

# 52. IWK

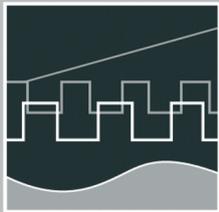
Internationales Wissenschaftliches Kolloquium  
International Scientific Colloquium



**PROCEEDINGS**

| 10 - 13 September 2007

## **FACULTY OF COMPUTER SCIENCE AND AUTOMATION**



## **COMPUTER SCIENCE MEETS AUTOMATION**

### **VOLUME II**

**Session 6 - Environmental Systems: Management and Optimisation**

**Session 7 - New Methods and Technologies for Medicine and  
Biology**

**Session 8 - Embedded System Design and Application**

**Session 9 - Image Processing, Image Analysis and Computer Vision**

**Session 10 - Mobile Communications**

**Session 11 - Education in Computer Science and Automation**

**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](mailto: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](http://www.tu-ilmenau.de/universitaetsverlag)
- Herstellung und  
Auslieferung: Verlagshaus Monsenstein und Vannerdat OHG  
Am Hawerkamp 31  
48155 Münster  
[www.mv-verlag.de](http://www.mv-verlag.de)
- Layout Cover: Torsten Weilepp
- Bezugsmöglichkeiten: Verlagshaus Monsenstein und Vannerdat OHG  
Am Hawerkamp 31  
48155 Münster  
[www.mv-verlag.de](http://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 52<sup>nd</sup> 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 |
|--|------|
| <b>6 Environmental Systems: Management and Optimisation</b>  |      |
| T. Bernard, H. Linke, O. Krol<br>A Concept for the long term Optimization of regional Water Supply Systems<br>as a Module of a Decision Support System | 3    |
| S. Röhl, S. Hopfgarten, P. Li<br>A groundwater model for the area Darkhan in Kharaa river<br>Th. Bernard, H. Linke, O. Krol basin                      | 11   |
| A. Khatanbaatar Altantuul<br>The need designing integrated urban water management in cities of Mongolia  | 17   |
| T. Rauschenbach, T. Pfützenreuter, Z. Tong<br>Model based water allocation decision support system for Beijing   | 23   |
| T. Pfützenreuter, T. Rauschenbach<br>Surface Water Modelling with the Simulation Library ILM-River   | 29   |
| D. Karimanzira, M. Jacobi<br>Modelling yearly residential water demand using neural networks   | 35   |
| Th. Westerhoff, B. Scharaw<br>Model based management of the drinking water supply system of<br>city Darkhan in Mongolia                                | 41   |
| N. Buyankhishig, N. Batsukh<br>Pumping well optimi ation in the Shivee-Ovoo coal mine Mongolia   | 47   |
| S. Holzmüller-Laue, B. Göde, K. Rimane, N. Stoll<br>Data Management for Automated Life Science Applications  | 51   |
| N. B. Chang, A. Gonzalez<br>A Decision Support System for Sensor Deployment in Water Distribution<br>Systems for Improving the Infrastructure Safety   | 57   |
| P. Hamolka, I. Vrublevsky, V. Parkoun, V. Sokol<br>New Film Temperature And Moisture Microsensors for<br>Environmental Control Systems                 | 63   |
| N. Buyankhishig, M. Masumoto, M. Aley<br>Parameter estimation of an unconfined aquifer of the Tuul River basin Mongolia                                | 67   |

M. Jacobi, D. Karimanzira 73  
Demand Forecasting of Water Usage based on Kalman Filtering

## **7 New Methods and Technologies for Medicine and Biology**

J. Meier, R. Bock, L. G. Nyúl, G. Michelson 81  
Eye Fundus Image Processing System for Automated Glaucoma Classification

L. Hellrung, M. Trost 85  
Automatic focus depending on an image processing algorithm for a non mydriatic fundus camera

M. Hamsch, C. H. Igney, M. Vauhkonen 91  
A Magnetic Induction Tomography System for Stroke Classification and Diagnosis

T. Neumuth, A. Pretschner, O. Burgert 97  
Surgical Workflow Monitoring with Generic Data Interfaces

M. Pfaff, D. Woetzel, D. Driesch, S. Toepfer, R. Huber, D. Pohlers, 103  
D. Koczan, H.-J. Thiesen, R. Guthke, R. W. Kinne  
Gene Expression Based Classification of Rheumatoid Arthritis and Osteoarthritis Patients using Fuzzy Cluster and Rule Based Method

S. Toepfer, S. Zellmer, D. Driesch, D. Woetzel, R. Guthke, R. Gebhardt, M. Pfaff 107  
A 2-Compartment Model of Glutamine and Ammonia Metabolism in Liver Tissue

J. C. Ferreira, A. A. Fernandes, A. D. Santos 113  
Modelling and Rapid Prototyping an Innovative Ankle-Foot Orthosis to Correct Children Gait Pathology

H. T. Shandiz, E. Zahedi 119  
Noninvasive Method in Diabetic Detection by Analyzing PPG Signals

S. V. Drobot, I. S. Asayenok, E. N. Zacepin, T. F. Sergiyenko, A. I. Svirnovskiy 123  
Effects of Mm-Wave Electromagnetic Radiation on Sensitivity of Human Lymphocytes to Ionizing Radiation and Chemical Agents in Vitro

## **8 Embedded System Design and Application**

B. Däne 131  
Modeling and Realization of DMA Based Serial Communication for a Multi Processor System

|  |     |
|--|-----|
| M. Müller, A. Pacholik, W. Fengler<br>Tool Support for Formal System Verification  | 137 |
| A. Pretschner, J. Alder, Ch. Meissner<br>A Contribution to the Design of Embedded Control Systems  | 143 |
| R. Ubar, G. Jervan, J. Raik, M. Jenihhin, P. Ellervee<br>Dependability Evaluation in Fault Tolerant Systems with High-Level<br>Decision Diagrams   | 147 |
| A. Jutmann<br>On LFSR Polynomial Calculation for Test Time Reduction   | 153 |
| M. Rosenberger, M. J. Schaub, S. C. N. Töpfer, G. Linß<br>Investigation of Efficient Strain Measurement at Smallest Areas<br>Applying the Time to Digital (TDC) Principle                  | 159 |
| <b>9 Image Processing, Image Analysis and Computer Vision</b>  |     |
| J. Meyer, R. Espiritu, J. Earthman<br>Virtual Bone Density Measurement for Dental Implants   | 167 |
| F. Erfurth, W.-D. Schmidt, B. Nyuyki, A. Scheibe, P. Saluz, D. Faßler<br>Spectral Imaging Technology for Microarray Scanners   | 173 |
| T. Langner, D. Kollhoff<br>Farbbasierte Druckbildinspektion an Rundkörpern   | 179 |
| C. Lucht, F. Gaßmann, R. Jahn<br>Inline-Fehlerdetektion auf freigeformten, texturierten Oberflächen im<br>Produktionsprozess   | 185 |
| H.-W. Lahmann, M. Stöckmann<br>Optical Inspection of Cutting Tools by means of 2D- and 3D-Imaging Processing   | 191 |
| A. Melitzki, G. Stanke, F. Weckend<br>Bestimmung von Raumpositionen durch Kombination von 2D-Bildverarbeitung<br>und Mehrfachlinienlasertriangulation - am Beispiel von PKW-Stabilisatoren | 197 |
| F. Boochs, Ch. Raab, R. Schütze, J. Traiser, H. Wirth<br>3D contour detection by means of a multi camera system  | 203 |

|   |     |
|---|-----|
| M. Brandner<br>Vision-Based Surface Inspection of Aeronautic Parts using Active Stereo  | 209 |
| H. Lettenbauer, D. Weiss<br>X-ray image acquisition, processing and evaluation for CT-based dimensional metrology   | 215 |
| K. Sickel, V. Daum, J. Hornegger<br>Shortest Path Search with Constraints on Surface Models of In-the-ear Hearing Aids  | 221 |
| S. Husung, G. Höhne, C. Weber<br>Efficient Use of Stereoscopic Projection for the Interactive Visualisation of Technical Products and Processes                           | 227 |
| N. Schuster<br>Measurement with subpixel-accuracy: Requirements and reality   | 233 |
| P. Brückner, S. C. N. Töpfer, M. Correns, J. Schnee<br>Position- and colour-accurate probing of edges in colour images with subpixel resolution                           | 239 |
| E. Sparrer, T. Machleidt, R. Nestler, K.-H. Franke, M. Niebelschütz<br>Deconvolution of atomic force microscopy data in a special measurement mode – methods and practice | 245 |
| T. Machleidt, D. Kapusi, T. Langner, K.-H. Franke<br>Application of nonlinear equalization for characterizing AFM tip shape   | 251 |
| D. Kapusi, T. Machleidt, R. Jahn, K.-H. Franke<br>Measuring large areas by white light interferometry at the nanopositioning and nanomeasuring machine (NPMM)             | 257 |
| R. Burdick, T. Lorenz, K. Bobey<br>Characteristics of High Power LEDs and one example application in with-light-interferometry  | 263 |
| T. Koch, K.-H. Franke<br>Aspekte der strukturbasierten Fusion multimodaler Satellitendaten und der Segmentierung fusionierter Bilder                                      | 269 |
| T. Riedel, C. Thiel, C. Schmallius<br>A reliable and transferable classification approach towards operational land cover mapping combining optical and SAR data           | 275 |
| B. Waske, V. Heinzl, M. Braun, G. Menz<br>Classification of SAR and Multispectral Imagery using Support Vector Machines   | 281 |

|  |     |
|--|-----|
| V. Heinzl, J. Franke, G. Menz<br>Assessment of differences in multisensoral remote sensing imageries caused by discrepancies in the relative spectral response functions                                 | 287 |
| I. Aksit, K. Bunger, A. Fassbender, D. Frekers, Chr. Gotze, J. Kemenas<br>An ultra-fast on-line microscopic optical quality assurance concept for small structures in an environment of man production | 293 |
| D. Hofmann, G. Linss<br>Application of Innovative Image Sensors for Quality Control  | 297 |
| A. Jablonski, K. Kohrt, M. Bohm<br>Automatic quality grading of raw leather hides   | 303 |
| M. Rosenberger, M. Schellhorn, P. Bruckner, G. Lin<br>Uncompressed digital image data transfer for measurement techniques using a two wire signal line   | 309 |
| R. Blaschek, B. Meffert<br>Feature point matching for stereo image processing using nonlinear filters  | 315 |
| A. Mitsiukhin, V. Pachynin, E. Petrovskaya<br>Hartley Discrete Transform Image Coding  | 321 |
| S. Hellbach, B. Lau, J. P. Eggert, E. Korner, H.-M. Gro<br>Multi-Cue Motion Segmentation   | 327 |
| R. R. Alavi, K. Brie<br>Image Processing Algorithms for Using a Moon Camera as Secondary Sensor for a Satellite Attitude Control System   | 333 |
| S. Bauer, T. Doring, F. Meysel, R. Reulke<br>Traffic Surveillance using Video Image Detection Systems   | 341 |
| M. A-Megeed Salem, B. Meffert<br>Wavelet-based Image Segmentation for Traffic Monitoring Systems   | 347 |
| E. Einhorn, C. Schroter, H.-J. Bohme, H.-M. Gro<br>A Hybrid Kalman Filter Based Algorithm for Real-time Visual Obstacle Detection   | 353 |
| U. Knauer, R. Stein, B. Meffert<br>Detection of opened honeybee brood cells at an early stage  | 359 |

## 10 Mobile Communications

|  |     |
|--|-----|
| K. Ghanem, N. Zamin-Khan, M. A. A. Kalil, A. Mitschele-Thiel<br>Dynamic Reconfiguration for Distributing the Traffic Load in the<br>Mobile Networks        | 367 |
| N. Z.-Khan, M. A. A. Kalil, K. Ghanem, A. Mitschele-Thiel<br>Generic Autonomic Architecture for Self-Management in<br>Future Heterogeneous Networks        | 373 |
| N. Z.-Khan, K. Ghanem, St. Leistritz, F. Liers, M. A. A. Kalil, H. Kärst, R. Böringer<br>Network Management of Future Access Networks                      | 379 |
| St. Schmidt, H. Kärst, A. Mitschele-Thiel<br>Towards cost-effective Area-wide Wi-Fi Provisioning   | 385 |
| A. Yousef, M. A. A. Kalil<br>A New Algorithm for an Efficient Stateful Address Autoconfiguration<br>Protocol in Ad hoc Networks                            | 391 |
| M. A. A. Kalil, N. Zamin-Khan, H. Al-Mahdi, A. Mitschele-Thiel<br>Evaluation and Improvement of Queueing Management Schemes in<br>Multihop Ad hoc Networks | 397 |
| M. Ritzmann<br>Scientific visualisation on mobile devices with limited resources   | 403 |
| R. Brecht, A. Kraus, H. Krömker<br>Entwicklung von Produktionsrichtlinien von Sport-Live-Berichterstattung<br>für Mobile TV Übertragungen                  | 409 |
| N. A. Tam<br>RCS-M: A Rate Control Scheme to Transport Multimedia Traffic<br>over Satellite Links  | 421 |
| Ch. Kellner, A. Mitschele-Thiel, A. Diab<br>Performance Evaluation of MIFA, HMIP and HAWAII  | 427 |
| A. Diab, A. Mitschele-Thiel<br>MIFAv6: A Fast and Smooth Mobility Protocol for IPv6  | 433 |
| A. Diab, A. Mitschele-Thiel<br>CAMP: A New Tool to Analyse Mobility Management Protocols   | 439 |

## 11 Education in Computer Science and Automation

|   |     |
|---|-----|
| S. Bräunig, H.-U. Seidel<br>Learning Signal and Pattern Recognition with Virtual Instruments  | 477 |
| St. Lambeck<br>Use of Rapid-Control-Prototyping Methods for the control of a nonlinear MIMO-System  | 353 |
| R. Pittschellis<br>Automatisierungstechnische Ausbildung an Gymnasien   | 459 |
| A. Diab, H.-D. Wuttke, K. Henke, A. Mitschele-Thiel, M. Ruhwedel<br>MAeLE: A Metadata-Driven Adaptive e-Learning Environment                      | 465 |
| V. Zöppig, O. Radler, M. Beier, T. Ströhla<br>Modular smart systems for motion control teaching   | 471 |
| N. Pranke, K. Froitzheim<br>The Media Internet Streaming Toolbox  | 477 |
| A. Fleischer, R. Andreev, Y. Pavlov, V. Terzieva<br>An Approach to Personalized Learning: A Technique of Estimation of Learners Preferences       | 485 |
| N. Tsyrelchuk, E. Ruchaevskaia<br>Innovational pedagogical technologies and the Information educational medium in the training of the specialists | 491 |
| Ch. Noack, S. Schwintek, Ch. Ament<br>Design of a modular mechanical demonstration system for control engineering lectures                        | 497 |







T. Bernard / H. Linke / O. Krol

## **A Concept for the long term Optimization of regional Water Supply Systems using a reduced Finite Element Groundwater Model**

### **ABSTRACT**

This paper presents an optimal control approach for the management of the surface and groundwater resources in a region with extreme water scarcity. For this purpose realistic models of the complete surface water system as well as a groundwater model of a large model area (6300 km<sup>2</sup>) have been developed. The key point of the proposed approach is the reduction of the detailed 3D-Finite Element groundwater model to a linear state space model of sufficient low order (~20). In combination with the highly efficient numerical solution algorithm IPOPT the optimal control problem can be solved even for a long term horizon in a user-friendly response time. The results of two optimization scenarios demonstrate the performance of the implemented concept. Furthermore, it is shown that the approximation quality of the reduced linear model is very good.

### **1 Introduction**

The very dynamic economic development in combination with a fast population growth has led to a huge water demand in the region of China's capital, Beijing. This region belongs to the semi-arid climate zone with a highly uneven distribution of the rain throughout the year. More than 80% of the annual precipitation is falling during the summer (July – September). The formerly abundant groundwater resources in the North China plain have been overexploited over the last decades resulting in a strong decline of the groundwater head (up to 40 m).

In a current research project a Decision Support System (DSS) for the management of the water resources in Beijing is being developed and implemented. The issues dealt with the DSS outlined in Figure 1 range from the optimization of reservoir extraction to the long-term strategic distribution of available water resources for industry, agricultural use and private households. For the design of the DSS a model of the entire water supply system of the Beijing region was developed. This complex model includes all

essential elements to collect, store and distribute the surface and groundwater resources on a regional scale. Model based DSS for optimal water supply have been developed for more than twenty years [1], but many of them are dealing with very simplified models, especially for the groundwater system. In our approach the groundwater behavior has been modeled by spatially distributed 3D Finite Element model of the considered region. However, the optimization of the regional water resources management with respect to different assumptions for future exterior changes (e. g. economy, climate changes) is one key item of the DSS to be developed. The solution of optimal control problems based on a high-resolution groundwater model is time-consuming and would prevent the practical application of the DSS. Therefore the complexity of the 3D finite element groundwater model has to be drastically reduced. Hence in this paper a trajectory based scheme for groundwater model reduction is proposed.

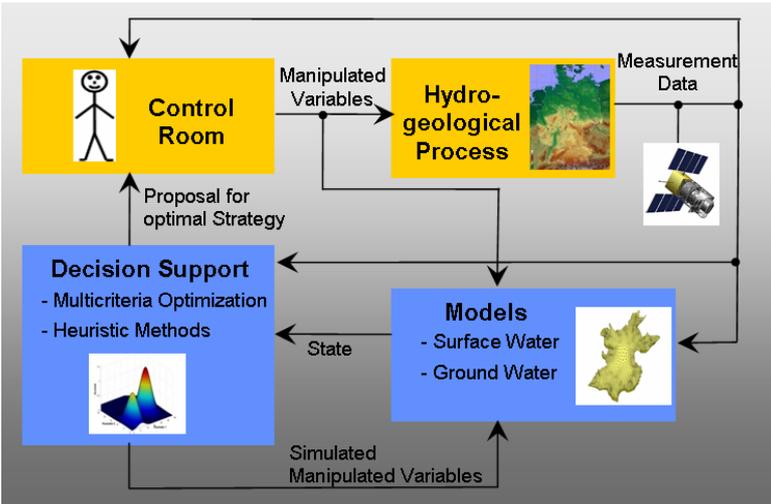


Figure 1: Basic Concept of the Monitoring and Decision Support System

This paper presents an overview of the developed process models for the surface water system (section 2) and groundwater resources (section 3) including model reduction. Aspects of the formulation and numerical solution of the optimal control problem are discussed in section 4. First results of the proposed optimal control approach for water management are presented in section 5.

**2 Surface water model**

Five reservoirs with a total storage capacity of roughly  $9 \cdot 10^9 \text{ m}^3$  are important for the management of the natural surface water resources in the considered area and are therefore incorporated in the decision support system. The water is distributed to the customers using rivers and artificial transport ways (channels, pipes) of a total length of

about 400 km. The water distribution system is formally described as a directed graph. The nodes represent reservoirs, lakes, points of water supply or water extraction and simple junction points. The water distribution process is described by the edges (e.g. river reaches, channels), whereas the flow dynamics is neglected according to the specified time step for decision support. The sole nonlinearity of the surface water model results from the volume-surface relation of the reservoirs, which is included to model the evaporation from the water surface and is described by a piecewise polynomial approach. Figure 2 shows the structure of the surface water system including the connections to the groundwater system.

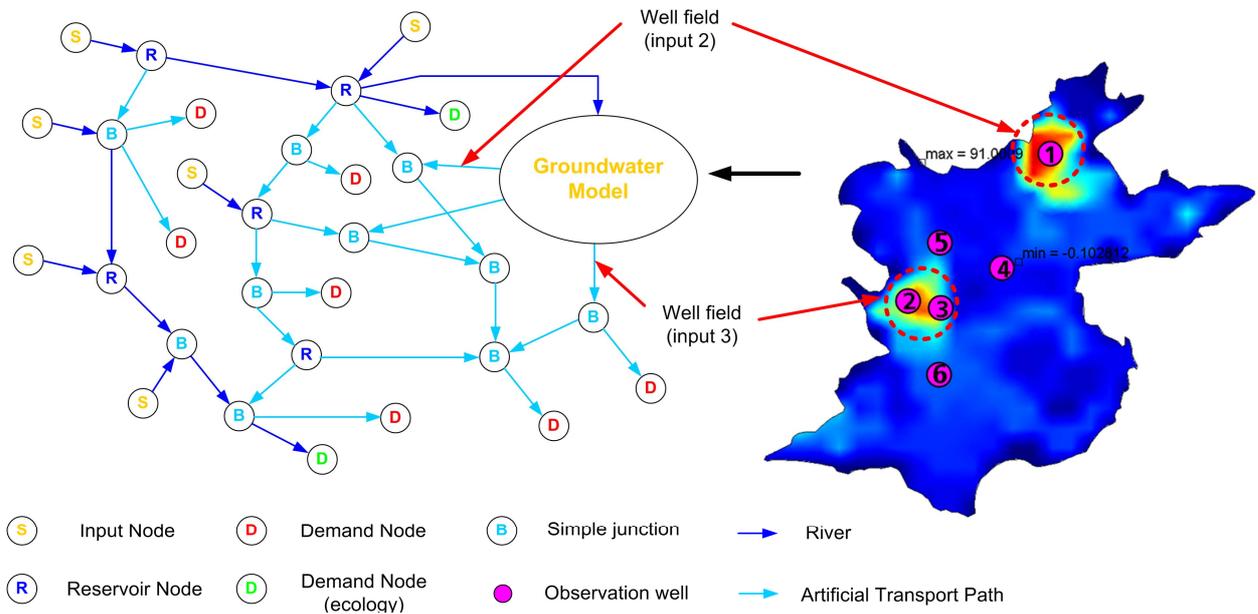


Figure 2: Structure and coupling of the reduced surface and groundwater model

### 3 Ground Water Model

#### 3.1 Equations and Parameters of the 3D Finite Element Model

The second important water resource is the groundwater in the considered area that is modelled by a spatially distributed finite element groundwater model. The governing equation for groundwater flow is Darcy's law  $k_f \cdot \text{grad } h = \dot{q}$  [L/T] [2] describing slow streams through unconfined aquifers. It denotes  $h$  [L] the hydraulic head (which corresponds to the groundwater level) and  $k_f(x,y,z)$  [L/T] is the hydraulic conductivity that governs the hydrogeological properties of the soil. Combining Darcy's law with mass conservation one yields the partial differential equation (1) which is a diffusion equation.

$$S_0 \frac{\partial h}{\partial t} - \text{div}(k_f \cdot \text{grad } h) = Q \quad (1)$$

(1) is an initial-boundary value problem which has to be solved for  $h$  in the 3 dimensional model domain  $\Omega$ . The initial condition is  $h(\Omega, t_0)$  (groundwater surface) at the initial time

$t_0$  in  $\Omega$ . Boundary conditions are chosen of Dirichlet type, i.e.  $h(\partial\Omega)$  at the boundary  $\partial\Omega$  which are assumed to be time-independent. In (1)  $S_0(\Omega)$  [1/L] denotes the specific storage coefficient.  $Q(\Omega,t)$  [1/T] summarizes all sources and sinks that coincide with the groundwater recharge and the exploitation in  $\Omega$ . The main task with respect to the groundwater model is the parameterization of the large-scaled model covering an area of 6300 km<sup>2</sup>. On the one hand the time independent parameters  $k_f$ ,  $S$  and boundary condition  $h(\partial\Omega)$  have to be estimated and generalized for the whole domain  $\Omega$  by a set of measured values. On the other hand the source / sink term  $Q(\Omega,t)$  has to be calculated time-dependent.  $Q$  can be divided in the source term  $Q_{rech}(\Omega,t)$  which describes the groundwater recharge  $Q_{rech}(\Omega,t)$  as a sink which corresponds to the flow into the groundwater storage e. g. due to precipitation. It holds

$$Q = Q_{rech} - Q_{expl} \quad (2)$$

The calculation of  $Q_{rech}$  and  $Q_{expl}$  (especially exploitation by agriculture) is performed by means of a geographical information system (GIS) that is appropriate to compute spatially distributed data. For these calculations only a spatial distributed precipitation and water demand is needed (these data are in general available). The water demand is split into the three user groups households, industry and agriculture, whereby a part of the agricultural water demand is covered by precipitation. The rest of the agricultural demand has to be fulfilled by irrigation, whereby a part of it is covered by waste water.

### 3.2 Reduction of the Ground Water Model

For the optimization of the water allocation system the full 3D Finite Element model is not very suited due to the large computational time. As for optimization a prediction of the hydraulic head (groundwater level) at a set of representative points is sufficient an input-output model (e.g. a linear state space model) with considerably smaller order  $n < 50$  than the original FEM model has to be derived. The sources and sinks of the groundwater system are regarded as input of the model as the water allocation and irrigation has to be optimized. The spatially distributed allocation of water is realized by a withdrawal for agriculture, which covers the whole model area (input 1) and two well fields (input 2 and 3, see Figure 2) which cover only a small fraction of the model area. Finally, the irrigation is also considered as input 4 of the reduced groundwater model. By a trajectory and identification based approach the original 3D-Finite Element model with ~100.000 nodes has been reduced to a state space model with only 18 states. The

reduced model has only a small approximation error (cf. results in section 5). For details regarding the model reduction see [3].

#### 4 Setup and solution of the optimization problem

The water management problem is stated as discrete time optimal control problem:

$$\min_{\mathbf{u}^k, k=1, \dots, N} \left\{ F(\mathbf{x}^N) + \sum_{k=0}^{N-1} f_0^k(\mathbf{x}^k, \mathbf{u}^k, \mathbf{z}^k) \right\} \quad (3)$$

subject to

$$\mathbf{x}^0 = \mathbf{x}(t_0), \quad \mathbf{x}^{k+1} = \mathbf{f}^k(\mathbf{x}^k, \mathbf{u}^k, \mathbf{z}^k), \quad \mathbf{h}^k(\mathbf{x}^k, \mathbf{u}^k, \mathbf{z}^k) = \mathbf{0}, \quad \mathbf{g}^k(\mathbf{x}^k, \mathbf{u}^k, \mathbf{z}^k) \leq \mathbf{0} \quad (4a,b,c,d)$$

where it denotes:  $\mathbf{x}$ : state variables,  $\mathbf{u}$ : control variables,  $\mathbf{z}$ : uncontrollable inputs (e.g. inflow to the reservoirs, evaporation),  $N$ : optimization horizon in time steps. The process equations (4b) contain the balance equations of the reservoir nodes and the reduced groundwater model. The balance equations of non-storage nodes are formulated as general equality constraints (4c) and technical capabilities of the water supply system (e.g. maximum capacity of channels or pipes) are described by inequality constraints (4d). The goals of the water management are transformed into the objective function (3) (which evaluates e. g. demand fulfillment or sustainability) and inequality constraints (e.g. time varying maximum storage volume of reservoirs, minimum rate of demand fulfillment, minimum groundwater head). The initial state  $\mathbf{x}^0$  and predictions for the non-controllable inputs  $\mathbf{z}$  over the full optimization horizon  $N$  are a prerequisite to solve the optimal control problem.

The discrete time optimal control problem (3) - (4) is numerically solved as a large-scale nonlinear programming problem in the state and control variables. Due to the high dimension of this solution approach for the discrete-time optimal control problem, it is essential to take the special sparsity structure of the problem into account. The optimization solver *IPOPT* is used as core numerical routine [4]. The problem interface for multistage optimal control problems of the optimization solver *Hqp* [5], which provides an efficient way for the problem formulation, is linked to *IPOPT*. The nonlinear programming problem resulting from a time horizon of 5 years with a sample time of 5 days has about 23,000 optimization variables and is solved in approximately 120 steps requiring about one 15 min time (Pentium 2.8 GHz).

#### 5 First Results with the Optimal Control Approach

The performance of the proposed optimal water management concept is demonstrated by means of two optimization scenarios. In scenario A it is assumed that no constraints with respect to the groundwater observation points  $\{h_i\}$ ,  $i=1, \dots, 6$  are defined, while in

scenario B a time dependent lower limit (constraint) to the hydraulic head of observation point 5 is applied as it is shown in Figure 4 on the upper subplot on the right side. In both cases the objective function contains terms to reduce the difference between water demand and supplied water as much as possible (i. e. fulfill demand to highest possible degree). The optimization horizon is  $t_h = 5 \text{ years}$ .

In Figure 3 results of the surface water system are presented. On the left the time plot of the water level of one exemplary reservoir is shown. It can be seen that for scenario B the water of the reservoir is used during  $t = 0 \dots 1.5 \text{ years}$  for groundwater recharge fields in order to fulfill the applied groundwater constraint. Note that in this management strategy the lower bound for the reservoir water level is reached for several times (at  $t = \{1.5, 2.5, 3.5, 4.5\} \text{ years}$ , cf. Figure 3), whereas in scenario A the water level of the reservoir is only slightly decreased. On the right of Figure 3 the time plot of the global surface water supply is shown. It is obvious that the water demand of the different customers is fully supplied in scenario A, while in scenario B the required increase of the groundwater head at point 5 yields a demand deficit of about 25 %.

Figure 4 shows on the upper left side the time plot of ground water extraction from the well field 2 and on the lower left side the artificial groundwater recharge. In the middle of Figure 4 the hydraulic head  $h$  at the observation points 5 and 6 (= output 5, 6) can be seen. In the right of Figure 4 the approximation error  $\Delta h$  of the reduced model (i.e. difference of output of full FEM model and reduced state space model) is shown. Obviously  $\Delta h$  of output 6 ( $|\Delta h_{max}| \sim 0.1 \text{ m}$ ) is smaller than for output 5 ( $|\Delta h_{max}| \sim 0.35 \text{ m}$ ), but both errors are small compared to the changes of these outputs which means that the approximation quality of the reduced model is good.

In scenario B the allocation of the well field, which is not located close to the point 5 and hence has only a little impact to this point, is increased to its upper limit. For the groundwater recharge (irrigation) the result of scenario A is a nearly constant value over time while in scenario B an interesting decrease at  $t = 0 \dots 1.5 \text{ years}$  can be observed. This decrease of groundwater recharge corresponds to the increased allocation from the reservoir which has been discussed. So obviously the optimization changes the management strategy in a transparent way. It is also obvious that the applied constraint of scenario B at point 5 is not violated, i.e. the hydraulic head is increased by the optimal control.

Summarizing, it can be stated that the water management system works in a transparent and user-friendly way as it is easy to parameterize with time-dependent

constraints on the hydraulic head at certain locations. The optimal control concept is planned to be launched as a component of a Decision Support System in a first version during the Olympic Games 2008 in Beijing.

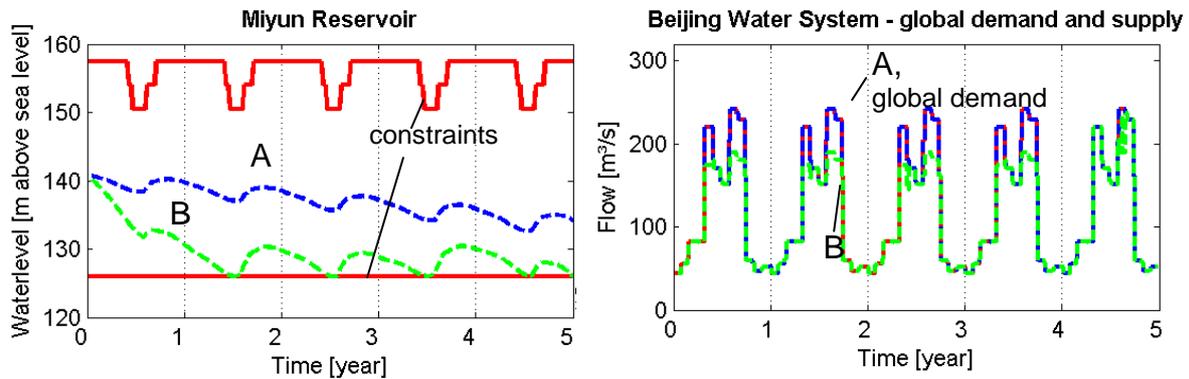


Figure 3: Results of the surface water system of scenarios A and B.

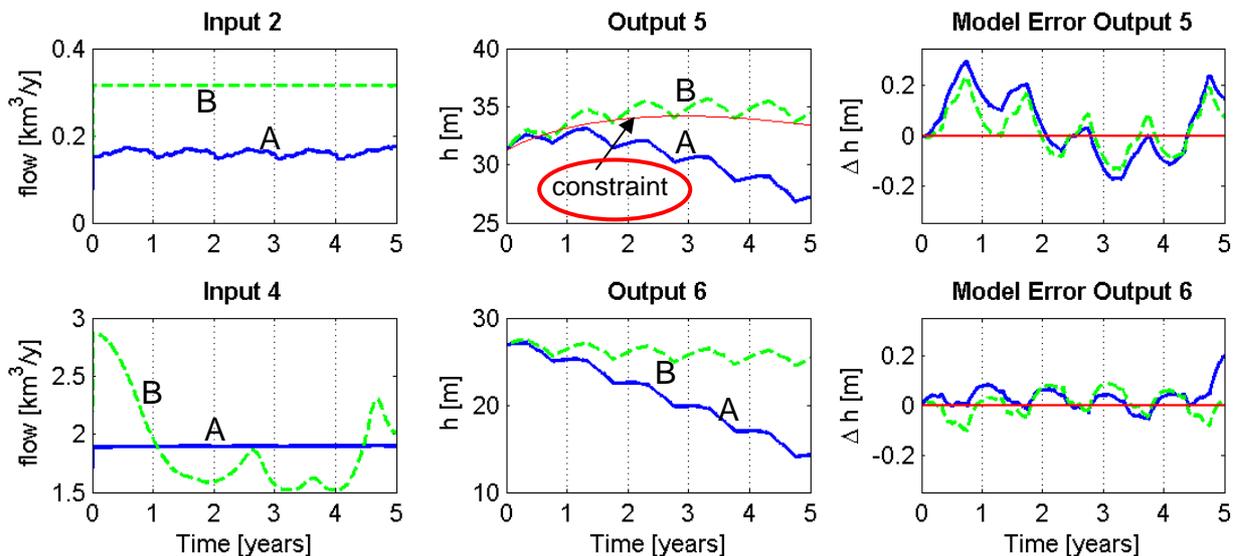


Figure 4: Results of the groundwater system of scenarios A (solid —) and B (dashed - - -)

#### References:

- [1] D. C. McKinney: International Survey of Decision Support Systems for integrated Water Management, Technical Report, IRG project No. 1673-000, 2004
- [2] J. Bear: Dynamic of Fluids in Porous Media, Dover Publications, Inc., New York, USA, 1988
- [3] T. Bernard, H. Linke, O.Krol: A Concept for the long term Optimization of regional Water Supply Systems using a reduced Finite Element Groundwater Model. VDI/VDE GMA-Kongress 2007, VDI-Berichte 1980, pp. 751-762, Baden-Baden (Germany), 12./13.6.2007
- [4] A. Wächter, L. T. Biegler. On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. Mathematical Programming, 106(1):25–57, 2006.
- [5] R. Franke, E. Arnold: The solver Omuses/HQP for structured large-scale constrained optimization: algorithm, implementation and example application, Sixth SIAM Conference on Optimization, 1999, Atlanta, <http://hqp.sourceforge.net/index.html>

#### Authors:

Dr.-Ing. Thomas Bernard, Dr.-Ing. Hartmut Linke, Oliver Krol  
 c/o Fraunhofer-Institut für Informations- und Datenverarbeitung IITB  
 Fraunhoferstraße 1, 76131 Karlsruhe  
 Phone: +49 721 6091-360, Fax: +49 721 6091-413  
 E-mail: {thomas.bernard, oliver.krol}@iitb.fraunhofer.de; hartmut.linke@ast.iitb.fraunhofer.de



S. Röhl / S. Hopfgarten / P. Li

## **A groundwater model for the area Darkhan in Kharaa river basin**

### **Introduction**

Water, especially drinking water, is a valuable substance worldwide. And also worldwide we are faced with the problem of decreasing drinking water resources. In the BMBF project “Integrated Water Resources Management for Central Asia: Model Region Mongolia (MoMo)” this serious problem is taken account. Based on the interaction of surface water, ground water and water distribution systems, influenced by climatic change, an integrated approach dealing with special problems of each system has been established. The project considers the Kharaa river basin in Mongolia, an arid region with typical problems like decreasing water resources and drinking water quality.

In this project the research at Technical University of Ilmenau is responsible for developing optimal strategies for the groundwater extraction in the area Darkhan in the Kharaa river basin. A simulation model is to be developed as a base for the optimization. The first step of our research is to develop an adequate mathematical model of the groundwater system and all required model parameters will be determined according to the data from the local area. This model will be used to calculate the current amount of groundwater, to forecast the future capacity and to develop strategies for an optimal, sustainable water extraction.

### **Model Area**

The model area is located in the Kharaa river basin, Mongolia, close to the town Darkhan, the third largest town of Mongolia with around 70,000 inhabitants. Because of only a few surface water resources available drinking water is extracted from ground water. The area is characterised by an arid, continental climate with a yearly precipitation between 300 and 400 mm. Only a small amount of the precipitation contributes to ground water recharge. Most of it is lost by run-off or evaporation.

Consequently, the main recharge of ground water is done by the river Kharaa, with a length of 300 km and a catchment area of around 15,000 km<sup>2</sup>. In the determined model area, as shown in Figure 1, the river has a width between 10 and 20 m and is fanning out. The water level highly depends on the precipitation. Thus the level is high in spring and low in summer and autumn. In winter the river is frozen, partly up to the ground because of the low temperature. This phenomenon has also to be taken into account.

The flow of ground water mainly occurs in a band along the river which determines the ground water system. The width is around 20 km. This unconfined aquifer consists of porous media which is characterised by alluvial sand and gravel with interlaced sandy loam [4], [5]. The depth of the water transporting layer is between 65 and 70 m dividing up in different layers. The level of ground water is dependent on the season, in average around 3 m below the ground.

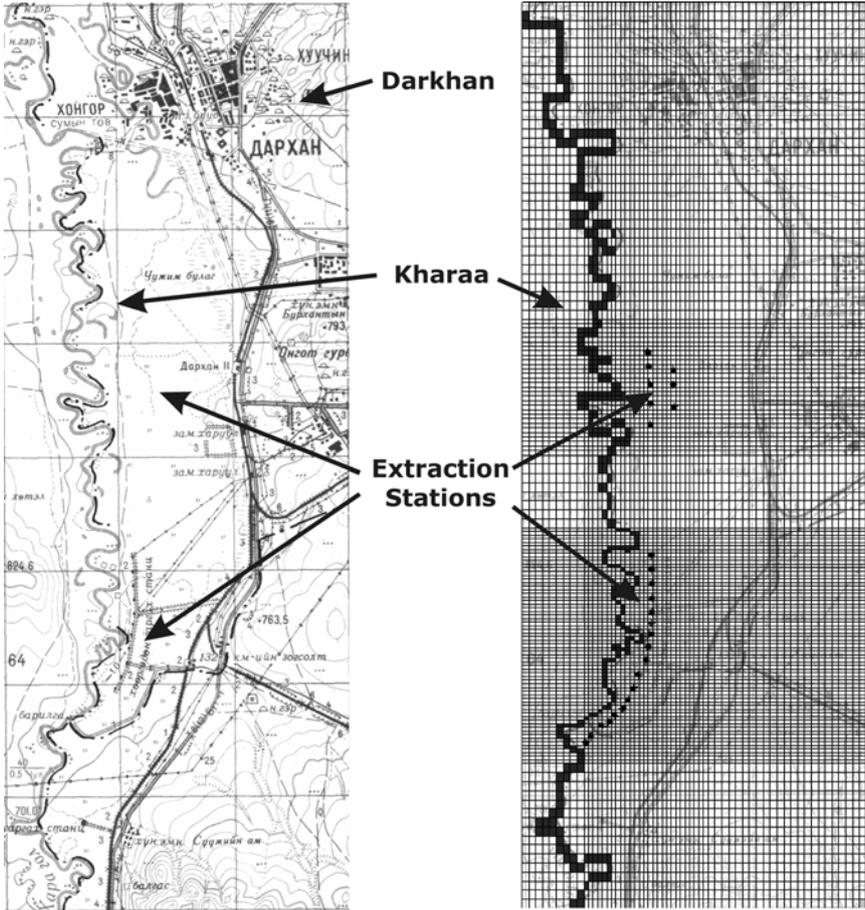


Figure 1 - Overview of the model area Darkhan

As mentioned above, ground water is used for drinking water purpose because of low

surface water resources. There are two main extraction systems with 18 and 8 wells. These systems are between 250 and 700 m far away from the Kharaa River. Each well has a depth of around 60 m and a capacity of 160 m<sup>3</sup> per day. The water is extracted from a depth of 30 m below ground. Based on this analysis of the region a mathematical model will be developed.

### **Modelling**

In a porous medium like this type of aquifer the ground water flow could be described by the following partial differential equation (1). This equation is a combination between the Darcy's equation which describes the velocity of water in porous media and the continuity equation which describes the mass balance of water [2], [1].

$$\frac{\partial}{\partial x}\left(k_x \frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(k_y \frac{\partial h}{\partial y}\right) + \frac{\partial}{\partial z}\left(k_z \frac{\partial h}{\partial z}\right) = S_0 \frac{\partial h}{\partial t} + q(x, y, z, t) \quad (1)$$

The level of ground water is described by the variable  $h$  which can be obtained at the solution of this equation. The parameter  $k$  represents the conductivity of the porous medium, individual for each of the three directions of space. The parameter  $S_0$  represents the specific yield of the aquifer. The water extraction of the aquifer is described by the parameter  $q$ .

Because the project is in the beginning phase and gaining data from the local area is very difficult, now we have only a few data available. In this case we have to make some assumptions. The first assumption commonly used in modelling such processes is that the aquifer is treated as isotropic and homogeneous. Second, a typical conductivity for alluvial sand combined with gravel [2] can be chosen. Third, the horizontal conductivity is equal in each direction and the vertical conductivity is chosen ten times smaller because usually vertical flow is slower. For the parameter  $S_0$  the same assumption is made and thus one value is used for the whole model. As a next step the model layers are to be defined. To better reflect the effect of ground water flow we use 3 layers with same parameters. We assume that the recharge only occurs in the first model layer.

To solve the partial differential equation (1) boundary conditions are necessary. We assume the head from the river as a constant one, and a no-flow boundary condition to the neighboured aquifer, which has only a little recharge to the aquifer. Due to

cutting the aquifer we define an inflow and outflow to the neighbour aquifer in north and south. As extraction parameters the two extraction sites are modelled by pumps with a depth of 70 m with an extraction of 160 m<sup>3</sup> per hour for each well.

A grid is defined based on a topographical map of the model area. We build up the grid which is refined near the river and extraction stations.

### **Calibration and Simulation**

After the parameterisation discussed above, the model calibration can be made. Here parameters are changed to fit best to the measured results using the method of least squares. Based on this result, a first model with the assumptions can be achieved. A detailed model will be developed step by step if new data are available.

One of the first steps was therefore the installation of data loggers which measure the level of ground water. The measured level can be used to calibrate the model. Based on this measurement data we also can determine more model parameter, e.g. through pumping tests. Acquiring data and improving groundwater simulation model are important tasks in the near future.

For modelling and simulation we use the software Visual MODFLOW [3]. It uses the technique of finite differences to solve the partial differential equations. The most advantage of MODFLOW is the open source. Also the input and output files can be accessed by other software tools. Thus the simulation model created with MODFLOW can be coupled to other software, e.g. optimisation software.

A steady-state simulation for the model area describing monthly water balance is being made. The next step will be the simulation based on a transient model which represents the real ground water flow. The results of steady-state as well as transient simulation will be presented.

### **Conclusions**

After a short introduction to the project and related facts the model area Darkhan was explained. The model area considered was described in detail. The concept of modelling is explained by making some assumptions because only a few data is available. Thus only a rough simulation model of the area Darkhan in Kharaa river basin is presented and simulation results will be presented.

Based on these results the model will be developed further and will become more detailed, so that it can be used for doing forecasts. Based on this detailed model of

the groundwater system, an interface to optimisation software for gaining optimal strategies for the water extraction will be established. This will allow analysing the potential of optimising the groundwater capacity and of the investigation of the water resource management in the area considered.

### **Acknowledgements**

The authors would like to acknowledge the support from BMBF (project number 0330764F) and extend their thanks to all Mongolian and German co-workers of the project.

### **References:**

- [1] Stone, William Jay: „Hydrogeology in Practice: A Guide to Characterizing Groundwater Systems“, Prentice hall, New Jersey, (1999)
- [2] Busch; Luckner; Tiemer: „Geohydraulik - Lehrbuch der Hydrogeologie Band 3“, 3. Auflage, Gebrüder Bornträger, Berlin, (1993)
- [3] Waterloo Hydrogeologic, Inc.: “Visual MODFLOW Professional: User’s Manual”, Waterloo Hydrogeologic, Inc., Waterloo (2006)
- [4] Jadambaa, Nanjiliin; Grimmelmann, Wolfgang; Kampe, Aribert:“Explanatory Notes for the Hydrogeological Map of Mongolia 1:1000000”, Ed.: Institute of Geology and Mineral Resources, Ulanbaatar, Mongolia, IGMR; Federal Institute for Geosciences and Natural Resources, BGR (2004)
- [5] Jadambaa, Nanjiliin; Grimmelmann, Wolfgang; Kampe, Aribert:“ Hydrogeological Map of Mongolia 1:1000000”, Ed.: Institute of Geology and Mineral Resources, Ulanbaatar, Mongolia, IGMR (2004)

### **Authors:**

Dipl.-Ing. Stefan Röll  
Dr.-Ing. Siegbert Hopfgarten  
Prof. Dr.-Ing. habil. 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, Germany  
Phone: +49 3677 69-2869, -1418, -1423  
Fax: +49 3677 69-1415  
E-mail: {Stefan.Roell,Siegbert.Hopfgarten,Pu.Li}@TU-Ilmenau.de



A. Khatanbaatar Altantuul

## The need designing integrated urban water management in cities of Mongolia

### SECTION 6.

Mongolia locates in Central Asia, has severe continental climate, insufficient stock of surface water supply for centralized system of urban areas. Average rainfall is in the range of 200 to 250 mm and significant part of this water evaporates or drains out and remained small part of the water feeds underground water. As a source of the centralized water supply only underground water, fed by water in stratum and their cracks. Water consumption depends on climate condition and living specifics of users.

But in recent years customary usage of water is changed by different reasons. The tendency to increase in the quantity is observed for all kinds of consumers, and it is especially noticed for municipal water supply, which is 70 percent of all consumed water.

In the opinion of experts, it became clear that the increase of consumed water in conditions of deficiency of water resources can be limited by number of actions including: implementing management of segment tariffs with norm of water consumption; reducing water loss; re-usage of water; forecasting of water need based on natural demography of the population. To reveal the causes of growing water consumption in the present situation of city Ulaanbaatar, natural experiments have been done. In this study, hourly measurements of the chosen factors ( $H$ ,  $Z$  are made,  $q_n$ ) were taken within several months (December, March, June, August) on 12 central thermal points that supplied 43 apartments and their 33633 inhabitants. The research work was conducted through the following stages:

1. Statistical sampling of objects (apartments) by the degree of well-being, which differ in height, population, pressure on input of water system, level of operation of system.
2. Continuous measurement of hourly charges of water, pressure in the system and amount of night consumption were taken on water-measured units.

Based on the findings, the mathematical model that shows relationships between basic influencing factors as the below:

$$Q(q_n, z, H) = a_0 + a_1 \cdot q_n^{p_{qn}} + a_2 \cdot z^{p_z} + a_3 \cdot H^{p_H} \quad (1)$$

In where:

$Q(q_n, z, H)$  = values of the daily specific consumption of water (liter per day-person)

$q_n, z, H$  = factors of influence on process of water consumption

$q_n$  - loss of water in system of water supply,

$z$  - population of apartments,

$H$  - water tariff

$p_{q_n}, p_z, p_H$  - degrees of corresponding modeling factors which are from literature review equals as a rule  $p_{q_n} = p_z = p_H = 1$ .

Using the method of the least squares, from the average values of the water consumption and major factors, coefficients and degrees of given regression equations (1) were determined as showed below:

$$a_0=120.96 \quad a_1=26.45 \quad a_2=0.576 \quad a_3=0.175 \quad p_{q_n}=0.5 \quad p_z=3 \quad p_H=1.5$$

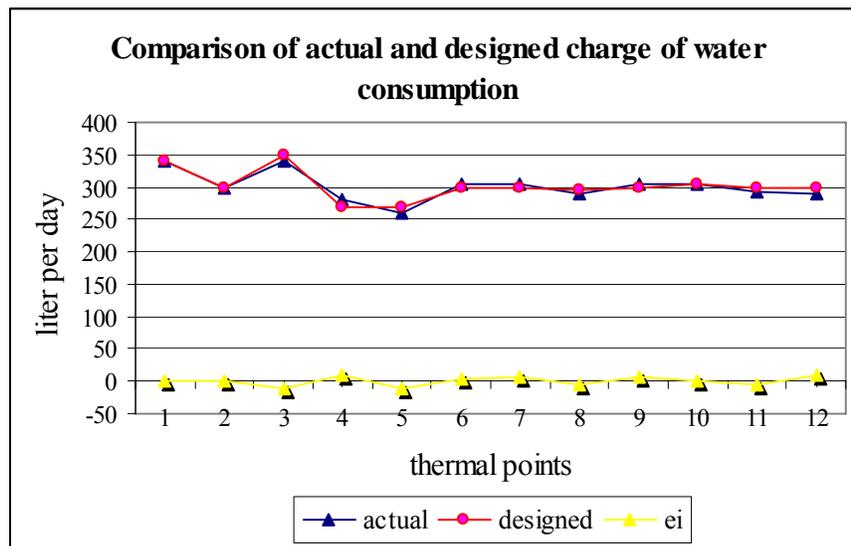
Finally, mathematical model that describes specific average daily water consumption becomes:

$$Q(q_n, z, H) = 120.96 + 26.45 \cdot q_n^{0.5} + 0.576 \cdot z^3 + 0.175 \cdot H^{1.5} \quad (2)$$

Values of  $Q_{fi}$ - factual average daily water consumption and values of  $Q_i$ , calculated through given model ( 2 ) are shown in Figure 1, and values are no viscous:

$$e_i = Q_{fi} - Q_i$$

**Figure 1**



The table below shows test results of coefficients  $a_0, a_1, a_2, a_3$ .

**table 1**

|   |               |              |              |              |
|---|---------------|--------------|--------------|--------------|
| <i>Coefficients at factors – <math>a_j</math></i>                           | <i>120.96</i> | <i>26.45</i> | <i>0.575</i> | <i>0.175</i> |
| <i>Estimations of quaintly distributions of Students – <math>t_j</math></i> | <i>8.513</i>  | <i>2.08</i>  | <i>0.041</i> | <i>0.012</i> |

From the table it is visible, that  $a_0$  and  $a_1$  are significant only. Therefore, generally it is possible to use simplified model ( 3 ), in which  $\Phi (q_n)$  – dependence function of water consumption on water loss ( $q_n$ ):

$$\Phi(q_n)=120.96+26.45 \cdot q_n^{0.5} \quad (3)$$

The empirical conclusion of significant water loss in municipal sector is also approved by the balance report of activities of the water supply and sewerage authority of city Ulaanbaatar.

**table 2**

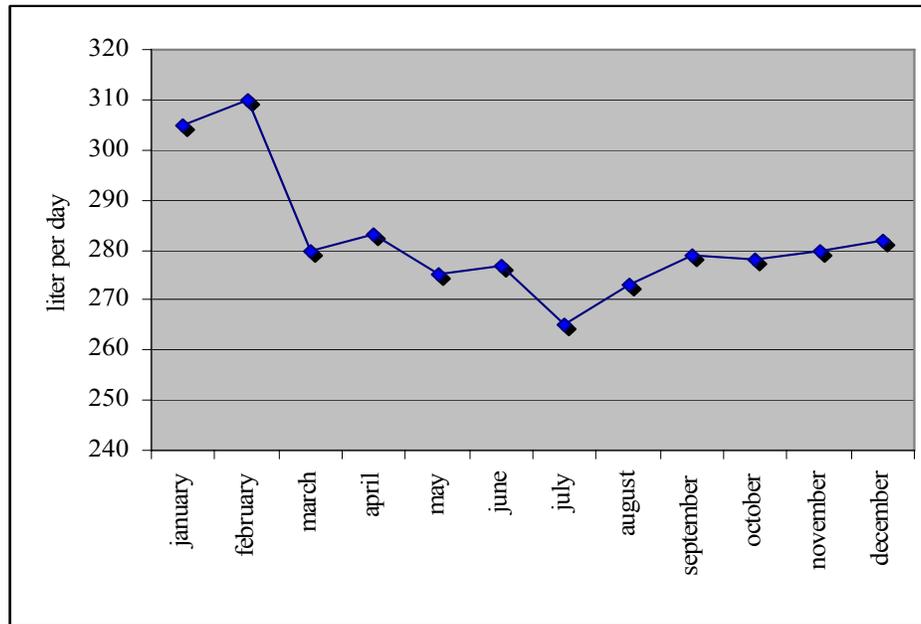
*The amount of loss water in percentage*

| 1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 |
|------|------|------|------|------|------|------|------|
| 56.6 | 34.4 | 34.3 | 35.7 | 34.7 | 24.6 | 26.5 | 29.1 |

On the basis of the taken measurements some prominent features of water consumption of population of city Ulaanbaatar are determined:

- Water consumption during cold season is higher than that of warm season. This can be explained by the specifics of lifestyle of the population, custom of water usage and fluctuating demography of the city population (Figure 2)

Figure 2



In Mongolia's situation, the change of the water consumption of population also can be explained by its dependency on the change of air temperature. Relationship between specific water consumption and air temperature can be stated as below:

$$Q(T)=266,18-0,646\cdot T+0,04\cdot T^2 \quad (3)$$

In which,  $T$ -an ambient temperature.

Analysis of full-scale investigation and received empiric dependence shows a necessity of adjusting the following factors having an influence on water consumption.

Firstly, loss of water in the water supply system basically caused by unadjusted pressure in the water supply system, technical status of system and lack of real calculation of water by consumers.

Secondly, occupancy of houses or free occurrence number of consumers which are changeable due to seasons of year. According to statistics, population of Ulaanbaatar has been increasing by 3% a year which causes difficulties in water supply.

Thirdly, tariffs of water at the present time requires an urgent solution of problem. Distinctive features of water supply to the populated area of Mongolia are classified in to 2 types: stationary and delivery. About 40 percent of population of Ulaanbaatar are fed from central water supply / 1 cube meter is valued at 300 tugrig\*, a remained share of water supply felt to delivery supply of which tariff is 500 tugrig/ cube meter. Tariffs in connection with segmentation of service market will assist to settle a lot of problems of water supply.

Fourthly, Establishment of fixed rates of water to the houses connected with centralized water supply. Recently, rate of water allocation which is 230 liters per one person a day and night acts officially throughout capital city. It acts for people who are fed from centralized water supply. In comparison with world standard, this rate is overvalued. 120 liters a day and night are taken a priori by us (formula 1).

The specialist have determined that one of the ways of economy of water in housing sector of urban area is decreasing an amount of consumable water to technologically required level in other words to social rate of water consumption by population. But water need of everyone is individual's "accidental quantity" and stipulated for influencing a lot of factors which predetermine different kinds of study of this problem.

At that time, by bringing a rate of water supply close to world standard we can optimize hydraulic calculation of network, make forecast for amount of water in the plan of resource-economy.

Current situation and analysis of the conducted full-scale investigation show a necessity of implementing an integrated management in the sector of water supply to the cities of Mongolia.

---

\* tugrig- Mongolian currency

#### References:

1. Works Novosibirsk State University of Architecture and Civil Engineering . p 49-54, Russia "A condition of water distribution to the population in Ulaanbaatar City" 2000.
2. News of VYZ. Civil engineering-, №12 p. 63-66, Russia "Definition of a model of water consumption on the basis regression analysis of the population" 2003
3. 18-21 November 2003 , Kemerovo, p. 86-88, Russia "The influence of operational characteristics on the optimization calculations for water supply and distribution systems" 6-th the international scientific - practical conference.

Author:

Mrs. Khatanbaatar Altantuul

Mongolian University of Science and Technology  
Peace Avenue, P.O.B-1419  
211213, Ulaanbaatar, Mongolia

Phone: 976-99747498 (m)

Fax: 976-342885

e-mail: [atuul70@yahoo.com](mailto:atuul70@yahoo.com)



Thomas Rauschenbach / Torsten Pfützenreuter / Zhang Tong

## **Model based water allocation decision support system for Beijing**

### **INTRODUCTION**

Beijing is a fast growing city. For several years the total population has increased steadily as well as the economy has grown very fast. The trend will continue in the future. An increasing demand of several resources, especially water, follows from this development. In connection with the semi-arid climate, these are the causes for the scarcity of the resource water in Beijing. Without consideration of sustainable resources use, the very positive development of the region is endangered.

Central management of all usable water resources for the city is urgently needed. That is why the Beijing Water Authority (BWA) is developing a "Capital Water Resources Allocation Decision Supporting System" to assist in the management of all the water resources of the capital of the People's Republic of China.

The general objectives of the joint Chinese - German project "Toward Water-Scarcity Megalopolis's Sustainable Water Management System" are as follows:

- To build rainfall - runoff models for the catchment areas of the most important reservoirs in the Beijing region with the aim to improve precision of runoff forecast.
- To establish water demand prediction models for a large city to solve the problem of water demand prediction under different circumstances.
- To built the joint operation model of multi water resources including surface water, groundwater, recycled water and water diversion as well as to primarily solve the allocation problem of Beijing water resources.
- To primarily establish operation-orientated Beijing water resources integrated allocation decision supporting system (DSS) and provide techniques for water resources management and decision-making.

The project is under the charge of BWA, German Fraunhofer Institute for Information and Data Processing and Fraunhofer Center for Applied Systems Technology. The project will be completed until 2008.

The DSS will realize an optimal allocation of existing water resources such as surface

water, ground water, recycled water and transferred water. That means an multiple optimization problem has to be solved [5]. Therefore, a goal function and a simulation model of the water resources as well as of the water supply system are necessary. In the first phase of the project the simulation model of the Beijing water supply system will be developed. Starting point for this development is a rough model for the simulation of the surface and groundwater resources. This model meets the demands for a decision support system with respect to accuracy and simulation speed [6]. The model takes account of

- the Miyun, Guanting, Huairou and Baihebao reservoirs and their catchment areas,
- the groundwater storage,
- the rivers and channels linking the reservoirs to the city and
- the pattern of consumption in households, industry and agriculture.

This paper presents the simulation model and the first experiences with this model as well as a first definition for a goal function for use in the DSS.

The authors thank the German Ministry of Education and Research (BMBF) and the Chinese Ministry of Science and Technology (MOST) for funding this project.

**STRUCTURE OF DECISION SUPPORT SYSTEM**

The objective of the project is to establish an integrated intelligent water resources allocation decision support system based on GIS. It uses advanced computer and network technology and it implements a man-machine-interface between decision makers and the system. Fig. 1 shows the logical structure of the DSS which will be developed.

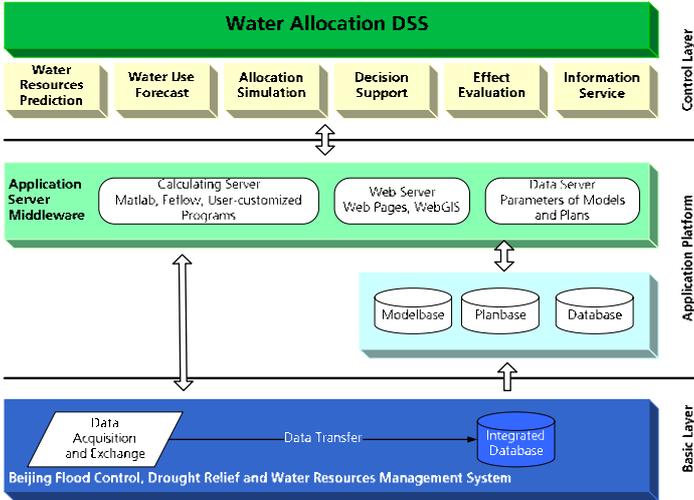


Fig. 1 The logical structure of the "Capital Water Resources Allocation Decision Supporting System" The DSS is designed as a distributed system, which consists of three layers. These are the basic layer, the application layer (application platform) and the control layer. In the

basic layer, the functions for data management and data storage are concentrated. The simulation models and the optimization strategies for decision-making are located in the application layer. Man-machine-interface is provided by the control layer. From here, the locally distributed users can access data and simulation results as well as optimization results. Therefore different rights of access can be assigned to the users.

### STRUCTURE OF THE WATER SUPPLY SYSTEM

The structure of the system is shown in Fig. 2. All essential parts of the Beijing water supply system will be considered in the model [1]. There are the four reservoirs Miyun, Huairou, Baihebao and Guanting. The catchments area models are integrated in this system in order to take into account the precipitation and the evapotranspiration. Further sources are groundwater storages. Furthermore there are the water transportation systems, the waterworks and the customers.

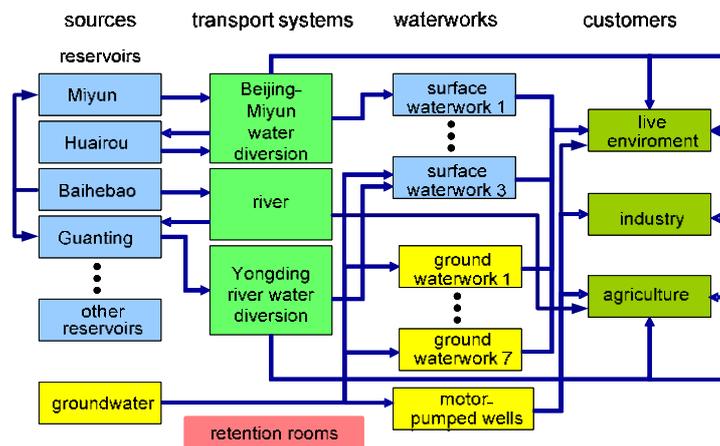


Fig. 2 Structure of the Beijing water supply system

### SAMPLE SIMULATION

To prove that the simulation model works, a rough approximation of the Beijing situation was taken as a sample for simulation. The Library “ILM-RIVER” is used for simulation [3]. Fig. 3 shows time series of precipitation and evapotranspiration for the year at hourly intervals. The sum of precipitation for the year is 578 mm.

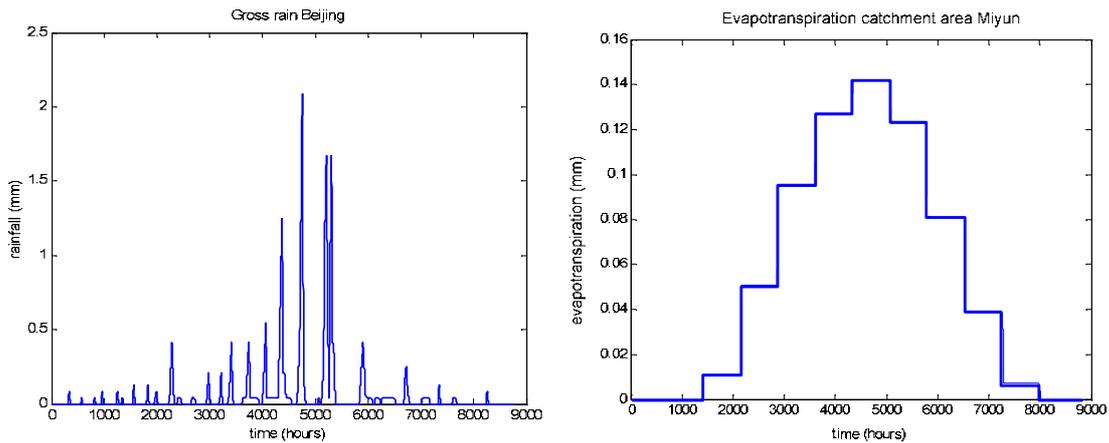


Fig. 3 Assumed annual rainfall and evapotranspiration for the Beijing area

Some simulation results shall be shown now. The water flowing from the Beijing catchment area in the form of groundwater flow as well as the water level in the groundwater storage are represented in Fig. 4. The maximum of the catchment output is postponed by approximately 1000 hours from the time of maximum precipitation and considerably smoothed. At the beginning of the year the water level is still 20 m. At the end it is 19.38 m and so has dropped by 62 cm a year. Changes to the water table are positive under the influence of additions from the catchment area and from channels and the river network, and negative under the influence of outflow to waterworks and pumping stations.

Fig. 5 shows the water inflow into the Miyun reservoir and the appropriate water level. The peaks of the inflow can get up to practically 1000 m<sup>3</sup>/s. The inflow from the groundwater is characterised by a large delay-time and dead-time. The figure taken for the beginning of the year was 24 m, the maximum level. At the end of the year the reservoir does still have a level of 22.5 m. The downward slope in the graph between 0 and 4300 hours is to be attributed to the effect of evaporation, of water diversion into the Beijing-Miyun channel at 50 m<sup>3</sup>/s, and into the river Chaobai at 30 m<sup>3</sup>/s, and of the demand from the No. 9 waterworks at approx. 10 m<sup>3</sup>/s. The little additional groundwater coming from the catchment areas fails to compensate for these outflows. The start of the rainy season becomes apparent at 4300 hours. The inflows are then greater than the outflows. The water level rises until about at the 7000th hour. After that, the outflow again dominates and the level at the end of the year is 22.5 m. If the outflow regime selected is continued, a similar inflow figure over 5 years would mean that considerable restrictions must be set on consumption.

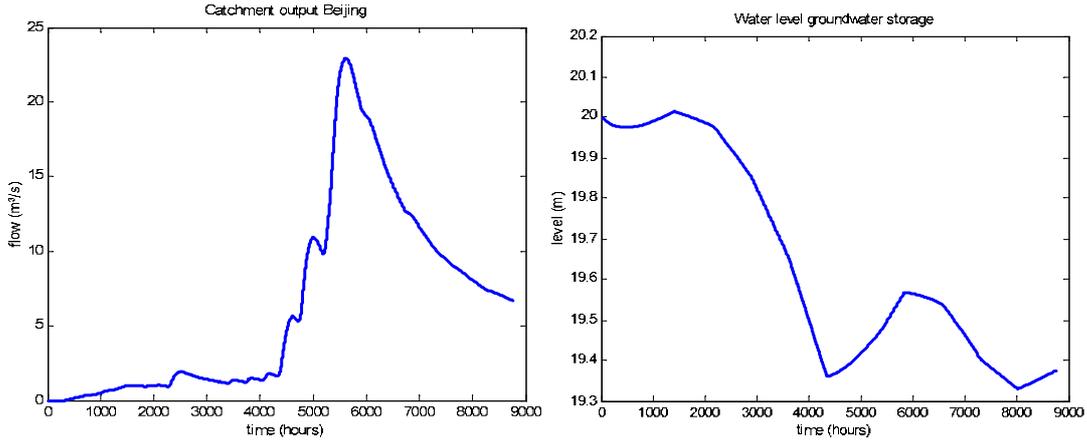


Fig. 4 Inflow of the city catchment area into the groundwater storage of the city region and the water level of this groundwater storage

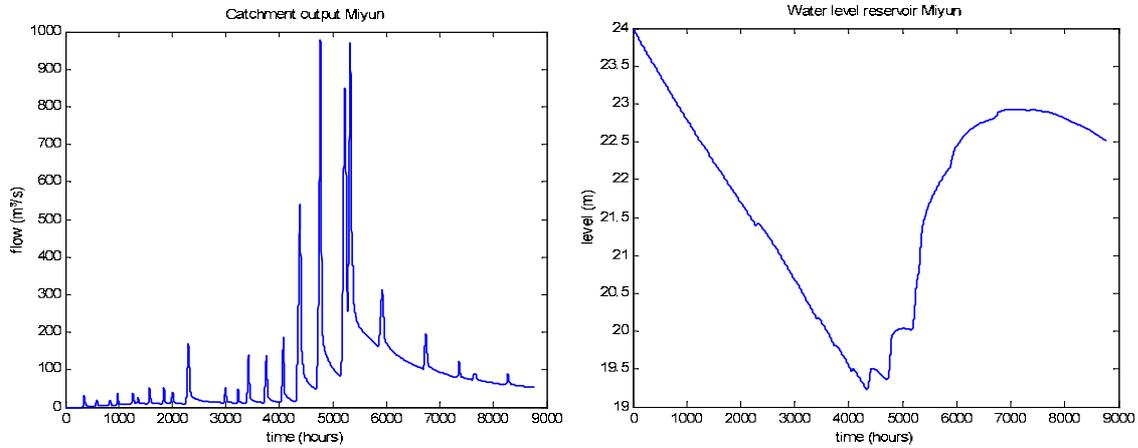


Fig. 5 Runoff from the catchment to Miyun reservoir and water level in the Miyun reservoir

### GOAL FUNCTION FOR OPTIMIZATION

The essential aims of the project are to prevent the water resources (especially groundwater) from decreasing and to guarantee the water supply of households, industry and agriculture. In order to achieve these goals a multiple criteria optimization problem has to be solved [2], [4], [7]. A first approach for a goal function  $I(t)$  to be minimized is defined as follows:

$$I(t) = \alpha_1 \cdot I_1(t) + \alpha_2 \cdot I_2(t) + \alpha_3 \cdot I_3(t) + \alpha_4 \cdot I_4(t) \quad \text{with} \quad \sum_{i=1}^4 \alpha_i = 1. \quad (1)$$

In eq. (1) the sub criteria  $I_i(t)$  have the following meaning:

1. The groundwater level change  $\Delta H_G(t)$ :

$$I_1(t) = \begin{cases} \min(-\Delta H_G(t)) & \text{for } 5 \text{ to } 9 \text{ years } (\Delta H_G(t) \leq 0) \\ \Delta H_G(t) & \text{for } 10 \text{ to } 14 \text{ years } (\Delta H_G(t) \approx 0) \\ \max(\Delta H_G(t)) & \text{for } \geq 15 \text{ years } (\Delta H_G(t) > 0) \end{cases}$$

2. The supply deficits of households  $D_H(t)$ , industry  $D_I(t)$  and agriculture  $D_A(t)$  in the following two possible forms:

a. Mean deficit:

$$I_{2,3,4}(t) = D_{H,I,A}(t) = \frac{1}{n} \sum_{i=1}^n (d_i - s_i)^+$$

$$\text{with } (X)^+ = \begin{cases} X & \text{for } X > 0 \\ 0 & \text{for } X \leq 0 \end{cases}$$

with  $d_i$  demand per time unit and  $s_i$  supply per time unit.

b. Days or months with deficit:

$$I_{2,3,4}(t) = D_{H,I,A}(t) = \sum_{j=1}^n g^j$$

$$\text{with } g^j = \begin{cases} 1 & \text{for } s_i < d_i \\ 0 & \text{otherwise.} \end{cases}$$

In the next phase of the project, this goal function will be used for the optimization of water resources allocation in Beijing. That means the DSS delivers optimal trajectories (in the sense of the goal function) for the control of reservoirs, sluices pumping stations etc. An essential precondition for this way of proceeding is the simulation model above described.

