für ein Fahrzeug an Hand bekannter globaler Informationen über die Start- und Zielpositionen sowie der vorhandenen Hindernisse im gesamten Aktionsraum eine oder mehrere mögliche Gesamtrouten, die die Startposition mit der Zielposition unter Vermeidung der Hindernisse verbindet. Bei einer dynamischen Bahnplanung wird die Hindernisvermeidung dynamisch durchgeführt, indem zuerst das Ziel direkt angesteuert wird, bis das Fahrzeug auf ein Hindernis trifft. Die dynamische Bahnplanung besteht nun darin, eine Bahn für ein Ausweichmanöver unter Berücksichtigung der Zielposition zu berechnen [2].

Im Bereich der mobilen Robotik werden zur Trajektorien-Planung bei Vorhandensein von statistischen und / oder dynamischen Hindernissen unterschiedliche Lösungsansätze verfolgt (Tu and Yang, 2003; Bennewitz et al., 2002; Melchior et al., 2003). Eine der populärsten Bahnplanungsmethoden sind die artificial potential fields (Tsuji et al., 2002). Diese Methode liefert in einem statischen Umfeld lediglich eine Lösung, die jedoch nicht die kürzeste Trajektorie liefert [1].

In der künstlichen Intelligenz (KI) findet man zur Lösung des hier genannten Problems weitere Lösungsansätze [3]. Seit 1968 findet die A*-Methode [4, 5] für die statische Bahnplanung als Graphensuchalgorithmus Anwendung [6]. Der Algorithmus sucht die beste Bahn von einer gegebenen Start- zu einer gewünschten Zielposition. Hierbei wird der Arbeitsraum durch einen Graphen modelliert. Der Algorithmus durchläuft den

Graphen und untersucht die Nachbarknoten der bereits erreichten Positionen.

In dieser Veröffentlichung wird eine neue Methode für die Berechnung von Bahnen für mobile Roboter, die in einem Aktionsraum von einer Start- zu einer Zielposition unter Vermeidung von statischen Hindernissen geführt werden sollen, vorgestellt. Diese Bahnplanung erfolgt offline für die gesamte Route und berücksichtigt dynamische Hindernisse nicht, die während der Fahrt entstehen können. Eine solche Bahnplanung wird beispielsweise bei der vollautomatischen Lagerhaltung benötigt, um fahrerlose Fahrzeuge flexibel im gesamten Lagerbereich automatisch zu manövrieren. Die Methode stützt sich dabei auf Messdaten eines Kamerasystems und berechnet mittels Ausweichecken mehrere mögliche Bahnen in Form von approximativen Strecken, die sich durch ihre Gesamtlänge sowie Anzahl und Stärke der Lenkvorgänge unterscheiden. Die Ergebnisse dieser Methode werden an Hand einer MATLAB-Implementierung diskutiert.

2. BAHNPLANUNG MITTELS APPROXIMATIVER STRECKEN

Die entwickelte Methode stützt sich auf die Verwendung einer oder mehrerer Kameras, die den Aktionsraum erfassen. Abb. 1 zeigt den Aktionsraum (Laborraum) mit drei Fahrzeugen. Durch Auswertung der Kameraaufnahmen werden die Positionen und Ausrichtungen der Fahrzeuge sowie die Positionen und Ausdehnung der Hindernisse automatisch erkannt.

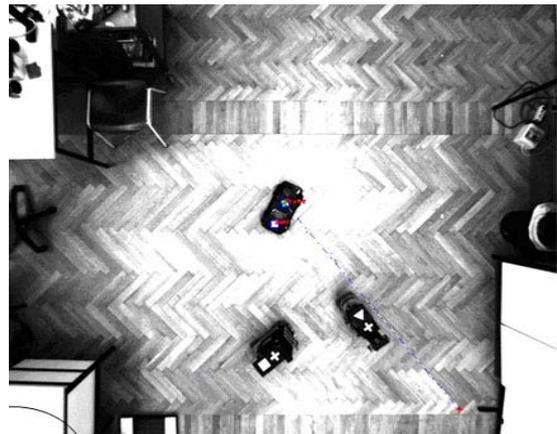


Abb. 1: Aufnahme des Aktionsraumes

Anschließend werden Ausweichecken berechnet, die als Kurspunkte dienen. Zwischen diesen befinden sich befahrbare kollisionsfreie Wege in Form gerader Strecken. Diese Methode basiert auf der Beschreibung des Arbeitsraumes in Form von Rasterpunkten (Pixeln), wie dies üblicherweise bei Kamera-Aufnahmen der Fall ist. Die Idee des hier vorgestellten Verfahrens stützt sich zur Berechnung der approximativen Strecken auf die Konturapproximation aus dem Bereich der Bildanalyse.

Zur Ermittlung von Ausweichecken wird jedes Hindernis von einem Rechteck eingeschlossen (s. Abb. 2). Die Eckpunkte der Rechtecke aller Hindernisse im Aktionsraum dienen hierbei als Ausweichecken für die Bahnplanung. Dadurch kann der Verlust an befahrbarem Aktionsraum gering gehalten werden.

Die Umgebung um jedes Hindernis herum wird in 8 Oktanten eingeteilt. Die

auszuwählenden Ausweichecken hängen von den Positionen den Start- und Zielpositionen entsprechend Tabelle 1 ab. Um einem Hindernis auszuweichen werden maximal zwei Eckpunkte verwendet.

Liegt beispielsweise die Startposition im dritten Quadranten und die Zielposition im siebten Quadranten (s. grau unterlegter Bereich der Tabelle an Position 3, 7), wird der Eckpunkt „c“ als Ausweichecke gewählt. Liegt hingegen die Startposition im zweiten Quadranten und die Zielposition im siebten Quadranten, dann werden die Eckpunkte „a“ und „d“ oder „b“ und „c“ ausgewählt. Abb. 2 illustriert die verwendete Strategie bei Vorliegen von zwei Hindernissen.

Wenn sich in der Umgebung mehrere Hindernisse befinden, wird zuerst das der Startposition am nächsten liegende Hindernis berücksichtigt. Es wird festgelegt, in welchen Oktanten sich Start- und Zielposition befinden, um die benötigten Ausweichecken auszuwählen. Der letzte ermittelte Ausweichpunkt wird als neuer Startpunkt betrachtet und gestützt auf die Oktantenauswertung des ihm am nächsten liegenden Hindernisses werden weitere Ausweichecken ausgewählt. Dieser Vorgang wird so lange wiederholt, bis sich zwischen der letzten ausgewählten Ausweichecke und der Zielposition keine Hindernisse mehr befinden (s. nebenstehende Abbildung).

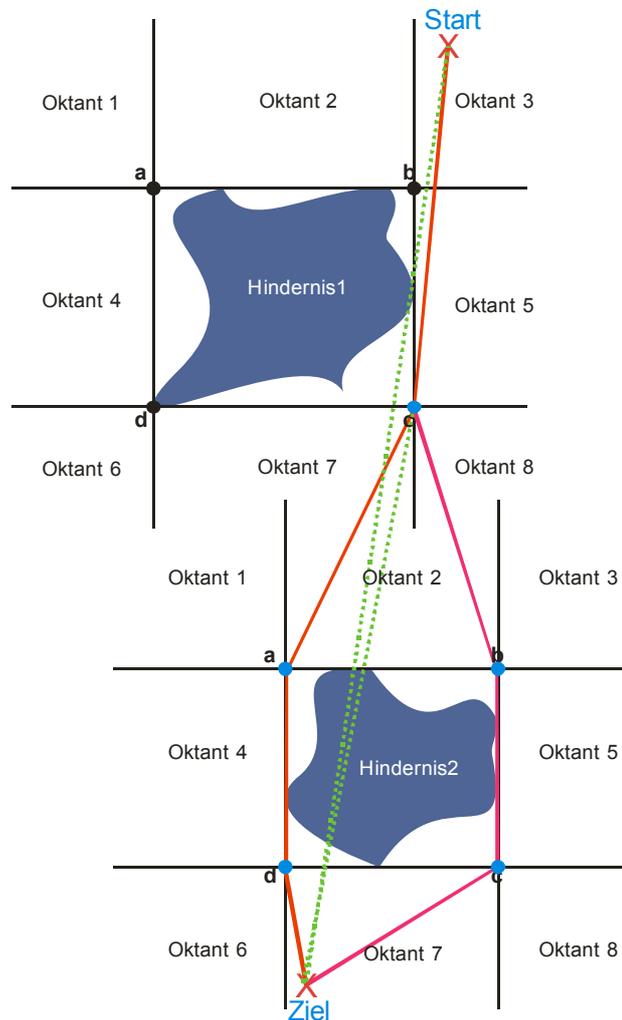


Abb. 2: Bahnplanung mittels Ausweichecken

		Start								
		1	2	3	4	5	6	7	8	
ZIEL	Oktant	1	0	0	0	0	b	0	d	b V d
	2	0	0	0	a	b	a	da V cb	b	
	3	0	0	0	a	0	a V c	c	0	
	4	0	a	a	0	ba V cd	0	d	d	
	5	b	b	0	ab V dc	0	c	c	0	
	6	0	a	a V c	0	c	0	0	0	
	7	d	ad V bc	c	d	c	0	0	0	
	8	d V b	b	0	d	0	0	0	0	

Tabelle 1: Ausweicheckentabelle

3. MATLAB-IMPLEMENTIERUNG:

Die präsentierte Methode wurde zum Test in MatLab implementiert. Es wurde ein Aktionsraum mit 12x12 Gitterpositionen festgelegt. Die Gitterpositionen wurden mit unterschiedlichen Werten, wie folgt, belegt:

- Startposition: rot
- mit Hindernis belegte Gitterposition: gelb
- freie Gitterposition: weiß
- Zielposition: blau
- Eckenposition: zyan

Abbildungen 3 und 4 zeigen zwei Testbeispiele.

In Abb. 3 befinden sich die Startposition im ersten Oktanten und die Zielposition im sechsten Oktanten. Laut der Ausweicheckentabelle, muss der Weg über den Eckpunkt (10, 3) verlaufen. Wie das Beispiel zeigt, wird eine einzige kollisionsfreie Bahn erzeugt, die im Bild durch einen grünen Pfeil markiert ist.

Im zweiten Beispiel liegen zwischen der Start- und der Zielposition zwei Hindernisse. Hier wird zuerst das der Startposition am nächsten liegende Hindernis betrachtet. Es ist ersichtlich, dass sich die Startposition im zweiten Oktanten und die Zielposition im siebten Oktanten befinden. Entsprechend der Ausweicheckentabelle werden die Eckpunkte (3, 4) und (6, 4) oder (3, 8) und (6, 8) ausgewählt.

Die zuletzt gefundenen Eckpunkte (6, 4) und (6, 8) werden jeweils als neue Startpunkte s_1 und s_2 betrachtet, und gestützt auf die Oktantantenauswertung werden die resultierenden Strecken, wie oben beschrieben, untersucht. In diesem Beispiel resultieren die Eckpunkte (8, 3) und (11, 3) oder (8, 10) und (11, 10).

Zwischen den letzten ausgewählten Ausweichpunkten (11, 3) und (11, 10) und der Zielposition befinden sich keine weiteren Hindernisse mehr. Die Bahnen werden nun durch kombinatorische Bildung von Wegen über Start- und Eckpunkte zu der

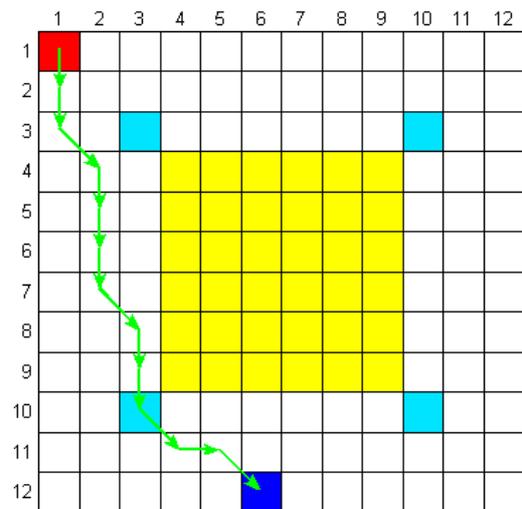


Abb. 3: Aktionsraum mit einem Hindernis

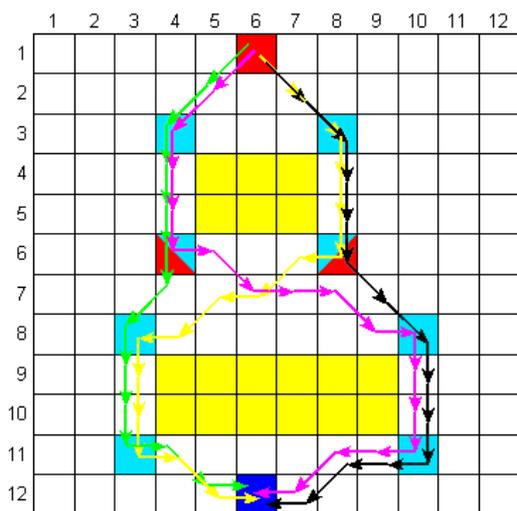


Abb. 4: Aktionsraum mit zwei Hindernissen

Zielposition gebildet, wobei Verbindungen von Eckpunkten der gleichen Ebene ausgeschlossen werden.

Für das vorliegende Beispiel sind vier kollisionsfreie Bahnen generiert worden, die im Bild grün $\{(1, 6), (3, 4), (6, 4), (8, 3), (11, 3), (12, 6)\}$, magenta $\{(1, 6), (3, 4), (6, 4), (8, 10), (11, 10), (12, 6)\}$, gelb $\{(1, 6), (3, 8), (6, 8), (8, 3), (11, 3), (12, 6)\}$ und schwarz $\{(1, 6), (3, 8), (6, 8), (8, 10), (11, 10), (12, 6)\}$ eingezeichnet werden.

4. SYSTEMAUFBAU

Das zur Verifizierung der Ergebnisse eingesetzte Robot-Vision-System besteht aus einem zentralen Leitreechner mit WLAN-Kommunikation und integrierter Bildauswertung, einer CCD-Kamera mit Weitwinkelobjektiv für die Bilderfassung, und mehreren mobilen Testfahrzeugen, die jeweils mit einem Mikrocontroller und einem PDA bestückt sind. Als Testfahrzeug wurde ein Modellfahrzeug verwendet, welches um einige Module erweitert wurde.



Abb. 5: Robot-Vision System

Es wurden Abstandssensoren für die Erfassung dynamischer Hindernisse und ein Mikrocontroller-Evaluationsbord verwendet, das als Schnittstelle zwischen der Elektronik des Fahrzeugs und dem im Fahrzeug integrierten PDA dient.

Der Aktionsraum mit einer Größe von 5m x 5m wird durch das Kamerasystem erfasst. Der Leitreechner wertet das aufgenommene Bild aus, indem er die Positionen der Fahrzeuge sowie die Positionen der statischen Hindernisse ermittelt. Eine kollisionsfreie Bahn wird dann, wie oben gezeigt, generiert und in entsprechende Fahrbefehle umgesetzt. Wie aus Abb. 5 zu entnehmen ist, kommunizieren die Fahrzeuge (Clients) mit dem Leitreechner (Server) über ein WLAN-Netz. Die Fahranweisungen werden an den PDA des betreffenden Modellfahrzeugs übertragen, welcher sie seinerseits an den Mikrocontroller weiterleitet. Der Mikrocontroller verarbeitet die Fahrbefehle nacheinander bis der berechnete Weg abgefahren ist. Im Falle einer Kollisionswarnung durch die Abstandssensoren stoppt das Fahrzeug schnellstmöglich und veranlasst eine neue Aufnahme der Kamera durch Senden einer Hindernismeldung an den Server. Eine Standortbestimmung anhand von Position und Ausrichtung sowie eine Neuberechnung

der Route wird dann vom Server durchgeführt, um mit einem entsprechenden Ausweichmanöver auf die neu entstandene Hindernissituation reagieren zu können.

5. ZUSAMMENFASSUNG

Mit der vorgestellten Bahnplanungsmethode, Bahnplanung mit Ausweichecken, können in Aktionsräumen mit beliebiger Anzahl von Hindernissen kollisionsfreie Fahrbahnen für mobile Roboter ermittelt werden, die jeweils eine Start- mit einer Zielposition verbinden. Diese Bahnen haben die Form eines Polygonzuges mit mehreren Kanten, die jeweils eine geradlinige Fahrtstrecke repräsentieren.

Die Methode zeichnet sich durch eine einfache Implementierung und schnelle Ausführung aus, da hier die Bahnen nicht durch kombinatorische Bildung von Wegen aus allen zur Verfügung stehenden Gitterzellen ermittelt werden, die dann auf die Kollisionsfreiheit geprüft werden. Ein weiterer Vorteil der Methode liegt in der bereits approximativen Form der generierten Bahnen.

Es ist an dieser Stelle noch einmal zu betonen, dass die erzeugten Bahnen sich in ihrer Gesamtlänge sowie Anzahl und Stärke der vorhandenen Lenkungen unterscheiden und dadurch auch in ihrer Qualität. Anhand dieser Eigenschaften können die erzeugten Bahnen u.a. unter Verwendung eines Optimierungskriteriums beurteilt werden.

References:

- [1] Cezar A. Sierakowski and Leandro dos S. Coelho. Study of two Swarm Intelligence Techniques for Path Planning of Mobile Robots. IFAC Robotics 'Robot Control I', 16th IFAC Juli 3-8, 2005, Pontifical Catholic University of Parana, Brazil.
- [2] Jean-Claude Latombe. Robot Motion Planning. Kluwer Academic publishers, ISBN 0-7923-9129-2 1991.
- [3] S. J. Russell, P. Norvig, International Edition, Artificial Intelligence: A Modern Approach, second Edition, ISBN: 0137903952, 2003.
- [4] G. Görz, C.-R. Rollinger, J. Schneeberger (Hrsg.), Handbuch der Künstlichen Intelligenz: 3. Auflage, Oldenbourg Verlag 2000, ISBN 3-486-25049-3, Kap. 4: Suche.
- [5] U. Lämmel, J. Cleve, Lehr- und Übungsbuch Künstliche Intelligenz. Lehr- und Übungsbuch. Fachbuchverlag Leipzig. 2004. ISBN: 3-446-22574-9. Kap. 3 , Problemlösung mittels Suche.
- [6] A. Steger. Diskrete Strukturen, Band 1: Kombinatorik, Graphentheorie, Algebra. Springer-Lehrbuch. Berlin: Springer, 2002. ISBN 3-540-67597-3. Kap. II, Graphentheorie.

Authors:

Prof. Dr.-Ing. Bernd Tibken
Dr.-Ing. Adnan Abou-Nabout
M.Sc. Ghaida Al Zeer
Bergische Universität Wuppertal, Lehrstuhl für Automatisierungstechnik/Regelungstechnik,
Fachbereich Elektrotechnik, Informationstechnik, Medientechnik
Rainer-Gruenter.Str. 21, 42097 Wuppertal
D – Wuppertal
Phone: (+49) 202 / 439 1754
Fax: (+49) 202 / 439 1756
E-Mail: alzeer@uni-wuppertal.de

C. Schröter, H.-M. Gross

Efficient Gridmaps for SLAM with Rao-Blackwellized Particle Filters

Abstract

Simultaneous localization and mapping (SLAM) has been an important field of research in the robotics community in recent years. A successful class of SLAM algorithms are Rao-Blackwellized Particle Filters (RBPF), where the particles approximate the pose belief distribution, while each particle contains a separate map. So far, RBPF with landmark based environment representations as well as gridmaps have been shown to work. Existing gridmap approaches typically used laser range scanners, because the high accuracy of that sensor keeps the state uncertainty low and allows for efficient solutions. In this paper, we present a combination of our previous work on map-matching with RBPF, which enable us to solve the SLAM problem also with low-resolution sonar range sensors. Furthermore, we introduce a simple and fast but very efficient shared representation of gridmaps which reduces the memory cost overhead caused by inherent redundancy between the particles.

1 Introduction and Related Work

In order to navigate autonomously, a basic requirement for a mobile robot is the ability to build a map of the environment. Because mapping depends on a good estimate of the robot's pose w.r.t. the environment, while localization needs a consistent map, the localization and mapping problems are coupled in applications where an unknown area has to be explored without an external position reference like GPS. The term Simultaneous Localization And Mapping (SLAM) has been coined for this problem [1]. SLAM can be seen as a generalization of the map building problem, as it describes the objective of acquiring a map of the environment without assuming any additional position information apart from those that can be derived from the mapping process itself.

There are two main criteria that can be used to categorize existing SLAM techniques: the kind of model used to describe the robot and environment state and the algorithm that is utilized to estimate the state belief.

In many SLAM approaches, the map representation is assumed to be a vector of point-like feature positions [2], also called landmarks. The attractiveness of feature/landmark-based representations for SLAM lies in their compactness. However, they rely on *a priori* knowledge about the structure of the environment to identify and distinguish potential landmarks. Furthermore, a data association problem arises from the need to robustly recognize landmarks. In contrast to landmark representations, gridmaps [3] do not make assumptions about specific features to be observable in the environment. They can represent arbitrary environment structures with nearly unlimited detail. However, they require a large amount of memory.

An effective means of handling the high-dimensionality in the SLAM problem has been introduced in the form of the Rao-Blackwellized Particle Filter (RBPF): in this approach the state space is partitioned into the pose and map state. A particle filter approximates the pose belief distribution of the robot, while each particle contains a map which represents the model of the environment, assuming the pose estimation of that specific particle to be correct.

Our aim here is to use a RBPF for grid mapping using no other sensory input than robot odometry and low-resolution sonar range scans. Since this requires a relatively large number of particles, we have to emphasize the efficient representation of the maps carried by the particles. To this purpose we present a short analysis of map redundancy between particles and a map storing scheme that exploits that redundancy in order to save memory.

The rest of the paper is organized as follows: We give a short introduction to the RBPF approach for SLAM in the next section. Section 3 will explain the specific details of our Sonar-SLAM implementation, while section 4 deals with the shared gridmap representation. Experiments with real robot data are presented and discussed in section 5, section 6 closes with a short summary and outlook.

2 Rao-Blackwellized Particle Filter for SLAM

As already described before, the complexity of the SLAM problem arises from the very high-dimensional state space, consisting of the variables describing the robot pose and the variables describing the environment state. In the case of gridmaps, the map alone usually contains a few thousands up to several million cells, each of which corresponding to a state variable. Obviously, a full posterior over the state is extremely costly to estimate. The idea of the RBPF in application to SLAM is to use a particle filter to estimate the robot trajectory distribution $p(x_{1:t}|z_{1:t}, u_{0:t})$ given the sequence of odometry measurements $u_{0:t}$ and environment observations $z_{1:t}$. This trajectory estimate is then used to estimate the desired distribution over map and trajectory:

$$p(x_{1:t}, m|z_{1:t}, u_{0:t}) = p(m|x_{1:t}, z_{1:t})p(x_{1:t}|z_{1:t}, u_{0:t}) \quad (1)$$

The particle filter works analogous to Monte-Carlo-Localization [7], except that instead of one given map each particle contains a separate map. To calculate the importance weights for $p(x_{1:t})$, each particle uses its own map. The map, in return, is built from the estimated trajectory of that corresponding particle. The effect is that a number of hypothesis maps are built, each corresponding to a possible trajectory. Importance weighting is performed with the weight for particle i following

$$w^{(i)} \simeq \frac{p(x_t^{(i)}|z_{1:t}, u_{0:t})}{\pi(x_t^{(i)}|z_{1:t}, u_{0:t})} \quad (2)$$

Here, $\pi(x_t^{(i)})$ denotes the proposal distribution. Typically, the motion model is used to generate the proposal distribution from the last particle generation (again, in analogy to localization), in which case the weight formula simplifies to

$$w^i \simeq p(z_t|x_t^{(i)}, m^{(i)}) \quad (3)$$

By repeatedly calculating importance weights followed by resampling to adapt the particle distribution to the estimated distribution, particles are preferred whose maps match new observations best, therefore the most likely map is selected.

3 Sonar Grid SLAM

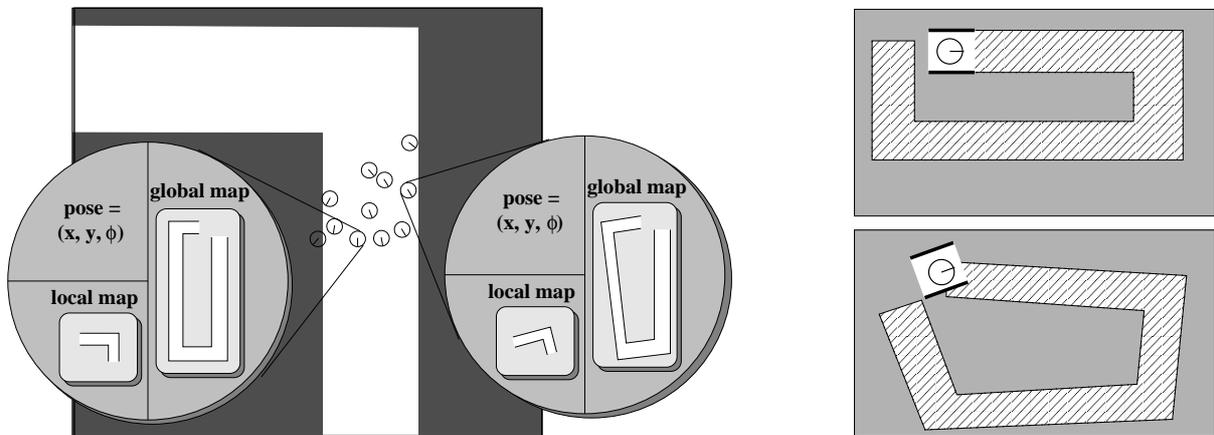


Figure 1: Left - Data representation overview: The particles model the distribution of the robot pose belief. Each particle carries a full map of the environment, which is a combination of the full particle trajectory and the sonar range measurements, and a local map, which only contains the most recent measurements. Right - Map matching: For the upper particle the local map (clean white) is aligned to the global map (hatched) very well, while for the lower particle, which does not contain a position belief consistent with the environment, the local map conflicts with the global map. This situation would result in a higher weight for the upper particle.