#### References:

- [1] BWA (2003) The introduction of Water Resources in Beijing and the Demand Analysis. Technical Report, Beijing Water Authority, Beijing, P.R. China.
- [2] Ehrgott, M. (2005) Multicriteria optimization. 2<sup>nd</sup> ed. Springer, Berlin.
- [3] Pfuetszenreuter, T. & Rauschenbach, Th. (2005) Library ILM-River for simulation and optimal control of rivers and hydropower plants. In: River Basin Management III (ed. by C. A. Brebbia), 277-285. WIT Press, Ashurst Lodge, Ashurst, Southampton, UK.
- [4] Rauschenbach, Th. & Gao, Z. (2005) Development of the "Capital Water Resources Allocation Decision Supporting System" for the city of Beijing. In: River Basin Management III (ed. by C. A. Brebbia), 277-285. WIT Press, Ashurst Lodge, Ashurst, Southampton, UK.
- [5] Rauschenbach, Th. (2001) Simulation and Optimal Control of Rivers and Hydropower Plants. In: Proceedings of the IASTED International Conference on Intelligent Systems and Control (ed. by M. H. Hamaza), 85-89. Clearwater, Florida, USA.
- [6] Rauschenbach, Th. & Wernstedt, J. (1999) ILM-RIVER - A General Design Tool for Simulation and Control of Rivers and Hydropower Plants. In: European Control Conference ECC'99, Karlsruhe, Germany.
- [7] Soncini-Sessa, R. & others (2000) Use of Multi-criteria Analysis to Resolve Conflicts in the Operation of a Transnational Multipurpose Water System – The Case of Lake Verbano (Italy – Switzerland). In: International Water Resources Association, Water International, Volume 25, Number 3, 334-346.

#### Authors:

PD Dr.-Ing. habil Thomas Rauschenbach  
Dr.-Ing. Torsten Pfuetszenreuter

Fraunhofer Center for Applied Systems Technology, Am Vogelherd 50  
98693 Ilmenau / Germany  
Phone: +49-3677-461124  
Fax: +46-3677-461100  
E-mail: [thomas.rauschenbach@ast.iitb.fraunhofer.de](mailto:thomas.rauschenbach@ast.iitb.fraunhofer.de)

Prof. Zhang Tong

Beijing Institute of Water, No. 21 Chegongzhuang West Road of Haidian District  
Beijing 100044, P.R. China  
E-mail: [zt@bjwateri.com](mailto:zt@bjwateri.com)

T. Pfützenreuter / T. Rauschenbach

## Surface Water Modelling with the Simulation Library ILM-River

### ABSTRACT

The MATLAB / Simulink toolbox ILM-River is a library for simulation and controller design of surface water systems consisting of rivers, channels, reservoirs and hydropower plants. Started with simulation objects for run-of-river reservoirs and river sections, in the last time the library was enhanced with models for complex surface water systems, chemical pollution propagation and an interface to the finite element groundwater simulator FEFLOW.

### INTRODUCTION

Designing controllers for reservoirs or channels typically requires an analytical model of the system. Beside the mathematical calculation of the controller law and the associated parameters, simulations methods are often used to test and optimize control strategies. This requires a model of the controlled process.

The toolbox ILM-River contains models for the most important elements of surface water systems and is used in different projects. It consists of two modules called River-MOD and River-CON.

River-MOD implements simulation models needed for modelling channels, reservoirs or river sections with their special characteristics. The control-oriented hydrodynamic models developed for River-MOD are well suitable for the design and the verification of control strategies. With the aid of a graphic editor, it is straightforward to build the entire model for one or more reservoirs or river sections.

The newest additions to the library are a set of simulation models for complex surface water systems, an interface to the groundwater simulator FEFLOW allowing the combined simulation of surface and ground water resources and the possibility to simulate chemical pollution in water bodies. These enhancements will be presented beside the already existing elements in detail.

The module River-CON implements different control concepts for reservoirs. Beside traditional control according to operating rules, several advanced concepts are integrated:

- a fuzzy control strategy for smoothing of discharge,
- a concept for the coordinated control of multiple river sections (cascades) to improve the flood control.

These methods allow to test and compare different strategies for controlling the output of reservoirs and river sections.

As examples, this paper presents the simulation model for a run-of-river reservoir and the propagation of benzene entering a river. Both pollution types, the continuous and the on-time pollution, are shown with their simulation results.

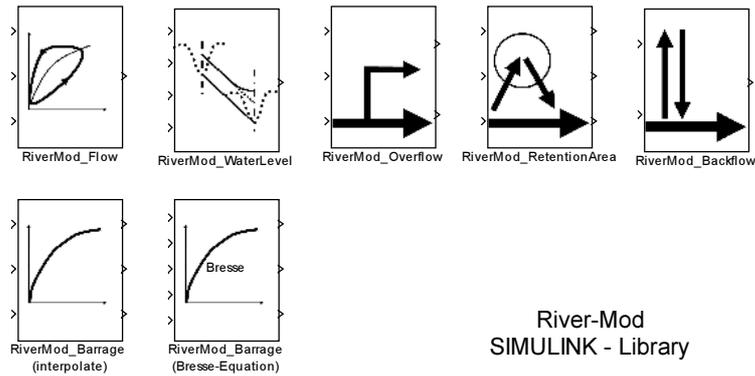


Fig. 1 The run-of-river simulation library

## MODEL LIBRARY RIVER-MOD FOR SURFACE WATER MODELLING

### Run-of-river Reservoir Modelling

The simulation models for the run-of-river part of River-MOD rely on an analytical description of the hydrodynamic behaviour of reservoirs. For control, it is sufficient to know the behaviour at gauges for water level and flows. Typically, simulation models describing instationary hydrodynamic behaviour of rivers and channels base on the Saint-Venant-equations [Abbott 1998]. However, such models are not suitable for an on-line optimization, because solving the differential equations system requires a lot of computational time. Furthermore, it is difficult to simulate overflow regions and backflow behaviour with these models. That's why new control-oriented hydrodynamic simulation models were developed, which meet all demands with respect to accuracy and simulation speed. More information on the particular blocks shown in fig. 1 is given in [Pfuetzenreuter 2005].

### Complex Surface Water System Modelling

Modelling a complex surface water system often results in a detailed simulation model with hundreds of parameters and a long simulation duration. The new development for River-MOD reduces the modelling effort by using simplified simulation blocks. Using these blocks results in a fast, reliable simulation model, that is sufficient for control and optimization tasks [Rauschenbach 2005].

The library contains the most important elements of a typical surface water system:

- Catchment areas: advanced models based on methods introduced by Lorent and Gevers [Lorent 1976],
- River and channels: elements like river and channel sections, weirs, sluices, control gates with their dynamic behaviour,
- Water supply systems: ground and surface water works, customer demand estimation
- Groundwater: groundwater storage models, interface to finite element simulator FEFLOW.

Using these elements, major parts of surface water systems can be modelled (fig. 2). Typically, the data driven parameterization is done using optimization techniques. All elements of this library are able to simulate the chemical pollutant transport and elimination processes described in the next section.

### Chemical Exposure Modelling

Chemical pollutants reaching water bodies often represent risks for people and environment. That's why it is necessary to model transport, sedimentation and

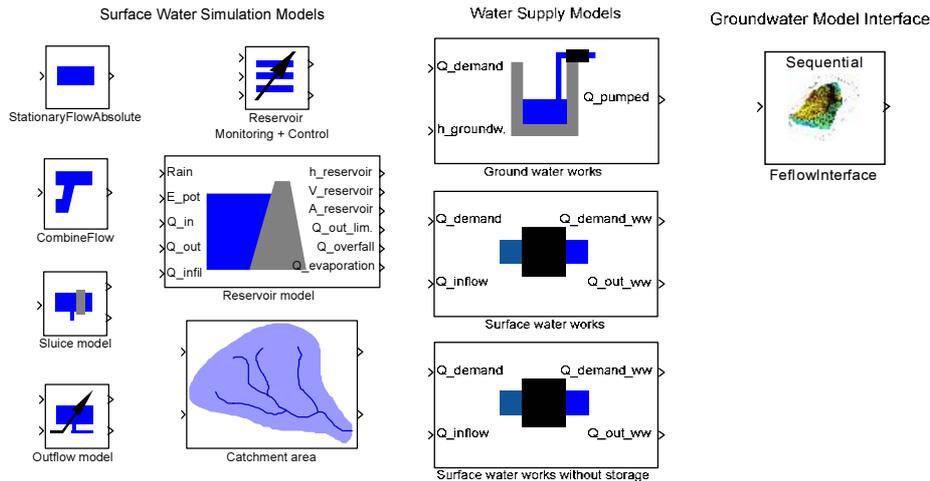


Fig. 2 New blocks of the River-MOD simulation library (only important blocks are displayed)

elimination of such substances.

The simulation blocks of the River-MOD library are able to simulate these processes and implement one-dimensional methods for simulation of two different scenarios:

- continuous pollution as consequence of natural and human life, industrial or agricultural production,
- one-time pollution, typically due to accidents or terrorist attacks.

The continuous pollution is modelled according to the methods developed by the GREAT-ER project as described in [Schowanek 2001]. The method assumes an instantaneous mixture of river and pollutant flows and a constant concentration along the cross section of the river. This leads to simple but efficient computation methods for the concentration down the river. The one-time pollution simulation was newly developed at Fraunhofer Center and takes into account the transport processes diffusion, dispersion and advection [Trapp 1997]. Both methods are designed with regard to a high computation speed. This allows a seamless integration of quality modelling aspects into existing simulation models.

### Coupling Surface and Ground Water Simulation

A current project at the Fraunhofer Center for Applied Systems Technology deals with the water allocation of the city of Beijing, China [Rauschenbach 2005]. One project emphasis is on the combined simulation of surface and ground water in the wider area around the city. The groundwater simulation is performed with the finite element simulator FEFLOW while the surface water, lumped parameter model runs on MATLAB / SIMULINK using the simplified simulation blocks. Coupling both models requires to analyze the physical interconnection and to develop an appropriate software concept.

The physical coupling concentrates on the most important interchange processes:

- groundwater withdrawal at wells,
- groundwater infiltration from rivers, lakes, reservoirs, irrigation,
- artificial recharge of groundwater at seepage fields.

The surface water simulation model computes all flow rates (withdrawal, infiltration, recharge) from its internal states. These time-dependent rates are transmitted to FEFLOW und used for simulation of a predefined period. As output, FEFLOW sends the hydraulic heads at all locations the surface water simulator is interested in. During simulation, the surface water model is responsible for starting and controlling the groundwater simulator as well as for the coordination of data transfer.

## MODEL LIBRARY RIVER-CON FOR RESERVOIR CONTROL

Especially run-of-river reservoirs are often built to use the environment-friendly hydropower for generation of electric energy. In most cases, this fact leads to the construction of reservoir cascades in rivers. The reservoirs in a cascade influence one another by hydraulic link-up. As a result, the natural flow behaviour is strongly affected by the control strategies of the reservoirs. This leads to an increasing demand of advanced control strategies for reservoir cascades.

Reservoir cascades have further tasks in addition to energy generation, e.g. to prevent overflows and to guarantee ship navigation. To avoid dangers for people and real values, the energy production has to subordinate to these tasks. Rules were created to support the operators in implementing the predefined strategies. Two modules of the model library River-CON realize a control according to such operating rules, a PID-controller with multiple parameter sets and a similar Fuzzy-concept. However, these concepts cannot optimally solve the mentioned multi-criterial tasks necessary for a coordinated operation of reservoir cascades. Instead, a newly developed concept called MEFURO is well suitable for such problems. Furthermore, a module for smoothing of discharge was integrated into River-CON. For more information to the particular control strategies see [Pfuetzenreuter 2005].

## SIMULATION MODEL APPLICATIONS

### Simulation Model for a Run-of-river Reservoir

Using the ILM-River toolbox different simulation models for river reservoirs were developed. As an example, this paper presents the model for the Danube reservoir Ybbs (fig. 3). Primary aim is to model the dynamic behaviour of the flow section between the two hydropower plants Wallsee and Ybbs. The model must describe the real inflow and discharge behaviour as well as the water level trajectories at the gauges [Rauschenbach 2001].

The total length of the reservoir Ybbs is approx. 34 km, the width ranges from 150 m to 1.000 m. The bottom slope in the reservoir is  $3,4 \cdot 10^{-4}$ . Downstream the gauge Au (in the middle of backwater storage) water bypasses the hydropower station Wallsee at flow rates higher than  $6000 \text{ m}^3/\text{s}$ . Hence, the downstream gauge of station Wallsee does not register the total flow to the reservoir Ybbs. For this reason, it is necessary to consider the effect of bypassing in the model. The activation of the retention rooms shown in fig. 3 occurs by opening floodgates and using pumping stations at a flow rate of  $4700 \text{ m}^3/\text{s}$ . Starting at a flow of  $5400 \text{ m}^3/\text{s}$ , the water enters the retention rooms directly. Fig. 4 shows the rough simulation model of the reservoir Ybbs.

With this model, good simulation results are achievable in all flow ranges. As an example, fig. 5 shows the simulated discharge of the hydropower station Ybbs during

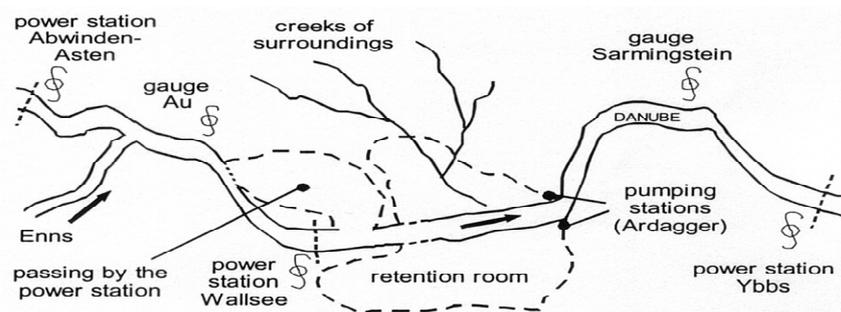


Fig. 3 River section between hydropower stations Abwinden and Ybbs



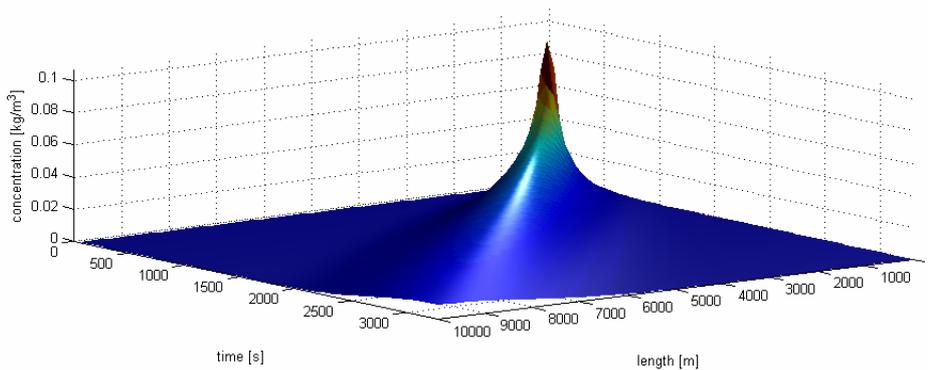


Fig. 7 Simulation of on-time pollution along a river section, length 10 km

interface to the finite element groundwater simulator FEFLOW. The chemical exposure modelling assists in estimating risks if pollutants enter a surface water system. The example of modelling a run-of-river reservoir in the Austrian Danube shows the practical usability of the simulation blocks.

The further development of the library is an ongoing work. Especially the pollution simulation methods have to be tested intensively. In addition, the models for the water supply network like pumping stations and waterworks as well as models for wastewater treatment are actively under investigation.

#### References:

- [Abbott 2005] Abbott, M. B. & Minns, A. W.: *Computational hydraulics*. Ashgate Publishing, Aldershot, Hampshire, UK, 2005.
- [Lorent 1976] Lorent, B.; Gevers, M.: *Identification of a rainfall-runoff process, application to the Semois river (Belgium)*. IVth IFAC Symp. on Identification, USSR, North Holland, pp. 735-744, 1999.
- [Pfuetzenreuter 2005] Pfuetzenreuter, T.; Rauschenbach, T.: *Library ILM-River for simulation and optimal control of rivers and hydropower plants*. In: *River Basin Management III* (ed. by C. A. Brebbia), 277-285. WIT Press, Ashurst Lodge, Ashurst, Southampton, UK, 2005.
- [Rauschenbach 2001] Rauschenbach, T.: *Simulation and Optimal Control of Rivers and Hydropower Plants*. In: *Proceedings of the IASTED International Conference on Intelligent Systems and Control* (ed. by M. H. Hamaza), 85-89. Clearwater, Florida, USA, 1999.
- [Rauschenbach 2005] Rauschenbach, T.; Gao, Z.: *Development of the "Capital Water Resources Allocation Decision Supporting System" for the city of Beijing*. In: *River Basin Management III* (ed. by C. A. Brebbia), 277-285. WIT Press, Ashurst Lodge, Ashurst, Southampton, UK, 2005.
- [Schowanek 2001] Schowanek, D.; Fox, K. et al.: *GREAT-ER: a new tool for management and risk assessment of chemicals in river basins*. *Water Science and Technology* 2(43), 179-185, 2001.
- [Thibodeaux 1996] Thibodeaux, L. J.: *Environmental Chemodynamics – Movement of Chemicals in Air, Water, and Soil*. John Wiley & Sons Inc., New York, USA, 1996.
- [Trapp 1997] Trapp, S.; Matthies, M.: *Chemodynamics and Environmental Modeling - An Introduction*. Springer Verlag Heidelberg, Germany, 1997.

#### Authors:

Torsten Pfützenreuter  
 Thomas Rauschenbach  
 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  
 thomas.rauschenbach@ast.iitb.fraunhofer.de

D. Karimanzira/ M. Jacobi

## **Modelling yearly residential water demand using neural networks**

### **ABSTRACT**

Demand management plays an increasingly important role in dealing with water scarcity in Beijing. It is important to understand the level and pattern of water use in various sectors across the regions for any measures being put into effect. The objective of this study is to enhance the understanding of the factors that influence water demand by examining closely the water use in the domestic sector and to develop a predictor for the yearly domestic water demand. Neural network techniques are investigated in an application to identify and subsequent offline prediction of a process "residential water demand" exhibiting nonlinearities (coupled determinants) and typical disturbances. The design and development of neural network process model from measured data is described, and practical aspects of the identification procedure are discussed. Results demonstrate that the developed neural network representation of the process behaviour is sufficiently accurate to be used independently from the process, emulating the process response from only process input information. Accurate long range predictions from the neural network model are mainly due to the use of a novel spread encoding technique for representing data in the network. Implementation of a predictive strategy incorporating the identified neural network model is described. Results illustrate the improvements in prediction performance that can be achieved when compared to regression models.

### **INTRODUCTION**

As water supply fails to meet the demand in many areas, careful analysis of decisions on the allocation of water is of great significance. The past policy responses to water scarcity are mainly supply oriented and aim at fostering the development and exploitation of new sources and expansion of the network infrastructure to guarantee the water supply. In recent year water policies have increasingly addressed demand management, which means development of water conservation and management programs to influence water demand. Demand driven measures include adoption of water-saving technologies and appliances, awareness raising and economic instruments such as price and tax. Population growth and urbanization and overall expansion in economic activities are the major factors underlying the increase in water consumption.

This work is part of the Chinese - German joint project "Towards Water-Scarcity Megalopolis' Sustainable Water Management System"[6]. This project takes the challenge of water shortage, the outstanding conflict between water supply and demand. It aims at a decision support system (DSS) for the sustainable development of economics and community in Beijing. An essential requirement for such a DSS is a simulation model of the water resources/supply system. The simulation model comprises the water supply, optimization and the water demand systems. On focus in this paper is

part of the water demand system, namely the domestic water demand. Compared to the agricultural and the industrial water demand is the domestic water demand very difficult to model, because it is determined by several subjective factors.

This paper proceeds as follows. Section 2 provides a brief introduction to modelling with neural networks. Section 3 describes the issues related to water demand estimation. Section 4 provides the results of the performance of the developed prediction system in comparison to regression techniques applied in several literature, followed by a short summary and conclusions.

## NONLINEAR MODELLING WITH NEURAL NETWORKS

In this work it has been chosen to restrict the attention to the so called multilayer perceptron neural networks (MLP) for modelling nonlinear processes. However, the particular choice of nonlinear model description is not vital for the predictor design. Other types of generic nonlinear model structures might be used instead.

Many different types of MLP based model structures can be considered when identifying nonlinear processes. However, it is typically assumed that the process can be described by the general model

$$y(t) = \hat{y}(t | t-1, \theta) + e(t) = g(\varphi(t), \theta) + e(t) \quad (1)$$

where  $\varphi(t)$  is the regression vector composed of the past information,  $g$  is the function realized by the MLP network,  $\theta$  is the model parameters (the weights), and  $e(t)$  is additive noise which is assumed to be white and independent of the past information. By inserting  $g$ , which is assumed to be a two-layer MLP network with tangent hyperbolic activation functions in the hidden units and a linear output unit, the predictor takes the form:

$$\hat{y}(t | \theta) = \sum_{j=1}^q W_j \tanh \left( \sum_{i=1}^{n_\varphi} w_{ji} \varphi(t) + w_j \right) + W_0 \quad (2)$$

where the components of  $\theta$ ,  $W_j$  and  $w_{ji}$ , specifies hidden-to-output layer and input-to-hidden layer weights, respectively. To reduce the degrees of freedom and because it is advantageous in many predictor system designs, it is common to consider model structures that are natural extensions of well-known linear model structures like *ARX* and *OE* in a similar regression vector is considered. The weights are estimated from a set of corresponding input-output pairs  $z^N = \{[u(t), y(t)]: t=1, \dots, N\}$  acquired in a practical experiment with the residential water demand process.

## DETERMINANTS OF THE RESIDENTIAL WATER DEMAND SYSTEM

Possible influencing factors for the domestic water demand for the Beijing region were selected and are listed in Figure 2 for the period from 1996-2005 [5]. They are factors from the weather, population and economy, which are thought to be obviously linked to the domestic water demand. The main objective of this study is to find out, which factor has the greatest influence and which ones are negligible in the models, and therefore simplify them. Correlation, regression and significance analysis described in the previous section were performed for the domestic water demand with respect to all input variables including the previous domestic water demand  $W_{prev}$ .

All estimated cross-correlation coefficients  $r_{xy}$  which were significant at the 5% level according to the three step procedure described above are summarized in Table 2 (bold) and Figure 3. Surprisingly, the magnitude of the total cross correlation coefficient of the domestic water demand and the number of employment is quite large (-0.777). A decreasing domestic water demand by increasing employment can not be explained logically. The partial correlation coefficient (a measure for the dependence of two variables after switching of the linear influences of other variables) between employment and the domestic water demand confirms this.

After switching of the linear influences of other variables the remaining partial correlation coefficient is only -0.065, which practically shows no linear dependence between the two variables. Also unexpected is the minimal correlation of the domestic water demand and the population (correlation coefficient of -0.016). Normally, one would think that the domestic water demand increases with increasing population growth. The correlation coefficients of the Beijing panel data were tested for significance and only the previous water demand, GDP per capita, employment, time and number of households (H) were significant at the 5% level (see Table 1). Therefore, it is recommended to include these variables as inputs in reduced forecasting models.

Regression results also show that several combinations of these variables are possible to obtain a reliable model. Most of the coefficients of the explanatory variables have expected signs. The positive value of temperature suggests domestic consumers use more water when the weather is relatively warm. Precipitation contributes negatively to water consumption, meaning that households tend to use less water when there is enough rainfall. Family size and water price (not shown) are not significant at any level, which may be due to the fact that both variables vary little with time.

|      | T            | P            | GDP          | H            | Tmp    | prec   | E             | W             | gdpc          |
|------|--------------|--------------|--------------|--------------|--------|--------|---------------|---------------|---------------|
| T    | 1.000        | <b>0.934</b> | <b>0.991</b> | <b>0.993</b> | 0.047  | -0.546 | 0.406         | 0.183         | -0.294        |
| P    | <b>0.934</b> | 1.000        | <b>0.928</b> | <b>0.947</b> | -0.199 | -0.413 | 0.548         | -0.016        | -0.537        |
| GDP  | <b>0.991</b> | <b>0.928</b> | 1.000        | <b>0.996</b> | 0.079  | -0.489 | 0.501         | 0.078         | -0.373        |
| H    | <b>0.993</b> | <b>0.947</b> | <b>0.996</b> | 1.000        | 0.002  | -0.486 | 0.499         | <b>0.676</b>  | -0.384        |
| Tmp  | 0.047        | -0.199       | 0.079        | 0.002        | 1.000  | -0.243 | -0.133        | 0.332         | 0.382         |
| prec | -0.546       | -0.413       | -0.489       | -0.486       | -0.243 | 1.000  | 0.069         | -0.500        | -0.199        |
| E    | 0.406        | 0.548        | 0.501        | 0.499        | -0.133 | 0.069  | 1.000         | <b>-0.777</b> | <b>-0.899</b> |
| W    | 0.183        | -0.016       | 0.078        | <b>0.676</b> | 0.332  | -0.500 | <b>-0.777</b> | 1.000         | <b>0.824</b>  |
| gdpc | -0.294       | -0.537       | -0.373       | -0.384       | 0.382  | -0.199 | <b>-0.899</b> | <b>0.824</b>  | 1.000         |
| Eva  | —            | —            | —            | —            | —      | —      | —             | <b>-0.726</b> | —             |

Table 1: Correlation coefficients of the variables

Important was also to find the robustness and the generality of the influencing parameters. Therefore, the correlation coefficient for the previous water demand, price, employment, time and number of households to domestic water demand were calculated for other regions of different nature where data could be obtained. Data for Germany and Canada was present and the calculated correlation coefficients were almost similar to that of Beijing,  $T$ ,  $E$  and  $H$  had the highest correlation with the domestic water demand, which suggests that models for forecasting domestic water demand that include these variables are quite reliable.

To avoid multicollinearity in the models, the population density  $P$  is excluded to be an explanatory variable as it is highly correlated with  $GDP$  and  $H$ .

## MODEL DEVELOPMENT AND TEACHING

### NN MODEL FOR THE RESIDENTIAL WATER DEMAND PREDICTION

It was found that the yearly water demand in the previous year, the employment rate, the number of households and the *gdp* were more significant variables affecting the yearly water demand in any given year [3]. Therefore, the neural network model consisted of four neurons in the input layer representing water demand in the previous year, the employment rate, the number of households and the *gdp*. The whole data set was divided into two parts, namely, “training set” and “testing set”. The training set was used to train the neural network model, whereas the testing set was used to test the performance of the neural network model in terms of some statistical parameters explained later.

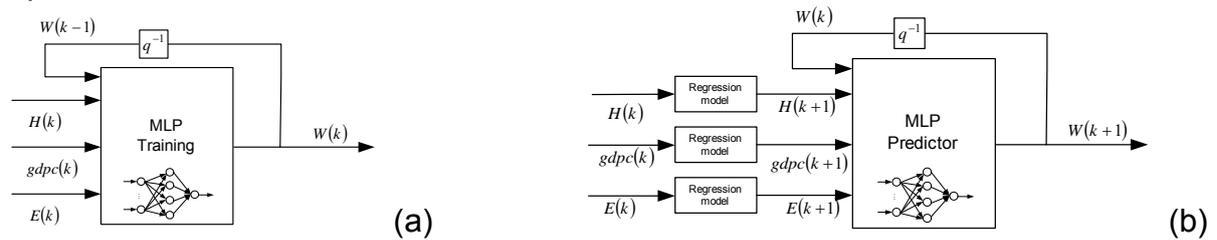


Figure 1: Structure of the neural networks (a) Training and (b) Prediction

### REGRESSION MODEL FOR THE RESIDENTIAL WATER DEMAND PREDICTION

A linear multiple regression model was also developed using the yearly water demand in the previous year, the employment rate, the number of households and the *gdp* as the explanatory variables. The structure of the model can be expressed as follows:

$$W = \alpha + \beta H + \gamma E + \delta W_{prev} + \mu gdp \quad (3)$$

where  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$  and  $\mu$  are the regression coefficients to be determined,  $H$  is the number of households,  $E$  is the employment rate,  $W_{prev}$  is the previous yearly water demand and  $gdp$  is the gross domestic product per capita. The linear multiple regression model was calibrated to determine the values of the regression coefficients using the same data set used for training the neural network model, and was tested using the same data set used to test the neural network model. Due to the fact that there could be some effects that are correlated to some explanatory variables the OLS is biased and inconsistent. Therefore, the parameter estimation was done using feasible generalized least squares analysis. The parameters in Table 2 were estimated for the regression model. Their standard deviation and coefficient of variation in % were also calculated.

| Parameter | Best Estimate | Standard Deviation | Coefficient of variation (%) |
|-----------|---------------|--------------------|------------------------------|
| $\alpha$  | 12.799        | 0.7511             | 586.817                      |
| $\beta$   | -8,9471E-4    | 0.0019             | -2.108.406                   |
| $\gamma$  | -6,8e74E-4    | 9,71115E-4         | -142.036                     |
| $\delta$  | -0.2142       | 0.486              | -2.269.412                   |
| $\mu$     | 0.0038        | 0.0016             | 41.849                       |

Table 2: Estimated parameters of the regression model

## RESULTS

### MODEL PERFORMANCE

Figure 2 and Table 3 shows the qualitative and quantitative results of the regression/neural network model, respectively. The sum of the squares of the residuals, the multiple correlation coefficients for I/O data and the linear correlation coefficient of measured output and the calculated output show very good results.

|   | Model results |           |
|---|---------------|-----------|
|   | Regression    | NN        |
| Sum of squares                                      | 0.0011        | 2.90e-007 |
| <b>Correlation: x - y data</b>                      |               |           |
| Multiple Correlation Coefficient                    | 0.9879        |           |
| <b>Correlation: y(experimental) - y(calculated)</b> |               |           |
| Linear Correlation Coefficient                      | 0.9879        |           |
| Linear Correlation Coefficient Probability          | 2.2028E-6     |           |

Table 3: Regression model results

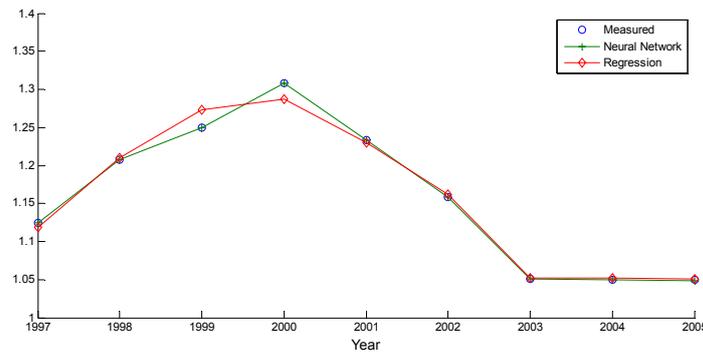


Figure 2: Results of the two models (Water Demand (100mil m<sup>3</sup>))

To compare the two models, the performance of each was quantified using two different statistical measures. The first statistical measure was “average absolute relative error” ( $f$ ). The average absolute relative error may be calculated using the following equations:

$$f = (1/N) \cdot \text{abs}(\text{sum}(RE_t)) \quad (4)$$

$$RE_t = \{(DO_t - DF_t) / DO_t\} * 100 \quad (5)$$

where RE, is the relative error in the forecasting during year t,  $DO_t$  and  $DF_t$ , are the observed and forecasted water demand during year t. The second statistical measure was the “threshold statistic”, which quantifies consistency in the forecasting errors from a particular model. The threshold static is calculated for a particular level of relative error (say, x %) in forecasting. A threshold statistic may be calculated using the following equation:

$$TS = \{n / N\} * 100 \quad (6)$$

where TS, is the threshold statistic for a level of x %, n is the number of data points forecasted having relative error in forecasting less than x %, and N is the total number of data points forecasted. In this study, the threshold statistics were calculated for levels of 1%, 2%, 5%.

| Level of threshold $x(\%)$ | Average Absolute Relative Error $f$ |          | TS (%)           |          |
|----------------------------|-------------------------------------|----------|------------------|----------|
|                            | Regression model                    | NN-Model | Regression model | NN-Model |
|                            | 0.0498                              | 0.0095   |                  |          |
| 1                          | -                                   | -        | 100              | 71.43    |
| 2                          | -                                   | -        | 100              | 100      |
| 5                          | -                                   | -        | 100              | 100      |

Table 4: Statistical measures for the models

## DISCUSSION OF RESULTS

The results in terms of various statistical measures obtained from all the models developed in this study are presented in Table 4. Observed and forecasted yearly water demands from the neural network and regression models for the testing period are shown in Figure 2. It is clear from the Table 1 that the neural network model performed the best in terms of and all levels of threshold statistics. The lowest average absolute relative error achieved by the best model was 0.0095%. In 71.43% of the forecasted water demands, the absolute relative error in forecasting was less than 0.3541%. All forecasted errors from the best model were less than 1.88 %.

## CONCLUSIONS

In the study presented here, a neural network and a regression model have been developed to forecast yearly water demand for Beijing, China. In this study several variables have been tested for their influence on the domestic water demand. It has been shown that to predict domestic water reliably at least the *gdpc*, the previous water demand, employment rate, the time and the number of households must be included. The estimation can be improved by using panel data covering a longer time period or more disaggregated sub-regional level analyses. It would also be useful to extend the study with more adequate data especially regarding time series water prices for the domestic sector. Well-designed household surveys would provide richer information and greater insights into the factors influencing domestic water demand. It was observed that the water demand in Beijing is a function of past water demand, the employment rate, the *gdpc* and the number of households. Based upon the results obtained in this study, the following conclusions can be made: the results obtained in this study are highly promising demonstrating the superiority of artificial neural networks over the conventional techniques such as regression analysis.

## References

- [1] Rosegrant, M., Ringler, C., Msangai, S., Cline, S. *IMPACT-WATER: Model Description*. Int. Food Policy Research Inst, Washington, D.C., 2005.
- [2] *Beijing Municipal Bureau of Statistics. Beijing Statistical Information net*, Mai 2005. <http://www.bjstats.gov.cn/english/index.html>.
- [3] Rauschenbach T. *Water Resources in Beijing and the Demand Analysis*, Juli 2005.
- [4] Brockwell, P.J. and Davis R. A. *Time Series: Theory and Methods*. Springer Verlag, New York. 1991

## Authors:

Divas Karimanzira  
 Marco Jacobi  
 TU-Ilmenau, Gustav Kirchhoff 1, P.O.B. 1005065  
 98684, Ilmenau  
 Phone: 03677691421  
 Fax: 03677691434  
 E-mail: divas.karimanzira@tu-ilmenau.de

Th. Westerhoff / B. Scharaw

# Model based management of the drinking water supply system of city Darkhan in Mongolia

## 1. Abstract

Within the framework of realizing the German BMBF-program objectives, Mongolia is proposed as a model region (MoMo) for the development, solution and implementation of integrated water resources management (IWRM). The working package 5 (WP) of MoMo project is dealing with the management of drinking water supply system of city Darkhan. The management of the water supply system is considered as three general working steps: modelling of the drinking water supply system (DWSS) calibration of the model and system parameters and management of the daily operating strategies as a decision support system.

## Key words

Model region Mongolia, drinking water supply system, pump stations, pipeline network, simulation, management

## 2. Introduction

The purpose of the water supply plan of Darkhan with a population of around 120,000 inhabitants is to define the development and management policy of the water supply facilities of the city in order to meet the expected requirements until 2020, including those likely to arise from the progressive urbanisation of the *ger* (or yurts) areas. One specificity of city Darkhan is the existence, beside a conventional urban area commonly referred to as *the core area* and equipped with all the facilities of modern cities, of large informal housing areas combining small houses and *ger* (Mongolian name for yurt). These *ger* areas regroup more than one half of the population of Darkhan, mainly the traditional and newly settled families and are not connected to DWSS. The DWSS structure in Darkhan is compared to other cities relatively simple. The tank area is located between the pump field and the city on a hill and the city on the opposite side of

this hill is supplied by the tanks without any additional pump in free fall. So the hydraulics calculation time can be reduced by subsumption of all demands in these loops to one sum demand for each loop and so reducing the network structure to a manageable amount of main pipes and hydraulic elements. The operation of inefficient pumps has resulted in some problems, including high

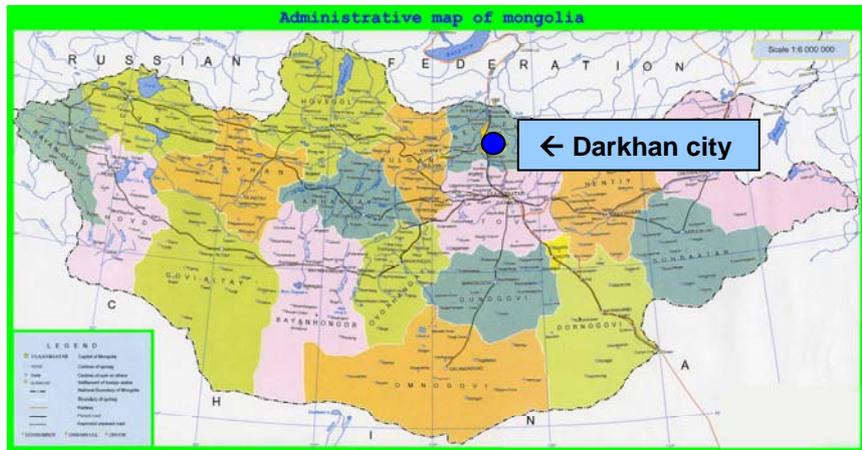


Figure 1: Map to show position of Darkhan city

electrical expenses and water supply shortages in *ger* areas. In addition, the existing water supply facilities remain in need of repair. Its water supply system consists of 225.3 km pipelines, 3 tanks and 18 well-pumps located one pumping field. The daily water demand is about 28.000 m<sup>3</sup>. The average elevation of the nodes is between 670 m and 780 m aSL.

In DWSS of city Darkhan we have several contrary goals of managing under the extreme climatic conditions of Mongolia. There is the planning for improvement of the technical situation and the expansion of the DWSS for *ger* areas. Therefore, it is desired to reduce the operation costs by optimal planning and operations of DWSS. Other task of WSS in Darkhan is to fulfill the water demand of the consumers. Different approaches have been proposed in the literature for simulation and optimization of DWSS. The software packet EPANET was used by Rossmann L.A. (1994), for simulating water networks and an augmented Lagrangian method was used for solving the NLP problem. Linear programming (LP) was proposed to solve large-scale WSS optimization problems (Diba et al. 1995, Sun et al. 1995). The optimization algorithms have been used for solving the optimal control of DWSS in Berlin (Burgschweiger, J., Gnädig, B. and Steinbach, M.C. (2004)).

### 3. Problem Formulation

The production and supply of drinking water for basic necessities of life requires a long-term careful approach to the resource water. To fit the permanent growing demands on

drinking water quantity and quality, a continuous work of practical orientated research is necessary for further development and application of new methods and their combinations. We consider dynamical management for DWSS with the purpose of planning the reconstruction of DWSS and minimizing its total operating costs. A typical WSS consists of source reservoir, from which water is pumped into pipe networks, of storage tank that floats on the system, of the various components of pumps and valves and of a collection of pipes connected to nodes. The water consumption at the nodes and the initial water level of tanks, the setting and parameters for pumps and valves are needed for the run of simulation. Such DWSS can be modelled by model-based decision support system, which can perform the hydraulic behaviour within pressurized pipe networks, the flow rate with velocity in each pipe, the pressure at each node and the water level in each tank in static and dynamic form under the given system conditions.

### 3.1 Hydraulic Formulation

The water from pump stations will be transported to consumers through the pipeline network. The operation task is to fulfill water demands and pressure restrictions at specific locations in the network, i.e.

$$Q_{\min l} \leq Q_l(t) \leq Q_{\max l} \quad l = 1, \dots, L \quad (1)$$

$$p_{\min k} \leq p_k(t) \leq p_{\max k} \quad k = 1, \dots, K \quad (2)$$

Decision variables in this problem are the outlet flow rates  $Q_i^P(t)$  and pressures  $p_i^P(t)$  of the pump stations. The pipeline network may be very large and complex. The formulation of relations between the input variables  $Q_i^P, p_i^P$  and output variables  $Q_l, p_k$ , leads to a large-scale nonlinear algebraic equation system. A feasible solution to the equation system means to look for proper  $Q_i^P, p_i^P$  so that (1) and (2) are satisfied. However, it is not a trivial task to find a feasible solution, since (1) and (2) are multiple constraints with a high dimension.

Let the flow-headloss relation in a pipe between nodes  $i$  and  $j$  be given as:

$$p_i - p_j = h_{ij} = rQ_{ij}^n + mQ_{ij}^2; \quad (3)$$

where  $p$  = nodal head,  $h$  = headloss,  $r$  = resistance coefficient,  $Q$  = flow rate,  $n$  = flow exponent, and  $m$  = minor loss coefficient.

The second set of equations that must be satisfied is flow continuity around all nodes:

$$\sum_j Q_{ij} - V_i = 0; \quad (4)$$

where  $V_i$  is the flow demand at node  $i$  and by convention, flow into a node is positive.

For a set of known heads at the fixed grade nodes, we seek a solution for all heads  $p_i$  and flows  $Q_{ij}$  that satisfy Eqs. (1) and (2).

The Gradient solution method begins with an initial estimate of flows in each pipe that may not necessarily satisfy flow continuity. At each iteration of the method, new nodal heads are found by solving the matrix equation:

$$\mathbf{JH} = \mathbf{F} \quad (5)$$

where  $\mathbf{J}$  = an (NxN) Jacobian matrix,  $\mathbf{H}$  = an (Nx1) vector of unknown nodal heads, and  $\mathbf{F}$  = an (Nx1) vector of right hand side terms.

The diagonal elements of the Jacobian matrix are:

$$A_{ii} = \sum_j \rho_{ij} \quad (6)$$

while the non-zero, off-diagonal terms are:

$$A_{ij} = -\rho_{ij} \quad (7)$$

where  $\rho_{ij}$  is the inverse derivative of the headloss in the link between nodes  $i$  and  $j$  with respect to flow. Each right hand side term consists of the net flow imbalance at a node plus a flow correction factor:

$$F_i = \left( \sum_j Q_{ij} - V_i \right) + \sum_j y_{ij} + \sum_f \rho_{if} H_f \quad (8)$$

where the last term applies to any links connecting node  $i$  to a fixed grade node  $j$  and the flow correction factor  $y_{ij}$  is:

$$y_{ij} = \rho_{ij} \left( r|Q_{ij}|^2 + m|Q_{ij}|^2 \right) \text{sgn}(Q_{ij}) \quad (9)$$

After new heads are computed by solving Eq. (5), new flows are found from:

$$Q_{ij} = Q_{ij} - (y_{ij} - \rho_{ij}(H_i - H_j)) \quad (10)$$

If the sum of absolute flow changes relative to the total flow in all links is larger than some tolerance (e.g., 0.001), then Eqs. (5) and (10) are solved once again. The flow update formula (10) always results in flow continuity around each node after the first iteration.

### 3.2 Pumping operation

The task of a pump station is to supply a given flow rate with a required pressure to the water network. We consider frequency-controlled pumps in parallel operation in a pump station. The time-dependent operating cost of pump  $j$  ( $i = 1, \dots, J$ ) in station  $i$  can be expressed as

$$Q_{i,j}^P(t) = \left( a_{i,j} + b_{i,j} Q_{i,j}^P(t) \frac{n_{i,j}^S}{n_{i,j}(t)} \right) \left( \frac{n_{i,j}(t)}{n_{i,j}^S} \right)^3 \quad (11)$$

Minimization the total operating costs of the pump station means to search for an optimal selection of pumps and optimal flow rates to be distributed on each individual pump. The total flow will be

$$Q_i^P(t) = \sum_{j=1}^J Q_{i,j}^P(t) \quad (12)$$

The flow rate of a single pump is a function of its pressure drop  $\Delta p_{i,j}$  and operating frequency  $n_{i,j}$  that can be described as

$$Q_{i,j}^P = \frac{n_{i,j}}{n_{i,j}^S} \left( \frac{1}{\beta_{i,j}} \left( 1 - \frac{1}{\alpha_{i,j}} \frac{\Delta p_{i,j}}{\Delta p_{i,j}^S} \left( \frac{n_{i,j}^S}{n_{i,j}} \right)^2 \right) \right)^{1/\gamma_{i,j}} Q_{i,j}^S \quad (13)$$

The parameters  $\alpha_{i,j}, \beta_{i,j}, \gamma_{i,j}$  and standard operation values  $Q_{i,j}^S, n_{i,j}^S, \Delta p_{i,j}^S$  in (10) can be obtained from the characteristic lines provided by pump manufacturers. The constraints in the operation of a pump include limitations of its flow rate, pressure drop and motor rotary frequency, i.e.

$$Q_{\min i,j}^P \leq Q_{i,j}^P(t) \leq Q_{\max i,j}^P \quad (14)$$

$$\Delta p_{\min i,j} \leq \Delta p_{i,j}(t) \leq \Delta p_{\max i,j} \quad (15)$$

$$n_{\min i,j} \leq n_{i,j}(t) \leq n_{\max i,j} \quad (16)$$

It should noted that constraint (16) is usually strict, since in practice the operating frequency is desired to be set near its standard value. It means the frequency ratio in (13) is often restricted by

$$u_{\min i,j} \leq \frac{n_{i,j}}{n_{i,j}^S} \leq u_{\max i,j} \quad (17)$$

During operation some pumps may be switched off, due to the varying total flow rate requirement. Thus, the flow rate of a pump should be zero, if it is not selected. To deal with this problem, an integer variable  $y_{i,j} \in \{0, 1\}$  for each pump has to be introduced ( $y_{i,j} = 0$  denotes the pump is off and  $y_{i,j} = 1$  it is on), such that

$$Q_{i,j}^P = \frac{n_{i,j}}{n_{i,j}^S} \left( \frac{1}{\beta_{i,j}} \left( 1 - \frac{1}{\alpha_{i,j}} \frac{\Delta p_{i,j}}{\Delta p_{i,j}^S} \left( \frac{n_{i,j}^S}{n_{i,j}} \right)^2 \right) \right)^{1/\gamma_{i,j}} Q_{i,j}^S y_{i,j} \quad (18)$$

This leads to a problem with (12) and (18) as equality and (14) and (17) inequality constraints. The objective function of this optimization problem is defined as the minimization of the total energy consumption of the pump station during the time period considered

$$\min f_i = \sum_{t=1}^{24} \sum_{j=1}^J \left( a_{i,j} + b_{i,j} Q_{i,j}^P(t) \frac{n_{i,j}^S}{n_{i,j}(t)} \right) \left( \frac{n_{i,j}(t)}{n_{i,j}^S} \right)^3 y_{i,j}(t) \quad (19)$$

where  $a_{i,j}, b_{i,j}$  are parameters in the correlation of power consumption of a pump.

Decision variables are the flow rate and operating frequency of each pump. The

pressure drop and total flow of the pump station are defined by solving the optimization problem of the water network. Another important issue to be considered is that frequent on-and-off of pumps sometimes is not allowed. This problem can be addressed by adding extra constraints or a penalty term in the objective function (19).

#### **4 Conclusions**

To solve such problems, the model based computer program HydroDyn is developing. HydroDyn allows the calculation of management strategies for pumps and valves while regarding the mechanical parameters, the behaviour of the consumers and the energy costs. Furthermore HydroDyn can be used for tracking the water network status (design of networks, calculation of hydraulic and quality in pressurized pipe networks and distribution system. We will apply the proposed approach to the management of reconstruction planning of the drinking water supply system of Darkhan city and to minimize the operation costs of large-scale water supply networks. Operating policies of water processing plants and pump stations can be developed by the proposed approach.

#### **References**

- Burgschweiger, J., Gnädig, B. and Steinbach, M.C. (2004), Optimization models for operative planning in drinking water networks, ZIB Report ZR-04-48, Zuse-Institut Berlin
- Diba A., Louie P.W.F. and Yeh W. W-G. (1995), Planned operation of large-scale water distribution system, *J. Water Resour. Plng. Mgmt.* ASCE, 121, 260-269.
- Putz, H., Scharaw, B. and Planke, H., (2001), Management of reservoir systems under quality requirements, Sixth international conference on computing and control for the water industry, pp. 479-489.
- Rossmann L.A. (1994), *EPANET Users Guide*, US Environmental Protection Agency, Cincinnati.
- Sakarya A.B.A. and Mays L.W. (2000), Optimal operation of water distribution pumps considering water quality, *J. Water Resour. Plng. Mgmt.* ASCE, 126, 210-220.
- Scharaw, B., (2005): Aufbau einer modellgestützten Bewirtschaftung der Trinkwasser- und Gasversorgung der HEW HofEnergie+Wasser GmbH, Jahresbericht 2005 Fraunhofer IITB, S: 48-49
- Scharaw, B., Westerhoff, Th., Putz, H. and Wernstedt, J., (2000), Model based management of a reservoir system, River Flood Defence, Volume 2, Kassel Reports of Hydraulic Engineering, pp. G-211-G216.

N. Buyankhishig / N. Batsukh

## Pumping well optimization in the Shivee-Ovoo coal mine Mongolia

### INTRODUCTION

The Shivee-Ovoo coal mine is located in Sumber sum of Gobi-Sumber aimag. Surface is hilly with elevation 1180-1230m above sea level (Fig 1). The study area is characterized by extremely continental climate with short hot summer and cold winter, seasonal and annual variation is high, always windy, and sometimes wind velocity reaches to 20m/s, dominant wind direction is from northwest. Total rainfall varies from mm in 150. Approximately 90% of the rainfall occurs between June and August. Average monthly temperature ranges from -3.0 °C in January to 15.7°C in August.

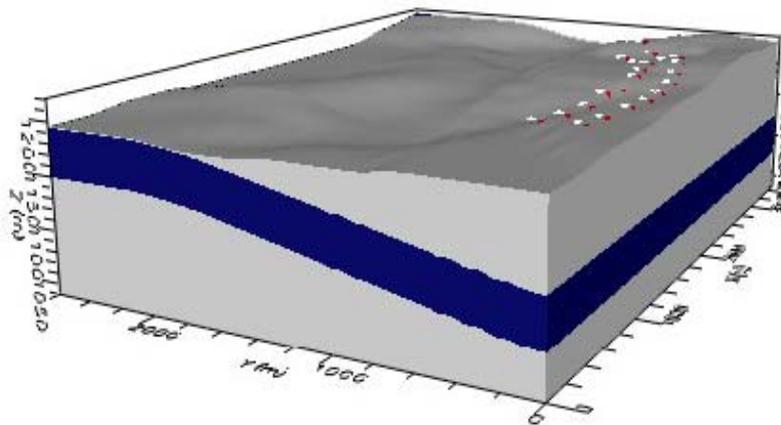


Fig 1. Coal deposit of the Shivee-Ovoo

#### Hydrogeological condition

Geologically the area belongs to Mesozoic formations, which are filled Choir graben basin, originated proterozoic geosynclinal and Dundgobi uplift zone. Foundation of Choir basin, coal bearing formation, cut by faults and consist of Jurassic volcanic sedimentary, Permian volcanic rocks [2].

Choir basin filled by Lower Cretaceous volcanic, sedimentary rocks and upper Cretaceous, sedimentary rocks, Shainshand formation. The Cretaceous rocks in the study area consists of, from oldest to youngest, the Bayanerkhet, the Tevshyn gobi and the Sainshand formations. Cretaceous rocks covered by Quaternary deposits. Tevshyn

Gobi formation, which belongs to the Lower Cretaceous, consists of alternating beds of sandstones, siltstone and coal. Sand grains range from fine to coarse and the degree of cementation vary. Upper Cretaceous rocks are distributed in high elevated areas, and consists of gray sand, brown to light gray clay and yellow gray sandstones. The groundwater in the aquifer occurs under confined condition that overlying confining layers of clay stone and siltstone [1]. Geologists of Central geological expedition carried out exploration in coal mine and evaluated coal resources in 1986 [1].

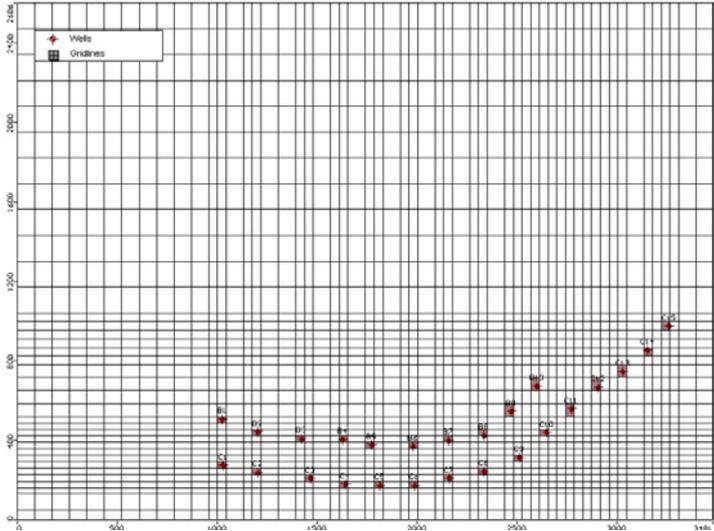


Fig 2. Dewatering well location

Based on aquifer test, Jadambaa.N determined hydraulic conductivity ranges from 0.0081 to 0.1m/d in siltstone and clay stone. Hydraulic conductivity in sandstone varies between 0.12-0.75m/d. For the model, hydraulic conductivity in overlying layer of coal bearing formation is taken averages 0.054m/d,  $6.24 \times 10^{-7}$  m/s. No one has investigated vertical hydraulic conductivities, although vertical hydraulic conductivities are likely lower than horizontal hydraulic conductivities in sedimentary rocks due to compaction. Vertical hydraulic conductivity is taken  $10^{-8}$  m/s based on previous investigation. Horizontal hydraulic conductivity of coal bearing formation is estimated by pump test value is  $5.035 \times 10^{-6}$  m/s, vertical hydraulic conductivity is  $10^{-7}$  m/s [2]. Geocological institute is estimated confined storativity of 0.0065 for pump test in the coal bearing formation.

According to the hydrostratigraphy and conceptual model, three layers are designed in the model. Each layer has 57 rows 36 columns for a total of 2052 cells in the model (Fig 2). Cell thickness depended on elevation of contact between the different layers.

The boundary conditions are constant head in coal bearing formations and uniform

recharge in upper boundary, whereas no-flow conditions were imposed on the lower boundary.

Report of hydrogeological investigation done by Geoecological Institute is notes that 6-7 percentage of precipitation recharges groundwater system. Recharge is 10.5mm/y.

Main purpose of the study is to determine pumping well rate in the Shivee-Ovoo coal mine. We used well optimization code of Visual MODFLOW Pro, a widely used finite difference flow code developed by Dr.Chunmiao Zheng and Patrick Wang from the University of Alabama [5]. Pumping well optimization was developed under steady state condition [4]. Objective of this optimization model was to maximize the total pumping rate. Strongly Implicit Procedure (SIP) and Slice-Successive Overrelaxation Package (SOR) solvers are used for solving numerical equation. Both solvers are solving a large system of simultaneous linear equations by iterations. Results solving by SOR and SIP methods are shown in table 1 and 2.

Table 1

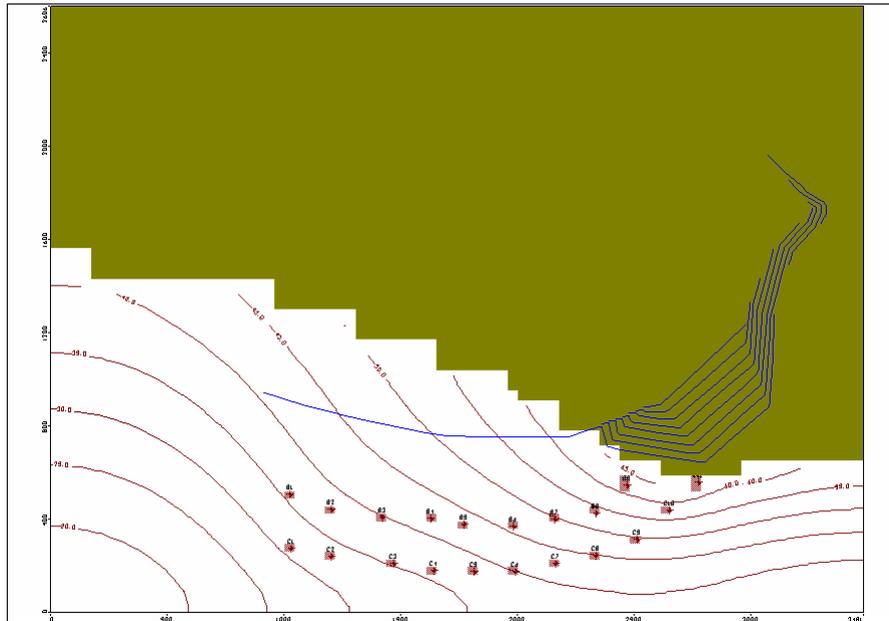
| Well                         | Ñ1   | C2   | C3  | C4  | C5   | C6  | C7  | C8  | C9  | C10 | C11 | C12 | C13  | C14 | C15 |
|------------------------------|------|------|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|------|-----|-----|
| Discharge, m <sup>3</sup> /d | 1000 | 1000 | 741 | 838 | 1000 | 967 | 768 | 903 | 967 | 838 | 999 | 903 | 1000 | 999 | 999 |

After optimization a new dewatering system was installed parallel to the B dewatering wells 160m. Distance between wells is 160m. When total extraction is 13 922m<sup>3</sup>/d drawdown will be 25-65m near open pit (Fig 3).

Table 2

| Well                         | Ñ1   | C2  | C3  | C4 | C5  | C6  | C7  | C8 | C9  | C10 | C11 | C12 | C13  | C14 | C15 |
|------------------------------|------|-----|-----|----|-----|-----|-----|----|-----|-----|-----|-----|------|-----|-----|
| Discharge, m <sup>3</sup> /d | 1000 | 871 | 870 | 0  | 709 | 741 | 935 | 0  | 516 | 999 | 806 | 96  | 1000 | 999 | 515 |

If 10057m<sup>3</sup> water pumps per day from 15 wells groundwater drawdown will be lowered from 10 to 40m.



S. Holzmüller-Laue / B. Göde / K. Rimane / N. Stoll

## Data Management for Automated Life Science Applications

Laboratory information management systems (LIMS) have been established for supporting project management, process planning, evaluation, and long-term management of the results in life science laboratories. The conventional function of a LIMS is the database-supported management of all business process information which results from laboratory operations [1-3]. The complete system integration of a LIMS covers recent topics of automated, bidirectional data communication with an alterable process control layer in laboratories, embedded high-parallel High Throughput Screening (HTS) / High Content Screening (HCS) process tracking, secure e-business, as well as embedded or off-line post processing, data mining or ERP and quality management.

The stream of data is continuously growing as a result of increasing laboratory automation, high throughput experiments, and modern analysis methods. Therefore, the central, user-friendly management of these data is of crucial importance for quality assurance and information availability.

Current challenges for information management solutions of life science automation focus on:

- Demand-oriented, arbitrarily hierarchical flexible process mappings of interdisciplinary applications in changing automation environments
- Guarantee of a cost-efficient, complete horizontal and vertical operating automation
- Support of the global e-business in compliance with quality standards, legal regulations, as well as data security
- Accomplishment of the automatically generated data flood of high-parallel laboratory processes (HTS, HCS), inclusive support of analysis of experiments and information discovery

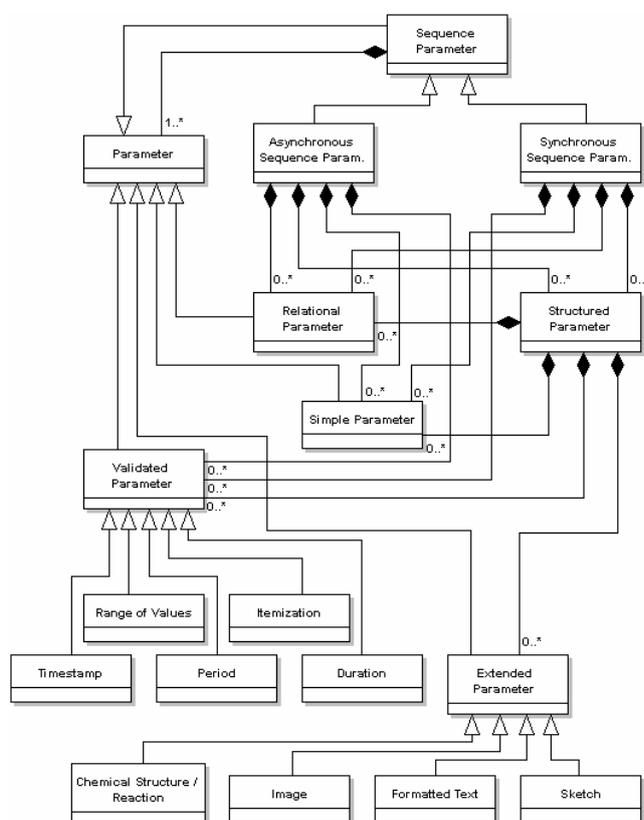
LIMS and electronic laboratory notebooks (ELN) are both used for mapping especially parts of sequential workflows of laboratories. For issues of data consistency, dedicated solutions of LIMS and ELN have to be coupled by using reliable synchronization algorithms. Our approach consequently assumes LIMS and ELN as integrated.

### **Flexible prozess hierarchie mapped different laboratory workflows completely**

The presented LIMS supports the interdisciplinary character of experiments in life sciences by a free mapping of laboratory processes and parameter definitions. This is a demand from a R&D

LIMS particularly in fast changing automation environments. A hierarchical process mapping which flexibly adapts to manual and automated process structures, working tasks and workflow phases based on an open system and parameter concept. Thereby, freely definable parameters with arbitrary data structure secure adjustment flexibility and application independence of the LIMS for flexible laboratory automation.

Important structuring elements are a variable project classification hierarchy in any number of discretionary steps, parallel sequences oriented on laboratory reservoirs, processes, process steps, process parameters, structured process parameters, as well as parameter sequences based on structured process parameters for the coverage of asynchronous and synchronous records. Expectations to ELN can be considered by an extension of the parameter system with multimedia data types and adapted user interfaces sufficiently in the LIMS.



**Figure 1** Parameter model of the presented LIMS

Figure 1 shows the class model of process parameters used in sequential and parallel activity structures to meet the requirements of comprehensive system integration within LIMS. Structured parameters can summarize for example multi-parameter data for individual measuring points (wells on microtiter plate or the like) or process variables and their meta-data (e.g., parameters of image acquisition, surroundings conditions, and process notes). Time series can be formed synchronously or asynchronously from any number of several parameters. Special data types offer particular user interfaces within LIMS: image browser for photos, integrated appropriate editors for the generation of sketches or formatted texts, tools for processing and search for specialized information like chemical structures, etc.

### **Automated communication between LIMS and laboratory instruments**

The secure and efficient control of immense data amounts generated by drug discovery requires a bi-directional automated communication between LIMS and laboratory instruments (e.g., sensors, analytical devices, robot process control systems, cell handling systems). Therefore, a framework concept with loose coupling process mappings based on configurable fast process database coupling and XML communications for horizontal and vertical system integration of

LIMS is observed. The developed LIMS interfacing to laboratory processes is based on a configurable converter for syntax and semantic adaptation of an open process hierarchy for project master data up to structured multimedia-based measurement parameters [4].

## Complete experiment documentation with integrated ELN

The complete documentation of the laboratory workflow combines all process parameter, measured values and results of the laboratory process supplemented to notes, documents, literature, etc. for a complex and consistent documentation, evaluation, as well as decision making. Figure 2 presents several user interfaces and information relations using an example of HTS experiments based on processes of plate positions.

The screenshot displays a complex LIMS interface with several key components:

- dynamic process documentation:** A text area for recording experimental details like 'Ziel', 'Reaktionsbezeichnung', and 'Kommentar'.
- assayed compounds:** A list of compounds with their IDs and names, such as '64\_VV\_Compound'.
- plate layout:** A grid representing a 96-well plate with colored circles indicating the status of each well (e.g., green for 'OK', red for 'Error').
- result parameter:** A list of numerical results for each well, such as 'CV: (13) 4.9664424299123 %'.
- process notes:** A bar chart showing 'Vitalität' (Vitality) for different compound IDs.
- compound library:** A sidebar showing a list of compounds and their chemical structures.

**Figure 2** Complex experiment documentation within LIMS

Attached barcodes allow a complete sample tracking. Compounds from compound libraries, which are also managed in LIMS, can be used in the biological experiment. Therefore, these experiments are related to additional information like chemical attributes and chemical structures. The management of compound libraries is supported by integrating special applications for drawing, showing, and searching of chemical structures.

The integrated ELN functionalities (text and graphic editor, including pen computing) allow the storage of unstructured data at any point within the hierarchical documentation as done in usual paper based laboratory notebooks. The electronic documentation of the development including motivations for methods and parameters is available for other team members and secures long-term reproducibility.

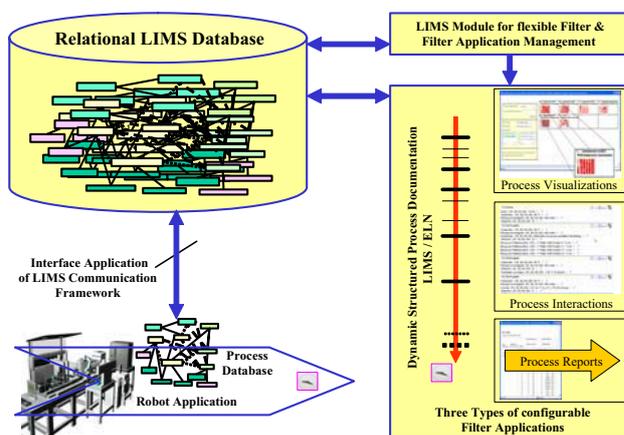
These functionalities are complemented by a flexible data processing option for operative using

of mathematical and logical algorithms between workflow parameters and process variables based on a spreadsheet oriented processor. Free programmable calculation templates link inputs from LIMS or laboratory instruments in terms of a complex instruction of mathematical and logical instructions to outputs which are new data sets within LIMS. This concept creates a self-consistent calculation process which can be automatically executed by a task manager of compatible operating systems on the server side.

### Data reduction by selection

Keeping the overview and the detailed analysis of these immense, miscellaneous data amounts can become the bottleneck. The diversity and complexity of mapped process data can result in losing clarity of process documentation. Therefore, another main focus is the supply of far-reaching possibilities for data selection, data verification, and information visualization, as well

as information supply by using standardized interfaces.



**Figure 3** Architecture of a filter- and filter application management

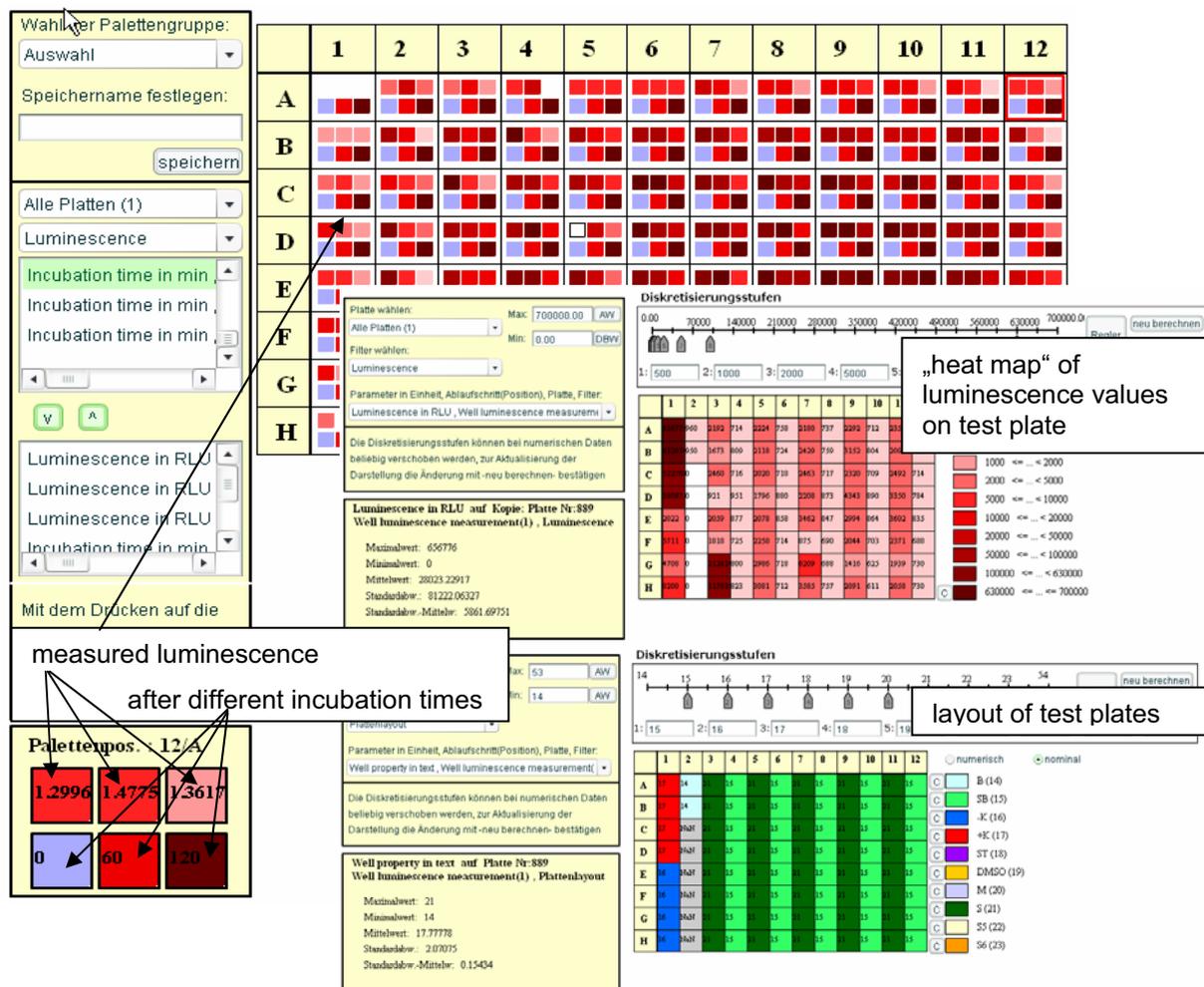
An adaptive filter concept (figure 3) allows a user-defined selection of process parameters in the flexible process documentation and a combination with different applications for visualization and reporting. The applied LIMS generates user interfaces or output routines depended of filter definitions within the process hierarchy. Important configurable types of filter application are embedded process visualizations, interactive user interfaces of process documentation, and reporting tools.

### Visualizations, reports and interfaces

The direct communication with the laboratory devices allows ad hoc generated process visualizations for fast process tracking as well as automated import of primary and calculated data from the devices. Embedded interactive visualizations with abstraction degrees can offer optimal support to extract all relevant information from the primary data for further decisions as soon as possible [5]. Topics of visual analysis method can be the exploration of process data, the comparison of process values, correlation, identification of hits or errors as well as keeping the overview. Manifold visualization techniques for nominal valued (plate layout down right in figure 4) and numerical data are integrated in the concept of user configurable filter application. The quality of experiment results can be conceived fast and supports time near decisions on following experiments (e.g., identification of hits of luminescence in primary screening data for

secondary screening experiments, as shown in the middle of figure 4). Figure 4 presents also the visualization of several process variables (here luminescence and incubation times). One up to six parameters per well of a microtiter plate (MTP) or a virtual MTP can be displayed in this manner for a comparing or uncovering correlations.