The base of our Sonar SLAM approach is a particle filter, where each particle contains a pose estimate as well as a map estimate. Without loss of generality we can assume the robot to start mapping at position $(0,0,0)$. While the robot moves, the particles move as well, according to the odometry readings and the probabilistic odometry motion model, which describes the uncertainty in the actual robot motion. Due to this uncertainty, the motion model contains a stochastic component, which effects in the particles spreading out and generating slightly different trajectories. Additionally, during motion the robot observes the environment by means of sonar range sensors. A map update is triggered frequently (approx. every 0.2m). In that map update, each particle adds the new environment observation to its own map, at its own estimated current position. Since the position estimates of the particles are slightly different, the maps differ as well (Fig. 1).

In order to determine the likeliness of a map hypothesis, we need to calculate particle weights by comparing expected and sensed measurement.

We already presented a way of comparing expectation and observation from sonar range sensors in a previous work on mapping [6]. There, we proposed an approach we called map matching: a local map was built from only the most recent sonar measurements and the resulting local map was matched against the global map to find the most likely position w.r.t. that global map. In order to be able to use map matching, each particle must not only know its global map, but also a local map. We exclude the most recent range measurements from the global map, and use those measurements for the local map. That way, global and local map are built from different data and we avoid comparing certain measurements against themselves. The local map can either be rebuilt from the pose and scan queues for each weight calculation or be persistent in the particle by just adding every new scan and forgetting old scans. Making the local map persistent is more efficient but less flexible.

The calculation of the match value between the local and global map is quite simple: For each occupied cell in the local map the occupancy value of the corresponding cell in the global map is tested. If the global map cell also is occupied, that cell contributes with a value of +1. If the global map cell is free, it contributes with a value of -1. Cells with unknown or undecided occupancy do not contribute. That way, the match value is positive if local and global map are very similar, and it is negative if many objects exist in the local map where there is free space in the global map. To obtain the actual particle weight $match^{(i)}$, an exponential function is applied as follows:

$$w^{(i)} = e^{\frac{match^{(i)}}{f}} \quad (4)$$

with f being a free parameter to influence the spread in the particle weights and therefore the speed of convergence.

4 Shared Gridmaps

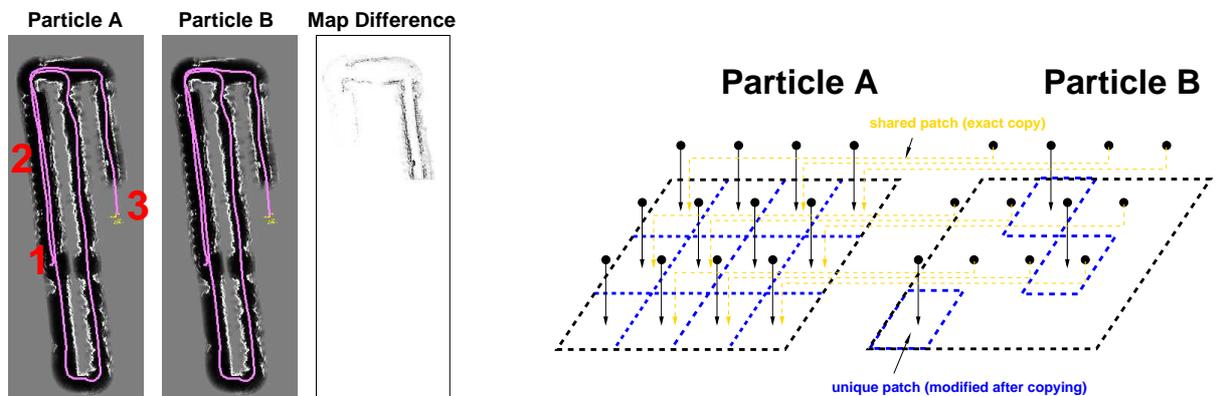


Figure 2: Left - Analysis: The robot started at position 1, closed the loop and moved onward to position 3. Particle B was generated as a copy of A during resampling approx. at position 2. Therefore, the major part of the map is identical between particles A and B. Right - Improved map representation: Particle A contains a map which consists of a number of separate patches. Particle B is created as a copy of A - the map of B consists of references to the patches in A. Only when A or B modifies a certain patch, it creates a real copy, so Particle A and B then have a separate instance of that patch.

A major problem with using gridmaps in RBPF is the memory cost: In a naive implementation, the number of cell values to be stored would be the product of grid size and particle number. However, the maps of the individual particles are not completely independent: In the resampling as part of the observation update, particles with low weights are deleted and replaced by copies of particles with higher weights. This results in multiple identical copies of the same map. Afterwards, each of the particles will modify its respective map differently, according to the path assumed through the probabilistic motion model: The copies will not remain identical, but it is important to notice the changes often only affect a small area of the already acquired map (see Fig. 2). The idea to save wasting memory for redundant information therefore is to split up the map into smaller patches and share those patches across the particles. When a particle A is cloned, each "copy" of a map patch belonging to the clone particle B is just a reference to the original patch. Only when either A or B modify a map patch later, a real copy is created in the local memory of the respective particle.

The effect of this representation is that the memory cost is not determined by the map area, but by the size of path loops. As long as a loop is not closed yet, particles are diverging and many path hypotheses are maintained. When the loop gets closed, only the best particles survive, and new particles are generated as copies of those few best fitting hypotheses. While a loop is open, each particle holds an own independent map of that loop, but when it is closed, only few unique maps of that specific loop (the best fitting ones) continue to exist. Therefore, the "residual" memory cost is determined by the entire map area (the sum of all loops) and nearly independent of the particle number, while the peak memory cost is determined by particle number and maximum length of a single loop.

5 Experiments

To test our approach we built maps of a home store which is the regular test environment for our navigation algorithms. This environment is very well suited for our proposed SLAM approach as it essentially consists of a large number of small circles of hallways (50 to 100 m loop length). Fig. 3 shows the resulting map and the the overall memory usage for all particles over time. The data shows that our SLAM approach using map matching and shared gridmaps builds a consistent map with a bounded amount of memory. Only robot odometry and sonar range sensors were used in those experiments.

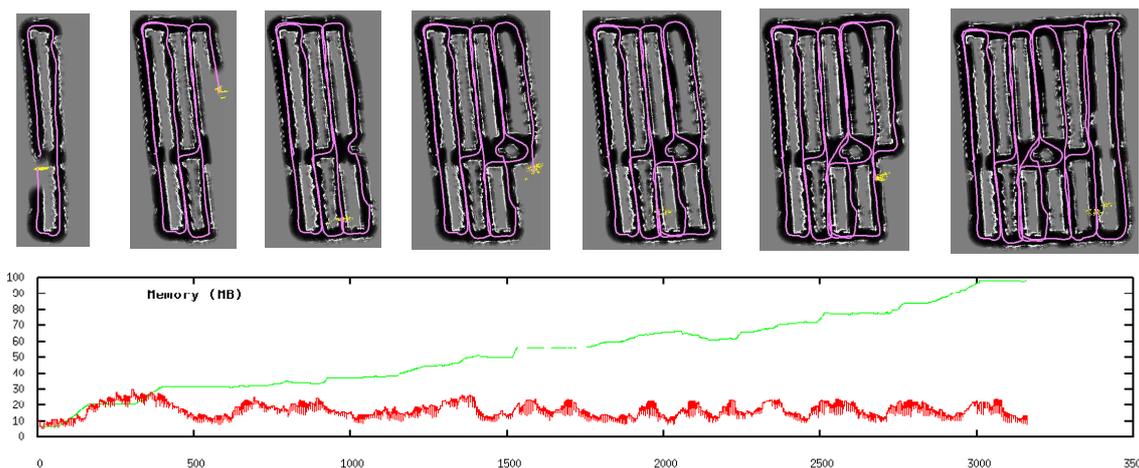


Figure 3: Row 1: Several steps of mapping (500 particles): Yellow dots denote the particle positions, path (magenta line) and map for one selected particle are shown.
 Row 2: Map memory cost for plain gridmaps (green) and shared maps (red, see section 4). It is clearly visible that the memory for plain maps is growing monotonously, while for the shared maps the cost collapses with each loop closure.

6 Summary & Outlook

We presented an implementation of RBPF with gridmaps which is able to solve the SLAM problem with low-resolution sensors such as sonar range finders. Furthermore, we introduced a shared map representation for particle filters which effectively makes the maximum memory cost depend on the loop size instead of the overall map size. Experiments show that our approach is well suited for large-scale environments consisting of many loops.

References

- [1] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*, The MIT Press, 2005.
- [2] R. Smith, M. Self, and P. Cheeseman, “A stochastic map for uncertain spatial relationships,” in *4th International Symposium on Robotic Research*. 1987, MIT Press.
- [3] H. Moravec, “Sensor fusion in certainty grids for mobile robots,” *AI Magazine*, vol. 9, no. 2, pp. 61–77, 1988.
- [4] K. P. Murphy, “Bayesian map learning in dynamic environments,” in *Advances in Neural Information Processing Systems 12 (NIPS99)*, 1999, pp. 1015–1021.
- [5] D. Haehnel, W. Burgard, D. Fox, and S. Thrun, “An efficient FastSLAM algorithm for generating maps of large-scale cyclic environments from raw laser range measurements,” in *Proc. IROS-2003*, pp. 206–211.
- [6] C. Schroeter, H.-J. Boehme, and H.-M. Gross, “Robust map building for an autonomous robot using low-cost sensors,” in *Proc. SMC-2004*, pp. 5398 – 5403.
- [7] D. Fox, W. Burgard, F. Dellaert and S. Thrun, “Monte Carlo Localization: Efficient Position Estimation for Mobile Robots” in *Proc. AAAI Natl. Conf. on Artificial Intelligence*, 1999, pp. 5398 – 5403.
- [8] C. Schroeter, A. Koenig, H.-J. Boehme, and H.-M. Gross, “Multi-sensor Monte-Carlo-Localization combining omnivision and sonar range sensors,” in *Proc. ECMR-2005*, pp. 164–169.

Author Information:

Christof Schröter, Horst-Michael Gross Department of Neuroinformatics and Cognitive Robotics, Faculty of Computer Science and Automation,
 Ilmenau Technical University, PO Box 10 05 65, 98684 Ilmenau
 Tel: +49 3677 69 1306
 Fax: +49 3677 69 1665
 E-mail: christof.schroeter@tu-ilmenau.de

S. Müller / A. Scheidig / A. Ober / H.-M. Gross

Making Mobile Robots Smarter by Probabilistic User Modeling and Tracking

Abstract

This paper considers the problem of tracking and modeling users of a mobile service robot. In order to adapt the behavior of an interacting robot to the user's preferences, the system has to know about the people in its surrounding and to model their properties. To that purpose, a consistent probabilistic model is introduced here, which is realizing the tracking process and storing the information.

1 Introduction

The weak acceptance of a robot actively offering services is a hard problem. People are not willing to interact with a stupid computer. One way to increase the rate of interactions is to make the robot smarter by selecting a behaviour more appropriate to the specific person. One essential step on that way, is to model the state of a user properly. In particular, the robot has to know where people are in its surrounding and what are their objectives. The task of people tracking typically is done using different sensors which are integrated in a probabilistic model, e.g. a particle or kalman filter. In former work [3] we developed a probabilistic multi-cue people tracker, which successfully runs on our shopping robot SCITOS. Besides the presence of possible users, further information like their gender, age and if they are in a hurry or willing to interact are necessary for an adequate reaction of the robot. Therefore, the simple model of people's positions has been improved by modeling different people and their properties. Because many of these properties can not be observed continuously, it is necessary to remember and recognize people, which typically is done by analyzing face images [5] or by a color model of a person. In our model both methods are used as cues for identification, which is done implicitly in the model. Based on a knowledge base like that, the robot can infer various facts about its situation for an intelligent action selection.

On a mobile robot there are different sensory systems gathering information about the robots environment (see fig. 1). A probabilistic model is the central place where all the

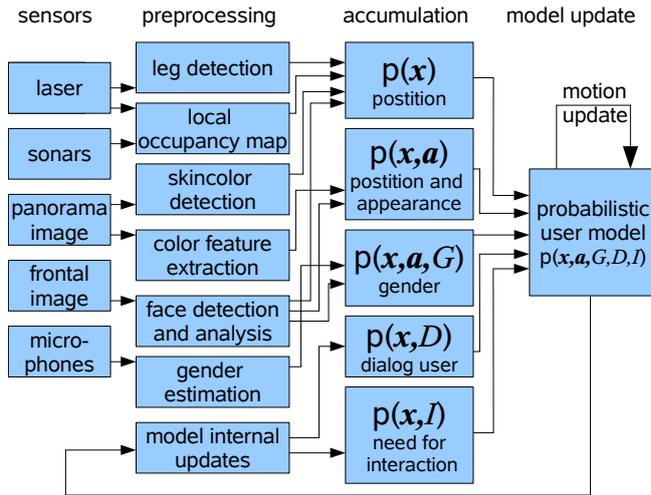


Figure 1: Layers of our systems architecture: (left) sensors providing raw data, (middle) preprocessing modules extracting information, (right) short time aggregation of information for model update

data are merged together. Our robot HOROS, a Pioneer platform based interacting robot, is equipped with some range sensors like a laser and sonars, two cameras yielding a panorama image and frontal images, and a pair of microphones for sound analysis mounted at the head.

In the following section the architecture of our system is described in more detail, before the probabilistic user model is explained. At the end of the paper some ongoing experiments are shown which are made for learning a classification of people's movement trajectories.

2 System Architecture

The environment recognition system of HOROS is based on a couple of sensors providing raw data as ranges, visual inputs, and a stereo audio signal. Based on these, different preprocessing modules are extracting information on people hypotheses in the surrounding of the robot. A brief overview on these modules is given below. After single observations of possible people's positions \mathbf{x} , person's color profile and facial features \mathbf{a} or their gender G are extracted, they are aggregated in the accumulation layer. Similar to the human brain here the different modalities are combined with their sensory context. Thus, observations on the appearance of a person are combined with a position, bearing only position estimation

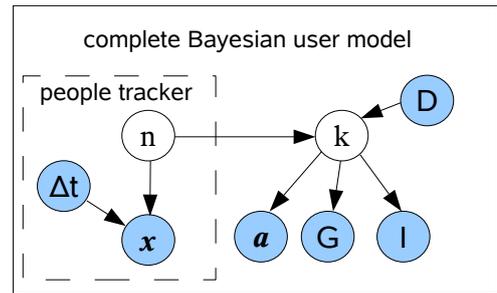


Figure 2: Bayesian network of the probabilistic model: \mathbf{x} position of hypothesis n , Δt time difference for representing movement trajectories, \mathbf{a} appearance vector (face and color) of person k , D is the person in dialog with the robot, I need to interact

from color detection are combined with distance measurements in the local occupancy map. Further, an extracted color pattern or an appearance vector from face analysis is assigned to a position and gender estimation from face analysis as well as from sound analysis are associated to an appearance \mathbf{a} and a position. As a last source of information there are model internal observations. E.g. the need of a person to interact with the robot can be estimated from the movement trajectory (see below) and the person standing in front of the robot is supposed to be that user who is in dialog with the robot. In addition to the improved quality of observations by combining them, the accumulation layer compensates the asynchronously occurring observations using a time slicing mechanism. In each time slice all the observations are accumulated before one update at the model is done.

Knowing all the aggregated observations, which are represented as a joint probability, the probabilistic model is updated. In fig. 2 the Bayesian network is shown. Here, besides the shaded circles, representing observed variables, the non shaded circles contain discrete hidden variables for the component n in a mixture of Gaussians for the position \mathbf{x} and the id k of the known people. The user properties are represented as conditional probabilities $p(\mathbf{a}|k)$ for the appearance, which is modelled as a Gaussian, $p(G|k)$ for the Bernoulli distributed gender, and $p(I|k)$ for the also Bernoulli distributed property of need for interaction with the robot. In order to decide who is the person, which is currently in dialog with the robot, $p(k|D)$ is representing a discrete distribution over the person ids k . There is one k standing for a non person or nobody, such that it is possible to express that nobody is in dialog, as well as that a position hypothesis n belongs to a non human object.

On each time step, the model is updated in three steps. In order to realize a Bayesian state estimation, similar to a Bayes-filter, first we have to perform a motion update. Here the positions are propagated according to the history which is represented by Δt and a motion model of the robot. Further the property of being dialog user $p(k|D)$ and the need for interaction $p(I|k)$ are diffused in order to allow a change over time. Second step is to infer the hidden and the unobserved variables given the current observations from accumulation layer. From that step we get a new prior distribution for the third step, which is a MAP estimation of the distribution parameters given the new observation and the prior model from last time step. Following this update regime, the model always contains the complete knowledge about people in the surrounding of the robot, but also the knowledge on absent people inspected before. With the current model the robot can easily infer e.g. the position of the dialog user by evaluation of $p(\mathbf{x}|D = 1)$ or the gender of a person trying to interact

by inferring $p(G|I = 1, D = 0)$. Due to the simple structure of the Bayesian network, it is easy to introduce further variables for user properties. All it takes is a source of information producing observations on this variable. Due to the ability of our face analysis system to estimate the age of a person, we would easily be able to add a variable A and a conditional probability $p(A|k)$ to represent the age of a user.

3 Recognition subsystems

As mentioned above, there are different preprocessing modules generating observations on the various state variables. A first module is consuming the laser scans, while classifying the segments of the scan as a pair of legs. Using a simple collection of criteria explained in [1], segments with a limited variance in distance and a defined size are classified as possible legs. If there are two legs within a distance of less than $0.5m$ a Gaussian hypothesis for a position \mathbf{x} observation is generated and sent to the accumulation layer. The laser scan and the noisy sonar measurements furthermore are integrated into an occupancy gridmap representing the local environment of the robot. This map is used by the accumulation layer to query the distance of objects observed in a known direction, as the skin color detections in the panorama image. Using a multi-instances particle filter [6] which is tracking the skin colored regions in the image, there are Gaussian hypotheses for directions of people which are limited in their distance by a map lookup.

To get a feature for distinguishing people, a color pattern is taken from the panorama image by determining the average UV value (luminance independent parts of YUV color space) and the variance of that pixels of three regions on the upper part of the body of people in the image. These are the first components a_1, \dots, a_{12} of the appearance vector. To find these regions, a parametric contour and skincolor model is fitted into the panorama image using a Monte-Carlo gradient descent. Following the list of modules in fig. 1, the next is the face detection and analysis. Here for detection the well known Viola and Jones detector [4] is used. For analyzing the face then an Active Appearance Model is used, which is tracking the frontal face while delivering parameters for position shape and texture of the face. The parameters for shape and texture are used as further components of the \mathbf{a} vector. Furthermore, the direction of the face and the distance from local map are used to generate a Gaussian position hypothesis. By classifying the appearance parameters it is possible to classify the gender of the person tracked, which is used to generate a $p(G, \mathbf{x}, \mathbf{a})$ observation. A further sensor based cue is the sound analysis. Here sound source localization is utilized to

get a hint for the direction of a speaker, before a speech detection and a gender classification is done [2]. The classification is based on Mel Frequency Cepstral Coefficients and on the fundamental frequency of the speaker. Also using a lookup of the distance the result of that module is a probability of the gender and a given position.

In contrast to the other variables, the observations for D and I do not result from external sensoric inputs. For updating these parts of the model, information from the model itself is used. The property of being the user who is in dialog with the robot, is estimated from the users position (similar to our shopping assistant SCITOS [3]). Thus each person standing in or approaching the region in front of the robot is supposed to be in dialog with the robot. The need for interaction also is only extracted from the movement trajectory of people. For generating observation distributions, first for each object $i \in \{1, \dots, n\}$ in the model the movement trajectory $p(\mathbf{x}, \Delta t | n = i)$ is inferred. Then using a heuristic measurement model $p(D|\mathbf{x}, \Delta t)$ and $p(I|\mathbf{x}, \Delta t)$ is used to infer the current distribution of D and I . Together with the current position $p(\mathbf{x}|\Delta t = 0, n = i)$ a new observation for updating the model is composed.

4 Experiments and Results

Because it is difficult to quantify such a probabilistic model, experiments have been done for single parts of the model separately. To evaluate the accuracy of the position tracking system (fig. 2 left), a simple experiment with a given reference from a top view camera has been done. By comparing the tracking hypotheses to the reference, we got a distance error below $0.5m$. Fig. 3 shows some exemplary trajectories taken during the experiments. The solid line is the estimated path and the dotted one is the reference. The shape of graphs suggest that the movement trajectories are quite specific for people who are interested in an interaction (graphs a and b) and those who are not (graphs c and d). For evaluating the abilities of the model to estimate I currently a complex experimental session has been accomplished. The robot there was running an information terminal application, while capturing the movement trajectories of the people in its surrounding. After a person completed its interaction or when a person only passed the robot, a questionnaire has been filled during an interview. By asking, if people are in a hurry or not and if they had an intention to use the information terminal, we intend to create an empiric measurement model $p(I|\mathbf{x}, \Delta t)$. First findings on the labeled trajectories are disillusioning. There are many factors of influence, which are not included in the model yet. Thus for increasing the reliability of an empiric model, the

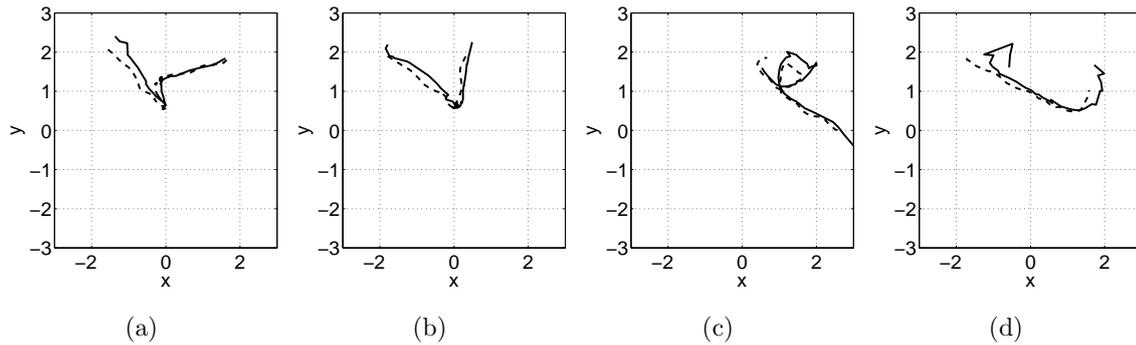


Figure 3: Exemplary trajectories of people moving in the surrounding of the robot, robot standing at $(0, 0)$ facing upwards, solid line: estimated path, dashed line: top down reference

specifics of the environment have to be taken into account. Depending on the position of doors and other points of interest in the room, the model will be suitable only for a fixed position. Therefore, higher effort will be necessary on encoding the trajectories for the small number of train data (157 non interacting people, 53 interactions at three different locations) to allow a satisfying generalization.

5 Conclusion

In this paper an overview of our probabilistic user model is given. We could show that tracking of positions works well, but evaluation of the model parts for user identification and estimation of user properties based on trajectories are still in progress.

References

- [1] K. O. Arras, Ó. M. Mozos, and W. Burgard. Using boosted features for the detection of people in 2d range data. In *Proc. of 2007 IEEE International Conference on Robotics and Automation*, Roma, Italy, 2007.
- [2] R. Brueckmann, A. Scheidig, C. Martin, and H.-M. Gross. Integration of a sound source detection into a probabilistic-based multimodal approach for person detection and tracking. In *Proc. Autonome Mobile Systeme (AMS 2005)*, pages 131–137. Springer, 2005.
- [3] S Müller, H.-M. Gross, A. Scheidig, and H.-J. Boehme. Are you still following me? In *Proc. of the 3rd European Conference on Mobile Robots (ECMR2007)*, 2007.
- [4] P. Viola and M. Jones. Fast and robust classification using asymmetric adaboost and a detector cascade. In *NIPS 2001*, pages 1311–1318, 2001.
- [5] Böhme H.-J Gross H.-M. Wilhelm, T. Classification of face images for gender, age, facial expression, and identity. In *Artificial Neural Networks: Biological Inspirations - ICANN 2005, LNCS 3696*, volume I, pages 569–574, 2005.
- [6] T. Wilhelm, H.-J. Boehme, and H.-M. Gross. A multi-modal system for tracking and analyzing faces on a mobile robot. In *Robotics and Autonomous Systems*, volume 48, pages 31–40, 2004.

Dipl.-Inf. Steffen Müller, Dr.-Ing. Andrea Scheidig, Antje Ober,

Prof. Dr. Horst-Michael Gross

Ilmenau Technical University, Neuroinformatics and Cognitive Robotics Lab, POB 10 05
65, 98684 Ilmenau

A. Swerdlow / T. Machmer / K. Kroschel / A. Laubenheimer / S. Richter

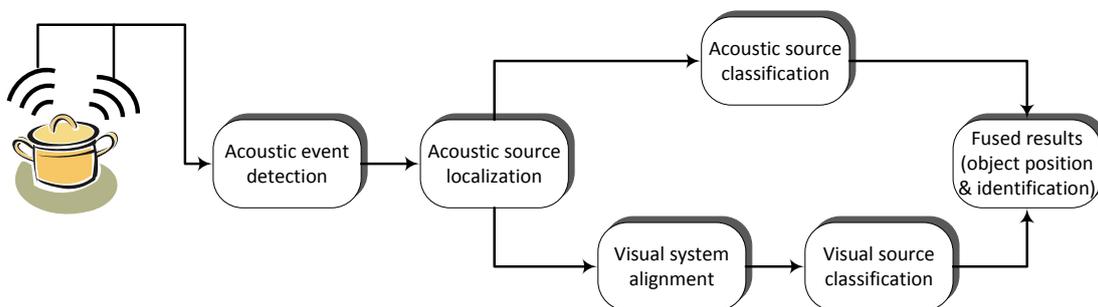
Opto-acoustical Scene Analysis for a Humanoid Robot

INTRODUCTION

The goal of the paper is to add the ability of a human to scan a scene acoustically and optically to a humanoid robot whose head is equipped with a visual system and a microphone array. The skill to scan and analyze a scene is an urgent precondition for a robot in order to cooperate with a human.