Another well-known visualization technique of process data are trend curves for time series. Further methods can expand this concept [6], in particular HCS analyses are assisted with configurable image browser.



**Figure 4** Visualization of experiment results based on microtiter plate format for overview and comparison

Process reports accumulating for post processing examined data records are a further example for system integration. These reports are made available by LIMS via standardized interfaces (XML, data bases, text, or csv file format), including automated communication for external processing to facilitate detailed analysis or presentations.

## Conclusion

Complex and flexible laboratory automation requires adaptive laboratory information management systems and an according LIMS system integration. Therefore, the direct, automated, and bi-directional communication between LIMS and laboratory components is key to secure and

efficient control of generated data.

Fast process tracking embedded in LIMS creates a new integration level for process control systems. Process data filters and modern scientific data visualizations are components in a flexible concept for the integration of diverse applications.

The achieved scalable and flexible process mapping in a web-based LIMS ensures the performance of arbitrarily challenging obligations to documentation and evidence as well as a high system availability in cooperation networks which also includes modern mobile computing.

#### References:

- [1] R.D. McDowall, et al., "Laboratory information management systems - part I. Concepts," J Pharm Biomed Anal, vol. 6, no. 4, 1988, pp. 339-359.
- [2] R.D. McDowall, "Not Another LIMS Project?," LCGC Europe, vol. 9, no. 1, 1996, pp. 21-26.
- [3] R. Jonak, "LIMS - New Trends in Software," G.I.T. Laboratory Journal Europe, vol. 6, no. 4, 2002, pp. 160-161.
- [4] B. Göde, et al., "Adaptive Online-Steuerung von Gerätesystemen - Die direkte Ansteuerung von Geräten aus einem LIMS," LITUS - Laboratory IT User Service, vol. 2, no. 2, 2005, pp. 12-14.
- [5] Schumann, H.; Müller, W., Visualisierung - Grundlagen und allgemeine Methoden. 1. ed.; Springer: Berlin, 2000.
- [6] Brodlie, K. W.; Carpenter, L.; Earnshaw, R. A.; Gallop, J. R.; Hubbold, R. J.; Mumford, A. M.; Osland, C. D.; Quarendon, P., Scientific Visualization: Techniques and Application. 1. ed.; Springer: Berlin, 1992.

#### Authors:

Dr.-Ing. Silke Holzmüller-Laue<sup>1</sup>

PD Dr.-Ing. Bernd Göde<sup>2</sup>

Dipl.-Ing. K. Rimane<sup>2</sup>

Prof. Dr.-Ing. Norbert Stoll<sup>1</sup>

<sup>1</sup>celisca – Center for Life Science Automation,

F.-Barnewitz-Str.8

18119, Rostock, Germany

Fax: +49(0)381 498 7802

Phone Dr. Holzmüller-Laue: +49(0)381 498 7721

E-mail Dr. Holzmüller-Laue: silke.holzmuller-laue@celisca.de

Phone Prof. Stoll: +49(0)381 498 7804

E-mail Prof. Stoll: norbert.stoll@celisca.de

<sup>2</sup>University of Rostock,

R.-Wagner-Str.31

18119, Rostock, Germany

Fax: +49(0)381 498 7702

Phone Dr. Göde: +49(0)381 498 7717

E-mail Dr. Göde: bernd.goede@uni-rostock.de

Phone Rimane: +49(0)381 498 7721

E-mail Rimane: kristina.rimane@uni-rostock.de

N. B. Chang/ A. Gonzalez

## **A Decision Support System for Sensor Deployment in Water Distribution Systems for Improving the Infrastructure Safety**

### **ABSTRACT**

Meeting security objectives for a water utility is a challenging task. The need for thoughtful and comprehensive planning for prevention, mitigation, response and recovery from an event became even clearer after the attacks of 9/11. In response, an Emergency Response Plan (ERP) for water supply systems is now required under the Bioterrorism Act of 2002 in the US. As part of this process, each US utility will incorporate the results of their Vulnerability Assessment into an existing or new ERP. The goal of this paper is to introduce the architecture of an intelligent decision support system in this emerging area for decision support in emergency preparedness and response for water utilities. It presents the integration of knowledge acquisition, ontology modeling, expert system evaluation, and environmental assessment for structuring an environmental decision support system.

### **INTRODUCTION**

There are nearly 60,000 community water supplies in the United States serving over 226 million people [1]. In an uncertain world, the prepared water utilities must provide safe and reliable water supplies in spite of natural and man-made disasters. In general, the framework for water utility protection comes in six parts, from environment to convey, to treatment, to storage, to distribution, and to point of use [2]. Natural disasters of concern may include earthquakes, floods, wind, public health, drought, hurricane, fire, and volcanoes. Human-caused disasters of concern may include electric power failures, communication and information failures, accidents and fire, hazardous material releases and contamination, security threats, equipment failure or line breaks. They are circumscribed by dotted lines alongside the service system to help identify the boundaries of the impacts that could be either conventional or non-conventional.

Meeting the security objectives for water utilities is an obviously challenging task. The need for thoughtful and comprehensive security planning for preventing an event, mitigating the impact of an unavoidable event, responding to an event, and recovering from an event became even clearer after the terrorist attacks of September 11, 2001 [3]. In response, an Emergency Response Plan (ERP) for water supply systems is now required under the Bioterrorism Act of 2002 in the United States. Each U.S. utility will need to incorporate the results of their Vulnerability Assessment into an existing or new ERP. The US Environmental protection Agency identifies three major steps in developing procedures - response, recovery, and remediation, with a list of initial and recovery notification required [4]. It should be consistent with the crisis management structure set out under the National Incident Management System enacted by the US Department of Homeland Security on March 1, 2004, which provides standardized protocols and procedures to coordinate the efforts of water systems and first responders during emergencies. Having a well-thought-out intelligent decision support system (DSS) in place before an extraordinary event strikes allows public officials to quickly respond to and stabilize chaotic situations.

The goal of this paper is to introduce the architecture of an intelligent DSS in this emerging area and delineate how to develop integrated ontological engineering analyses, expert systems' evaluation, and environmental modeling platforms for decision support in emergency preparedness and response for water utilities. It particularly presents the process of knowledge acquisition, ontology modeling, expert system evaluation, and environmental assessment for structuring the environmental decision support system (EDSS). This EDSS will help the users learn from experts with practical experience in developing the sensor network plans for Vulnerability Assessment (VA) and putting them into effect in all aspects. Based on the elements that go into this EDSS, it may help to determine a sound monitoring process to effectively implement ERP in a water utility.

## **METHODOLOGY**

A key concept when determining or evaluating a potential threat to a water utility is the "design-basis threat," which is the credible threat that a utility's security systems are designed to protect against [5]. Identifying the design-basis threat should be based on knowledge of the threat profile in a specific area, as well as past events

from which a good source of information on the region's threat profile. Prior to the 9/11 terrorist attacks, most water utilities would have identified an operational accident, such as chlorine gas leak, or natural disaster, such as storm events resulting in high turbidity source water, as the primary threat to their utility. Nowadays, the possibility of deliberate disruptive terrorist attacks must also be considered. Achieving this goal could call for an advanced planning, design, and operation of sensors and sensor network in response to the acute needs for building an EDSS.

In the past, advanced early warning systems around the world can be found in Ohio River (USA), Paris (France), St. Clair River (Canada), River Trent (United Kingdom), River Dee (United Kingdom), Yodo River (Japan), Rhine River (multiple countries), Edmonton (Canada), etc [6]. There are also numerous traditional exterior and interior detection sensor technologies available for various type of VA in these early warning systems. For example, monitoring stations installed in River Trent, United Kingdom, generate alarms based on water quality, and telemeters this data to treatment facilities for emergency response. In this case, the following are considered to be the most significant pollutant risks, including pesticides, ammonia, dissolved oxygen, chlorate, bromide, nitrate, boron, nickel, antimony, Giardia, Cryptosporidium, total phenols, polycyclic aromatic hydrocarbons, and hydrocarbons. In addition to traditional instruments, emerging monitoring techniques have been developing with unprecedented speed. Most notable in this development is the biosensors for testing the bacteria/viruses/protozoa, given that waterborne pathogens had been a long standing threat to human societies. A critical issue in water utility is how to utilize the knowledge about those newly-developed sensors to aid in the design of emergency preparedness and response systems in the near future, including the early warning systems. This paper aims to present a unique viewpoint from an Artificial intelligence (AI) perspective for promoting risk-based DSS in emergency events.

### Monitoring Technologies

The accidental discharge of contaminants from various industrial sources and natural floods might occur at unpredictable times. In the past, online physical and chemical sensors were installed to detect physical and chemical parameters, such as temperature ammonia, turbidity, color, conductivity, pH, dissolved oxygen, conductivity, oxidation reduction potential (ORP), chloride, hardness, alkalinity, particle numbers, volatile organic compound (VOC) (e.g., such as Methyl, Tertiary

Butyl Ether, MTBE), toxicity (Microtox monitor), total organic carbon (TOC), some inorganic matter (UV Kontron), water quality (e.g., Hydrolab depth sampling), and flavor profile (e.g. methylisoborneol (MIB) and geosmin). Plant operators normally count on color and turbidity measurements for chemical dosing and control. For example, both powdered activated carbon (PAC) and alum are added in response to high levels of color and turbidity. Yet, detection and identification of microorganisms on a real-time basis would significantly further reduce the risk of contaminated water reaching the end-users. In recent years, there are some bio-monitoring stations that had been deployed in river systems, such as the cases of river Rhine, Japan and South Korea, to supplement their extensive system of online gas chromatographs in source water. Several types of fish tests, daphnia tests, mussel tests, bacteria tests, and algae tests were applied in parallel with any online physical and chemical sensors. On a routine basis, analyses are performed for samples taken at different depths in the reservoirs, lakes, and rivers to detect Chlorophyll-a, plankton, zooplankton, etc. in response to public health concerns.

In response to special needs, there are many emerging monitoring techniques that are on the horizon. Some of them have started becoming commercially available while the others are still in an exploratory phase. The new technologies, such as an electronic nose using acoustic wave sensor, quartz micro balance (QMB), metal oxide (MOX), and polymer composite sensor (PCS), are emerging. The last, has been used in California to detect algae blooms and odor problems. In addition, methods for real-time measurement of trihalomethanes (THMs) and haloacetic acids (HAAs) in water distribution networks are highly desirable because of the promulgation of Maximum Contaminant Levels (MCLs). Some emerging methods are becoming gradually available [7]. Yet, there is a growing concern of bioterrorism in water supply, such as the introduction of *Cryptosporidium* [8]. Some rapid bacteriological methods, such as DNA-gene chip technology, flow cytometry, immunoassays, polymerase chain reaction (PCR), electrochemiluminescence, ribotyping, bioluminescence, and immunomagnetic separation (IMS) have been devised to meet the real-time needs for online measurement [6]. The Oak Ridge National Lab has devised a new tool, AquaSentinel, to detect a broad spectrum of poisons. AquaSentinel will be commercialized soon [9]. These new and emerging monitoring techniques may aid in vulnerability assessment by providing an additional dimension of crucial information for decision making.

### Environmental Decision Support System

The proposed rule-based expert system for the customized sensor network in a GIS platform will be derived from extensive experience in modeling water transmission and distribution networks. The primary concern is the disruption of supply to a large number of customers in rural communities with all the consequences for disrupting fire fighting capabilities and public health risks. The optimal locations for sensor deployment may then be drawn from multi-dimensional simulation outputs with respect to associated statistical indices. Contaminated water can have potentially catastrophic impacts on the health of end users, such as happened in Milwaukee, Wis., in April 1993 where a cryptosporidiosis outbreak infected more than 400,000 water consumers. In Gideon, Mo., in Nov., 1993, a Salmonella outbreak resulted in the death of seven individuals [12]. Integrating modeling and mapping systems to realize the contaminant kinetics and modeling strategies becomes the recent focus, promoting system reliability and the ability to present the risks to customers in a more amiable way [10, 11]. Specific uses generally require particular model synthesis and integration. For example, the assessment of THMs and HAAs in water distribution networks must rely on the combination of water quality models and hydraulics models (e.g., such as Hardy-Cross method). Vulnerability assessment can be further made possible by combining sensors with models to determine what would be the optimal locations in the water supply systems to deploy and which type of sensor is required for corresponding location. Models required in this area include hydraulic models, steady state water quality models, dynamic water quality models, optimization models, and display/visualization models. The models encompass a wide range of methods that can be used to predict the water quality in source water, treatment trains, and water distribution networks under a set of conditions. Ontology modeling can be integrated with environmental models for decision making. Some ontologies issues have been investigated for groundwater contamination and remediation in which dynamic semantics and task-oriented knowledge are often represented by rules or problem-solving methods [13, 14, 15]. The appropriate use of sensors/sensor networks in an EDSS would be able to combine all early warning information from source water monitoring systems to process monitoring of water treatment plants, and to the monitoring of water quality in distribution systems. On a design-basis, striving to link all critical components together, the advances of ontological engineering may further promote the functionality of the EDSS by providing a friendly

design tool for the collection and selection of sensors and predictive models. Seamless integration of the EDSS into current SCADA systems can be anticipated by appropriate applications of water quality and demand forecasting, process modeling and assessment, inferential sensing, and process control [16].

## CONCLUSION

Current regulatory pressures are driving the drinking water industry to more sophisticated levels due to the recent homeland security and public health concerns. An EDSS integrating AI and traditional DSS techniques can be developed for most drinking water treatment process, covering from source water, transmission system, treatment facilities, finished water storage, distribution system, and supporting infrastructure. It is also essential that the importance of online analyzers in the advanced real-time monitoring system be reflected in the development of an EDSS as well as the implementation of effective analyzer quality assurance and error detection programs.

## References:

- [1] U.S. EPA (1999). 25 year of the Safe Drinking Water Act: History and Trends, EPA 816-R-99-007, Office of Water, Dec. 3, 1999.
- [2] Grigg, N. S. (2002). Surviving Disasters in Water Utilities: Learning from Experience. AWWA Report, ISBN 1-58321-204-3.
- [3] Hogan, A. Jr. and DeBoer, J. (2005). Security Planning in an Unstable World: A Public Official Guide. AWWA Report, ISBN 1-58321-252-3.
- [4] U.S. EPA (2002). The Guidance for Water Utility Response, Recovery & Remediation Actions for Man-Made and/or Technological Emergencies, Office of Water (4610M) EPA 810-R-02-001, April 2002.
- [5] American Water Works Association (AWWA) (2002). Water System Security: A field Guide. ISBN 1-58321-193-4.
- [6] Grayman, W. M., Deininger, R. A., Males, R. M. (2001). Design of Early Warning and Predictive Source-Water Monitoring Systems. AWWA Report, ISBN 1-53821-172-1.
- [7] Emmert, G. L., Cao, G., Geme, G., Joshi, N., and Rahman, M. (2003). Methods for Real-time Measurement of THMs and HAAs in Distribution Systems. AWWA and EPA Report, ISBN 1 84339 8982.
- [8] Foran, J. A. and Brosnan, T. M. (2000). "Early warning systems for hazardous biological agents in potable water." Environmental Health Perspectives, 108(10), pp. 993-996.
- [9] Krause, C. (2005). "Bioterrorism Early Warning System-Algae?" Journal of Environmental Health, p43.
- [10] Mays, L. W. (editor) (1989). Reliability Analysis of Water Distribution Systems. ASCE, New York, ISBN 0-87262-712-8.
- [11] Mays, L. W. (editor) (2000). Water Distribution Systems Handbook. McGraw-Hill, New York, ISBN 0-07-134213-3.
- [12] Clark, R. M. and Grayman, W. M. (1998). Modeling Water Quality in Drinking Water Distributing Systems. American Water Works Association, Denver, CO, ISBN0-89867-972-9.
- [13] Chan, C.W., Peng, Y., and Chen, L. L. (2002). "Knowledge acquisition and ontology modeling for construction of a control and monitoring expert system." International Journal of Systems Science, 33(6), 485-503.
- [14] Antoniou, G. and van Harmelen, F. (2004). A Semantic Web Primer, Cambridge, MIT Press.
- [15] Omelayenko, B., Crubezy, M., Fensel, D., Benjamins, R., Wielinga, B., Motta, E., Musen, M., and Ding, Y. (2005). UPML: The language and tool support for making the semantic web alive, in Fensel et al., (eds.), MIT Press, 140-170.
- [16] American Water Works Association (AWWA) (2002). Online Monitoring for Drinking Water Utilities. Edited by Erika Hargesheimer, Osvaldo Conio, and Jarka Popovicova, ISBN 1-58321-183-7.

## Authors:

Prof. Ni-Bin Chang and Prof. Avelino J. Gonzalez  
University of Central Florida  
Orlando, FL, 32816 USA  
Phone: (407) 823-1375 and (407) 823-5027  
E-mail: [nchang@mail.ucf.edu](mailto:nchang@mail.ucf.edu) and [gonzalez@ucf.edu](mailto:gonzalez@ucf.edu)

P. Hamolka / I. Vrublevsky / V. Parkoun / V. Sokol

## New Film Temperature And Moisture Microsensors For Environmental Control Systems

### Environmental Systems: Management and Optimisation

Progress in electronic technologies and such advantages as small dimensions and weight, high thermosensitivity have resulted in rapid development of a new type of sensors - film temperature and moisture sensors. Such types of microsensors can be easily embedded in diverse computerized control systems, used, for example, for management of environmental of various agriculture storage installations, which generally must function with strictly predefined temperature and humidity ranges.

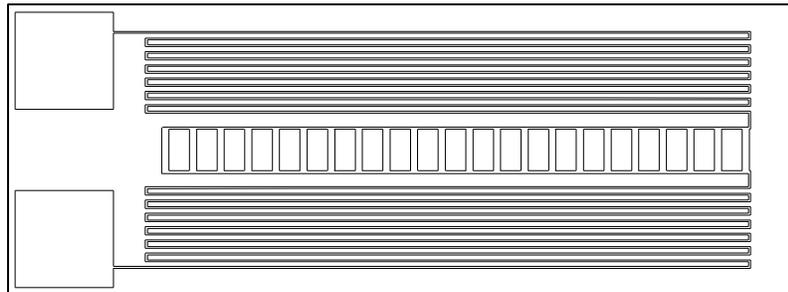
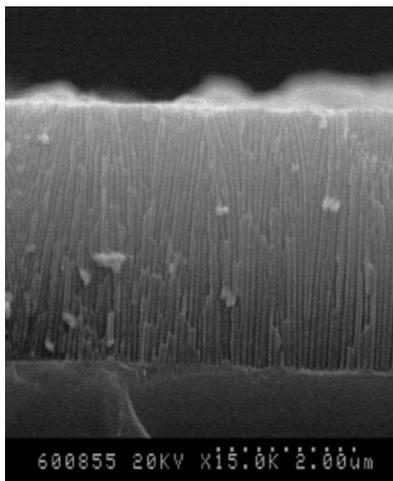
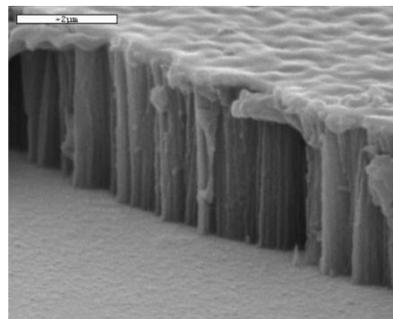


Fig. 1. Topology of copper film sensor structure

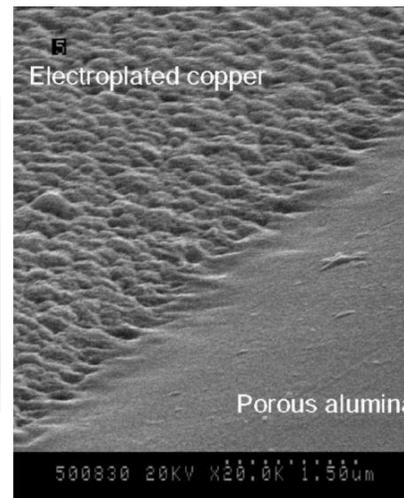
Our sensors represent microelectronics devices with area of 2x5 mm with two leads each (Fig. 1). They are developed on the basis of the hybrid nanoporous alumina technology (Fig. 2).



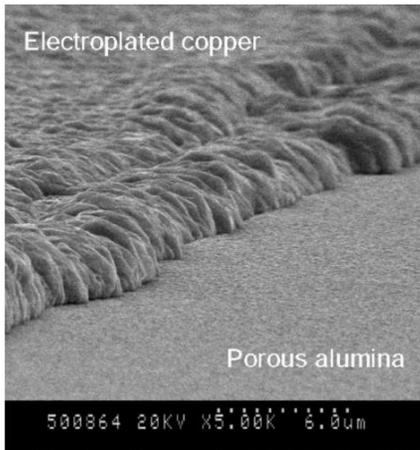
a



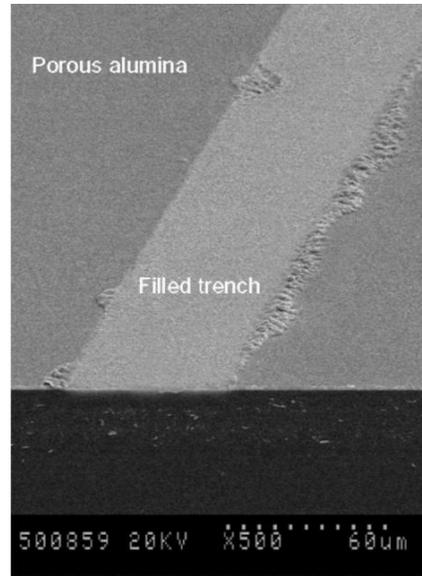
b



c



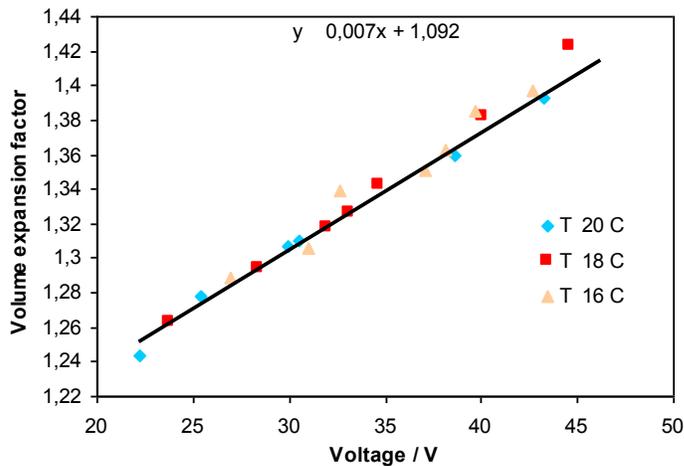
**d**



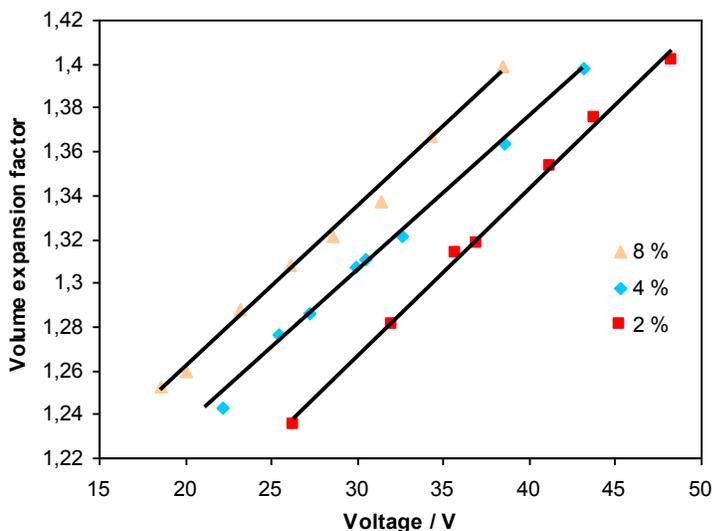
**e**

Fig 2. Hybrid nanoporous alumina technology: **a** - SEM image of nanoporous alumina; **b** - SEM image of electrolessly deposited Cu layer in the bottom of the trench; **c** - SEM image of planar ECP copper – nanoporous alumina interface; **d** - SEM image of ECP copper - nanoporous alumina interface; **e** - SEM image of Cu-filled trench (sensor structure).

Temperature sensors have electric resistance of 50 Ohm and protecting dielectric based on nanoporous alumina made by electrochemical anodizing in water solution of oxalic acid (Figs. 3 and 4).



**(a)**



(b)

Fig. 3. Volume expansion factor ( $k$ ) vs anodizing voltage in the steady-state growth region of the porous alumina film as a function of temperature (a) and electrolyte concentration (b): (a) 4 % solution of oxalic acid at the temperature of 16, 18 and 20 °C; and (b) 2, 4 and 8 % solution of oxalic acid at 20 °C.

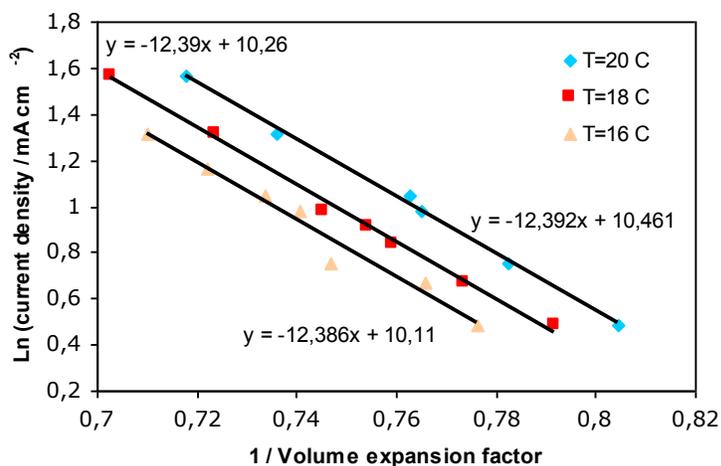


Fig. 4. Dependence of anodic current density logarithm on the inverse volume expansion factor of the porous alumina film as a function of temperature. 4 % solution of oxalic acid at the temperature of 16, 18 and 20 °C.

Their electric characteristics were studied in the temperature range of -60 to 160°C. The principle of operation of developed moisture sensors lies in periodical monitoring of capacitance of thin (2,5 micrometers) film of anodized alumina. At the end of each measurement cycle that porous film can be dried by means of embedded heater layer. Thus the critical low lag effect is achieved.

Main distinguishing feature of our devices in comparison with similar film sensors is the usage of nanoporous alumina as a material for protecting and insulating coating. The main problem at manufacture of such film temperature sensors is a mutual co-ordination between elastic properties of thermosensitive material and protecting insulating coating. An important feature of nanoporous alumina is a low elastic modulus of about 90-140 GPa comparing with the value of 340-380 GPa for non-porous alumina [1,2]. This feature allows to reduce mechanical stresses arising in film structures, for example at cycle heating and cooling. Thus, the long-term stability of electric characteristics of such film sensors improves. Another important technological feature – usage of electrochemical deposited copper as a thermosensitive material in temperature sensors and as a material of interconnections in moisture sensors [3]. Electrochemical deposition of Cu provided high purity of films and thus - high reproducibility of measurements.

**References:**

- [1] Sokol, I. Vrublevsky, V. Parkoun, K. Moskvichev, Anal. Bio. Chem. 375 (2003) 968.
- [2] Z. Xia, L. Riestler, B.W. Shanon, W.A. Curtin, J. Liang, A. Yin, J.M. Xu, Rev. Adv. Mater. Sci. 6 (2004) 131.
- [3] M. Traving, I. Zienert, E. Zschesh, G. Schindler, W. Steinhögl, M. Engelhardt, Appl. Surf. Sci. 252 (2005) 11.

**Authors:**

Professor Dr. Vitaly Sokol,  
Dr. Igor Vrublevsky  
Vladimir Parkoun  
Post graduated student Piotr Hamolka

Hybrid Technology Lab, Department of Nano and Microelectronics, Belarusian State University of Informatics and Radioelectronics, 6 Brovka str  
Minsk, 220013, Belarus  
Phone: +375 17 2938024  
Fax: +375 17 2313908  
E-mail: zapad3@tut.by

N. Buyankhishing, K. Masumoto, M. Aley

## **Parameter estimation of an unconfined aquifer of the Tuul River basin Mongolia**

### **INTRODUCTION**

Population growth in the Ulaanbaatar area, Mongolia has been rapid, and demand for potable water will continue to rise. The water requirements of Ulaanbaatar city are generally met from groundwater originating from the Tuul River. Groundwater use in the area has consistently increased since 1960. 274.56 thousand m<sup>3</sup>/day groundwater supplies in the area derived from alluvial deposit. Totally, there are 159 production wells for the water supply in Ulaanbaatar city. Averagely about 110-130 production wells work and pump 150 000 m<sup>3</sup>/day water [1]. Central extraction system has been working since 60's, groundwater level (22-25m) in the area is declined and created depression cone that indicates groundwater shortage. The reliability of predictions using groundwater models depends on parameter estimation.

### **Study Area**

The study area is situated in the Ulaanbaatar city area (Fig 1). The Ulaanbaatar area is located in the Tuul River basin, which is surrounded by mountains. The mountains in the south have general altitude 2000 m above sea level, and the highest peak is 2256.3 m a.s.l. The north side of these mountains has sharp edges, which range in elevation from 600 to 900 m above sea level, and forms several parallel valleys. The Tuul River originates in the Hentii Mountains and flows generally from northeast to southwest in a variably meandering channel. The total length of the Tuul River is 819 km, and the total catchment area is 50.4 km<sup>2</sup>. The catchment area near Ulaanbaatar is about 6300 km<sup>2</sup>. The width of the river in the city area is 45-50 m, but during dry season falls to 5-18 m. River depth during droughts is about 0.9-1.2 m, average velocity is 0.31-2.24 m/sec and maximum velocity reaches 4 m/sec. The river is recharged mainly by precipitation (JICA, 1995). Daily mean discharge data has collected at the Ulaanbaatar station from 1946 to 1991, with values ranging from 0 m<sup>3</sup>/sec (winter time) to 627 m<sup>3</sup>/sec (during flooding). The river at Ulaanbaatar is completely frozen from December to February. Flow gradually begins in March and discharge gradually increases to peak in the rainy seasons from July to August. River water levels during flooding are typically 1.8-2.0 m higher than in drought periods. The study area is characterized by a semi-arid climate, with a hot, dry summer and cold winter. Average monthly temperature in the river basin varies from a minimum of – 2.3°C in January to maximum of 16.7°C in August, as recorded at Ulaanbaatar station. Annual precipitation in the area varies from 242.7 mm to 396.7 mm, depending on the altitude of the station sites. Nearly 80 percent of annual precipitation falls between October and March.

## Hydrogeological Setting

Geologically, geology around the study area consists mainly of Carboniferous sediments, which are intruded by Jurassic to Triassic granitoids rocks and locally covered by Cretaceous sediments

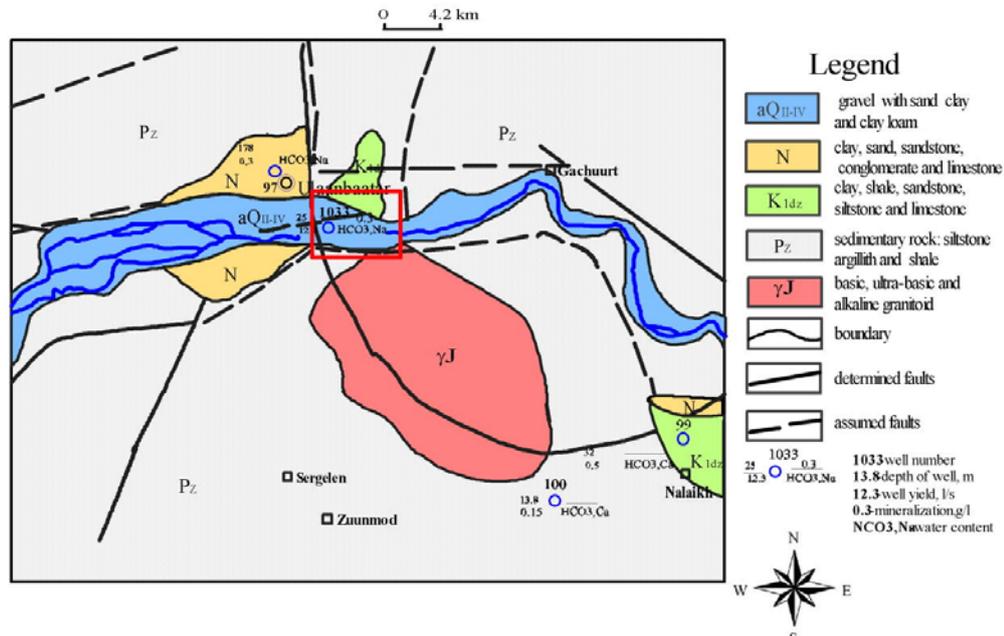


Figure 1. Study area

are contained in the Paleozoic to Quaternary formations. The geologic units are further classified into hydrogeologic units depending on hydraulic and storage properties of the rocks. Broad alluvial basin has been filled by more than 5-65m of gravel with sand and clay and it is commonly irregular in thickness and composition. Average thickness is 30m. The alluvial deposits are divided into two layers: upper (Later Quaternary) and lower (Middle to Later Quaternary). Upper coarse-grained deposit makes up the most permeable zones of the unconfined aquifer system. Hydraulic property data for the alluvium-unconfined aquifer was derived from aquifer pumping tests. Hydraulic conductivities of coarse-grained alluvium in the upper layer range from 44 to 55 m/day; however, values of 100 m/day are more typical. Specific yield ranges 0.13 to 0.2. Lower finer-grained alluvium, it is typically a fine sand with silt and clay and has hydraulic conductivity values of about 5 to 35 m/day. Depth to groundwater from monitoring wells varies in the range 0-11m below ground level. The water table is 0- m below land surface in the basin.

## Boundary condition

The bottom of the unconfined aquifer system is defined, as the top of Tertiary clay is a low-permeability unit that does not store or transmit significant quantities of groundwater. This is considered as no-flow boundary condition in the numerical model. The northern and southern boundaries of the alluvium are bounded by Carboniferous sedimentary and Jurassic granite. These boundaries are considered as no-flow boundaries. The western and eastern parts constant flux boundary condition is chosen because groundwater flows from west to east. Source of data accumulated from previous researchers.

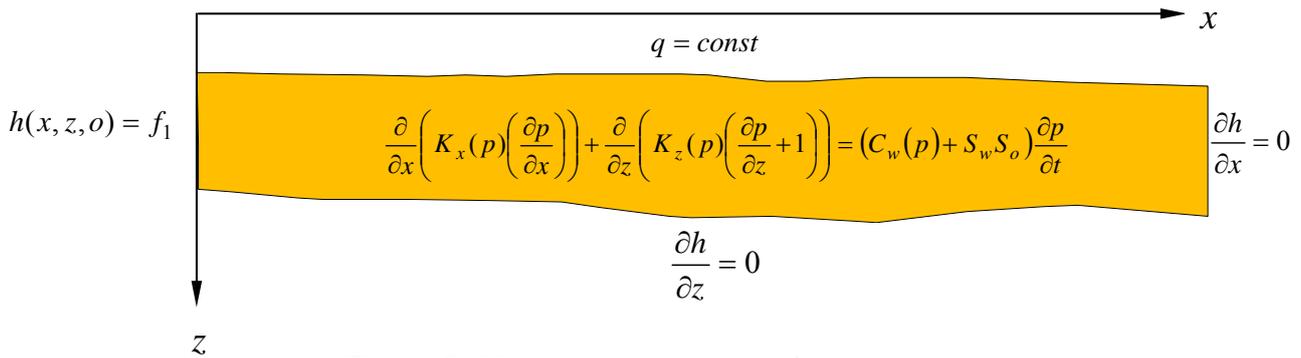


Figure 2. Mathematical model of the area

### Grid design

All the grids have uniform lateral dimensions of 50 m by 50 m. The model grid covering km<sup>2</sup> of the study area was discretised into 238 cells with 37 rows and 7 columns and vertically by two layers (Fig 2). The cell is irregularly spaced.

### Input Parameters

The lower layer is defined by material of relatively low- permeability between upper layer and impermeable layer. Upper zone of lower layer consists of fine gravel and finer sand that is yellowish-gray in color, fine grained, and with clay. Thickness of this zone is 1-4m. Below this zone is sand with fine gravel and clay. Thickness of this zone is 1-4m. Third zone is same as first zone of the lower layer. Thickness ranges 2-5m.

### Calibration

The conceptual model is converted to a numerical representation of the system for simulation using the groundwater flow code '3-D unsaturated flow' (Masumoto K, 1997). Calibrations are performed by systematically varying parameters within reasonable ranges and investigating the effects of these changes on the results. Errors between measured and simulated heads are minimized during the calibration.

Results from the steady-state model can easily be used as initial conditions in the transient model. For our case we used transient simulation model. First need to estimate steady-state time duration. We developed the steady-state model for aquifer conditions of November 1976. The head at that time was chosen because the water levels were near equilibrium.

Thus, the steady-state model was used to investigate initial conditions (steady-state time), hydraulic conductivity and recharge rate.

The inverse problem must be solved to find appropriate model structure and model parameters, and then solve the forward problem to obtain required prediction results. The inverse problem or calibration involves adjustment of the model structure and model parameters of a simulation model simultaneously or sequentially so as to take the input-output relation of the model fit any observed excitation-response relation of the real system.

The calibration targets of the study were:

- To minimize errors between measured and simulated water levels in

November.

- To find head distribution between observation points.
- To estimate initial conditions.
- To estimate recharge rate and hydraulic conductivities.
- In trial-and-error calibration, parameter values are initially assigned to each grid.
- During calibration, parameter values are adjusted in sequential model runs to match simulated heads to the calibration targets

Firstly we developed an initial condition (steady-state time) appropriate for November 1976. To develop this initial condition, we run the transient simulation model using the recharge rate. The first guess of the recharge rate was zero, derived from meteorological observation data. The study area belongs to a semi-arid region, in which evaporation largely exceeds precipitation. Hydraulic conductivity was assumed as 0.01cm/sec.

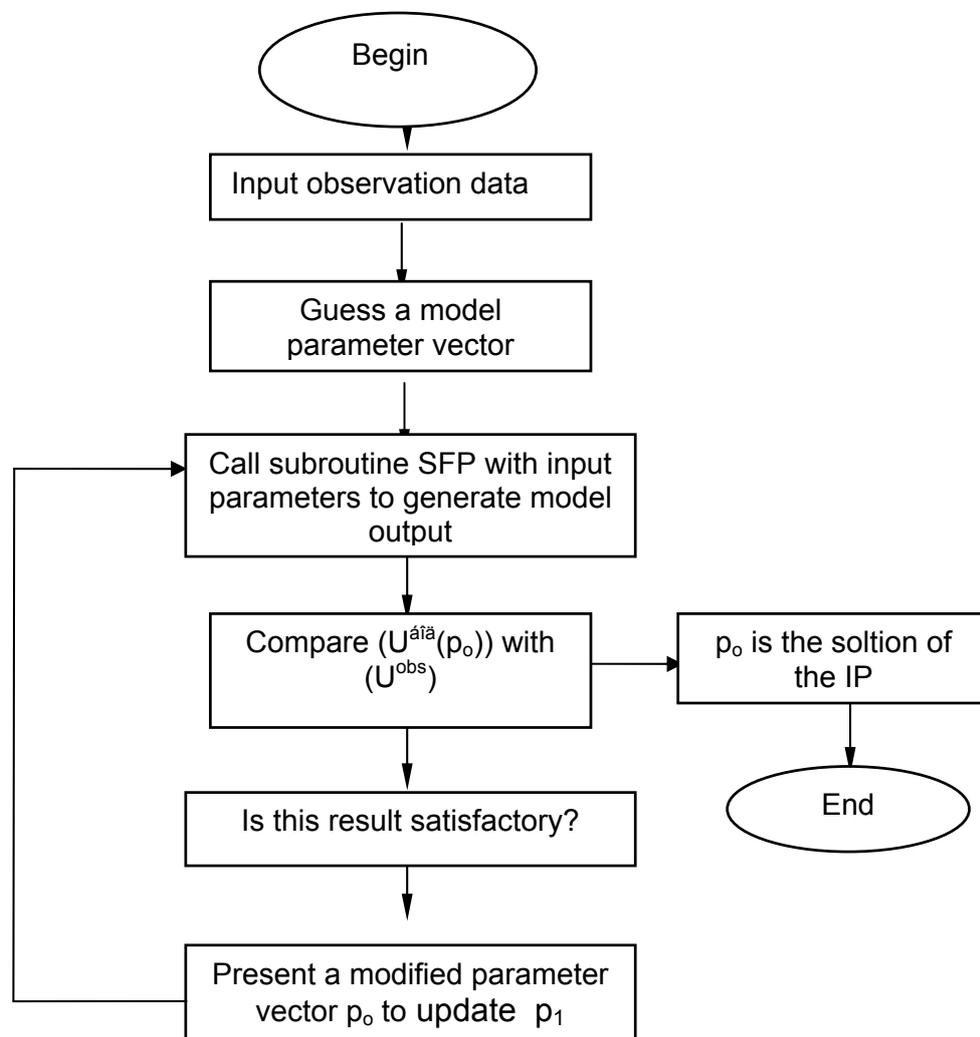


Figure 3. Flow chart of the trial and error method

State variables of a groundwater system can be easily measured. This measurement, such as water level, is usually obtained from observation wells. Water levels are used for the study, but the accuracy of the result depends on the accuracy of the

information and the magnitude and distribution of the aquifer permeability. About 50 simulations were made in order to obtain a set of recharge rate and hydraulic conductivities. Recharge is not only highly dependent on climate, but also on surface and subsurface conditions. We estimated 25 hydraulic conductivities.

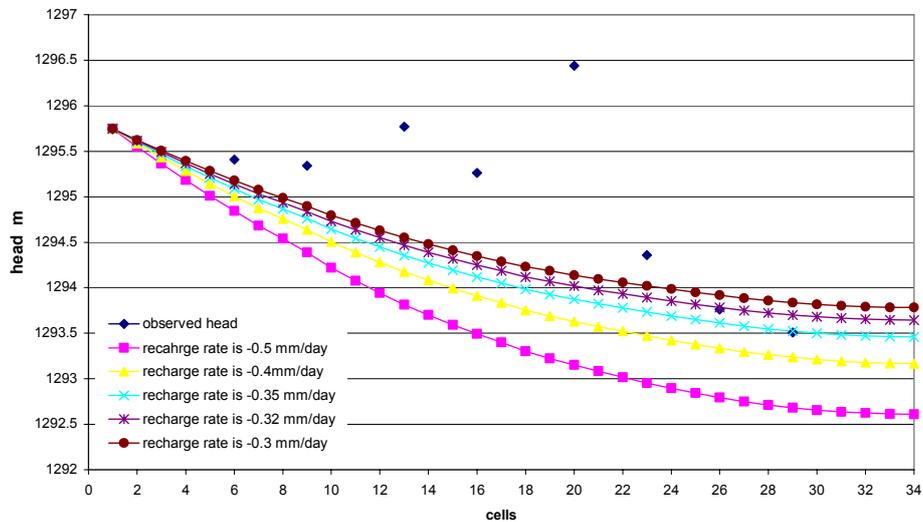


Figure 4. Estimation of recharge rate

Heads after 6 months was chosen as an initial condition because that stable condition continues until April. This time period belongs to winter season.

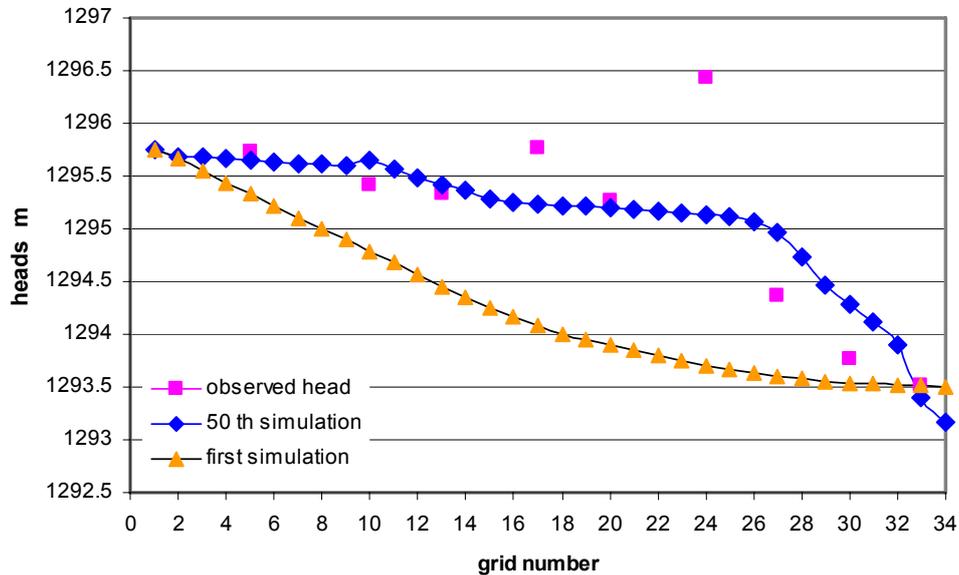


Figure 5. Estimation of hydraulic conductivity

Then we got calibrated recharge rate. Result was compared with meteorological data. Calibrated recharge rate was  $-2.9$  mm/day. We also tested heads after 3 months, and recharge difference was 0.3.

After the 50th simulation, the simulated heads were close to the observed ones. The calibrated hydraulic conductivity is show in Fig 5. About 50 simulations were made in order to obtain a set of recharge rate and

hydraulic conductivities. Each simulation, simulated heads were compared with observed heads were adjusted to minimize head difference. The final set of model hydraulic conductivity provided “best fit” between observed and simulated heads. Calibrated heads almost similar to the observed head.

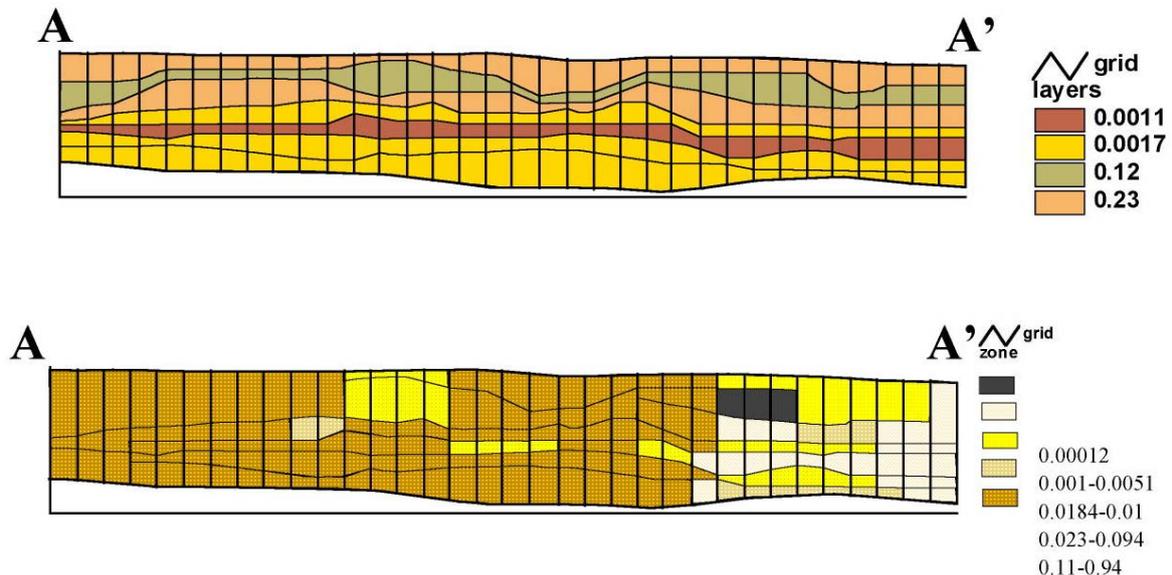


Figure 6. Comparison of initial and calibrated hydraulic conductivity

For this study we used water-level information, but the accuracy of the result depends on the accuracy of the information and the magnitude and distribution of the aquifer permeability.

Recharge is not only highly dependent on climate, but also on surface and sub-surface conditions.

**References:**

[1] Jadambaa.N. 2004. Problems recovery and protect of Tuul River water. p.75

[2] PNIIS (USSR), 1983. Technical study and hydrogeological investigation on new water resoures for Ulaanbaatar city of Mongolia. Moscow.(in Russian)

[3] Japan International Cooperation Agency (JICA), 1995. The study on water supply system in Ulaanbaatar and its surroundings. Final report. Ulaanbaatar. (in English)

**Authors:**

N.Buyankhishig

M.Masumoto

M.Aley

Company, street, P.O.B. Mongolian University of Science and Technology, O.O.Box-520/46

Zip code, city

Phone: 976-11315650

Fax: 976-11-312291

E-mail:bbn@yahoo.com

M. Jacobi / D. Karimanzira

## **Demand Forecasting of Water Usage based on Kalman Filtering**

### **ABSTRACT**

In this work a method of demand forecasting of water usage which is based on Kalman filtering is presented. The Kalman filter is a method that provides an efficient computational solution in least square sense. It has thus received much attention for estimation purposes. To demonstrate the efficiency of the proposed technique, it is used to predict the water usage demand of Beijing. The data employed in the forecasting process are the previously forecasted demand and several exogenous influences. This method employs a simple mathematical process with less computational consumption.

### **INTRODUCTION**

Signal modeling is an attractive topic and plays an important role in many fields such as analyzing stock market, forecasting water level in dams, forecasting car sales volume, weather forecasting or even in water usage demand. In many applications, it is preferable to select the model parameters so as to minimize a sum of squared error criterion. This technique often leads to satisfactory results and has a closed form solution to the interested problem. Therefore, least squared error (LSE) modeling has been widely used in various of data analysis problems. Kalman filtering is one of well-known techniques, which is based on least squared error minimization. Basically, Kalman filtering behaves as a tool to analyze and solve the problem of estimation that has either one or more unknown variables at a time. It can be applied for filtering problem, smoothing and also for prediction [3]. Therefore, it is widely applied in many fields [5,6,11,14]. In [8], forecasting of structural time series models and the Kalman filter were described. It is thus initiating the idea to apply the concept of the Kalman filtering for forecasting in water usage demand in this work.

The process of forecasting using this method is much simpler than those approaches [1,4]. In addition, the computational consumption of this method is quite less. The

concept of the proposed technique will be given in section of “basic principles”. Later, the forecasting results of water demand using the proposed approach compared with linear regression model will be given in section “simulation results”. Finally, some conclusions will be drawn.

## BASIC PRINCIPLES

### Multiple linear regression

In this work, a multiple linear regression model is selected to be a method of statistical analysis. Let  $W_k$  be a predicted water usage demand of  $k^{\text{th}}$  year in which a multiple linear regression model can be expressed as

$$W_k = \alpha + \beta H_{k-1} + \gamma W_{k-1} + \phi E_{k-1} + \phi P_{k-1} \quad (1)$$

where  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\phi$  and  $\phi$  are modeling parameters,  $W_k$  is the actual water demand,  $H_{k-1}$  is the previous number of households in the  $k^{\text{th}}$  year,  $W_{k-1}$  is the previous water demand,  $E_{k-1}$  is the previous number of employment and  $P_{k-1}$  is the previous water price. Based on the concept of signal estimation, the modeling parameters are selected, so that least squared error is minimized (equation 2). Due to the fact that there could be some effects that are correlated to some explanatory variables the ordinary least mean squares method (OLS) is biased and inconsistent. Therefore, in this paper, the weighted least mean squares method is used, so that points with a greater weight contribute more to the fit.

$$\min \sum_{i=1}^n w_i (\hat{Y}_i - Y_i)^2 \quad (2)$$

In matrix notation, the weighted least squares estimator of  $\hat{\theta}$  (vector of all parameters) is

$$\hat{\theta} = (X^T W X)^{-1} X^T W Y \quad (3)$$

$w_i$  is given as the inverse of the variance, giving points with a lower variance a greater statistical weight:

$$w_i = \frac{1}{\sigma_i^2} \quad (4)$$

and

$$W = \text{diag} \{w_1, \dots, w_n\} \quad (5)$$

The parameters in (1) were estimated for the regression model based on the data

from 1996 to 2002 found at [2,10]. Their standard deviations were also calculated (table 1).

| Parameter     | $\alpha$ | $\beta$   | $\gamma$  | $\varphi$ | $\phi$ |
|---------------|----------|-----------|-----------|-----------|--------|
| Best Estimate | 1.2799   | -8.95E-04 | -6.874E-4 | -0.2142   | 0.0038 |
| $\sigma$      | 0.7511   | 0.0019    | 9.71E-04  | 0.486     | 0.0016 |

**Table 1: Statistical measures for the estimate parameters**

### Proposed Kalman filtering technique

In 1960, Kalman published his well-known paper [9] describing a recursive solution to the discrete-data linear filtering problem. Since then, due in large part to advances in digital computing; the Kalman filter has been the subject of extensive research and application, particularly in the area of autonomous navigation. A good overview on an introduction to the Kalman filter can be found [12]. Here the Kalman filtering concept will be briefly given as follows. In the Kalman filtering concept, the general problem of estimating the state  $x \in \mathfrak{R}^n$  of a discrete-time controlled process is described by the following equation

$$x_k = A x_{k-1} + B u_k + w_{k-1} \quad (6)$$

where  $A$  is transition matrix of prediction system,  $B$  is matrix of optional control input,  $u_k$  and  $w_k$  represent the state input and process noise, respectively.

In addition, a measurement  $z \in \mathfrak{R}^m$  of the state is

$$z_k = H x_k + v_k \quad (7)$$

where  $H$  is transition matrix of measurement system and  $v_k$  is measurement noise.

By employing a form of feedback control, the Kalman filter estimates the process state at some time and then obtains feedback in the form of measurements. With this in mind, the Kalman filtering equations are classified into to groups

- time update (prediction)
- measurement update (correction)

For describing the time series a simple AR(1) model is used. The general description for the AR model is shown in equation (13). This equation will be adapted to the AR state space model (equations (6), (7)) [7,13]. The parameters  $a_i$  excluding  $a_1$  are set to zero.

$$y_k = a_1 y_{k-1} + a_2 y_{k-2} + \dots + a_n y_{k-n} \quad (13)$$

The described method was implemented by using Matlab.

Additionally, the Kalman filtering technique is applied with regression method mentioned in previous subsection. The regression method is used for estimation of the  $a_1$  of the AR(1) model. It is also used for parameter estimation of the ARX model (AR model with external influences) described in the previous subsection (equation 1). The general ARX model description is shown in equation 16.

$$y_k = a_1 y_{k-1} + a_2 y_{k-2} + \dots + a_n y_{k-n} + b_{1,1} u_{1,k} + \dots + b_{1,m} u_{1,k-m-1} + b_{2,1} u_{2,k} + \dots + b_{2,p} u_{1,k-p-1} + \dots \quad (16)$$

Basically, by using ARX model, the number of households, employment and water price in each year is required in the prediction process. In summary, there are four approaches to be compared in this paper:

- Regression based ARX model (model 1),
- Kalman filtering technique using AR(1) model with  $a_1 = 1$  (model 2),
- Kalman filtering technique using AR(1) model with previous identified  $a_1$  parameter by regression (model 3) and
- Regression based ARX model with additional use of the Kalman filter (model 4).

The four concepts are implemented in Matlab, which the results are illustrated in the following section.

## SELECTION OF MEASUREMENT NOISES

Long term measurement noise strength can be approximated by looking at a short time period of the time series curve. If we assume that we would approximate it by a dynamic linear model, we can approximate the standard derivation of the model measurements noise by looking at the strength of residuals. The selected variance of noise was  $\sigma_x^2 = 0.001$ , which quite well fits to the observed residuals.

The choice of measurement noises in long term models can be done, for example, by visual inspection, because the exact choice of noise strength is not crucial. The process noise strength is selected based on cross validation, which implicitly corrects also the choice of measurement noise strength. By visual inspection a suitable measurement noise for the AR-estimation model was  $\sigma_{ar}^2 = 0,1$ . Because we are interested in the missing parts of data in prediction with Kalman filter based model. The best way to do this is to follow the measurements exactly whenever there are measurements and use AR models for prediction only when there are no measurements.

This happens when the measurement noise level is set to as low as possible and the process noise is set to a moderate value. Our choice for the noise level in the Kalman filter model was  $\sigma_p^2 = 0.01$ .

## SIMULATION RESULTS

Raw data (water demand, population, gross domestic product) from 1996 to 2002 are employed in the forecasting process [2,10]. The data from 1997 to 2003 were employed to examine the prediction efficiency of the three techniques. The results are shown in figure 1. It can be seen that the methods using the ARX model outperforms the other two methods and the Kalman filter provides also advantages to the simple regression based ARX model.

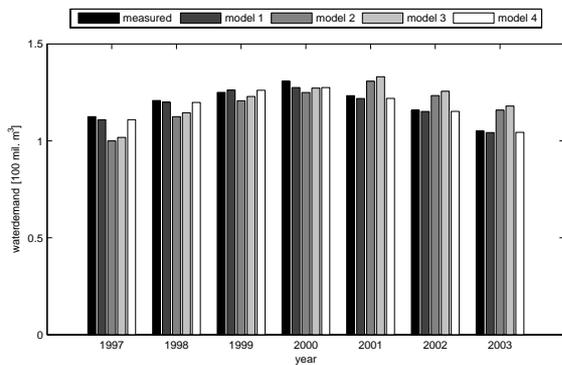


Figure 1: Water demand forecasting.

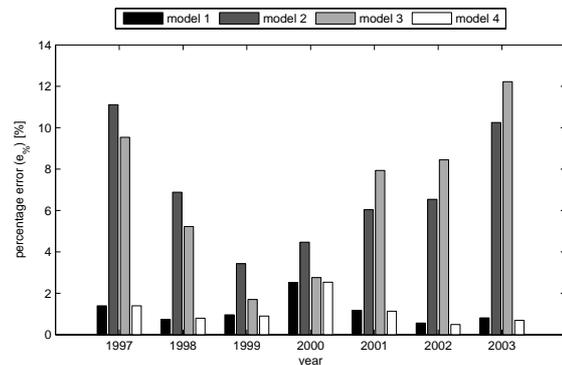


Figure 2: Percentage forecasting errors

The comparison of the forecasting error among these are shown in Figure 7 and also in Table 7, where the amount of error is calculated by

$$e_{\%} = \left( \frac{|D_k - \hat{D}_k|}{D_k} \right) \cdot 100\% \quad (10)$$

where  $e_{\%}$  is the percentage error,  $D_k$  the actual water demand and  $\hat{D}_k$  the forecasted water demand.

| Year    | 1997  | 1998 | 1999 | 2000 | 2001 | 2002 | 2003  |
|---------|-------|------|------|------|------|------|-------|
| Model 1 | 1.39  | 0.74 | 0.96 | 2.52 | 1.17 | 0.56 | 0.81  |
| Model 2 | 11.11 | 6.88 | 3.43 | 4.47 | 6.04 | 6.54 | 10.25 |
| Model 3 | 9.54  | 5.23 | 1.69 | 2.76 | 7.93 | 8.45 | 12.23 |
| Model 4 | 1.39  | 0.79 | 0.89 | 2.53 | 1.13 | 0.49 | 0.69  |

Table 2: Table of the forecasting errors of the different methods

## CONCLUSIONS

This paper introduced a technique based on Kalman filtering, for forecasting water usage demand.

Based on the raw data from year 1997 to 2003, the forecasting results of using regression technique and the proposed approach, also different time series models (AR(1) and ARX) are compared. With the obtained results, it is found that the proposed method based on the AR model with external influences is the best method. Also the use of the Kalman filter with the ARX model has advantages because of the correction step; the filter can adapt new situations without new parameter identification used the regression method. It can be assumed that this effect will increase with a wider time horizon of the data deployed in the forecasting process. The proposed approach thus provides an alternative and efficient tool for forecasting water demand.

### References:

- [1] G. G. Archibald. Forecasting water demand - a disaggregated approach. *Journal of Forecasting*, 1983.
- [2] Beijing Municipal Bureau of Statistics. Beijing statistical information net. <http://www.bjstats.gov.cn/english/index.html>, May 2005.
- [3] R.G. Brown and P. Y.C. Hwang. *Introduction to Random Signals and Applied Kalman Filtering*. John Wiley & Sons, Inc., New York, NY, 2nd edition, 1983.
- [4] Z. Cook, S. Urban, M. Maupin, R. Pratt, and J. Church. Domestic, commercial, municipal and industrial water demand. Assessment and Forecast in Ada and Canyon Counties, Idaho; Idaho Department of Water Resources IDWR, 2001.
- [5] W. Du and P. Driessen. Speech enhancement based on kalman filtering and algorithm. In *IEEE Conf Communications, Computers and Signal Processing*, pages 142--145, May 1991.
- [6] M. Efe, D.P. Atherton, and J.A. Bather. Adaptive kalman filters for manoeuvring target tracking. In *IEE Colloquium Target Tracking and Data Fusion*, pages 1--7, June 1998.
- [7] A.C. Harvey. *Time Series Models*. Harvester Wheatsheaf, 2nd edition, 1993.
- [8] A.C. Harvey. *Forecasting structural time series models and the Kalman filter*. Cambridge University Press, Great Britain, 1990.
- [9] R.E. Kalman. A new approach to linear filtering and prediction problems. *ASME Trans. Basic Engineering*, pages 35-45, March 1960.
- [10] National Bureau of Statistics of China. China statistical information net, May 2005.
- [11] B. Sridhar, P. Smith, R.E. Suorsa, and B. Hussien. Multirate and eventdriven kalman filters for helicopter flight. *IEEE Mag. Control Systems*, 13:26--33, Aug. 1993.
- [12] G. Welch and G. Bishop. An introduction to the kalman filter. Technical Report 95-041, Dept. of Computer Science, University of North Carolina, 2006.
- [13] J.-Wernstedt. *Experimentelle Prozessanalyse*. Verl. Technik, Berlin, 1st edition, 1989.
- [14] S.-S. Xiong and Z.-Y. Zhou. Neural filtering of colored noise based on kalman filter structure. *IEEE Trans. Instrumentation and Measurement*, vol.52, pages 742-747, June 2003.

### Authors:

Marco Jacobi  
Divas Karimanzira  
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, divas.karimanzira}@tu-ilmenau.de





Jörg Meier / Rüdiger Bock / László G. Nyúl / Georg Michelson

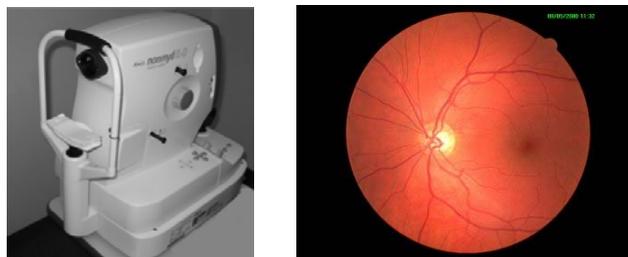
## Eye Fundus Image Processing System for Automated Glaucoma Classification

### INTRODUCTION AND VISION

Glaucoma is an eye disease that threatens the eyesight of the patients. As the disease progresses, nerve fibers in the retina die, which, if left untreated, leads to blindness. In Germany around 5% of the population (5 mill.) live with a glaucoma risk while around 800.000 people suffer from glaucomatous damages [1]. Although glaucoma cannot be healed, the progression can be stopped. Therefore, early detection of the disease is essential for preventing one of the most common causes of blindness. Glaucoma screenings based on digital images of the retina have been performed in the past few years in the clinics but they still lack robust automated assistance.

We devised an automated system that detects glaucomatous eyes based on acquired fundus images. In contrast to other approaches [2-5], we use image-based features of fundus photos that do not depend on exact measurements gained by segmentation techniques. This appearance based approach is new in the field of retina image processing.

Our vision is to establish a screening system that allows fast, robust and automated detection of glaucomatous changes in the eye fundus. Such a system could even be deployed in everyday environments, like shopping malls, to reach many people. It helps to discover suspected glaucomatous cases and warn the subject, so that careful evaluation can be done in time to control disease progression. This would not only reduce health care costs of treating glaucoma but would also prevent affected patients from vision loss. An acquisition device (Kowa NonMyd digital fundus camera) and an example image of the retina are shown in Figure 1.



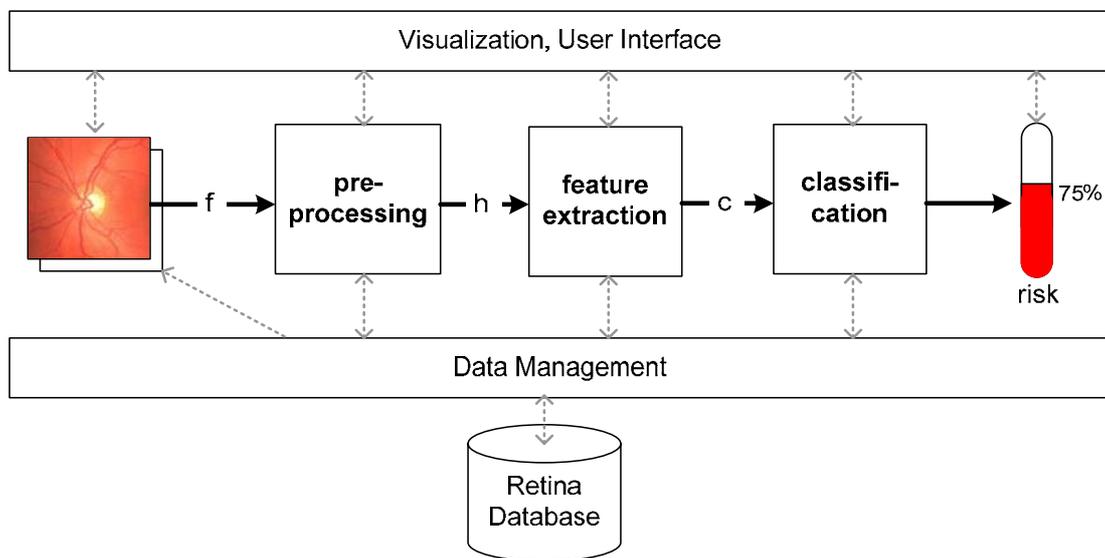
**Figure 1** Kowa Digital Fundus Camera and an acquired image example.

First, we briefly describe our processing system and the used methods. It is followed by an evaluation based on 200 images. By applying a 2-stage classification scheme we achieve a total classification correctness of 86 %.

## SYSTEM OVERVIEW

We devised a system for computer aided detection of **eye** diseases (called **CatEye**). It is a database driven framework to process, analyze and classify retina images. Existing functionality from tools, such as the ITK image processing toolkit [6] or Matlab can be easily integrated while the C++ framework provides image (and derived) data access (read and write) to the retina database. There are interfaces to implement image processing filters or methods to compute classification features. Programs can be fairly easily created to process large image sets from the database by applying certain filters or to visualize results.

The image processing pipeline follows the standard three-stage structure of (i) pre-processing, (ii) image-based feature extraction, and (iii) classification (see Figure 2). A brief description of the methods used in each step follows.



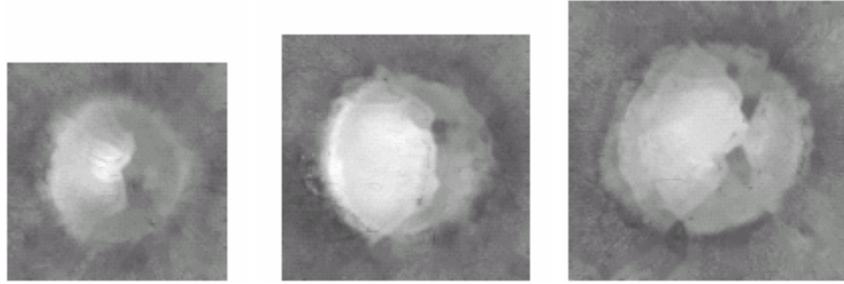
**Figure 2** CatEye (Computer Aided Detection of Eye Diseases): System Overview.

## PREPROCESSING

On one hand, nonuniform illumination is a general problem in retina image analysis. It is due to the small size of the objects and the complexity of the optical system (including both the camera and the eye) involved in the imaging process. Such inhomogeneities are corrected by robust homomorphic surface fitting [7].

On the other hand, blood vessels introduce a high variation in retina images which seems to be a distracting feature when diagnosing glaucoma. In our study, blood vessels are removed by computing a vessel mask and interpolating the missing pixel values by morphological inpainting [8].

The neuroretinal rim around the optic disc (papilla) is the most important region for glaucoma detection [9]. We normalize the images such that the papilla is centered and appears in the same size. This normalized input is required for the feature computation by appearance based approaches. Localization and size estimation is done by the method of [5]. Finally, all images are scaled to a fixed size of 128x128 pixels for feature extraction. Figure 3 shows example images after all steps are done except the final scaling step.



**Figure 3** Examples of images after preprocessing (illumination corrected, cropped ROI, and excluded vessels) showing different papilla sizes. First: a healthy retina, second: a glaucomatous optic disc, third: a healthy retina with macro papilla.

## FEATURE EXTRACTION

To capture different aspects of the image information, we use four types of feature extraction. (i) The first set of features is obtained by taking the pixel values directly as input to principal component analysis (PCA) which is used here as a dimensionality reduction technique. (ii) The second feature group comes from 28 Gabor texture filter responses [10] that represent spatial and spatial-frequency information of the data. The filter output is also compressed using PCA. (iii) The third set of features is computed from the coefficients of the Fast Fourier Transform (FFT) which contains translation invariant global frequency information. Again, PCA is used for data reduction. (iv) Histograms provide a compact summary of the intensity distribution in an image. In this application, they also show structural parts of the images (background, papilla rim, optic cup). A tri-modal Gaussian mixture is fitted to the histogram using maximum likelihood estimation and then the found distribution parameters, such as the mean and the variance values, serve as features.

## CLASSIFICATION

We found that a 2-stage classification scheme performs better than classification using any of the four feature sets alone or using a single pooled feature set. First, the four sets of feature values are classified separately by support vector machines (nu-SVM [11]). Then, the probability score of belonging to the glaucoma class, obtained from each of the four classifiers is taken as a new feature vector input to another classifier. This final SVM-classifier decides whether the sample is considered glaucomatous or not.

## EVALUATION AND RESULTS

For evaluation, we took 200 images (50 images each of healthy and glaucomatous eyes for training and a similar mixture for separate testing; age of the subjects:  $57 \pm 10$  years) randomly selected from the Erlangen Glaucoma Registry (EGR) that contains thousands of records of the eye ground of healthy subjects and patients having glaucoma. Diagnosis was done by an ophthalmologist using anamnesis, image data and other measurements. The images were acquired by a Kowa NonMyd digital fundus camera.

We computed the overall classification correctness and also the F-measure, which is the harmonic mean of sensitivity and precision, for healthy and glaucomatous eyes. The experiments were performed with a cross-validation on the separated test set. The performance of the classifications using one feature set only varies: total correctness of 73% with the histogram features, 76% with the FFT coefficients, 80% with the Gabor

textures and 83% with the pixel intensities. When applying the second classification step, 86% classification correctness is achieved with an F-measure of 83% for healthy and 88% for glaucomatous samples. This is similar to what was achieved by experienced human observers. According to [12], experts achieved by qualitative assessment of optic disc stereo photographs (63 normal and 29 glaucomatous subjects) an average F-measure of 91% for detecting normals and 79% for detecting glaucoma.