A typical peculiarity of the auditory system is that it senses its environment almost omnidirectional, whereas the visual system is only focused on a section of the environmental scene.

Figure 1: Block diagram of the opto-acoustical evaluation of the environment.



The humanoid robot extracts the direction of acoustic events by the use of the acoustic sensory system. This is realized by a microphone array consisting of four microphones which are positioned at the ears, the forehead and the chin of the robot. Since the distance information which might be gained by analysis of the acoustic signals, is too imprecise, it is not used. Instead, the head is moved into the detected direction and the optical signal can be picked up by the visual system. Thus, by fusing acoustical and optical information, the coordinates of the event can be determined and the object identified by evaluating the acoustical and optical information. For exemplification, such an opto-acoustical evaluation of the environment is shown in Figure 1.

THE AUDIO SYSTEM

Acoustic Sound Source Localization

The technique of choice in most passive acoustic sound source localization systems using a microphone array is a two-step procedure. First, the time difference of arrival (TDOA) of sound signals in a pair of spatially separated microphones is estimated. Then the estimated TDOA is used in combination with the known microphone array geometry for the localization of the sound source in the environment. The most popular approach for determining the TDOAs is the Generalized Cross Correlation (GCC) method [1, 2], which was also used for this work.

Acoustic Sound Source Identification

In addition to the acoustic localization, the identification of localized persons and ambient sound sources is another major part of the acoustic scene analysis. The interaction between man and machine gains more and more importance. Typical applications are for instance the identification of speakers by humanoid robots or the identification of passengers within a car to adjust position and speaker specific properties.

We use the Mel Frequency Cepstral Coefficients (MFCC) as spectral features in combination with the Gaussian Mixture Model (GMM) to identify speakers [3, 4]. For the classification of ambient sound sources we present another approach. The sampled instationary signal $s(k)$ requires a short time spectral analysis based on segments of 16 ms each and an overlap by the factor 0.5, within which the signal is assumed to be stationary. Like speaker identification, these sources are usually instationary. In contrary to the speaker identification, data processing takes place in the time domain.

In order to be able to detect an acoustic event, the energy within a frame is calculated for each frame. The energy $en(\kappa)$ in the frame κ of length $L = 256$ is

defined as $en(\kappa) = \frac{1}{L} \cdot \sum_{k=n_\kappa}^{n_\kappa+L-1} s^2(k)$.

For the classification of acoustic events, autoregressive (AR) models are used. The estimation of the detected sound class is done in the following way:

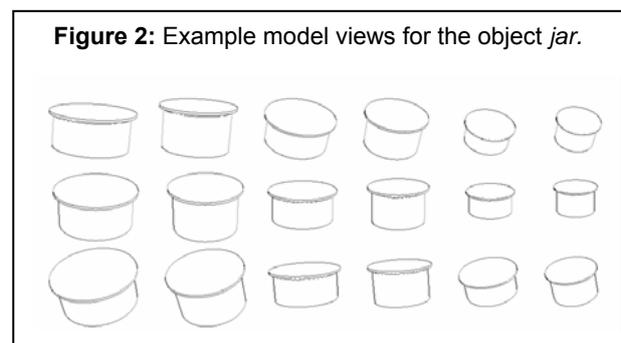
$$\hat{K}(\kappa) = \arg \min_{c=1, \dots, N_K} \left[\min_{j=1, \dots, P^{(c)}} \sum_{i=(\kappa-1) \cdot L}^{\kappa \cdot L-1} \left(s(k) - \sum_{\ell=1}^M p_{j,\ell}^{(c)} \cdot s(k-\ell) \right)^2 \right].$$

For each sound class $K^{(c)}$ with $c = 1, \dots, N_K$ to be recognized, one or more AR models $\mathbf{p}_j^{(c)}$ with $j = 1, \dots, P^{(c)}$ of order M are appointed. For every sound class $K^{(c)}$ and the associated prediction coefficients $\mathbf{p}_j^{(c)}$, the prediction error $e_j^{(c)}(k)$ for the sample $s(k)$ is determined. To be able to determine which model fits the currently handled frame κ at best, the energy of the prediction error signal is calculated for every sound class $K^{(c)}$ and the associated models $\mathbf{p}_j^{(c)}$ over the entire frame κ . Subsequently, the value of the prediction error of the model $\mathbf{p}_j^{(c)}$ and the sound class $K^{(c)}$ which represents the frame κ at best is calculated by finding the minimal prediction error. Finally, the frame κ is assigned to the estimated sound source class \hat{K} by calculating the index of the class with the lowest prediction error.

THE VISUAL SYSTEM

For pose and shape estimation, we have developed an appearance based method. Like [5] und [6], we use different model views to calculate the parameters of the object's pose. However, instead of using a 2-D-similarity transformation, we apply a chain of different filtering steps to obtain the best fitting model view.

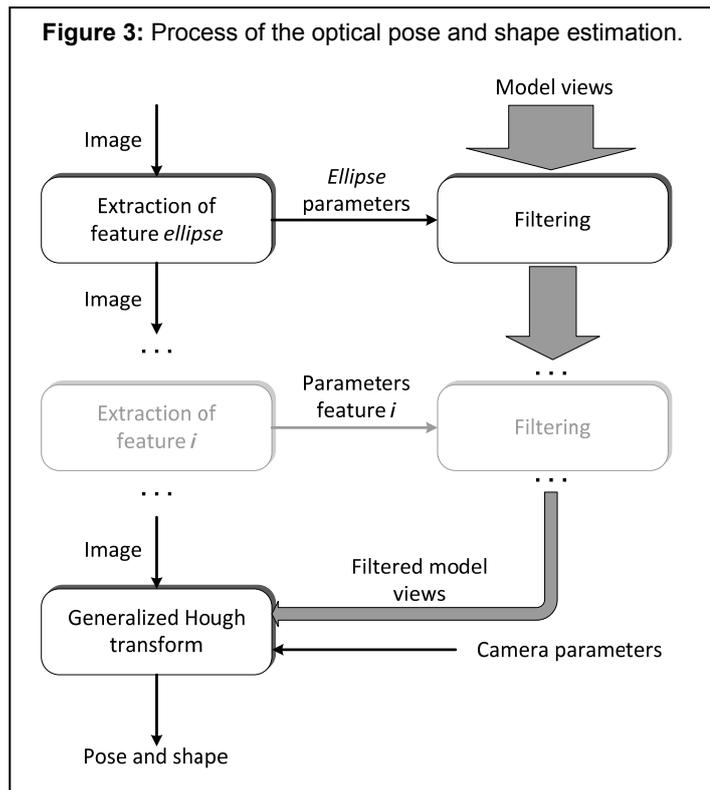
The generation of virtual views of the objects, which are to be detected, is accomplished offline. Therefore CAD models are used. A virtual camera is placed on the surface of a sphere surrounding the CAD model of a given object. By changing radius r , azimuth and polar angles θ and ϕ , respectively, the model views are generated. By using deformable models, based on [7], we can not only detect static objects, but we are also able to deal with objects with different shapes. Figure 2 shows some example model views created for the object *jar* that we have used for our experiments.



In order to provide an accurate object pose, an adequate large number of model views is required. To obtain the model view which fits best to the object in the camera image, several steps of filtering are applied to the set of model views. For this purpose features, for instance ellipses, edge segments or junctions are extracted in both the model view images and the camera image. By calculating a distance vector between these features in the model view images and the camera image, the most

likely model views to fit the camera image's object are obtained and at each step the number of model views is reduced. Using efficient algorithms for feature extraction, the time needed for each step can be kept low.

The last step we apply is an extended form of the generalized Hough transform [8]. The model views passing all previous filters are processed and the one with the best fit to the camera image's object is obtained. Using this



model view and its position in the camera image, which is provided by the generalized Hough transform, the final object pose can be calculated. Figure 3 illustrates the process of the optical pose and shape estimation.

EXPERIMENTAL SETUP AND SELECTED RESULTS

For data recording, omni-directional electret condenser microphones in combination with a colour camera (resolution 1392 x 1040 pixels) were used. Since in the project described in this paper a typical scene from a kitchen is sensed, real experiments were carried out in a test environment. For both, the acoustic and the optical analysis, we used jars with different shapes for our experiments. Additionally, each detection system was tested on further objects which are characteristic for a kitchen environment.

For the acoustic analysis, various kitchen appliances¹ in combination with two untrained sound sources² were used. The percentage of correct frame classifications and the required number of AR models of order 16 for each ambient sound source state are summarized in Table 1; the standard deviation is given in brackets.

¹ **KC(P)**: kitchen clock (programming), **KC(E)**: kitchen clock (expiration), **CG(A)**: coffee grinder (activity), **T(D)**: toaster (down), **T(U)**: toaster (up), **TP(R)**: telephone (ringing), **J(B)**: jar (boiling)

² **US(S)**: untrained source (speech), **US(KN)**: untrained source (knocking noise)

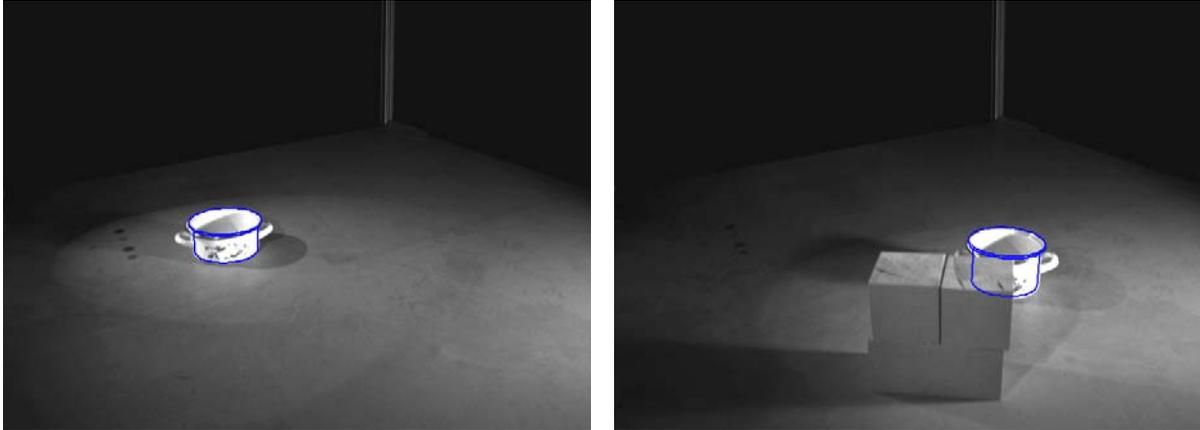
Table 1: Percentage results of the frame based classification with AR models of order 16 for kitchen appliances.

Sound class\AR model	KC(P)	KC(E)	CG(A)	T(D)	T(U)	T(R)	J(B)
KC(P)	98.10 (1.95)	1.90 (1.95)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
KC(E)	1.15 (0.26)	98.85 (0.26)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
CG(A)	1.31 (0.41)	0.00 (0.00)	80.24 (2.33)	12.99 (1.68)	4.99 (1.17)	0.48 (0.30)	0.00 (0.00)
T(D)	1.03 (0.09)	0.00 (0.00)	2.53 (1.10)	84.55 (3.37)	11.84 (3.42)	0.04 (0.09)	0.00 (0.00)
T(U)	1.07 (0.01)	0.00 (0.00)	0.40 (0.50)	11.52 (3.01)	87.01 (3.14)	0.00 (0.00)	0.00 (0.00)
TP(R)	1.03 (0.09)	0.00 (0.00)	1.50 (0.36)	0.32 (0.33)	0.12 (0.18)	97.03 (0.52)	0.00 (0.00)
J(B)	0.99 (0.00)	0.00 (0.00)	0.95 (0.66)	8.12 (2.20)	1.43 (0.55)	0.00 (0.00)	88.51 (1.93)
US(S)	3.33 (0.77)	0.00 (0.00)	81.03 (3.90)	4.00 (1.53)	3.09 (0.82)	8.16 (3.32)	0.40 (0.34)
US(KN)	1.07 (0.18)	0.08 (0.11)	36.87 (6.30)	44.24 (5.52)	11.64 (1.40)	6.10 (1.69)	0.00 (0.00)
Average number of needed AR models	5.12	5.68	16.36	17.48	17.40	15.20	20.04

As can be seen, the classification with AR models is a multiple detection issue. This is the reason why also untrained sound sources (speech, knocking noise) are always classified. To avoid this deficiency, frames can be aggregated into blocks of defined size. A trade-off has to be made between a high percentage of correct classification results and a high number of estimates which is crucial for the continuous real-time classification. The entire acoustic event within the actual block matches the sound source class, which prevails in this block. This approach increases the detection rate to nearly 100%. Additionally, blocks can be labelled as invalid (rejected) in case that less than a defined percentage of frames within one block classify the same sound class.

For the optical analysis, we used several kitchen objects like cups, bowls, dishes, and the above-mentioned jars. Figure 4 shows two different-sized jars. By means of the calculated pose and shape parameters, a CAD model of the jar is projected into the respective camera image. As one can see, the presented method is able to detect even partially occluded objects.

Figure 4: Results of the optical pose and shape detection.



CONCLUSION

Fusing the acoustical and optical long-term information leads to what we call an opto-acoustical map which can be provided to the humanoid robot and therewith enhances the robots environment analysis capabilities, even under bad illumination conditions and in acoustically challenging environments.

ACKNOWLEDGMENT

This work has been supported by the German Science Foundation DFG within the Sonderforschungsbereich 588 *Humanoid Robots*.

References:

- [1] C. H. Knapp and G. C. Carter. The generalized correlation method for estimation of time delay. *IEEE Trans. Acoustics, Speech, and Signal Proc*, 24(4):320–327, 1976.
- [2] D. Bechler and K. Kroschel. Confidence scoring of time difference of arrival estimation for speaker localization with microphone arrays. In 13. Konferenz Elektronische Sprachsignalverarbeitung (ESSV), Dresden, 2002.
- [3] D. O’Shaughnessy. *Speech Communication - Human and Machine*. IEEE Press, New York, 2000.
- [4] K. Kroschel and D. Bechler. Demonstrator for automatic text-independent speaker identification. In DAGA 2006, Braunschweig, 2006.
- [5] A. R. Pope. Learning to recognize objects in images: acquiring and using probabilistic models of appearance. Dissertation, University of British Columbia, Canada, 1995.
- [6] C. Plagemann. Ansichts-basierte Erkennung und Lokalisierung von Objekten zur Initialisierung eines Verfolgungsprozesses. Diplomarbeit, Universität Karlsruhe (TH), 2004.
- [7] A. Laubenheimer. Automatische Registrierung adaptiver Modelle zur Typerkennung technischer Objekte. Dissertation, Universität Karlsruhe (TH), 2004.
- [8] D. H. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2): 111-122, 1981.

Authors:

Alexej Swerdlow, Timo Machmer, Kristian Kroschel
Universität Karlsruhe
Institut für Nachrichtentechnik
Kaiserstr. 12, 76131 Karlsruhe, Germany
Phone: +49 - 7 21 - 608 3348
Fax: +49 - 7 21 - 608 3799
E-mail: swerdlow@int.uni-karlsruhe.de

Astrid Laubenheimer, Steffen Richter
Fraunhofer Institute IITB
Autonomous Systems and Machine Vision
Fraunhoferstr. 1, 76131 Karlsruhe, Germany
Phone: +49 - 7 21 - 6091 411
Fax: +49 - 7 21 - 6091 233
E-mail: laubenheimer@iitb.fraunhofer.de

A. Ahranovich / S. Karpovich / K. Zimmermann

Multicoordinate Positioning System Design and Simulation

ABSTRACT

The approach of multicoordinate positioning control system with parallel kinematics level division is described in the paper. By such system hierarchy definition, the information integrity is achieved. One of the most effective motion program building approaches is based on holonomic automatic systems, whereas this approach presupposes some definite structure of differential analyzer working-out. This structure enables motion program building in a uniform and transparent form enabling affix dynamics control.

TRIPLANAR CONTROL SYSTEM HIERARCHY

Triplanar [1] is a multicoordinate complex positioning system with 6 degrees of freedom. Like any other complex equipment's control system, Triplanar's one is divided into several levels aiming to represent the information and functional integrity of each of them. Each control level is based on the lower one, so by such means the information integrity is achieved. Fig. 1 represents the block-diagram of Triplanar control system.

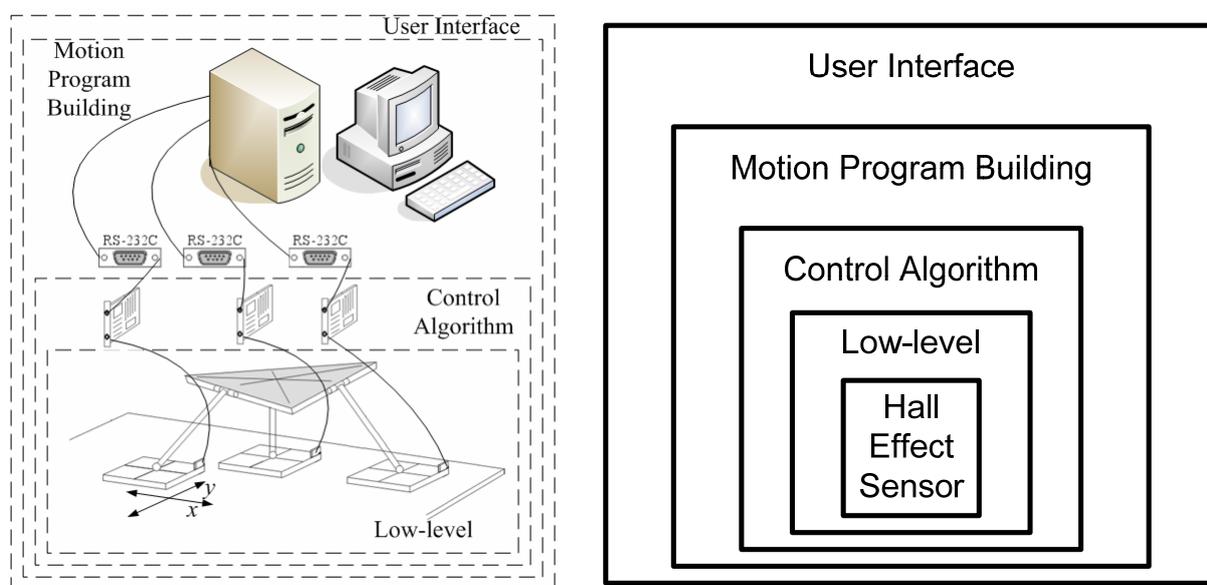


Fig. 1 Triplanar Control Levels

The general upper level software aims to bring user commands to the lower level. The input interface of the general software is user interface. The Motion Program Building level is supposed to build motion program basing on user input data and pass control signals to the lower system. While plenty of motion program building algorithms exist, the current work discusses one which is based on holonomic automatic systems.

MOTION PROGRAM BUILDING

The motion program building is based on differential analyzers synthesis. This approach enables easy parameterization of initial equations connecting system variables and therefore, it enables position demands generation on any topological manifold which is described with the help of equation:

$$F_j(t, x_1, x_2, \dots, x_n) = 0; \quad j = 1, 2, \dots, m \quad (1)$$

The first-order system of differential equations:

$$\frac{dx_i}{d\varphi} = f_i, i = 1, 2, \dots, n, \quad (2)$$

the solution of which satisfies the equation (3) in established range of variables M is built according to the following parameterization condition:

$$\sum_{i=1}^n \frac{\partial F}{\partial x_i} \frac{dx_i}{d\varphi} = 0 \quad (3)$$

This problem has a solution set, at that functions f_i in all cases depend on partial derivatives $\frac{dF}{dx_i}$. For analytical algorithm simplification it is concerned that functions f_i are

linear functions of mentioned above partial derivatives. This method of differential analyzers synthesis has an essential advantage: the argument φ , which is concerned to be a system parameter, can be any analytical function, what specifically lets realize the argument control, which is necessary for differential analyzer structure simplification.

The building of multi-coordinate control systems based on differential analyzers gives such advantages as:

- Simplification of control algorithm and automatic control system structure;
- Possibility to control the speed of affix movement without considerable complication of structure of automatic control system;

- Possibility of change-over of control system parameters for reproduction of affix movement on different topological manifold without considerable complication of structure of automatic control system;
- Possibility of optimal curves programming, i.e. such control design that provides maximum performance of manufacturing equipment while surface treatment;
- Possibility of universal software development for machines control for surface treatment.

Therefore, the motion program building level is built on the basis of differential analyzers, the block diagram of worked out control system and several examples of generated trajectories are presented on fig. 2.

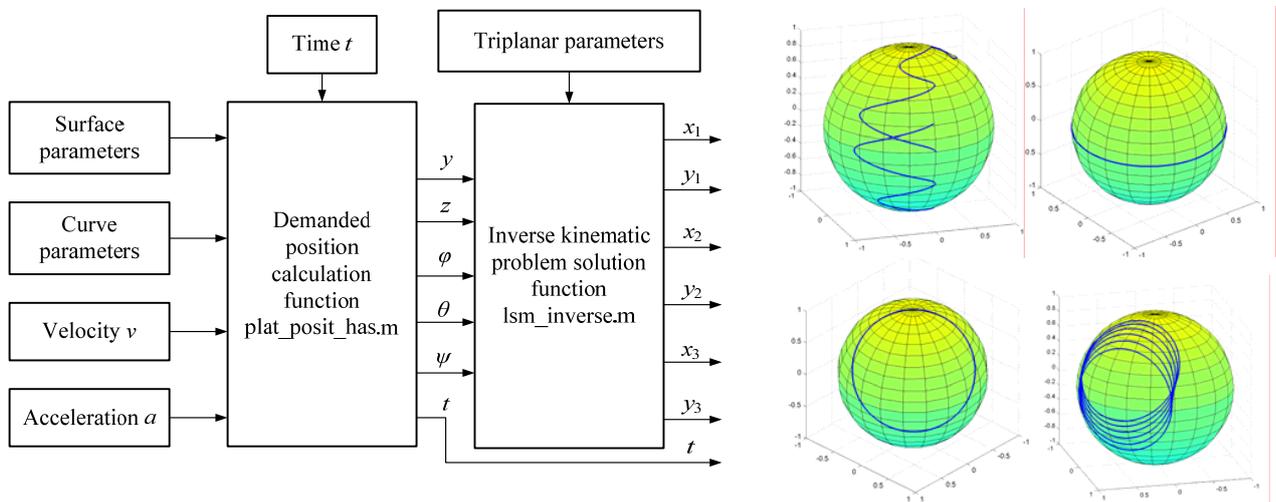


Fig. 2 Motion Program Building Block Diagram and Simulation Results

As it is depicted on fig. 2, the motion program building procedure generates trajectories on quadric surfaces with predefined velocity profile [4].

INVERSE KINEMATIC CALCULATION ALGORITHM

The inverse kinematic problem is based on geometrical Triplanar model presented on fig. 3. The appropriate secant planes CDR_1 , AER_2 and BFR_3 determine the variety of linear stepper motors positions R_1, R_2, R_3 on the stator.

Therefore, the following equations are built for the purposes of inverse kinematic problem solution:

$$A_1x + B_1y + C_1z + D_1 = 0 \quad (4)$$

where $A_1 = x_B - x_A$; $B_1 = y_B - y_A$; $C_1 = z_B - z_A$; $D_1 = -A_1x_D - B_1y_D - C_1z_D$.

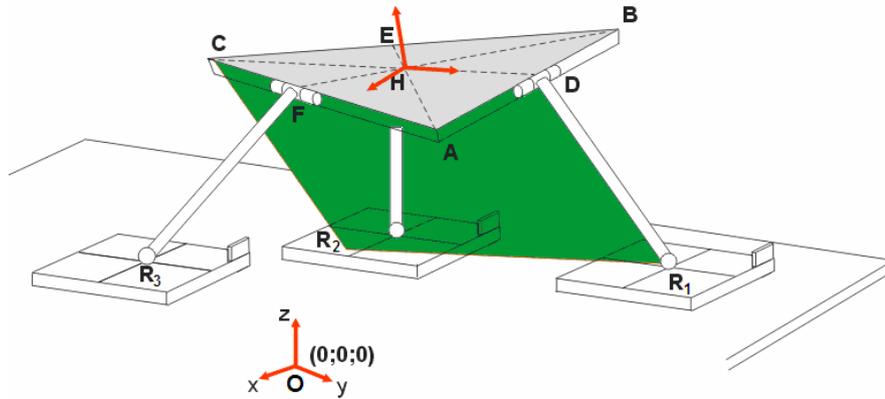


Fig. 3 Geometrical Triplanar Model

Taking into account z-coordinate of LSM, $z=0$ it is possible to build a system of equation which gives the following equations the solution of which is inverse kinematic problem solution:

$$\begin{cases} l = \sqrt{(x_D - x_{R1})^2 + (y_D - y_{R1})^2 + (z_D - z_{R1})^2}, \\ A_1 \cdot x_{R1} + B_1 \cdot y_{R1} + D_1 = 0. \end{cases} \quad (5)$$

$$\begin{cases} l = \sqrt{(x_E - x_{R2})^2 + (y_E - y_{R2})^2 + (z_E - z_{R2})^2}, \\ A_2 \cdot x_{R2} + B_2 \cdot y_{R2} + D_2 = 0. \end{cases} \quad (6)$$

$$\begin{cases} l = \sqrt{(x_F - x_{R3})^2 + (y_F - y_{R3})^2 + (z_F - z_{R3})^2}, \\ A_3 \cdot x_{R3} + B_3 \cdot y_{R3} + D_3 = 0. \end{cases} \quad (7)$$

The worked out inverse kinematic problem solution algorithm enables to implement real-time calculation structure which is based on analytical solution of equations (5...7).

LSM MODEL DESIGN AND VERIFICATION

For the purposes of closed loop control system implementation, the mathematical, computer models of linear stepper motors (LSM) LSM PF-211.HS was worked out and verified. The mathematical model of LSM was constructed of three basic sub models including electrical, electromagnetic and mechanical parts. Every sub model describes the appropriate energy conversion with non-linear transformation laws.

The model is based on description of electromagnetic module of LSM, the cross-section

of which is presented on fig. 4.

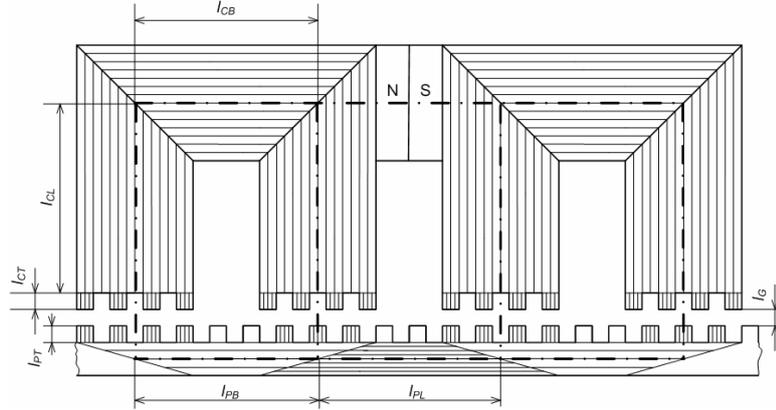


Fig. 4 The Electromagnetic Module Cross-Section

Based on physical electromagnetic module description, the following equation was found for the purposes of LSM PF-211.HS force description:

$$F = R_{G0} \frac{4\pi k_G}{p} \left(\Phi_{23} \Phi_{pm.A} \cos \frac{2\pi x}{\tau} - \Phi_{67} \Phi_{pm.B} \sin \frac{2\pi x}{\tau} \right) \quad (8)$$

The state space mathematical model was build on the basis of the following equations:

$$\begin{aligned} \dot{\Phi}_{23} &= f_1(\Phi_{23}, R_i(x, \Phi_i(i_A(\Phi_{23}, R_j(\dots)), R_k(\dots))), i_A, t) = \tilde{f}_1(\Phi_{23}, x, i_A, t) \\ \dot{\Phi}_{67} &= f_1(\Phi_{67}, R_i(x, \Phi_i(i_A(\Phi_{67}, R_j(\dots)), R_k(\dots))), i_B, t) = \tilde{f}_1(\Phi_{67}, x, i_B, t) \\ \dot{v} &= f_3(F(\Phi_{23}, \Phi_{67}, \Phi_{pm.A}(\Phi_i(i_A(\Phi_{23}, R_j(x, \Phi_j(\dots))), R_k(\dots))), \\ &\quad \Phi_{pm.B}(\Phi_i(i_B(\Phi_{67}, R_j(x, \Phi_j(\dots))), R_k(\dots))), F_d(v), t) = \tilde{f}_3(\Phi_{23}, \Phi_{67}, x, t) \\ \dot{x} &= v = \tilde{f}_4(v, t) \end{aligned} \quad (9)$$

Basing on these equations, one can mention that variables Φ_{23} , Φ_{67} , x and v are state space variables of direct drive. While the number of states are unique for each system, it should be emphasized that the choice of variables that are declared as state variables is rather a matter of convenience [1]. Using these equations, it is possible carry out computer simulation of direct drive. The inputs of the system are power amplifier currents i_A and i_B . The outputs which are to be monitored depends entirely on the application. In case presented x , v and F are output variables.

To verify worked out model, Identification Toolbox of MATLAB was used. This tool helps verify mathematical model with high accuracy. During experiments, the model of motor LSM-211PF.HS was verified. Identification Toolbox needs some input and output data which represent control signals and system response. Basing on these data, the transfer

function of the system is built, and, using one of identification techniques, the model transfer function is constructed. For the case of LSM-PF.211.HS motor verification a hall-effect sensor was used, as this is a standard option of this class of drives.

For the purposes of experimental data acquisition, a test bench was implemented, and direct drive transient process signals were measured and processed. The signals were measured and processed with different acceleration, velocity and travel tasks, to provide as much as possible information for analysis. The measurements were carried out using specified equipment on Ruchservomotor enterprise. Basing on experimental data and data acquired during computer simulation, transfer functions of direct drive were built. The analysis of model adequacy was carried out using ARX and PEM methods [6]. Transient responses of model worked out and real direct drive were acquired and compared. As the result, the built mathematical model is adequate up to 91%.

CONCLUSION

The hierarchical control system of Triplanar is described in the paper. It includes motion program building level based on holonomic automatic systems, inverse kinematic problem algorithm built on the basis of analytical geometry approaches used for procedure simplification, and LSM control system design and verification with the help of MATLAB software set.

References:

- [1] Homepage Mechatronik Laboratorium Paderborn, Universität Paderborn
- [2] Karpovich S., Rusetsky A., Liashuk Y., Matiushkov V. The Optimal Designing of Precision Coordinate Systems and Mechanisms of Electronic Technology Production Equipment.: Minsk, "Integral", 1999 (in Russian)
- [3] Karpovich S., Mezhinsky Y. Design Aspects of Flexible Manufacturing Systems with Linear Stepping Motors // Proceedings of 41th Internationales wissenschaftliches Kolloquium. Ilmenau, Germany, 1996. - P. 166-170.
- [4] Zimmermann K., Ahranovich A., Karpovich S. Analysis and Synthesis of Differential Analyzers for Holonomic Automatic Systems // Scientific Proceedings. - Vol.1. - Aachen: Shaker Verlag, 2004. - p.246-251
- [5] Ahranovich A., Zimmermann K., Karpovich S. Investigating the Ability of Holonomic Automatic Systems` Dynamics Control // Proceedings of 50th International Scientific Colloquium, Ilmenau (Germany), September 19-23, 2005. - TU-Ilmenau, 2005. - p.127-128
- [6] Ljung L. System Identification: Theory for the User. Prentice-Hall, Englewood Cliffs, NJ, 1998. 150P

Authors:

Dipl.-Ing., Post graduate student Aliaksandr Ahranovich
Prof., Dr.-Eng. Habil Svyatoslav Karpovich
Belarusian State University of Informatics and Radioelectronics, P.Brovki Str. 6,
220027 Minsk Belarus
Phone: +375 (17) 293-88-30
E-mail: mmts@bsuir.unibel.by

Uni. prof. Dr.-ing. habil. Klaus Zimmermann
Technical University Ilmenau
D-98684 Ilmenau
E-mail: klaus.zimmermann@tu-ilmenau.de

A. Balkovoy, V. Cacenkin, G. Slivinskaia

Statical and dynamical accuracy of direct drive servo systems

ROBOTICS AND MOTION SYSTEMS

The modern direct drive servo systems based on brushless AC motors (BLACM) have been gaining popularity owing to their high torque to current ratio, high efficiency and robustness. The high stiffness of mechanical coupling and high resolution of digital control and measuring systems, allow the considering of BLACM direct drive servo systems as continuous systems. In these systems, the BLACM may be considered as double integrator. The position closed loop with this actuator may be designed with proportional-integral-differential (PID) controller or with a state-space controller. So, the comparison of various control structures with PID control and state-space control render the interest.

From the example of the drive system with PID controller, it may be shown that direct drive servo system with BLACM has the astaticism of the third order and brings the acceptable transients. The structure of such system is in Fig. 1 represented.

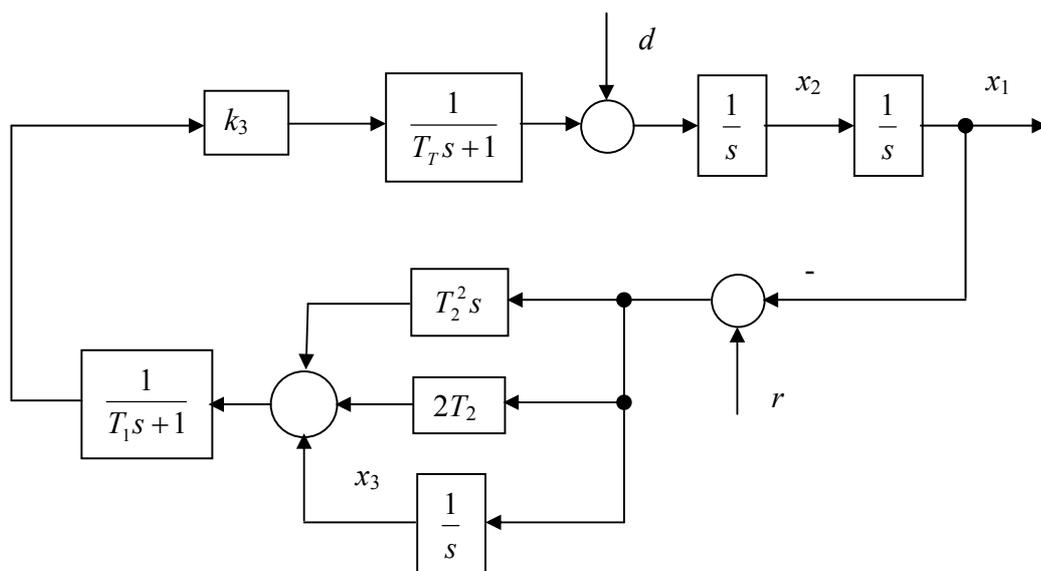


Fig. 1. The drive system with PID controller

The first-order link with time-constant T_T represents the non-ideal conversion of BLACM currents in the current stiff inverter (CSI) into the motor torque and the first-order link with time-constant T_1 represents the lag-effect in digital controller [1]. The transfer function of the "ideal" controller is designed with "critical" damping in numerator that makes the design of controller parameters sufficient easier [2]:

$$G_C(s) = k_3 \left(\frac{1}{s} + 2T_2 + T_2^2 s \right) = \frac{k_3(1 + 2T_2 s + T_2^2 s^2)}{s} \quad (1)$$

The closed loop transfer functions from command and from disturbance of the system in Fig. 1 are:

$$G_{CLr}(s) = \frac{k_3(1 + 2T_2 s + T_2^2 s^2)}{T_1 T_T s^5 + (T_1 + T_T) s^4 + s^3 + k_3 T_2^2 s^2 + 2k_3 T_2 s + k_3} \quad (2)$$

$$G_{CLd}(s) = \frac{s(1 + (T_1 + T_T)s + T_1 T_T s^2)}{T_1 T_T s^5 + (T_1 + T_T) s^4 + s^3 + k_3 T_2^2 s^2 + 2k_3 T_2 s + k_3} \quad (3)$$

The error of the system is:

$$e(s) = (1 - G_{CLr}(s))r(s) - G_{CLd}(s)d(s). \quad (4)$$

Now, the errors coefficients from command and from disturbance may be calculated [1]:

$$C_r(k) = \frac{d^k}{ds^k} (1 - G_{CLr}(s))|_{s=0}; C_d(k) = -\frac{d^k}{ds^k} G_{CLd}(s)|_{s=0}; k = 0, 1, 2, \dots \quad (5)$$

As a result, the errors coefficients are:

$$C_r(0) = 0; C_r(1) = 0; C_r(2) = 0; C_r(3) = \frac{6}{k_3}; C_d(0) = 0; C_d(1) = -\frac{1}{k_3}. \quad (6)$$

So, in the servo system with BLACM and PID-controller, the astaticism from command of the third order may be achieved.

The controller parameters are selected with "critical" damping in numerator. In this case, the time constant T_2 of the controller may be calculated from the index of oscillation and base frequency $\omega_0 = \sqrt[3]{k_3}$:

$$T_2 = \frac{2}{\omega_0} \sqrt{\frac{M^2 - M\sqrt{M^2 - 1}}{M^2 - 1}}. \quad (7)$$

Usually, index of oscillation is $M=1, 2 \dots 1, 5$ in dependence of stability of system parameters [3].

The simplest structure of system with state-space control is in Fig. 2 represented. The closed loop transfer functions of the system in Fig. 2 are:

$$G_{CLr}(s) = \frac{x_1(s)}{r(s)} = \frac{l_1}{s^2 + l_2 s + l_1}; G_{CLd}(s) = \frac{x_1(s)}{r(s)} = \frac{1}{s^2 + l_2 s + l_1}; \quad (8)$$

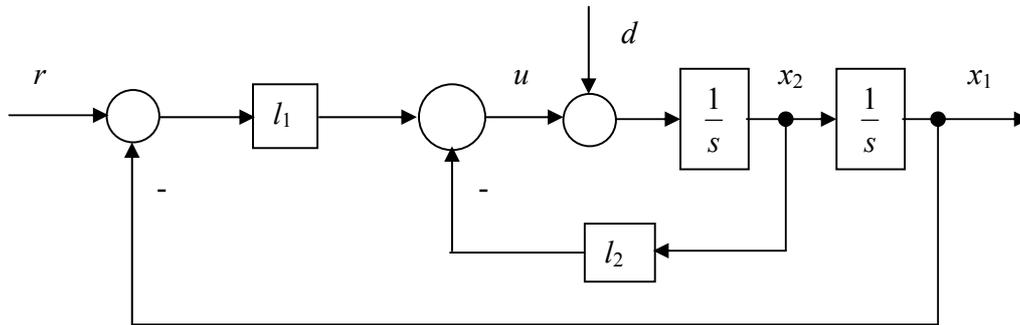


Fig. 2. The drive system with state-space controller

the error of the system is:

$$e_r(s) = \frac{s^2 + l_2s}{s^2 + l_2s + l_1}. \quad (9)$$

In accordance with (5), in the servo system Fig 2, the astaticism from command of the first order may be achieved:

$$C_r(0) = 0; C_r(1) = \frac{1}{l_1 l_2}; C_d(0) = -\frac{1}{l_1}. \quad (10)$$

The further development of the system with state-space control is the structure where the full-order observer is added (Fig. 3).

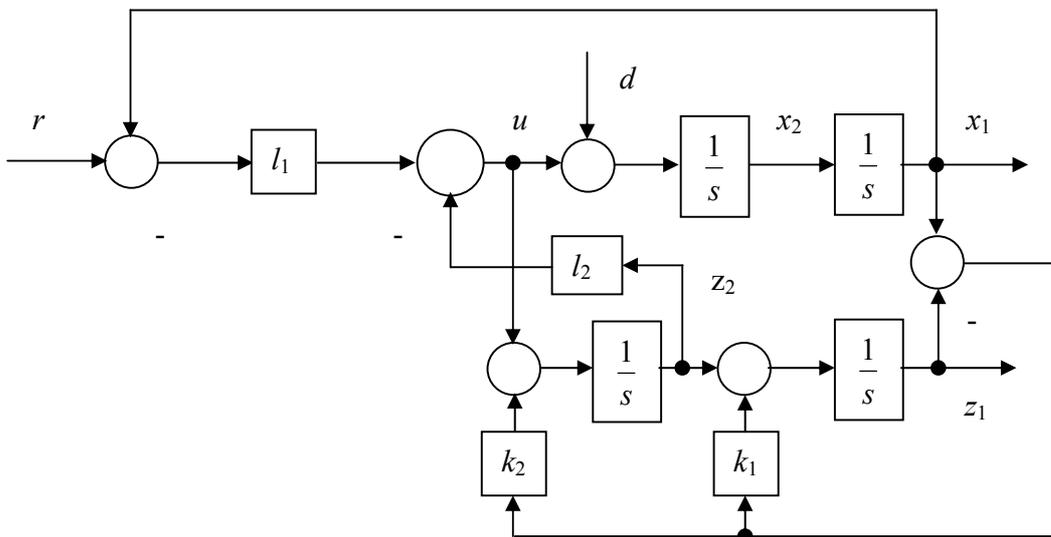


Fig. 3. The drive system with full-order observer

For the state-space control, the velocity from observer is used. The closed loop transfer functions of the system in Fig. 3 are:

$$G_{CLr}(s) = \frac{(s^2 + k_1s + k_2)l_1}{s^4 + (k_1 + l_2)s^3 + (k_1l_2 + k_2 + l_1)s^2 + (k_1l_1 + k_2l_2)s + k_2l_1};$$

$$G_{CLd}(s) = \frac{s^2 + (k_1 + l_2)s + k_1l_2 + k_2}{s^4 + (k_1 + l_2)s^3 + (k_1l_2 + k_2 + l_1)s^2 + (k_1l_1 + k_2l_2)s + k_2l_1}.$$
(11)

Using (4) and (5), the errors coefficients may be calculated:

$$C_r(0) = 0; C_r(1) = \frac{1}{l_1l_2}; C_d(0) = -\frac{1}{l_1} - \frac{k_1l_2}{k_2l_1}.$$
(12)

So, the steady-state errors of the system in Fig 3 are approximately the same as in system from Fig. 2.

It is expected that similar to PID controller, the definite advantages may be obtained from the system with state-space controller and state-space errors observer. The structure of such a system is in Fig. 4 depicted. The state-space errors are from the outputs of observer acquired.

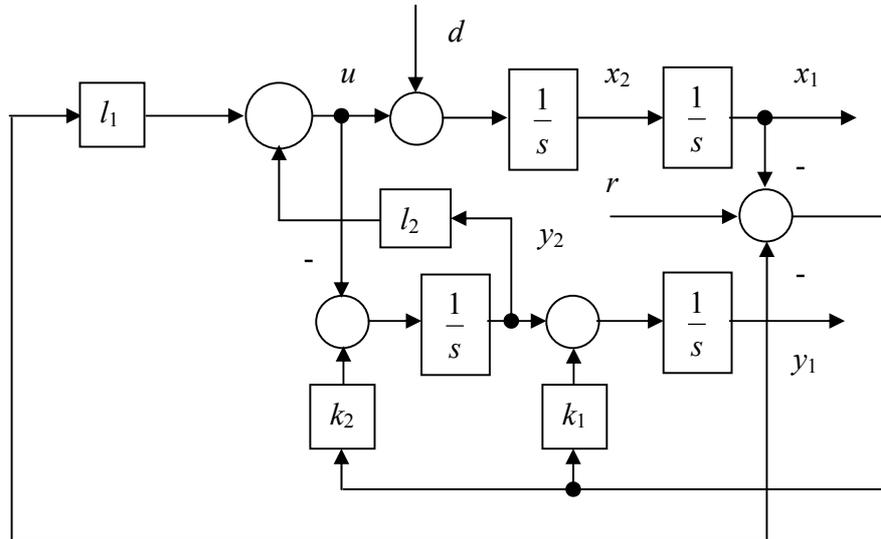


Fig. 4. The system with state-space controller and state-space errors observer

The closed loop transfer functions of the system in Fig. 4 are:

$$G_{CLr}(s) = \frac{(k_1l_1 + k_2l_2)s + k_2l_1}{s^4 + (k_1 + l_2)s^3 + (k_1l_2 + k_2 + l_1)s^2 + (k_1l_1 + k_2l_2)s + k_2l_1};$$

$$G_{CLd}(s) = \frac{s^2 + (k_1 + l_2)s + k_1l_2 + k_2 + l_1}{s^4 + (k_1 + l_2)s^3 + (k_1l_2 + k_2 + l_1)s^2 + (k_1l_1 + k_2l_2)s + k_2l_1}.$$
(13)

The errors coefficients are:

$$C_r(0) = 0; C_r(1) = 0; C_r(2) = \frac{2}{k_2} + \frac{2(k_2 + k_1l_2)}{k_2l_1}; C_d(0) = -\frac{1}{k_2} - \frac{k_2 + k_1l_2}{k_2l_1}.$$
(14)

So, in the servo system from Fig 4, the astaticism from command is the second order but the steady state errors due to acceleration or load are remaining.

The further development of the system properties is the compensation of the constant disturbance. The simplest model of the constant disturbance is the "slow" disturbance in compare to system and observer transients. The model of such "constant" value is the output of integrator with zero random initial condition [4]:

$$\begin{aligned} \frac{dd}{dt} &= 0; \\ d(0) &= d_0, \end{aligned} \tag{15}$$

where d is the disturbance; d_0 is the scalar random disturbance.

The equations for system variables are:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} u; \quad x_1 = C = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \tag{16}$$

In equations (15), the controlled states are x_1 and x_2 . The disturbance x_3 is uncontrolled. All the parameters are observable. So, the closed loop controller and observer may be designed. The structure of the system with state-space controller, state-space errors observer and disturbance model is in Fig. 5 depicted.

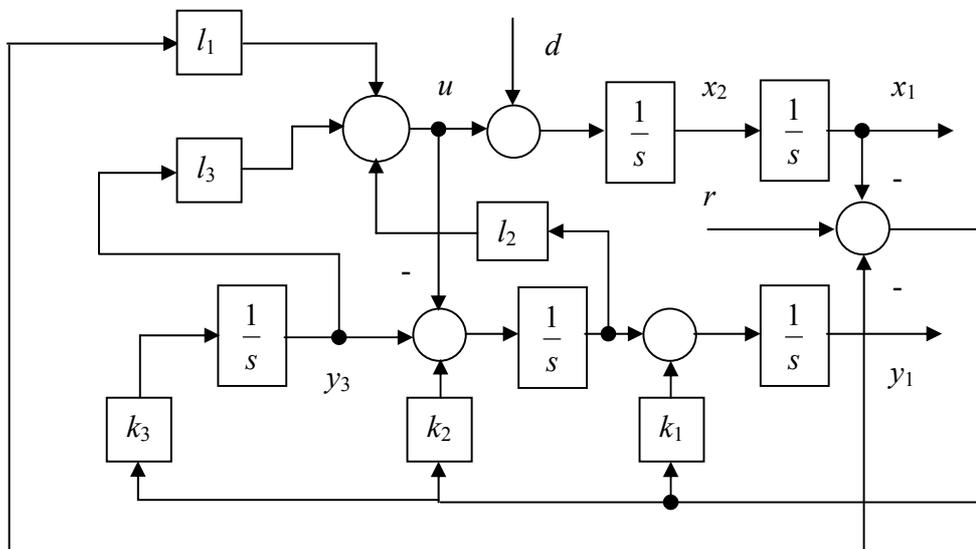


Fig. 5. The system with state-space controller, state-space errors observer and disturbance model

The disturbance compensation means $l_3 = 1$. The closed loop transfer functions of the system in Fig. 5 are:

$$\begin{aligned}
G_{CLr}(s) &= \\
&= \frac{(k_1 l_1 + k_2 l_2 + k_3) s^2 + (k_2 l_1 + k_3 l_2) s + k_3 l_1}{s^5 + (k_1 + l_2) s^4 + (k_1 l_2 + k_2 + l_1) s^3 + (k_1 l_1 + k_2 l_2 + k_3) s^2 + (k_2 l_1 + k_3 l_2) s + k_3 l_1}; \\
G_{CLd}(s) &= \\
&= \frac{s(s^2 + (k_1 + l_2) s + k_1 l_2 + k_2 + l_1)}{s^5 + (k_1 + l_2) s^4 + (k_1 l_2 + k_2 + l_1) s^3 + (k_1 l_1 + k_2 l_2 + k_3) s^2 + (k_2 l_1 + k_3 l_2) s + k_3 l_1}.
\end{aligned} \tag{17}$$

The errors coefficients are:

$$\begin{aligned}
C_r(0) = 0; C_r(1) = 0; C_r(2) = 0; C_r(3) &= \frac{6}{k_3} + \frac{6(k_2 + k_1 l_2)}{k_3 l_1}; \\
C_d(0) = 0; C_d(1) &= -\frac{1}{k_3} - \frac{k_2 + k_1 l_2}{k_3 l_1}.
\end{aligned} \tag{18}$$

The astaticism of the system has third order from reference and first order from disturbance.

The analysis of steady-state errors has shown the advantage of two structures: with PID controller and with state-space controller, state-space errors observer and disturbance model. The PID controller needs the digital differentiation that usually has noise due to discretisation. Also, the precision of the system depends from sampling period. In the drive system with observer, there is no differentiation and the precision of integration has weak dependence from sampling period. But the sampling period has to be small because of observer has to be "faster" as observed system.

The another criterion of the servo system selection is the quality of transients. The parameters of system with PID controller may be first evaluated using (7) for selected base frequency and index of oscillation. In the servo system with observer, the synthesis of controller parameters may be done independent.

For the structure in Fig. 5, the simple method of controller parameters selection is the use of modal control method [4]. The characteristic polynomial of controller is:

$$D_r(s) = s(s^2 + l_2 s + l_1) = 0. \tag{19}$$

Here, the zero root is the result of non-controlled disturbance x_3 . Let the desirable characteristic polynomial of the plant as:

$$s(s^2 + 2\zeta_p \omega_p s + \omega_p^2). \tag{20}$$

From (19), (20) and the gain $l_3 = 1$ (selected to give perfect disturbance cancellation) following the equations:

$$l_1 = \omega_p^2; l_2 = 2\zeta_p \omega_p. \tag{21}$$

The characteristic polynomial of observer is:

$$D_{ob}(s) = s^3 + k_1 s^2 + k_2 s + k_3, \quad (22)$$

The method of observer parameters selection is again the use of modal control method. The desired characteristic polynomial is:

$$(s + a_o)(s^2 + 2\zeta_o \omega_o s + \omega_o^2) = s^3 + (2\zeta_o \omega_o + a_o)s^2 + (2\zeta_o \omega_o a_o + \omega_o^2)s + \omega_o^2 a_o. \quad (23)$$

The coefficients of observer are calculated as:

$$k_1 = 2\zeta_o \omega_o + a_o; k_2 = 2\zeta_o \omega_o a_o + \omega_o^2; k_3 = \omega_o^2 a_o. \quad (24)$$

The frequencies ω_p , ω_o depend of bandwidth of drive system and observer. The self-oscillation frequency of observer ω_o has to be higher as self-oscillation frequency of plant ω_p . The coefficients ζ_p , ζ_o determine the oscillations of transients. The coefficient a_o assign the speed of transients.

The further comparison of statical and dynamical errors of direct drive servo systems with BLACM demonstrates the advantages of system with PID controller and system with error state space observer. The PID controller is more sensitive to sampling frequency and error state space observer inserts the additional lag effect. The analytical research and simulation of the servo systems dynamical errors show the negative action of "small" time constants of PID-control and relative slow dynamics of system observer on servo system transients. For dynamics improvement of such systems, the "small" time constants of PID-control have to be reduced and the processing speed of the error state observer has to be enhanced. Some simulations and real systems dynamics validating these resume. The further research of control structures and controller and observer parameters has to be done in accordance with obtained here conclusions. The final choice has to be done after research of real systems.