## CONCLUSION

We presented our automatic system for computer aided detection of eye diseases (CatEye) used to identify glaucomatous eyes in fundus photographs. The images can be acquired quickly and without any inconvenience to patients. The classification success rate of the system is comparable with that of experienced human observers. Thus, such a system can be deployed in large scale screening examinations for early detection of the disease. To our knowledge this is the first data-driven feature computation and classification system for glaucoma detection from retina images.

### Acknowledgments:

Images and diagnoses obtained from the Erlangen Glaucoma Registry. R. Bock was supported by the SFB 539, A4 (German Research Foundation). J. Meier is a member of the Int. Max Planck Research School on Optics and Imaging. L.G. Nyúl is a fellow of the Alexander von Humboldt-Foundation.

### References:

- [1] Initiativkreis zur Glaukomfrüherkennung, Germering, Germany, [www.glaukom.de](http://www.glaukom.de)
- [2] Lester, M., Swindale, N.V., Mikelberg, F.S.: Sector-based analysis of optic nerve head shape parameters and visual field indices in healthy and glaucomatous eyes. *J Glaucoma* 6(6) (Dec 1997) 370–376
- [3] Swindale, N.V., Stjepanovic, G., Chin, A., Mikelberg, F.S.: Automated analysis of normal and glaucomatous optic nerve head topography images. *Investig Ophthalmol Vis Sci* 41(7) (2000) 1730–1742
- [4] Zangwill, L.M., Chan, K., Bowd, C., Hao, J., Lee, T.W., Weinreb, R.N., Sejnowski, T.J., Goldbaum, M.H.: Heidelberg retina tomograph measurements of the optic disc and parapapillary retina for detecting glaucoma analyzed by machine learning classifiers. *Invest Ophthalmol Vis Sci* 45(9) (Sep 2004) 3144–3151
- [5] Chrástek, R., Wolf, M., Donath, K., Niemann, H., Paulus, D., Hothorn, T., Lausen, B., Lämmer, R., Mardin, C., Michelson, G.: Automated segmentation of the optic nerve head for diagnosis of glaucoma. *Med Image Anal* 9(4) (2005) 297–314
- [6] Yoo, T.: *Insight into Images: Principles and Practice for Segmentation, Registration, and Image Analysis*. A.K. Peters 2004
- [7] Narasimha-Iyer, H., Can, A., Roysam, B., Stewart, C.V., Tanenbaum, H.L., Majerovics, A., Singh, H.: Robust detection and classification of longitudinal changes in color retinal fundus images for monitoring diabetic retinopathy. *IEEE Trans Biomed Eng* 53(6) (2006) 1084–1098
- [8] Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C.: Image inpainting. In: *SIGGRAPH'00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. (2000) 417–424
- [9] Lester, M., Garway-Heath, D., Lemij, H.: *Optic Nerve Head and Retinal Nerve Fibre Analysis*. European Glaucoma Society (2005)
- [10] Jain, A., Farrokhnia, F.: Unsupervised texture segmentation using gabor filters. In: *Systems, Man and Cybernetics, 1990. Conference Proceedings.*, IEEE International Conference on. (4-7 Nov. 1990) 14–19
- [11] Chen, P.H., Lin, C.J., Schlkopf, B.: A tutorial on v-support vector machines. *Applied Stochastic Models in Business and Industry* 21(2) (2005) 111–136
- [12] Greaney, M.J., Hoffman, D.C., Garway-Heath, D.F., Nakla, M., Coleman, A.L., Caprioli, J.: Comparison of optic nerve imaging methods to distinguish normal eyes from those with glaucoma. *Invest Ophthalmol Vis Sci* 43(1) (2002) 140–145

### Authors:

Dipl.-Inf. Jörg Meier  
Dipl.-Inf. Rüdiger Bock  
Dr. László G. Nyúl  
Prof. Dr. Georg Michelson

University of Erlangen-Nuremberg, Institute of Pattern Recognition  
Martensstr. 3, 91058 Erlangen  
Phone: +49 9131 85 27882, Fax: +49 9131 303811  
E-mail: [joerg.meier@informatik.uni-erlangen.de](mailto:joerg.meier@informatik.uni-erlangen.de)

L. Hellrung / M. Trost

## **Automatic focus depending on an image processing algorithm for a non mydriatic fundus camera**

### **ABSTRACT**

This work presents an algorithm to focus a non mydriatic fundus camera automatically using an image processing approach to evaluate the actual state of an opto-mechanical focus aid. At present, the user has to adjust a focus and a positioning aid simultaneously. An automation of the focusing procedure would simplify the use of the device, especially for unpracticed users, and help to optimize the achievable image quality. A three-step algorithm is presented for detection of the interesting parts by adjustments of pattern matching, gradient search and blob analysis approaches. From image processing results a control sequence is generated for the focus motor of the camera device. The used methods were optimized and verified with a database of representative images. The efficiency and practical use of the implemented solution could be demonstrated by integration into the commercial fundus camera VISUCAM<sup>PRO NM</sup> by Carl Zeiss Meditec AG.

### **INTRODUCTION**

The fundus of a human eye means the complete inner surface of the eyeball characterizing the retina, optic nerve and macula. In medicine, fundus imaging is used for the diagnosis of different diseases. According to the patient's eye, the quality of an image depends on the positioning and focusing accuracy of the fundus camera [2]. For non-mydriatic fundus imaging, i.e. undilated pupils, the fundus is illuminated with infrared light to avoid eye reactions. Due to the lack of contrast and structure, it's not possible to visually focus the system in this mode. Therefore, aids for positioning and focusing are projected onto the retina [3]. With the help of the two aids, the correct adjustment has to be found by user.

The principle task to focus the camera correctly can be seen from Fig. 2-1)). The two white focus bars have to be brought to alignment by changing the focus motor position

manually.

Automation of the focusing procedure would simplify the handling of a fundus camera and help to optimize the achievable fundus image quality. A convenient automation approach has to deal with the tasks shown in Fig.1 in analogy to manual mode.

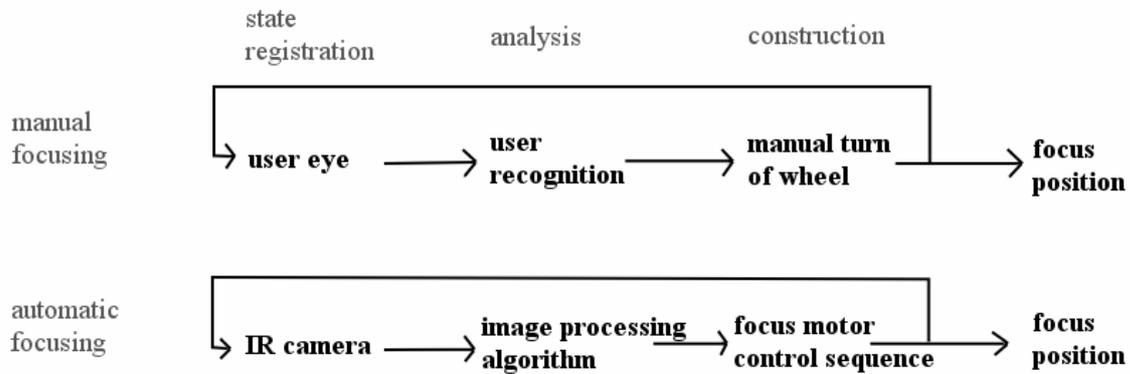


Fig. 1 Analogy of the tasks for manual and automatic focus procedure.

## METHODS

For specifying the position of the focus bars a three-step detection algorithm was developed. For each step of the algorithm significant classification parameters are identified. Therewith, all intermediate results are classified by a Linear Discriminant Analysis (LDA). The first step is a pre-classification by histogram parameters. Pictures not containing the focus aid because of wrong positioning are ejected.

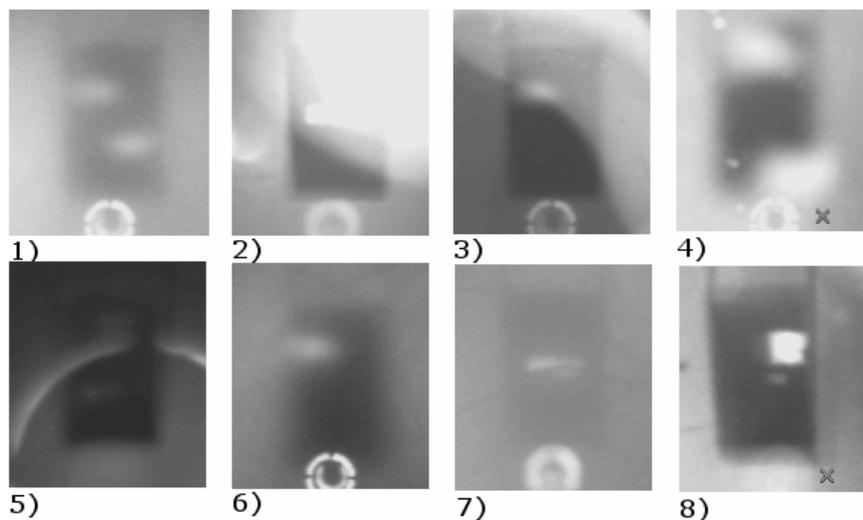


Fig. 2. Automatic evaluation of the IR images has to take different aspects into account. The principle task is to bring the two white focus bars to alignment (1). But the focus aid can be completely or nearly hidden (2,3), deformed (8) or the edges can be hidden by the focus bars (4). Additionally, the focus bars are not always visible (3, 6), have different appearances (8), are strongly smeared (4,6) or hidden (5,7).

Second, the algorithm seeks for the darker rectangle of the focus aid background. The detection of the focus aid is complicated mostly by missing information about the

position of the focus bars inside the focus aid and, additionally, by smearing effects and deformation with increasing ametropie of the eye (see Fig. 2-3,4,8)). The pattern matching uses a rectangular pattern with smoothed edges. For being invariant to focus bars positions, the normalized cross correlation is just calculated for the edge areas, leaving out the inner part of the pattern and the image part. Fig. 3 shows the principle rectangular pattern (1), the left out area when calculation correlation (2) and an exemplary correlation image (3).

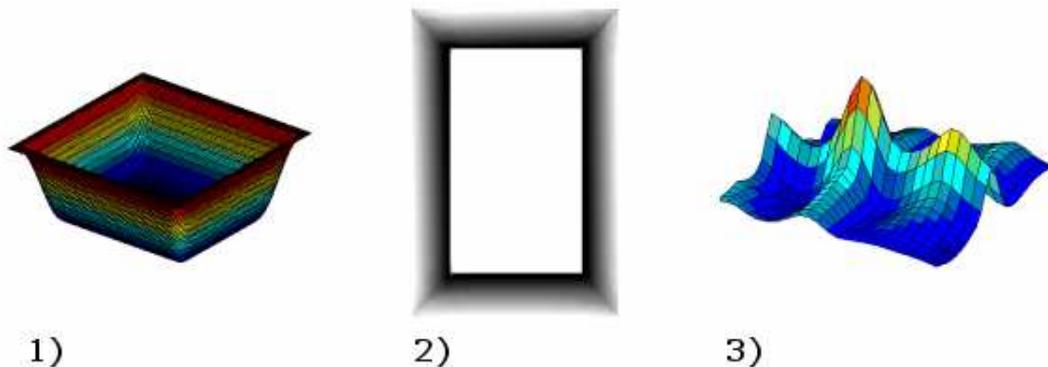


Fig. 3 Pattern matching approach with a smoothed rectangular pattern (1) for focus aid detection. For being invariant to the unknown position of the lightish focus bars inside, the normalized cross correlation is just calculated for the inner part (2). The result shows the correlation of the edge areas (3).

Additionally, a fast gradient search supports the decision of the pattern matching result. A gradient describes the neighbored grey value differences and can be minimal for light-dark crossing and maximal otherwise. That means, the horizontal (vertical) edges of the focus aid can be described ideally as minimum and maximum in the averaged gradient  $g(x)$  for all rows (columns). Additionally, an assumption about the approximated distance  $\Delta$  of the edges can be made (see Fig. 4-1)). Due to the influence of the focus bars and other noise, it's not enough to detect the global minimum and maximum of  $g(x)$  (see Fig. 4-3)). Therefore, this gradient search detects similar extrema with the defined distance by the following. The signal  $g(x)$  is mirror-inverted, i.e. two similar maxima with distance  $\Delta$  are expected. Then the signals  $g(x)$  and  $-g(x)$  are shifted about  $\Delta$  and the minimum operator is used to remove all extrema with incorrect distance (see Fig. 4-2)). The resulting maximum of the curve describes the position of the edge pair. Finally, the inaccuracy of an approximated distance is cleared by searching the neighbored minimum and maximum of the averaged gradient.

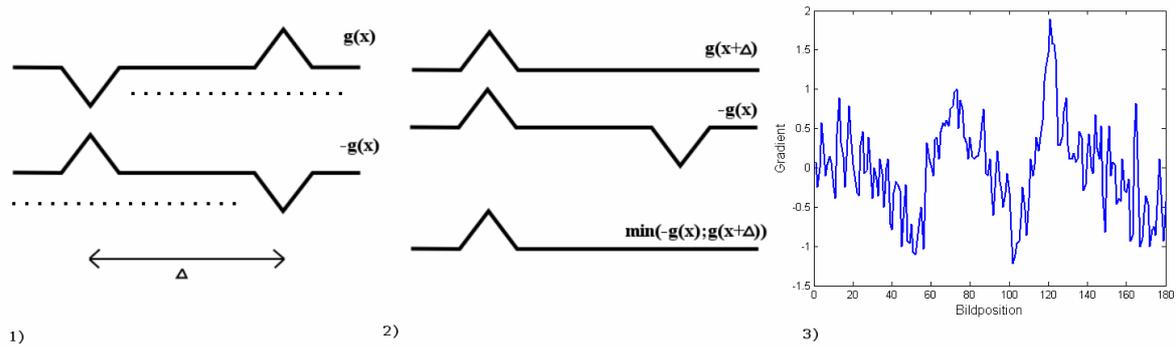


Fig.4. 1) The edge detection by gradient search is expecting ideally a minimum and maximum with distance  $\Delta$  for the averaged gradient of image columns (rows). The inverted signal  $-g(x)$  is showing a similar maximum with distance  $\Delta$  (pointed lines). 2) The signals  $g(x)$  and  $-g(x)$  are shifted about  $\Delta$ . Using the minimum operator leads to significant maximum for the correct edge position. 3) This approach is necessary because of the influence by focus bars and noise to the real gradient signal.

The resulting rectangle of the focus aid search borders the Region-of-Interest for the third algorithm step, the detection of the focus bars. Following aspects have to be taken into account: different forms (see Fig.1-4), 7) and 8)), the unknown number of visible focus bars (see Fig. 1-6)) and interfering objects (Fig. 1-5)). A normalized cross correlation with a gaussian filter, see Fig.5-1), forces all approximately elliptical elements of the chosen filter size. The correlation image, see Fig. 5-2), is binarized whereas correlation peaks produce areas possibly being a focus bar, Fig.5-3). Parameters for the classification of the possible focus bars are calculated by blob analysis. From this, different form parameters are used for LDA –classification to identify the correct focus bars.

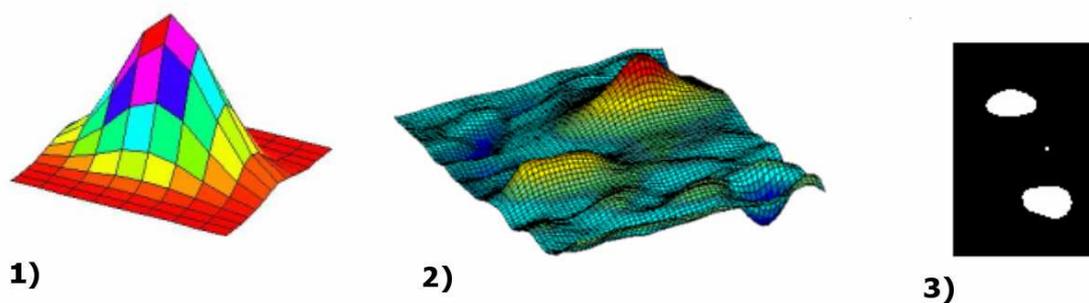


Fig. 5 Detection of the focus bars with normalized cross correlation with a Gaussian filter (1). The correlation image (2) is binarized and the resulting parts evaluated by blob analysis parameters (3).

If both focus bars could be detected, the positions are used to generate a control sequence for the focus motor. The new position is set and the resulting image is used for next control iteration until the focus bars are brought to alignment.

## RESULTS

To avoid wrong control sequences for automation the reliability of the found focus bar positions is most important in establishing the objectives of the algorithm. The reason for this is given by the preferably not possible than wrong automation result. In this case, it means the minimization of false-positive detected focus bars. Therefore, more than 700 pictures from different patients, including an artificial lens and a cataract diseased eye, and an artificial eye with variable ametropie were taken. The pictures were evaluated manually und divided for training and validation data sets to identify significant parameters differing correct and incorrect detections used by Linear Discriminant Analysis. Also, used filter parameters, distances and the LDA coefficients were optimized.

The possible quality of the discrimination between correct and incorrect detections for each algorithm part is shown in Fig. 6. All figures show an adequate possible differentiation between the correct and incorrect detections. Most important is the high reliability of focus bar detections in part 3) to avoid false-positive detections of the focus bars not suffering from extremely high false-negative detection rate.

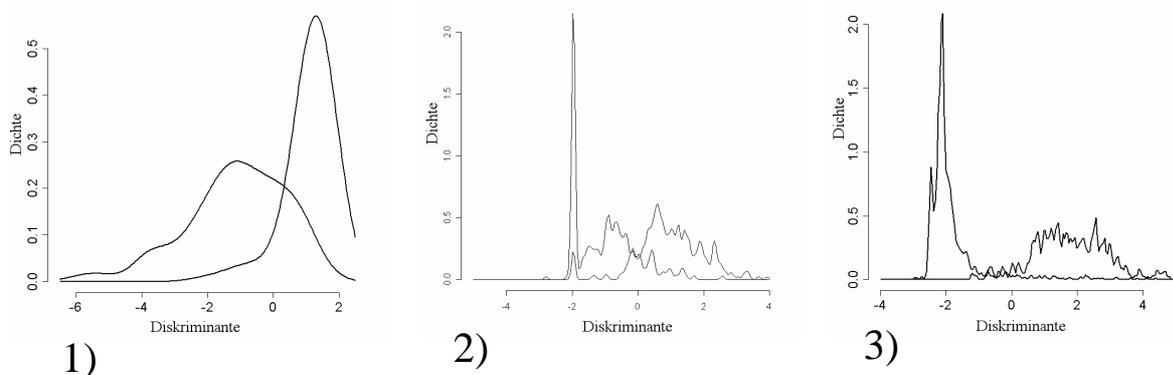


Fig. 6 Separability of significant parameters by Linear Discriminant Analysis (LDA) for each step of the algorithm, the pattern matching (1), gradient search (2) and focus bar detection by blob analysis parameters (3). Most important is the high discrimination of the real focus bars and detected noise.

After optimization, the entire algorithm leads to following results. The positive prediction value, meaning the correctness of a positive classification, is 95%. Additionally, the negative prediction value, meaning the correctness of a negative classification is 78,6%. Researches into the relation of focus bar distance and focus motor position addicted to a linear interrelationship to solve the generation of a control sequence.

## DISCUSSION

An implementation of the presented algorithm is integrated into the existing device software for VISUCAM<sup>PRO NM</sup> by Carl Zeiss Meditec AG. It focuses a non mydriatic fundus camera automatically without any hardware changes of the system device. The image processing results shows that the detection of the focus aid by pattern matching is well separable into correct and false results. The reliability is increased by the gradient search. Due to the high separability of the focus bar detection, an adequate accuracy for practical use is possible. The promising theoretical positive and negative prediction values are supported by first clinical validations.

A linear equation describes sufficiently the interrelationship of focus bar distance and focus motor position for a fast and accurate control.

Problems of the algorithm arises from patients with strongly diseased eyes, e.g. by cataract and very small pupils. In this case, just one focus bar is reliably visible or the background of the focus aid is not cognizable anymore because of very strong smearing. Then, the proposed three-step algorithm cannot detect the focus bars.

### References:

- [1] L. Hellrung, Echtzeit-Regelung einer Fokussiereinrichtung basierend auf der Analyse des Fokussierzustandes im Live-Bild einer Funduskamera, Diplomarbeit, Institut für Biomedizinische Technik und Informatik, TU Ilmenau, 2006
- [2] P. Weymar, Untersuchungen zu Einstellhilfen an non-mydriatischen Funduskameras, Diplomarbeit, Fachhochschule Jena, 2004
- [3] G. Link, D. Biernat, Fokussiervorrichtung für ophthalmologische Geräte, insbesondere Funduskameras und Verfahren zu dessen Verwendung, Offenlegungsschrift DE 10 2005 003 440 A1, Deutsches Patent- und Markenamt, 2005
- [4] R. C. Gonzalez and R. E. Woods, Digital Image Processing, Prentice Hall, Inc. Upper Saddle River, New Jersey, 2002

### Authors:

Dipl.-Inf. Lydia Hellrung  
Technische Universität Ilmenau  
Institut für Biomedizinische Technik und Informatik  
PF 100565  
98684 Ilmenau  
E-mail: lydia.hellrung@stud.tu-ilmenau.de

Dipl.-Ing. Michael Trost  
Carl Zeiss Meditec AG  
Göschwitzer Strasse 51-52  
07745 Jena  
E-mail: m.trost@meditec.zeiss.com

Matthias Hamsch, Claudia H. Igney, Marko Vauhkonen

## **A Magnetic Induction Tomography System for Stroke Classification and Diagnosis**

### **Abstract**

About 15 million people worldwide suffer a stroke every year. 5 million people die as a consequence of the stroke and another 5 million people become permanently disabled. While an ischemic stroke can be treated by a thrombolytic drug therapy within some hours after the stroke, the therapy would be lethal for a patient who had suffered a hemorrhagic stroke. Therefore a fast and reliable diagnosis is necessary to provide the patients with an adequate therapy.

One of the promising technologies for the fast classification of strokes is magnetic induction tomography (MIT). Magnetic induction tomography allows the reconstruction of conductivity distribution images for a wide variety of industrial and medical applications. In a medical application the MIT technique is used for acquiring information of conductivity distributions and conductivity changes in human tissue. The advantage of this technique is the contactless and non-invasive way of data collection. It is therefore also applicable for collecting information about the brain tissue.

The paper presents the setup of a 16 channel MIT system for analysis of conductivity distributions in human tissue. In addition to the description of the system measurement results of phantom objects and first reconstructed images are presented.

### **I. Introduction**

Annually, worldwide 15 million people suffer a stroke. Five million of these people die as a consequence of the stroke and another 5 million people become permanently disabled, placing a burden on family and community [1]. Beside reducing the major modifiable risk factors for stroke, like high blood pressure and tobacco use, it is important to provide a patient who suffered a stroke with an adequate therapy as fast as possible. While an ischemic stroke can be treated by a thrombolytic drug therapy within

some hours after the stroke, the therapy would be lethal for a patient who had suffered a hemorrhagic stroke. Therefore a fast and reliable diagnosis is necessary before providing the patient with the therapy.

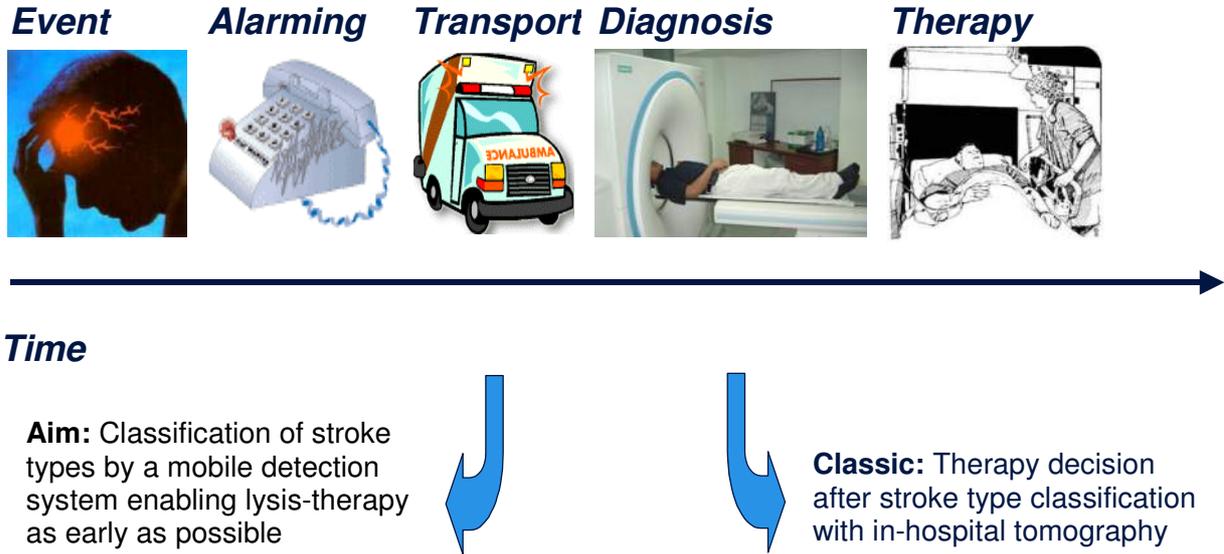


Fig.1 Care chain after a stroke event.

Figure 1 presents the classic care chain after a stroke event. After alarming the ambulance the patient is carried to a hospital for diagnosis and treatment. Depending on the distance to the next hospital and the availability of CT or MRI machines in the hospital several hours are lost before the first diagnosis. A mobile stroke diagnosis system could reduce the time from alarming the ambulance to the first diagnosis dramatically. The reduced time to therapy increases the probability of a complete recovery from the stroke and reduces costs for rehabilitation. Magnetic Induction Tomography (MIT) offers a new contactless technique for a decision support system inside the ambulance by enabling the mapping and imaging of the electromagnetic tissue properties.

## II. System Description

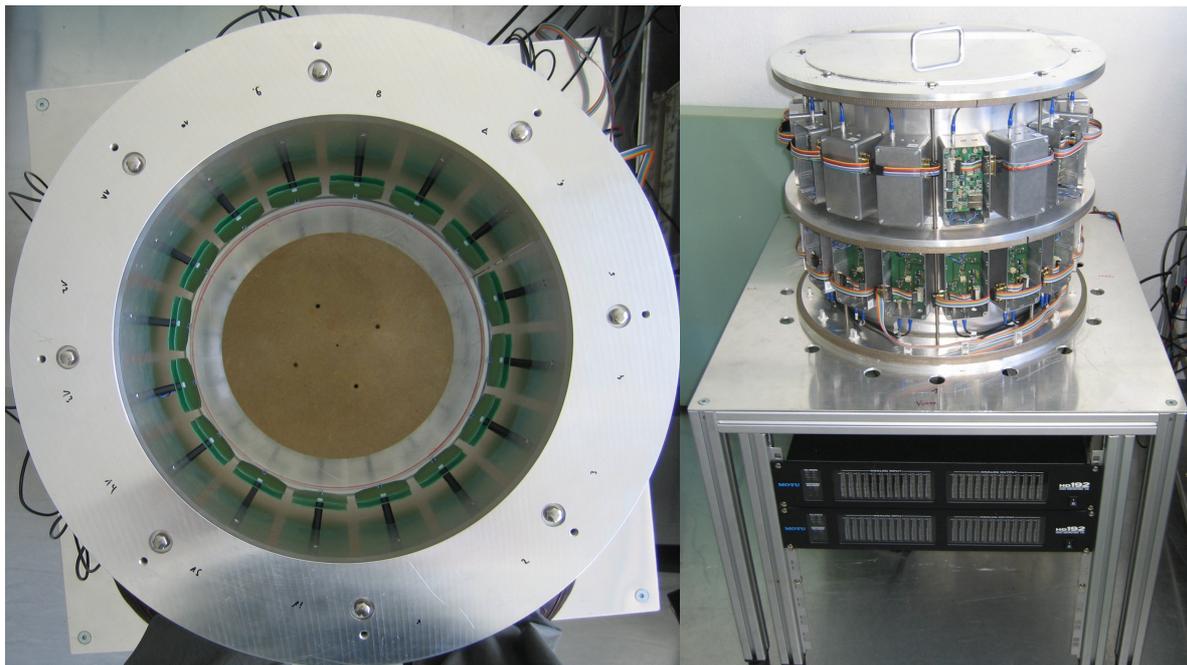
### A. Measurement Principle

The multi-coil system consists of excitation and receiver coils. A sinusoidal current flows through the excitation coil and generates a primary alternating magnetic field. Typical frequencies of the sinusoidal current are in the range from 1 MHz to 10 MHz. In a conductive medium, the alternating field induces eddy currents, which are proportional

to the conductivity of the medium. The eddy currents generate a secondary alternating electromagnetic field that is also proportional to the conductivity. If the skin depth is large compared to the thickness of the sample, which is generally true for biological tissues, the secondary field is nearly  $90^\circ$  phase shifted to the primary field [2] and can be detected by a receiver coil or a gradiometer. The measured secondary magnetic field gives an indication for the conductivity of the sample. If multiple coils are used for measuring the signal a conductivity distribution can be reconstructed.

A multi-frequency magnetic induction tomography system using planar gradiometers for the detection of the secondary magnetic field was introduced in [3]. The measurement system presented in this paper uses planar detection coils similar to the systems shown by Watson [4] or Korjnevsky [5], but improves the measurement speed by enabling a parallel readout of the 16 measurement channels.

### ***B. Mechanical setup***



*Fig.2 MI-Tomograph: top-view (left) and side-view with electronic modules (right)*

The Philips MI-Tomograph consists of a cylindrical shaped tank made of aluminum with excitation and receiver coils mounted on the inner wall of the tank. The inner diameter of the tank is 35 cm. In the current setup 16 excitation coils are arranged circularly, shaping the outer ring of coils. The inner circle of coils is formed by 16 receiver coils (see Fig. 2 left). The coils are made of PCB material (FR4, 1 mm thick). In the current



The second signal generator provides the sinusoidal signal (LO) for the downconversion of the measurement signal in the receiver module. It is split up by a power splitter and then distributed to the receiver modules, too. These modules include fixed gain amplifiers for the measured signal received by the receiver coil. The receiver modules implement a downconversion of the received radio frequency (RF) signal to a lower intermediate frequency (IF). The current setup uses an RF signal of 1 to 10 MHz and an IF signal of 10 kHz. The IF outputs of the receiver modules are connected to an A/D converter device. In the current setup an audio sampling equipment is used. It allows the parallel and synchronized sampling of up to 48 channels. In addition to the possible parallel sampling of the channels it features low noise (0.0005% THD+N) and 24-bit resolution of the sampled values at a sampling rate of 192 kHz. The sampled values are transferred to a PC for further signal processing. The PC also controls the entire measurement setup. The control software is written in National Instruments Labview, a commonly used graphical programming language for measurement automation. Beside controlling the signal generators and the excitation modules it also processes the sampled data. By using pre-calculated matrices the Labview software is capable of creating images of the conductivity distribution based on the measured data set directly.

### III. Reconstructed image

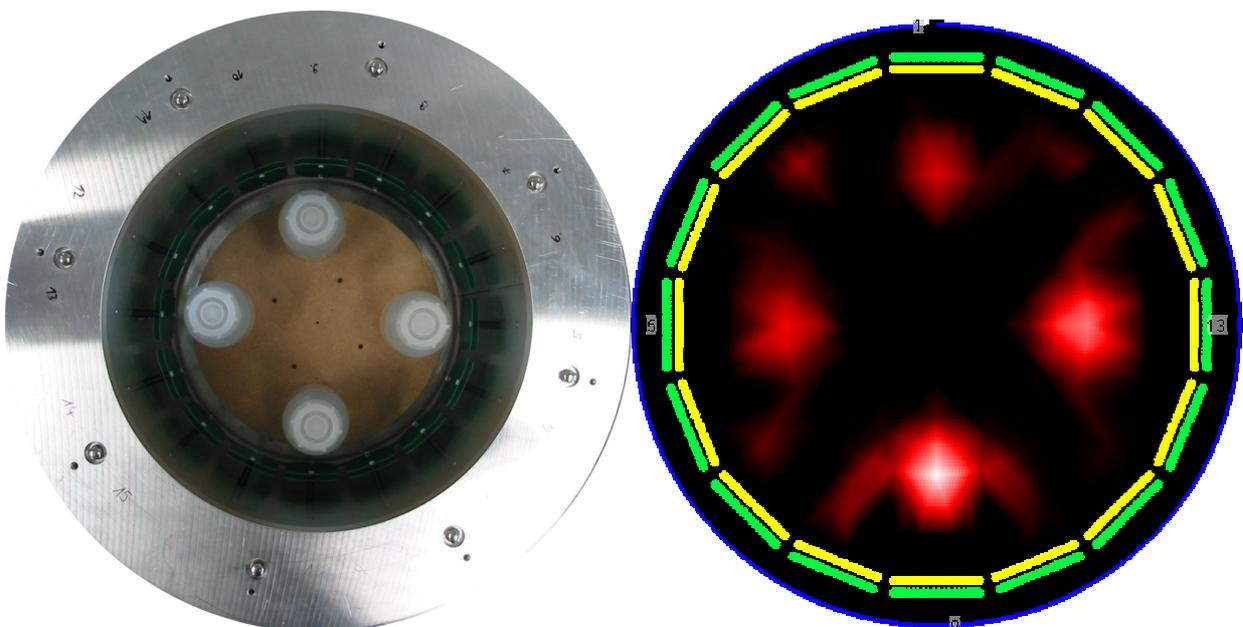


Fig.4 Measurement setup (left) and reconstructed image (right) of 4 plastic bottles filled with saline solution (conductivities 0.5, 0.75, 1 and 1.25  $\text{Sm}^{-1}$ )

Figure 4 shows the result of an experiment with four plastic bottles, each filled with 250 ml saline solution of different conductivities. The upper bottle was filled with  $0.5 \text{ Sm}^{-1}$  saline solution, the left one with  $0.75 \text{ Sm}^{-1}$ , the right one with  $1 \text{ Sm}^{-1}$  and the lower one with  $1.25 \text{ Sm}^{-1}$  solution. The left picture shows the arrangement of the plastic bottles in the tank. The right picture shows the reconstructed conductivity distribution. Details on the image reconstruction algorithm can be found in [6]. The rising conductivity from the upper to the lower bottle is clearly visible by the increasing intensity of the color.

#### IV. CONCLUSIONS

Magnetic induction tomography has the potential to enable a faster stroke diagnosis in the future. A prototype of an MIT system was presented in the paper. The prototype offers the possibility to analyze phantom objects and calculate conductivity distributions from the measured values. A reconstructed image of four phantom objects was shown. Compared to other existing MIT systems the parallel readout of all 16 measurement channels enables faster measurements. The measurement cycle needs only one-sixteenth of the time of a sequentially working system, which also leads to more reliable measurement results. Main focus of future work will be the reduction of temperature drift as well as phase noise in the system.

#### References:

- [1] Judith Mackay, George Mensah: The Atlas of Heart Disease and Stroke. World Health Organization. ISBN 978-9241562768
- [2] Griffiths H (2005) Magnetic induction tomography Electrical impedance tomography, Methods, history and Application. London IOP Press 213-238
- [3] Rosell-Ferrer J, Merwa R, Brunner P, Scharfetter H (2006) A multifrequency magnetic induction tomography system using planar gradiometers: data collection and calibration. *Physiol. Meas.* 27 271-280
- [4] Watson S, Williams RJ, Morris A, Gough W, Griffiths H (2002) The Cardiff magnetic induction tomography system. *Proc. Int. Fed. Med. Biol. Eng. EMBEC02, Vienna, Austria, Dec 4-8 (ISBN 3-901351-62-0) vol. 3 Part 1 pp116-7*
- [5] Korjenevsky A, Cherepenin V, Sapetsky S (2000) Magnetic induction tomography : experimental realisation. *Physiol. Meas.* 21 no 1 89-94
- [6] Vauhkonen M, Hamsch M, Igney CH (2007) Image reconstruction approaches for Philips magnetic induction tomograph. *Proceedings ICEBI, 2007, in press*

#### Authors:

Dr. Matthias Hamsch  
Dr. Claudia H. Igney  
Dr. Marko Vauhkonen  
Philips Research Europe, Weissshausstrasse 2  
D-52066 Aachen, Germany  
Phone: +49 241 6003754  
Fax: +49 241 6003518  
E-mail: [Matthias.Hamsch@philips.com](mailto:Matthias.Hamsch@philips.com)

T. Neumuth / A. Pretschner / O. Burgert

## **Surgical Workflow Monitoring with Generic Data Interfaces**

### **Introduction**

During the next years the digital operating room (DOR) and its future requirements will emerge and influences the technical development in the medical engineering field [1]. Workflow management systems (WfMS) supporting the data flow between surgical assist systems (SAS) during the surgical treatment of the patient are strongly involved in the future scenarios. WfMS can contribute to intraoperative interventions by synchronizing the surgical treatment of the patient with surgical process models (SPM) [2, 3] and by providing the relevant parts of patient models. The SPM represent knowledge of the general treatment support and the support of patient specific surgical decisions.

Due to the multiplicity of operating room equipment, and thus multiple heterogeneous data sources, no unique methodology for the acquisition, storage and analysis of data in operating rooms was established in practice. The already available data structures and communication protocols originate from the field of Picture Archiving and Communication Systems (PACS) related to radiology, and are standardized in DICOM [4], an ISO-standard for the description and exchange of images that aims at the interoperability of vendor-specific PACS systems. The DICOM-standard is currently extended for the use in surgical scenarios, currently ongoing as *Surgical-DICOM*-developments [1, 5].

One emerging issue related to WfMS is the situational awareness in the OR [1]: The WfMS has to recognise which part of the surgical procedure is being performed at the moment, and which parts have already been performed.

The definition of generic interfaces to the surgical workflow monitoring system and the connection to the peripheral systems are essential. Both of them are caused by the integration efforts for the infrastructure in the operating room and the need for workflow monitoring on a higher abstraction level during surgical interventions.

The objective of this publication is the proposal of a subpart for the infrastructure for the integration of device-specific metadata and higher-level sensor data into the Surgical

Workflow monitoring system offering a generic interface to the process.

### Infrastructural Scenarios and Use Case

Currently, the interoperability between SAS in operating rooms is very limited. Existing network capabilities are restricted either to data transfer capabilities such as DICOM, or to vendor specific conformance initiatives for interfaces between a few number of devices such as tracking and navigation devices. The existing infrastructure cannot be used for surgical workflow monitoring because of missing interfaces and networking capabilities of the systems.

Lemke and Vannier [1] proposed a Therapy Imaging and Model Management System (TIMMS, cf. Fig. 1) for the digital operating room (DOR). The functional support of model guided therapy and the added user value originating from advanced networking capabilities are essential to this infrastructure.

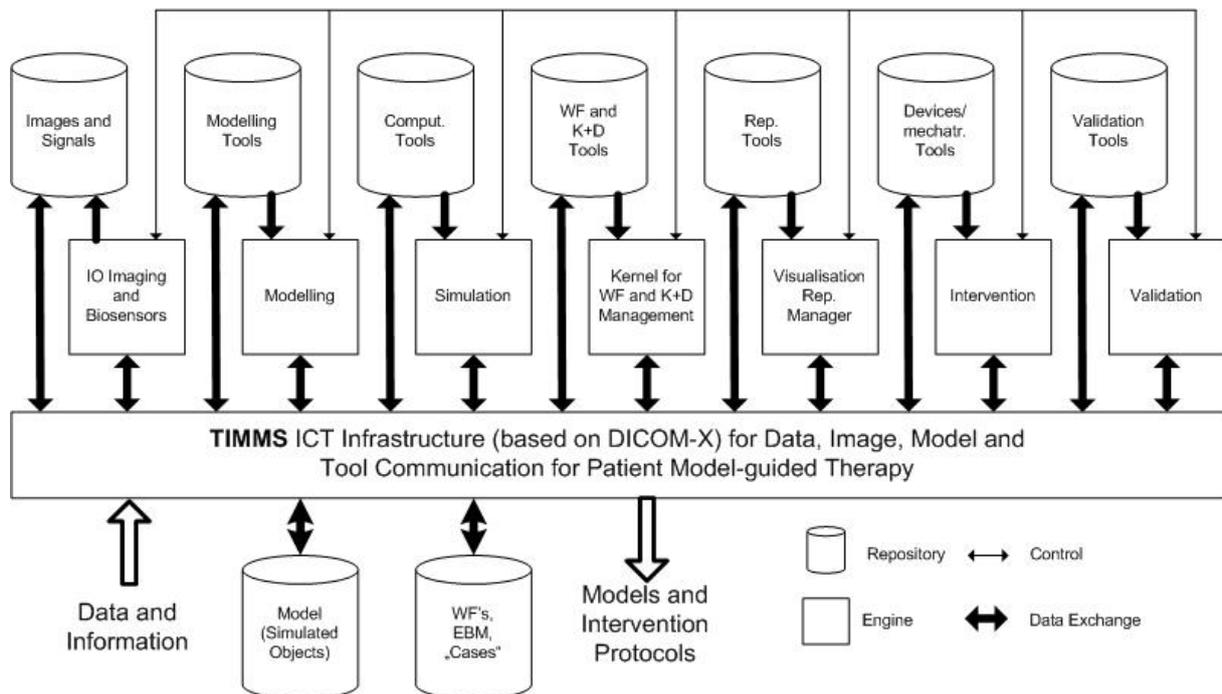


Figure 1: Therapy Imaging and Model Management System (TIMMS) [4]

One functional unit of the TIMMS infrastructure should be a module providing situational awareness [1]. For an intraoperative treatment this can be realized by a surgical workflow monitoring system.

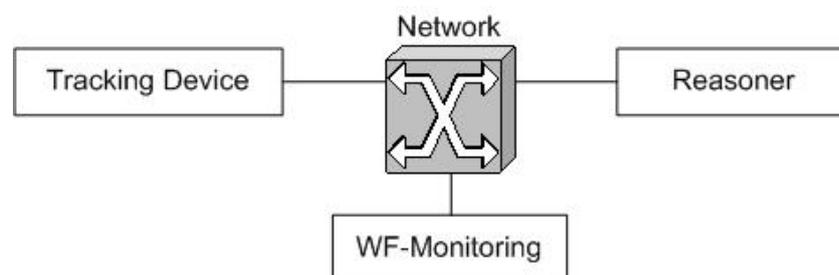
The surgical workflow monitoring system receives messages from peripherally distributed sensor systems and relates them to one another to form a higher description

context of the surgical treatment. The methodology is based on a structured transfer of metadata from stand-alone devices into the Surgical Workflow Monitoring system. Metadata designates the capabilities and interfaces of devices as well as (higher level) parameters and data types.

A simplified cut-out from the TIMMS infrastructure is shown in Fig. 2. To illustrate the cooperation of a surgical workflow monitoring system with the peripheral equipment in the OR, we present our approach with a use case of platform-variant tracking devices. The use case example adds distribution capabilities to situation recognition approaches such as proposed in [6, 7]. In the surgical workflow both approaches aim at the detection of phases based on machinery sensory data. The low level tracking data are obtained by sensor systems and need to be interpreted in a higher context. The decision which data have to be transferred over the network relies on the design of the overall system.

For the example set-up shown in Fig. 2 the following communication scenarios are possible:

1. Non-interpreted low-level sensory data are transmitted from the tracking device to the surgical workflow monitoring system and they are integrated into the actual surgical process model.
2. Non-interpreted low-level sensory data are transmitted from the tracking device to the reasoner and interpreted by the reasoner.
3. Interpreted higher level data are transmitted from the reasoner to the surgical workflow monitoring system and are integrated into the surgical process model.



**Figure 2: Infrastructural Scenario for Surgical Workflow Monitoring**

Although it is possible to transmit non-interpreted low-level sensory data to the monitoring system, it is not useful in every case. A motion tracking system monitoring the movement of an instrument in the surgeons main working hand might either transmit detailed data of the x-/y-/z-trajectories over time or just the context “cutting from ... to ...” that was interpreted by the reasoner. To transmit abstracted data such as the context

would require reasoning capabilities of the tracking system. If no reasoning capabilities are available, the low level sensor data might be transmitted to a reasoning module in the same or different network, interpreted, and then transferred into the monitoring system. In the latter case, one more component participates in the “process chain”.

### **Technical Approach to Solution**

The proposed use case demands a high flexibility originating from the multiple roles every participating system can play. Units need to behave as clients or servers, for some scenarios as both. The high complexity requires the use of established technologies. A technical approach designed for this challenge is OPC Unified Architecture [8]. OPC stands for *Object Linking and Embedding for Process Control* and was formerly based on the Microsoft component model (COM/DCOM). OPC Unified Architecture (OPC UA) is a platform independent standard for the communication of a multitude of systems and applications. Servers and clients are able to communicate beyond network environments. OPC UA defines service groups that are placed at disposal by a server which can furthermore be supported by clients. The specification of OPC UA especially emphasizes the independence of platforms and operating systems. A special communication interface enables server/client communication as well as a communication amongst various servers. The OPC UA server application programming interface (API) and the OPC client API divide the client and the server application on the one hand, and the communication stack on the other. Thereby, a high level of modularity is made available to the user. OPC - UA allows for the choice of the protocol: if both devices support OPC, than the binary protocol might be chosen; if that is not the case, one can fall back on Simple Object Access Protocol (SOAP). The interface lies underneath the user program and it is implemented via software. Due to this fact, it does not compete with field bus systems such as PROFIBUS or CAN, but combines application programs and sub-assembly drivers. The fact that the protocol between clients and servers can be chosen as binary protocol or web-service provides the system engineer with advanced design opportunities. Thus, the decision of using only pure middleware approaches such as Remote Procedure Calls (RPC), Common Object Request Broker Architecture (CORBA), Component Object Model (COM/DCOM), or Java Remote Method Invocation (Java RMI) or a pure web-services solution such as Simple Object Access Protocol (SOAP) turns into a choice for combining the advantages

of both worlds. The design choice relies especially on the required flexibility, advanced networking for passing firewalls, and the time-criticalness of transmission.

## **Conclusions**

The operating room of the future has a need for enhanced inter-device communication. Surgical Workflow Monitoring systems for high-level monitoring and data collection will become an important future part of the OR. These systems need to deal with a variety of devices, messages, and processes.

We propose the use of OPC UA as Service Oriented Architecture for Surgical Workflow Monitoring systems. The use of OPC UA enables the implementation of a standardized software interface for data exchange between applications of different vendors. OPC services can be extended to any platforms with XML and HTTP support as web service. Complementary to OPC UA several technologies necessary for networked ORs can be applied from already existing device description languages such as Field Device Tool (FDT) or Electronic Device Description Language (EDDL) [9]. That would also contribute to an integration of devices. Furthermore, a uniform interface for data acquisition and data analysis is available to the Surgical Workflow Monitoring system via OPC UA.

The approach highlights the abstraction of the device- and vendor-specific operating systems. That has several consequences, such as the unification of the heterogeneity of the used computer platforms; the concealment of the distribution of applications due to local transparency and uniform namespaces; and basic mechanisms for distribution, such as directory service, tracing service, and interface adoption.

The application is only achievable by agreement of different manufacturers. The design of the boundary conditions should reach two aims: on the one hand it should ensure the specific internal know-how of the manufacturers and on the other hand the surplus value of this new interface should lead to increased user value or saving of expenses.

## References:

- [1] Heinz U. Lemke, Michael W. Vannier. The Operating Room and the Need for an IT Infrastructure and Standards. *International Journal of Computer Assisted Radiology and Surgery* 1(3): pp.117-121, November 2006
- [2] Thomas Neumuth, Gero Strauß, Jürgen Meixensberger, Heinz U. Lemke, Oliver Burgert. Acquisition of Process Descriptions from Surgical Interventions. In: S. Bressan, J. Küng, R. Wagner (eds.). *Database and Expert Systems Applications. LNCS 4080* Springer, 2006: pp. 602-611.
- [3] Thomas Neumuth, Nico Durstewitz, Milos Fischer, Gero Strauß, Andreas Dietz, Jürgen Meixensberger, Pierre Jannin, Kevin Cleary, Heinz U. Lemke, Oliver Burgert. Structured Recording of Intraoperative Surgical Workflows. *SPIE Medical Imaging 2006 - PACS and Imaging Informatics, SPIE, 2006*, pp. CID 61450A.
- [4] DICOM. Digital Imaging and Communications in Medicine. <http://medical.nema.org/>, 2007.
- [5] Oliver Burgert, Thomas Neumuth, Michael Gessat, Stephan Jacobs, Heinz U. Lemke. Deriving DICOM Surgical Extensions from Surgical Workflows. *SPIE Medical Imaging 2007 - PACS and Imaging Informatics, SPIE, 2007*, pp. CID651604.
- [6] Gunther Sudra, M. Ott, S. Speidel, R. Dillmann. Situation Modelling and Situation Recognition for a Context-aware Augmented Reality System, *Proc. of CARS 2007*, to appear
- [7] Seyed-Ahmad Ahmadi, Tobias Sielhorst, Ralf Stauder, Martin Horn, Hubertus Feussner, Nassir Navab. Recovery of Surgical Workflow Without Explicit Models, *Proc. of MICCAI 2006, LNCS 4190*, pp. 420-428
- [8] OPC Foundation. OPC unified architecture specification part1: Concepts version 1.00. Technical report, [opcfoundation.org](http://opcfoundation.org), July 28, 2006.
- [9] Klaus Bender, Daniel Großmann, Benjamin Danzer. FDT+EDD+OPC UA = FDD UA Die Gleichung für eine einheitliche Gerätebeschreibung. *atp 2/2007, 2*: pp. 48-54, Februar 2007.

## Authors:

MEng Thomas Neumuth<sup>a</sup>  
Prof. Dr.-Ing. Andreas Pretschner<sup>b</sup>  
Dr.-Ing. Oliver Burgert<sup>a</sup>

<sup>a</sup>University of Leipzig, Innovation Center Computer Assisted Surgery, Ph.-Rosenthal-Str. 55, 04103 Leipzig

<sup>b</sup>HTWK Leipzig, Institute for Process Information Technology and Control Systems, Wächterstr. 13, 04109 Leipzig

Phone: +49-341-97-12010 / Fax: +49-341-97-12009

E-mail: [thomas.neumuth@medizin.uni-leipzig.de](mailto:thomas.neumuth@medizin.uni-leipzig.de)

M. Pfaff, D. Woetzel, D. Driesch, S. Toepfer, R. Huber, D. Pohlers, D. Koczan,  
H.-J. Thiesen, R. Guthke, R.W. Kinne

## **Gene Expression Based Classification of Rheumatoid Arthritis and Osteoarthritis Patients using Fuzzy Cluster and Rule Based Methods**

### **7 NEW METHODS AND TECHNOLOGIES FOR MEDICINE AND BIOLOGY**

With gene expression data becoming increasingly available in medicine there is an increasing need to adequately analyse this kind of data and use it for biomarker identification and classification with respect to specific diseases in order to support medical diagnosis.

We report here on an investigation into the use of gene expression information for the classification of rheumatoid arthritis (RA) and osteoarthritis (OA) patients applying fuzzy cluster and rule based methods.

The aim of this investigation was to identify gene expression patterns that can be used to classify RA and OA patients.

RA and OA both belong to the chronic rheumatic diseases. Whereas RA represents an inflammatory joint disease with an aggressive, joint destructive character (affecting approximately 1% of the population in industrialised countries), OA is a degenerative rheumatic disease with superimposed inflammatory flares.

Current medical diagnosis of the two diseases is primarily based on clinical data. In the future, this diagnosis may be supported by a gene expression based diagnosis. Such an approach may not only aid future differential diagnosis, but also provide clues to disease mechanisms and an improved therapy.

The study described here focused on the application of fuzzy cluster and rule based methods to reveal interrelations between gene expression in synovial tissue and the clinical diagnosis of RA and OA as well as a control group (CG; neither RA nor OA, mostly joint trauma or normal).

The original database used in part in this study was assembled at the Rudolf Elle Hospital Eisenberg, Germany, and covers 268 patients (107 RA, 118 OA, 36 CG and 7 other). Gene expression data was available for 33 of these patients (13 RA, 10 OA, 10 CG) and for 22,283 Affymetrix® U133A gene fragments each.

The data analysis approach applied consists of five consecutive steps: data pre-processing, clustering, rule extraction, rulebase construction with optimisation as well as cluster and rule based classification with validation.

Data pre-processing included logarithm calculation and median based normalisation of the gene expression intensities as well as outlier detection and removal.

Clustering of the data into two clusters (low and high expression of the respective gene) was performed applying a modified fuzzy c-means algorithm.

Extraction of uni-conditional rules from the clustered data was carried out using a modified *Kiendl* relevance index to rate and rank the relevant rules with the conclusions RA, OA, CG.

Rulebases were then constructed and optimised using the highest ranked rules and used for the classification of patients with respect to the diseases/control.

The obtained results were finally subjected to leave-one-out cross-validation.

The rule extraction step yielded three ranked rule lists consisting of 45, 27, 7 rules with the conclusion CG, RA, OA, respectively.

Figure 1 shows the main results of the analysis. Patients are represented in the rows and genes in the columns. Genes, i.e. uni-conditional rules with the respective gene in the conditional part, are arranged from left to right for the respective patient group (CG, RA, OA) according to the *Kiendl* relevance index as calculated in the rule extraction step (highest *Kiendl* values left). The figure only includes relevant genes, i.e. those with a *Kiendl* relevance index >0. The patients are arranged from top to bottom according to the clinical diagnosis (CG, RA, OA; see also Status). For each patient and each relevant gene, the fuzzy membership value calculated in the preceding fuzzy c-means clustering step is plotted ('Heatmap' values between 0 and 1; see left scale).

Since Figure 1 shows clearly contrasting blocks for the three patient groups along the diagonal based on the extracted relevant genes, these genes can be used to reliably classify the large majority of CG, RA, OA patients. It can be seen that only few patients, e.g. the last patient in the RA group (no. 23) and the last one in the OA group (no. 33), can not be reliably classified based on these genes.

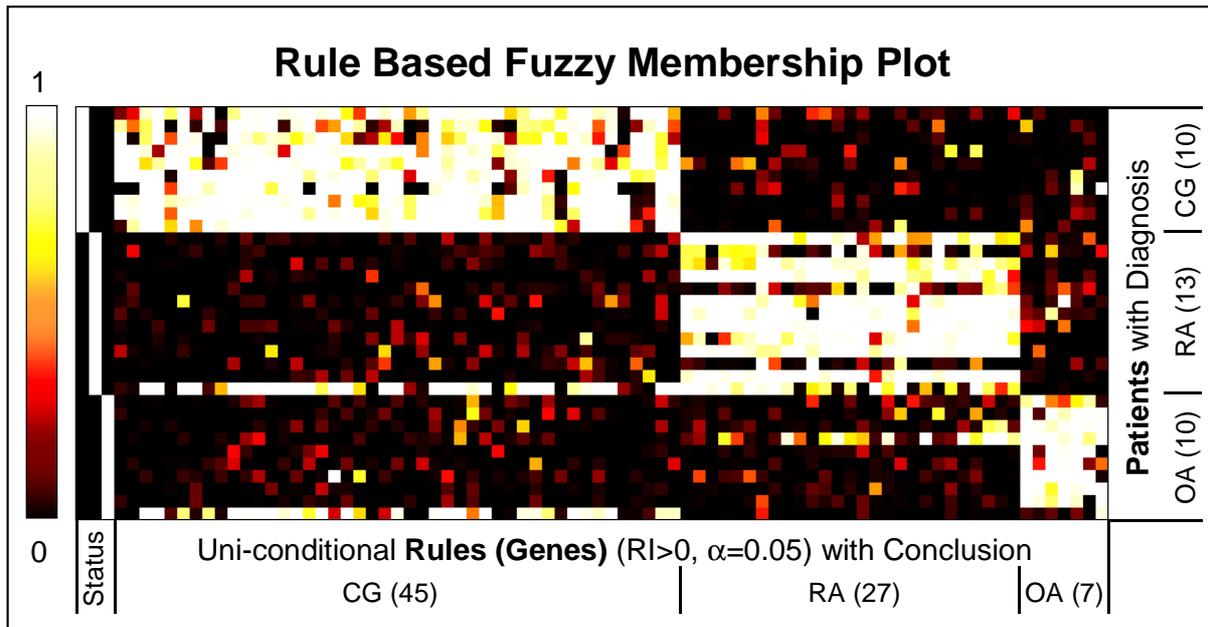


Figure 1: Rule based fuzzy membership plot for 33 patients and all 79 relevant genes

More specifically, the results are as follows:

Classification using the rulebase with all 79 extracted relevant rules (45 for CG, 27 for RA, 7 for OA) as shown in Figure 1 yielded a modelling error of 9%. 3 out of the 33 patients were incorrectly classified: no. 23 (actually RA, classified CG), no. 27 (actually OA, classified RA) and no. 33 (actually OA, classified CG), i.e. specificity was 100% and overall sensitivity 91%, with RA sensitivity 92% and OA sensitivity 80%.

Classification using an optimised rulebase with the 7 top ranked rules each for CG, RA, OA, i.e. altogether only 21 rules (Figure 1: the 7 rules furthest left for CG and for RA and all 7 rules for OA) resulted in a modelling error of just 3%. Now only 1 patient out of the 33, no. 33 (actually OA), was incorrectly classified as CG. Specificity here was also 100%, while overall sensitivity improved to 96%, with RA sensitivity improved to 100% and OA sensitivity to 90%.

The results obtained using the optimised rulebase were validated by leave-one-out cross-validation. This yielded a generalisation error of 24%. The incorrect classifications (8) occurred when patients no. 7 (actually CG), no. 12, 15, 21, 23 (all actually RA) and no. 27, 30, 33 (all actually OA) were left out. Among these 8 patients were those 3 who were, in the same way, also incorrectly classified when the original rulebase was used (no. 23, 27, 33). The additional 5 incorrectly classified patients were no. 7 (actually CG, classified OA), no. 12, 15, 21 (all actually RA, all classified OA) and no. 30 (actually OA, classified CG).

In summary, the study established a viable classification approach for this particular kind of gene expression data in conjunction with clinical diagnosis data employing a blend of unsupervised methods (data pre-processing and clustering) and supervised methods (rule extraction and rulebase construction). Although the analysis of 33 patients already provided good classification results, further validation with additional gene expression data for larger numbers of patients is required to finally establish gene expression based decision support systems for the differential diagnosis of rheumatic diseases.

**Authors:**

Dr Michael Pfaff <sup>1</sup>, Dirk Woetzel <sup>1</sup>, Dominik Driesch <sup>1</sup>, Dr Susanne Toepfer <sup>1</sup>,  
Dr René Huber <sup>2</sup>, Dr Dirk Pohlens <sup>2</sup>, Dr Dirk Koczan <sup>3</sup>, Prof Hans-Juergen Thiesen <sup>3</sup>,  
Dr Reinhard Guthke <sup>4</sup>, Prof Raimund W. Kinne <sup>2</sup>

<sup>1</sup> BioControl Jena GmbH  
Wildenbruchstr. 15  
D-07745 Jena, Germany  
Phone: +49-3641-527831  
Fax: +49-3641-527832  
E-mail: michael.pfaff@biocontrol-jena.com

<sup>2</sup> Friedrich Schiller University Jena  
Medical Faculty  
Department of Orthopaedics  
Experimental Rheumatology Unit  
Waldkrankenhaus 'Rudolf Elle' gGmbH  
Klosterlausnitzer Str. 81  
D-07607 Eisenberg, Germany

<sup>3</sup> University of Rostock  
Medical Faculty  
Institute for Immunology  
Schillingallee 70  
D-18055 Rostock, Germany

<sup>4</sup> Leibniz Institute for Natural Product Research  
and Infection Biology e.V. - Hans Knoell Institute  
Department of Molecular and Applied Microbiology  
Beutenbergstr. 11a  
D-07745 Jena, Germany

Susanne Toepfer, Sebastian Zellmer, Dominik Driesch, Dirk Woetzel,  
Reinhard Guthke, Rolf Gebhardt, Michael Pfaff

## **A 2-Compartment Model of Glutamine and Ammonia Metabolism in Liver Tissue**

### **7 NEW METHODS AND TECHNOLOGIES FOR MEDICINE AND BIOLOGY**

#### **Introduction**

The aim of this modelling study was to develop an adequate model of glutamine and ammonia metabolism in perfused rat livers under particular consideration of the metabolic zonation in the liver lobulus.

The liver is the main organ of intermediary metabolism. It consists of thousands of identical small sub-units, the liver lobuli. Each lobulus consists of a central vein and is confined by portal fields in a hexagonal arrangement. From each portal field, the blood flows into the lobulus and is drained through the central vein. The area surrounding the portal field is called the periportal region and the area around the central vein is called the perivenous region. During passage from the portal field to the central vein, blood passes a number of hepatocytes. Hepatocytes realize almost all metabolic functions, e.g. synthesis and degradation of amino acids, proteins, lipids and carbohydrates [1]. Due to highly regulated enzyme expression [2], most metabolic pathways are only present either in the periportal or the perivenous compartment of the lobulus. This is called metabolic zonation. As a consequence, synthesis and degradation are simultaneously carried out but in different compartments of the lobuli. One example where this metabolic zonation has an important impact is glutamine and ammonia metabolism. In the organism toxic ammonia is removed by converting glutamate into glutamine. The glutamine-rich blood enters the liver at the periportal region and the enzyme glutaminase (G) degrades glutamine to ammonia and glutamate. Subsequently, toxic ammonia is eliminated by the enzyme carbamoyl-phosphate synthetase (C) in the course of urea synthesis [1, 3]. Remaining ammonia is detoxified in periportal hepatocytes by converting glutamate into glutamine by the glutamine synthetase (GS).

## Experimental Data

The model identification presented here is based on experimental data published in [3]. In [3], the metabolic zonation was investigated in perfused rat livers by comparing metabolic flux rates depending on the perfusion direction. The isolated rat livers were perfused without recirculation either in physiological antegrade (from portal to central vein) or in retrograde (from central to portal vein) direction with a weight-specific constant perfusion flow  $F_{mL}$ . For several experiments, either glutamine synthetase or carbamoyl-phosphate synthetase were inhibited by addition of methionine sulfoximine to the perfusion fluid or by using a bicarbonate-free buffer, respectively. Modelling is based on measured glutamine (GLN), ammonia (NH<sub>4</sub>), urea (UREA) and glutamate (GLU) concentrations in the influent and effluent of the perfused rat livers recorded from 16 perfusion experiments (Fig. 2, blue bars). All data were obtained in metabolic steady-state and are means from 4-12 separate runs. Data preprocessing was performed by scaling measured effluent concentrations in order to guarantee that the overall sum of nitrogen in the influent together with the estimated endogenously arising nitrogen is equal to the overall sum of nitrogen in the effluent.

## Knowledge- and Data-driven Modelling

Model development comprises the knowledge-driven determination of the model structure and the subsequent data-driven estimation of the model parameters.

The model consists of eight differential equations with one equation for each concentration  $c_{GLN,i}$ ,  $c_{NH_4,i}$ ,  $c_{UREA,i}$ ,  $c_{GLU,i}$  and for each compartment ( $i = 1,2$ ), respectively. Both compartments are run through by the perfusion flow  $F$  carrying the concentrations  $c_{j,Input}$  in influent ( $j = GLN, NH_4, UREA, GLU$ ). During antegrade perfusion, the effluent concentrations  $c_{j,Output}$  are equal to the concentrations  $c_{j,2}$  in compartment 2 (Fig. 1), while during retrograde perfusion, they are equal to the concentrations  $c_{j,1}$  in compartment 1 (not shown).

The periportal compartment 1 with the volume  $V_1$  includes glutamine degradation by glutaminase and urea synthesis by carbamoyl-phosphate synthetase. In contrast, the perivenous compartment 2 with the volume  $V_2$  contains glutamine synthesis by the enzyme glutamine synthetase (Fig. 1). The corresponding reaction equations consider stimulation (Michaelis-Menten equation with Michaelis-Menten constant  $K_M$ ) and inhibition (noncompetitive inhibition with dissoziation constant  $K_I$ ) by substrates and

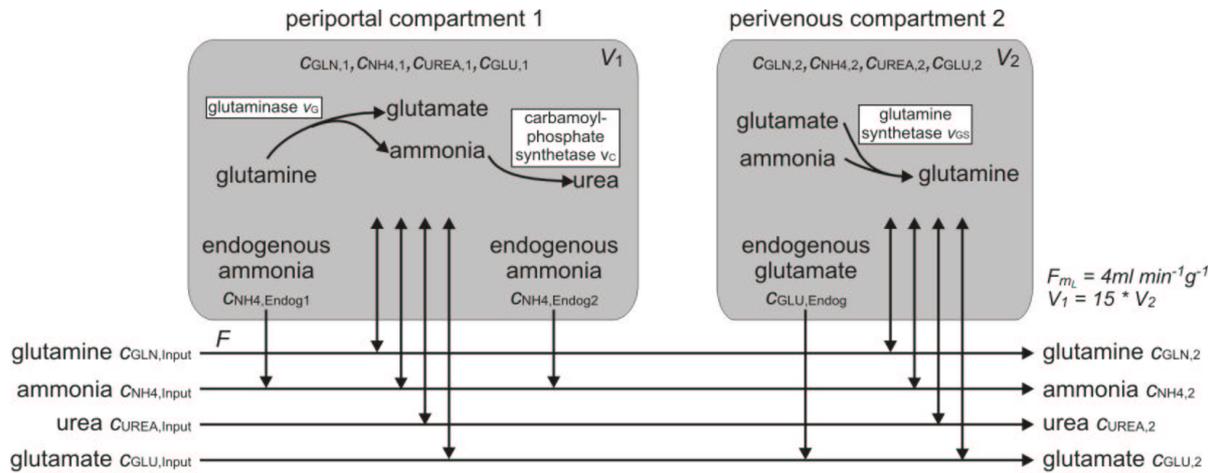
products. Thus, glutamine degradation  $v_G$  has a maximum velocity  $v_{MAX,G}$  and is stimulated by glutamine ( $K_{M,G,GLN}$ ) and portal ammonium ions ( $K_{M,G,NH4}$ ) [3, 4]. Urea synthesis  $v_C$  is characterized by the maximum velocity  $v_{MAX,C}$  and the stimulation by ammonia ( $K_{M,C,NH4}$ ). Glutamine synthesis  $v_{GS}$  has the maximum velocity  $v_{MAX,GS}$  and is stimulated by ammonia ( $K_{M,GS,NH4}$ ) and glutamate ( $K_{M,GS,GLU}$ ). Furthermore, it is inhibited by the product glutamine ( $K_{I,GS,GLN}$ ) [4]. Consequently, each differential equation can consist of three parts. The first part describes concentration changes caused by synthesis and degradation from the three metabolic reactions (G,C,GS) that were taken into account. The both other parts describe the change of concentrations caused by the inflow and the outflow of the perfusion stream  $F$ .

Endogenous ammonia arises from endogenous protein and amino acid breakdown [5]. Endogenous ammonia production is located mainly in the periportal compartment, which consists of more than 95 % of all hepatocytes of the glutamine/NH<sub>4</sub> system. Modelling the observed effects caused by endogenous ammonia requires two positions where the endogenous ammonia portions  $c_{NH4,Endog1}$  and  $c_{NH4,Endog2}$  operate. Furthermore, several experiments point to an endogenous source of glutamate  $c_{GLU,Endog}$  in the perivenous compartment 2 (Fig. 1).

During experiments with inhibition of glutamine synthesis or urea synthesis, the respective metabolic reactions are only active to a very small degree. In order to model this effect, the synthesis rates  $v_{GS}$  and  $v_C$  are multiplied with the additional activity factors  $e_{GS}$  and  $e_C$ . During the simulation of experiments without inhibition, these factors are set to one and, thus, have no influence on the synthesis rates. In case of simulating experiments with inhibition, the activity factors possess values much smaller than one and, therefore, decrease the synthesis rates.

A common model for all 16 experiments was identified. The parameters of the model were estimated by one of the following methods: (i) Parameters are known or can be calculated from known experimental settings. (ii) Parameters can be directly calculated from the measurement data. (iii) Parameters are optimized by fitting the model output to the measurement data using a nonlinear optimization technique.

The first method is applied for the yield coefficients  $Y$  and the terms  $F/V_i$  that are constant for all experiments due to the weight-specific perfusion flow  $F_{mL}$ . The second method is used to estimate the total levels of endogenous ammonia and glutamate. All remaining model parameters are optimized by minimizing the root square error (RSE)



periportal compartment 1

$$dc_{GLN,1}/dt = -1/Y_G \cdot V_G - F/V_1 \cdot c_{GLN,1} + F/V_1 \cdot c_{GLN,Input}$$

$$dc_{NH4,1}/dt = V_G - 1/Y_C \cdot V_C \cdot \theta_C - F/V_1 \cdot c_{NH4,1} + F/V_1 \cdot (c_{NH4,Input} + c_{NH4,Endog1})$$

$$dc_{UREA,1}/dt = V_C \cdot \theta_C - F/V_1 \cdot c_{UREA,1} + F/V_1 \cdot c_{UREA,Input}$$

$$dc_{GLU,1}/dt = V_G - F/V_1 \cdot c_{GLU,1} + F/V_1 \cdot c_{GLU,Input}$$

perivenous compartment 2

$$dc_{GLN,2}/dt = v_{GS} \cdot \theta_{GS} - F/V_2 \cdot c_{GLN,2} + F/V_2 \cdot c_{GLN,2}$$

$$dc_{NH4,2}/dt = -1/Y_{GS} \cdot v_{GS} \cdot \theta_{GS} - F/V_2 \cdot c_{NH4,2} + F/V_2 \cdot (c_{NH4,1} + c_{NH4,Endog2})$$

$$dc_{UREA,2}/dt = -F/V_2 \cdot c_{UREA,2} + F/V_2 \cdot c_{UREA,1}$$

$$dc_{GLU,2}/dt = -1/Y_{GS} \cdot v_{GS} \cdot \theta_{GS} - F/V_2 \cdot c_{GLU,1} + F/V_2 \cdot (c_{GLU,1} + c_{GLU,Endog})$$

$$V_G = V_{MAX,G} \cdot c_{GLN,1} / (K_{M,G,GLN} + c_{GLN,1}) \cdot c_{NH4,Input} / (K_{M,G,NH4} + c_{NH4,Input})$$

$$V_C = V_{MAX,C} \cdot c_{NH4,1} / (K_{M,C,NH4} + c_{NH4,1})$$

$$v_{GS} = V_{MAX,GS} \cdot c_{NH4,2} / (K_{M,GS,NH4} + c_{NH4,2}) \cdot c_{GLU,2} / (K_{M,GS,GLU} + c_{GLU,2}) \cdot K_{i,GS,GLN} / (K_{i,GS,GLN} + c_{GLN,2})$$

Fig. 1: Structure of the 2-compartment liver model (antegrade perfusion)

between the measured and preprocessed concentrations  $c_j^k$  and the simulated model output  $c_{j,Output}^k$  for experiments  $k$  ( $k = 1, \dots, 16$  and  $j = GLN, NH4, UREA$ ).

It should be mentioned, that the model output  $c_{GLU,Output}$  can not be adapted with an acceptable quality to the measured glutamate concentrations. It is assumed, that in contrast to the GLN, NH4 and UREA release, the glutamate level is strongly influenced by other pathways not taken into account here. However, the model can simulate the effects of the three considered metabolic reactions on the release of glutamate.

## Results

The modelling results are shown in Fig. 2. The 2-compartment model developed is able to cover all main effects with respect to glutamine, ammonia and urea release observed during the experiments with changing perfusion direction. The following effects are described in [3] and are captured by the model:

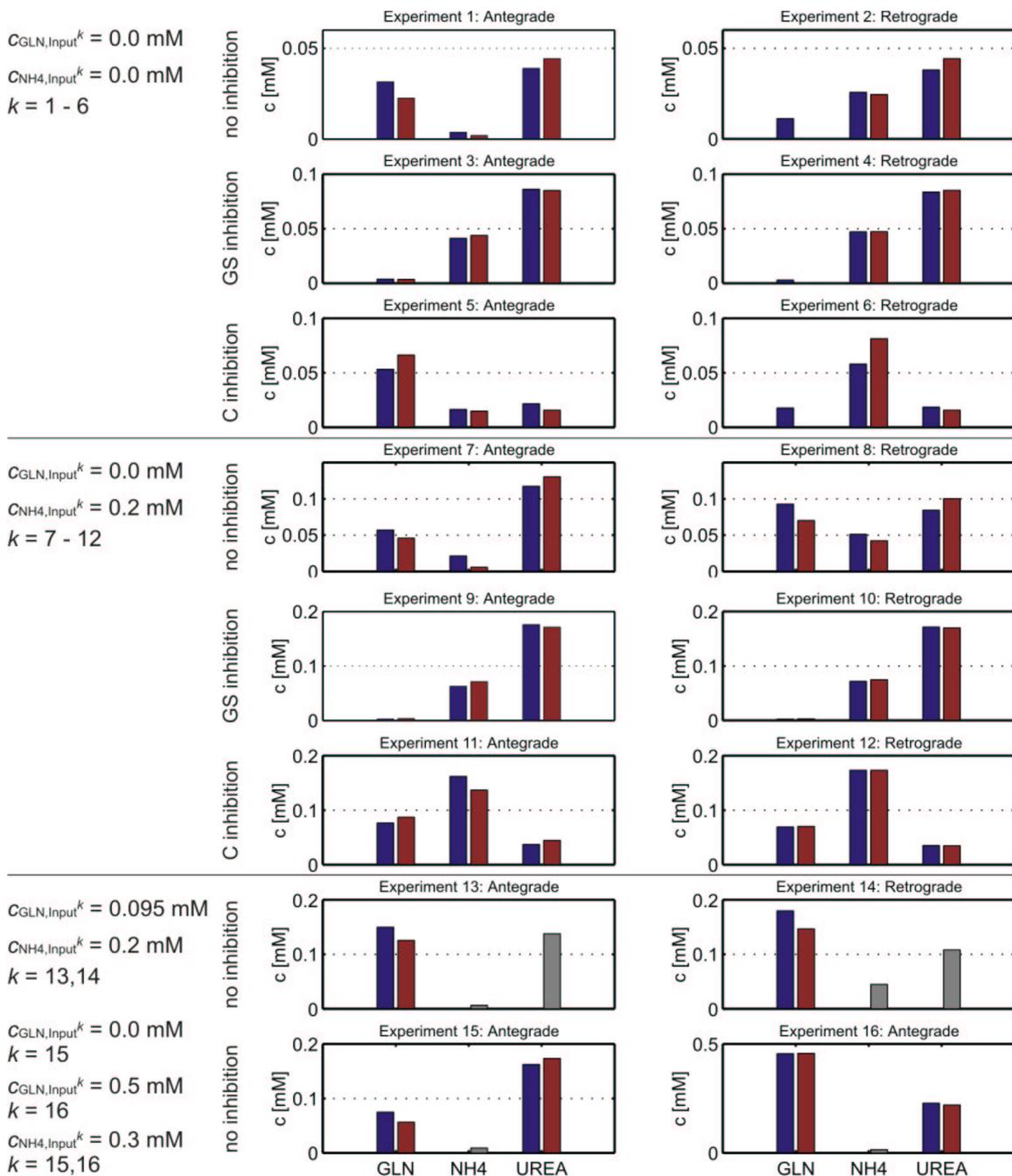


Fig 2: Results of the 2-compartment model (RSE = 0.356); blue = measurement data, red = model output, gray = model output, but no measurement data available

- Ammonium ions can be very efficiently eliminated during antegrade perfusion but are released by the liver to a considerable extent during retrograde perfusion (Exp. 1 and 2, Exp. 7 and 8 in Fig. 2).
- Endogenous ammonium ions are utilized for synthesis of glutamine during antegrade perfusion but are washed out during retrograde perfusion (Exp. 5 and 6).
- Added ammonium ions are mainly converted to urea during antegrade perfusion but are converted to glutamine and urea during retrograde perfusion (Exp. 7 and 8).
- During the inhibition of the competing pathways urea and glutamine synthesis, the

direction of perfusion does not influence the release of glutamine or urea (Exp. 3 and 4, Exp. 9 and 10, Exp. 11 and 12).

- During physiological antegrade perfusion, urea synthesis can remove high ammonia concentrations but is restricted to levels above approximately 0.05 mM (Exp. 3 and 9). The subsequent glutamine synthesis removes very efficiently low ammonia concentrations which were not eliminated by urea synthesis (Exp. 1 and 7). This feature is obtained by the Michaelis-Menten constants with  $K_{M,C,NH_4} \gg K_{M,GS,NH_4}$ .

## Conclusions

Comparisons of these results with a 1-compartment model show that this simpler model is structurally unable to cover the observations made during the perfusion experiments. That means, the 1-compartment model (even if including the same metabolic reactions as the 2-compartment model) can not be fitted to the measurement data with an acceptable modelling error. Consequently, an adequate liver model of glutamine and ammonia metabolism requires the structural mapping of metabolic zonation. By using expert knowledge and experimental data, a 2-compartment model covering all main observed effects could be developed. This 2-compartment model can be used as a general tool for simulation, e.g. defects in metabolic reactions that decrease the ammonia uptake.

## References:

- [1] Gebhardt, R.: Metabolic Zonation of the Liver: Regulation and Implications for Liver Function. *Pharmac. Ther.* 53, 275-354, 1992.
- [2] Braeuning, A., Ittrich C., Köhle, C., Hailfinger, S., Bonin, M., Buchmann, A., Schwarz, M.: Differential Gene Expression in Periportal and Perivenous Mouse Hepatocytes. *FEBS Journal* 273, 5051-5061, 2006.
- [3] Häussinger, D.: Hepatocyte Heterogeneity in Glutamine and Ammonia Metabolism and the Role of an Intercellular Glutamine Cycle during Ureogenesis in Perfused Rat Liver. *Eur. J. Biochem.* 133, 269-275, 1983.
- [4] Sies, H., Häussinger, D.: Hepatic Glutamine and Ammonia Metabolism. In: *Glutamine Metabolism in Mammalian Tissue*. Häussinger, Sies (Eds.) Springer-Verlag Berlin Heidelberg, 1984.
- [5] Schimassek, H., Gerok, W.: Control of the Levels of Free Amino Acids in Plasma by the Liver. *Biochemische Zeitschrift.* 343, 407-415, 1965.

## Authors:

Dr Susanne Toepfer<sup>1</sup>, Dr Sebastian Zellmer<sup>2</sup>, Dominik Driesch<sup>1</sup>, Dirk Woetzel<sup>1</sup>,  
Dr Reinhard Guthke<sup>3</sup>, Prof Rolf Gebhardt<sup>2</sup>, Dr Michael Pfaff<sup>1</sup>

<sup>1</sup> BioControl Jena GmbH  
Wildenbruchstraße 15, D-07745 Jena, Germany  
Phone: +49 - 3641 - 527831, Fax: +49 - 3641 - 527832  
E-mail: susanne.toepfer@biocontrol-jena.com

<sup>2</sup> University of Leipzig, Medical Faculty  
Johannisallee 30, D-04103 Leipzig, Germany

<sup>3</sup> Leibniz Institute of Natural Product Research and Infection Biology.V. - Hans Knoell Institute  
Beutenbergstr. 11a, D-07745 Jena, Germany

J.C. Ferreira / A.A. Fernandes / A.D. Santos

## **Modelling and Rapid Prototyping an Innovative Ankle-Foot Orthosis to Correct Children Gait Pathology**

### **ABSTRACT**

Nowadays, the conventional Ankle Foot Orthosis (AFO) models existents in the market cannot be use by about 30% of the patients, and have some limitations to all the users, because should be custom-made. Knowing these problems the work developed has a main goal to design a new type of AFO, adaptable as far as possible to all the needs of the children's patients to have a normal life. This project is a partnership created between Instituto Superior Técnico, Departamento de Engenharia Mecânica and the children's Hospital D. Estefânia having as objective to develop and model an Ankle Foot Orthosis applied to children with gait pathology considering neurological problems such as drop foot stroke, spinal cord injuries (example: Spina Bífida), cerebral palsy, multiple sclerosis and trauma. The light weight AFO developed has in consideration all the patient needs, and a new concept idea of orthosis, attended to achieve the standardization of all possible components, never forgetting that always exist some personalised components that are tailored for specific persons. The design and concept development of a lightweight orthosis for the human lower limb that could comfortably provide plantar flexion and dorsiflexion energy during walking was accomplished. With all the improvements that are achieved is possible to have a larger target of people with pathological gait using these innovative AFO's products, which could allows increasing their comfort.

### **INTRODUCTION**

There are millions of individuals with gait disabilities, requiring either rehabilitation or permanent assistance. Passive dynamic ankle-foot orthosis (AFOs) are often prescribed to improve gait performance for those with various neuromuscular disorders. Patients with various neuromuscular disorders often lose control and strength in their lower-limbs, which impairs their mobility. Drop foot pathology consists of the inability to

properly lift the foot due to weakness or non-functioning dorsiflexor muscles. The current assistive technology for drop foot is a mechanical brace called an Ankle Foot Orthosis (AFO), which has been gaining increased usage over the last few years [1]. Recent biomechanical analyses of AFO's have shown an increase in walking speed and overall gait performance for patients with post-polio syndrome [2, 3]. Designs and materials used for AFO's range from simple polypropylene braces to advanced custom carbon composite fiber designs. AFO's design varies in the shape and length of the foot component as well as the stiffness and length of the tibial component depending on the desired functional outcomes. Rapid prototyping is an ideal technology for AFO development and manufacture because of its inherent customizing capabilities, which make fabrication of subject-specific AFO's feasible. In addition, the rapid prototyping has already been successfully used in the fabrication of prosthetic sockets for lower limb amputees [4]. The storage and release of elastic energy within the structure of passive dynamic AFO's is an important design characteristic that helps compensate for the neuromuscular disorders. During walking, elastic parts of the AFO deforms elastically under the influence of the body weight, thus storing elastic energy. During the second half of the stance when the AFO is unloaded, the AFO releases the stored energy to help satisfy the energetic demands of walking [5]. Although AFO's offer some biomechanical benefits, there are many disadvantages that can be improved on [6].

The main contribution of this research work is to apply ankle biomechanics, engineering principles and current rapid prototyping technology to develop an improved adaptive and autonomous system to assist drop foot gait.

### **MODELLING THE ANKLE-FOOT ORTHOSIS**

The design model of the ankle-foot orthosis (AFO) consists of ten parts, including a foot section made in thin-polymer material. A light leg section articulately attached to the foot section on each side of the ankle by two rotational mechanical axes in order to facilitate dorsiflexion. The leg section has a two flanged rim-frames linked by two carbon composite bars overlapping the upper edge of the foot section. The overlap provides a stop for the backward movement of the leg section around the articulation (flexible joint) axe, thus limiting plantar flexion and drop-foot. Two elastic material parts store and release the elastic energy within the structure of the passive dynamic AFO. One metallic rear bar limits the foot rotational angular movement when walking. The AFO's Design Model and the constitutive parts are shown in figure 1.



Figure 1: AFO's Design Model and constitutive parts.

### FEM ANALYSIS OF THE ANKLE-FOOT ORTHOSIS

The stress and strain of the main components of the AFO was investigated by Finite Element Method analysis. The critical situation occurs for an excessive dorsiflexion during the initial contact that needs a functional limitation. This limit function is of importance to prevent knee flexion with a incorrect posture and with inclination of the leg, which promotes an extraordinary muscular effort and consequently a great fatigue with exhaustion.

The stress and strain FEM analysis of the AFO's light leg section articulately attached to the foot section is shown in figure 2.

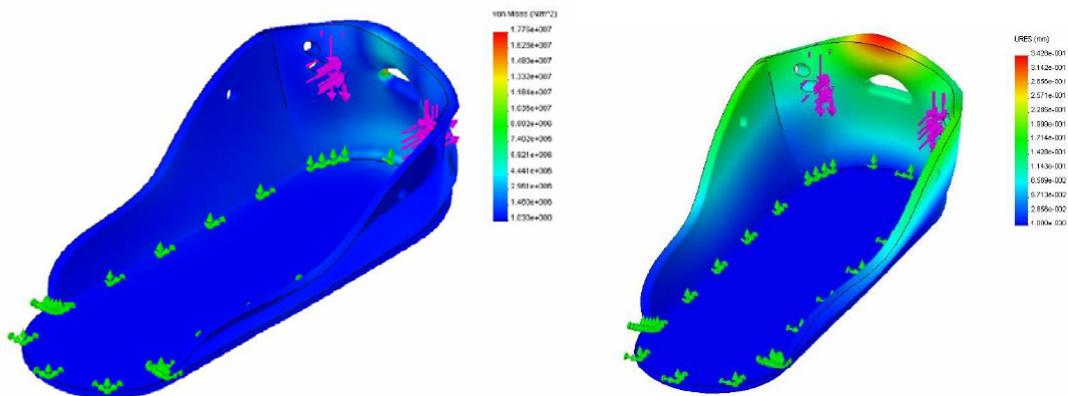


Figure 2: Von Mises stress analysis and strain analysis of AFO's foot component.

The FEM analysis allows identifying a critical Von-Mises stress located near the open hole utilised to anchorage the blocking belt with an average value of 17,8

MPa. The maximum part deformation is also located at the rear of the foot part in the vicinity of the hole utilized to fix the blocking belt.

The stress and strain FEM analysis of the AFO's light leg section articulately attached to the foot section is shown in figure 3.

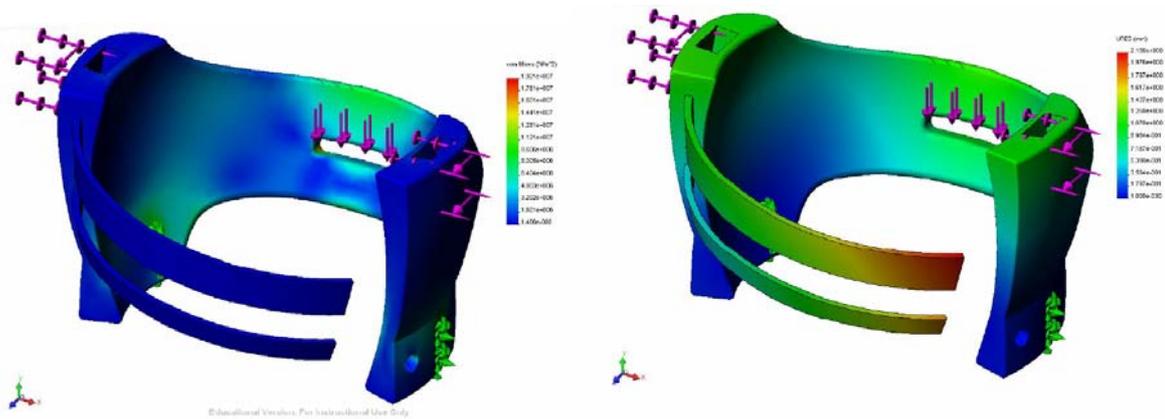


Figure 3: Von Mises stress analysis and strain analysis of AFO's leg section articulately.

Analysing these figures for an applied load in the order of 250N and a tension strength factor owing the actuating angles at the oval extremities could be verified that the critical zones, with higher tension values attain 19MPa that is inferior to the tensile strength – 21MPa, and are located in the anchorage zone of the blockage band. In the constrains neighbourhood there are also tension values with a magnitude in the order of 11 MPa since in this zones the part is fixed, i.e. where the load reactions are located. The maximum strains attain 1, 25 mm where the blockage band is fixed, i.e. where is applied a load in the order of 250N. For the other part zones the deformations are insignificant. The stress and strain FEM analysis of the AFO's upper leg section is shown in figure 4.

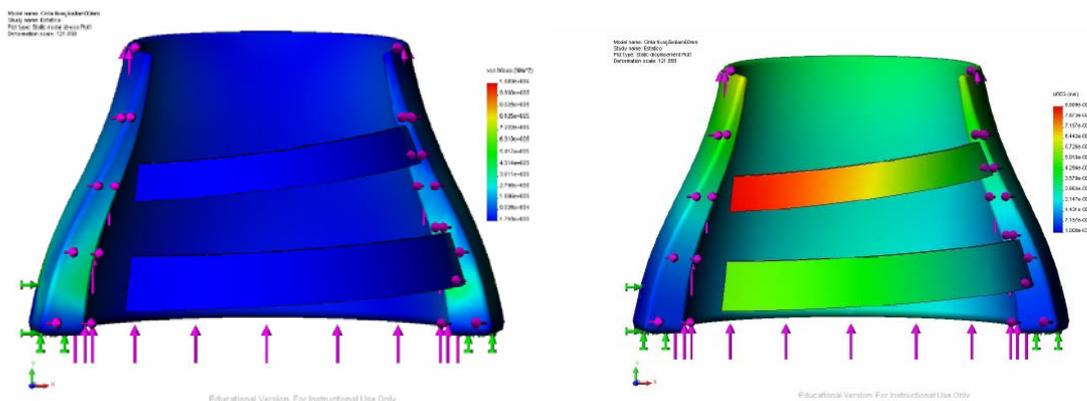


Figure 4: Von Mises stress analysis and strain analysis of AFO's leg section.

The maximum Von-Mises stress values registered are in the order of 1MPa and are located in the constrict part zone. In spite off this stress value is considerably inferior to the material tensile strength. The maximum values of the Von-Mises strain analysis indicates that the deformation values are reduced in the order of 0,05mm at the top zone of the fixation belt.

### **AFO PROTOTYPES MADE BY RP TECHNOLOGY**

The AFO's prototypes were made by Rapid Prototyping (RP) technology from the design and taking into account the improvements from FEM analysis. The several stereolithographic models were built in a Spectrum Z510 Color 3D Printers (Z Corporation-Burlington, USA), with zp130 powder and zb58 binder. The different stereolithographic models and the complete AFO's prototype are shown in figures 5a and 5b.



Figure 5 – Rapid Prototyping: (a) Prototypes of the AFO and (b) complete new ankle-foot orthosis prototype.

## CONCLUSIONS

The light weight AFO developed has in consideration all the patient needs, and a new concept idea of orthosis, attended to achieve the standardization of all possible components, never forgetting that always exist some personalized components that are tailored for specific persons.

The design and concept development of a lightweight orthosis for the human lower limb that could comfortably provide plantar flexion and dorsiflexion energy during walking was accomplished.

With all the improvements that are achieved is possible to have a larger target of people with pathological gait using these innovative AFO's products, which could allows increasing their comfort.

### Acknowledgment:

The authors would like to thanks to Fundação para a Ciência e Tecnologia the support of projects POCI/2010 and POCI/EME/60650/2004 "Prototipagem Rápida pelo Processo de Impressão 3D para o Desenvolvimento de Ferramentas e Próteses Médicas".

### References:

- [1] Brown, S. J., 1995, "Marketing Trends in Managed Care". Baylor Institute for Rehabilitation, Dallas, TX,
- [2] Halstead, LS, Crittenden, DT, Nielsen, J-P, Loke, M., 2003, "Dynamic bracing: a novel approach to lower extremity bracing for individuals with incomplete spinal cord injury and other neuromuscular disorders," J Spinal Cord Med 26 (1): 16.
- [3] Nelson KM, Kepple TM, Siegel KL, Halstead LS, Stanhope SJ, 2003, "Ankle foot orthosis contribution to net ankle moments in gait," Proceedings of American Society of Biomechanics.
- [4] Faustini, M., Crawford, R., Neptune, R.R., Rogers, W.E. and Bosker, G., 2005, "Design and analysis of orthogonally compliant features for local contact pressure relief in transtibial prostheses," J of Biomechanical Engineering 127(6): 946-951.
- [5] Faustini, MC, Neptune, RR, Crawford, RH, Stanhope, SJ, 2006, "Selective Laser Sintering of Passive Dynamic Ankle-Foot Orthoses". Proceedings of Solid Free-Form Symposium, pp124-129.
- [6] Carlson, W E, Vaughar, C L, Damiano, D L, and Abel, M F, 1997, " Orthotic management of gait in spastic diplegia." American Journal of Physical Medicine & Rehabilitation, 76, pp219-225.
- [7] Winter, David A., 1991, "The Biomechanics and Motor Control of Human Gait: Normal, Elderly and Pathological". University of Waterloo Press book, Second Edition, Ontario, Canada.

### Author:

Prof. José Carvalho Ferreira  
Eng. A.A. Fernandes and Eng. A.D. Santos  
IDMEC, 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](mailto:joseferreira@ist.utl.pt)

H. T. Shandiz, E. Zahedi

## Noninvasive Method in Diabetic Detection by Analyzing PPG Signals

*Abstract:* As the report says, every 10 second 2 persons become diabetic patients. It is a medical condition caused by a lack of insulin, which makes the patient produce a lot of urine and feel very thirsty. If it is not recognized on time, blindness is one of the results of this disease. Recognizing diabetics by invasive and simple method can help the national health center to alert people. By filtering and using Kurtosis function on PPG signals a method for recognizing diabetic is proposed.

### 1. Introduction

Noticeable numbers of people are dying worldwide because of diabetic disease each year. The cause of this disease is due to lack of insulin or its effectiveness. The result of this disease is high level of glucose in the blood. Short and long-term complications are due to this change in glucose level. As the amount of blood volume in the figure changes, this variation can be measured by photoplethysmography (PPG). When a fixed source of infrared radiation is used, the variation of blood volume act as a phototransistor and the receive signal is changed. This is why we use the PPG signal for recognizing the diabetic.

In the following we review, the researchers try for using PPG in monitoring health conditions of people, as an invasive method.

A. Emilio et al, use PPG signal to measure intravascular volume change and indicate the blood pressure. They formed two groups according to age. Each group had three healthy volunteers: three males who are younger age group (24-26 years) and two males and one female for older age (58-68 years) group. They found statistically significant difference in intravascular finger compliance between the younger and older people over a range of  $P_{mean}$  from 40-60 mmHg [1].

M. Joseph et al, explored the feasibility of employing PPG and pulse oximetry to assess the status of the blood circulation in the dental pulp. Radiographic and subjective test for determination the validity of a tooth is painful and invasive [2].

A. Johansson et al, made a new respiratory device based on PPG signals. They found the respiratory signals from PPG device were compared to impedance plethysmographic signals [3].

J. Allen and A. Murray, have used signal processing algorithms to calculate beat to beat measures of pulsatility, and determined normative data for multi-site PPG pulse assessments. They collected waveforms from right and left ear lobes, thumb pads and toe pads. They measured median coefficients of variation for pulse amplitude, difference in pulse transition time [4].

V. S. Murthy et al, presented the result of the spectral analysis of PPG signals for normal and patients with various cardiovascular disorders. They show the PPG can be used to analyze cardiovascular disorder and supplement existing methods of analyzing heart rate variability [5].

K. w. Chan and Y. T. Zhang, used a LMS filter with automatic step size control to mitigate the effects of motion artifact in PPG recording for long term patient monitoring. They proof their LMS filter with variable step size has better performance than fixed step size [6].

As Kurtosis function is used in this work, briefly the normal distribution and deviation of it is

explained.

A random variable is a function in which outcome of a probabilistic experiment is its domain and real number is the range. If density function of a random variable is as follow:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

In which  $\sigma$  is standard deviation and  $\mu$  is mean of data, then it is normal [7].

The kurtosis of a distribution is defined as follow:

$$k = \frac{E(x-\mu)^4}{\sigma^4} \quad (2)$$

It is a measure of how a distribution is deviated from normal. The kurtosis of the normal distribution is 3. Distributions that are more outlier-prone than the normal distribution have kurtosis greater than 3; distributions that are less outlier-prone have kurtosis less than 3 [8].

By putting the PPG signal in array  $X$  and using the following MATLAB command

$$k = \text{kurtosis}(X) \quad (3)$$

The deviation from normal in signal is measured.

## 2. Data acquisition

There are two array groups of PPG signals. Pathologic arrays contain data from all subject tested in the hospital who were diabetic (39-64 years ages). Healthy arrays contain all data from healthy subjects (22-52 years old). Figure 1 shows a recorded PPG signal for a diabetic patient.

In each file, the only variable is a (50x24750) which is the raw PPG data:

- Each row is one particular lead.
- First column: subject number: Sb
- Second column: lead number: Ld
- Third column: Age
- Fourth column up to end: raw PPG data
- The length of the files has been limited to 90 sec (sampled at 275 Hz, this gives 24750 sample points)
- The number of rows (records) has been limited to 50 per file to limit file size
- The format of the data is uint32 to save on space.

## 3. Signal processing and result

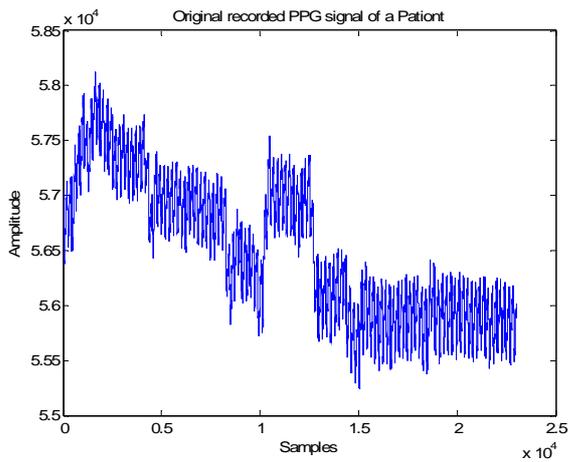
Some parts of the PPG signal may be affected by motion artifact noise, which is due to the movement of the probe. In order to avoid these episodes, it is necessary to inspect the signal visually and select that part of the signal that is "stable". Our experience shows a length of about 1000 samples seems sufficient to extract the value of the Kurtosis or negentropy.

Then the dc level of signals are removed and data are normalizing between  $\pm 12$ . As the PPG signal that is recorded may contain noise and baseline drift, it is good to preprocess it using a band pass filter. It has to be FIR type to conserve the shape. As our experiences show, the recommended values for the BPF are 0.6 Hz to 15 Hz. The original sampling frequency of the data was 275 Hz. Figure 3 shows the frequency characteristic of the filter. The result of applying this filter to data in figure 2 is representing in figure 4. Figure 5 shows the result for healthy and patient data.

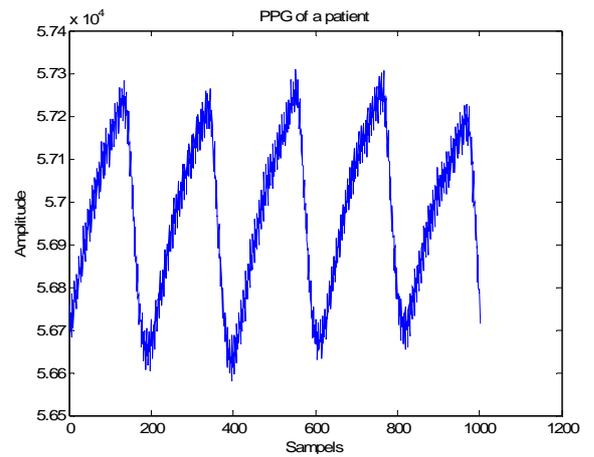
## 4. Conclusion

The result of applying Kurtosis function on filtered selected real data shows the Kurtosis for patient is around 20.6 while for healthy is 20 and under. For healthy people the accuracy of the

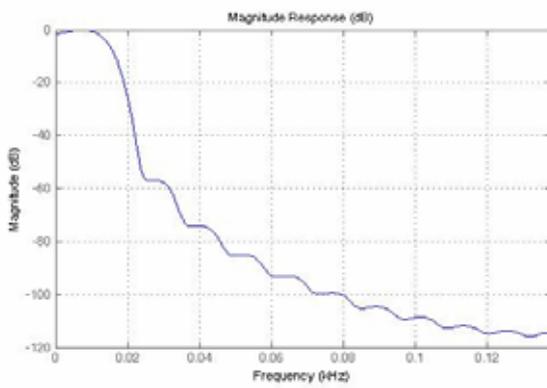
method is about %85 and for patient is about %76.



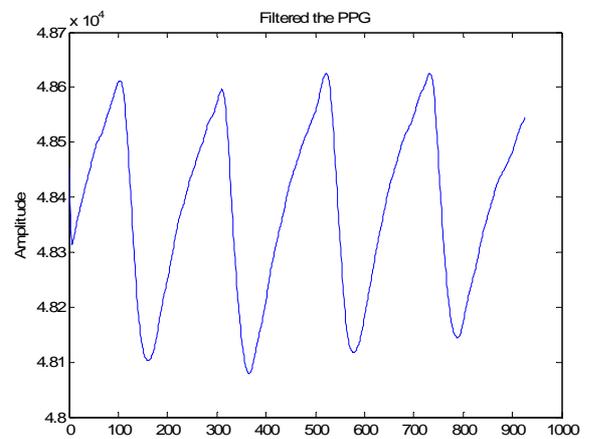
**Figure 1.** A patient recorded PPG signal



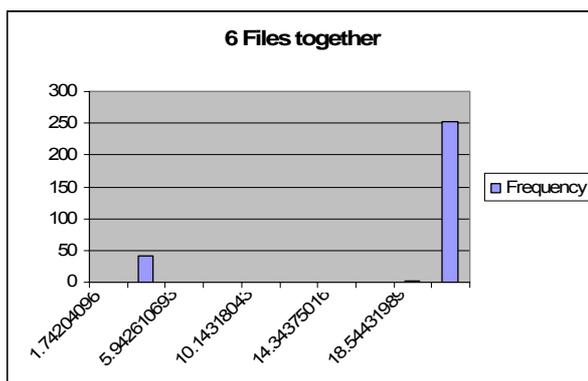
**Figure 2.** One stable part with 1000 samples of figure 1



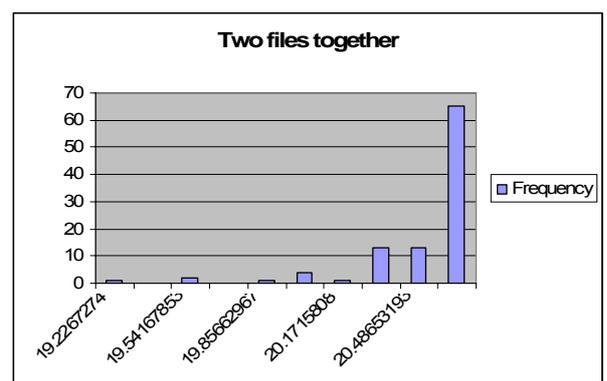
**Figure 3.** Frequency domain characteristics of BP filter



**Figure 4.** Filtered data in Figure 2



(a)



(b)

**Figure 5.** Kurtosis function: a) Healthy and b) Diabetics people

**References:**

- [1] A. Emilio, M. S. Lopez-Beltran, L. Perry and M. Stanley, A Noninvasive Method for Measuring Regional Peripheral Compliance, *IEEE 11<sup>th</sup> Annual International Conference in Medical And Biology*, 1989.
- [2] M. J. Schmitt, R. L. Webber and E. C. Walker, Optical Determination of Dental Pulp Vitality, *IEEE Transactions on Biomedical Engineering*, Vol. 38, No. 4, April 1991.
- [3] A. Johansson, L. Nilsson S. Kalman and P. A. Oberg, Respiratory Monitoring using Photoplethysmography, *Proceedings of the 20<sup>th</sup> Annual International Conference of the IEEE Engineering in Medical and Biology*, Vol. 20, No 6, 1998.
- [4] J. Allen and A. Murray, Variability of Photoplethysmography Peripheral Pulse Measurements at the Ears, Thumbs and Toes, *IEE Proceedings on Science, Measurements and Technology*, Vol. 147, No. 6, November 2000.
- [5] V. S. Murthy et al, Analysis of PPG Signals of Cardiovascular Patients, *IEEE Proceedings of the 23<sup>rd</sup> Annual EMBS International Conference*, October 25-28, Istanbul Turkey, 2001.
- [6] K. w. Chan and Y. T. Zhang, Adaptive Reduction of Motion Artifact from PPG Recording Using a Variable Step Size LMS Filter, *IEEE*, 2002
- [7] A. Papoulis and P. S. Unnikrishna, *Probability, Random Variable and Stochastic Process*, 4<sup>th</sup> edition, McGraw Hill, 2002.
- [8] MATLAB User Guide

**Authors:**

Dr HEYDAR TOOSSIAN SHANDIZ

Dr EDMOND ZAHEDI

Shahrood University of Technology, Electrical Engineering

Faculty

7 Th Tir Square, P.o.Box 36155-316, Shahrood, IRAN

36155-316, Shahrood

Phone: 0098-9121733733

Fax: 0098-273-3334419

E-mail: htshandiz@shahroodut.ac.ir or

h\_t\_shandiz@hotmail.com

S. V. Drobot, I. S. Asayenok, E. N. Zacepin, T. F. Sergiyenko, A. I. Svirnovskiy

## **Effects of Mm-Wave Electromagnetic Radiation on Sensitivity of Human Lymphocytes to Ionizing Radiation and Chemical Agents *in Vitro***

### **TOOLS FOR DIAGNOSIS AND THERAPY**

#### **INTRODUCTION**

Research into effects of mm-wave electromagnetic radiation (EMR) on human health has attracted considerable interest in recent years. Mm-wave EMR is a component of electromagnetic pollution, on the one hand, and one of key mechanisms in adjusting vital functions [1, 2], on the other. Apoptosis of human peripheral blood lymphocytes represents a suitable test system for assessing harmful environmental impact on human body [3] and may be used for biological dosimetry of ionizing radiation (IR) [4]. In view of this, metabolic activity of peripheral blood lymphocytes is a convenient indicator of mm-wave EMR effects on human body. The impact of low-intensity mm-wave EMR on blood corpuscles and peripheral blood plasma has been researched recently [5]. However, no research is yet available of the EMR effects on lymphocytes *in vitro*, although, as noted above, lymphocyte cellular preparations may present a suitable test system to assess harmful environmental impact on vital human functions.

In this paper a technique and results of research into effects of mm-wave EMR on vitality of human lymphocytes *in vitro* and an possibility of EMR use for prophylaxis or regeneration of the damages caused by ionizing radiation and chemical agents are presented.

#### **RESEARCH AND RESULTS**

The research was conducted on donor lymphocytes separated from heparinized peripheral blood by centrifugation. The selected lymphocytes were subsequently placed in culture medium and subjected to the 38.0-55.0GHz EMR. For this purpose sterile capsules, which have been made of radiotransparent polyethylene, were filled with the prepared cells suspension and airproofed. Then the capsules were located in a pyramidal horn aperture in the sizes  $10 \times 20 \text{ mm}^2$ , which corresponds to the sizes of the

capsules. The EMR exposure varied between 5 and 45 minutes. The EMR power flow density employed in the research amounted to 1-2 mW/cm<sup>2</sup>, which corresponds to a low-intensity or 'informative' EMR impact. The cellular suspension was subjected to ionizing radiation in total doses of 0.17, 2.0 and 5.0 Gy. Effects of the EMR and ionizing radiation were investigated using different sequences: X-rays followed by EMR or vice versa. Chemical agents such as Dexamethazonum and Cladribine were added in the culture medium to research into a toxic action on lymphocytes of the chemical agents with ionizing radiation also. The lymphocytes, following their irradiation by EMR and X-rays, were incubated for 44 hours in the same culture medium. Irradiation results were assessed using a biochemical technique of MTT testing to evaluate variations in cellular metabolic activity. The level of metabolic activity in cells not subjected to EMR was taken as reference at 100 %. Figures 1 and 2 show the lymphocytes metabolic activity dependences on the EMR exposure.

The mm-wave low-intensity EMR effect results in increase of peripheral blood normal lymphocytes metabolic activity (the solid line at fig. 1 and 2). The expressed stimulation is observed only at short-term five-minute exposure. Statistically proved distinctions in comparison with not irradiated cells are not revealed at longer EMR exposure.

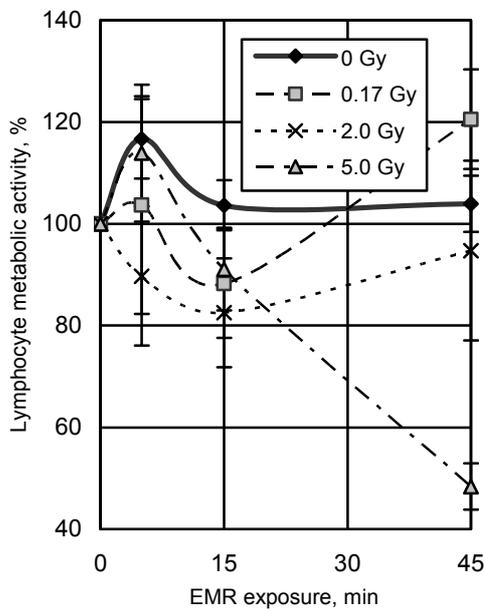


Fig. 1. Metabolic activity of lymphocytes previously subjected to various doses of ionizing radiation dependences on the EMR exposure

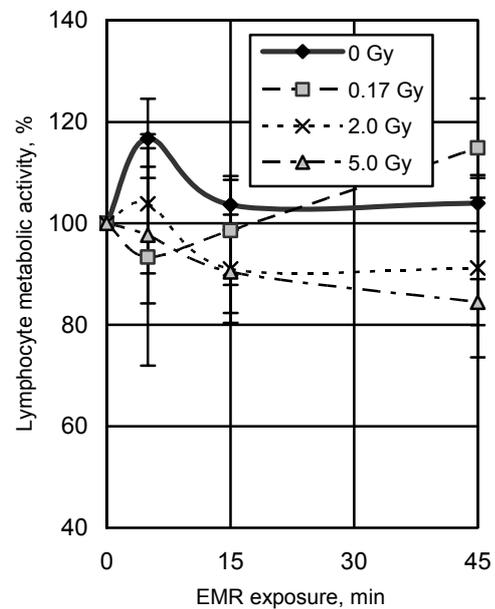


Fig. 2. Metabolic activity of lymphocytes subsequently subjected to various doses of ionizing radiation dependences on the EMR exposure

The effect of the low-intensity EMR on the lymphocytes in a combination with IR is selective. It is observed both protective and regenerative impact, and the effect of EMR is determined by the doses and the order of application of the external factors.

The "therapeutic" effect of EMR within 5-15 minutes on the lymphocytes following their irradiation by X-rays *in vitro* does not result in a significant change of its metabolic activity (fig. 1). The increase of the EMR exposure up to 45 minutes results in the strongly pronounced change of the lymphocytes viability. The effect of EMR in the combination with IR by a small dose (0.17 Gy) is stimulating, with IR by a dose 2.0 Gy – regenerating, with IR by a dose 5.0 Gy – oppressing.

The research of the "therapeutic" effect of EMR on the cells damaged by X-rays has confirmed a "cumulative" character of its action. The significant effect of EMR by a small exposure (5-15 minutes) on the lymphocytes, irradiated by IR, it is not revealed (fig. 1). The expressed cellular reactions are observed when the EMR exposure increase up to 45 minutes.

In a case of the "preventive" effect of EMR there is no a statistically significant decrease in the lymphocytes metabolic activity (fig. 2). The preliminary effect of EMR within 5-45 minutes promotes practically full recovery of the peripheral blood lymphocytes viability after irradiative on them IR in doses 0.17-5.0 Gy.

The effect of EMR prevents a damage of the lymphocytes by X-rays. Probably, the preliminary irradiation on the lymphocytes *in vitro* by EMR results in an increase of biological processes intensity in the lymphocytes that prepares the cells to stress – reactions before the subsequent X-ray irradiation.

The low intensity mm-wave EMR effect at the wide frequency band can prepare an organism for the subsequent adverse impacts, for example, to effect of IR and to make an organism less susceptible to it. The adaptation to the subsequent adverse impact can be compared to the effect which is being given by training, but is being carried out during short time. The informative EMR impact provides a correction at a level of physiological functions.

The research of the combined effect of the low-intensity EMR, IR and the chemical agents on the normal lymphocytes has revealed a various character of their viability changes (fig. 3 and 4). A effect of Dexamethazonum and Cladribine in the concentration close to therapeutic, results in a suppression of the peripheral blood lymphocytes viability up to the level  $90,98 \pm 5,6$  % and  $56,5 \pm 2,4$  % in comparison with the reference values, accordingly.

The irradiative on the blood cells by EMR within 15 minutes in a combination both with Dexamethazonum, and with Cladribine raises its metabolic activity up to 20 % and more in comparison with the reference level, and practically does not effect with other intervals of exposure.

The effect of EMR within 45 minutes, an irradiation of IR by a small doze (0.17 Gy) in a combination with the both chemical agents restores the lymphocytes viability.

The irradiation of IR in doze 2.0 Gy in a combination with Dexamethazonum and EMR within 15 minutes reduces the lymphocytes metabolic activity. The increase of the EMR exposure time up to 45 minutes practically restores the cells viability up to the initial size. In part restoring effect of EMR on the cells, which are exposed by IR in a doze 2.0 Gy and Cladribine is observed. However, the initial level of the cells metabolic activity is not achieved by the increase of EMR exposure time up to 45 minutes.

The low-intensive EMR increases the damaging effect of IR in doze 5.0 Gy in a combination with the chemical agents. The increase of the EMR time exposure up to 45 minutes results in a reduction of the cells metabolic activity in the case.

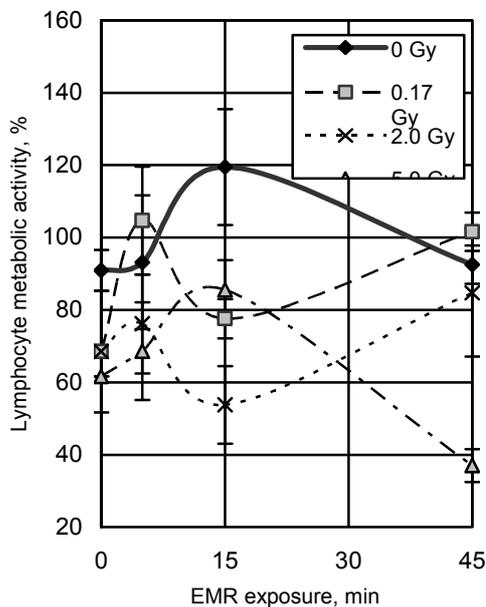


Fig. 3. Metabolic activity of lymphocytes previously subjected to various doses of ionizing radiation in combination with Dexamethazonum dependences on the EMR exposure

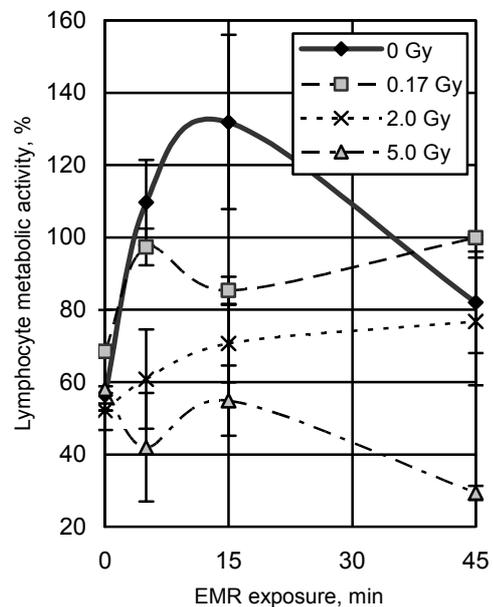


Fig. 4. Metabolic activity of lymphocytes previously subjected to various doses of ionizing radiation in combination with Cladribine dependences on the EMR exposure

## CONCLUSION

The low-intensity mm-wave EMR produces stimulating effect on peripheral donor blood lymphocytes *in vitro*. Damages of lymphocytes caused by small doses of ionizing radiation and the chemical agents are compensated by long EMR exposures. Preventive exposure of lymphocytes by EMR provides protective effect against ionizing radiation. The fixed effects can be used both for protective, and for the combined therapy of various nature diseases.

The low intensity mm-wave EMR is the physical factor which changes the viability of the healthy persons peripheral blood lymphocytes are damaged by the impact of IR and chemical agents *in vitro*. The reaction of the lymphocytes to the effect of EMR in a combination with IR and the chemical agents is defined by duration of its exposure, the sequence of the factors impact and a doze of IR.

### References:

- [1] Frohlich H. The biological effects of microwaves and related questions // Advances in Electronics and Electron Physics, Vol.53, 1980, pp.85-152.
- [2] Deviatkov N. D., Golant M. B., Betskiy O. V. Mm-waves and their role in vital processes. Moscow, Radio and Communication, 1991. (in Russian).
- [3] Svirnovskiy A. I., Sheleg S.V., Govela T.E. et al. Spontaneous and induced by chemical agents apoptosis in human lymphocytes. Ecological anthropology. Minsk, 1999, pp. 216–220. (in Russian).
- [4] Menz R., Andres R., Larsson B. et al. Biological dosimetry: the potential use of radiation-induced apoptosis in human T-lymphocytes. Rad. and Environ. Biophysics, Vol. 36, pp. 175–181, Mar. 1997.
- [5] Walleczek J. Electromagnetic field effects on the cells of the immune system: the role of calcium signaling. FASEB J., Vol. 6, pp. 3177-3185, 1992.

### Authors:

Dr.-Ing. S. V. Drobot  
Dr. med. I. S. Asayenok  
Dr.-Ing. E. N. Zacepin  
Belarussian State University Informatics and Radioelectronics, 6 P.Brovki Street  
220013, Minsk, Belarus  
Phone: +375 17 2938945  
Fax: +375 17 2021033  
E-mail: hanms@bsuir.unibel.by

Dr. med. A. I. Svirnovskiy  
PhD student T. F. Sergiyenko  
Republican Scientific and Practical Center of Hematology and Transfusiology,  
160 Dolginovskiy Trakt, Minsk, 220053, Belarus  
Phone: +375 17 2898684  
E-mail: asvirnov@yandex.ru







B. Däne

## Modeling and Realization of DMA Based Serial Communication for a Multi Processor System

### 1. INTRODUCTION

Performance of communication is very important for multi processor systems. This paper will present a communication solution for a multi DSP (Digital Signal Processor) system using DSPs from *TMS320C67x* family [1] of *Texas Instruments*. On-chip serial high-speed ports and DMA (Direct Memory Access) units are used. Primary goals are small latency, small protocol overhead even for short messages and minimized influence on filter calculations and block transfers that isochronously run on the system. The on-chip serial ports, designed for point-to-point communication only, should be enabled for multi-point shared-medium communication using small hardware extensions only.

The design process has been carried out in model based manner. Hierarchical models represented hardware and software functions of communication process. Behavior of on-chip hardware has been determined, modeled and validated using existing documentation [2] and experimental results. In order to get quantitative performance results from simulation several communication scenarios have been modeled. These results have later been compared to practical results from the system realized.

So the project is a case study in model based design. It investigates the use of design and evaluation methods with a nontrivial practical example [3]. Extending another paper [4] this paper additionally discusses modeling of the DMA units. Modeling tool has been MLDesigner [5,6] of MLDesign Technologies Inc.

### 2. HARDWARE SETUP

Hardware is based on the multi DSP system described in [7], using five DSPs from *TMS320C67x* family. It already used a communication interface based on parallel memory busses. The goal was to implement additional serial communication based on on-chip duplex synchronous serial interfaces. These interfaces, named McBSP (Multi Channel Buffered Serial Ports), perform at a gross data rate of approximately 83 Mbit/s [2,8]. Multi channel feature is available by time multiplex of up to 128 logical channels.

External hardware structure was designed as a simple serial bus that shares one impedance-controlled line for each node's data, clock and frame impulse signals, respectively. This non-documented regime is enabled by software controllable tri-state properties of on-chip signal sources involved here. Protocol support or middleware supporting this regime was not available.

Access to control and data registers of McBSP units has to be done by DMA rather than CPU cycles due to CPU core performance loss when interrupting loops and accessing slow memory or peripherals [1]. Coordination between CPU core and DMA includes usage of interrupts in order to organize concurrent timing behavior.

Motivation for the additional serial communication feature with respect to the parallel bus already available is an independent channel for short but urgent messages and alarms rather than additional bandwidth. Parallel bus is used for isochronous block transfers that should not be affected by such messages and in turn should not introduce additional latency into them.

### 3. MODELING CONCEPT AND GOALS

As mentioned above the goal of the case study is studying and evaluating model based design methods with an real example rather than the technical solution itself. Especially timing analysis and quantitative performance measurement should be performed with the model for later comparison to practical results. One typical challenge for such projects is including not only components under design (here external hardware structure and a number of software modules) but also relevant parts of the embedding system (here on-chip compo-

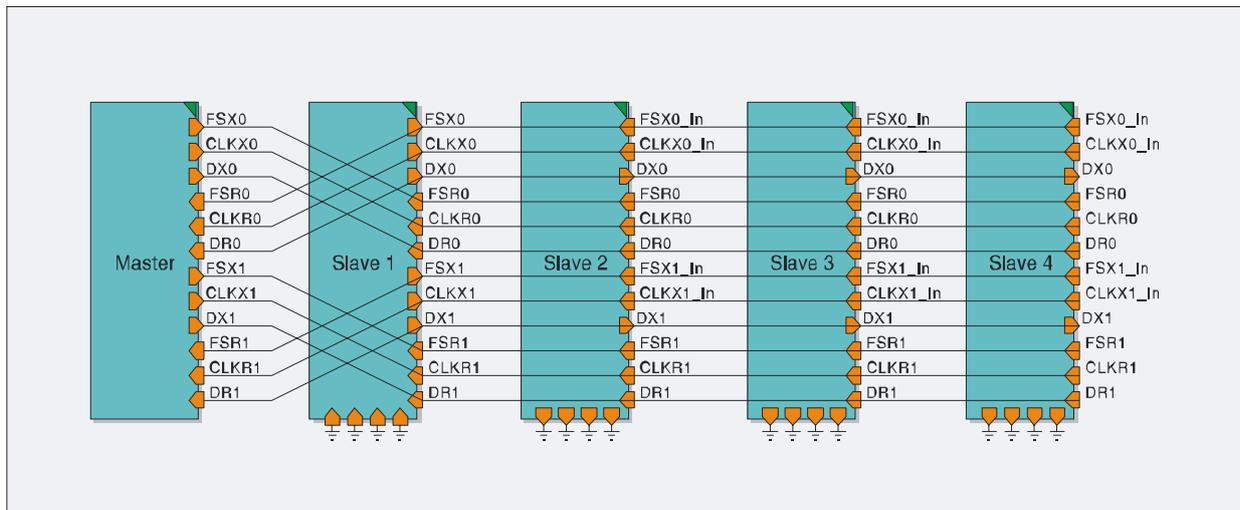


Fig. 1: Top level of model

nents and timing behavior of CPU core) into the model. Latter parts of the model are available from different sources using different representations and have to be integrated into one consistent model.

The model uses the DE (discrete event) domain [10] of the MLDesigner modeling tool mentioned above. The main paradigm of this domain is based on atomic messages between ports of blocks that may be structured hierarchically. This kind of domain also can be found in the open-source modeling tool Ptolemy Classic [11] from University of Berkeley that precedes the MLDesigner tool. Also included are blocks from the FSM (finite state machine) domain that has been substantially extended during development of MLDesigner [12].

Investigation of timing behavior and performance values in this case study mainly is based on simulation techniques. This requires equipping the model with communication scenarios and observation elements. Generating implementation components (especially software modules) directly from the model is not subject of this case study but will be considered in future work.

#### 4. MODEL STRUCTURE

The structure of the realized model will be shown partly using some example modules. At top level the model structure resembles the overall hardware structure of the target system with five DSP nodes (see fig. 1). One node acts as master, the other four as slaves. Events of DE domain at this level mainly represent edges of binary electrical signals that are included in the bus structures. The inside models of the DSP nodes also derive from their block structure, including two duplex McBSP units and two DMA channels each. Another block per DSP node encapsulates software functions of this node. Events at this level represent signal edges, port accesses, abstract trigger events and others.

The software functions are modeled using elements of the FSM domain. Included are sequences for port access and interrupt management, simple protocol implementations and communication scenarios. Cycle accurate modeling at instruction level has not been used. Software functions instead have been modeled at the level of abstract operations with constant time consumption including several processor instructions each.

A large part of modeling effort lies in the blocks that refine DMA and McBSP units. Due to lack of some kind of model or design documents from the manufacturer this information had to be extracted from manuals and application notes [2,9,13,14]. Some open questions have been solved by isolated experiments with evaluation hardware. Contributing to the goals of this case study it has been shown that such an approach can lead to behavioral models of sufficient accuracy.

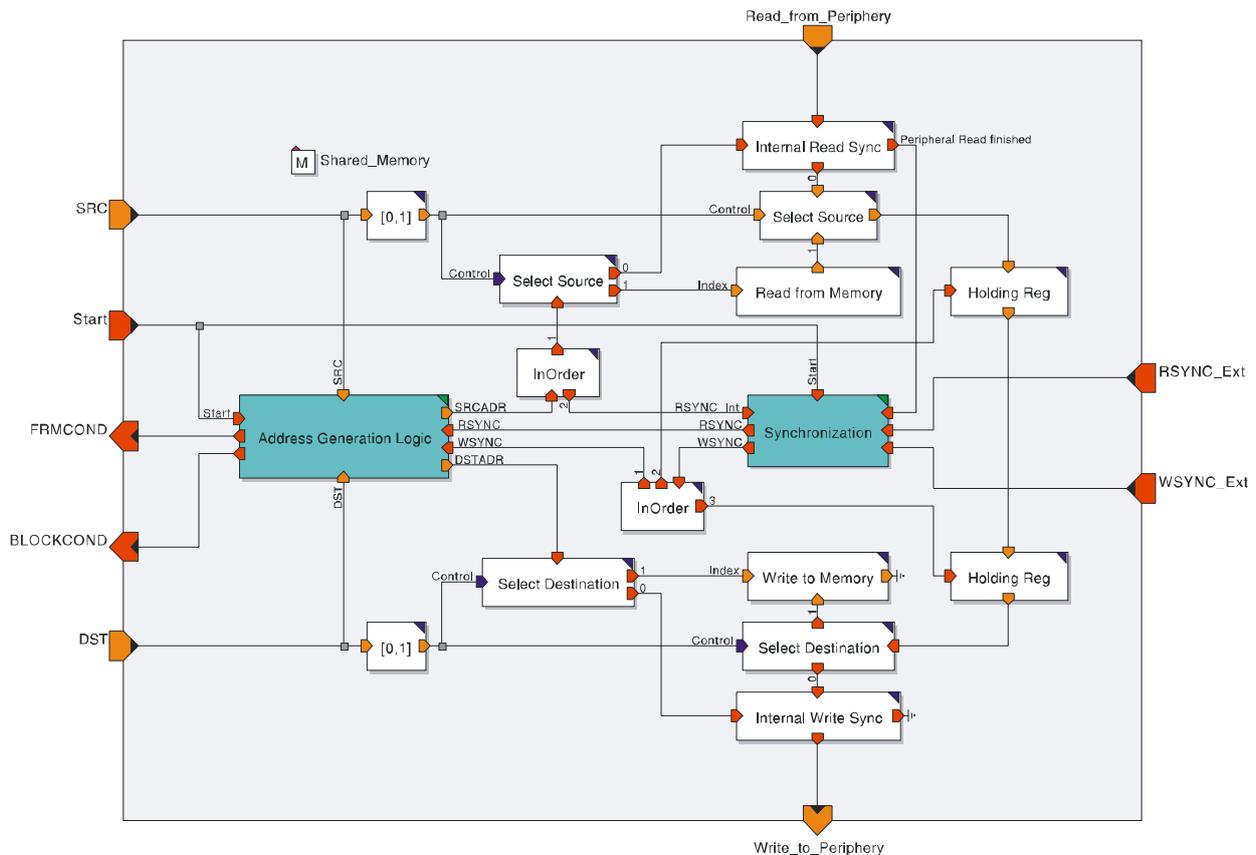


Fig. 2: Module for one DMA channel

For both DMA and McBSP units models with several hierarchical levels have been realized and tested. At bottom level all functions are built from basic elements (primitives) of DE domain. For example fig. 2 shows the inside of one DMA channel model. The signals at the border of the block correspond to logical conditions that start, synchronize and acknowledge transfers of the DMA channel. They have been derived from documentation and belong to interrupt conditions, control bits and others. Inside are blocks for organizing the operation, obtaining parameters such as addresses and block lengths and routing synchronization signals. The block "Address Generation Logic" (not shown) contains other modules for counting elements and address values with respect to element size and address counting mode.

As an example for a bottom level module fig. 3 shows the inside of the block "Synchronization". Visible are logical functions for handling read and write cycles. The upper part generates trigger events (output port at right) that initiate read cycles, the lower part does so for write cycles. At left there are control inputs that may be assigned to several trigger sources. The block among other functions handles the start of the block transmission and the alternating of read and write cycles. The shaded elements represent options controllable by software that in real hardware appear as control registers. In the model the modeling tool's parameter concept is used for this task.

## 5. RESULTS AND COMPARISON

The model and its functions with respect to expected behavior have been verified by simulation using prepared communication scenarios. Behavior has been observed by interactive model elements. Furthermore signal sequences of several external signals have been recorded for comparison. Later comparable communication scenarios have been investigated using the real target system. Sequences of same signals have

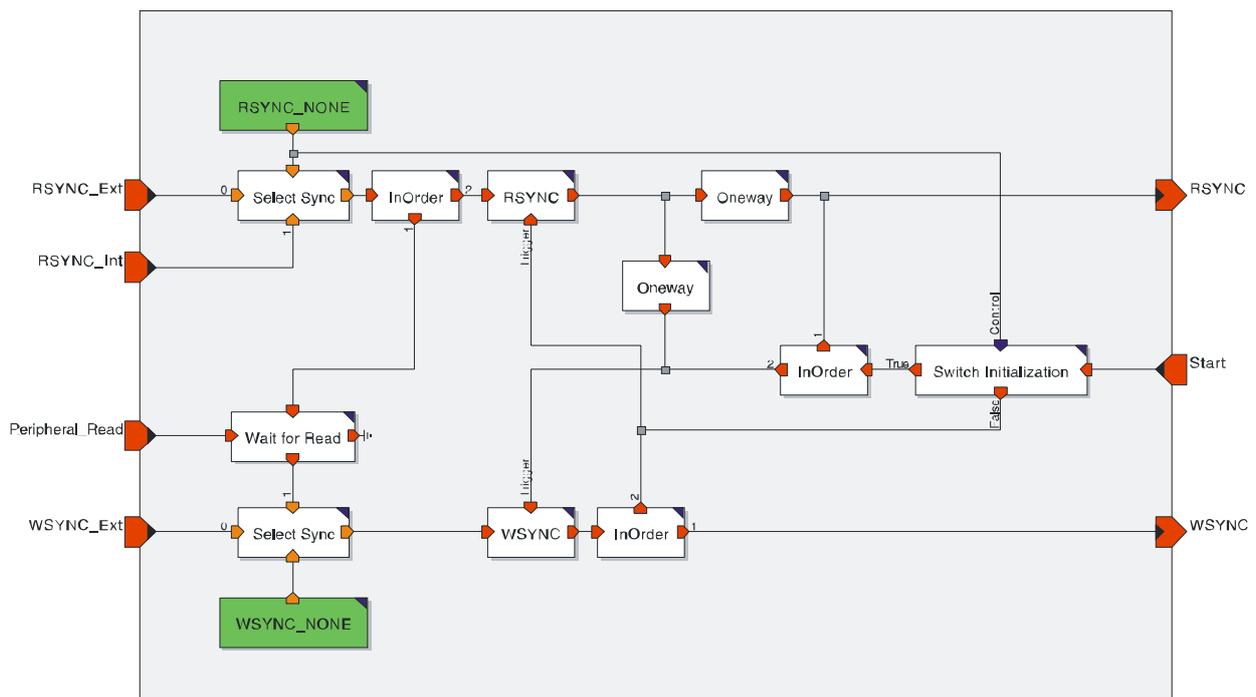


Fig. 3: Module “Synchronization” of one DMA channel

been recorded by logic analyzer. Comparison of both recordings showed good match. This fact serves as indication for the validity of the modeling approach.

Another significant part of the task was determining and comparing quantitative performance parameters. During simulation using the communication scenarios mentioned above net data rates and message latencies have been measured. While doing so several parameters such as block lengths and channel counts have been varied. Furthermore the tests have been carried out for two different members of the *TMS320C67x* DSP family, modifying the model according to their individual properties.

Measuring the same values with the real target system leads to good match with maximum divergence inside a five percent margin. Table 1 shows some values for both the C6701 device and the C6713 device. Note that different frequency division is used by these devices. Tests have been carried out for small and large block sizes and for single and multi channel modes. Data rates for multi channel modes are sum values for all channels. As expected small block sizes introduce a small overhead over large block sizes. Furthermore multi channel mode leads to some overhead over single channel mode.

However a systematical divergence towards too optimistic values is visible. This is caused by relatively coarse grain models of some software components that do not include time losses such as wait states or memory conflicts in internal data memory.

## 6. CONCLUSION

This case study demonstrates potentials and limits of this model based approach for planning and realizing a communication solution for an embedded multi processor system. The building of one consistent model using information that derived from different sources and that was available in different representation successfully has been carried out. Of-the-shelf-components for which no detailed design documents are available have been included in the model structure. Results from evaluating the model have been compared to experimental results from the real target system.

Further work will include refinement of the modeling process, leading to more precise prediction of quantitative values. Furthermore methods for generating actual hard- and software from parts of the model will be investigated and evaluated.

| Configuration |                       |                    |                    | Results                            |                                     |                     |
|---------------|-----------------------|--------------------|--------------------|------------------------------------|-------------------------------------|---------------------|
| Device        | Clock frequency [MHz] | Number of channels | Block size [Bytes] | Data rate from simulation [Mbit/s] | Data rate from measurement [Mbit/s] | Relative difference |
| TMS320 C6701  | 167                   | 1                  | 32                 | 51.6                               | 50.7                                | + 1.8%              |
|               |                       |                    | 65536              | 52.9                               | 52.8                                | + 0.2%              |
|               |                       | 4                  | 32                 | 50.8                               | 49.6                                | + 2.4%              |
|               |                       |                    | 65536              | 52.4                               | 51.6                                | + 1.5%              |
| TMS320 C6713  | 300                   | 1                  | 32                 | 68.7                               | 67.2                                | + 2.2%              |
|               |                       |                    | 65536              | 71.1                               | 71.2                                | - 0.1%              |
|               |                       | 4                  | 32                 | 68.0                               | 65.2                                | + 4.3%              |
|               |                       |                    | 65536              | 71.2                               | 70.4                                | + 1.1%              |

Table 1: Data rates from simulation and from measurement

## 7. ACKNOWLEDGEMENTS

This work has been supported by Deutsche Forschungsgemeinschaft (DFG) under grant SFB 622.

MLDesigner © 2005 MLDesign Technologies, Inc. All rights reserved. <http://www.mldesigner.com/>  
Fig. 1, 2 and 3 are taken from this tool.

### References:

- [1] N. Dahnoun: Digital Signal Processing Implementation Using the TMS320C6000 DSP Platform. Prentice Hall 2000.
- [2] C. Turner, R. Ma, B. Cobb: TMS320 C6000 and TMS320C5000 McBSP Features Summary. Application report SPRA638A. Texas Instruments Inc., Dallas, 2005.
- [3] S. Herrig: Konzeption, Modellierung und Realisierung eines Kommunikationssystems für eine Multi-DSP-Architektur (Concept, Modeling and Realization of a Communication System for a Multi-DSP-Architecture). In German. Diploma thesis, TU Ilmenau/IA, Ilmenau 2005.
- [4] B. Däne, W. Fengler: Modelling Hardware and Software for Fast Serial Interprocessor Communication. In: M. Adamski, L. Gomes, M. Węgrzyn, G. Łabiak (Eds.): DESDes'06 - Discrete Event System Design 2006. A Proceedings volume from the 3rd IFAC Workshop. Rydzyna, Poland, 26-28 September 2006. University of Zielona Góra Press, 2006, S. 155-160.
- [5] V. Zerbe, U. Freund, H. Salzwedel: Mission Level Design of Control Systems. In: SCI/ISAS'99 Multiconference on Systemics, Cybernetics, Informatics, vol. 7, pp. 237-243. Orlando 1999.
- [6] N. N.: MLDesigner Documentation (Draft), Version 2.5. MLDesign Technologies, Inc., Palo Alto, 2004.
- [7] B. Däne, F. Berger: A Multiprocessor DSP System for a High Throughput Control Application. In: EDERS-2004: The European DSP Education and Research Symposium, 16th November 2004, Birmingham, UK.
- [8] S. Anjanaiah, V. Soteriou: Using the TMS320C6000 McBSP as a High Speed Communication Port. Application report SPRA455A. Texas Instruments Inc., Dallas, 2001.
- [9] D. Bell, E. Biscondi: TMS320C620x/TMS3206701 DMA and CPU. Data Access Performance. Application report SPRA614A. Texas Instruments Inc., Dallas, 2000.
- [10] N. N.: Introduction to Discrete Event Modeling. MLDesign Technologies, Inc., Palo Alto, 2004.
- [11] J. Buck, S. Ha, E. A. Lee, D. G. Messerschmitt: Ptolemy. A Framework for Simulating and Prototyping Heterogeneous Systems. Int. Journal of Computer Simulation, 22(4)1994, S. 155-182.
- [12] H. Rath, H. Salzwedel: ANSI C Code Synthesis for MLDesigner Finite State Machines. In: H. Kern (Ed.): Synergies between information processing and automation: 49th International Scientific Colloquium at Ilmenau TU, 27.-30.9.2004. Shaker Verlag Aachen, 2004.

- [13] N. N.: TMS320C6000 DSP Multichannel Buffered Serial Port (McBSP) Reference Guide. User manual SPRU580D. Texas Instruments Inc., Dallas, 2004.
- [14] N. N.: TMS320C620x/C670x DSP Program and Data Memory Controller/Direct Memory Access (DMA) Controller Reference Guide. User manual SPRU577A. Texas Instruments Inc., Dallas, 2004.

**Author:**

Dr.-Ing. Bernd Däne  
TU Ilmenau  
P. O. Box 100565  
D-98693 Ilmenau  
Phone: +49-3677-69-1433  
Fax: +49-3677-69-1614  
E-mail: [bernd.daene@tu-ilmenau.de](mailto:bernd.daene@tu-ilmenau.de)  
Web: <http://www.tu-ilmenau.de/ra>

M. Müller / A. Pacholik / W. Fengler

## Tool Support for Formal System Verification

### Abstract

The design of complex real time systems requires verification within all stages of the development process. To ensure quality already in early design stages, formal verification has to be conducted on high abstraction levels and for potentially incomplete system definitions. In this paper we present an approach to formally verify the temporal behaviour of simulation models by introducing a tool chain based upon the system design environment MLDesigner and the verification tool UPPAAL.

### 1 Introduction

The design process of a complex embedded system aims to construct the right system in a correct way. To build the right system means to integrate the expected behaviour related to the users needs. This is ensured by simulation based validation [1]. For a system to be correct means, it is free of faults in a way, that specified properties can be ensured. This can be proved by formal verification [2].

Today simulation based validation is applied on system level, whereas verification is applied on implementation or block level. There are no techniques available to prove temporal properties in a high level validation model. A complete verification can only be conducted, if block level implementations are available. For example, SystemC is often proposed to be used for system level specifications. However, verification is either conducted dynamically, during simulations, or formally for SystemC models, for which RTL<sup>1</sup> code generation is possible. By the knowledge of the authors formal verification and constraint verification is not possible for generalized SystemC models.

To overcome this gap, we present a constraint driven approach to verify system level properties in early design stages. The verification flow realized in our work is depicted in Figure 1. The simulation model is supplemented with formal constraint information. Then this validation model is transformed into a verification model, which is used as input for a verification tool. Results from the verification tool can be back-annotated into the simulation model.

This paper is organized as follows. First related work is presented. Then we describe the verification flow and the transformation tool in detail. The paper ends with our conclusions and an outlook to the future work.

<sup>1</sup>Register Transfer Language

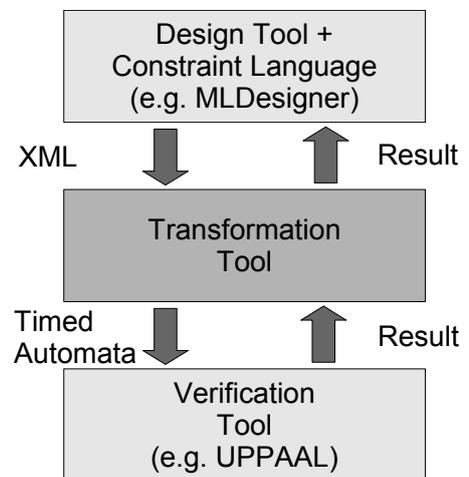


Fig. 1: Verification Flow

## 2 Related Work

Based on the requirements identified and proposals made in [3] an approach for verifying a system model's temporal properties has been developed. As proposed, a structured constraint language has been defined and the transformation of time constraint expressions to systems of timed automata has been developed and implemented in [4]. By exploring the state-space of resulting automata the verification of the time constraints can be conducted. The tool UPPAAL is supposed to be utilized for this verification.

Two approaches for designing the systems have been considered, using MLDesigner discrete event models, as proposed in [3], or UML activity, sequence and state diagrams, as proposed in [4]. The tool presented in this paper has implemented the MLD-based approach.

### 2.1 System Modelling

MLDesigner [5] is a tool based on the ideas of the Ptolemy project [6]. System models can be easily created by using existing customizable functional blocks and connecting them. New blocks can be created using the Ptolemy Language, a facility to describe functional blocks in terms of C++ code fragments. The tool MLDesigner organizes the structural connection, creation, execution and debugging of models. The tool provides the ability to link different models of computation within one system. This enables to create efficient simulation models, but limits the capabilities of implementation code generation.

According to the underlying ideas we can find a lot of similarities between SystemC 2.0 [7] and the MLDesigner Discrete Event (DE) modelling domain. A scheduler is used to organize the execution (simulation) of functional blocks. Timing information is added to the simulation in terms of delay. Data is exchanged using memories and/or pointers. Signals are event based, using the scheduler to inform the functional blocks. When describing hardware, there are some drawbacks according to hardware specific support (e.g. data types) compared to SystemC. In our opinion MLDesigner represents the more general approach on resource and systems modelling. Furthermore it provides powerful debugging opportunities according to the discrete event modelling approach.