References:

- [1] Цыпкин Я.З. Основы теории автоматических систем. - М.: "Наука", 1977.
- [2] Бесекерский В.А., Попов Е.П. Теория систем автоматического регулирования. – М.: Наука, 1972.
- [3] Иващенко Н.Н. Автоматическое регулирование. Теория и элементы систем. Учебник для вузов. Изд. 4-е, перераб. и доп. М.: Машиностроение, 1978.
- [4] Квакернаак Х., Сиван Р. Линейные оптимальные системы управления. – М.: Мир, 1977.

Authors:

PhD Alexander Balkovoy
 PhD Victor Cacenkin
 Dipl. Math. Galina Slivinskaia
 Department of Electric Drive, Moscow Power Engineering Institute (TU), Krasnokazarmennaya 14,
 111250, Moscow, Russia
 Phone: (495) 673-0285
 Fax: (495) 673-1348
 E-mail: balk@aep.mpei.ac.ru

Y. Litvinov / S. Karpovich / A. Ahranovich

The 6-DOF Spatial Parallel Mechanism Control System Computer Simulation

ABSTRACT

The 6-DOF spatial parallel mechanisms are frequently used in industrial robots, motion systems, flight training, test systems, etc. The 6-DOF spatial parallel mechanisms enable development of controllable dynamical electro-mechanical systems, by providing sophisticated 3D movements. However, such systems' control system simulation problem still remains one of the main difficulties in the sphere of robotics [1].

The described 6-DOF spatial parallel mechanism control system simulation is intended for using in motion system construction for working platform 3D movement.

THE 6-DOF SPATIAL PARALLEL MECHANISM

A 6-DOF (degree-of-freedom) spatial parallel mechanism is composed of 6 independent legs connecting the mobile platform with the base. Each of these legs is a serial kinematic chain that is controlled by one motor which actuates one of the joints. The structure chart of the investigated spatial parallel mechanism, presented on Fig. 1, is capable of realizing sophisticated 3D movements of mobile platform. The $A_1A_2A_3A_4A_5A_6$ mobile platform is directly connected to O_1, O_2, \dots, O_6 motors driving shafts by $l_1r_1, l_2r_2, \dots, l_6r_6$ kinematic joints. The 6-DOF spatial parallel mechanism supports three coordinate axial displacements (x, y, z) and three angular rotations (ψ – yaw, θ – pitch, φ – roll) of the mobile platform about the appropriate coordinate axes respectively.

THE INVERSE KINEMATIC PROBLEM SOLUTION

The output position and orientation of the platform directly correspond to the input actuation from motors driving shafts. Therefore, the 6-DOF spatial parallel mechanism

control system modeling often starts with the inverse kinematic problem (IKP) solution, that implies the determination of the input variables (motors shafts rotation angles $\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6$) out of the output variables (platform position x, y, z and orientation ψ, θ, φ) [2].

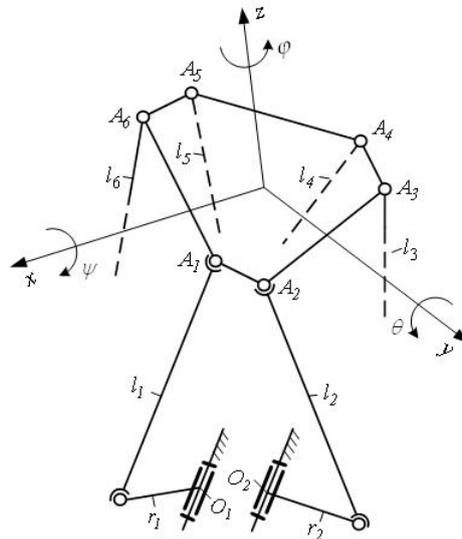


Fig. 1. The 6-DOF spatial parallel mechanism structure chart

The input variables are connected with the output variables by the appropriate equations. The initial equations analysis showed that the inverse kinematic problem doesn't always have general solution. The reason is, that for a given set of output variables there are several valid sets of input variables. A new developed program with implementation in MATLAB/Simulink environment software enables inverse kinematic problem solution for mechanism type proposed on Fig. 1. The inverse kinematic problem computing program screenshot is presented on Fig. 2.

The program provides both immediate and step-by-step solution of the inverse kinematic problem for the 6-DOF spatial parallel mechanism. The step-by-step mode enables consecutive computing of the inverse kinematic problem for specified incremental changes of $x, y, z, \psi, \theta, \varphi$ variables on every step. Moreover, the platform's movement boundary region can be build on the basis of the inverse kinematic problem computing algorithm. For instance, the platform's movement boundary regions for different ψ, θ, φ constant angular values and x, y, z variables incremental changes are presented on Fig. 3.

As it can be seen on Fig. 3 the platform's movement boundary region volume has a nonlinear dependence of angular values ψ, θ, φ .

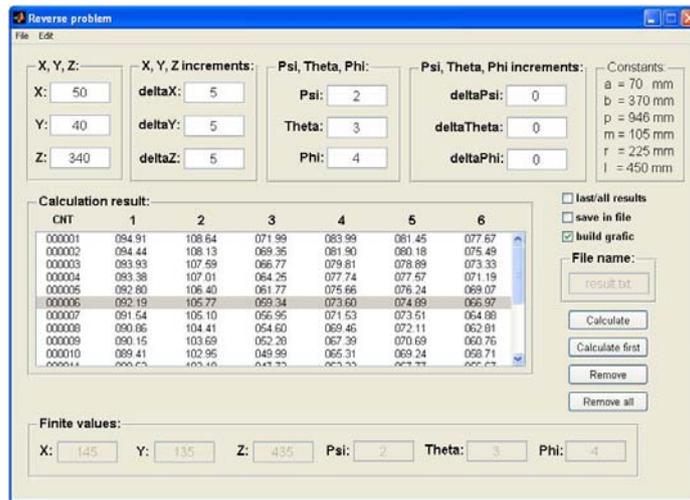


Fig. 2. The inverse kinematic problem computing program

The experiment carries out resulted in boundary region with maximum volume determination with zero angular values ψ, θ, φ .

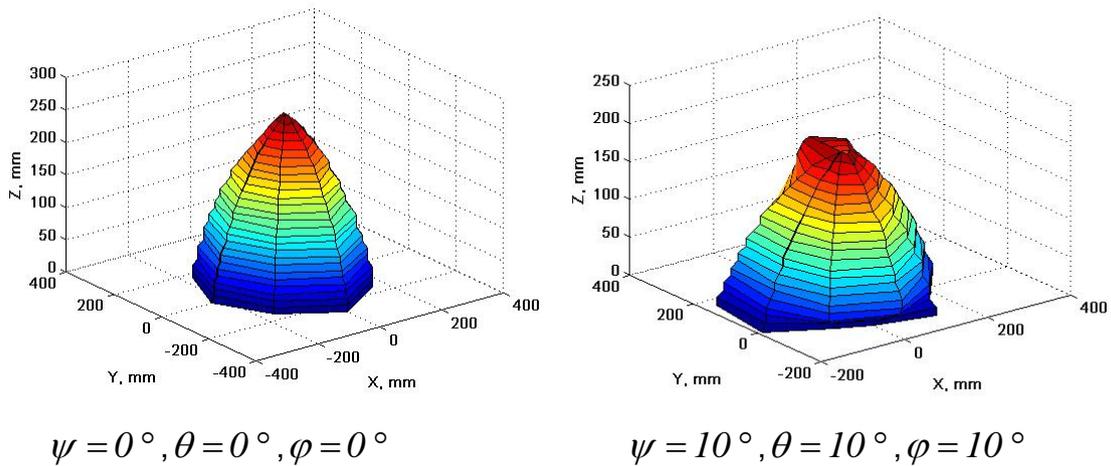


Fig. 3. The platform's movement boundary regions

In order to build proper 6-DOF spatial parallel mechanism control system simulation the six control motors shafts rotation angles changes should be investigated which directly correspond to the determined platform movement. The inverse kinematic problem computing program provides possibility for building motors shafts rotation angles changes trajectories according to platform's movement from initial to final position and orientation (Fig. 4). The motors shafts rotation angles trajectories are obtained by the inverse kinematic problem solution for specified incremental changes of $x, y, z, \psi, \theta, \varphi$ variables on every step. On Fig. 4 can be seen that the rotation angles has nonlinear variation trajectories and different total angular changes.

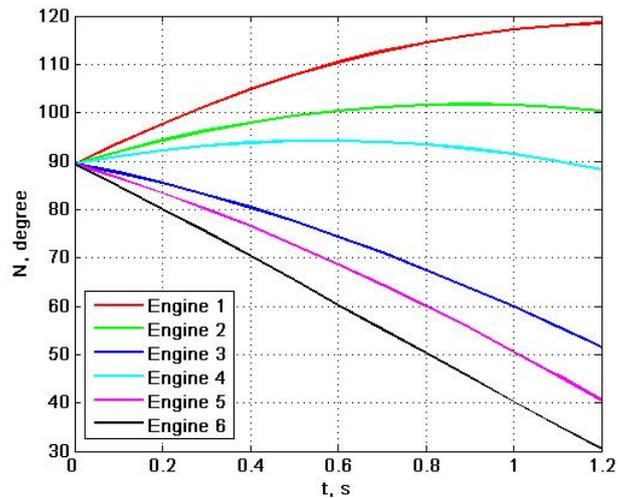


Fig. 4. The motors shafts rotation angles changes trajectories

This means that all six motors to actuate the 6-DOF spatial parallel mechanism must be controlled independently, bringing the demand of six regulators implementation.

THE 6-DOF SPATIAL PARALLEL MECHANISM CONTROL SYSTEM COMPUTER SIMULATION

The 6-DOF spatial parallel mechanism control system consists of the inverse kinematic problem computing module (IKP module), regulators (one per motor) and hardware components to control six motors. However, regulators synthesis and, therefore, control signals tuning are among the main problems that deal with working out such a control system. The rapid solution is to reconstruct simulation loop feedback system in MATLAB/Simulink environment, composed of the control system model and motor model, to carry out simulation (Fig. 5).

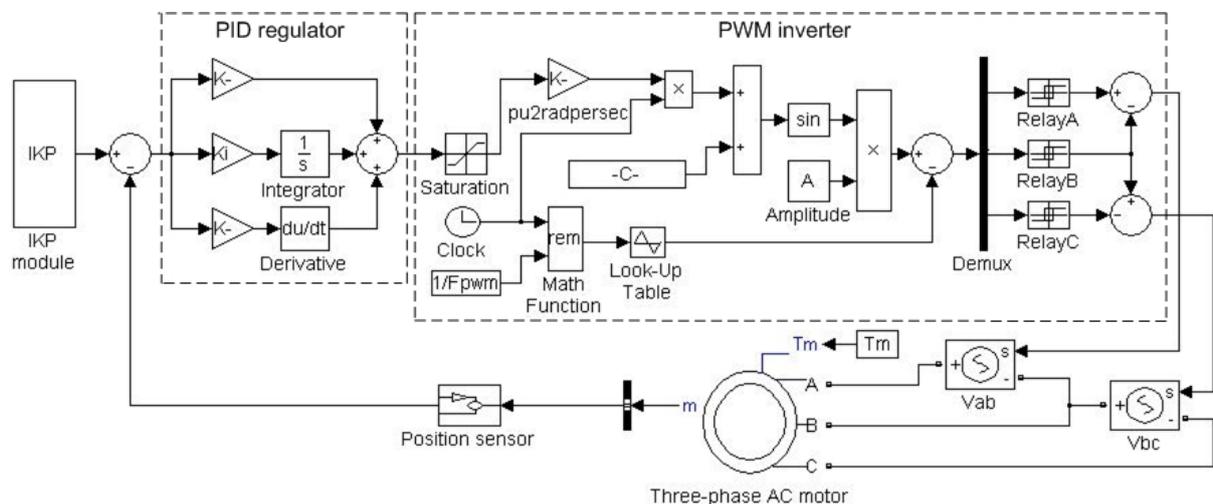


Fig. 5. Simulation system block diagram for one three-phase AC motor

The 6-DOF spatial parallel mechanism control system modeling in MATLAB/Simulink environment starts with such simulation loop feedback system reconstruction. Computer simulation was carried out with usage of PID regulators and three-phase AC motors. The simulation block diagram of one three-phase AC motor control system is presented on Fig. 5. As it can be seen on Fig. 5, AC motor is fed by PWM signals produced by PWM inverter. The three-phase AC motor simulation model is implemented on the basis of the following equation sets [3]:

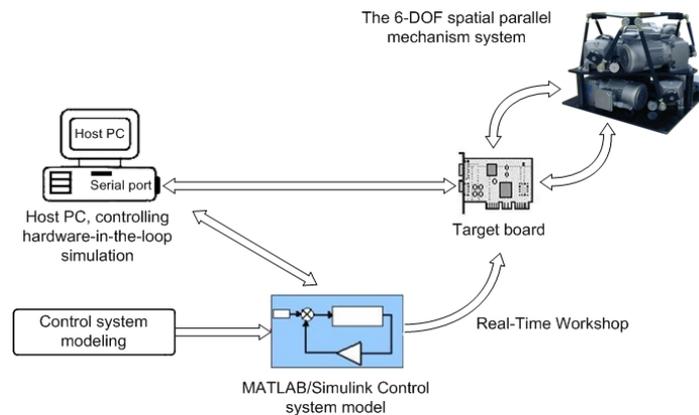
$$\begin{cases} V_{qs} = R_s i_{qs} + \frac{d}{dt}(L_s i_{qs} + L_m i_{qr}) + \omega(L_s i_{ds} + L_m i_{dr}); \\ V_{ds} = R_s i_{ds} + \frac{d}{dt}(L_s i_{ds} + L_m i_{dr}) - \omega(L_s i_{qs} + L_m i_{qr}); \\ V_{qr} = R_r i_{qr} + \frac{d}{dt}(L_r i_{qr} + L_m i_{qs}) + (\omega - \omega_r)(L_r i_{dr} + L_m i_{ds}); \\ V_{dr} = R_r i_{dr} + \frac{d}{dt}(L_r i_{dr} + L_m i_{ds}) - (\omega - \omega_r)(L_r i_{qr} + L_m i_{qs}); \\ T_e = p(i_{qs}(L_s i_{ds} + L_m i_{dr}) - i_{ds}(L_s i_{qs} + L_m i_{qr})). \end{cases} \quad (1)$$

$$\begin{cases} \frac{d}{dt} \omega_m = \frac{1}{2H}(T_e - F \omega_m - T_m); \\ \frac{d}{dt} \theta_m = \omega_m. \end{cases} \quad (2)$$

In accordance with the (1,2) description, the simulation can be carried out with different three-phase AC motors depending on electrical and mechanical parameters which can be set directly via new developed interface. The developed simulation system, implemented in MATLAB/Simulink, enables PID regulator tuning according to motor output characteristics (rotor speed, rotor rotation angle, electromagnetic torque, rotor and stator currents).

The 6-DOF spatial parallel mechanism control system computer simulation is implemented as the hardware-in-the-loop control model simulation in MATLAB/Simulink modeling environment. MATLAB/Simulink in conjunction with Real-Time Workshop can automatically generate, package, and compile source code from Simulink models to create real-time software applications that can immediately be executed on a variety of systems and hardware platforms thus enabling the 6-DOF spatial parallel mechanism control system hardware-in-the-loop simulation as well. The 6-DOF spatial parallel mechanism control system hardware-in-the-loop simulation structure is presented on Fig. 6. As the result of hardware-in-the loop simulation (Fig. 6), the 6-DOF spatial parallel mechanism control system hardware-in-the-loop simulation model has been

developed in MATLAB/Simulink for simulation on TI TMS320C2000 DSP debug hardware platform (target board).



The 6-DOF spatial parallel mechanism control system hardware-in-the-loop simulation block diagram

The developed hardware-in-the-loop simulation model enables the execution of the MATLAB/Simulink control system model directly on physical hardware debug platform, giving that way the possibility of rapid control prototyping for the 6-DOF spatial parallel mechanism control system.

CONCLUSION

Control system simulation for 6-DOF spatial parallel mechanism is described in the paper. It covers mathematical and computer models of inverse kinematic problem, AC electrical motor regulator development and simulation with controller code generation using rapid control prototyping approach.

References:

- [1] Karpovich S., Rusetsky A., Liashuk Y., Matiushkov V. The Optimal Designing of Precision Coordinate Systems and Mechanisms of Electronic Technology Production Equipment.: Minsk, "Integral", 1999 (in Russian)
- [2] Lung-Wen Tsai. Robot Analysis: The mechanics of serial and parallel manipulators.– New York: John Wiley&Sons Inc., 1999. – 505C
- [3] www.mathworks.com

Authors:

Student Yegor Litvinov
 Prof., Dr.-Eng. Habil Svyatoslav Karpovich
 Dipl. –Ing. Aliaksandr Ahranovich
 Department of Mathematics
 Laboratory of Mathematical Modeling of Technical Systems
 Belarusian State University of Informatics and Radioelectronics
 P.Brovki Str. 6, Minsk 220027 Belarus
 Phone: +375 (17) 239-88-30
 Fax: +375 (17) 231-09-14
 E-mail: mmts@bsuir.unibel.by

V. Lysenko/ W. Minchenya/ K. Zimmermann

Minimization of the number of actuators in legged robots using biological objects

Bionically Inspired Robotics Biomechanics

Abstract

A new approach of special problem-solving methods at the initial design stages is presented. The methods are based on the analysis and the combination of technical or biological objects and a legged robot. Described techniques allow us to create several new legged robots. A new class of micro robots and a new class of legged mechanisms is chosen to present the possibilities of the method. Merging the kinematics of a salamander with the kinematics of an octoped allows us to develop a new eight legged robot with only three actuators. Combining a flying insect and a piezotransducer with extremities supplies a new object - the piezomicrorobot. For movement of multi-legged robot through a pipe we use the travelling wave of the Holothouria.

Introduction

Biological objects as prototypes are used preferably due to the fact that during millions of years of evolution their principles of motion have been developed contemplating minimal energy wasting. [1]

The essential design stage, which is discovering ideas for new functional principles of technical systems, is almost entirely based on the know-how of the engineer [2, 3].

The subject of our work is the development of new functional principles of legged robots.

By using the working principle and the kinematics of biological prototypes it is possible to develop new ideas for mobile robots. Some biological objects use unusual ways of moving of the extremities to obtain the necessary trajectory. They change form and sizes of the body to create the necessary movement of legs.

By applying the introduced method new robots can be created. It is based on the

combination of biological and technical objects. The developed method is based on the well-known principle known as the combination of alternative systems. It enables the transfer of characteristics and structure from one object (i.e. its kinematics) to another object leading to new desirable characteristics or to the optimization of existing technical objects[4].

Biologically inspired robots

A multi-legged mobile systems classification is represented (Fig.1.). In our opinion, there exist only 4 - 5 main principles of functioning of biological objects for providing the necessary trajectory of the legs movement. The suggested classification and the analysis of biological prototypes have allowed us to create some new mobile robots. In known walking machines the several actuators for moving each leg are used. Our design principles allow us to use each actuator for moving several legs. Thus, we managed to minimize number of actuators at the robot. It opens new possibilities of the considerable miniaturization of mobile robots in future.

Thus, the ability to develop new functional principles of legged robots (i.e. new motion principles, new kinematics etc.) is provided. The analysis is used to realize the transition from known (in biological objects) to new (for legged robots) forms of motion.

Minimization of number of actuators multi-legged robots can be reached through:

- use of periodical changing the shape of the body of the robot in horizontal dimension (salamander, lizard)
- use of periodical changing the size of the body of the robot in vertical dimension (flying insects)
- use of anisotropy of friction (snake)
- use of periodical character or feature of trawling wave (holothouria)
- use of multidimensional resonance swinging of elastic extremities (mosquito)
- reducing of number of bearing legs (kangaroo, basilisk, birds).

For micro robots it is possible to use a principle of movement as at Polichetae (Fig.2). In this biologic object the legs have no actuators and no degrees of freedom relative to a body. They are rigidly attached perpendicularly to a surface of a body, so they move and incline together with deformation of this surface. To create necessary trajectory of a distal end of a leg, Polichetae and Holothouria uses deformation of the case as trawling wave.

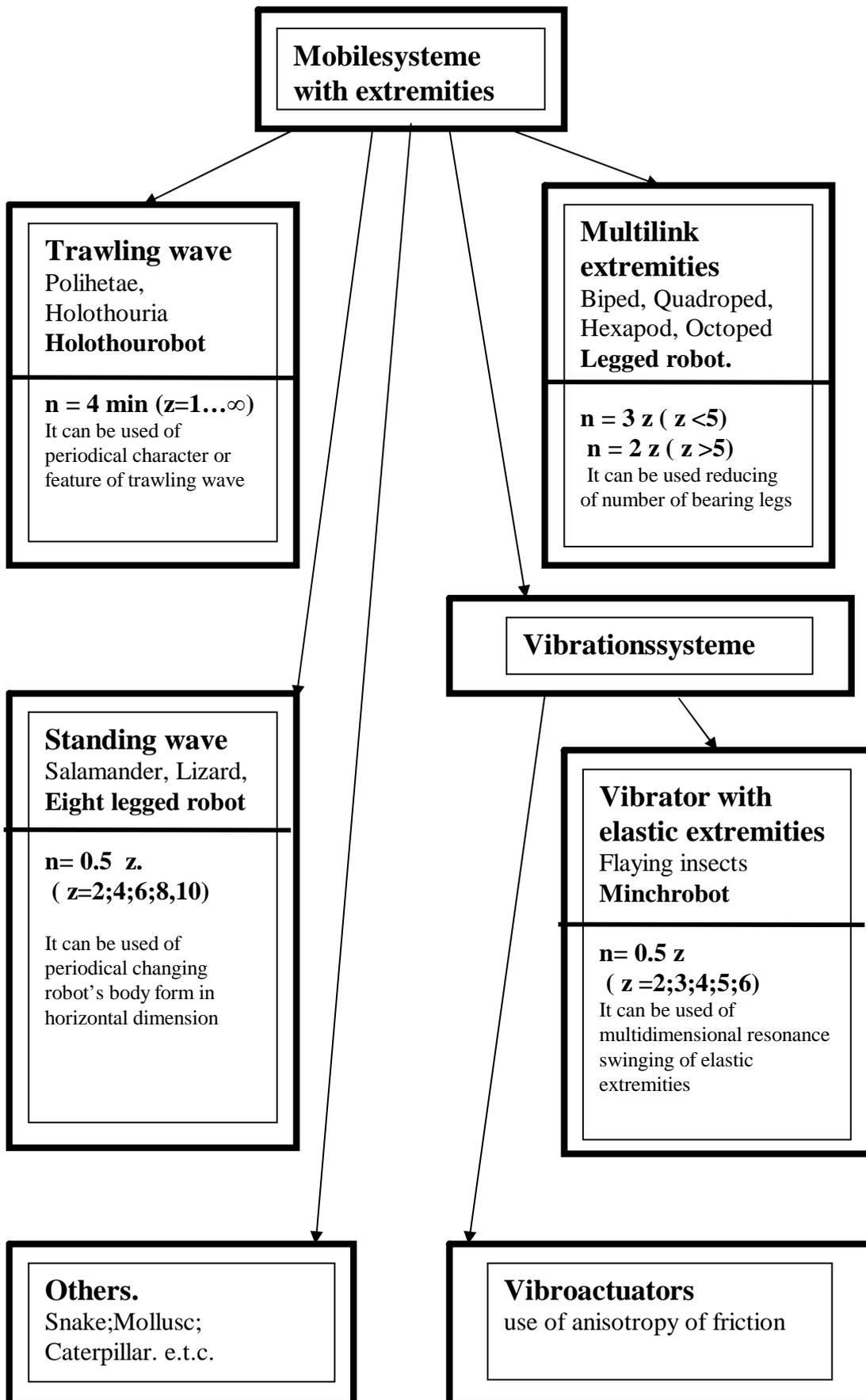


Fig.1 Multy-legged mobile systems classification
(n- number of actuators, Z- number of legs.)

The number of legs-needles is not limited, but number of the actuators enabling deformation of the case, is minimal. It is possible to create tiny robot with a plenty of legs and with low number of small-sized actuators.

We have developed the moving robot-probe with 100 legs and with only four actuators - "Holothourobot" (Fig.2). It can be used in medicine for minimally invasive surgery.

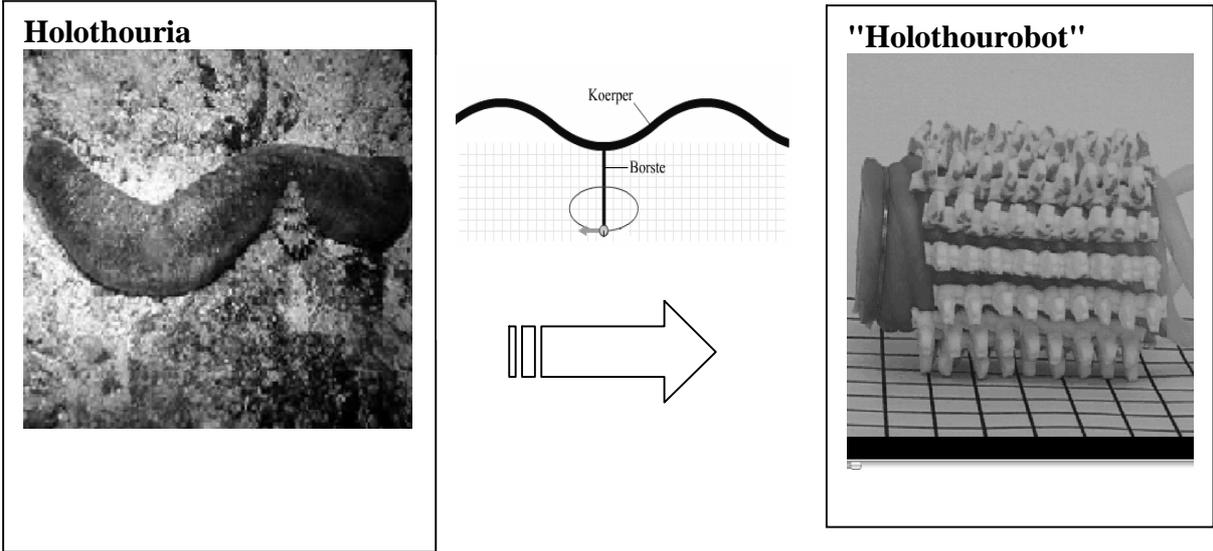


Fig.2. Using of periodical character or feature of trawling wave

The salamander bends its body in a horizontal plane and due to this, the body moves relative to the points of support (Fig.3). By using the deformation of a robot body in a horizontal plane it is possible to provide it's motion due to a minimum number of actuators. The actuators are not connected to legs and they are necessary only for the deformation of the robot body. Having as few as three actuators it is possible to realize the motion of the robot with eight legs "Eightleggedrobot" (Fig.3).