The MLD system models are optimized for fast simulation speeds. Due to these optimizations (extensive use of pointers, abstract data types, simulation management, describing functional behaviour in terms of C++ code), it is hard to extract a state model to apply model checking techniques directly. In our approach we substitute the C++ code with timing constraints, a structured language defined in EBNF describing the behaviour based on events, abstracting from data structures. Of course, finite state modules (domain FSM) can be embedded in discrete event models (domain DE).

### 2.2 Constraint Languages

There are different ways to classify properties. One is to distinct between constraints and assertions where constraints are limitations to both the environmental behaviour (regarding to the inputs) and the expected system behaviour (regarding to system outputs) [8]. Another is to state all side effect free specifications as system constraints, as done in the Object Constraint Language (OCL) [9].

In this paper every specified behavioural property is named *constraint*. The role of such a property depends on the view on the system or its fragments. A property can be expected to be already fulfilled or asserted to be fulfilled by the current device under verification. Regarding to

the system modelling environment depicted above, we need to specify the use of (sub)systems within their environments. This includes to specify restrictions on inputs and outputs (safety properties), following called *specification constraints*. We also need to specify behavioural constraints between inputs and outputs, representing a (preferably minimal) subset of the systems expected behaviour, following called *implementation constraints*.

### 2.3 Verification Approach

There exist different kinds of models, that are suited for formal verification: clocked state models, timed automata, Petri nets, interval Petri nets, process algebra models, etc. These models provide different possibilities to model data, synchronization and timing. On the other side there are different, optimized internal data structures to represent state space and algorithms to apply state traversal and property checking, e.g. BDD<sup>2</sup> based symbolic verification. Because of the use of zero timed events in system models, we cannot use clocked state models. We need a state model involving timing in terms of clocks. Thus we decided to use timed automata, providing channel based communication, a synchronizing scheme, which is very similar to events. For verification we use the tool UPPAAL [10]. It provides a good support and is freely available.

## 3 Enhanced system models

Although well suited for developing event based behavioural system models, MLDesigner does not feature adequate means to describe the temporal behaviour of model elements [3]. So MLD model elements have to be artificially enhanced by Constraint Language expressions. There are four different model elements to be considered for applying temporal constraints: on one hand *System*, *Module* and *FSM* as complex (sub)model descriptions, on the other hand the subordinate model element *Instance*, which represents a functional part inside a complex model.

Of importance for model consistency is to keep the right *perspective*, when formulating the constraint expressions for a certain model element. This concept of perspective is originated in the design perspective of MLD, which limits the view to the elements residing in one hierarchical level and does neither provide nor require complete knowledge of the full model hierarchy. This has a major influence on the complexity of the event identifiers used within the constraint expressions.

Since the functional behaviour within a system is supposed to be provided by *Instance* blocks, they have to carry **implementation** constraint expressions, that describe the temporal dependencies between their input and output signals. The perspective is thereby limited to the boundaries of the particular instance.

The *System* and *Modules* as complex system definitions contain instances and their relations. The perspective is a *local perspective* set to the system's own hierarchical level, so the constraints expressed on this level have to be **specifications** related to the (sub)system's inner signals. Concerning verification the necessity may arise to make assertions on the occurrence of any signal from any instance on any subsystem level. So, for the *System* an additional *global perspective* needs to be defined, which exceeds the MLD provided view and requires knowledge of the complete model hierarchy. Here **global specification** constraints have to be formulated, that refer to signals throughout all model hierarchy levels.

---

<sup>2</sup>binary decision diagram

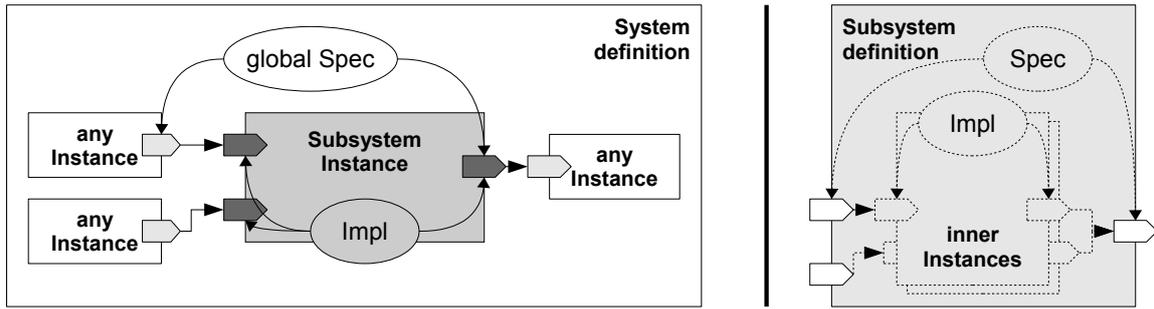


Fig. 2: MLD model elements and constraints

An *FSM* block contains an automaton structure (states and transitions). The perspective is limited to the boundaries of the FSM. The internal behaviour of an FSM depends on the *states*, so the **FSM implementation** constraints timing the state transitions are stored there. *Transitions* contribute to the internal behaviour by featuring *conditions* and *actions* - expressing the dependence of the transition from input events on one hand and the dependence of output events from transitions on the other.

The MLD modelling paradigm enables a once defined module or primitive to be instantiated in systems or modules various times on different hierarchy levels. Since the transformation produces a flat model, system wide unique identifiers have to be introduced. The concept of distinct constraint perspectives combined with respectively enhanced event identifiers eases the element handling and prevents conflicts on the resulting single system level.

#### 4 The tool MLD2UPPAAL

The tool MLD2UPPAAL conducts the transformation from the enhanced MLD model to the UPPAAL timed automata system by passing three main conversion stages. Two intermediate model formalisms have been developed to support the transformation and provide a basis for different model manipulations. A summary of this workflow is shown in Figure 3.

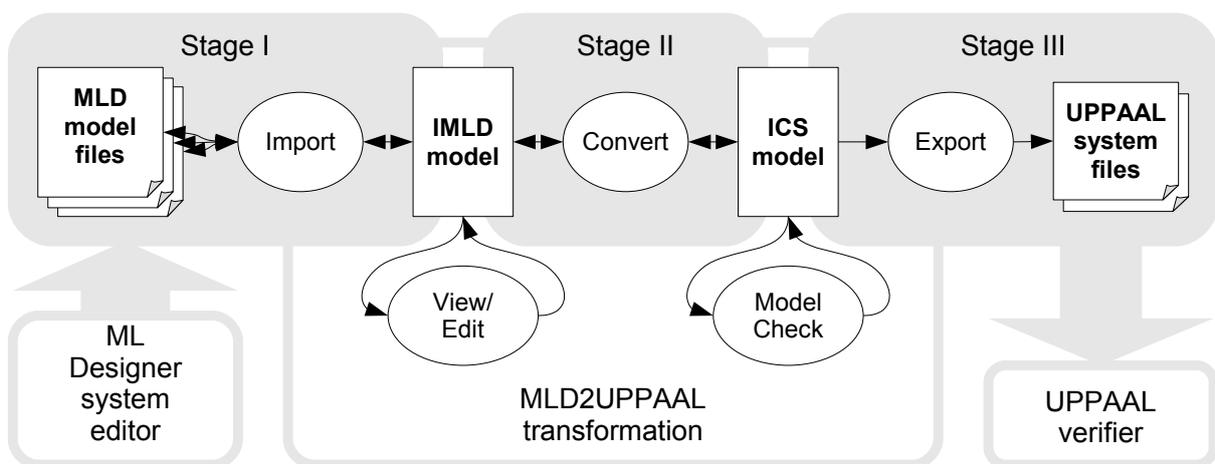


Fig. 3: MLD2UPPAAL workflow

Since an MLD model features a hierarchical structure, the model and each subsystem and element definitions are stored in separate files. In the first processing stage a given set of model

files, including the top level system file and all referenced files, are imported into an *Intermediate MLD Model (IMLD)* instance. Main purpose of this intermediate formalism is to join the MLD system and subsystem definitions into one model while maintaining the hierarchy information. At this stage the application allows varying the level of hierarchy resolution and the revision of the constraint expressions. As the workflow indicates, the original model definition is kept consistent to the intermediate models by backpropagating content changes. The second stage converts the intermediate MLD model into an instance of the *Intermediate Constraint System Model (ICSM)*. During this conversion the system structure is flattened, which results in the substitution of all subsystems according to whether they are to be resolved or not. Unresolved subsystems are treated as simple instance blocks with implementation constraints, whereas resolved subsystems instances are replaced by their inner structure and local specification constraints to provide assertions about that system partition. The ICS model notation allows the check of syntactic and semantic constraint consistency in consideration of the concept of perspective mentioned above. Once the check is completed successfully, the ICS model is transformed to timed automata using UPPAAL templates to map the constraint properties. An UPPAAL-compliant XML-notation is stored in model and query files, that serve as input for the UPPAAL verifier.

The tool is implemented as a Java application, that appears as shown in Figure 4, and was successfully tested against selected examples.

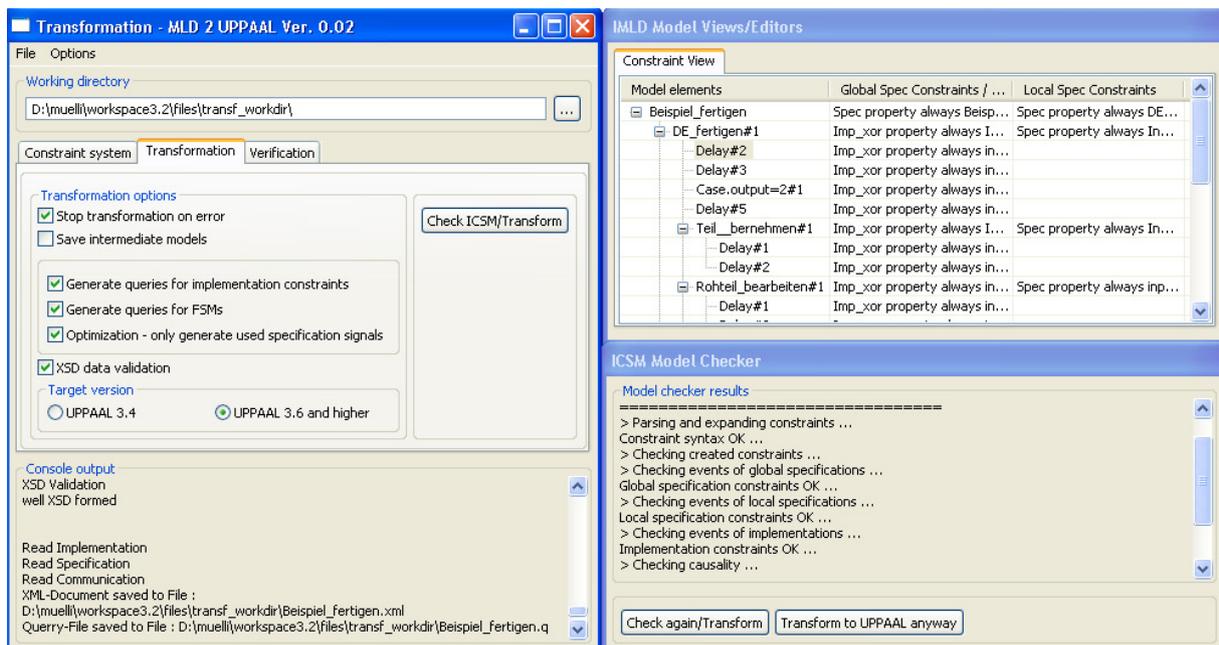


Fig. 4: MLD2UPPAAL screenshot

## 5 Conclusions and future work

By introducing the temporal constraint language and applying it to MLD designer system models we showed the possibility to add formally verifiable properties to models with high degrees of abstraction. The tool implemented in this work's context gives an example of how to bridge the gap between design and verification formalisms. Since it has been derived from the application framework introduced in [11], a customization to various design sources

and verification targets is easily possible. The given transformation of constraints to UPPAAL automata can be easily extended by adding custom templates. Currently we are going to apply our verification method to a number of design models to validate and benchmark it regarding model complexity. Using completely formally interpretable simulation models will make the models more directly verifiable and thus reduce the amount of additional information to be supplied in form of constraints. An according approach concerning extended finite temporal properties has already been presented in [12].

### Acknowledgements

This work has been funded by edacentrum e.V. within project FEST.

### References

- [1] SCHORCHT, G. ; UNGER, P. ; GEORGE, A. ; TROXEL, I. ; ZINN, D. ; SALZWEDEL, H. ; FARHANGIAN, K. ; MICK, C.K.: System-Level Simulation Modeling with MLDesigner. In: *IEEE / ACM MASCOT 2003 - 11th ACM / IEEE International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems, 12.-15. Oktober 2003, Orlando, Florida, USA* (2003)
- [2] T.KROPF: *Introduction to Formal Hardware Verification*. Springer Verlag, 1999. – ISBN 3–540–65445–3
- [3] PACHOLIK, A. ; FENGLER, W. ; SALZWEDEL, H. ; VINOGRADOV, O.: Real Time Constraints in System Level Specifications Improving the Verification Flow of Complex Systems. In: *Net.ObjectDays 2005* (2005), September
- [4] KLÖCKNER, J.: *Transformation of UML System Descriptions to Verify Temporal Properties*. 2006. – Diploma Thesis, in German language
- [5] MLDESIGN TECHNOLOGIES, Inc.: *MLDesign Technology Homepage*. : MLDesign Technologies, Inc., 2006. – <http://www.mldesigner.com/>
- [6] DEPARTMENT OF EECS, UC B.: *Official Homepage of the Ptolemy Project*. : UC Berkeley, EECS, 2006. – <http://ptolemy.eecs.berkeley.edu/>
- [7] GHENASSIA, Frank: *Transaction-Level Modeling with SystemC*. Springer, 2005. – ISBN 0–387–26232–6
- [8] FOSTER, H.D. ; KROLNIK, A.C. ; LACEY, D.J.: *Assertion-Based Design 2nd Edition*. 2. Kluwer Academic Publishers, 2004. – ISBN 1–4020–8027–1
- [9] WARMER, J. ; KLEPPE, A.: *Object Constraint Language 2.0*. mitp-Verlag Bonn, 2004. – in german language. – ISBN 3–8266–1445–3
- [10] DEPARTMENT OF INFORMATION TECHNOLOGY, Sweden: *Official Homepage of the Uppaal Tool*. : Department of Information Technology, Uppsala University, Sweden, 2006. – <http://www.uppaa1.com>
- [11] MÜLLER, M.: *Metamodel based Transformation of System Models - an Application Framework*. 2006. – Diploma Thesis, in German language
- [12] PACHOLIK, A. ; FENGLER, W.: A System Model for Formal Verification of TLM based Transaction Properties. In: *10th Communications and Networking Simulation Symposium (CNS) 2007* (2007)

### Autors:

|  |                                     |
|--|-------------------------------------|
| Dipl.Inf. Marcus Müller,               | Department of Computer Architecture |
| Dipl.Ing. Alexander Pacholik,          | Technical University of Ilmenau     |
| Prof. Dr.-Ing. habil. Wolfgang Fengler | 98684 Ilmenau, Germany              |

{marcus.mueller, alexander.pacholik, wolfgang.fengler}@tu-ilmenau.de

A. Pretschner / J. Alder / Ch. Meissner

# **A contribution to the design of embedded control systems**

## **1. Introduction**

The area of automated control engineering increasingly tends towards distributed autonomous systems. Thus rises the requirement for control modules that can be implemented in a simple and cost-effective way by means of ordinary PC technology. An obvious case is reliable Linux-based systems. Such systems, also in form of embedded systems, represent an alternative to the conventional control systems. This development is accelerated by the rapid growth of the hard and software components. The introduction of the International Electrotechnical Commission 61131-3 (international standard of modern software programming of programmable logic controllers) has likewise a positive influence on the use of off-the-shelf PC technology in a market that has been dominated by proprietary industrial technology. Even the IEC 61499 provides a framework and architecture for describing the functionality in distributed control systems. The engineering tool [3, 5] can be used to fill out these standards with executable software. With the developed software tool [6] the software design process of automatic controls can be accomplished in particular for users of embedded systems error free. The Tool- Chain from the process behavior to the process visualization, developed by the example of the ARM9 processors, illustrates the typical engineering process for control-system applications.

## **2 Preceding work**

For the automated design of control programs several approaches exist. We choose an automata-theoretic approach as the basis of our design tool. The structure of the program package is governed by the IEC 61131 standard. Access to variables is implemented by a data base, declaration of classes is implemented by function blocks, and their instantiation and integration into the program are also implemented. The system requirements are described and converted into an appropriate software tool. Transfer to practice is done using conventional control applications. The intensified employment of distributed automation systems led to the use and to the design of control systems based on PC and/or microcontroller. The necessary information of the description of the software packages are referred on the home site of the project. Among them are both a detailed on-line assistance, and the system manual as pdf file [4]. The executable files of the project are freely available.

## **3 Function block design**

The PLC is provided as automatically generated code by the design tool SPaS [3,4]. The graphic editor enables to create process flow charts. Each process flow chart syntactically corresponds to a function block. The function block design corresponds to

the design of the IEC 61131-3 function block design

The behavior of the function block is defined in terms of the algorithms and state information. In extension to the IEC 61499 the algorithms are expressed as state machines and the Execution Control Chart is coupled with the state machine. The *WITH* qualifier will be omitted with assumption of a sequential behavior of the scheduler

All function blocks have to be programmed completely by graphics. The project structure in the design tool is governed by the IEC 61131-3 structure guidelines, i.e., there are project nodes for the resource, the programs, as well as for the needed instances. The function blocks can be linked by simply connecting event / data flow connections between block input and output variables. A special graphical net-editor does not exist at time yet. The development is under way. Linking the function block instances by variables you have to use a text editor manually. The execution order of the function blocks is determined by the user choosing the appropriate level in the project tree. The IEC 61131-3 software model does consider configuration that have multiple resources. To communicate between them you can use global variables or communication function blocks (IEC 61131-5)

From the graphic representation, syntactically correct code (in this case C and C++) is produced. This code may be integrated as an independent task. Here also the binding to the I/O of the controller takes place. The design is transparent and permits the user to map to equivalent source code which later has to be compiled to binary code.

For the start-up of the control a back representation of the graph can be created. It can visualize at run time the current state of the automaton.

#### **4 Extension to distributed systems**

Applying to distributed systems you have to take care about the execution order. Unfortunately, such mechanisms are not defined in the IEC61131-3. The main restriction using the IEC 61131-3 software model is described as following [22]:

- Applications in the IEC 61131-3 model are not distributable over multiple resources.
- The function block execution order is not always clearly defined.

The function blocks expressed in IEC 61131-3 language can be encapsulated as IEC 61499 function blocks as a first approach.

The application has to be separated into different resources necessarily at this modeling stage. Further you have to bring in communication function blocks to the model. In this application an interface function block was designed. It is a composite function block containing appropriate Publisher / Subscriber function blocks and / or IO\_Writer / IO\_Read function blocks. This modeling manner is often sufficient for ordinary PLC's, which are driven sequentially by their operating system. The execution order of the function blocks depends on the calling level in the main thread of the operation system. The execution control of the composite function blocks may be deferred to the mechanisms described by the IEC 61499.

The execution order of the function blocks and the validity of the sampled data is not clearly defined for more advanced applications realized with microcontrollers. You have to provide an execution control chart (ECC) to the composite function block at each resource to achieve a well defined coordination of the net behavior. Service interface function blocks may be necessary to communicate to the devices.

The deployment of the software modules on distributed resources is possible. The use of the flow Charts of " SPaS" in connection with the function block model of the IEC 61499 allow the administration of several resources in one project. Communication between the resources is not the subject of the draft process.

## 5 Communication device

Once the program is translated to C or C++ the communication to the inputs/ outputs of the process has to be established. This is done via a common bus system: CAN. Using the standard CANopen application interface, the system provides the master slave-behavior between the master control device (this) and the I/O devices. To obtain a modular system, we have chosen the USB-interface as breakout to the periphery. We developed several layers, which handle bus protocols and software interfaces to provide an easy high abstracted interface of I/O access (see figure 1). Even you can see the software module deployment between the kernel and the user space of Linux.

The first two layers present themselves as linux kernel modules. First *CANu* connects to the CAN controller device and distributes CAN messages and second *CANopenDrv* implements the *CANopen* protocol. *CANu* receives incoming CAN messages and delivers them to specific applications, which are registered on this module. With this approach, the applications, which are user space programs or other kernel modules, share one resource. Internal *CANu* follows completely the idea of CAN to broadcast messages, because a message received by a CAN controller is delivered to every on this device registered application and the other way around. The advantage of this is, that the limited access of one device is now overridden and an abstracted CAN Interface is provided. *CANopen Drv* is a virtually *CANopen* Master/Slave conforming the CiA 301 DS V4. 0.2 specification. It's protocol defines broadcast and client-server communication objects by partitioning the available range of CAN identifiers. It processes received messages and handles network events and states. According to this the I/O data is provided through the internal object dictionary, which can be accessed by the control application. The control application itself runs in userspace and communicates with the *CANopen* driver over syscalls. That means, all important time critical *CANopen* protocol functions are running in kernelspace with very low latency. To improve the handling, an additionally developed C library is provided as a third layer, which makes it easier to use the *CANopen* driver. On top of this another C++ library is implemented to encapsulate the functionality one more time. Based on this, it is possible to derive own classes to be used in the control applications *iopl* interface. The control program is developed as a plugin module of the Lintouch server application. If the control system application is divided in parts over the communication network, each intelligent network node should correspond to these hard and software requirements.

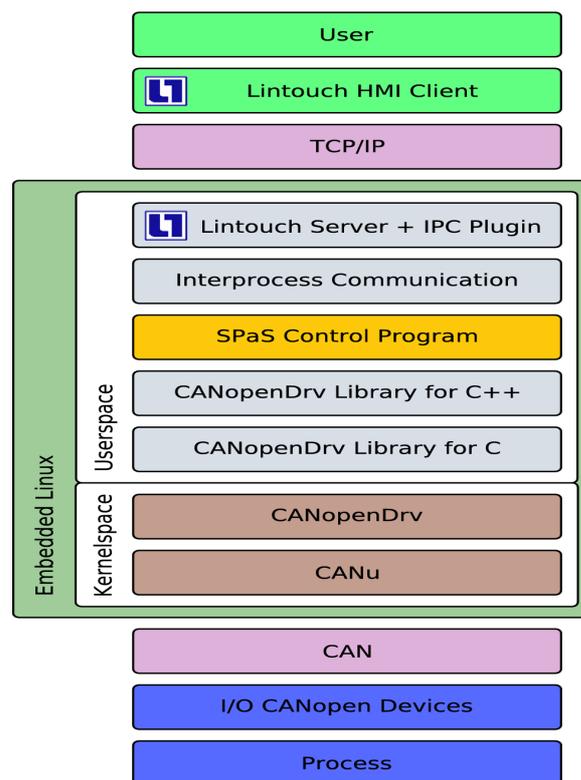


Figure 1: Software modules and communication layers

## 6 Conclusion

As you can see the aim to provide an easy way to access the data of I/O devices in the *CANopen* network is fulfilled. In order to that it is also possible to exchange *CAN* controllers and I/O devices, without changing the control applications code. The next step is the communication between the control program and the user. This is done by the open source human machine interface (HMI) *Lintouch*. It contains a TCP/IP server on the control system that collects and distributes process variable values. A user is now able to connect from a plugged ethernet network to this server. That lintouch client downloads a previous generated lintouch project file from the server, which provides an interface that displays the current process state and allows the user to interact with the process. This interface could run on a panel or something else. To gather the current process variables the lintouch server uses plugins, which are building the bridge between the server and the process. In this case we developed our own, which is able to receive and transmit data from the *SPaS* control program. Our decision leads to interprocess (OS process) communication. This means, that we use the *System V IPC* of linux, especially the shared memory capabilities, to distribute data to the server and the other way around. Because *SPaS* uses data vectors as data input, output and memory, we followed this approach and map these vectors to shared memory. The plugin now reads and converts the data to lintouch conform formats. The interaction with the process only happens via the memory vector, so that the direct feedthrough setting of output variables is not possible.

Because of the high abstraction of the I/O access and process visualization, it is imaginable easy to exchange the used control program, which also can be generated very easy by the *SPaS* software.

### References:

- [1] IEC 61131-3. Programmable controllers - part 3: Programming languages. Technical report, DIN Deutsches Institut für Normung e.V., 2004.
- [2] Robert Lewis. Modelling control systems using IEC 61499. IEE Engineering Series 59. IEEE, The Institution of Electrical Engineers, 2001.
- [3] A. Pretschner. Softwareprojektierungsumgebung für automatische Steuerungen SPAS. In International Congress SPS/IPC/Drives Nürnberg 2000, pages S. 99–103, 2000.
- [4] A. Pretschner. Spas - online - dokumentation. Technical report, University of Applied Sciences Institut for Prozessinformation Technology and Controls Systems, 2005.
- [5] VDI/VDE 3681. Einordnung und Bewertung von Beschreibungsmitteln aus der Automatisierungstechnik. Technical report, VDI/VDE-Handbuch Regelungstechnik, 2004.
- [6] Alder,J.;Pretschner,A.:Prozess-Steuerungen- Projektierung und Inbetriebnahme mit dem Softwaretool SpaS, Springerverlag 2007, ISBN 978-3-540-71083-7

### Authors:

Prof. Dr. Andreas Pretschner  
Dr.-Ing. Jochen Alder  
HTWK Leipzig, FBEIT, PSF 30 11 66.  
04251, Leipzig  
Phone:+49 341 3076 1135  
Fax:+49 341 3076-1243  
E-mail:andreas@pretschner.com

R.Ubar / G.Jervan / J.Raik / M.Jenihhin / P.Ellervee

## **Dependability Evaluation in Fault-Tolerant Systems with High-Level Decision Diagrams**

### **INTRODUCTION**

To achieve high dependability, systems today are often designed with fault tolerance features to first detect errors and then to mask or recover from the effects of those errors. Thus, testing of these features is extremely important in understanding how dependable the systems are with the incorporated fault tolerance mechanisms and in gaining insight into the success of error detection and recovery. Fault injection is a means to effectively test and stress the error handling and fault tolerance mechanisms, so that the system behavior can be studied prior to their actual deployment.

Fault injection techniques can be classified in three main categories [1]: physical (or hardware Implemented fault injection), software implemented and simulation-based. Simulation-based fault injection is a useful experimental way to evaluate the dependability of a system during the design phase.

Recently many simulation based fault injection techniques have been proposed for system dependability evaluation [2-6]. They are targeted for execution-based models of working systems. Simulating faults in a system model (e.g. based on hardware description language VHDL, Verilog, System C etc) assures high flexibility.

However, there exists a problem of selecting faults to be injected. Erroneous responses in a system, in many cases, do not necessarily lead to a failure at the application level, even when the discrepancy with the nominal behavior has a long duration. An accurate but high-level fault analysis in the complete system is therefore required to discriminate real failure conditions from non-critical errors. Such an analysis is very difficult to carry out on the execution-based models using languages like VHDL, Verilog, System C. In this paper a method is proposed based on high-level modelling of systems with Decision Diagrams (DD) to select faults for injection targeting dependability evaluation.

## **MALICIOUS FAULTS AND FAULT INJECTION**

Traditionally in simulation-based fault injection techniques, the fault location, fault type and fault insertion time are typically selected at random. The drawback of randomly selected faults is the large probability that the injected fault will remain latent; that is, produce a no response fault injection experiment. This will reduce the quality of dependability evaluation procedure. Thus, there is a need to locate faults which do not belong to the no response category.

The goal of the fault injection experiment is to exercise the system's fault processing capabilities. Faults which fail a system in the absence of system fault detection capabilities are defined to be malicious [8,9]. Malicious faults systematically exercise the fault processing attributes of the system. A malicious fault if undetected will fail the system under test. Likewise, a malicious fault is guaranteed to produce an error which will produce a failure if it is not properly processed by the system. Thus, using malicious faults to estimate the dependability of the system eliminates the possibility of the no response fault injection experiment.

Malicious fault list can be generated by creating fault trees using fault dependency analysis by reverse implication and providing fault collapsing. To cope with the complexity of fault analysis when carried out at plain low (logic) levels, we propose a hierarchical approach where the fault tree creation and fault collapsing will be carried out first at a higher level, thereafter to refine the fault injection points at a lower level. To carry out the fault dependency analysis and creation of malicious fault lists we propose to use the method of diagnostic modelling digital systems by decision diagrams (DD) [10-13].

## **MODELLING DIGITAL SYSTEMS WITH DECISION DIAGRAMS**

DD-s allow to investigate and solve the problems of fault dependency analysis at different abstraction levels of digital systems depending on the complexity of the system description. The well known Binary Decision Diagrams (BDD) [10] can be considered as a particular case of this model for using at logic level.

For lower (logic) level fault analysis we will use a special class of BDDs – structurally synthesized BDDs (SSBDD) [11] to represent the topology of digital circuits in terms of signal paths at the fanout free region (or macro) level. Macro level fault analysis can be carried out more efficiently than traditional gate level analysis [13].

In high-level architectural descriptions, we usually partition the system into high level components or subsystems where the descriptions of subsystems in general case can be modeled by control and data paths. These paths can be represented by a set of DDs where the nonterminal nodes in DDs correspond to the control variables (instruction codes, addresses, control words, logic or timing conditions etc), and the terminal nodes correspond to data variables, functional expressions or more complex behavioral or algorithmic descriptions of other subsystems. The terminal nodes will be represented again by a set of DDs to allow entering a lower hierarchical level for disclosing the label of the terminal node [12,13].

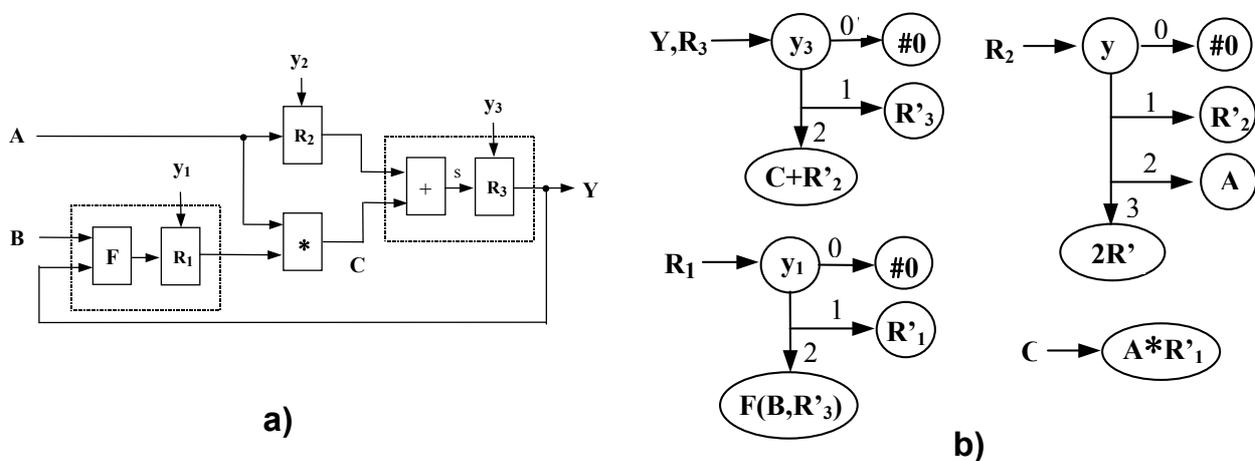


Fig.1. Simple data-path represented by DDs

In Fig.1 a simple register-transfer level data path consisting of 4 subcomponents (three registers with input logic and a multiplier) each represented by its DD is depicted. Each graph has a single nonterminal node labelled by a control variable whereas the terminal nodes are representing reset, hold, load and data manipulation (addition, shift, multiplication and a nondisclosed  $F(B,R'_3)$ ) functions. The values on edges correspond to the values of control signals. The apostroph shows that the value of the register variable is taken from the previous clock cycle.

Each path in a DD describes the behavior of the subsystem represented by DD in a specific mode of operation (working mode). For example, the path from the node  $y_3$  to the node  $C + R'_2$  in the graph  $R'_3$  represent the addition operation  $C + R'_2$  activated by the control signal  $y_3 = 2$ .

The faults having effect on the operation associated with activated path can be related to the nodes along the path. A fault may cause incorrect leaving of the

activated path. From this point of view we can introduce a very general fault model for DDs related to faulty behavior of nodes. In the case of SSBDDs the faults of a binary node are equivalent to the traditional logic level stuck-at-0(1) fault model.

In more general case of DDs the faults of the nonterminal nodes can be interpreted as the addressing fault model of microprocessor cores [13]: output edge is either always activated (stuck) or always broken, or instead of the given edge, another edge or a set of edges is activated. This is a very general fault model that allows to represent most of control errors in the case of microprocessor cores.

It is easy to extend the described fault model of a DD node to model complex faults like shorts, delays, crosstalks etc. All of these faults can be represented by a checkpoint where the error manifests at certain conditions i.e. by a pair (fault site; fault condition). Since the nodes of DDs can be regarded as checkpoints the fault model on DDs can be represented as a pair (node; fault condition). In such a way the fault model on DDs can be seen as a very powerful model able to cover a very wide class of faults in digital systems.

The formalism of DDs allows to implement simple algorithms to analyze the cause-effect and fault dependency relationships which would be impossible in case of descriptions written in languages like VHDL, Verilog, System C.

### **MALICIOUS FAULT LIST GENERATION WITH DECISION DIAGRAMS**

Malicious faults will be generated for a given set of test sequences. A test activates a process in a system consisting of a sequence of iterations. Each iteration (clock cycle, microinstruction or instruction cycle, transaction etc.) activates paths in DDs of the corresponding DD-model of the system. The faults having effect on the behavior of the system at a given iteration can be related to the nodes of the activated path.

For each observable checkpoint  $C_i$  of a test sequence, a Fault Tree (FT) with a root  $N_i$  will be created. Assuming an erroneous signal is detected at  $C_i$ , a set of candidate faulty checkpoints  $C_k$  explaining the misbehavior in  $C_i$  is created by tracing the activated path on the corresponding DD for  $C_i$ . This set of checkpoints  $C_k$  will be included into FT as successors of  $N_k$ . For each node  $N_k$  in FT, a list of malicious faults associated with  $C_k$  is determined. For all leaf nodes in FT to be constructed the same procedure is repeated until the input events for the test sequence are reached. If  $C_k$  represents stored data (register variables), we will continue fault dependency analysis in the previous test iteration. Each node in FT characterised with a list of

malicious faults and the number of iteration can be used as a point for fault injection. To reduce the number of malicious faults, the FT can be compressed (collapsed) by using fault equivalence and fault dominance relationships. The fault model is not restricted by the approach. In general, arbitrary changes  $R + \epsilon$  of the values of a variable  $R$  can be accepted. The set of accepted values of  $\epsilon$  can be given by a fault list.

An example of FT created for the data path in Fig.1 and for the test sequence in Table 1 is shown in Fig. 2. The malicious fault list selected from FT is depicted in Table 2.

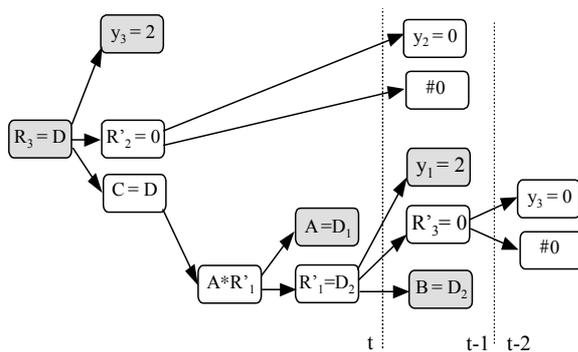


Fig.2. Fault tree

| t | y <sub>1</sub> | y <sub>2</sub> | y <sub>3</sub> | A              | B              | C | R <sub>1</sub> | R <sub>2</sub> | R <sub>3</sub> | Y |
|---|----------------|----------------|----------------|----------------|----------------|---|----------------|----------------|----------------|---|
| 1 |                |                | 0              |                |                |   |                |                | 0              |   |
| 2 | 2              | 0              |                |                | D <sub>2</sub> |   | D <sub>2</sub> | 0              |                |   |
| 3 |                |                | 2              | D <sub>1</sub> |                | D |                |                | D              | D |

Table 1. Test sequence

| t | y <sub>1</sub> | y <sub>3</sub> | A                             | B                             | R <sub>2</sub>      | Y              |
|---|----------------|----------------|-------------------------------|-------------------------------|---------------------|----------------|
| 2 | 0              |                |                               | D <sub>2</sub> + $\epsilon_2$ | $\epsilon_3 \neq 0$ |                |
| 3 |                | 0,1            | D <sub>1</sub> + $\epsilon_1$ |                               |                     | D = $\epsilon$ |

Table 2. Malicious fault list

## CONCLUSIONS

A new approach based on fault dependency analysis with Decision Diagrams is proposed for selecting faults to be injected targeting fault tolerance evaluation in digital systems. Such a fault analysis is impossible when the execution-based models written in VHDL, Verilog or System C are used. The proposed method allows to avoid no response fault injection experiments and as the result to increase the quality of dependability evaluation.

**Acknowledgement:** The work has been supported by Estonian Science Foundation grants 5910, 6717, 6829 7068, EC 6th framework IST project VERTIGO, Estonian Information Technology Foundation (EITSA) and Enterprise Estonia.

## References:

- [1] A.Benso and P.Prinetto, eds. Fault Injection Techniques and Tools for VLSI reliability evaluation. Kluwer Academic Publishers, 2003.
- [2] E.Jenn, J.Arlat, M.Rimén, J.Ohlsson, J.Karlsson, "Fault injection into VHDL models: the MEFISTO tool", in Proceedings 24th International Symposium on Fault-Tolerant Computing (FTCS-24), pp. 356-363, Austin, TX, USA, 1994.
- [3] V.Sieh, O.Tschäche, F.Balbach, "VERIFY: Evaluation of Reliability Using VHDL-Models with Embedded Fault Descriptions", in Proceedings 27th International Symposium on Fault Tolerant Computing (FTCS-27), pp. 32-36, Seattle, WA, USA, 1997

- [4] D.Gil, J.Gracia, J.C.Baraza, P.J.Gil, "A Study of the effects of Transient Fault Injection into the VHDL Model of a Fault-Tolerant Microcomputer System", in Proceedings 6th IEEE International On- Line Testing Workshop (IOLTW'2000), pp. 73-79, Palma de Mallorca, Spain, 2000.
- [5] R. Leveugle and K. Hadjiat, "Multi-Level Fault Injections in VHDL Descriptions: Alternative Approaches and Experiments," J. Electronic Testing: Theory and Applications (JETTA), vol. 19, no. 5, pp. 559-575, Oct. 2003.
- [6] J.C. Baraza, J. Gracia, D. Gil and P.J. Gil. Improvement of Fault Injection Techniques Based on VHDL Code Modification. 10th IEEE International High-Level Design Validation and Test Workshop, 30 Nov.-2 Dec. 2005 Page(s):19 – 26.
- [7] J.Sosnowski, P.Gawkowski, P.Zygulski, A.Tymoczko. Enhancing Fault Injection Testbench. Proceedings of the International Conference on Dependability of Computer Systems (DEPCOS-RELCOMEX'06), May 2006 Page(s):76 – 83.
- [8] D.T.Smith, B.W.Johnson, J.A.Profeta. System Dependability Evaluation via a Fault List Generation Algorithm. IEEE TRANSACTIONS ON COMPUTERS, VOL. 45, NO. 8, AUGUST 1996, pp.974-979.
- [9] A.Benso, P.Prinetto, M.Rebaudengo, M.Sonza, R.Ubar. A New Approach to Build a Low-Level Malicious Fault List Starting from High-Level Description and Alternative Graphs. Proc. IEEE European Design & Test Conference, Paris, March 17-20, 1997, pp.560-565.
- [10] R.E.Bryant. Graph-based algorithms for Boolean function manipulation. IEEE Trans. on Computers, Vol.C-35, No8, 1986, pp.667-690.
- [11] A.Jutman, A.Peder, J.Raik, M.Tombak, R.Ubar. Structurally Synthesized Binary Decision Diagrams. 6th International Workshop on Boolean Problems, Freiberg, Germany, Sept. 2004, pp.271-278.
- [12] G.Jervan, R.Ubar, Z.Peng, P.Eles. Test Generation: A Hierarchical Approach. In "System-level Test and Validation of Hardware/Software Systems" by M.Sonza Reorda, Z.Peng, M.Violante. Springer Series in Advanced Microelectronics, Vol.17, 2005, pp. 63-77.
- [13] O.Novak, E.Gramatova, R.Ubar. Handbook of Electronic Testing. CTU Printhouse, Prague, 2005, 400 p.

**Authors:**

Prof. Raimund Ubar

Dr. Gert Jervan

Dr. Jaan Raik

Mr. Maksim Jenihhin

Prof. Peeter Ellervee

Tallinn University of Technology, Raja 15

12617, Tallinn, ESTONIA

Phone: +372 6202252

Fax: +372 6202253

E-mail: {raiub, gerje, jaan, maksim}@pld.ttu.ee, lrv@cc.ttu.ee

A. Jutman

## On LFSR Polynomial Calculation for Test Time Reduction

### ABSTRACT

Testing of nanoscale semiconductor devices, nowadays, faces a strong shift from classical methodologies towards self-test, self-diagnosis, and self-repair solutions – all to improve manufacturability and reliability characteristics of the final product. Due to these efforts, it is possible to keep production costs at reasonable levels when moving to finer and less reliable manufacturing technologies. Linear Feedback Shift Registers (LFSR) and other Pseudo-Random Pattern Generators (PRPG) have become one of the central elements used in embedded testing and diagnosis of contemporary complex electronic systems like processors, controllers, and high-performance integrated circuits. Current paper presents a mathematical framework of LFSR polynomial calculation for fault coverage improvement and test cost reduction. The proposed technique allows shortening test runtimes and increasing fault coverage by embedding specific pre-calculated test patterns into the PRPG sequence.

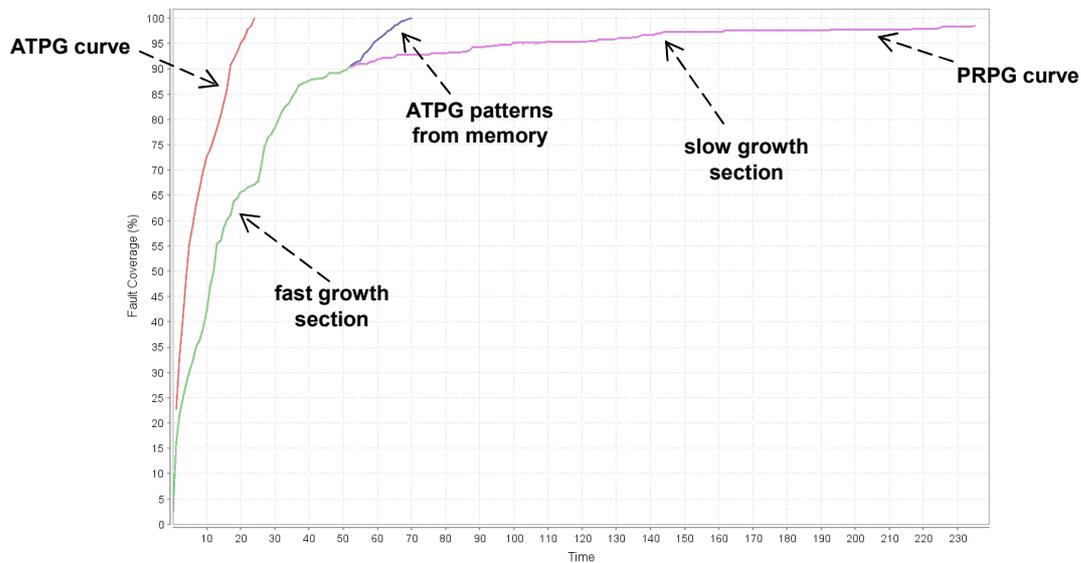
### INTRODUCTION

Accordingly to the International Technology Roadmap for Semiconductors (ITRS) [1], the increasing complexity of recent VLSI circuits and transition to multi-core System-on-Chip (SoC) and Network-on-Chip (NoC) paradigms has made testing (including planning, test generation and scheduling) one of the most complicated and time-consuming problems in the domain of digital design.

During the last several years, ITRS reports indicate that the semiconductor manufacturing industry is inevitably moving towards test compression and self-testing approaches that allow either to efficiently feed test data to individual system cores or to initially design self-testable cores. The main driving force of such a transition is the bandwidth gap between the I/O frequency and very high internal clock rates of modern semiconductor devices.

As the result, built-in self-test (BIST) becomes widely recognized as one of the promising approaches for testing modern nanoscale devices – especially SoC-s. Unlike external test pattern generators, embedded test facilities have a good access to any internal core or unit and can be used for self-test and self-repair purposes. Moreover, they can work at the same speed as the system providing detection of defects, which do not manifest themselves at lower clock rates.

Pseudo-random test pattern generation (TPG) techniques are the main instrument used in BIST. Linear feedback shift registers (LFSR) represent the simplest and most commonly used pseudo-random TPG (PRPG) hardware. However, in terms of fault coverage and testing runtimes, the efficiency of an LFSR is far from optimum. A test



**Figure 1. PRPG patterns and Fault Coverage trends**

generated by an LFSR is usually an order of magnitude longer (see Fig. 1) than the test that would be calculated externally by a model-based automated test pattern generator (ATPG). In general, PRPG fault coverage trend is characterized by such peculiarities like fast initial growth and too long time to complete. Figure 1 illustrates this fact clearly showing two sections of PRPG curve. The slow growth section of PRPG curve in Figure 1 is mostly caused by existence of hard-to-test faults (HTTF) which are usually very hard to handle by PRPG-based methods (see Fig. 2.a)

As the result, there are many works that target improvement of PRPG efficiency. A big portion of research is devoted to study of alternative PRPG types [2,3] that have better saturation properties compared to the one of LFSR.

Much more gain is provided, however, when combining PRPG and ATPG patterns together. For example, one can get much shorter test sequence if he applies ATPG patterns right after the breakpoint between fast and slow sections of PRPG curve (Fig. 1 and Fig. 2.b). In case of self-testing, these patterns have to be stored in memory (Hybrid BIST) [4]. Hence, the more ATPG patterns we use, the bigger the memory overhead is.

This way of combining PRPG and ATPG patterns is not the only one possible. For instance, the Bit-Flipping BIST method adds extra circuitry to PRPG outputs. This circuitry modifies selected bits of selected PRPG patterns in such a way that these modified patterns become equivalent to ATPG patterns [5] (Fig. 2.c). However, the size of the bit-flipping controller represents a serious drawback of this method. Sometimes it occupies up to 30-40% of the circuit under test (CUT) area.

Another approach, called Reseeding, is considered as a promising one [6]. It allows for generation of several PRPG sequences where each one is optimized for covering a certain portion of HTTFs (Fig. 2.d). In terms of the test time, hardware cost, and fault coverage, the efficiency of this method is very close to the one of the Hybrid BIST.

This paper proposes a novel solution where ATPG patterns are used to form initial PRPG sequence. In this sequence, HTTF patterns are initially placed as close to each other as possible (Fig. 2.e). Then a special polynomial is calculated so that this sequence becomes feasible. In its pure form, this approach does not require nor additional memory for test patterns neither extra circuitry, but still it improves HTTF coverage and test length of PRPG sequence. For achieving extreme results, this method can be even combined together with all the methods mentioned above.

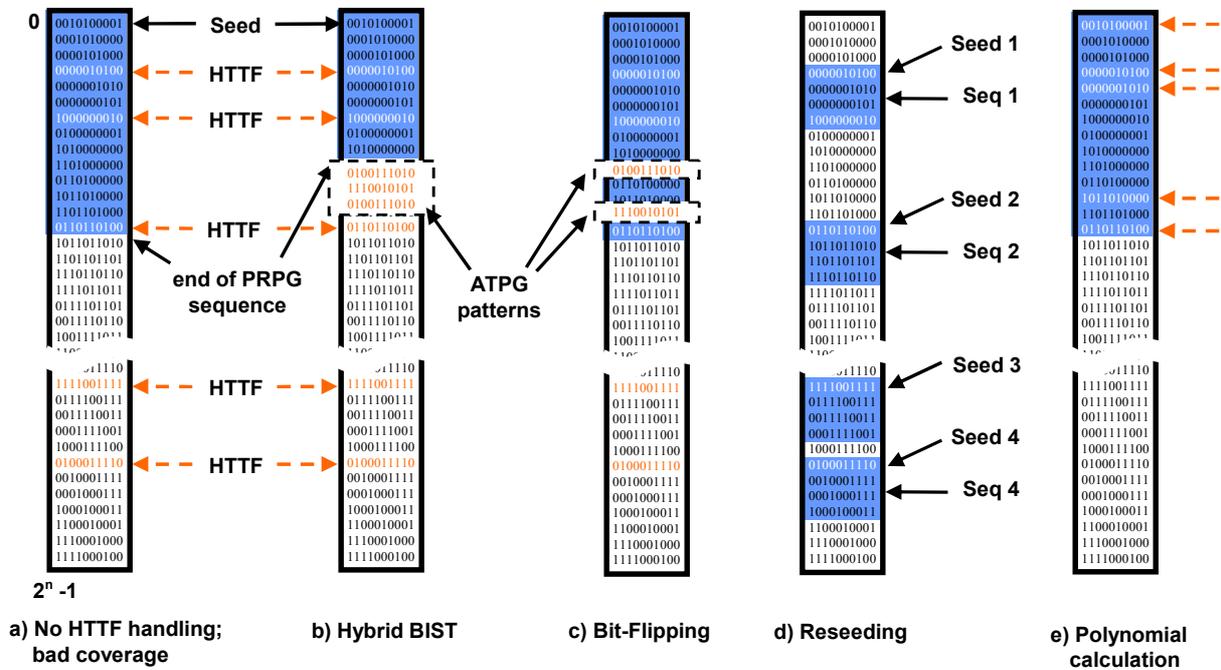


Figure 2. PRPG sequence and hard-to-test-fault (HTTF) coverage handling

## DESCRIPTION OF THE METHOD

In this section we will first look deeper into LFSR structure and properties and then describe the main idea behind the method of polynomial calculation.

In Figure 3, a common internal structure of LFSR is given. It consists of  $D$  flip-flops connected in series and feedback loops collected by an XOR gate. This forms a simple shift register with a special kind of feedback. The presence or absence of the feedback loops is described by a so-called *generator polynomial*. The state of the LFSR at the beginning of test generation is determined by its initial state parameter called *seed*.

The main useful property of LFSR circuits is such that if clocked repeatedly, they go through a fixed sequence of unique states, which has a number of explicit properties of randomness and can be used, therefore, as a TPG in a BIST scheme [7]. The maximum number of such unique states is  $(2^n - 1)$ , where  $n$  is the length of the LFSR (i.e. no. of flip-flops). However, the actual length of this sequence depends on selected polynomial and seed. Figure 3 shows a configured LFSR and the sequence it generates. One can see that the sequence has 4 unique patterns only and the grey patterns are just a repetition of black ones. Hence, polynomial and seed have direct influence to the resulting test quality and, therefore, they play an important role in TPG. In most of modern approaches, a fully configured LFSR is used. Its configuration is based on a

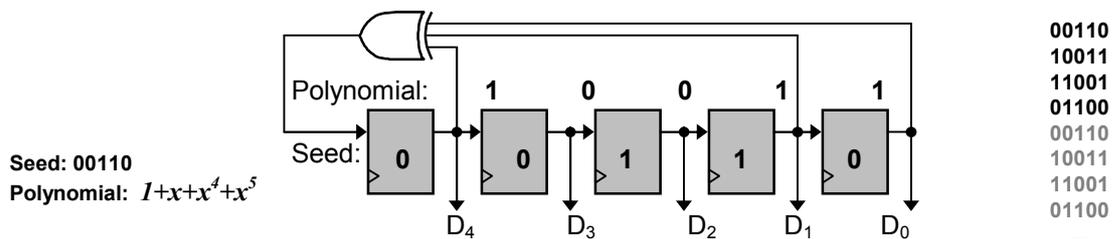


Figure 3. Internal structure of LFSR; the sequence it generates

*primitive polynomial* – a polynomial that guarantees  $(2^n - 1)$  unique states.

In our work, we neglect this common practice due to following reasons. At first, a longer LFSR with a big number of flip-flops can generate enough unique patterns to fully test a CUT even if it is not based on a primitive polynomial. Secondly, our method allows calculating a good polynomial that targets most of HTTFs by a possibly shorter sequence.

The overall strategy of the proposed method is the following:

1. Define set of HTTFs;
2. Select ATPG patterns for given HTTFs;
3. Form a continuous sequence;
4. Calculate polynomial;
5. Let configured LFSR run until all the rest faults (non-HTTF) covered.

The first and the second steps can be done in many different ways. The choice of a particular method depends on many factors and it is, therefore, left behind the scope of the current paper. Both steps are preparational for the main phase, that is calculation of sequence and polynomial. However, the efficiency of the proposed method depends dramatically on the choice of proper HTTFs and the way the corresponding ATPG patterns were generated or selected. Selection of most difficult HTTFs and usage of patterns with don't cares is highly recommended. The method limitations with respect to these facts are further considered in the next section.

The novelty of the method mainly lies in the next two steps of the procedure: forming a continuous feasible LFSR sequence and calculation of the corresponding polynomial.

The former task means adaptation of existing patterns into a form that in principle can be generated by an LFSR. We will look at this task deeper.

If LFSR of size  $n$  generates a sequence of  $m$  states, then this sequence can be represented as a binary  $m$ -by- $n$  Toeplitz matrix  $A$  with repeating entries  $a_i$ , where  $1 \leq i \leq k$  and  $k = n + m - 1$ .

$$A = \begin{bmatrix} a_n & a_{n-1} & \dots & a_3 & a_2 & a_1 \\ a_{n+1} & a_n & \dots & a_4 & a_3 & a_2 \\ a_{n+2} & a_{n+1} & \dots & a_5 & a_4 & a_3 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n+m-1} & a_{n+m-2} & \dots & a_{m+2} & a_{m+1} & a_m \end{bmatrix}$$

The LFSR seed is represented by the first row of the matrix  $(a_n \dots a_2 a_1)$ . The polynomial is not directly visible but it can be calculated by solving a system of linear equations. This task is solved in the next step of our procedure.

Let us denote LFSR feedbacks by vector  $(x_1, x_2 \dots x_n)$ , where  $x_j = 1$  if the corresponding feedback exists, otherwise  $x_j = 0$ . From the definition of the LFSR structure, we can derive a system of  $m-1$  linear equations that describes the LFSR operation.

$$\begin{aligned} a_n x_1 \oplus a_{n-1} x_2 \dots a_2 x_{n-1} \oplus a_1 x_n &= a_{n+1} \\ a_{n+1} x_1 \oplus a_n x_2 \dots a_3 x_{n-1} \oplus a_2 x_n &= a_{n+2} \\ a_{n+2} x_1 \oplus a_{n+1} x_2 \dots a_4 x_{n-1} \oplus a_3 x_n &= a_{n+3} \\ \dots & \dots \\ a_{n+m-2} x_1 \oplus a_{n+m-3} x_2 \dots a_m x_{n-1} \oplus a_{m-1} x_n &= a_{n+m-1} \end{aligned}$$

This set of equations can be unambiguously constructed from the matrix  $A$ . In other words, if we have a sequence of LFSR states of sufficient length (in general:  $m \geq n$ ) we can unambiguously calculate the generator polynomial of that LFSR by solving the corresponding system of linear equations. Despite the fact that the system is based on modulo-2 operations, we still can use Gaussian elimination to solve it. For faster results, special techniques developed for Toeplitz matrices can be used, e.g. Levinson

recursion, which complexity is  $O(n^2)$ .

In practice, in order to construct LFSR sequence from given ATPG patterns, one has to put them into a bit-stream of length  $k$ :  $(a_k a_{k-1} \dots a_1)$ , then construct a system of linear equations and solve it in order to find variables  $(x_1, x_2 \dots x_n)$  that represent the target LFSR polynomial.

Let us consider the small example in Figure 4. The leftmost part of this figure (4.a) shows the ATPG patterns to be processed. Here we assume that these patterns were generated for selected HTTFs (steps 1 and 2). Due to don't care bits, these patterns can be combined into a bit-stream of length 11 (Fig 4.b) and being then consequently transformed into a 7-pattern LFSR sequence and a 6-by-5 matrix in Figure 4.d. The solved system of linear equations and the hardware implementation of this solution are given in Figure 4.e and Figure 5 correspondingly.

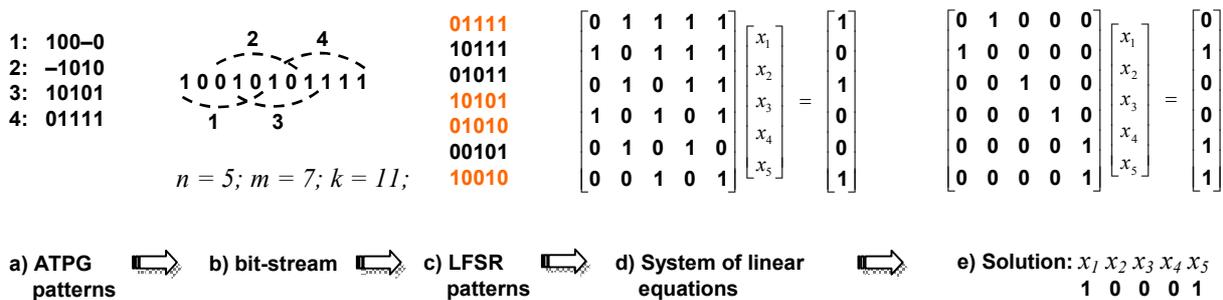


Figure 4. Solution flow: from ATPG to LFSR

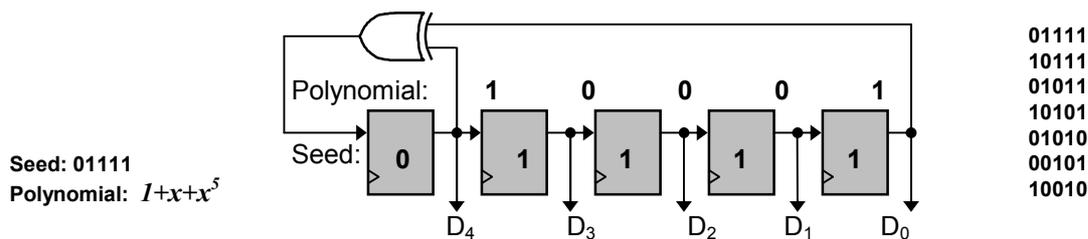


Figure 5. Hardware implementation of the solution from Figure 4

## LIMITATIONS AND APPLICABILITY OF THE METHOD

Due to general properties of linear equation systems, a solution exists if rank of the extended matrix is equal to the rank of the main matrix. In our case, if this rank is also equals  $n$ , a single solution exists. In practice, it doesn't make any problem if rank is less than  $n$ . Then, one just has to choose a solution with lower hardware cost. The problem arises if the system of linear equations is inconsistent. This might happen if  $m$  is larger than  $n$ . Statistically, the expectation value for the border between system consistency and inconsistency is close to  $n$ . As the result, the polynomial calculation problem will reliably have a solution for input data that is compactable to a bit-stream of length  $k \leq 2n$ . If this is not the case, one either has to increase  $n$  or change the ATPG vectors. Since LFSR length increase, in general, is not desirable, a careful selection of target HTTF set and existence of don't-care bits in ATPG vectors are very important.

The 7-pattern LFSR sequence in Figure 4.c contains 4 ATPG vectors and 3 extra ones. Since in practice a test has to cover much more faults than just some HTTFs, the

existence of these extra vectors is beneficial for they are covering additional faults. In general, LFSR can be let generating even a longer test sequence for the sake of getting higher fault coverage. This fact can be also used in cases when the system of equations becomes inconsistent. Then such a solution should be found, that satisfies the largest part of the system. HTTFs left aside as the result of this simplification will be covered later in the LFSR sequence. Such an approach will still help to reduce the overall test time since it puts at least some of HTTF test vectors into the beginning of the test sequence. There are other ways to cope with this issue. The simplest one is to split the system of linear equations into several consistent subsystems of length  $h \leq n$ . Each such subsystem has to be solved then separately producing a separate polynomial. This would result in more complicated multiple-polynomial LFSR hardware that, similarly to reseeding, has to change its configuration at each  $h^{\text{th}}$  step [6]. Instead of changing the state it would change the polynomial.

Another solution would be to correct the inconsistent equations using a correcting vector matrix. In practice, such a correcting vector can be implemented similarly to the bit-flipping hardware. The difference is such that in the proposed method a single bit needs to be flipped instead of flipping arbitrary bits in the test sequence as in case of [5].

The proposed method of polynomial calculation can be especially beneficial when combined with other techniques like e.g. reseeding or hybrid BIST.

## CONCLUSIONS

This paper presents a novel theoretical approach to LFSR polynomial calculation for the purpose of test time reduction through fitting ATPG vectors into LFSR test sequence. The method allows to improve fault coverage characteristics of LFSR-based testing without using additional hardware resources. As the result, the method can be used together with other existing approaches of combining LFSR sequence with ATPG patterns (reseeding, bit-flipping, multiple-polynomial, etc.) for improving their efficiency.

### Acknowledgement:

This work has been supported by the Estonian Science Foundation grants 5910 and 7068, Estonian IT Foundation grant 07-03-00-05, and Enterprise Estonia funded ELIKO project.

### References:

- [1] *The International Technology Roadmap for Semiconductors, 2006 Update: Test and Test Equipment*. <http://public.itrs.net/>
- [2] G. Mrugalski, J. Rajski, J. Tyszer, "Cellular Automata-Based Test Pattern Generators with Phase Shifters," in *IEEE Trans. on CAD/ICAS*, vol. 19, no. 8, pp. 878–893, 2000.
- [3] S. Chidambaram, D. Kagaris, D. K. Pradhan, "A Comparative Study of CA with Phase Shifters and GLFSRs," in *Proc. IEEE International Test Conference (ITC)*, Austin, USA, Nov 8-10, 2005, pp. 926-935.
- [4] G. Jervan, Z. Peng, R. Ubar, O. Korelina, "An improved estimation methodology for hybrid BIST cost calculation," in *Proc. Norchip Conference*, Oslo, Norway, Nov 8-9, 2004, pp. 297 – 300,
- [5] H. Wunderlich, G. Kiefer, "Bit-Flipping BIST," in *Proc. Int. Conf. on Computer-Aided Design*, Nov 10-14, 1996, pp. 337-343.
- [6] A. A. Al-Yamani, S. Mitra, E. J. McCluskey, "BIST reseeding with very few seeds," in *Proc. 21<sup>st</sup> IEEE VLSI Test Symposium (VTS)*, Napa Valley, USA, Apr./May 2003, pp. 69–74.
- [7] M. Bushnell, V. Agrawal, *Essentials of Electronic Testing for Digital Memory & Mixed Signal VLSI Circuits*, Kluwer Academic Publishers, 2000, 712 p.

### Author:

Dr. Artur Jutman  
Tallinn University of Technology  
Raja 15, 12618 Tallinn, ESTONIA  
Phone: +372 620 22 63  
Fax: +372 620 22 53  
E-mail: [artur@pld.ttu.ee](mailto:artur@pld.ttu.ee)

Maik Rosenberger, Martin J. Schaub, Susanne C. N. Töpfer, Gerhard Linß

## **Investigation of efficient strain measurements at smallest areas applying the time to digital (TDC) principle**

### **ABSTRACT**

*Mechanical stresses can significantly exert a large negative influence on stability, accuracy, maximum allowable forces, operational reliability and durability. Such irregular forces occur also at manual electronic internal and external measuring gauges, for example so-called Quicktester. These measuring instruments are used for fast and precise measurements of inside flutes and inside diameters, as well as for thickness measurements. Through different operation positions and irregular loads, applied on the measuring gauge arm by the user, the measuring arm of the instrument is deformed within the  $\mu\text{m}$  range. Thus, the precise determination of the deflection is a decisive advantage for minimising measuring deviations.*

*The paper at hand presents a novel method for strain measurements exemplified at a Quicktester measuring gauge. The method is based on the application of miniature strain gauges and on the signal analysis with the time to digital principle (TDC). Typically the application of miniature strain gauges is characterised by minimal output signals. In addition the cost criterion and the energy efficiency play an important role for mobile measuring instruments. Thus, an application circuit based on new time to digital (TDC) procedures according to the Picostrain method was developed and successfully applied with miniature strain gauges. The paper presents experimental results regarding temperature characteristic, linearity and resolution. Deflections within the  $\mu\text{m}$  range could be stably measured.*

### **INTRODUCTION**

Each technical system is subjected to most different environmental conditions and loads during its whole life cycle. In most cases these loads cause mechanical stress within the technical system. Mechanical stress has a negative influence on several properties, e.g. mechanical strength, accuracy and maximal achievable life span.

Up to now it was common to investigate mechanical stresses during the development period and to try to eliminate any influence of them on the system. This implies increased safety factors for the mechanics, oversize of components and increased development and manufacturing costs. In order to minimise these disadvantages it is inevitably important to record data about the mechanical stresses during the whole product life cycle [1]. This applies similarly to large machines as well as to small mechanical measuring instruments such as an electromechanical measuring gauge by the company Kröplin for example. Aim is primarily the reduction of measuring uncertainty for these highly precise manual gauges but also efficient material usage, increased quality and product safety. First tests with strain gauges have been performed by A. C. Ruge in 1938. Thereby, the fundamental of this measurement method was the change of the electrical resistance of metals under strain or under compression [2].

The electromechanical, quick measuring gauges possess a fixed and a movable measuring arm. The movable measuring arm is pressed on the measuring object with a defined spring force. The fixed arm is pressed on the measuring object by the operator. The measuring arm is deformed by several µm due to the contact load. Thus, it is necessary to measure the load of the fixed measuring arm during each measurement. Special challenges for the integration of a suitable measuring unit into such an electromechanical, quick measuring gauge are minimum power consumption, for the measuring gauge is battery powered, and minimum installation place for the measuring sensors and its electronic circuitry.

**FUNDAMENTALS OF STRAIN MEASUREMENTS**

Measurements of mechanical stress using strain gauges are widely deployed in industrial applications. Typical problems, when applying strain gauges, are nonlinear material properties of the adhesives, temperature drift of the strain gauges itself and irregular behaviour of the material exposed to stress. According to Thomson and Wheatstone the basic resistance of a material is:

$$R_0 = \frac{\rho \cdot l}{A_{cross-section}} \quad \text{where:} \quad \begin{array}{l} \rho - \text{specific resistance} \\ l - \text{conductor length} \\ A - \text{conductor cross-section.} \end{array} \quad (1)$$

The strain dependent change in resistance  $R$  is decisive for measurements with strain gauges, see Eq. 2.

$$R = R_0 \cdot (1 + k \cdot \varepsilon) \quad \text{where:} \quad \begin{array}{l} R_0 - \text{basic resistance} \\ k - \text{strain sensitivity} \\ \varepsilon - \text{change in length (strain)}. \end{array} \quad (2)$$

Aim of the application is the maximisation of the sensitivity of the strain sensitive elements. Thereby  $R_0$  and  $k$  are constants. The stress dependent strain  $\varepsilon$  changes at mechanical loading. As far as constructively possible the strain  $\varepsilon$  at the measuring area should be maximal. The factor  $k$  of metallic strain gauges amounts to 2.05 (constantan). Basically a large factor  $k$  is linked to a high sensitivity. However, specifically for semiconductor strain gauges a large factor  $k$  also entails an unfavourable temperature characteristic, see [3] page 43. Often deployed measuring analysis methods utilising measuring bridges are detailed in [4].

A new method for the analysis of a strain gauge measuring bridges has been investigated with the help of TDCs, whereby the Picostrain principle was applied. This method was the experimental basis for further investigations. The discharge time of a capacitor is measured with a TDC. The change in discharge time of the capacitor is proportional to the strain [5]. The capacitor is discharged by the resistance of a strain gauge down to an arbitrary switching threshold. The discharge times are compared and the related strains are calculated [6].

## **PICOSTRAIN TDC TECHNIQUES FOR SMALL STRAIN GAUGES**

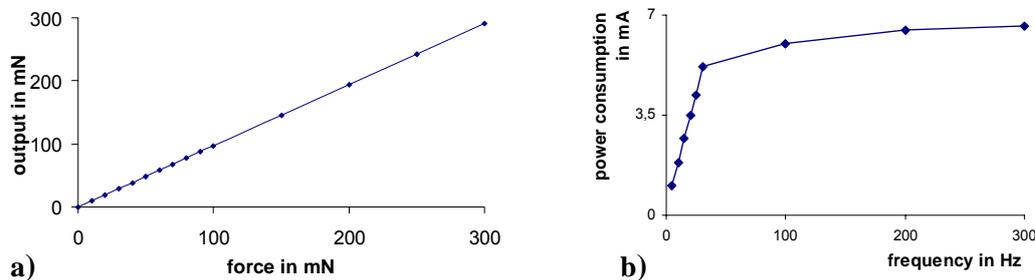
The experiments and the comparison of different measuring amplifier including an analogue-to-digital converter (ADC) with a measuring amplifier based on the TDC principle have proven, that the ACAM Picostrain system is the best choice. It is an integrated system for the analysis of the strain gauges. It requires no further high-consumption components, such as amplifiers, and is characterised by a low power consumption. A special advantage lies in the fact that no Wheatstone measuring bridge with 4 strain gauges is necessary. Two strain gauges suffice in order to attain maximum accuracy. Tests with four strain gauges did not result in a gain of accuracy. Investigations in suitable miniature strain gauges yielded the strain gauge 1-LY11-3/350 by the company HBM as most suitable [7] (Fig. 1a). This strain gauge has a basic resistance of 350 ohms which is a prerequisite for the application of the TDC

principle. Fig. 1b illustrates a mounted strain gauge on a measuring arm with an active measuring area of app. 4.5 mm<sup>2</sup>. Fig. 1c shows the general measuring setup.



**Fig. 1:** a) Strain gauge 1-LY11-3/350, b) Measuring area of the strain gauge c) Test setup

Aim of the experiments was the determination of the smallest measurable strain signal. Strains caused by a contact pressure of 3N exerted by the operator amount to up to 42  $\mu\text{m/m}$  at one measuring arm. Using the Picostrain amplifier enables measuring frequencies of up to 50 kHz [8]. The measuring software of the company ACAM was utilised for the analysis of the measurements at the PC. For the experiments two strain gauges were mounted on the top and bottom side of different measuring arms. Afterwards defined loads were applied (Fig. 2). The maximum load of the measuring arms equals 3 N. Each larger load is an inappropriate operation of the measuring device.

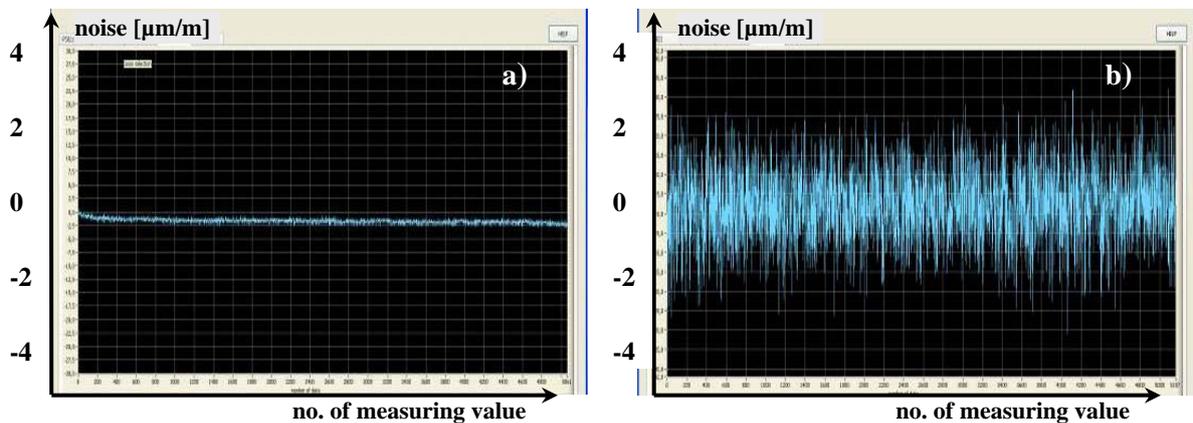


**Fig. 2:** a) Linearity of the measurement signal b) Power consumption vs. frequency of the TDC circuit

After the experiments investigating the fundamentals the overall system was optimised. The TDC principle offers the opportunity to define most different parameters before starting the measurement. For the integration into a battery powered measuring device an optimal compromise between measuring uncertainty, measuring speed and power consumption had to be identified. A larger measuring frequency is always linked to a larger power consumption (Fig. 2b). A standard delta-sigma-converter which is a special type of an ADC was applied for comparison measurements. In comparison to the TDC method the power consumption of this standard ADC was seven times larger.

Furthermore the TDC principle provides a number of filter functions which had to be adapted for the application at the Quicktester. With the help of test plans an effective resolution of the strain of 15 bits was achieved. Further experiments were focussed on the electrical cable connections between the strain gauges and the TDC Picostrain chip. Thereby screened and firmly fixed coaxial cables turned out to be the best solution.

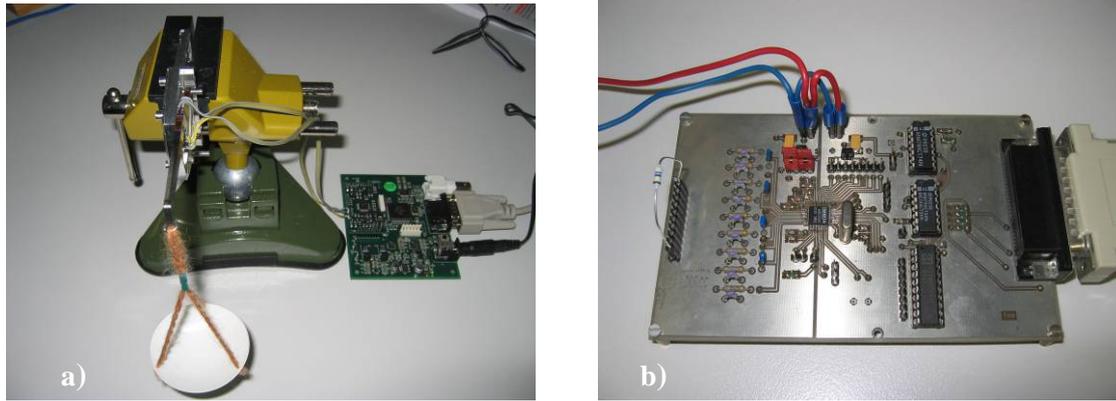
A stable voltage supply is decisive for enabling the highly precise time measurements with the TDC method. Therefore tests with different voltage supplies like a step-up-controller, a drop-down-controllers and various batteries were performed. The voltage supply based on a drop-down-controller with a battery delivered very good measuring results (Fig. 3a). On contrast the base noise is amplified by factor five when a step-up-controller is utilised (Fig. 3b).



**Fig. 3:** (a) Base noise of the battery  $\pm 0.07 \mu\text{m/m}$  b) Base noise with a step-up-controller  $\pm 3.5 \mu\text{m/m}$

## EXPERIMENTAL RESULTS

For the integration into Quicktesters by Kröplin an extremely low energy strain measuring system has been developed. The optimal position of miniature strain gauges, namely 1-LY11-350 by the company HBM, on the measuring arms were calculated and tested (Fig. 4a). A strain resolution of  $0.14 \mu\text{m/m}$  has been attained with the Picostrain measuring bridge. A minimal power consumption of 5.14 mA at a measuring frequency of 30 Hz poses an ideal compromise between these both parameters. The measuring results prove that the chosen measuring setup is optimal for the integration into Quicktesters. All relevant hard- and software parameters for the integration have been considered. A long battery life span and a high resolution of the measuring arm deformation is enabled.



**Fig. 4:** (a) Test setup for linearity measurements (b) Deployed standard ADC

## CONCLUSION

The experiments have proven that the deployment of the TDC principle for strain measurement is ideal for electromechanical measuring gauges such as the Quicktester. This method enables highly precise strain measurements at minimum power consumption. Furthermore the required installation space is minimal and no expensive and high-stable reference voltage supply and no additional measuring amplifiers are required. Comparisons with standard ADC solutions (Fig. 4b) yield the result that these are more cost-intensive and consume more power. When using the TDC principle special care must be put on the cable connection. Finally, the appropriate type of voltage supply (drop-down controller) must be used.

## References

- [1] Linß G.: Qualitätsmanagement für Ingenieure. 2. Auflage, Fachbuchverlag Leipzig, 2005.
- [2] Keil S.: Beanspruchungsermittlung mit Dehnungsmessstreifen. CUNEUS Verlag, Zwingenberg a.d. Bergstraße, 1995.
- [3] Hannah R.L., Reed S.E.: Strain Gage Users Handbook. Elsevier Applied Science, Cambridge, 1992.
- [4] Window A.L.: Strain Gauge Technology. Elsevier Applied Science, Essex, 1992.
- [5] Kirianaki N.V., Yurish S.Y., Shpak N.O., Deynega V.P.: Data Acquisition and Signal Processing for Smart Sensors. John Wiley & Sons Ltd, 2002.
- [6] ACAM Messelectronic GmbH: Applikationsschrift 11. Stutensee-Blankenloch, 2004.
- [7] Hoffmann K.: Eine Einführung in die Technik des Messens mit Dehnungsmessstreifen. Hottinger Baldwin Messtechnik GmbH, Darmstadt, 1987.
- [8] ACAM Messelectronic GmbH: Datasheet PS021. Stutensee-Blankenloch, 2006.

## Authors:

Dipl.-Ing. Maik Rosenberger, Dipl.-Ing. Martin J. Schaub,  
Dipl.-Wirtsch.-Ing. Susanne C.N. Töpfer, Univ.-Prof. Dr.-Ing. habil. Gerhard Linß

Technische Universität Ilmenau  
Quality Assurance Department  
P.O. Box 100 565  
98684 Ilmenau, Germany  
Phone: +49 3677 693941  
Fax: +49 3677 693823  
Email: maik.rosenberger@tu-ilmenau.de





Joerg Meyer / Raymund Espiritu / James Earthman

## Virtual Bone Density Measurement for Dental Implants

### MEASURING OSSEOINTEGRATION OF TITANIUM IMPLANTS

The term *osseointegration* describes the integration of a dental implant into the surrounding bone material. A method for virtual bone density measurement has been developed, which is based on digital image processing of CT scan data. In a cadaver study, a CT scan of the maxilla has been obtained, and two implant sites with somewhat complementary properties have been selected as objects of study.

The visualization procedure consists of 2-D cross-sectional CT imaging, 3-D gradient-based hardware-accelerated volume rendering using 3-D texture mapping, implant site extraction using 3-D selection of a 2-D cross-sectional, tri-linearly interpolated 2-D image, and computation of a bone density profile and line integral along the implant. By visually displaying the effects of variations in implant size, location of the implant site, bone density, and osseointegration, conclusions can be drawn for optimal placement and anchoring of dental implants, eventually leading to more stability, higher durability, and an increased lifetime of the implanted tooth.

#### Introduction

Virtual bone density measurement uses three-dimensional imagery from a CT scan of the specimen to determine X-ray absorption in tissues, which is directly correlated to bone density. In our method, an arbitrarily located two-dimensional cross-section is extracted from the 3-D scan, which represents an anatomical feature, in our case a dental implant. After choosing the location, a measurement is taken on the 2-D image. In the given study, the measurement is an indicator for the degree of osseointegration of a dental implant.

#### Background and Significance

For healthy teeth, the percussive energy produced by mastication processes is attenuated by the periodontal ligament at the bone-tooth interface. This ligament,

however, is lost when the natural tooth is replaced with an implant for reasons such as disease or irreparable damage. The implant transmits the percussive forces directly into the bone at the material-bone interface.

In this study, two different dental implant sites (figure 1) were evaluated using a cross-sectional image based method for the computation of bone density as a function of distance from the implant apex. Each cross-section is centered on the longitudinal axis of the implant reaching from the buccal to the lingual side. A sequence of density values along one side of the implant is called a profile. The average of such a density profile represents a line integral, which is an indicator for the degree of osseointegration.

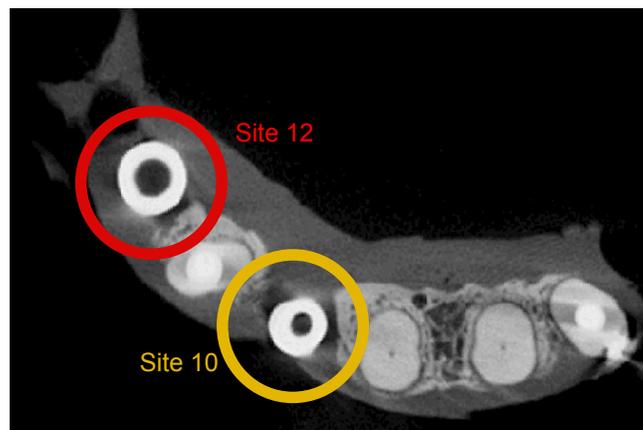


Fig. 1. Selected implant sites 10 and 12

### **Related Work**

Osseointegration is the permanent incorporation of an implant into bone. This direct and functional connection cannot be separated without fracture. Osseointegration is an important process, along with bone healing, that occurs after dental implants are placed and covered, after which the implants are uncovered and connected to an abutment to allow for mechanical loading [1].

A sufficient amount of loading is needed to strengthen bone through bone formation around the wound site. If this requirement is not met or significantly exceeded, osteoclastic activity will commence and the bone subsequently removed from the site [2, 3]. In therapeutic loading for dental implants, the implant design must replace the function of the periodontal ligament that is lost upon prosthetic placement by transmitting stress waves through the tissue near the natural level [4]. Also, the surrounding bone tissue must be stable to secure the implant prior to loading.

Given the importance of monitoring the development of bone that provides support, this paper suggests a virtual method of profiling the bone density gradient that surrounds

the implant.

The visualization methods include 2-D cross-sectional CT imaging, 3-D gradient-based [6], hardware-accelerated volume rendering using 3-D texture mapping [7, 8, 9], and implant site extraction using 3-D selection of a 2-D cross-sectional, tri-linearly interpolated 2-D image [10, 11].

## 2-D Measurement of Osseointegration

In order to measure bone density in the proximity of the implant site, a two-dimensional cross-section was extracted from the volumetric grid using tri-linear interpolation (figure 2). The cross-section intersects the longitudinal axis of the implant and reaches from the buccal to the lingual side. This way, a standardized coordinate system was defined for bone density measurements.

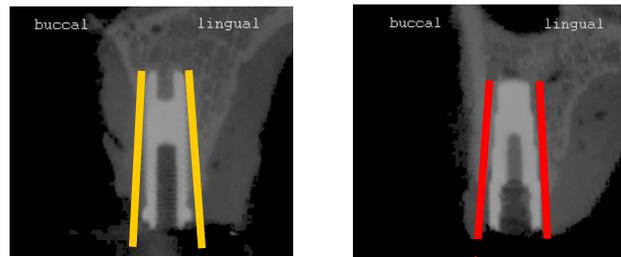


Fig. 2. Extracted cross-sectional image of implant sites 10 and 12 from 3-D scan

Using Bresenham's line algorithm, a straight line was drawn on both sides of the implants to collect density information near the implant. The line originates at the same vertical coordinate as the apex of the implant and maintains a predefined horizontal distance to the implant. This line was then used to collect the data for the bone density profile and to compute a line integral, i.e., a single number that is characteristic for the bone density of a particular implant and side (buccal or lingual).

In order to normalize the line integral and to make it independent of the size of the implant, only pixel values inside the bone and gums were included in the computation. Pixels with values characteristic for air (background pixels) were ignored. Also, if the line accidentally cut through the implant, those values would have been ignored as well. The sum of the pixel values was divided by the number of active pixels along the line in order to make the line integral independent from the size of the implant.

Typical profiles for implant sites 10 and 12 (buccal and lingual sides) are shown in figure 3. The horizontal axis represents the distance from the apex of the implant, and the vertical axis shows the intensity. The average, i.e. the value of the line integral, is shown as a horizontal line with a percentage value.

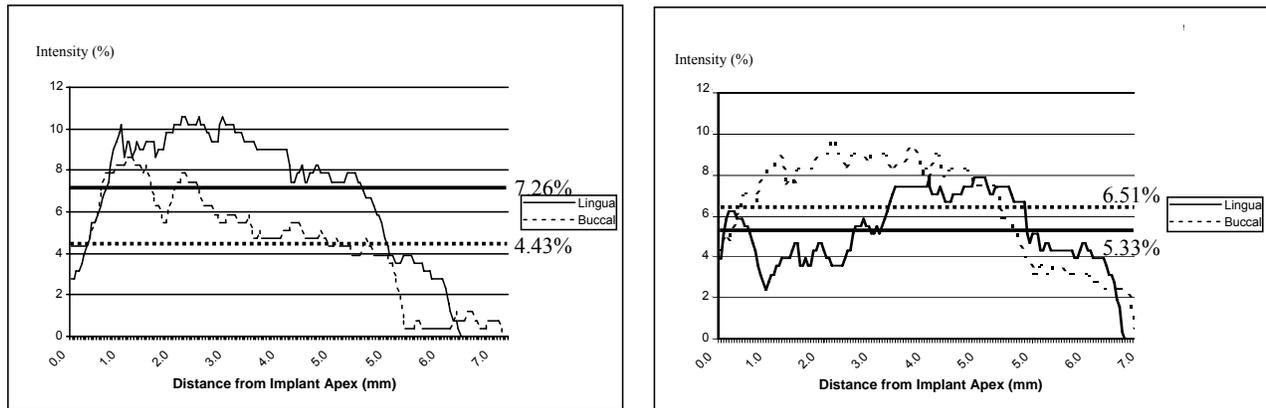


Fig. 3. Bone density (intensity) for the lingual and buccal sides of implant sites 10 and 12.

Please note that in implant site 10 the lingual side appears to be stronger, whereas in implant site 12 the buccal side appears to be stronger. This could have been caused by various effects, including aging, placement of the implant, and mechanical stress.

### Visualization

Instead of presenting the data in a diagram, the density profiles can be visualized directly on the 2-D cross-section (figure 4). The orientation of the profiles has been rotated to reflect the mutual support of the implant from the bone on the buccal and lingual side.

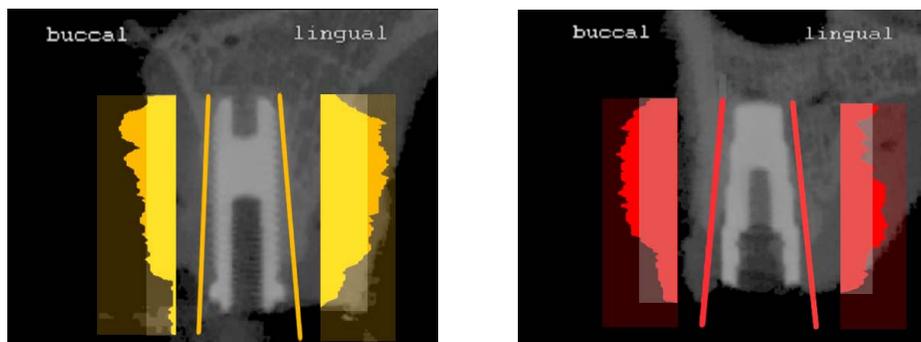


Fig. 4. Bone density profiles for the lingual and buccal sides of implant sites 10 (left) and 12 (right). The width of the highlighted areas shows the value of the line integral as a single

measure of osseointegration per site and per side (buccal or lingual). When shown in context, the profiles show the amount of horizontal support for each vertical position. The images clearly indicate that the support is usually better near the apex of the implant. The two examples also indicate that there may be differences in the buccal and lingual support of the implant.

### **Discussion of Results**

The following observations were made from these visualizations. For both implant sites the bone density values followed a general trend of an initial rise, peak plateau, and fall as distance increased from implant apex (figure 4). Early rise and peak in bone density was generally seen within 3 mm from the implant apex. This range of increased bone density may be explained by the presence of a high-density layer of cortical bone at the dorsal bone/implant interface [4].

It is important to note the exception to this trend in the data collected for the lingual side of site 12 as shown in figure 4 (right). Another important observation to note is the lack of prominent plateaus in the curves generated for the buccal and lingual sides of implant site 10 as shown in figure 4 (left). These occurrences were validated with a mechanical study (percussion testing) that was conducted to verify the results from the virtual study [4].

Following the peak plateaus of bone density for both implant sites 10 and 12, which correspond to the lateral incisor, and 1<sup>st</sup> bicuspid or premolar, respectively, the buccal regions generally displayed lower trends of decreasing bone density than the lingual sides. This incidence may be expected since the density of bone mass of the buccal cortex within the incisive and premolar region is known to be lower than the density of the corresponding lingual cortex [5].

### **Conclusions**

We presented a method for virtual computation of bone density for dental implants. Small sample images for each implant were extracted from a three-dimensional CT scan. Bresenham's line algorithm was implemented to collect density values along a designated distance (mm) from the apex of the implant. The density profiles were then mapped onto the 2-D sample image and shown together with the line integral as a general measure for osseointegration. Overall, the results show a good correlation between the virtual study and the mechanical test study [4].

From data collected by virtual and mechanical testing, it was evident that the bone density profile of the implant is site-specific and/or determined by the extent to which bone develops around the implant. The results obtained by the present study may serve as a platform for the future examination of the process of bone healing and development in vivo.

#### **Acknowledgements:**

The authors would like to thank the American Equilibration Society for their generous support. The authors wish to also thank Dr. Eric Herbranson for his assistance in acquiring the CT data used in the present work.

#### **References:**

- [1] R. Gapski, H.-L. Wang, P. Mascarenhas, N.P. Lang, Critical review of immediate implant loading. *Clinical Oral Implants Research* 14, p. 515, 2003.
- [2] M.R. Forwood, C.H. Turner, Skeletal adaptations to mechanical usage: results from tibial loading studies in rats. *Bone* 17, p. 1975, 1995.
- [3] A.G. Robling, K.M. Duijvelaar, J.V. Geever, N. Ohashi, C.H. Turner, Modulation of longitudinal and appositional bone growth in the rat ulna by applied mechanical force. *Bone* 29, p. 105, 2001.
- [4] L.R. VanSchoiack, J.C. Wu, C.G. Sheets, J.C. Earthman, Effect of bone density on the damping behavior of dental implants: An in vitro method. *Materials Science and Engineering* 26 ,pp. 1307-1311, 2006.
- [5] N. Wowern, Variations in bone mass and bone activity within the Mandible. *Calcified Tissue International* 21, pp. 397-404, 1975.
- [6] J. Kniss, G. Kindlmann and C. Hansen. Interactive Volume Rendering Using Multi-Dimensional Transfer Functions and Direct Manipulation Widgets. *Proceedings of IEEE Visualization 2001*, pp. 255-262, 2001.
- [7] F. Dachille, K. Kreeger, B. Chen, I. Bitter, and A. Kaufman. High-Quality Volume Rendering Using Texture Mapping Hardware. *SIGGRAPH Eurographics Graphics Hardware Workshop*, pp. 69-76, 1998.
- [8] R. Westermann and T. Ertl. Efficiently using graphics hardware in volume rendering applications. *Computer Graphics (SIGGRAPH '98)*, 32(4), pp. 169-179, 1998.
- [9] M. Meissner, S. Guthe, and W. Strasser. Higher Quality Volume Rendering on PC Graphics Hardware, Wilhelm Schickard Institute for Computer Science, Graphical-Interactive Systems (WSI/GRIS), University of Tuebingen, 2001.
- [10] J. Meyer, S. Gelder, K. Kretschmer, K. Silkenbauumer, and H. Hagen. Interactive Visualization of Hybrid Medical Data Sets. *Proc. of WSCG '97*, Vol. 2, Pilsen, Czech Republic, pp. 371 - 380, 1997.
- [11] R. Sengupta, J. Meyer, and Z. Zhang. Hybrid Pipelining Approach to Image Alignment for Large-scale Brain Image Data, 7th IASTED International Conference on Computers, Graphics, and Imaging (CGIM 2004), Kauai, Hawaii, pp. 78-83, 2004.
- [12] P. Magne, U. Belsler, Rationalization of shape and related stress distribution in posterior teeth: A finite element study using nonlinear contact analysis. *Int. Journal on Periodontics and Restorative Dentistry*, 22, pp. 425-433, 2002.

#### **Authors:**

Joerg Meyer, Electrical Engineering & Computer Science  
Raymund Espiritu, Biological Science  
James C. Earthman, Chemical Engineering & Materials Science  
University of California, Irvine  
Henry Samueli School of Engineering  
644E Engineering Tower  
Irvine, CA 92697-2625  
U.S.A.  
Phone: +1-949-824-9321  
Fax: +1-949-824-3203  
E-mail: jmeyer@uci.edu

F. Erfurth / W.-D. Schmidt / B. Nyuyki / A. Scheibe / P. Saluz / D. Faßler

## Spectral Imaging Technology for Microarray Scanners

### INTRODUCTION

In biotechnology fluorescence analysis methods are a growing area of interest. Many techniques have been developed, to investigate biological functions on a molecular level. But conventional methods of microarray analysis are, although suggesting high sensitivity at the detection of genetic material, rather inaccurate, unreliable [1] and slow. Origin of these deficiencies are, e.g., procedural variations at the production of microarrays. Because of the low, unidimensional information content obtained with conventional scanners, the variations can not be identified and assigned to a single source. Here, spectral imaging (SI) technology allows advances through fast, two-dimensional detection of whole fluorescence emission spectra. The benefits of higher information content could be shown in experimental studies for different situations.

### SPECTRAL IMAGING PRINCIPLE

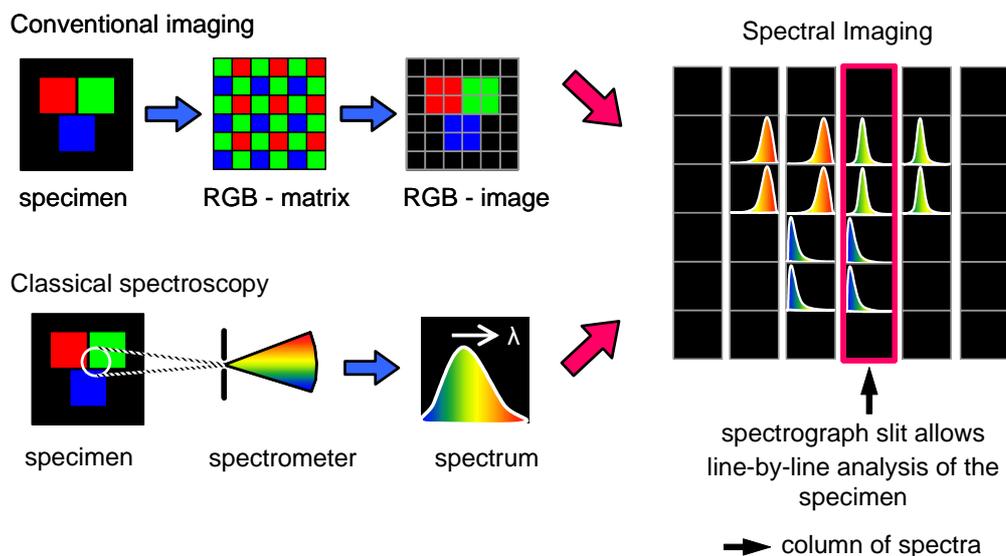


Fig. 1, Spectral imaging as a combination of imaging and spectroscopy

SI is a beneficial combination of optical spectroscopy and digital image processing. The technology originates from airborne remote sensing used by geologists. The emitted ra-

diation to be analyzed goes through an entrance slit and a dispersive element before it falls on the array detector. Here a spatial and a spectral dimension are recorded simultaneously (Fig.1). The third dimension is detected consecutively by scanning the sample. Because the number of spectral channels is in the range from tens to hundreds, our SI-system can be classified as hyperspectral.

## **MICROARRAY ANALYSIS**

The application of SI for microarray scanners allows probing more than one fluorescence label with a single laser excitation source. In gene technology, fluorescently labeled DNA samples are attached to a glass or plastic surface. Through specific binding, DNA probes are hybridized, e.g., to the fixed oligonucleotides. After excitation of the sample the fluorescent emission of different dyes yields biological information about the probes. Gene expression levels provide important information of the state of any organism. To measure it, the hybridized nucleic acids of the microarrays are scanned either point-by-point by a fluorescence microscope or by an imager. One significant drawback of the current technology is channel cross-talk, when the dye number increases [2,3]. Although microarrays are already widely used, solutions must be found to increase the sensitivity, accuracy and throughput. The application of multiple dyes can help with extensive studies. For gene expression experiments with a large number of specimens, including several microarrays, reference standards can be used for indirect comparison [4].

Microarrays of oligonucleotides with four different cyanine dyes Cy3, Cy3.5, Cy5, and Cy5.5 (MWG Biotech) were prepared by the Leibniz Institute for Natural Product Research and Infection Biology (HKI) in Jena and measured by a GenePix 4000B scanner (Molecular Devices) as a reference initially. Investigated, dye labeled oligonucleotides were spotted with dye concentrations ranging from 1 pmol/ $\mu$ l to 0.5 fmol/ $\mu$ l.

## **EXPERIMENTAL SETUP**

The experimental setup consists of three parts, excitation path, emission path and translation of the sample. The excitation is accomplished by two lasers, a frequency-doubled solid state laser (Soliton DLSOT-100) producing 100 mW of 532 nm radiation and a diode laser (Stocker Yale Lasiris Magnum 501L) with 250 mW at 635 nm in pulsed mode. The SI-system detects the emission within a line-shaped area. To illuminate only that region and to avoid unnecessary degradation of surrounding dye, the two lasers are

equipped with Powell lens line projection optics. This allows an excitation line with a width of less than 40  $\mu\text{m}$ , enhancing the resolution of the SI-system in scan direction.

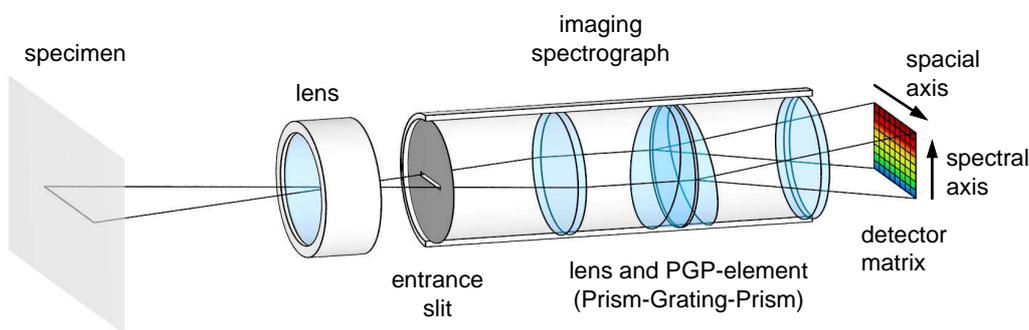


Fig. 2, Schematic diagram of the imaging spectrograph ImSpector (Specim / Finland)

The detection system comprises a lens with emission filter, spectrograph and camera. A macro lens (Schneider-Kreuznach Apo-CPN 4.0/45) is used to realize magnifications from 1:1 to 1:20. Two different emission filters (Semrock LP-01-532RS-25, LP02-647RU-25) suppress scattered light from the lasers. In Fig.2 the principle of the imaging spectrograph (Specim ImSpector V8E) is illustrated. The entrance slit with a length of 14.3 mm was chosen to be 80  $\mu\text{m}$  wide, which is the best compromise between wavelength resolution (6 nm) and adequate light intensity. The fluorescent light is collected by a highly sensitive cooled EMCCD-camera (Andor iXon DV887DCS-BV), capable of single photon detection. Typically, signal intensities of about 12,000 counts are achieved at an excitation energy of 300  $\mu\text{J}$ . Experiments showed, that the optimum conditions of irradiation power and camera integration time are at 3 mW and 100 ms.

The microarray, having size and shape of a microscope slide, is translated by a motorized translation stage for micrometer steps, perpendicular to the extent of the excitation and detection line. With one measurement, up to 512 spectra are detected simultaneously, having the potential of fast scans. The whole setup can be seen in Fig.3.

## COMPONENT ANALYSIS

Component analysis yields a clear discrimination of the fluorescent cyanine dyes, despite strongly overlapping emission spectra. Fluorescence signals of labeled biospots can be seen as linear superimposed. Therefore, data interpretation can be performed by component analysis, which determines the abundance of single spectral signatures [5,6]. The analysis of the whole spectral response of each biospot allows a separation of different dyes with neighbored fluorescence spectra, as is the case with multicolored spots.

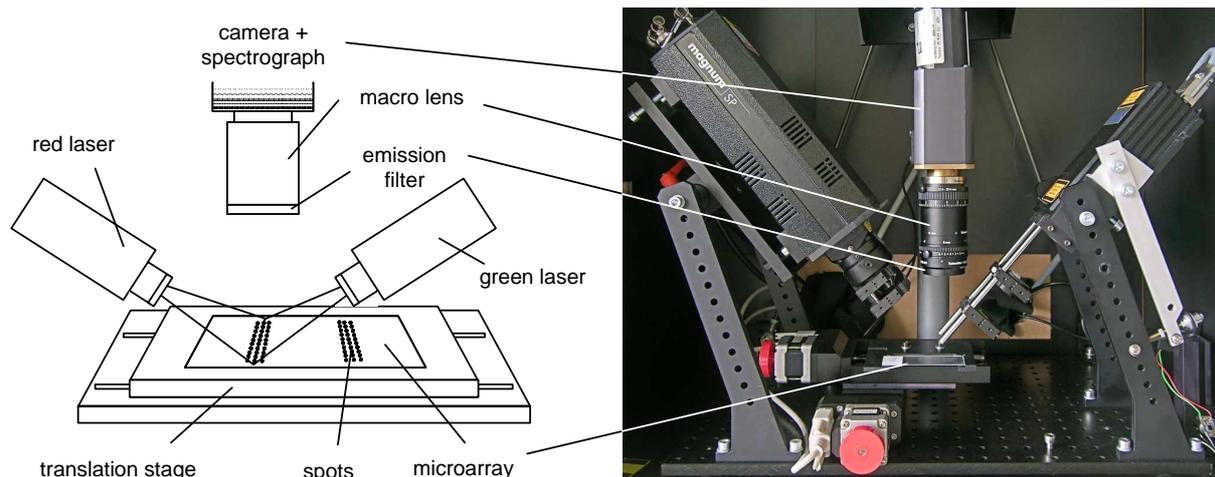


Fig.3, Experimental setup of the spectral imaging microarray scanner

One possible method is linear spectral unmixing, where the emission spectra of all dyes involved have to be known. To avoid the restriction of demanded *a priori* knowledge, a multivariate curve resolution – alternating least squares – algorithm (MCR-ALS) was developed in cooperation with the University of Barcelona [7]. The MCR-ALS only needs sparse *a priori* information, emission spectra are optional. In case they are known, the information can be used as a constraint to accelerate the ALS optimization process. This unmixing procedure does not only calculate quantified fluorescence emissions of the known dyes but also that from initially unknown sources, which can be identified at a later step with their signature. Both unmixing methods yield maps that contain the intensity of all discrete fluorescent emission sources. All calculations were done with Matlab.

## EXPERIMENTAL PROCEDURE AND RESULTS

Four-dyes labeled microarrays (HKI) were analyzed by SI. The fluorescent labels were applied in different configurations of both, co-labeling and concentration.

For component analysis of the SI-data, fixed (*a priori*) as well as a number of adjustable signatures were predefined and used with the MCR-ALS algorithm. For excitation at 532 nm three fixed signatures (Cy3/3.5/5) and four adjustable were admitted. At red excitation, three fixed (Cy3.5/5/5.5) and three variable were allowed. In both cases, a non-negativity constraint was assigned.

The analysis yielded clearly distinguishable intensity maps illustrated in Fig.4. The fundamental emission spectra (a) and (d) found on the biochip were used as signature libraries with the MCR-ALS unmixing procedure. Result of this calculation were intensity maps (b) and (e) for every signature. Two or three of these maps can be combined to a RG(B)-image (c) and (f) that easily shows the co-occurrence of concentrated dye.

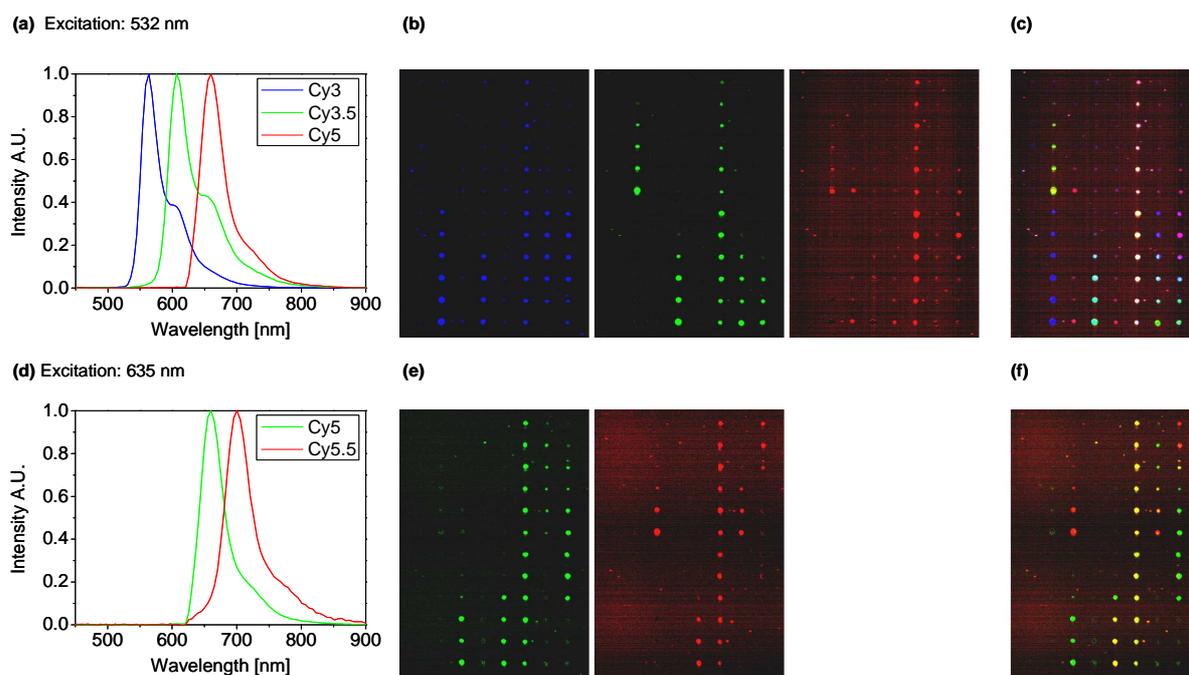


Fig. 4, Fluorescence spectra (a,d), raw intensity maps of Cy3, Cy3.5, Cy5 and Cy5.5 (b,e) and combined intensity maps (c,f), which show the co-occurrence of dye

We also investigated conventional twofold labeled spots. False-positive signals, based on channel crosstalk, were determined by the analysis of 616 spots from a microarray scanner calibration slide (Full Moon BioSystems DS01), 308 of which consisted of Cy3 and the other half of Cy5. The spots were concentrated from  $1.1 \cdot 10^{-3}$  to  $1.5 \cdot 10^5$  fluorophores/ $\mu\text{m}^2$ . The slide was analyzed with GenePix and SI-system. After component analysis by MCR-ALS (non-negativity and equality constraint, one variable and two pre-assigned signatures), the grey value images of both measuring methods were analyzed for average spot intensity. To obtain the relative false-positive error, the average intensity of all equally concentrated Cy5-spots was divided by the associated Cy3-spots after subtraction of the spots interspace signal (background).

The analysis of false-positive signals at 532 nm excitation yielded a relative error of  $6 \cdot 10^{-6}$  (SI) and  $1.5 \cdot 10^{-3}$  (GenePix) at high dye concentrations. At very low amounts of fluorescent label the error was above 10% for GenePix and 5% for SI. On average, the error of the SI-System was found to be smaller by a factor of 3.5 (Fig.5). Comparable results were found at the analysis of twofold labeled microarrays made by the HKI. Here the average false-positive error of 2,7% (GenePix) could be reduced to 0,8% (SI).

Besides the samples described, we have made successful investigations on alternative fluorescent dyes Rhodamine 6G and Tetramethylrhodamine. Furthermore, we analyzed high density (37,632 spots) microarrays from the Netherlands Cancer Institute (NKI).

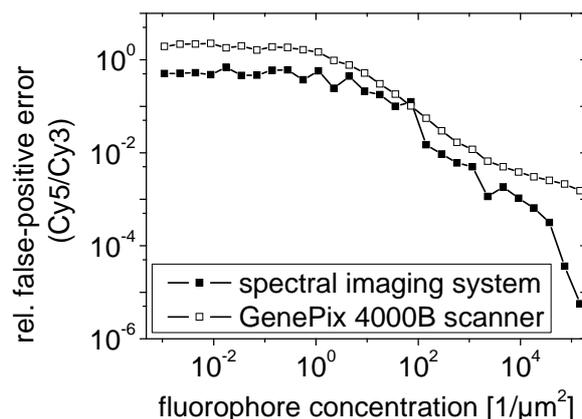


Fig. 5, Comparison of relative false-positive signals of different concentrated Cy5-spots

## CONCLUSION

The spectral resolved measurement of fluorescence emission of biospots by SI-technology shows clear advantages in comparison to conventional filter-based microarray scanners. We could show an efficient detection and discrimination of the fluorescent dyes Cy3, Cy3.5 and Cy5 by green laser excitation and Cy5 and Cy5.5 by red excitation, which is impossible with a conventional filter-based microarray scanner. We also found a reduction of false-positive results for twofold labeled microarrays. Summarized, our SI-technology is able to differentiate clearly between three overlapping spectral dye-signatures besides disturbances. SI has the potential to enhance the accuracy, reliability, and speed of microarray scanners, because the fluorescence of many biospots can be measured simultaneously as well as spectral disturbances of the background.

### References:

- [1] M. J. Martinez, A. D. Aragon, A. L. Rodriguez, J. M. Weber, J. A. Timlin, M. B. Sinclair, D. M. Haaland and M. Werner-Washburne: Identification and removal of contaminating fluorescence from commercial and in-house printed DNA microarrays, *Nucleic Acids Res.*, 31, No.4, e18, 2003.
- [2] M.B. Sinclair, J.A. Timlin, D.M. Haaland, M. Werner-Washburne: Design, construction, characterization, and application of a hyperspectral microarray scanner. *Applied Optics*, 43, No.10, 2079-88, 2004.
- [3] Y.C.M. Staal, M.H.M. van Herwijnen, F.J. van Schooten and J.H.M. van Delft: Application of four dyes in gene expression analyses by microarrays, *BMC genomics*, 6:101, 2005.
- [4] B.R. Peixoto, R.Z.N. Vêncio, C.M. Egidio, L. Mota-Vieira, S. Verjovski-Almeida and E.M. Reis: Evaluation of reference-based two-color methods for measurement of gene expression ratios using spotted cDNA microarrays, *BMC genomics*, 7:35, 2006.
- [5] W.D. Schmidt, F. Erfurth, D. Fassler: Fluorescence imaging spectroscopy of biochips. *Proc. 2nd International Spectral Imaging Workshop (Villach/Austria)*, 27-34, 2005.
- [6] W.D. Schmidt, F. Erfurth, A. Tretyakov, B. Nyuyki, G. Mrotzek, H.P. Saluz, D. Fassler: Spectral Imaging of Fluorescent Microarrays. *Proc. 3rd International Spectral Imaging Workshop. (Graz/Austria)*, 33-47, 2006.
- [7] R. Tauler: Group of Solution Equilibria and Chemometrics ; Chemical and Environmental Research Institute of Barcelona (IIQAB); University of Barcelona / Spain, private communication, 2006

### Authors:

Dipl.-Ing (FH) Florian Erfurth, Dr. Wolf-Dieter Schmidt, Dipl.-Ing (FH) Berla Nyuyki, Dipl.-Ing. Armin Scheibe, Prof. Dieter Faßler  
 GMBU e.V., Felsbachstraße 7, 07745 Jena, +49 (03641) 3667 31, erfurth@gmbu-jena.de

Prof. Hans- Peter Saluz  
 Hans-Knöll-Institut, Beutenbergstr. 11, 07745 Jena

T. Langner / D. Kollhoff

# Farbbasierte Druckbildinspektion an Rundkörpern

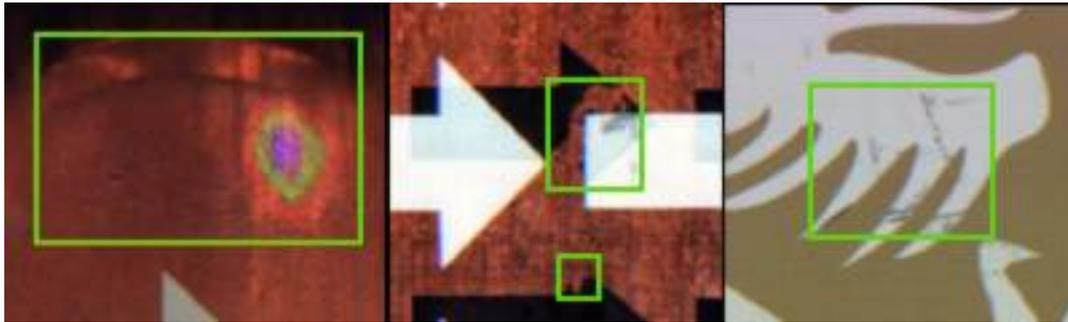
## Kurzfassung

*Die Druckbildinspektion ist eine seit längerem etablierte Methode, um im Bedruckungsprozess von Postern, Etiketten und Verpackungen eine gleichbleibende Qualität zu gewährleisten. Die meisten Inspektionsverfahren beschränken sich dabei auf die Prüfung von planaren Oberflächen. Drucke, welche direkt auf verschiedenartig geformte Objekte aufgebracht werden, erfordern teilweise andere Herangehensweisen. In diesem Beitrag wird ein Druckbildinspektionssystem vorgestellt, welches speziell für die Inspektion von Rundkörpern konzipiert wurde und die sich dabei ergebenden Schwierigkeiten löst.*

## Aufgabenstellung

Die dieser Arbeit zugrunde liegende Aufgabenstellung eines namhaften Trinkflaschenherstellers umfasst die Entwicklung eines Druckbildinspektionssystems zur Kontrolle von Bedruckungen von zylindrischen Flaschen und Tuben. Dabei teilt sich der Inspektionsvorgang in die Fehlererkennung von Druck- und Materialfehlern durch Vergleich mit einem „Golden Template“ und die Prüfung auf korrekten Farbdruck mittels Farbklassifikation. Folgende Fehler müssen sicher erkannt werden:

1. Druck- und Materialfehler
  - Oberflächenfehler (Dellen, Kratzer, Einschlüsse)
  - Spritzer, Verläufe, Versatz, Verwischungen
  - Zu starke, blasse oder fehlende Bedruckung
2. Farbfehler
  - Fehlfarben
  - Farbverläufe
  - Farbabweichungen vom Sollmuster



**Abb.1** Beispiele für Delle, Druckfehler und Kratzer

### **Probleme bei der Inspektion von zylindrischen Körpern**

Für die Aufnahme der zu prüfenden Bilder mit einer Zeilenkamera muss das Prüfobjekt gedreht werden. Die abgewickelte Oberfläche kann dann mit der Kamera erfasst werden. Hierfür wurde ein Hardwareverbund aus Antriebssystem, Beleuchtung und Kamerakopf konstruiert. Als Beleuchtung dient eine faseroptische Linienbeleuchtung in Verbindung mit einer HQI-Kaltlichtquelle.

Bei dieser Art der Bildaufnahme ergeben sich jedoch Probleme, welche größtenteils auf Schlupf im Antriebssystem und Unrundheiten zurückzuführen sind.

Schlupf äußert sich im Bild durch Verzüge wie Stauchung und Streckung des Druckmusters entlang der Aufnahme­richtung. Er entsteht durch unzureichend straffe Einspannung der Flasche / Tube und durch Vibrationen im Antriebsstrang. Aus technischen Gründen wird statt einem Direktantrieb ein Riemenantrieb eingesetzt.

Untersuchungen haben gezeigt, dass durch die Verwendung eines Drehgebers zur Triggerung der Kamera und Optimierung der Hardwarekomponenten der Schlupf minimiert werden kann. Weitere Verzerrungen werden durch einen ungleichmäßigen Rundlauf (dezentrierte Einspannung bzw. Formfehler der Flasche) verursacht, die nicht restlos beseitigt werden können. Die Auswertung der Druckbilder erfordert also ein Maß an Verzugtoleranz der Algorithmen gegenüber Stauchung, Streckung und Scherung.

Ein weiteres Problem bei der Inspektion von zylindrischen Objekten ist das Finden eines Startpunktes für den verzugstoleranten Golden-Template-Vergleich im Bild. Bedingt durch Druckprozess und Transport der Prüflinge, gibt es keinen zur Bedruckung synchronen Startpunkt der Bildaufnahme, so dass die aufgenommenen Druckbilder eine Translation in Aufnahme­richtung aufweisen. Die teilweise verwendete 360°-Bedruckung der Objekte erfordert deshalb ein geeignetes Offsetkorrekturverfahren zur Bestimmung eines reproduzierbaren Startpunktes. Von dem Rundkörper werden zwei vollständige

Umdrehungen aufgenommen, so dass bezüglich eines definierten Startpunktes mindestens eine vollständige Abwicklung im Bild vorhanden ist.

### Probleme bei der Inspektion von Farbdrucken

Für die Aufnahme der Farbbilder wird eine trilineare Farbzeilenkamera verwendet. Diese stellt einen guten Kompromiss zwischen Auflösung und Kosten dar. Die hier auftretenden Probleme resultieren zum Großteil aus dem Abstand der Farbzeilen des Sensors zueinander und der Rotation des Prüfobjektes.

Um die dadurch entstehenden Farbsäume an Kanten zu vermeiden, muss eine genaue Kalibrierung von Antrieb, Drehgeber, Belichtungszeit und Zeilenverzögerung erfolgen. Eine weitere Fehlerquelle ist die Beleuchtung von nicht planaren Oberflächen. Bedingt durch den Abstand der Farbzeilen muss ein relativ breiter Streifen homogen ausgeleuchtet werden. Durch die Wölbung der Oberfläche ergeben sich deshalb für jede Zeile unterschiedliche

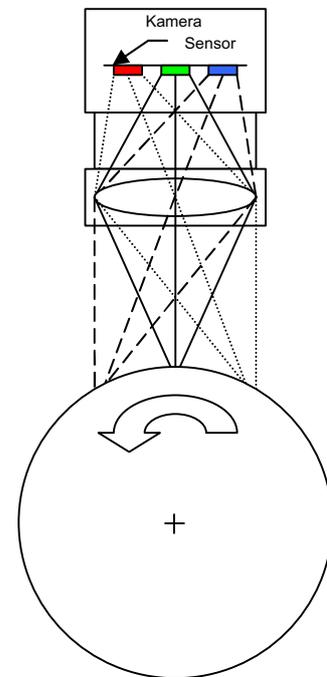


Abb.2 Abbildungsprinzip

Reflektionswinkel. Das Resultat sind auszugleichende Schwankungen der Intensität und Farbverfälschungen. Des Weiteren muss die Kurzzeit- und Langzeit-Stabilität der Lichtquelle beachtet werden. Um die Einflüsse von Schwankungen im Beleuchtungssystem auszugleichen, wird in der nächsten Ausbaustufe eine Online Farb-Kalibrierung mittels Drei- oder Mehrbereichssensoren angestrebt.

### Hardware

Für die Suche nach der optimalen Kombination der Hardwarekomponenten wurde ein intensiver Test von Farbzeilenkameras verschiedener Hersteller durchgeführt. Um Flaschen und Tuben unterschiedlicher Höhe inspizieren zu können, musste ein breiter Aufnahmebereich mit hoher Auflösung abgedeckt werden.

Um das Auftreten von Farbsäumen zu minimieren und den Einfluss der gewölbten Oberfläche auf die Beleuchtung so klein wie möglich zu halten, musste eine Kamera mit einem sehr kleinen Abstand der Farbzeilen gewählt werden.

Die Entscheidung fiel aufgrund einer Auflösung von 4096 Pixeln und einem Zeilenabstand von  $30\mu\text{m}$  auf die Dalsa Piranha Color PC-30-04K60 Kamera.

Für die Normierung des Helligkeitsverlaufes bietet diese zusätzlich eine eingebaute Shadingkorrektur, wodurch sich softwareseitig ein Verarbeitungsschritt einsparen lässt.

Um auf die sich ergebende Verschiebung der Fokusebene bei Änderung des Flaschendurchmessers reagieren zu können, wurde ein Aufbau konstruiert, welcher eine Höhenverstellung des gesamten Aufnahmesystems ermöglicht.

### Lösungsansätze am Beispiel der Offsetkorrektur

Für die Korrektur des Offsets zweier Bilder muss ein gegenüber Fehlern im Druck tolerantes Verfahren zum Einsatz kommen. Herkömmliche Korrelationsansätze oder Vergleiche direkt im Bildmaterial sind daher ungeeignet. Aus diesem Grund werden aus den Aufnahmen mittels Förstner-Algorithmus [1] Merkmale in Form von Eckpunkten mit den Parametern Position  $(x,y)$ , Betrag und Orientierung extrahiert. Das zu matchende Bild enthält nun zwei volle Umdrehungen des Rundkörpers, während das Referenzbild nur eine Umdrehung umfasst. Außerdem werden bei Druckfehlern (z.B. fehlende Farben) nicht alle Eckpunkte wieder gefunden. Die Offsetkorrektur lässt sich daher als Problem des Matchings zweier partiell übereinstimmender Punktmengen charakterisieren. Zur Lösung dieses Problems wird deshalb eine optimierte Form des Point-Pattern-Matchings verwendet.

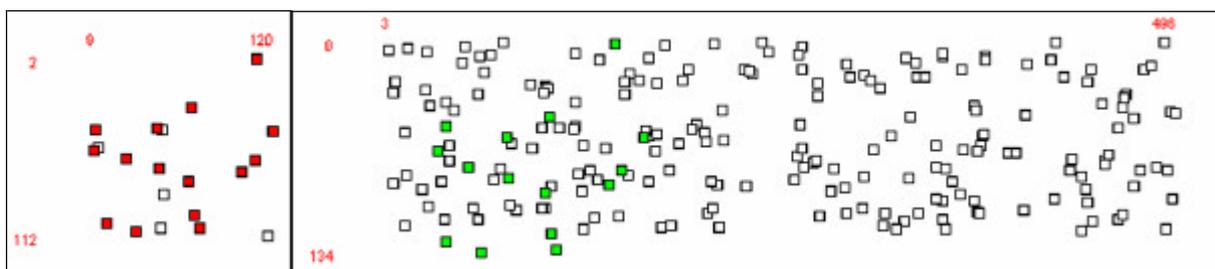


Abb.3 Punktmengen A und B mit gefundener Korrespondenz (mittlerer Distanzfehler = 3,6)

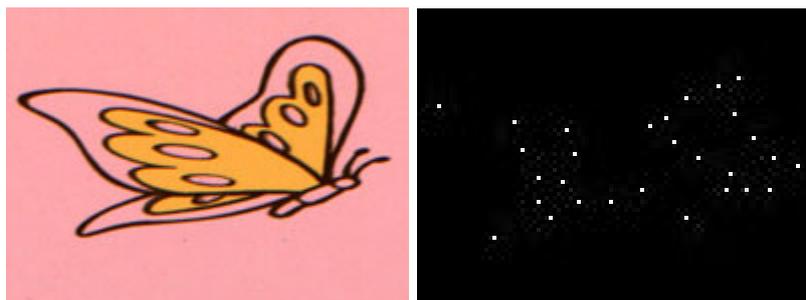


Abb.4 Druckmuster und daraus extrahierte Eckpunkte

Sind zwei Punktmengen  $A$  und  $B$  gegeben, dann löst das Verfahren folgende Aufgaben:

- Finde die korrekten Punktkorrespondenzen zwischen den Mengen  $A$  und  $B$  und berechne daraus die Transformation  $T$ .
- Löse das Minimierungsproblem: Suche die Transformation  $T$ , so dass der Abstand zwischen den Punktmengen  $T(A)$  und  $B$  minimal wird.

Als Suchstrategie kommt hier der Vergleich von Nachbarschaften nach *Wamelen* [2] mit

einer Komplexität  $O\left(m(\log n)^{\frac{3}{2}}\right)$  ( $m$  = Anzahl der Elemente in der Suchpunktmenge,  $n$  =

Anzahl der Elemente in der Referenzpunktmenge) zum Einsatz. Einen Überblick über weitere Strategien geben *Cox und Jager* [3].

Der Algorithmus arbeitet wie folgt:

- Suche nach lokalen Transformationen, die alle nächsten Nachbarn der Punkte aus  $A$  auf die nächsten Nachbarn der Punkte aus  $B$  abbilden
- Auswahl der, den Suchkriterien entsprechenden, besten globalen Transformation aus den lokalen Transformationen

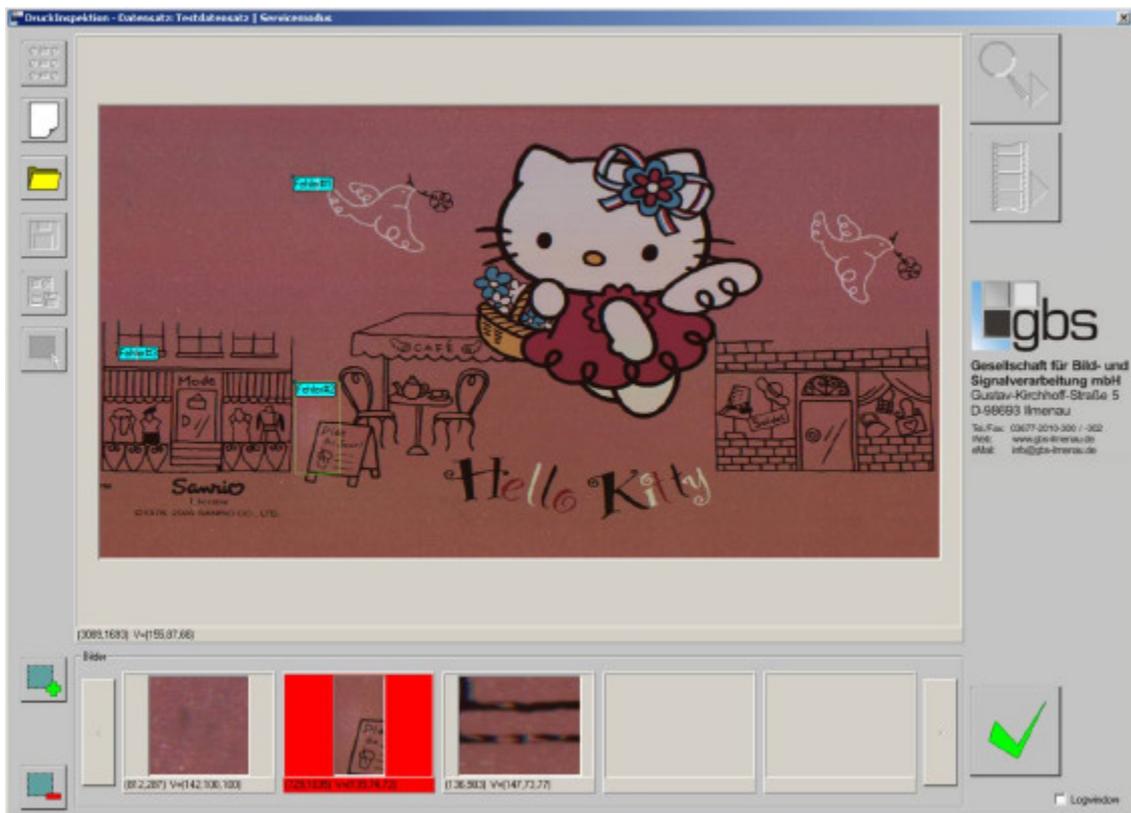
Um die Laufzeit des Verfahrens zu optimieren und die Geschwindigkeit bei der Suche nach den nächsten Nachbarn zu erhöhen, wurden einige Erweiterungen vorgenommen.

- Für eine schnelle Nachbarschaftssuche werden kD-Bäume verwendet.
- Es können Toleranzgrenzen für die einzelnen Transformationen festgelegt werden, um die Suche vorzeitig abubrechen ( es tritt nur geringe Translation senkrecht zur Aufnahmeaufrichtung auf, keine Rotation oder Skalierung ).
- In die Distanzberechnung zweier Eckpunkte fließen neben ihren Koordinaten weitere Eigenschaften, wie z.B. die Intensität der Ecke, ein ( die Wahrscheinlichkeit ist hoch, dass in den zwei Aufnahmen markante Punkte wiedergefunden werden )

Versuche führten wir mit zufallsgesteuert generierten Testdaten durch, wobei zwischen einem Referenzpunktsatz und einem Suchpunktsatz Manipulationen wie affine Transformation, Hinzu- und Wegnahme von Punkten und Verrauschen der Punktkoordinaten vorgenommen wurden. Dabei ergab sich eine sehr gute Übereinstimmung zwischen der vorgegebenen und der ermittelten Transformation. In der Praxis wurden aufgrund der nichtlinearen Verzerrungen des Bildes und des Versatzes der einzelnen Farbdrucke gegeneinander Abweichungen unterhalb von 4 Pixeln erreicht, die für die weitere Verarbeitung ausreichend sind.

## Software

Die Anwendungssoftware wurde so gestaltet, dass sie mit wenigen Schaltflächen per Touchscreen bedienbar ist und dem Druckmaschinenfahrer direkt an der Druckstraße einen schnellen Überblick über die aktuelle Qualität der Bedruckung bietet. Durch den modularen Aufbau der Software wird eine nachträgliche Erweiterung um Statistik- und Klassifikationsmodule ermöglicht. Somit kann die Anwendung optimal an die jeweiligen Anforderungen angepasst werden.



**Abb.5** Applikation zur Druckbildinspektion

### Literatur:

- [1] W. Förstner, E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features, ISPRS Intercommission Workshop on Fast Processing of Photogrammetric Data, 1987
- [2] P. van Wamelen, Z. Li, and S. Iyengar. A fast expected time algorithm for the point pattern matching problem, Technical Report 1999-28, Louisiana State University, Dept. of Mathematics, 1999
- [3] G.S. Cox and G. de Jager. A survey of point pattern matching techniques and a new approach to point pattern recognition, COMSIG 1992

### Autoren:

Dipl.-Inf. Tim Langner  
Department of Computer Science and Automation  
Computer Graphics Group  
TU Ilmenau  
98683, Ilmenau  
Phone: +49-3677-693141  
E-mail: Tim.Langner@tu-ilmenau.de

Dipl.-Ing. Dietmar Kollhoff  
GBS mbH,  
Gustav-Kirchhoff-Str 5  
98683, Ilmenau  
Phone: +49-3677-2010303  
Fax: +49-3677-2010302  
E-mail: Dietmar.Kollhoff@GBS-Ilmenau.de

C. Lucht / F. Gaßmann / R. Jahn

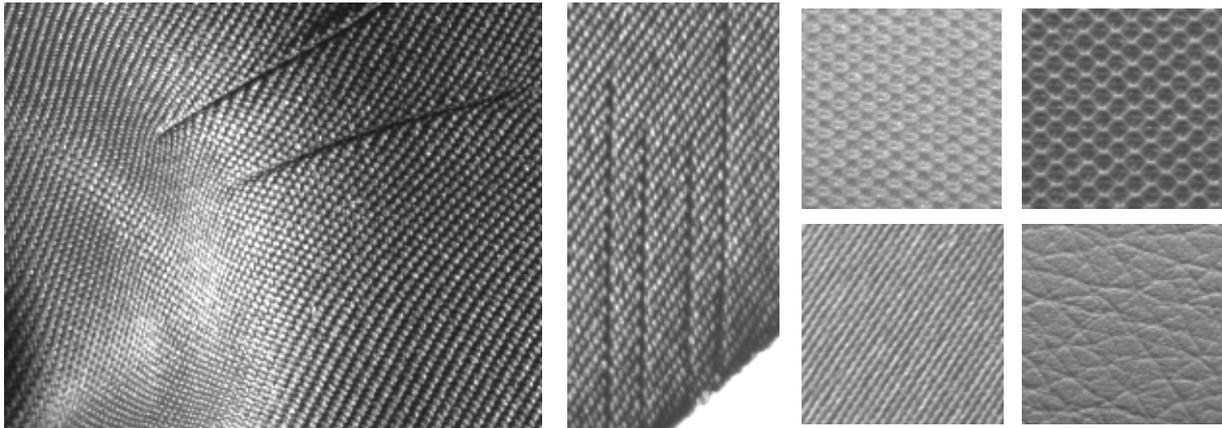
## **Inline-Fehlerdetektion auf freigeformten texturierten Oberflächen im Produktionsprozess**

### **Einleitung**

Die Gewährleistung eines hohen Qualitätsstandards bei gleichzeitiger Reduzierung der Produktionskosten erfordert die Ausweitung der maschinellen Qualitätskontrolle mit immer anspruchsvolleren Anforderungsprofilen. Zu diesen komplexen Prüfaufgaben ist die Inline-Inspektion von Bauteilen mit freigeformten, texturierten Oberflächen zu zählen. Dabei sind unter den im Fertigungsprozess bestehenden Randbedingungen sowohl die Maßhaltigkeit der 3D-Formgebung als auch die Fehlerfreiheit der Oberfläche mittels industrieller Bildverarbeitung zu kontrollieren. Insbesondere die Inspektion der Oberfläche, welche bereits im unverformten (planen) Zustand eine hohe Herausforderung darstellt, bedarf der Anwendung leistungsfähiger Bildverarbeitungswerkzeuge. Nachfolgend sollen am Beispiel von Bauteilen für die Kfz-Innenverkleidung ein ausgewählter Lösungsansatz mit Konzeption der Bildanalyse und Merkmalauswahl vorgestellt werden. Das ZBS verfügt über ein umfangreiches, in langjähriger Projektarbeit angeeignetes Know-how auf den Gebieten der Texturanalyse und 3D-Datenverarbeitung ([1]...[4]). Darauf aufbauend wird aktuell ein AiF-gefördertes Kooperationsprojekt zur genannten Prüfaufgabe bearbeitet. Dieser Beitrag stellt ausgewählte Zwischenergebnisse vor.

### **Problemstellung**

Zur Verkleidung von Kfz-Innenräumen werden aus Laminaten mit texturierter Oberflächenschicht vielfältige 3D-geformte Bauteile gepresst, zugeschnitten und für die Formhaltigkeit mit Kunststoff hinterspritzt. Die Oberflächen bestehen aus verschiedenartigen Materialien (z.B. textile Flächengebilde, Leder) mit typischer, aber vielgestaltiger Strukturierung und Farbgebung. Während der Fertigung können verschiedene Fehler auftreten, welche inline erkannt werden müssen. Hierzu zählen neben Abweichungen von der 3D-Form vor allem Oberflächendefekte wie Falten, Blasen, Fadenauslöser, Durchspritzungen und Verzüge. Dabei reichen die Fehlergrößen von der Mikrostruktur des Flächengebildes (z.B. Fadenfehler) bis zu großflächigen Oberflächenveränderungen.



**Abbildung 1** Typische Texturen und Oberflächenfehler; links: Textureindruck an Ausformung mit gleichzeitiger Faltenbildung als Fehler, Mitte: Fadenauslöser an einer Schnittkante, rechts: weitere Texturbeispiele (3x Textil, 1x Leder-Imitat)

Die Prüfung soll in separaten, fertigungsintegrierten Stationen erfolgen, was aufgrund des bestehenden Durchsatzes Prüfzeiten von < 20 s pro Bauteil erfordert. Die bedarfsgerechte Fertigung führt zu häufigen Wechseln der Bauteilform, des Materials und der Farbgebung, womit entsprechende Umrüstungen des Prüfsystems notwendig werden.

### Lösungsansatz

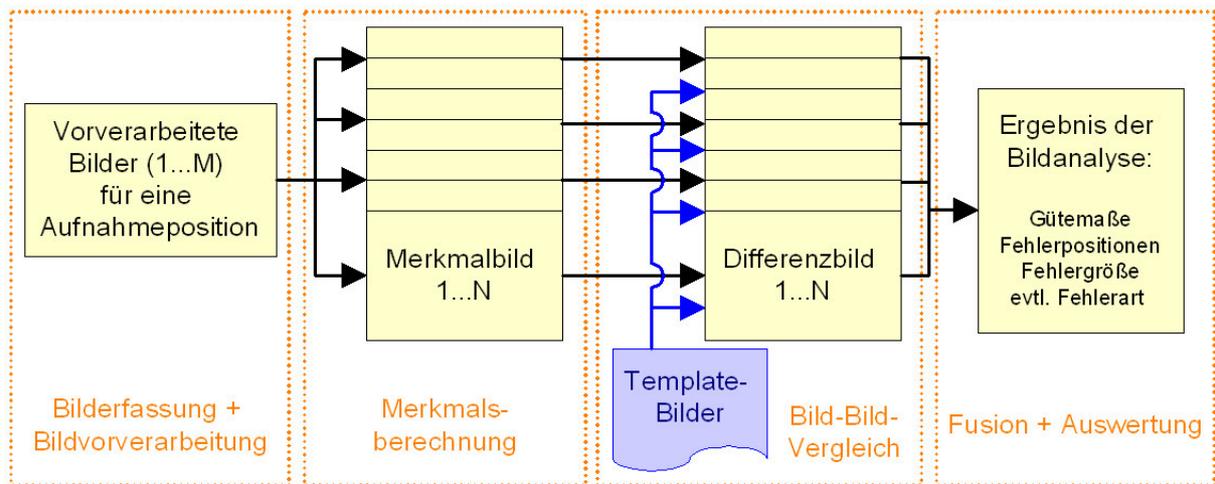
Die Prüfaufgabe erfordert den Einsatz der Texturanalyse. Dabei kann prinzipiell auf leistungsfähige, sich bei der Inspektion ebener, meist homogen texturierter Oberflächen bewährte Algorithmen zurückgegriffen werden. Da aber das enge Zeitregime der Fertigung nur wenige Bildaufnahmen zur Erfassung der komplexen Objektoberfläche erlaubt, müssen großflächige Abbildungen vorgenommen werden. Die resultierenden neigungsbedingten Verzerrungen führen dann zu einem inhomogenen Textureindruck unabhängig von der realen Oberflächenstruktur. Zusätzlich bewirken unterschiedliche Beleuchtungswinkel bei verschiedenen Flächenneigungen weitere Texturinhomogenität. Generell wird die Detektierbarkeit von Oberflächenfehlern hierdurch stark beeinträchtigt.

Einen möglichen Ansatz, diesen erschwerten Analysebedingungen zu begegnen, stellt der in der industriellen Bildverarbeitung vielfältig eingesetzte Bild-Bild-Vergleich dar. Dabei wird davon ausgegangen, dass Bilder von gleichen Oberflächenregionen bei identischen Abbildungs- und Beleuchtungsbedingungen zwischen einem fehlerfreien Prüfling und einem Gutmuster (Golden Template) den örtlich gleichen Textureindruck aufweisen. Im Fehlerfall sind Oberflächendefekte dann als möglichst deutliche Abweichungen detektierbar. Bei texturierten Oberflächen lässt sich dieses Verfahren jedoch nicht direkt auf den erfassten Bilddaten (Grauwerte) anwenden. Sinnvolle Ergebnisse sind nur bei einem Vergleich von Merkmalbildern mit bestimmten Invarianzeigenschaften (z.B. Phaseninvarianz bei periodischen Texturen, Invarianz bzgl. tolerierbarer Mikrostrukturab-

weichungen) möglich. Eine wesentliche Schwierigkeit dieses Ansatzes besteht offensichtlich in der Auffindung von Texturmerkmalen bzw. Merkmalsätzen, welche die jeweils betrachteten, erwartungsgemäß inhomogenen Texturen gut beschreiben und gegen mögliche Oberflächenfehler abgrenzen. Nachfolgend soll dieser templatebasierte Lösungsansatz detaillierter betrachtet werden.

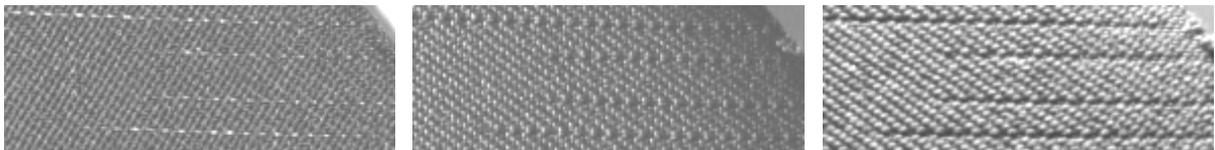
### Konzeption der Bildanalyse

Das Grundkonzept der Kannphase lässt sich gemäß Abbildung 2 in vier Blöcke gliedern.



**Abbildung 2 Grundkonzept für die Kannphase der Bildanalyse**

Die Bilderfassung der Objektoberfläche mit wenigen Bildaufnahmen bei maximaler Struktur- und Fehlerinformation kann aufgrund der 3D-Form nur gelingen, wenn das Objekt und/oder die Aufnahmeanordnung relativ zueinander in Optimallagen verkippt werden. Wegen des Beleuchtungseinflusses auf die Fehlerausprägung (siehe Abbildung 3) muss dabei ein besonderes Augenmerk auf die Beleuchtungsgeometrie gelegt werden.