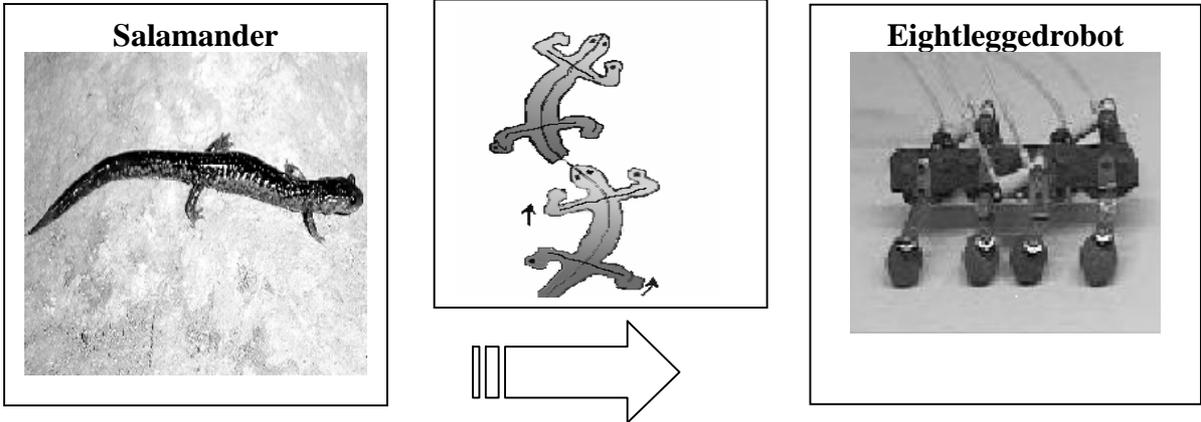


Fig.3. Motion due to periodical changing of the shape of the body in the horizontal plane

Some flying insects create resonant oscillations of the wings due to periodic change of the form and the sizes of the rigid body (Fig.4). The insects' muscles are connected not to the wings, but to the walls of a rigid body and deform it. The deformation of the body turns into swinging of wings.

It is possible to create the mobile robot at which the body vibrates, and legs have no actuators. The necessary trajectory of the distal part of a leg is formed due to excitation of the high-frequency swinging in proximal part of an elastic curvilinear leg and due to mechanical transformation of these swinging in low-frequency.

We developed an essentially new type of a moving system "Minchrobot" (Fig.4). As the body and as the actuator a piezo-bimorph-plate is used. It can cover 1 meter per 1 second.

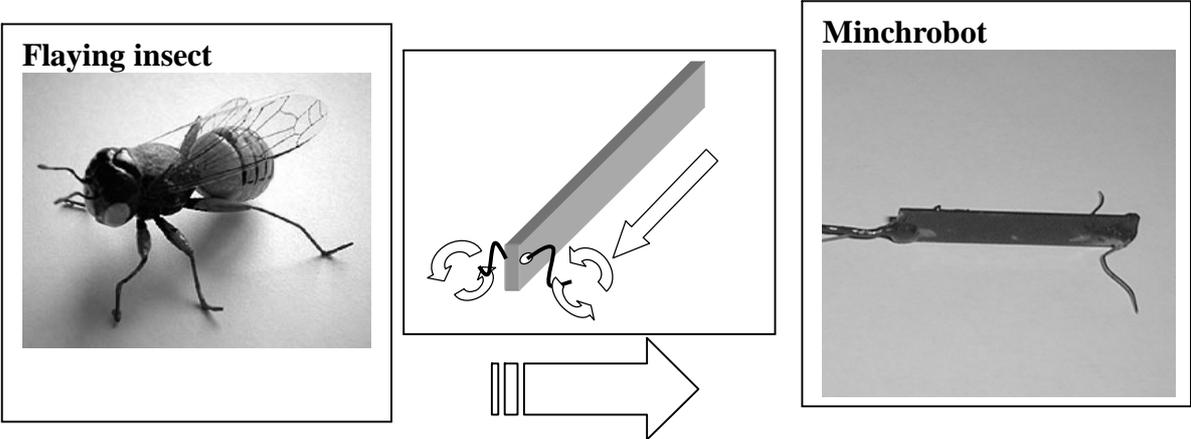


Fig.4. Motion by multidimensional resonance swinging of elastic extremities

Summary

The analyses of biological objects and alternative technical systems offer new opportunities for the engineers. That analysis is an indistinct provisional approach of solving a technical problem. In further stages of the design process the engineer formulates precisely this solution and verifies it by means of mathematical modeling and calculation.

The described technique does not supply convertible constructive drawings

immediately. However, it provides new solutions with new ideas. Furthermore, it is possible to develop essential new legged robots with minimal number of actuators.

References:

- [1] Zimmermann, K.: An approach to the modelling of biological and technical movement systems. 1. International Conference on Motion Systems, Univ. Jena, 1997
- [2] Lysenko, V.: Algorithmische Methode für die Entscheidung auf Anfangsstufen beim Entwerfen. 41. Internationales Wissenschaftliches Kolloquium. 1996. TU Ilmenau.
- [3] Lysenko, V.: Method for improving actuators by modelling the motion of an earthworm. 1. International Conference on Motion Systems, Univ. Jena, 1997
- [4] Lysenko, V.; Zimmermann, K.: New procedure for designing optimized technical systems with use of biological objects. 1st Intern. Conf. on Design & Nature, Udine, Sept. 2002, WIT Press Southampton 2002, pp.115-122

Authors:

Doz. Dr. Victor Lysenko
Doz. Dr. Wladimir Mintchenia
Belorussische Nationale Technische Universität Minsk
Kaf. SMIS, Prosp. Nesawisimosty 65. korp.17
220027 Minsk, Belarus,
Tel. ++375172311120
E-mail: victor_lysenko@mail.ru

Prof. Dr.-Ing. habil. Klaus Zimmermann
Technische Universität Ilmenau
Fakultät für Maschinenbau
PF 100565
098684 Ilmenau
Tel. ++49 3677 692478
Fax ++49 3677 691823
klaus.zimmermann@tu-ilmenau.de

J. Kroneis / T. Gastauer / S. Liu / B. Sauer

Flexible modeling and vibration analysis of a parallel robot with numerical and analytical methods for the purpose of active vibration damping

Abstract

In this paper three different methods for vibration analysis of a planar parallel robot are presented: The two numerical approaches (FEM and MBS) work with models of different discretization levels. In case of analytical modeling the manageability of the resulting equations for closed loop control concepts are besides a sufficient characterization of oscillation behavior a key factor for derivation of active vibration damping strategies for parallel robots. In this paper only the vibration analysis is performed applying the different methods and interpreting the results.

Parallel robots and demonstrator SpiderMill

The mechanical design of parallel robots, with its drives units mounted on the fixed base causes a better dynamical behavior than achievable for classical serial robots. Despite the generally higher structural rigidity of parallel robots (due to sustaining effects of the connected braces) in the workspace these structures still tend to vibrations, especially in case of lightweight constructions for fast pick and place applications. The oscillations cause unwanted errors in the trajectory of the tool center point (TCP) and therefore problems in handling of parts or their machining. As a result, analysis of oscillation behavior and development of appropriate strategies for vibration prevention or damping



Figure 1: The SpiderMill

are research focus. In order to find a strategy also for upgrading of existing mechanical structures no structural optimization of the robot should be performed or any kind of adaptionic components used, but rather a concept for active vibration damping developed. Therefore an extensive vibration analysis and adequate modeling of the oscillation modes have to be accomplished.

The planar two DOF parallel robot considered in this paper, called SpiderMill (see Figure 1), comprises a double redundant closed-chain structure, constructed with only revolute

joints and standard aluminum profiles. The robot is intended to be used for rapid prototyping. Measurements on the demonstrator exhibit its vibration tendency performing fast movements.

Modeling of the flexible parallel robot

The vibration analysis, which means the identification of eigenfrequencies (and natural modes), is accomplished for the parallel robot using three different methods: Two of them (FEM and partial flexible MBS models) are numerical solutions and one is a new analytical approach. The derived analytical model will also establish the basis for development of active vibration damping strategies for parallel robots.

FEM analysis with ANSYS: Using ANSYS static deformation behavior is analyzed by finite element method for different configurations of the robot. For the modeling of the parts Solid-Modeling method – generating stereoscopic solids – of ANSYS is used exclusively. The created single parts are combined to form a complete model, called basic model, in a second step. Changing the geometric relationships between elements different configurations of the robot are established. Beside different discretization levels for the braces also models with and without bearing stiffnesses are created. Effects of mass inertias and damping are neglected. In contrast to the other presented approaches all parts of the robot are modeled flexible in the used ANSYS model.

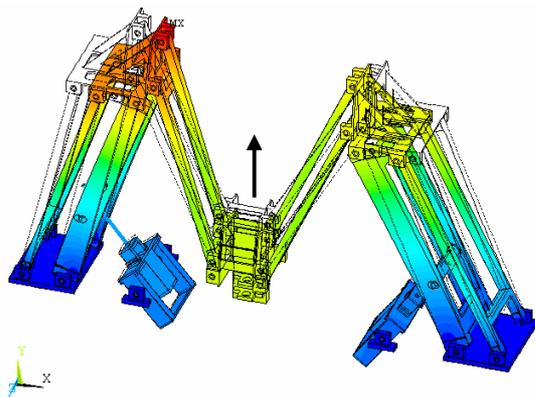


Figure 2: One natural mode

By means of modal analysis the configuration dependent eigenfrequencies and natural modes of the structure are analyzed. The modal analysis implemented in ANSYS is a linear analysis allowing the application of different calculation methods, depending on the application – in our case Block Lanczos method is used. The first 15 eigenfrequencies (corresponding to a

frequency range from 0 to 100 Hz) and natural modes are analyzed. Thereby a calculated linear bearing stiffness is underlying (75 000 M/mm radial, 60 000 N/mm axial). The consideration of more than the first six eigenfrequencies is not necessary due to the fact that these frequencies will cause a dumping of the moving platform which is only possible in a negligible range due to the mechanical structure. For higher natural

modes and therefore higher eigenfrequencies it can be recognized that the structural oscillations of the braces and their resulting modes exhibit the main part of TCP displacement. Furthermore, it can be concluded that the lower natural modes describe the movements of the whole structure in all directions (x-y-z), whereas the higher modes present only movements of the structure elements (braces) in working space (x-y-plane).

Flexible MBS in MSC.ADAMS: For complex mechanical structures constructed with many components a classical analytical modeling approach considering all interactions between the single parts is not possible. Also the usually high order of accurate FEM models makes the simulation and analysis of this class of systems, e.g. parallel robots, difficult. Therefore the modeling as a partial flexible multibody-systems (MBS) simulation model gives the opportunity to analyze the statical and dynamical behavior of a robot and also allows the visualization of movements and oscillations.

In the strategy carried out here, a combination between a classical multi-body-system simulation and a FEM approach is established. The rigid bodies of a basic model are partly substituted by flexible elements in different discretization steps. Due to the combination of the strategies most of the inherent disadvantage of both methods can be avoided. Classical MBS approaches can not study flexibilities of components and can perform only a vibration analysis in the lower and middle frequency range. Using FEM only, movements of bodies are reduced to small ranges. Also due to the usually many DOF the FEM approach is limited to linear models (nonlinear couplings of single system components are not regarded). Furthermore system dynamics is not exactly known (boundary conditions and loads are derived mostly from rigid body analysis). Therefore an adequate modeling – especially in the here occurring case of big nonlinear movements of components with a significant influence of their elasticity – requires a combination of both strategies.

For development of the rigid basic model, data from a CAD tool have been imported in MSC.ADAMS. The basic model is improved in several steps regarding (non-)linear elastic bearings (joints), elastic braces and elastic spindles. The calculated nonlinear bearing characteristics are defined as splines. For allowing a direct comparison especially to the analytical approach presented later bearing stiffnesses of the used MSC.ADAMS reference model have been raised. Braces are modeled as deformable bodies with linear-elastic characteristics. To integrate the flexible parts via finite

elements in the MBS model ADAMS/Flex is applied. With the subprogram AutoFlex the rigid braces of the basic model can be described as flexible bodies using Solid Existing Geometry Method or Extrusion Method. Furthermore, it is also possible to import ANSYS data. In our case only the long braces (not the cross members between them) are modeled flexible applying the first method. For the “flexible” modeling of the spindles a length-dependent General Force connecting rigid bodies has been used.

Applying modal approach an elastic deformation is superposed to the rigid body movement of the structure. Thereby the adequate choice of modes is of vital importance. For the model of the SpiderMill a combination of static and dynamic displacement functions (Component Mode Synthesis, CMS) is used. The Craig-Bampton method is applied as CMS technique, where the occurring component modes consist of static displacement modes (Constraint Modes) and modes of motions (Fixed-boundary Normal Modes; solution of the eigenvalue problem). The disadvantage of the modal approach compared to a discrete one is the only approximative modeling of the stiffness behavior of the robot. But for a mass dominated structure, containing low and high oscillation frequencies and performing linear deformations the modal approach is more suitable than the discrete one and therefore used in our case. As a result of an existing numerical damping, which depends on the solver step size, exact dynamic analysis of the stiffness is not possible. Therefore only analysis of static stiffness behavior is performed. Bearing forces are analyzed by means of statical simulation. Due to the fact that the stiffness highly depends on the configuration of the robot, the whole workspace has to be regarded for an entire analysis.

The implemented MSC.ADAMS models of the SpiderMill are used for dynamic studies regarding loadings of parts or their deformation, respectively. For statical vibration analysis MSC.ADAMS/Vibration is applied. It would also be possible to use models, implemented in a parameterizable manner, for optimization of the mechanical design in the construction phase. But this is not the goal of our approach.

Analytical flexible body model: In a third step an analytical flexible body model of the robot taking the concept of effective payloads into account is derived by applying a hybrid modeling approach. The strategy allows a transient analysis of the eigenfrequencies of the robot during movements, e.g. along different trajectories. For derivation and verification, especially of (rigid) kinematics and dynamics the above mentioned MSC.ADAMS models are used.

Based on a hybrid strategy an analytical description for rigid body kinematics and dynamics is derived: Using the MSC.ADAMS models centers of mass and corresponding point masses are determined especially for complex structural elements like the parallel crank mechanisms. Rigid kinematics is defined by use of standard frame transformations and holonomic constraints. Lagrange's equations of the first type are used for derivation of rigid dynamics. In order to model flexible effects the closed-chain

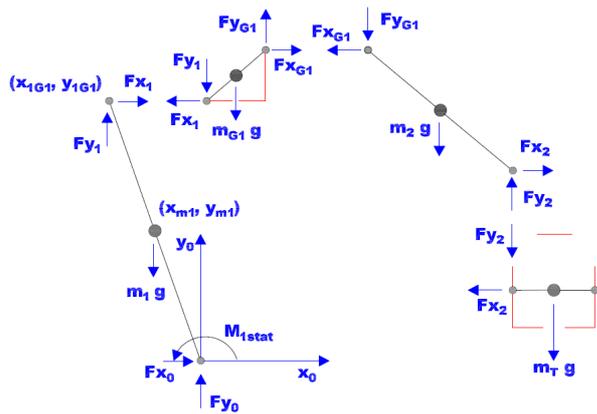


Figure 3: Static forces for the left side

structure is regarded as to serial robots each with two flexible links connected by the moving platform. Moreover, contrary to other flexible body models of parallel robots – where the flexible links have the (rough) form of bars or beams – each beam of the analogous model is regarded separately, taking a configuration dependent effective mass

at its tip into account. The masses are calculated using static torque balances of the robot by cutting free the links and solving the Newton-Euler equations for static forces based on equilibrium conditions (see Figure 3).

Using Euler-Bernoulli beams and assumed modes method, flexible kinematics and dynamics are derived. The eigenfrequencies of the separate beams as well as the oscillations at the TCP can be calculated for different trajectories, performing a transient analysis. Due to the fact that bearings are regarded as ideal stiff in this first analytical modeling approach their counterparts in the MSC.ADAMS reference model have been adapted for comparison purpose.

Analysis and damping of the critical eigenfrequencies

The critical eigenfrequencies of the parallel robot, especially at the TCP, can be analyzed with all three strategies: To illustrate vibration behavior of TCP eight positions on a reference circle in the motion plane of the robot (z-coordinate is zero, see Figure 4) have been defined and analyzed. The results for the first eigenfrequencies are exemplarily compared for position PK 1 in Table 1. Higher frequencies are not regarded due to the fact that only the low frequencies are critical for accuracy of path and positioning. In general (very) high frequencies can not be damped by active vibration damping concepts, because of their very small time constants.

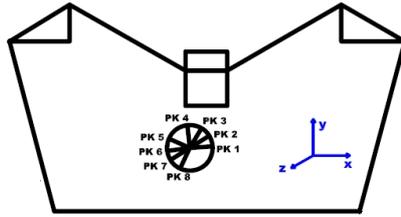


Figure 4: Sketch with reference circle

	x-direction	y-direction
ANSYS	16,48 Hz	20,78 Hz
MSC.ADAMS	42,50 Hz	42,50 Hz
Analytical model	44,74 Hz	44,74 Hz

Table 1: Eigenfrequencies for postion PK 1

Differences between the analytical description and especially the MSC.ADAMS reference model are primarily caused by unmodeled bearing behavior and disregarded effects of the elastic spindles, constraining the free oscillations of the actuated axes of the robot, in the analytical model. Furthermore the analytical model is entire planar. The differences between the frequencies in the ANSYS and MSC.ADAMS models mainly occur due to the full flexible modeling of all bodies and a different description of the bearing stiffnesses in ANSYS. One of the main advantages of the analytical model is also its very short simulation time.

Conclusion and Future Work

In this paper oscillation behavior of a parallel robot has been analyzed applying three different approaches. Thereby occurring (minor) differences of absolute values for the 1st eigenfrequencies have been discussed. To further verify the analytical description, measurements on the demonstrator especially regarding the bearing behavior are necessary. Assured knowledge about occurring frequencies than can be used in open loop control concepts e.g. input shaping of trajectories to damp or even suppress critical frequencies. Moreover, derived and further improved analytical descriptions of flexible kinematics and dynamics form the basis for closed loop control strategies of active vibration damping. Due to the fact that both damping concepts work with the available actuation no constructive changes of the robot are necessary and existing mechanical structures can be upgraded in case of a sufficient dynamic behavior of their drive system.

Authors:

Dipl.-Ing. Jens Kroneis - Institute of Control Systems
 Dipl.-Ing. Tobias Gastauer - Institute of Machine Elements, Gears and Transmissions
 Prof. Dr.-Ing. Steven Liu - Institute of Control Systems
 Prof. Dr.-Ing. Bernd Sauer - Institute of Machine Elements, Gears and Transmissions
 University of Kaiserslautern
 Erwin-Schrödinger-Straße / Gottlieb-Daimler-Straße
 67663, Kaiserslautern
 Phone: +49 (0) 631 205 3132 / +49 (0) 631 205 2858 / +49 (0) 631 205 4535 / +49 (0) 631 205 3405
 Fax: +49 (0) 631 205 4205 / +49 (0) 631-205-3716
 E-mail: kroneis@eit.uni-kl.de / gastauer@mv.uni-kl.de / sliu@eit.uni-kl.de / sauer@mv.uni-kl.de

Arvid Amthor / Tino Hausotte / Gerd Jäger / Pu Li

Friction Modeling on Nanometerscale and Experimental Verification

Abstract

This work concerns the modeling and experimental verification of the extremely non-linear friction behavior in positioning on nanometerscale. The main target of this work is to adjust and identify a simple dynamic friction model which allows a model-based estimation of the friction force in combination with the system inertia against displacement. Experiments in the pre-sliding and sliding friction regimes are conducted on an experimental setup. A hybrid two-stage parameter estimation algorithm is proposed to fit the model parameters based on the experimental data.

I. INTRODUCTION

Friction is a highly nonlinear phenomenon which is present in nearly all mechanical systems. It is induced by interactions between the two rubbing surfaces and depends on several parameters such as surface topography, surface materials or lubricant. Friction can be differentiated into two regimes: the presliding (micro-slip) and the sliding (gross sliding) regime. In the first regime adhesive forces are dominant and friction behaves like a nonlinear spring. In the sliding regime the contacts between the asperities are broken and the friction force depends on the shearing resistance of these asperities. The transition between the mentioned two regimes is continuous and depends on many effects like moving direction, rate of the applied force and others. During a controlled motion these non-linear characteristics lead to tracking errors, limit cycles, stick-slip motion and so on [2]. Due to this dominant nonlinear impact on movements with small displacements, modeling friction is essential to achieve a high-precision dynamic positioning. This is a quite challenging task, since accurate friction modeling based on physical principles and material/surface properties is not possible to date. Hence “Greybox” and “Blackbox” models in combination with efficient identification methods based upon experimentally observed data are often used to solve this problem [1].

II. MECHANICAL SYSTEM WITH FRICTION

A simple mechanical system with friction is considered (Fig.1) [7], [8]. It consists of a mass m , a linear spring k and a damper c . The system is stimulated by a force u and the (unmeasurable) friction force f resist the excited motion.

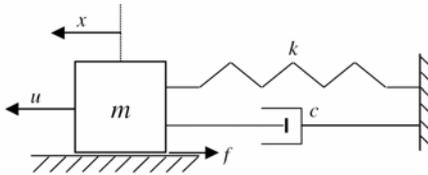


Fig. 1: Simple mechanical system with friction

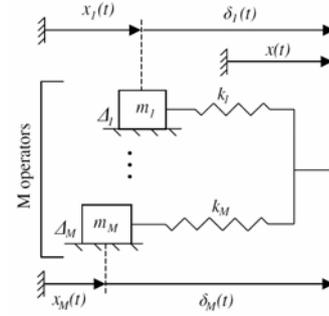


Fig. 2: Basic Maxwell-Slip model structure

This simple mechanical system can be modeled as:

$$m \cdot a + c \cdot v + k \cdot x = u - f \quad (1)$$

The acceleration $a(t)$ and the velocity $v(t)$ are typically obtained by numerical differentiation and this lead to numerical errors and phase shift. The basic idea is that these signals can be calculated via a moving average representation of the order n_v and n_a respectively:

$$v(t) \approx \sum_{j=0}^{n_v} p_j \cdot x(t-j) \quad (2)$$

$$a(t) \approx \sum_{j=0}^{n_a} q_j \cdot x(t-j) \quad (3)$$

A various number of dynamic friction models are available and in this case an extended version of the basic Maxwell-Slip model structure is chosen [1]. This so called Generalized Maxwell-Slip friction model is selected because it is simple and able to predict the friction force with a high accuracy. Furthermore, it can reflect friction phenomenons like the presliding hysteresis with nonlocal memory. This model consists of M elasto-slide operators in a parallel configuration (Fig. 2). Each operator has a negligible mass and a maximal spring deformation (treshold) Δ_i . All elements are subjected to an equal displacement $x(t)$. Hence the Maxwell-Slip structure can be modeled by a set of nonlinear state equations [5], [6]:

$$\delta_i(t+1) = \text{sgn}[x(t+1) - x(t) + \delta_i(t)] \cdot \min\{|x(t+1) - x(t) + \delta_i(t)|, \Delta_i\}; \quad i = 1 \dots M \quad (4)$$

The friction force can be approximated by summing up the spring force of each operator:

$$f(t) = \sum_{i=0}^M k_i \cdot \delta_i(t) \text{ with } t = 1, 2, \dots \text{ referring to discrete time} \quad (5)$$

According to [7], [8] it is possible to allow the friction force depending upon present and past values of the spring deflection, as well as upon values of the displacement. This can be achieved by calculating the friction force by using Finite Impulse Response (FIR) filters of order n' and n respectively:

$$f(t) = \sum_{j=0}^{n'} r_j \cdot x(t-j) + \sum_{j=0}^n \theta_j^T \cdot \delta(t-j) \text{ with } \delta(t) = [\delta_1(t) \dots \delta_M(t)]^T \quad (6)$$

The displacement filter has one dimension and the deflection filter M dimensions. r_j and θ_j^T are the vector coefficients of the FIR filters. Equations (2), (3) and (6) represent approximations of acceleration, velocity and friction force based on the past displacements. Placing of these expressions into equation (1) leads to a time discrete system model:

$$m \cdot \left(\sum_{j=0}^{n_a} q_j \cdot x(t-j) \right) + c \cdot \left(\sum_{j=0}^{n_v} p_j \cdot x(t-j) \right) + k \cdot x(t) = u(t) - \sum_{j=0}^{n'} r_j \cdot x(t-j) - \sum_{j=0}^n \theta_j^T \cdot \delta(t-j) \quad (7)$$

which can be rewritten as:

$$u(t) = \sum_{j=0}^{n_x} g_j \cdot x(t-j) + \sum_{j=0}^n \theta_j^T \cdot \delta(t-j) \quad (8)$$

$$\text{with : } \quad n_x = \max\{n_a, n_v, n'\} \\ g_j = m \cdot q_j + c \cdot p_j + k + r_j$$

This so called DNLRX (*Dynamic NonLinear Regression with direct application of eXcitation*) model consists of two finite impulse response filters. It can provide the current value of the applied force from the known displacement history. In addition, the DNLRX model reflects the *inverse* system behavior of the considered mechanical system. Thus it can be used in a model-based feedforward control system to compensate the disturbances introduced by the non-linear friction characteristics.

III. PARAMETER ESTIMATION ALGORITHM

The parameter identification algorithm uses pairs of displacement - applied force signals to determine the model parameters via a quadratic cost function:

$$J = \sum_{t=k}^N e^2(t) \quad (9)$$

$e^2(t)$ is calculated as difference between the measured, $\hat{u}(t)$, and the model provided, $u(t)$, friction:

$$e(t) = \hat{u}(t) - u(t) \quad (10)$$

Substituting equation (8) in equation (10) yields:

$$\hat{u}(t) = \theta^T \cdot \left[x(t) \dots x(t - n_x) ; \delta^T(t) \dots \delta^T(t - n) ; 1 \right]^T + e(t) \quad (11)$$

$$\text{with : } \theta = \left[g_0 \dots g_{n_x} ; \theta_0^T \dots \theta_n^T ; b \right]^T ; \text{ (b-Offset)}$$

where θ and the threshold-vector $d = [\Delta_1 \dots \Delta_M]$ are the parameters that will be identified. The model is nonlinear with respect to d and linear with respect to θ . A sequential two-stage optimization algorithm is used to identify the model parameters, i.e.

$$[d \ \theta] = \arg \min_d \left\{ \min_{\theta} J(\theta, d) \right\} \quad (12)$$

At the first stage a genetic algorithm is utilized to find the areas of local minimas in the parameter space [3]. At the second stage the Nelder-Mead-Simplex algorithm is used to locate the global minimum in the regions provided in the first stage [9]. For initialization of the proposed identification algorithm, initial values for the maximum deflection thresholds of the springs (Δ_i) are required. To find these initial values, a data-pair is selected where the system is in sliding regime. At this moment (t_{sl}) all Maxwell-Slip elements are sliding and the assumption $\delta_i = \text{sgn}[x(t_{sl})] \cdot \Delta_i$ is justified.

To obtain an optimal identification result, the ‘‘dominant’’ displacement extremum of the time series is selected. For the identification process only data pairs with $t > t_{sl}$ are used. To determine the quality of identification, the *Normalized Output Error* is utilized:

$$NOE = \left(\frac{\sum_{t=k}^N (\hat{u}(t) - u(t))^2}{\sum_{t=k}^N (\hat{u}(t) - \hat{m}_u)^2} \right) \cdot 100\% \quad (13)$$

where \hat{m}_u is the sample mean of the current friction signal and k is specified in equation (9).

IV. EXPERIMENTAL SETUP

For experimental verification a linear guideway driven by a voice coil actuator is used. The operating range of this system is 25 mm. The friction is introduced to the system by the ball bearings of the guideway. The position is measured by a laser interferometer SP-2000 of the company SIOS with a resolution below 0.1 nm. For data acquisition and control purposes a modular dSpace© system in combination with Matlab/Simulink© is utilized. The position is sampled with a rate of 25 kHz and the control algorithm works with a sampling rate of 6.25 kHz. For more detailed information about the experimental setup the reader is referred to [4], [10].

V. RESULTS

The identification process starts with a selection of the optimal model order (M, n, n_x) . This task can be done by evaluating the value of *NOE*. Table 1 shows the identification results of a given friction-displacement dataset with several model configurations. Due to

model order	NOE
DNLRX(2,2,1)	0.3276 %
DNLRX(3,2,2)	0.1417 %
DNLRX(4,2,2)	0.1079 %
DNLRX(5,3,3)	0.0837 %
DNLRX(6,3,3)	0.0598 %
DNLRX(7,4,4)	0.0621 %

Table 1: NOE with respect to the model order (M, n, n_x)

the fact that the cost value decreases practically insignificant *NOE* beyond five Maxwell-Slip elements with a FIR filter length of three, this model order is most suitable to reflect the friction behaviour with respect to the computational cost. Based on this result the mentioned model order has been used for the following identification process.

Fig. 3 presents the identification dataset composed of the displacement (a) and the related applied force (b). The dataset consists of displacements in the sliding as well as in the presliding regime. The data was lowpass filtered with a cutoff frequency of 50 Hz. To speed up the identification process the algorithm was carried out only at every hundredth data-pair. In comparison with a identification using every data sample the performance did not degrade significantly (about 0.01%) and the computing time was reduced by a factor of 30. As shown in Fig. 3 (c), the *NOE* after the identification is below 0.083%. This demonstrate the effectiveness of the proposed two-stage identification algorithm to find a nearly global minimum in the parameter space. Fig. 4 shows the the ability of the identified model to predict the system behaviour for a different validation dataset. In that case the *NOE* is 1.33%. It should be mentioned that t_{sl} is the first sample of the plots in Fig. 3 and 4.

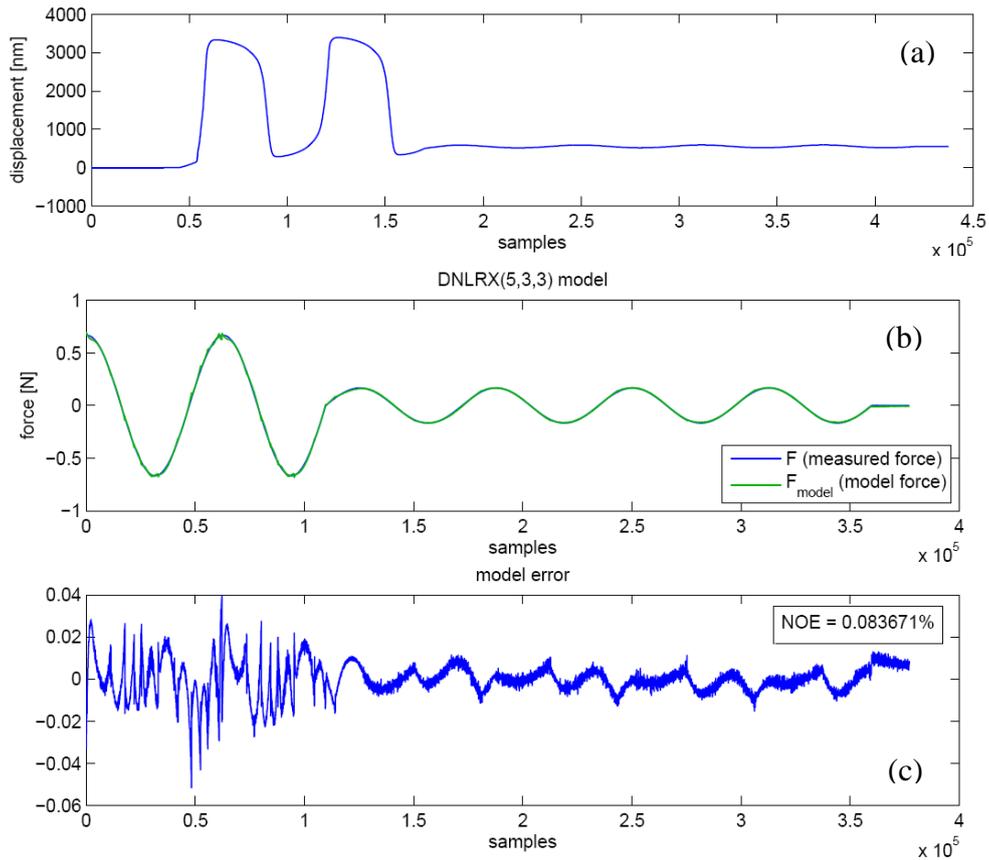


Fig. 3: Identification dataset: (a) displacement; (b) applied force; (c) model based error (NOE)

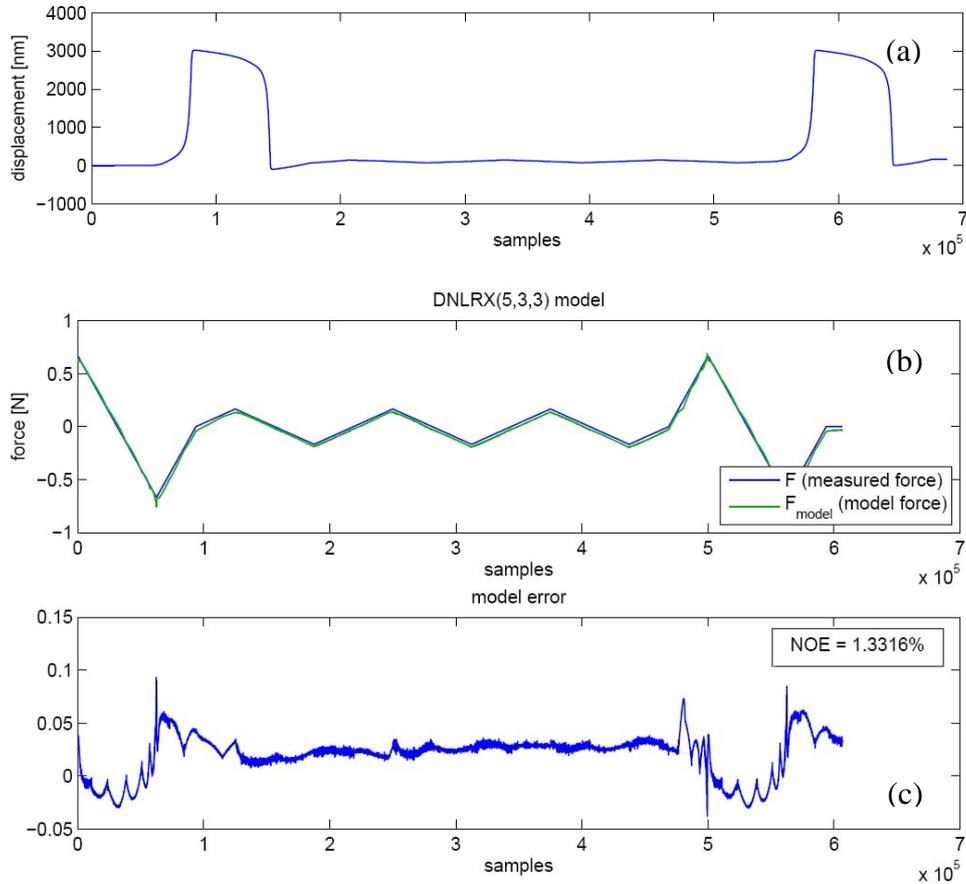


Fig. 4: Validation dataset: (a) displacement; (b) applied force; (c) model based error (NOE)

VI. CONCLUSION

The problem of dynamic friction modeling has been addressed in this work. A two-stage approach to parameter identification is proposed. One of the most commonly used friction models has been analyzed and identified with experimental data. It is shown that this so-called DNLRX model is able to reflect the friction behaviour of ball bearings on nanometerscale and thus can be used for positioning control.

VII. ACKNOWLEDGMENTS

The work has been done in the framework of the Collaborative Research Centre "Nanopositioning- and nanomeasuring machines" at the TU Ilmenau, which is supported by the German Research Foundation (DFG) and the Thuringian Ministry of Science. The authors would like to thank all colleagues who offered helps to the work presented.

References:

- [1] AL-BENDER, F. ; FASSOIS, D. G. ; PARLITZ, U. ; WORDEN, K.: Identification of Pre-Sliding and Sliding Friction Dynamics: Grey Box and Black Box Models. In: *Mechanical Systems and Signal Processing*, 2005
- [2] ARMSTRONG-HÉOUVRY, B. ; DUPONT, P. ; DE WIT, C.: A survey of Models, Analysis Tools and Compensation Methods for Control of Machines with Friction. In: *Automatica*, Vol. 30, No. 7, 1994
- [3] HOUK, CH. R. ; JOINES, J. A. ; KAY, M. G.: A Genetic Algorithm for Function Optimization: A Matlab Implementation <http://www.ise.ncsu.edu/mirage/GAToolBox/gaot/>.
Version: 1998, Abruf: 8.06.2007
- [4] HAUSOTTE, T.: *Nanopositionier- und Nanomessmaschine*, Technische Universität Ilmenau, Dissertation, 2002
- [5] LAMPAERT, V. ; AL-BENDER, F. ; SWEVERS, J.: Modification of the Leuven Integrated Friction Model Structure. In: *IEEE Transactions on Automatic Control*, Vol. 47, No. 4, 2002
- [6] LAMPAERT, V. ; AL-BENDER, F. ; SWEVERS, J.: A Generalized Maxwell-Slip Friction Model appropriate for Control Purposes. In: *IEEE International Conference for Physics and Control*, St. Petersburg, Russia, Proceedings, 2003
- [7] RIZOS, D.D. ; FASSOIS, D.: Maxwell Slip Model Based Identification and Control of Systems with Friction, In: *44th IEEE Conference on Decision and Control, and the European Control Conference 2005*, Sevilla, Spanien, Proceedings, 2005
- [8] RIZOS, D.D. ; FASSOIS, D.: Friction Identification based upon the LuGre and Maxwell Slip Models, In: *16th IFAC World Congress*, Prag, Proceedings, 2005
- [9] WERNSTEDT, J.: *Experimentelle Prozessanalyse*. Berlin : Verlag Technik, 1989, S.201–206
- [10] ZIMMERMANN, J. ; SAWODNY, O. ; HAUSOTTE, T. ; JÄGER, G.: Friction Modelling of a linear high-precision Actuator. In: *IFAC World Congress*, Prag, Proceedings, 2005

Authors:

Arvid Amthor
Tino Hausotte
Gerd Jäger
Pu Li

Technische Universität Ilmenau
Fakultät für Informatik und Automatisierung
Institut für Automatisierungs- und Systemtechnik
Fachgebiet Simulation und Optimale Prozesse
Postfach 10 05 65
98684 Ilmenau

Phone: +49 (0) 3677 69-2816
Fax: +49 (0) 3677 69-1415
E-mail: arvid.amthor@tu-ilmenau.de

V. Piwek / B. Kuhfuss / S. Allers

Feed drives – Synchronized Motion is leading to a process optimization

The mechanical and control superposition of axis motion is very common and used within industrial engineering. Further more a process optimization is in focus to synchronize the drive on the technological process itself. In this paper the mode of operation, the structure and especially the integration in the control system of feed axis is displayed on two manufacturing processes – an incremental cold forming process and a contour turning process.

1. Linear feed for the incremental forming

The characteristics and the potential of optimization of linear feed axis for high dynamic manufacturing processes with highly variable process loads are considered using the example of rotary swaging by the hitch-feed method [1]. In a rotary swaging machine the shaping of the workpiece is performed in small steps by oscillating action of the dies, **figure 1**.

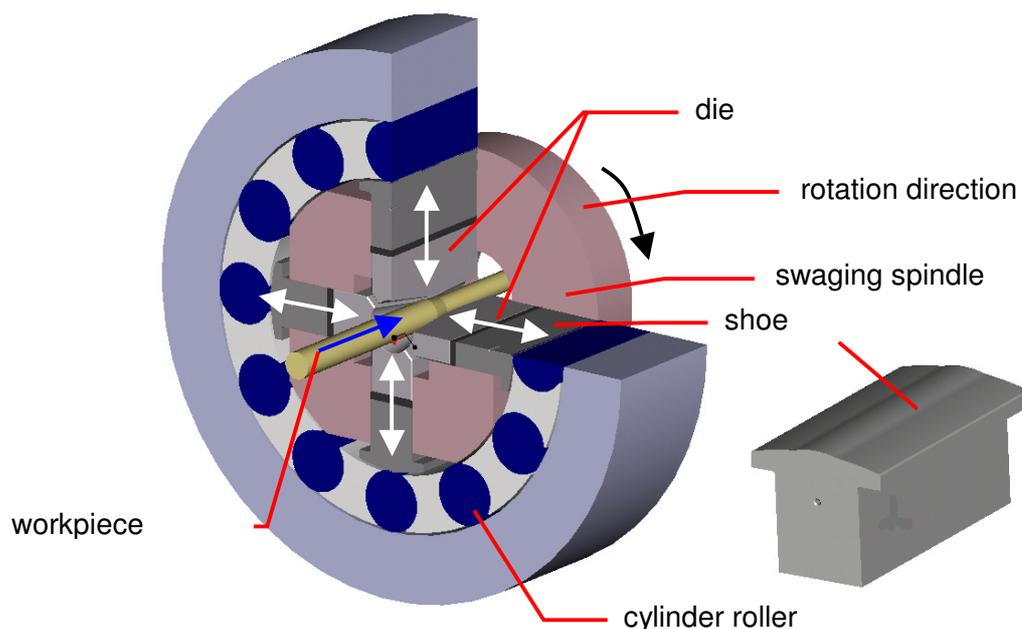


Figure 1: Rotary swaging machine (schematic)

The radial action of the die is generated by rotation of the driven swaging spindle due to the cam shaped outline of the shoe. Through the roll-off-process between the shoes and the cylinder rollers the dies are moved inwards simultaneously at every overrun of the cams. The cross section of the axially feeded workpiece will be reduced. Due to the rotation of the swaging spindle relative to the workpiece a constant circumferential shaping can be obtained.

The outcome of the forming process is an axial impact force which is acting against the feed direction. Furthermore the feed motion is deadlocked by the closing of the dies and the effect of friction and caused by the radial forming force. A significant advancement in the process performance is achieved by synchronized feed and die motion. With the synchronized incremental feed every incremental stroke includes a phase of acceleration and deceleration without axial reaction forces from the rotary swaging process. Because the feed is stopped during the shaping phase there is no additional axial force working towards the swaging machine. Performing this synchronized feed motion the rotary swaging machine is less stressed in axial direction. A feed caused expansion of the swaging machine which can cause a bigger deviation from the ideal diameter can be avoided. Furthermore the restraint feed kinematics can reduce abrasion on the die surfaces by reducing friction. The charts in figure 2 display the strokes of the feed $s(t)$ and the die stroke $h(t)$ as well as the profile of feed rate $v(t)$ for a synchronized incremental travel.

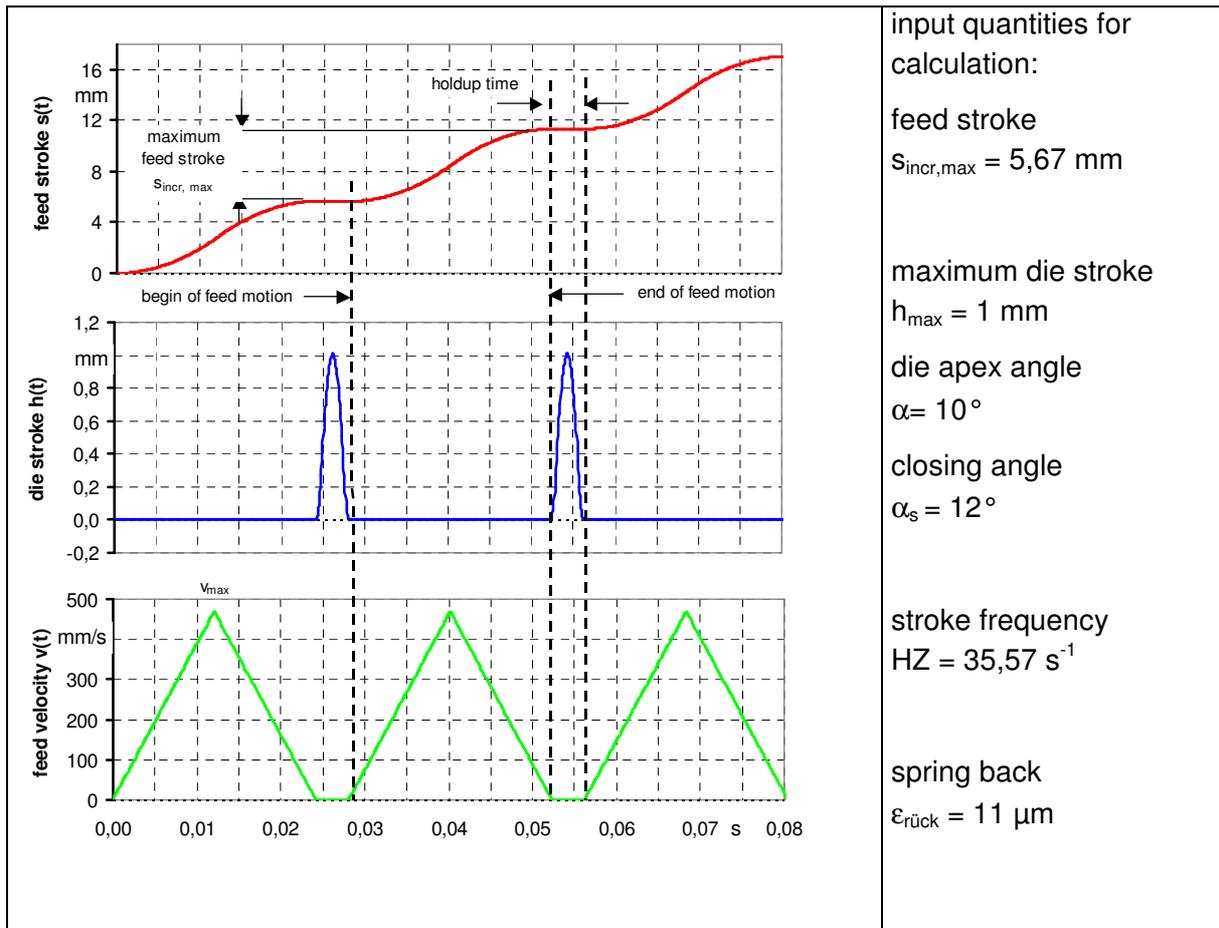


Figure 2: Calculated strokes and velocities for the synchronized incremental feed travel

The feed motion starts when the dies are opening and ends before the next closing phase. The velocity increases until the maximum speed v_{max} is reached at the half of the maximum incremental stroke $s_{\text{incr,max}}$. On a synchronized feed system there are extended dynamical requirements as are on a velocity controlled feed system. Constructively this is solved using full hydrostatic guideways. The power supply of the drive and the motion control is carried out by a power converter, which is externally loaded with the enable signal.

The prototype consists of the main components rotary swaging machine and the linear feed, **figure 3**.

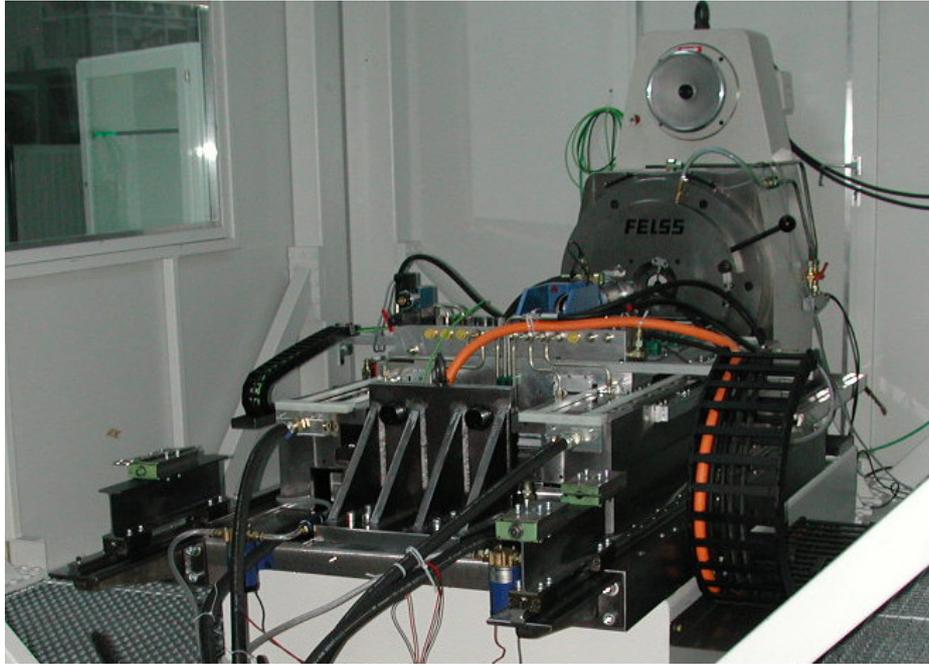


Figure 3: View into the workcore of the prototype

The rotary swaging machine and the feed unit with the bracket unit is mounted on the same substructure. The actual die movement as process factor is detected indirectly by strain gauges on the outer ring of the rotary swaging machine and the signal is conditioned. The extracted square pulse is switched on a hardware input of the power converter and interrupts the motion which is programmed for the entire feed stroke [2]. Thereby the desired cascaded profile of motion is generated.

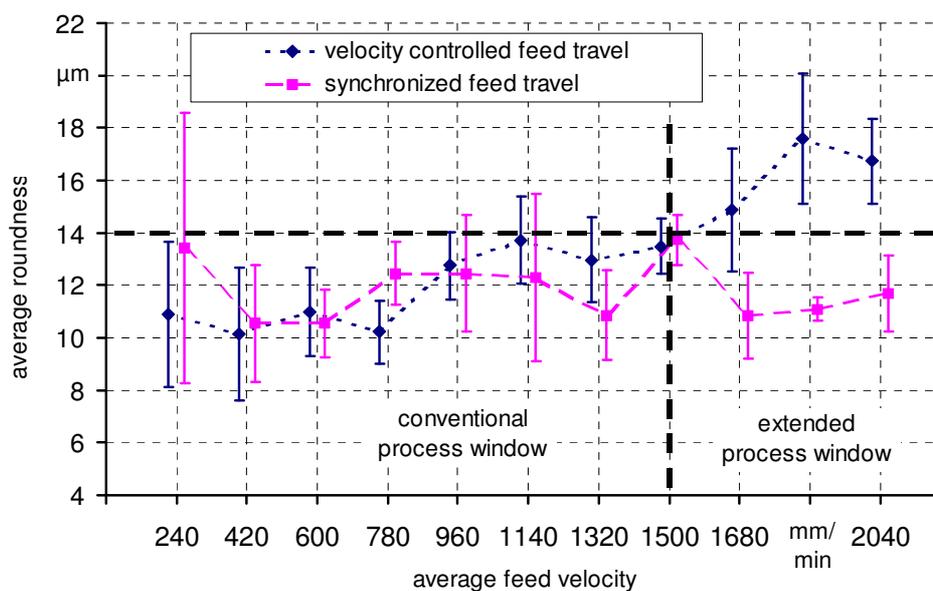


Figure 4: Average roundness

The effect of the synchronized feed system compared to a conventional feed with constant velocity is shown in figure 4. Relating to the quality feature "average roundness of the workpiece" an increase of productivity of about 20% can be observed.

2. Contour turning process with Fast-Tool-Servo

The second application is dealing with the manufacturing of contoured workpieces by a turning process [3]. Therefore the tool tip has to be tracked according to a given ovality or to compensate a deformation due to the work fixture clamping [4]. The complete system contains the in-process measurement technique for a continuous registration of the contour, the required controller and the feed systems, figure 5.

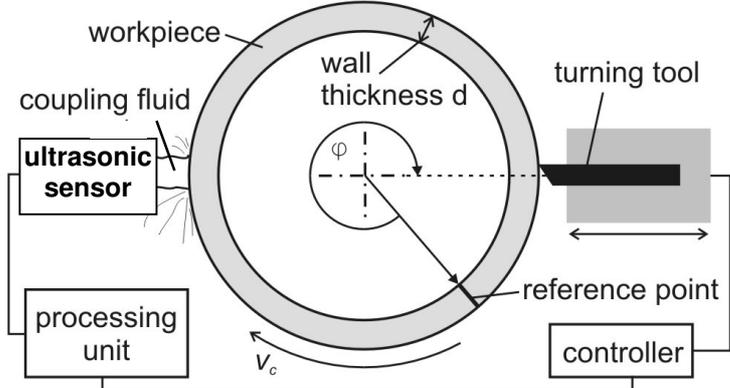


Figure 5: Complete system (schematic)

The cutting edge motion is realized using a linear direct drive guided by monolithic joints. The drive consists of two primary and two secondary parts in a double cam arrangement, figure 6.

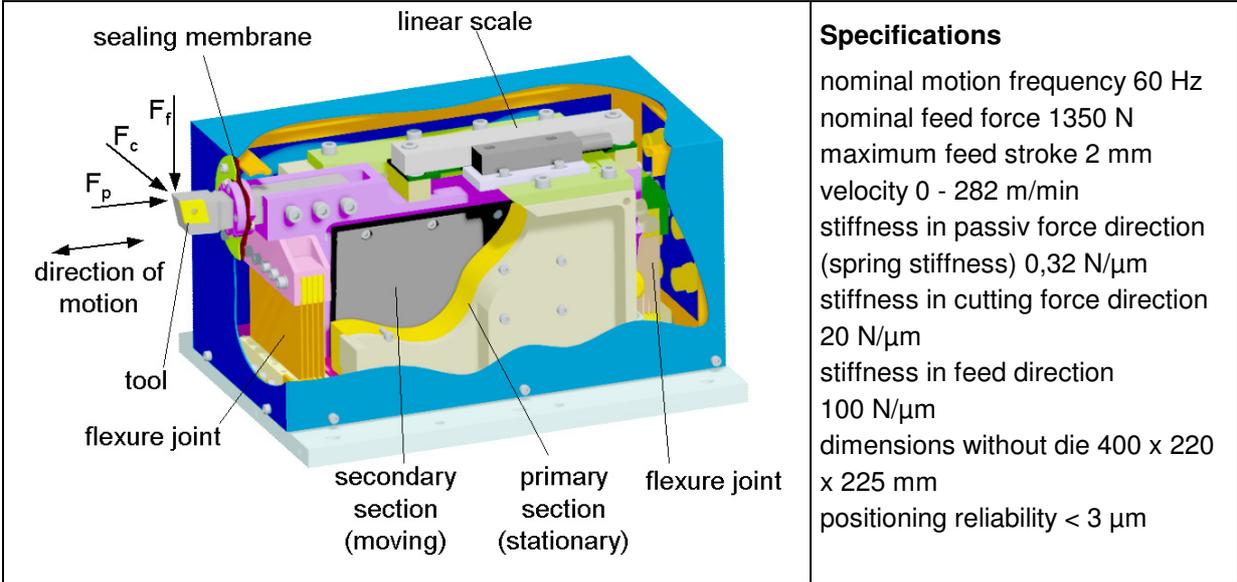


Figure 6: Linear direct drive with monolithic joints

The monolithic joints are designed as pendulum support to prevent tipping of the cutting edge along the stroke. The vertical displacement is relating to the interesting workpiece diameter a deviation of 2nd order. Considering the geometric design data the displacement amount of 0,38 µm is negligible. The joints are fatigue endurable up to feed stroke of two millimeters. For position control of the cutting edge the current position is measured by a linear measuring scale. Typical frequencies of motion are 60 Hz where by the drive can realize process loads up to 1350 N.

The setpoint setting of cutting edge position is realized by a DSP-Board and TTL-Signals. This board is also calculating the signal from the encoder of the spindle and is generating the synchronization with the feed motion. With a software surface feed motions with different amplitudes and different multiples of the workpiece rotation frequency can be programmed.

References:

- [1] Piwek, V.: Dissertation Dr.-Ing., Universität Bremen, Dezember 2006 Shaker Verlag Aachen, ISBN 978-3-8322-6148-1
- [2] Kuhfuss, B.; Piwek, V.: Untersuchungen am Lineardirektantrieb zur Synchronisation der Vorschubbewegung auf die Werkzeugbewegung eines Bearbeitungsprozesses Fachmesse und Kongress 23.-25.11.2004, Nürnberg
- [3] Kuhfuss, B.; Goch, G.; Allers, S.; Stoebener, D.; Dijkman, M.: In-process control of cutting depth during turning SICE Annual Conference 2007, International Conference on Instrumentation, Control and Information Technology, Japan
- [4] Kuhfuss, B.: Ultradynamische Drehmaschine zur Unrundbearbeitung Werkstatt und Betrieb 133 (2000) 9, S. 52 - 54

Authors:

Dr.-Ing. Volker Piwek
Prof. Dr.-Ing. Bernd Kuhfuss
Dipl.-Ing. Sebastian Allers
University of Bremen, Badgasteinerstr. 1,
28359 Bremen
Phone: 0049 421 218 4681
Fax: 0049 421 218 9685
E-mail: ypiwek@uni-bremen.de

René Güttler/Axel Schneider/Christoph Ament/Josef Schmitz

Using a reinforcement learning approach in a discrete event manufacturing system

Abstract

Up to date assembly lines and manufacturing systems use manufacturing cells and single machines which are loaded and unloaded by a gantry loader. The gantry loader is a rare resource that has to be shared by all manufacturing cells and machines (clients). In this context the loading and unloading strategy which is implemented in the controller of the gantry is essential to gain a high manufacturing efficiency. To our knowledge, the implemented control strategies are mostly based on fixed schedules that dictate the order in which the available clients are served. On the one hand a fixed schedule guarantees safe operation in the normal case. On the other hand a fixed schedule can not adapt to unknown situations. We propose a reinforcement learning approach to add the flexibility of lifelong learning to the classical controller's ability of keeping the process in well defined boundaries. The control approach introduced in this paper can also be used offline to train a new controller or it can be used online. To evaluate the adaptive properties of the flexible controller trained by means of reinforcement learning we present simulation data of a small setup consisting of n single machines and a gantry loader.

Introduction

Assembly and production lines consist of components like conveyors, machines for handling the work pieces, CNC-machines for drilling, milling and turning etc. The production processes are controlled by human workers who still, in some cases, fulfil production steps manually. In these processes, all members, be it humans or machines, have to interact in such a way that the work pieces are produced with a high efficiency at a desired quality. The work pieces have to pass different production steps in different machines. Each machine in this chain is specialized on certain production steps to minimize set-up time. Usually, there is a well defined number of machines at each production site in the line to process the desired operations for the work pieces in a desired cycle time. In a real world process, each machine is affected by different machine failures. In such a case, a human operator has to find the source of defect, eliminate it and restart the machine.

Especially when more than one machine accomplishes the same process (parallel) usually a gantry loader transports the work pieces from a storage area or a conveyor to the machines and back. In a well configured system this transfer time (including waiting time due to busy machines) is minimal because the process time of each machine and the arrival of new work pieces are synchronized. This balancing is disturbed by the above mentioned machine failures. A classical controller for the gantry loader orientates

itself to the given state vector of the machines at time t and reacts according to its implemented strategy. However, in some cases it is more suitable to alter the implemented strategy, for example in unpredictable situations like the sudden shutdown of machines due to failures or configuration changes. Therefore, an adaptive strategy for loading and unloading, even in the situations described above, would minimize the transport costs in error-prone real world scenarios. The loader should be able to decide on the action it has to take next for a given situation of the environment. This decision should consider past and current data respectively to adapt to changing situations. Strategies that use past data in current control actions can be regarded as learning strategies.

On the one hand, a controller with a live long learning algorithm is able to find time optimal solutions in new situations but stability can not always be guaranteed. On the other hand, a classical controller uses a robust strategy which is not time optimal in unpredictable situations. In this paper we propose a combination of these two control approaches to combine their advantages (see Fig. 1).

Different applications have already been introduced in the context of industrial manufacturing in which an optimal control strategy was designed by means of a reinforcement learning algorithm. Creighton and Nahavandi developed a reinforcement learning agent to determine the optimal operating policy in a multi-part serial line by using a discrete event simulation environment [3]. Ayedin and Oztemel successfully implemented a reinforcement learning agent for a job-shop scheduling problem [5]. Real world problems for elevator dispatching were presented by Crities and Barto [6, 7].

Using a Reinforcement Learning approach

The idea of Reinforcement Learning (RL) was derived from dynamic programming (DP) where a full model of the process has to be known (including all states and transition probabilities). The motivation for using a reinforcement learning approach for the control of a manufacturing process is that it does not need an explicit model or an a-priori strategy. In real world control problems a closed model of a complex process is difficult to derive, especially in manufacturing applications that consist of many components. A controller based on reinforcement learning develops its strategy by getting rewards for beneficial behaviour and punishments (or fewer rewards) for useless actions.

A control system that is based on reinforcement learning consists of the following components: A policy, a reward function, a value function and usually a model of the

environment [1]. During a learning phase the agent makes a decision (policy) and controls the environment by an action a so that the environment changes its state from s to s' . As a result of the change in the state the agent receives a specified reward r ($r \in R$) for the action he has taken. A cumulative reward is calculated by integration of all current rewards over time. The cumulative reward function implicitly defines the goal of the RL while the agent tries to maximize the overall received reward. The decision for a certain action of the agent for state s depends on the value function which describes the long term experience of the agent [1, 10 and 11].

As opposed to classical controller design, in which the developer has to choose the control strategy based on a model, in RL-based controller design the difficulty lies in designing the reward and value functions. Current research is focussed on finding suitable reward and value functions for a given application because there is no general framework for the design process [7, 8 and 9].

Common methods for calculating the value-function are so called Q-values [1]. These values are stored in large tables or are estimated by neural networks when the state space is too large. The following updating equation (1) describes how to estimate a Value-Function that is derived by the Bellman Equation. It is independent of the transition probability to get from state s to state s' ,

$$Q_{k+1}(s, a) = (1 - \alpha_{k+1})Q_k(s, a) + \alpha_{k+1}[r(s, a) + \gamma \max_{a'} Q_k(s', a')] \quad (1)$$

where $s \in S$ are the states and $a \in A$ are the actions of the agent. $r(s, a)$ denotes the reward for a given state s and action a . The parameter α describes the learning rate and γ the discount rate for learning. Discounting the reward determines the present value for future rewards [1]. The states of the environment include the position of the gantry loader, the loading states of the gripper, the state of the storage area or conveyor and the states of the machines. Actions that are processed by the crane are wait, turn left, turn right and pick-and-place for both grippers.

Architecture of the proposed system

The architecture of the system proposed in this paper comprises a classical controller, a flexible controller, a comparator and the event discrete model of the material flow in the real process. Fig.1 shows the control scheme of this approach. The new, flexible controller to be trained is arranged in parallel to the classical controller. During a training period only the classical controller works on the real process whereas the flexible

controller works on a model of the real process. Both the real and the model process receive work pieces at the input and release work pieces at the output after manufacturing. A second control input of the real and model process is connected to the output of the classical and flexible controller, respectively. Both controllers receive a state vector of their own process as input. Besides the states of the process, the state vector also contains a reward signal from a teacher that judges the efficiency of the real and the modelled process' actions. The classical controller does not benefit from the teaching signal. However, the flexible controller learns to improve its control strategy because controller actions which lead to undesirable results (e.g. low work piece throughput) are punished and actions that improve the work piece flow are rewarded. Therefore, the flexible controller can adapt to situations in which, for example, the productivity of manufacturing cells is decreased due to failures or is increased due to recovery of the cell. A comparator constantly monitors the reward signals from the process controlled by the classical controller and from the process controlled by the flexible controller. When the flexible controller achieves higher rewards, the comparator circuit can decide to swap the classical and the flexible controller (commutator switches in Fig. 1). As a result, the flexible controller governs the real process and the classical controller works on the process model. If the quality of the flexible controller's actions on the process decreases, the comparator can also reverse the swapping process and use the classical controller again until the flexible controller achieves better rewards.

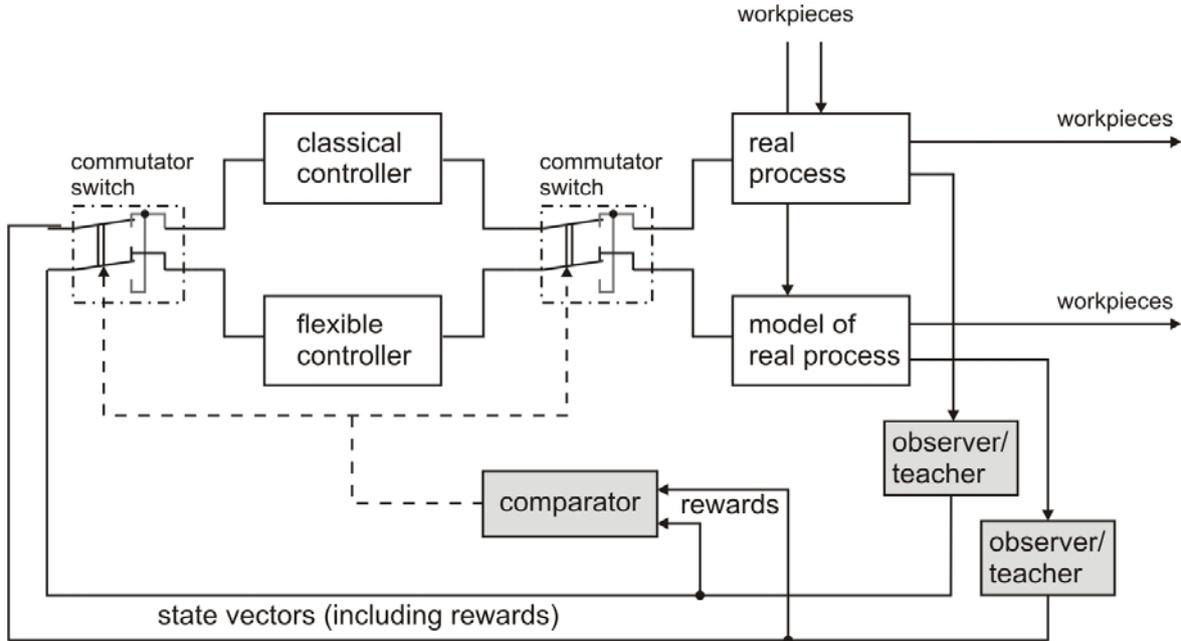


Fig. 1 Combined controller

Simulation data of a small configuration

In order to test our control approach we implement a model of a manufacturing system Matlab/Simulink (The MathWorks Inc., Natick, MA). We compared a simple FIFO (first in first out) strategy with a Reinforcement Learning strategy in throughput of workpieces and the cumultativ reward rate of the action of the gantry loader. The configuration has 4 machines in which work pieces are processed with a processing time of 120s, a gantry loader with two grippers, both for transporting work pieces from the machines to the storage area and back. Finally, there is a storage area where new work pieces arrive every 30s. The system dynamics are characterized as follows: The time to travel from one machine to another is 2.5s, the time for grabbing (pick or place) a work piece takes 6s. In an optimal case the gantry loader's position is above the machines or the storage area whenever the processing of a work piece is finished or a new work piece has arrived. Hence there is no waiting for transport and one can expect a maximal throughput. For each movement of the gantry loader the agent gets a static negativ Reward (penalty). In the successful state, where the crane places a processed work piece at the conveyor belt, the agent gets a reward of 1000 Points (see Fig. 2)

strategy	Output (workpieces)
FIFO	300
RL with exploration factor 0.1	350
RL with manual mode	453

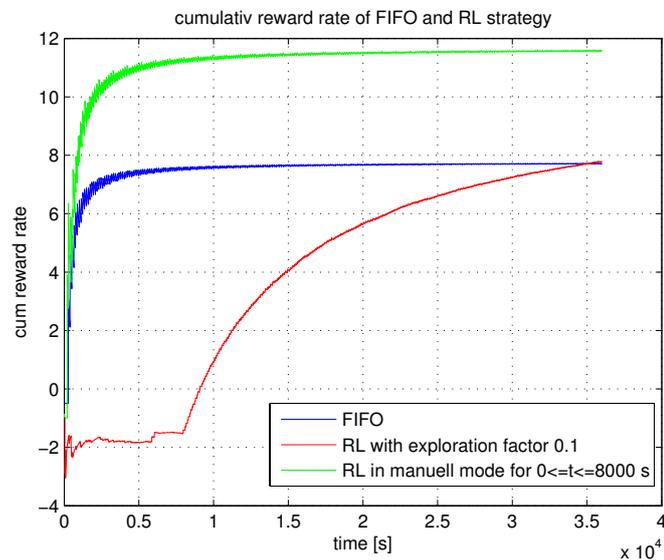


Fig. 2 Comparison between a simple FIFO strategy and a Reinforcement Learning approach in output and reward rate

For this architecture to become active in controlling the real production system the commutator switch is needed (Fig.1). At the branching point between the FIFO reward rate and the Reinforcement Learning reward rate the commutator have to switch from the classical strategy to the reinforcement learning strategy until the reward rate of the

new strategy becomes worse. We suppose that for a well trained flexible controller with Reinforcement Learning the average throughput time for both cases, undisturbed and disturbed process, is less than for the classical controller (e.g. FIFO). Thus, one can expect more output during a given time window.

Conclusion and Future Work

We propose a new architecture to handle control problems in a discrete manufacturing environment by implementing a reinforcement learning approach. We tested this approach separately on a simulated machine configuration with a control task for the gantry loader. In this simulation we compared a classical FIFO strategy with a reinforcement learning strategy. We show that the new approach works in an undisturbed environment. In the present paper we did not consider failures of the machines. Consideration of stochastic failures will be part of our future work.

References

- [1] Sutton, R.S.; Barto, A. G. Reinforcement learning: an introduction. Cambridge: MIT Press, 1998.
- [2] Ghavamzadeh; M., Mahadevan, S. Hierarchically Optimal Average Reward Reinforcement Learning. Proceedings of the Twentieth International Conference on Machine Learning (ICML-2002), pp.195-202. Sydney, Australia, 2002.
- [3] Creighton, D. C.; Nahavandi S. Optimizing Discrete Event Simulation Models Using a Reinforcement Learning Agent. Proceedings of the 2002 Winter Simulation Conference Vol.2,p. 1945-1950 , Sch. of Eng. & Technol., Deakin Univ., Geelong, Vic., Australia, 8-11 Dec 2002.
- [4] Mahadevan, S.; Theocharos, G.: Optimizing Production Manufacturing using Reinforcement Learning. Eleventh International FLAIRs Conference 372-377, 1998.
- [5] Aydin, M. E.; Öztemel, E. Dynamic job-shop scheduling using reinforcement learning agents. Robotics and Autonomous Systems 33, 169-178. 2000.
- [6] Crites, R. H.; Barto, A.G. Improving Elevator Performance Using Reinforcement Learning. Advances in Neural Information Proceedings Systems. Cambridge: MIT Press. 1996.
- [7] Crites, R. H.; Barto, A.G. Elevator Group Control Using Multiple Reinforcement Learning Agents. Boston. Kluwer Academic Publishers. 1997.
- [8] Hong, J.; Prabhu V. V. Distributed Reinforcement Learning Control for Batch Sequencing and Sizing in Just-In-Time Manufacturing Systems. United States. Kluwer Academic Publishers, 2004.
- [9] Ghavamzadeh, M.; Mahadevan, S. Hierarchical Average Reward Reinforcement Learning, CMPSCI Technical Report 03-19, University of Massachusetts, 2003.
- [10] Mitchell, T.M. 1997: Machine Learning. McGraw Hill.
- [11] Gosavi, A: Simulation-Based Optimization: Parametric Optimization Techniques and Reinforcement Learning. Kluwer Academic Publishers, Boston, 2003.

Authors:

Title René Güttler,
Title Axel Schneider,
Title Christoph Ament,
Title Josef Schmitz

Company, street, P.O.B.
98693, Ilmenau
Phone: 03677/691467, E-mail: rene.guettler@tu-ilmenau.de

Name	Page
A	
Abd rabou Kalil, M.	281
Abel, M.	47, 269
Aguiar, A. P.	359
Ahranovich, A.	463, 477
Al Zeer, G.	437
Allers, S.	503
Ament, Ch.	41, 173, 275
Amthor, A.	495
Arnold, E.	157
Arredondo, M. A.	395
Astakhov, V.	165, 169, 171
Avshalumov, A.	95
B	
Balkovoy, A.	469
Bayramkulov, K.	169
Behn, C.	245
Behrmann, T.	419
Bessenov, A. A.	21
Bodyanskiy, Ye.	27, 35
Borikov, V.	89
Bretschneider, P.	211, 217
Bretthauer, G.	257
Büchner, F.	317
C	
Cacenkin, V.	469
Chu, J.	185
Cormack, A.	395
D	
Davydau, G.	141
Debes, K.	431
Diab, A.	281
Doppelhammer, J.	193
Döring, N.	351
E	
Ebel, M.	237
Eichhorn, M.	377
Eifart, M.	287
Ekblad, J.	65
Engel, R.	365
Eylert, B.	47
F	
Fengler, W.	3

Fernlund, H.	65
Ferreira, J.	401
Filaretov, G.	95
Franke, R.	193
Friedemann, A.	351
Fujiki, N.	71
G	
Gastauer, T.	489
Gehr, A.	407
Gerbracht, H.	199
Ghabcheloo, R.	359
Glotzbach, T.	371, 377, 389, 407
Gocht, U.	9
Golz, M.	77
González, A.	65
Gorshkov, Ye.	35
Götze, M.	287
Groß, H.-M.	59, 425, 431, 445, 451
Gu, Y.	185
H	
Hajipoor, A.	135
Hausotte, T.	495
Heit, M.	229, 331, 345
Hensel, H.	53
Hilgers, P.	275
Höchemer, M.	425
Hofer, E.- P.	71
Hong, W.	199
J	
Jacobi, M.	371
Jäger, G.	495
Jawish, I.	179
Jenennchen, E.	229
K	
Kalwa, J.	365
Karpovich, S.	463, 477
Katernoga, O.	141
Kattanek, W.	287
Kaufmann, M.	257
Knauf, R.	113/ 119
Kobayashi, N.	71
Kochubey, T.	171
Koenig, A.	431

Kolodyazhnyi, V.	35
Konigorski, U.	129
Kopfstedt, T.	251, 263
Kratz, M.	311, 351
Krause, T.	413
Krone, J.	173
Kroneis, J.	489
Kroschel, K.	457
Krugljak, V.	345, 229
Kryvenka, M.	331
Kühn, M.	305
Kümmerling, S.	311, 351
Kuhfuss, B.	503
L	
Langbein, P.	173
Laubenheimer, A.	457
Lehmann, U.	173
Lemmel, M.	339, 419
Li, P.	199, 495
Litvinov, Y.	477
Liu, S.	489
Lohfelder, Th.	269
López, A.	129
Lu, Z.	223
Lucia, A.	401
Lüttkopf, K.	47
Lysenko, V.	483
M	
Ma, J.	293
Machmer, T.	457
Maia, P. B.	401
Marinov, M.	389
Martin, C.	59
Meissner, K.	53
Meyer, P.	311
Minchenya, W.	483
Mitschle-Thiel, A.	281
Moreno, P.	317
Müller, C.	41
Müller, F.	83, 407
Müller, P.	317
Müller, S. T.	451
N	
Nabout, A.	437
Nalepa, G. J.	107
Naumenko, J.	167
Nedelina, A.	3
Neupert, J.	157
Nissen, V.	15, 299
O	
Ober, A.	451
Otto, P.	21, 27, 35, 173, 377, 389
P	
Pascoal, A.	359
Pfützenreuter, T.	371
Pham, T. N.	205
Picini, A.	383
Piwek, V.	503
Pliss, I.	27
Popov, V.	141
Potapovich, A.	141
Protzel, P.	413
R	
Reinecke, H.	41
Richter, F.	305
Richter, S.	457
Ritter, S.	211
Rosbach, T.	287
Rozhenko, S.	331
Rudenko, O. G.	21
S	
Sakurai, Y.	119
Salzwedel, H.	101, 305
Sauer, B.	489
Sauer, O.	237
Saul, C.	281
Sawodny, O.	145, 157
Scheidig, A.	451
Schenderlein, M.	431
Schnatmeyer, M.	339
Schneider, K.	157
Schneider, M.	173, 371, 377, 389
Schreiber, A.	287
Schröckel, J.	173
Schröter, C.	425, 445
Shandiz, H.	135
Shaposhnikov, K.	165

Silvestre, C.	359
Slivinskaia, G.	469
Sokolov, O.	9
Sommer, D.	77
Stark, U.	173
Steege, F.- F.	59
Su, H.	185
Suchý, J.	151
Suzuki, R.	71
Swerdlow, A.	457
T	
Taru, Y.	71
Teslenko	27
Tibken, B.	437
Tsuruta, S.	119
V	
Vervoorst, J.	251, 263
Vanni, F.	359
W	
Wachten, Ch.	41
Wagenknecht, M.	9
Wenzel, A.	83, 407
Wernstedt, J.	83
Westermann, D.	217, 229, 311, 317, 331, 345, 351
Winkler, A.	151
Wojnicki, I.	107
Wu, Ji.	223
Wu, Yu.	223
Wutke, B.	205
Z	
Zangrilli, A.	383
Zapaniotis, A. I.	401
Zhong, Y.	223, 293
Zimmermann, J.	145
Zimmermann, K.	245, 463, 483