**Abbildung 3 Textur- und Fehlerausprägung bei unterschiedlich gerichteter Beleuchtung**

Den gestellten Anforderungen wird eine Bildaufnahmeeinrichtung gerecht, welche aus einer adaptierbaren Kamera-Beleuchtungs-Anordnung und einer Positioniereinrichtung für die Prüfobjekte besteht. Die Positioniereinrichtung, vorzugsweise ein Roboterarm, präsentiert die Bauteiloberfläche dem Bildaufnahmesystem in vorbestimmten Lagen. Das Bildaufnahmesystem besteht aus ein oder mehreren Matrixkameras sowie zugehörigen Strahlern für eine flächige Ausleuchtung. Für mehrer Aufnahmen in einer Lage werden zur Beleuchtungsvariation Strahler mit blitzbaren Lichtquellen (z.B. LED) verwendet. Der Bilderfassung folgt eine Bildvorverarbeitung, welche Arbeitsschritte wie Shadingkor-

rektur, Histogrammtransformationen und Vorfilterungen zur Verbesserung der Merkmalgewinnung enthält. Des Weiteren sind geschwindigkeits- und speicheroptimierende Maßnahmen wie die pyramidale Zerlegung der Quellbilder und die Maskierung von relevanten Oberflächenregionen hinzuzuzählen.

Aus den vorverarbeiteten Bildern werden im nächsten Arbeitsschritt mehrere Merkmalbilder berechnet. Die in einer Belehrungsphase durchgeführte Merkmalauswahl hängt wesentlich von der Oberflächentextur, den zu erwartenden Fehlern und dem Zeitrahmen für die Inspektion ab (siehe nachfolgenden Abschnitt). Für den Bild-Bild-Vergleich sollen Oberflächenfehler als Abweichungen zur fehlerfreien Textur kenntlich gemacht werden, gleichzeitig sollen tolerierbare Texturunterschiede keine Betonung erfahren.

Der Bild-Bild-Vergleich berechnet Differenzmaße zwischen den Merkmalbildern des Prüfobjekts und den als Golden Template abgelegten Merkmalbildern von Gutmustern. Um zu tolerierende Texturschwankungen besser zu berücksichtigen wird ein Bild-Bild-Vergleich mit geringer Ortsunschärfe [3] durchgeführt.

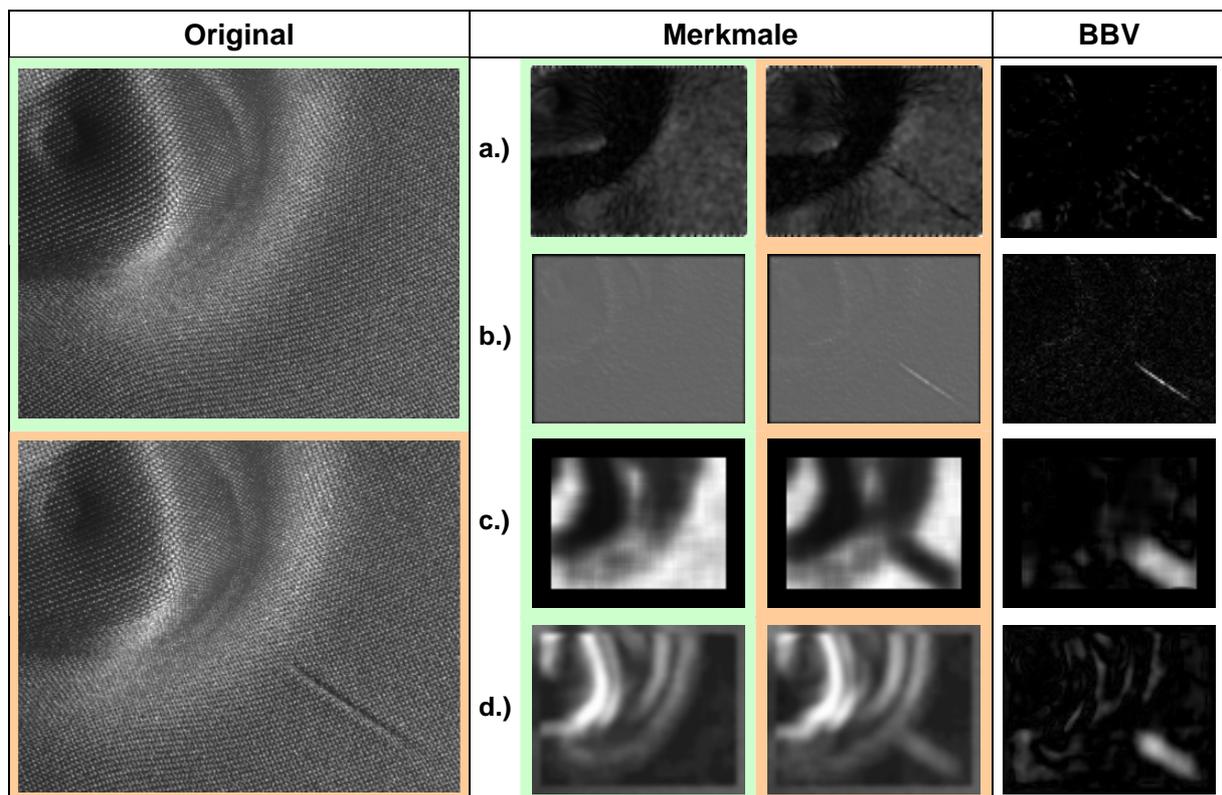
Im letzten Schritt der Bildanalyse werden die Differenzbilder der einzelnen Merkmalbilder mittels intelligenter Schwellwertverfahren ausgewertet und die Ergebnisse auf vorzugsweise höherem Abstraktionsniveau fusioniert. In Abhängigkeit von der Merkmalauflösung sind die Position und Abmessung einzelner Fehler sowie eingeschränkt die Fehlerart als Resultat verfügbar.

Die Belehrungsphase für dieses Konzept muss bei einem hohen Automatisierungsgrad sehr aufwendig gestaltet werden. Zur Belehrung bzw. Einrichtung zählen die Auswahl der Kamera-Beleuchtungs-Anordnung, die Festlegung von Bildaufnahmepositionen und Regionenmasken, die Selektion der Texturmerkmale pro Region, die Parametrierung der Merkmalberechnung, die Templategenerierung sowie die Ermittlung von Schwellwerten für die Ergebnisauswertung. Die Belehrungsergebnisse werden in prüfobjektspezifischen Parameterrezepten zusammengestellt und abgelegt. Damit kann auch bei kurzen Umrüstphasen in der Fertigung eine schnelle Einrichtung des Inspektionssystems auf bekannte Prüfobjekte erfolgen. Belehrungen auf neue Prüfobjekte (neue Form, Material, Texturorientierung) sind nach Möglichkeit offline vorzunehmen, da sonst hohe Geschwindigkeitsvorgaben erfüllt werden müssen.

### **Auswahl von geeigneten Texturmerkmalen**

Die große Bandbreite an Texturarten und Fehlern erfordert eine große Palette von Merkmalen. Einschränkend wirkt sich hierbei das enge Zeitregime für die Inspektion aus. Wenig rechenintensive Texturmerkmale können durch lineare Filteroperationen gewon-

nen werden. Bei geeigneter Parametrierung dieser Filter und Organisation zu einer Filterbank lassen sich eine Vielzahl von Texturen beschreiben und enthaltene Fehler detektieren. Die Parametrierung kann mit systematischen Abstufungen (z.B. nach Frequenz, Orientierung) oder texturadaptiv erfolgen. Durch Nutzung von pyramidal zerlegten Quellbildern und Maskierungen für relevante Regionen wird die Bildverarbeitung beschleunigt, was auch die Berechnung größerer Merkmalsätze ermöglicht.



**Abbildung 4 Bild-Bild-Vergleich von Merkmalbildern eines Gutmusters (grün hinterlegt) und eines fehlerhaften Prüflings (rot hinterlegt); Merkmale: a.) adaptierter Bandpass, b.) adaptierter Texturfilter, c.) Energiemaß aus NGLD-Matrix, d.) lokale Varianz**

Zur phaseninvarianten Merkmalgewinnung bei periodischen Texturen mit ausgeprägten Peaks im Fourierspektrum eignen sich richtungsselektive Bandpassfilter (z.B. Gaborfilter, siehe Abbildung 4a). Bei Adaption auf jeweils einen Peak lassen sich alle Texturstörungen (z.B. Falten, Fadenauslöser) detektieren, welche diesen Spektralanteil nicht oder nur geringfügig aufweisen. Zur Ausschöpfung des Potentials sollte die Parametrierung deshalb möglichst texturadaptiv, also durch Auswertung der Texturortsfrequenzen, erfolgen. Neben den Bandpässen sind auch adaptive Filter denkbar, welche zur Bildung der Filtermaske die komplette Strukturinformation aus beispielsweise lokalen Autokovarianzmatrizen nutzen. In Abbildung 4b sind exemplarisch die Merkmalbilder für einen adaptierten Texturfilter nach [5] dargestellt. Weitere effiziente Möglichkeiten für die Texturbeschreibung von periodischen aber auch nicht periodischen Texturen bilden Merkmale der lokalen Grauwertstatistik. Neben statistischen Momenten erster Ordnung (Mit-

telwert, Varianz,...) ist die Grauwertstatistik zweiter Ordnung in Form von integralen Merkmalen der Co-occurrence-Matrizen bzw. deren Modifikationen (z.B. NGLD-Matrix [2]) nutzbar. Die erforderlichen Invarianzeigenschaften werden zu Lasten der Bildauflösung durch die Integration in lokalen Umgebungen erreicht. Beispielhafte Merkmalsbilder zeigen die Abbildungen 4c und 4d. Der anschließende Bild-Bild-Vergleich lässt für alle Merkmale des betrachteten Fehlerbeispiels eine sichere Fehlerdetektion zu.

### **Zusammenfassung**

Die Inline-Oberflächeninspektion von freigeformten, texturierten Bauteilen für die Kfz-Innenverkleidung ist in jeder Hinsicht eine anspruchsvolle Bildverarbeitungsaufgabe. Sowohl Komplexität und Vielfältigkeit der Oberflächentexturen als auch die bestehenden Randbedingungen im Fertigungsprozess bilden hohe Anforderungen für die Problemlösung. Der vorgestellte Lösungsansatz für die Bildanalyse basiert auf dem Bild-Bild-Vergleich von Texturmerkmalsbildern. Als Texturmerkmale eignen sich vorrangig die Ergebnisse schneller Texturfilteroperationen und die integralen Merkmale der lokalen Grauwertstatistik erster und höherer Ordnung. Angesichts der großen Bandbreite von Texturen und Fehlern lässt sich kein allgemein nutzbarer Merkmalsatz benennen. Vielmehr muss für jede Bauteilform und Oberflächentextur ein eigenes Rezept erstellt werden, welches in einem aufwendigen Belehrungsprozess mit Adaption der Verfahrensparameter an Textur und potentielle Fehler gewonnen wird.

Die bisherigen Untersuchungen zum Lösungsansatz stimmen positiv, zeigen aber auch einen immensen Entwicklungsbedarf auf. Die Umsetzung des vorgestellten Bildanalysekonzepts sowie die Entwicklung, Erprobung und Evaluierung der erforderlichen Algorithmen sind Gegenstand der aktuellen Projektbearbeitung.

#### **Literatur:**

- [1] Franke K.-H.: Erfassung und Verarbeitung von 3D-Daten, Schriftenreihe des ZBS e.V., ISSN 1432-3346, Ilmenau, 2004
- [2] Franke K.-H.: Bildverarbeitung und Mustererkennung, Schriftenreihe des ZBS e.V., ISSN 1432-3346, Ilmenau, 2004
- [3] Franke K.-H., Lucht C., Nickolay B., Nowack C., Rösel G.-S.: Neue Bildverarbeitungstechnologien für die automatisierte optische Kontrolle strukturierter Oberflächen in der Produktion, Forschungskuratorium Maschinenbau e.V., Heft 250, 2000
- [4] Franke K.-H., Kempe H., Kollhoff D., Wienecke J.: Merkmalsbasierte Fehlerdetektion und -klassifikation bei der Inspektion von strukturierten Wafern, Proceedings 40. Int. Scientific Colloquium, TU Ilmenau, 1995, Seiten 642-647
- [5] Poschmann R.: Untersuchung eines Basisalgorithmus der statistischen Texturanalyse, Dissertation, TU Ilmenau, 1992

#### **Autoren:**

Dipl.-Ing. Carsten Lucht, Dipl.-Ing. Frank Gassmann, Dr.-Ing. Rainer Jahn  
Zentrum für Bild- und Signalverarbeitung e.V./(ZBS e.V.), Gustav-Kirchhoff-Str. 5  
98693, Ilmenau, Germany  
Phone: +49 3677 2010 303  
Fax: +49 3677 2010 302  
E-mail: [carsten.lucht@zbs-ilmenau.de](mailto:carsten.lucht@zbs-ilmenau.de) / [frank.gassmann@zbs-ilmenau.de](mailto:frank.gassmann@zbs-ilmenau.de) /  
[rainer.jahn@zbs-ilmenau.de](mailto:rainer.jahn@zbs-ilmenau.de)

H.-Wolfg. Lahmann / M. Stöckmann

## **Optical Inspection of Cutting Tools by means of 2D- and 3D-Image Processing**

### **Image Processing, Image Analysis and Computer Visions**

#### **ABSTRACT**

2D-and 3D-image processing systems become more and more important in the cutting tool industry. Because of increased requirements to the different machining processes (e.g. milling, turning, grinding) the used cutting tools must have a very high quality. Hence, it is necessary to evaluate the tools quality before they are used in the machine. The non-contact determination of quality parameters by means of image processing systems is therefore of special relevance. This paper presents a solution of 2D-inspection of cutting inserts for hard-whirling processes and a 3D-measuring approach in order to evaluate features of the surface topography of grinding tools.

#### **1 INTRODUCTION**

The hard-whirling process is the most economical milling procedure for the production of threads and thread-like profiles. Hard whirling uses profiled cutting inserts which must be specially prepared.

The profiling of new and worn cutting inserts for whirling tools is realised by a grinding process. Before the grinding process can start the correct processing parameters for the machine are to be determined. For the determination of processing parameters the geometry (i.e. the contour) and the wear of already used cutting inserts are to be measured by means of a vision system. In the first part of this paper a developed solution will be presented to evaluate the profile and face wear of cutting inserts. In the second part two investigated measuring principles for the evaluation of surface topography features of grinding tools, which are used in profiling process of cutting inserts et al., will be described.

## 2 DETERMINATION OF WEAR FEATURES OF CUTTING INSERTS

In order to detect and to evaluate the wear appearances it is necessary to use incident lights for acquisition of face wear and transmitted light for acquisition of profile wear.

Fig. 1 shows both applied measuring principles.

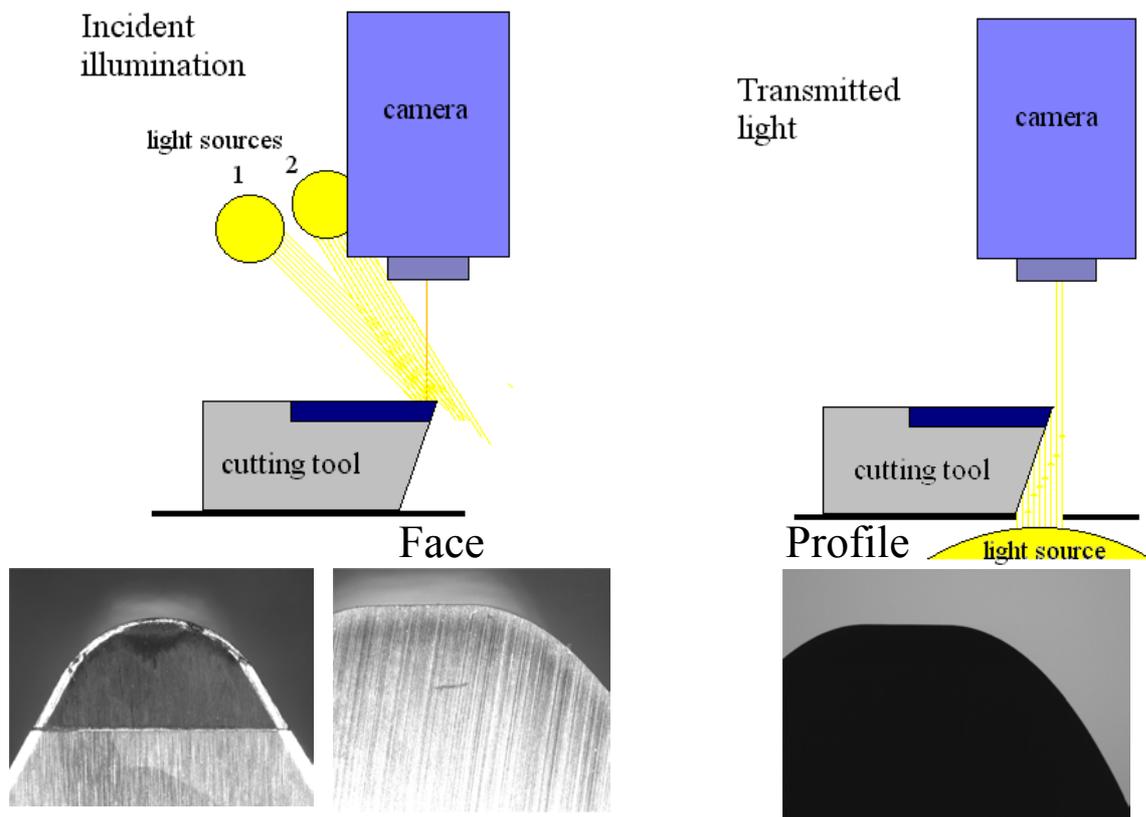


Fig.1: 2D-Measuring principles

For the detection of both kinds of wear appearances a coordinate measuring device with a high resolution CCD matrix camera and a special illumination unit are used. The measuring device is coupled to a computer system to control the grinding machine which profiles the tools to be used in the hard-whirling process. The incident illumination unit consists of two light sources which are aligned onto the focus of the camera system (Fig. 2). The light sources are automatically controlled by evaluating the contrast of the surfaces measured. Angles of the cutting surface can be measured to an accuracy of 2 degrees. This illumination allows image acquisition with reproducible quality and full contrast pictures. The separate radial and vertical positioning of the illuminator arms enables to light up any cutting face profiles for an evaluation.

On the basis of the dimension of cutting inserts some images of the cutting face are

captured and fit together. This is based on the necessary magnification of the cutting face for the take up of the wear dimension. For this reproduction you need a 3 times magnification on the coordinate measuring device (UNI-VIS 250, Mahr OKM GmbH) with an appropriate image size of 2.44mm x 2.46mm. In relation to this image field you need 3 up to 12 pictures depending on the whirling blade (cutting insert) size. These images have to be taken up overlapped, with the same contrast /1/.

The determination of profile wear is carried out in transmitted light by means of 2D-image processing approach. At this principle the cutting edge of whirling insert is projected as bright/dark transition on the camera. The bright/dark transition is detected via measuring line, located perpendicular to the cutting edge, and as scatter-plot of measuring data is recorded. About the measured x-y-positions of coordinate table of measuring device in relation to the CCD-camera the data will be exactly related to each other and evaluated. The measuring process occurs fully automated on the basis of CAD files. The end result is the profile deviation between actual- und nominal contour /2/.

The face wear is captured in incident light (Fig. 2). After image acquisition of all single-regions of the insert surface the captured images are fit together in order to evaluate the surface wear (Fig. 3). Afterwards follows the calculation of wear in several steps. Based on the cutting edge the position of the chamfer is determined at first, if presented, and then it will be determined the areas of wear. The face wear is subdivided in breakouts as well as in width and depth of wear. The determination of depth of wear is carried out separately by means of a 3D-image processing approach (according to confocal principle) /3/.



Fig. 2: Incident light source

(Source: Mahr OKM)



Fig. 3: Composed images

### 3 DETERMINATION OF SURFACE FEATURES OF GRINDING TOOLS

High precision and economical grinding of cutting tools as well as of workpieces can only be done by means of very sharp grinding tools. Therefore the topography of grinding tools plays a very high importance. Electroplated grinding tools (metal bonded grinding tools) have generally a single grinding layer. This thickness depends on the process of electroplating the grinding grains in the bonding matrix and reaches values between 30% up to 60% of the used grain size. The production process of electroplated grinding tools is determined by statistical procedures, which also influence the physical characteristics.

The key features for the topography of electroplated grinding tools are the average grain protrusion (Fig. 4) and the average deviation of the grain peaks. Based on a White-Light-Interferometer and a Confocal System with Super-Long-Working-Distance- Objectives 3D-image processing approach were investigated in order to determine the key features for the topography of grinding tools.

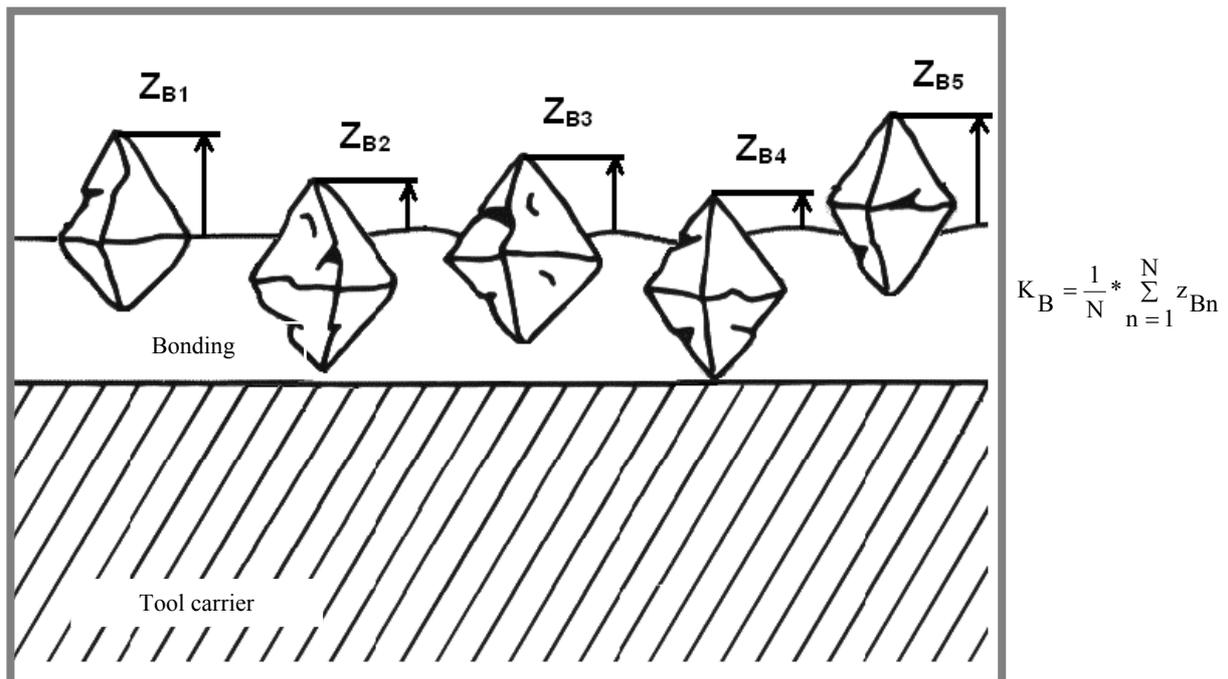


Fig. 4: Average grain protrusion ( $K_B$ )

At the White Light Interferometry the short coherence length of white light source is used. A beam of white light is split. One beam is projected on the tool surface to be measured, the other is projected on reference mirror. After reflection at the surface, both beams recombine and interfere (correlogram, Fig. 5). A CCD-camera captures the local intensity created as function of the tool topography. Moving the grinding tool during the

measurement changes the interference pattern, which allows the determination of maximal values of grain peaks and the bonding level. Based on this information a computer generates a 3D-image of the tool's topography. Following the 3D-image is used in order to calculate the average grain protrusion and the average deviation of the grain peaks.

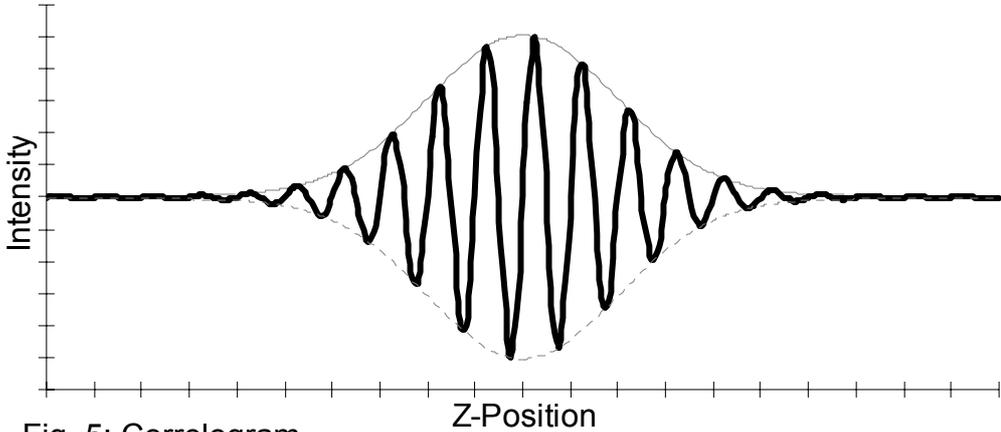


Fig. 5: Correlogram

At the Confocal System (Fig. 7) microscope objectives with high numerical apertures ( $0.6 < N < 1.4$ ) and super-long-working-distance are used. The optical unit (camera, illumination) of the Confocal System is advanced vertically in steps so that each point on the tool surface passes through the focus. The focus function (contrast in depending of vertical position) is determined for each pixel of the CCD-camera. Then the maximum of each focus function is determined and out of it the 3D-image is generated (Fig. 6). The 3D-image data is the basis for the calculation of the average grain protrusion and the average deviation of the grain peaks [4].

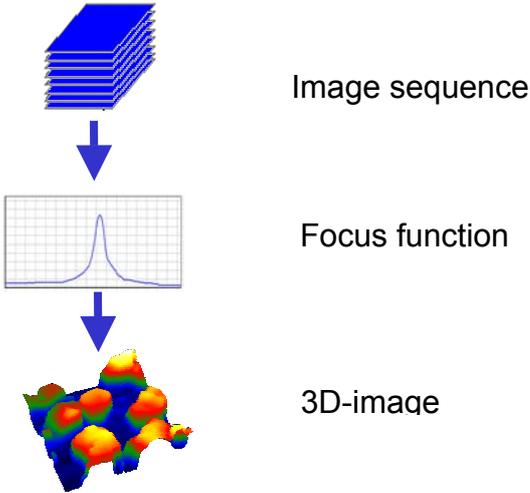


Fig. 6: Procedure



Fig. 7: Confocal System

Both the White Light Interferometry and the Confocal Microscopy may be used for the determination of the average grain protrusion and the average deviation of the grain peaks of electroplated grinding tools. The achieved results have shown that the topography of grinding tools with natural grains is better to detect by means of the Confocal Microscopy than by means of the White Light Interferometry. Investigations on grinding tools with synthetic grains showed good results based on the White Light Interferometry.

In consideration of required measuring time and accuracy of measurement by user of the industry the application of Confocal Microscopy is preferable compared to the White Light Interferometry.

#### 4 Conclusion

The fundamental investigations for the development of 2D-inspection solution for cutting inserts were carried out in close co-operation with partners of tool and measurement equipment industry /1/.

The first development works in conjunction with the White Light Interferometry and the Confocal Microscopy were supported by companies of the grinding tool industry /5/.

#### References:

- [1] Lahmann, H.-W.: Integration of a Vision System for the Profiling of Cutting Tools in a Grinding Process. EUTIST-IMV Meeting, Pori (Finland), 27./28.May 2002
- [2] Ernst, D.; Grüner, E.; Lahmann, H.-W.; Moser, W.: Bildverarbeitungssystem bewertet Schneidplattenverschleiß. Qualität und Zuverlässigkeit, 5/2004. S. 114-116
- [3] Lahmann, H.-W., Hage, H.: Optische Messtechnik an einem KMG steuert die Fertigung von Zerspanwerkzeugen. VDI-Berichte Nr. 1914, 2005, S.177-186
- [4] Lahmann, H.-W., Stöckmann, M.: Messtechnische Bewertung von Schleifwerkzeugoberflächen auf Basis der konfokalen Mikroskopie. 6. Oldenburger 3D-Tage, Oldenburg, 1./2. Februar 2006
- [5] Barthelmä, F.; Lahmann, H.-W.; Stöckmann, M.: Qualitätsbewertung von galvanisch gebundenen Diamant-Schleifwerkzeugen mittels Bildverarbeitung. IDR 37 (2003) Nr. 4, S. 302-307

#### Authors:

Dipl.-Phys. Heinz-Wolfgang Lahmann  
Dipl. Wirtsch.-Ing. (FH) Marius Stöckmann  
Gesellschaft für Fertigungstechnik und Entwicklung e.V., Näherstiller Str. 10  
98574, Schmalkalden  
Phone: +49 3683 69000  
Fax: +49 3693 690016  
E-mail: info@gfe-net.de

M. Melitzki / G. Stanke / F. Weckend

## Bestimmung von Raumpositionen durch Kombination von 2D-Bildverarbeitung und Mehrfachlinienlasertriangulation – am Beispiel von PKW-Stabilisatoren<sup>1</sup>

### ABSTRACT

A system for the measuring of space positions of elements relevant for mounting of mechanical parts is presented. The measurement is performed combining 2D image processing in parallel with multiple laser line triangulations. The development is done for a bending machine for the compensation of deformations of car stabilizers suffering deformations during cooling. Two measuring systems are used one for each end. Existing results of the 2D-image processing and of 3D-measurement were purposefully extended and the results of both procedures were merged and used for a bending simulation process. The implemented measurement concept proved as durably and sufficiently flexible, it resulted in the necessary accuracy in the range better 0.1 mm.

### MOTIVATION

In Fertigungsprozessen der Fahrzeugindustrie sind zunehmend Bauteilkomponenten mit komplizierten Geometrien anzutreffen. Typische Beispiele sind 3-dimensional gebogene Massivrundstäbe für PKW-

Stabilisatoren, die für die Querstabilität des Fahr-



Abb. 1: PKW-Stabilisator

zeugs verantwortlich sind. Die gebogenen Stäbe besitzen je nach Fahrzeugtyp und Einsatzzweck eine Länge zwischen 1200 und 2500 mm und einen Durchmesser von 15 – 100 mm. Bei der Warmformtechnologie wird das auf ein definiertes Längenmaß geschnittene Stabmaterial nach der Erwärmung in einer Biegemaschine in die gewünschte Geometrie gebracht. Die fertig gebogenen Teile werden vergütet, um die geforderten mechanischen Gebrauchseigenschaften einzustellen. Während des

---

<sup>1</sup> BMWi-gefördertes Kooperationsprojekt (PRO INNO II, KF0030704LF5) „Automatisches System zum Vermessen und Richten von 3-dimensional gebogenen Massivrundstäben (STAB)“

Abschreckens im Vergütungsvorgang kommt es zu einer schlagartigen Umwandlung des Gefüges und in Folge dessen auch zu Verzügen der Bauteile. Der eingetretene Verzug muss mit höchster Genauigkeit vermessen und bei Überschreitung der zulässigen Toleranzen nachgerichtet werden [1]. Der Mess- und Nachrichtvorgang passiert heute aufwendig von Hand. Dieser Prozess ist durch eine automatisierte Mess- und Biegevorrichtung, die neben der Messung und dem mechanischen Biegen auch eine Biegesimulation umfasst, effektiver und zuverlässiger zu gestalten.

## SYSTEMANSÄTZE / MESSAUFBAU

Für die Vermessung 3-dimensionaler Objekte liegen unterschiedliche Techniken vor. Von diesen wurden taktile und auf Stereokameras basierende ausgeschlossen. Versuchstechnisch realisiert und auf Eignung hin untersucht wurden zwei Ansätze. Zum einen wurde ein laserbasierter intelligenter Triangulationssensor entlang der bekannten Grobgeometrie des Stabilisators verfahren und zum anderen wurde mit zwei festen Systemen die Raumlagen der montagerelevanten Stabilisatorflanschbohrungen links und rechts vermessen. Der erste Ansatz liefert die komplette Stabisometrie, einschließlich der Bohrungspositionen. Vorteilhaft bei

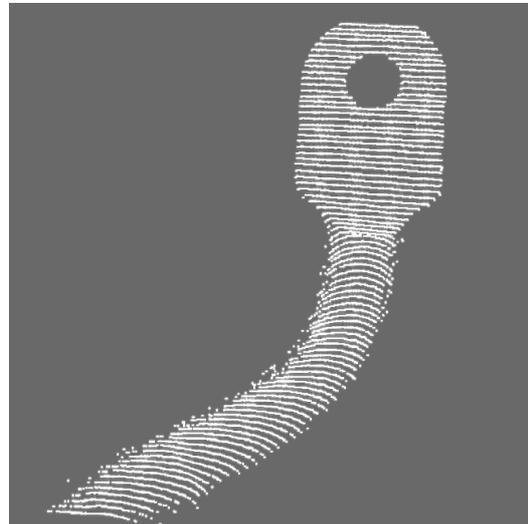


Abb. 2: Punktwolke eines Stabilisatorendes

diesem Ansatz ist, dass ein Umrüsten auf einen neuen Stabilisortyp lediglich durch eine Umprogrammierung der Bewegungsbahn realisiert werden kann. Zu beachten sind die Scanzeiten von einigen Sekunden und der Aufwand für das Verfahren des Triangulationssensors in mehreren Achsen. Die mit diesem Ansatz erreichten Genauigkeiten erfüllen die Anwenderanforderungen. Der zweite Ansatz geht von je einer Kamera und einem Dreifachlaserlinienprojektor an jedem Stabilisatorende aus. Mit diesem Aufbau kann sowohl das Bohrloch des Flansches im Durchlicht beobachtet, wie auch die Kamera/Projektor Kombination als Mehrfachlinientriangulator eingesetzt werden. Vorteil dieses Ansatzes sind sehr kurze Messzeiten sowie fest aneinander angebundene Geometrien der 2D-Messung und der Triangulation. Zu beachten sind die Auswahl der geeigneten Mehrfachlinienprojektoren (realisierte Winkel) und die rigide geometrische Anpassung an einen Stabilisortyp. Die Orientierungen der Beobach-

tungskameras sind so eingestellt, dass ihre optischen Achsen mit den Achsen der Bohrlöcher übereinstimmen. Auch mit diesem Systemansatz sind die Genauigkeitsvorgaben erfüllbar. Die Einfachheit, insbesondere die der Integration in eine Richtmaschine, sprach für die Realisierung genau dieses Ansatzes.

## BERECHNUNG DER RAUMPOSITIONEN

Die montagerelevante Raumposition der beiden Stabilisatorenenden sind durch die Mittelpunkte der Flanschbohrungen in  $x$ ,  $y$  und  $z$  sowie durch den „Aufbiegungswinkel“ **alpha** der Stabilisatorenenden sowie den Winkel deren Verschränkung **beta** bestimmt. Die Abbildung 3 zeigt ein Stabilisatorende aus zwei Sichtpositionen, wobei die jeweiligen Koordinaten und Winkel gekennzeichnet sind. Entsprechend dem gewähltem Ansatz werden  $x$  und  $y$  durch eine Approximation des Bohrlochmittelpunktes in einer (2D-)Durchlichtaufnahme bestimmt. Die  $z$ -Position des Bohrlochmittelpunktes liegt in der vorderen Flanschebene und ist über die Triangulation für die Position der oberen Laserlinie definiert. Der Winkel der Verschränkung **alpha** ergibt sich aus dem

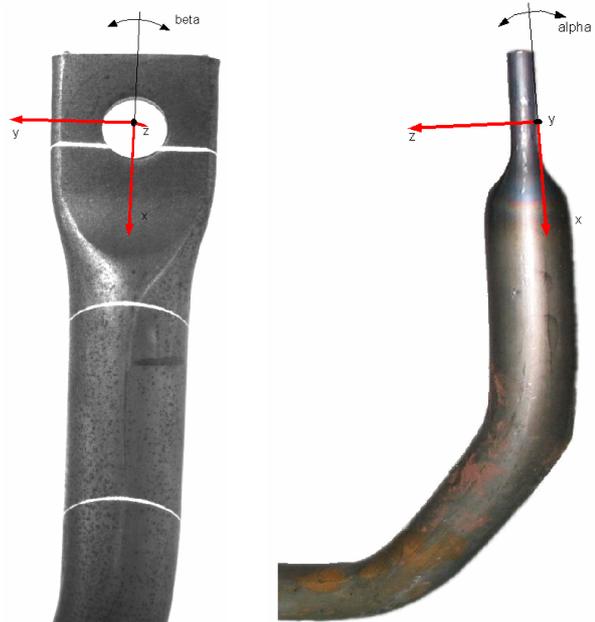


Abb. 3: Auf- und Seitensicht auf ein Flanschende des Stabilisators, Aufsicht mit 3 Laserlinien

Verlauf der zylindrischen Bereiche in der Durchlichtaufnahme. Der Winkel der Aufbiegung wird mit Hilfe der  $z$ -Positionen der unteren beiden Laserlinien gewonnen. Damit sind die für die Berechnung der Biegekorrektur notwendigen Werte komplett.

Das Messsystem ist in sich komplett kalibriert, sowohl hinsichtlich der Objektiv- und Kameraparameter wie auch hinsichtlich der drei realisierten Laserlinientriangulatoren. Für diese Kalibrierung wurden die Verfahren und Programmmodule aus [2], die auf [3] zurückgehen, genutzt und für den Mehrfachlinienansatz erweitert [4]. Bildverarbeitungstechnisch waren die sowohl im Labor wie auch am Aufstellort der Richtmaschine vorhandenen Nebenlichteffekte zu unterdrücken, um unverfälschte Messwerte zu erhalten. Geleistet wurde dies durch eine Schaltung der Laser, so dass sowohl Bilder mit wie auch ohne Laser gewonnen werden konnten und über Maskenbildung und Differenzbilder Spiegelungseffekte (vgl. Abb. 3) zu eliminieren waren. Die notwendige Binarisierung

erfolgte über eine adaptive Schwellwertbildung, um im Wesentlichen auf Spiegelungseffekte innerhalb der Bohrungszylinder zu reagieren. Die Homogenisierung der Linieninformation erfolgte über morphologische Filter [5]. Für die Detektion der Linien selber konnten die Suchbereiche stark eingeschränkt werden, da a priori bekannt war, dass die eintretenden Maßabweichungen der Stabilisatoren ein bekanntes Maß nicht überschreiten. Die Maßbestimmung selber erfolgt für  $x$  und  $y$  über eine Schätzung des Kreismittelpunktes der sichtbaren Bohrung, alle anderen Maße basieren auf einer subpixelgenauen Bestimmung der Positionen der einzelnen Laserlinienpunkte, bei  $z$  mit anschließender Linienapproximation, wobei hier zu Unterdrückung von Ausreißern ein schrittweiser Ansatz gewählt wurde.

Relevant für den praktischen Einsatz sind die erreichte Wiederholgenauigkeiten der Messungen sowie die Standardabweichungen der Messwerte. Sowohl mit einem Versuchsaufbau im Labor wie auch nach der Integration der Messsysteme in die Richtmaschine konnte über Messreihen nachgewiesen werden, dass mit dem ansatzgemäßen einfachen Messprinzip Wiederholgenauigkeiten, die bei 0,1 mm, meist aber wesentlich darunter, liegen, zu erreichen sind.

## EINBINDUNG IN DIE RICHTMASCHINE

Für die An- und Einbindung in die Richtmaschine waren neben den funktionellen Erfordernissen geometrische Restriktionen für die Positionierung wie auch die Vorgaben für die Kommunikation mit der Maschinensteuerung einzuhalten. Das wesentliche Integrationsproblem bestand aber in der Anbindung der Koordinatensysteme der beiden Messsysteme an das Koordinatensystem der Richtmaschine. Diese Anbindung war notwendig, da die Berechnung und Ausgabe der Werte für die

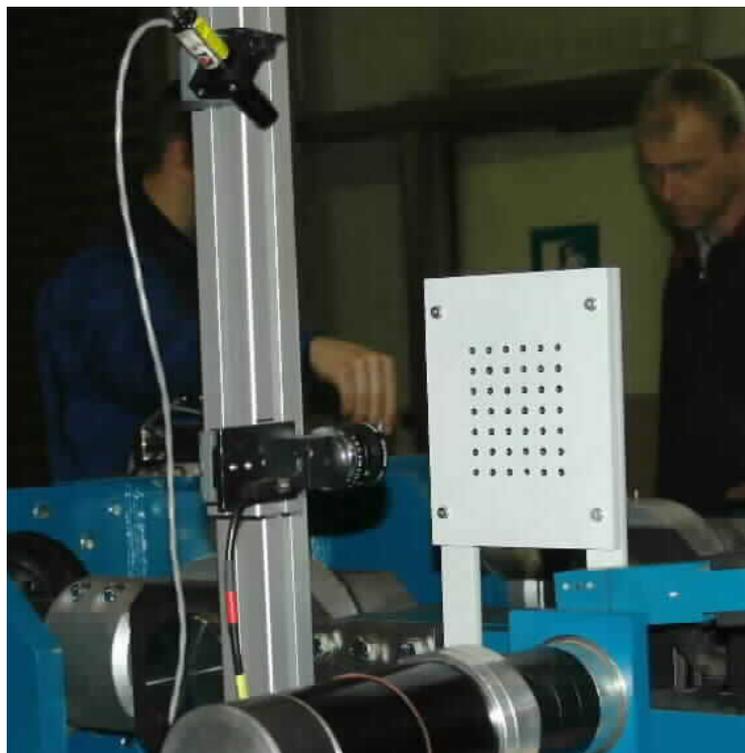


Abb. 4: Kamera, Laserlinienprojektor und Kalibrierplatte montiert in der Richtmaschine – eine Seite

Wege der Biegestempel im Koordinatensystem der Maschine erfolgen musste. Für die Anbindung der Koordinatensysteme wurden Kalibrierplatten nach [3] so adaptiert, dass ihre hochgenaue Fertigung, ihr Montage in der Richtmaschine wie auch das Einmessen ihrer Lage im Koordinatensystem der Maschine möglich wurden. Da Kamera und Lasertriangulation an die Koordinaten der Kalibrierplatte angebunden sind, die Lage der Kalibrierplatten im Maschinenkoordinatensystem bekannt ist, können alle Messwerte, die in Kamerakoordinaten gewonnen wurden, in Maschinenkoordinaten umgerechnet werden. Mit diesen Messwerten erfolgt durch externe Programme eine Simulationsrechnung, die im Resultat der Maschine die Wege für die Biegestempel vorgibt. Nach erfolgter Biegung übernimmt das Messsystem wieder die Prüfung des Ergebnisses. Das Diagramm in Abb. 5 veranschaulicht einen Prozess des „Aufbiegens“ der Stabilisatorschenkel. Auf der waagerechten Achse sind die Stellwege des Biegezyinders in mm aufgetragen, auf der senkrechten die erreichten „Verbiegungen“ am Flanschende, auch in mm. Relevant sind in den Kurven der horizontal verlaufende Abschnitt am Anfang, er ist durch die Elastizität des Materials bedingt, und der weitestgehende lineare Verlauf der beiden beeinflussten Größen (hier  $z$  und  $x$ ) im Bereich der plastischen Verformung. Die Inhomogenität im mittleren Teil der Anstiege (bei dem Stellweg 2,1 mm) ist durch den Versuch, zweimal mit den gleichen Stellparametern zu biegen, erklärt.

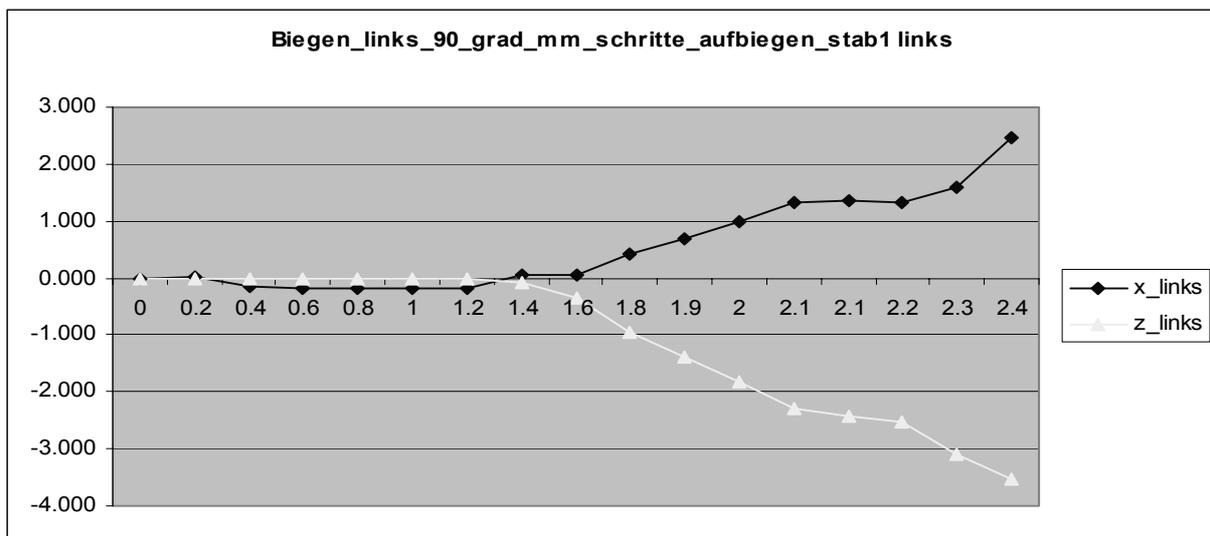


Abb. 5: Abhängigkeit der „Aufbiegung“ eines Stabilisators von den Stellwegen der Biegezyylinder

## ZUSAMMENFASSUNG

Vorgestellt wurde ein Verfahren für die Bestimmung von Raumpositionen (von Flanschbohrungen an beiden Enden von PKW-Stabilisatoren) durch Kombination von 2D-Bildverarbeitung und Mehrfachlinienlasertriangulation. Die Entwicklung erfolgte für die Ausrüstung einer Biegemaschine für das Richten von PKW-Stabilisatoren mit zwei Messsystemen. Vorhandene Ansätze der 2D-Bildverarbeitung wie auch der 3D-Vermessung wurden zielgerichtet erweitert und die gewonnenen Messwerte für eine Biegesimulation zusammengeführt. Das realisierte Gesamtkonzept erwies sich als robust und ausreichend flexibel, es brachte die erforderlichen Genauigkeiten im Bereich besser 0,1 mm.

### Referenzen:

- [1] "Automatisches System zum Vermessen und Richten von 3 dimensional gebogenen Massivrundstäben", Projekt Nr. KF0030704LF5, PRO INNO II, GFai, Berlin, 2005
- [2] Otto, M.: Entwicklung, Test und Optimierung eines Verfahrens für die Positions- und Orientierungsbestimmung von Betonfertigteilen, HUB, Institut f. Informatik, Berlin 2006
- [3] Zhang, Z.: A Flexible New Technque for Camera Calibration, <http://research.microsoft.com/~zhang/Papers/TR98-71.pdf>
- [4] Melitzki, M.: Entwicklung und Umsetzung eines bildbasierten Verfahrens zur räumlichen Vermessung von Kraftfahrzeugstabilisatoren, Diplomarbeit TFH Berlin, 2007
- [5] Sonka, M.; Hlavac, V.; Boyle, R.: Image Processing, Analysis and Machine Vision. Brooks/Cole, Pacific Grove, ISBN 0-534-95393-X, 1999

### Autoren:

Dipl.-Inf. Andreas Melitzki  
Prof. Dr. sc. techn. Gerd Stanke  
Dr. Frank Weckend  
Gesellschaft zur Förderung angewandter Informatik e. V.,  
Rudower Chaussee 30  
12489 Berlin  
Phone: +49 (0) 30 63921643  
Fax: +49 (0) 30 63921661  
E-mail: <name>@gfai.de

Frank Boochs, Christoph Raab, Rainer Schütze, Jörg Traiser, Holger Wirth

## 3D contour detection by means of a multi camera system

### ABSTRACT

In this paper we show a solution dedicated to measure spatial contours by means of multiple cameras. The contours allow to derive position and orientation, if necessary, of objects which have to be measured very precisely. The use of multiple cameras makes the system very flexible and robust. Set up, processing tasks and results of the system are presented.

### SET UP

The system has a solid optical head (cf. fig. 1), to be used as base for cameras and light source. This head can be mounted on a robot arm and then serves as effector which observes objects at predefined positions. This allows fast and flexible measurements of various spatial elements belonging to the object. But in general, the system could also be used in a stationary way, when numerous elements have to be measured under fixed conditions.

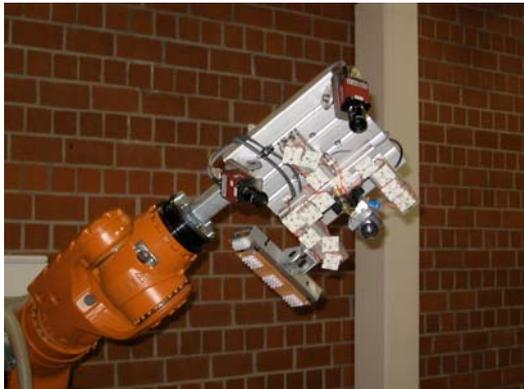


Figure 1 optical head

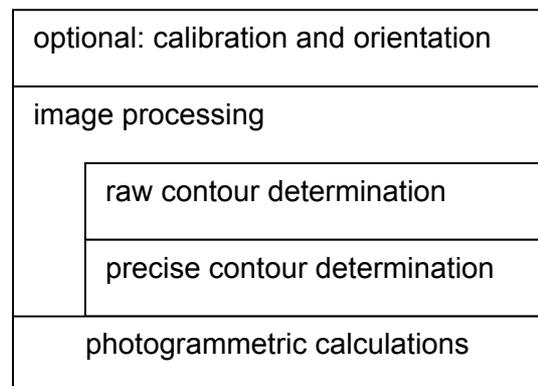


Figure 2 workflow of calculations

On the optical head at least 2 cameras have to be mounted, but this can be simply extended to more sensors, because robustness and accuracy then will profit. In the center of the head a light source is mounted which may project diffuse light but also can use special patterns if necessary to illuminate the object in a characteristic way.

In the first development stage we used 2 cameras in order to have the base for spatial calculations, but 3 cameras are more convenient due to check results more easily against blunder and the impact of viewing directions and provide a higher degree of

accuracy.

## Processing

As can be seen from fig. 2 the whole process consist of three steps. One of those only has to be used if necessary (calibration) and assures the geometrical quality of the system, the other ones have to be applied onto each image provide the final spatial information for each object observed.

### Camera Calibration and Orientation

A geometrical preprocessing of the system is necessary due it's flexibility with respect to the types of cameras used, to their mounting and orientation on the frame, which can be changed if necessary to adopt the configuration to the needs of the application. Furthermore possible forces (movements, thermal environment) may cause changes which have to be checked more or less frequently. The steps to be realized comprise a control of the internal camera geometry (calibration) and a check of the external relations on the frame (orientation).

A camera is calibrated when the focal length ( $c$ ), the principle point ( $H'$ ) and some distortion parameters ( $\Delta r$ ) are known. These parameters describing the camera properties are summarized in literature as the parameters of the interior orientation (IO) [1;2].

The determination of the parameter values for each camera is integrated into the workflow by matter of an automatic calibration procedure. The used calibration method allows a full automatic workflow to determine the IO parameters. Figure 3 shows the procedure of our calibration method.

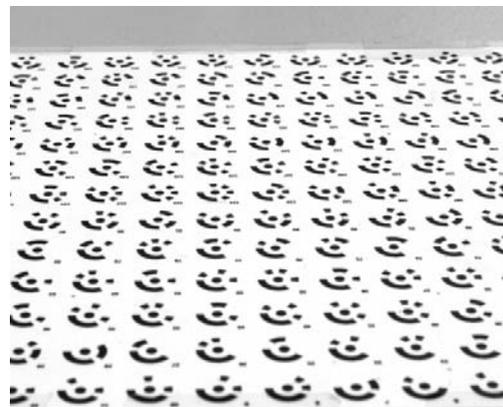
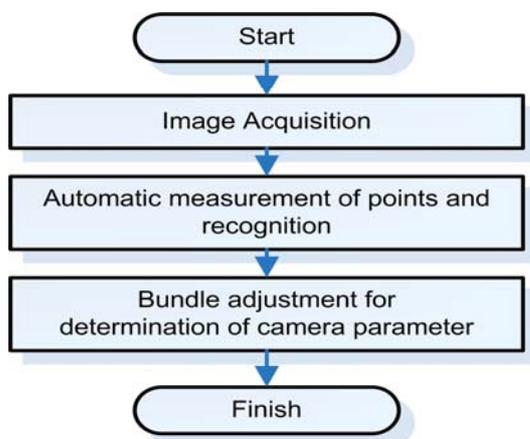


Figure 3 Procedure of the used calibration method      Figure 4 view onto point field used for calibration

About 32 images are acquired of the calibration panel (cf. Fig. 4) from different directions, inclinations and rotations of the camera around the optical axis. Within each image the center of all points are automatically measured, using the center of gravity (COG) method together with an analysis of the code ring [3] providing the point ID.

The measured image coordinates, the approximated object coordinates and the estimated exterior orientation [4] are brought together to calculate the parameters of the interior orientation with an adjustment library (AxOri [5]). The adjustment is free supported by the picture coordinates, the EO and the IO, whereas the object coordinates set the date (scale, rotation and translation) of the system.

In order to give an estimate of the geometrical quality of a calibration step Table 1 summarizes the root mean square values (RMS) of the object coordinates, the standard deviations of the EO and of course the standard deviations of the IO, for three AVT Marlin F-046B (b/w) cameras (pixel size 8.3 x 8.3 μm, 16mm objective). As can be seen, the results are sufficient to achieve precise results during the later measuring process.

|            | Camera (μm) |           |           | Image points (μm) |           | Object points (μm) |     |     |
|------------|-------------|-----------|-----------|-------------------|-----------|--------------------|-----|-----|
|            | c           | xH'       | yH'       | X                 | y         | X                  | Y   | Z   |
| Cam. A – C | 1.8 – 2.1   | 4.6 – 5.2 | 3.3 – 3.8 | 0.1 – 0.2         | 0.1 – 0.4 | 1.0                | 1.0 | 1.1 |

Table 1 Root mean squares of significant parameter of the IO bundle ([μm])

A calibration only should be necessary when cameras have been exchanged or modified, or other reasons for a change of their internal set up exist.

More often the external configuration might be changed, what then needs to check the geometrical configuration between the cameras or, if existing, in relation to an external coordinate system. This task uses the same strategy as with calibration, but allows a reduced number of images to be taken. Results of an orientation are show below:

|          | Image points (μm) |           | Object points (μm) |     |     |
|----------|-------------------|-----------|--------------------|-----|-----|
|          | X                 | y         | X                  | Y   | Z   |
| Cam. A-C | 0.1 – 0.2         | 0.1 – 0.2 | 1.0                | 2.5 | 3.6 |

Table 2 Root mean squares of significant parameter of the EO bundle ([μm])

**Contour Detection**

The calculation of the spatial contour is based on all images taken by a simultaneous shot of all cameras. These images are then individually processed in order to detect the 2D-contours of the object and are then combined in a photogrammetric calculation, which results in a spatial description of the object. Redundancy due to multiple images

allows to check the results and to eliminate blunder or regions where the contour is damaged.

For the 2D-contours it is assumed, that there exists a mathematical model, which can be used for a fitting process. This allows to process all contour elements with respect to this model giving a higher degree of accuracy. Actually elliptical contours are realized, but this can easily be substituted in case of other geometrical objects. The process of fitting is splitted in two parts. A raw determination of the 2D-contour (cf. Fig. 4) and a following exact determination by analyzing the gray scale value-profiles along the normal vectors of the contour (cf. Fig. 5).

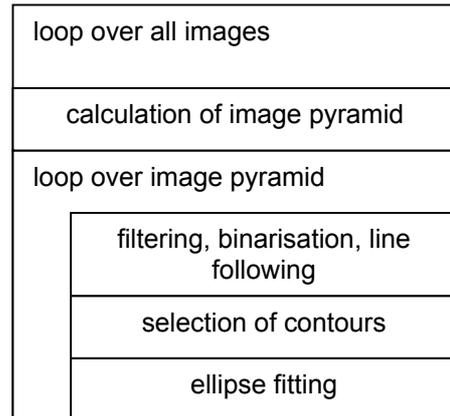


Figure 4 image processing (part1)

The **raw determination** consists of an edge detection by filtering, binarisation and line following process. The contours found during edge detection are evaluated and selected by means of geometric properties. All segments retained are used for an ellipse fitting. This process loops over an image pyramid allowing to reduce the impact of parasitic image elements appearing in the edge detection and increases the robustness.

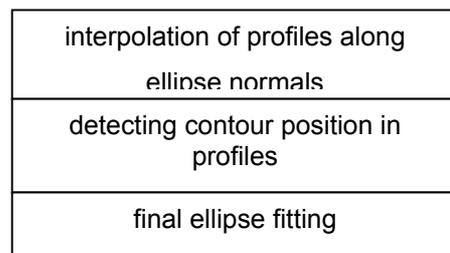


Figure 5 image processing (part2)

The **precise contour determination** happens only on the level of highest resolution. First, discrete gray scale values along local profiles normal to the contour are generated, in order to give a detailed local description of the transition from regions inside the contour to outside. In each profile the position of the maximal gradient close to the preliminary contour is determined, checked for plausibility and compared to the results of the neighborhood. The peaks found in all profiles are then used for a repeated ellipse fitting. The adjustment includes a blunder detection and finally provides the parameters of the elliptic contour in the image coordinate system.

As result of this process a parametric description of all contours is available and can be used for the final photogrammetric treatment of the data.

This is founded on the knowledge about the geometrical relationships between the images coming from the preceding calibration and

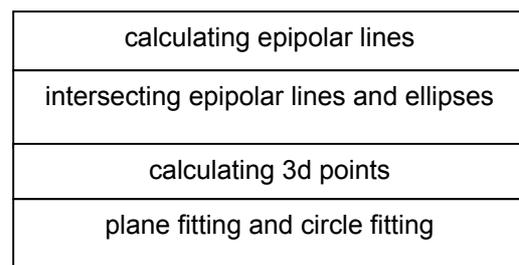


Figure 6 photogrammetric calculation

orientations steps. As the contours don't provide individual points to be used as input for spatial calculations these points are generated artificially using epipolar constraints. Starting with a first image pair, each contour is intersected with a number of epipolar lines for the respective image. In case of multiple images the points generated in one image are used to extend this process to all other contours. Each artificial point then has a number of counterparts – at least one – which then can be used to intersect the corresponding rays yielding a 3D point.

As matter of control all 3D-contour points are fitted into a plane, in order to check for deviations in the 2D-contours. If they are displaced, this results in changes of the distance to the spatial contour, what can be identified in the residuals of the plane. These residuals can be used to analyze erroneous contour parts and to reject points if necessary. As long as the plane gives acceptable results the spatial contour points can be fitted into the appropriate mathematical model, providing the final result.

## RESULTS

At the moment the processing chain is completely realized for the use of 2 images. Some minor steps concerning data control and blunder detection in case of more than two images have still to be implemented. We therefore present results of a two image situation.



Figure 7 test object shown in image A



Figure 8 detected contour plotted in image B

Figure 7 shows the object measured, Figure 8 displays the contour detected and Table 3 gives an idea of the quality. The example belongs to a car body and shows a hole used to fix a seat of a carriage.

From table 3 it can be deduced, that the quality achieved is satisfying. In relation to the size of an individual pixel the center of the contour object has been determined with a 40th of an image element, what is a satisfactory value.

|    |                             |                 |                             |                   |                |
|----|-----------------------------|-----------------|-----------------------------|-------------------|----------------|
| 2D | $\sigma$ ellipse of image A |                 | $\sigma$ ellipse of image B |                   | Cell size      |
|    | 0.003                       |                 | 0.003                       |                   |                |
| 3D | $\sigma$ plane              | $\sigma$ circle | $\sigma$ center x           | $\sigma$ center y | Pixel in space |
|    | 0.032                       | 0.021           | 0.003                       | 0.004             |                |

Table 3 standard deviations of one camera combination (images A and B); all values in mm

After integration of the remaining analysis steps concerning more than two cameras the system may serve as base for a flexible and reliable tool for the determination of spatial contours and their position and orientation in space.

#### References:

- [1] Luhmann T. (2003), "Nahbereichsphotogrammetrie - Grundlagen, Methoden und Anwendungen", 2nd Edition, Wichmann Verlag, Heidelberg, Germany, ISBN 3-87907-398-8
- [2] Remondino, F., Fraser, C. (2006), "Digital camera calibration methods: considerations and comparisons", International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVI, part 5, pp. 266-272. ISPRS Commission V Symposium, Dresden, Germany
- [3] Niederöst, M., Maas, H.-G. (1997), "Entwurf und Erkennung von codierten Zielmarken", Tagungsband der 16. Wissenschaftlich-Technischen Jahrestagung der Deutschen Gesellschaft für Photogrammetrie und Fernerkundung, Oldenburg 1996, Germany
- [4] Wrobel, B. (1999), "Minimum Solutions for Orientation" in "Calibration and Orientation of Cameras in Computer Vision", Editors Th. Huang/A. Gruen, Springer-Verlag
- [5] axios3d (2007), "Ax.Ori - Photogrammetrischen Bündelblockausgleichung", <http://www.axios3d.de/produkte/ori/> (Status 04-06-2007)
- [6] Demant, Streicher-Abel, Waszkewitz: Industrielle Bildverarbeitung: Wie optische Qualitätskontrolle wirklich funktioniert. Springer 2002

#### Authors:

Prof. Dr.-Ing. Frank Boochs  
Dipl.-Ing (FH) Christoph Raab  
Dipl.-Ing (FH) Rainer Schütze

Institut für Raumbezogene Informations- und Messtechnik, Holzstr. 36, 55116 Mainz  
Phone: + 49 6131 2859 666  
Fax: + 49 6131 2859 699  
Email: [i3mainz@geoinform.fh-mainz.de](mailto:i3mainz@geoinform.fh-mainz.de)

Dipl.-Ing Jörg Traiser  
Dipl.-Ing (FH) Holger Wirth

Metronom Automation GmbH, Max-Hufschmidt-Str. 4a, 55130 Mainz  
Phone: + 49 6131 250 83 85 - 0  
Fax: + 49 6131 250 83 85 - 8  
E-mail: [mail@metronom-automation.de](mailto:mail@metronom-automation.de)

M. Brandner

## Vision-Based Surface Inspection of Aeronautic Parts using Active Stereo

### Abstract

Dimensional inspection both during and after the manufacturing process is a valuable tool to identify defective parts and to isolate bad tolerance trends of the production line. Safety critical components as frequently encountered by the aeronautic industry require the inspection of 100% of the production leading to a demand for fast and reliable inspection tools. Vision-based inspection offers appealing properties such as a non-contacting measurement principle and the simultaneous measurement of multiple features on a given part. However, especially in the aeronautic sector the inspection system is often faced with difficult surface properties (e.g. non-coated shiny parts) and a large variety of object sizes ranging from small engine components to large fuselage elements. This work describes an optical sensor prototype based on two active stereo systems used to perform quality control measurements on small-scale aeronautic parts. Experimental results on aeronautic parts show the feasibility of the proposed approach.

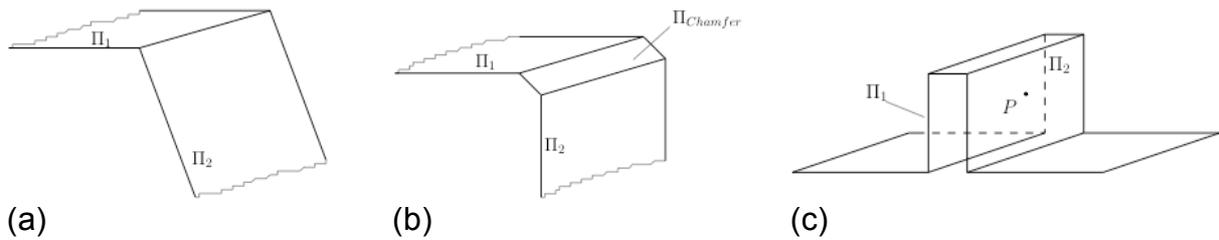
### Introduction

The detection of defective parts plays an important role in the production of aeronautic parts. Due to the high complexity and the demanded level of accuracy, inspection systems capable of performing highly accurate quality control measurements are required. Vision-based inspection systems have gained in importance especially for applications that require the inspection of a large number of geometric features within a short timeframe or for automated processing [1]. In contrast to standard mechanical inspection systems such as coordinate measurement systems and measurement arms, optical systems are only restricted by the line of sight.

The robustness of vision-based measurement systems in difficult situations such as changing ambient illumination or presence of highly reflective target surfaces can be improved by using a structured light approach. In this case, the illumination system is used to support the search for corresponding image features in different views of the scene.

Structured light can, however, be used to obtain active stereo setups once both the camera and the light projector are calibrated. Laser sources are frequently used in such applications as resultant projectors have small form-factors and are easy to integrate into the sensor design.

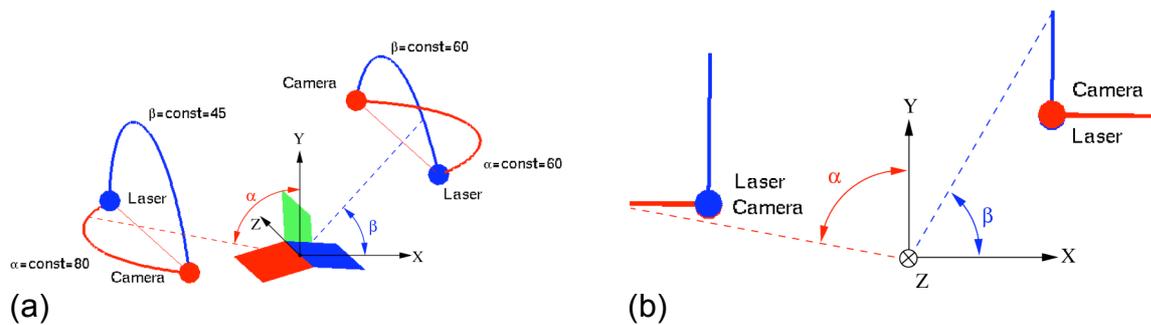
Vision-based inspection of aeronautic parts is confronted with a wide range of object scales. The development presented in this paper targets towards the robust measurement of the geometric entities *wall thickness*, *angle between planes*, and *edge between planes* as outlined in Figure 1. The proposed sensor has been developed as part of a larger project aiming at the improvement and adaptation of image processing tools to aid the quality control and assembly processes in the aeronautic sector.



**Fig. 1:** Target entities to be measured by the proposed sensor. (a) Angle between planes. (b) Edge between planes with chamfer. (c) Wall thickness.

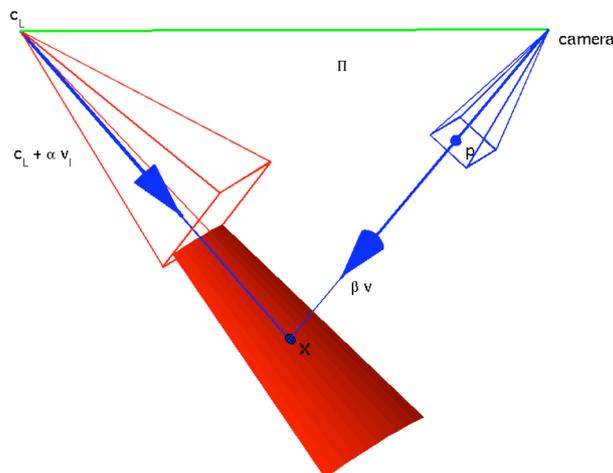
### Active Stereo Setup

A combination of two cameras and two laser projectors is used to capture object properties such as the parameters of planar surface patches, angles between adjacent surfaces, distances, and wall-thickness. The components of the proposed sensor are arranged as outlined in Figure 2. Simulations of different setups showed the maximum performance of the setup in terms of measurement uncertainty for  $90^\circ$  tilt angles between the components. The layout is further optimised so that each combination of an image sensor and a laser projector can act as active stereo pair.



**Fig. 2:** Geometry of the sensor components. (a) Two cameras (red dots) and two laser sources (blue dots) are placed relative to the sensor coordinate system. (b) Side-view of the setup.

We define an active stereo system as a geometric setup comprising an image sensor and a light source as outlined in Figure 3. Assuming calibrated components the unknown point  $\mathbf{X}$  can be reconstructed in 3D space using the geometry of the projected ray and the observation of the image sensor.



**Fig. 3:** 3D reconstruction using Active Stereo.

The laser projectors used in the sensor are equipped with line-generating lenses such that each laser projects a number of  $N=10 \dots 20$  parallel lines. In order to use these lines in the context of the active stereo a geometric model describing the surfaces projected by the laser is needed. Figure 4 depicts the deviation of the projected structures from their ideal straight lines. The different surfaces projected by the laser are modelled using their intersection with a plane orthogonal to the principle axis of the laser in a distance of 1m. These intersections are referred to as profiles and modelled by two polynomials where

$$s_i(t) = (p_{i,x}(t), p_{i,y}(t))^T \quad (1)$$

(cf. Figure 4b). Consequently, after substitution of a first and a second order polynomial, the  $i$ -th surface projected by the laser is represented by

$$S_i(t) = c_L + \alpha(t, a_2 t^2 + a_1, 1)^T \quad (2)$$

All measurements performed by the sensor rely on the estimation of planes or piecewise planar structures. As the reconstruction of points in 3D space and consequently the estimate of the plane parameters are straight forward to obtain using the projector model in Equation 2, we subsequently focus on the treatment of parameter uncertainty. The intersection of the  $i$ -th surface  $S_i$  with a planar surface patch  $\Pi_j$  can be reformulated to obtain a quadratic form [2]

$$p^T C_i(\Pi_j) p = (x, y, 1)^T C_i(\Pi_j) (x, y, 1) = 0 \quad (3)$$

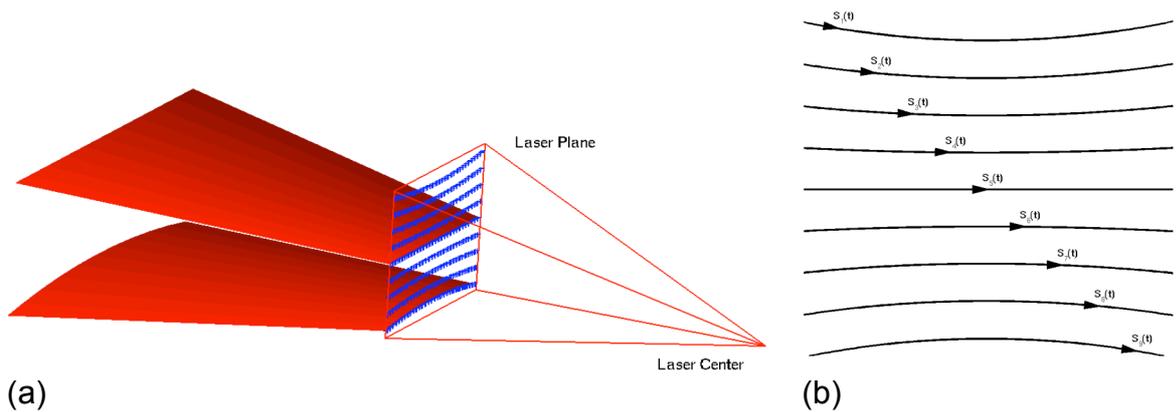
in image coordinates where the parameters of the  $3 \times 3$  matrix  $C_i$  capture the geometry of the setup relative to the planar patch  $\Pi_j$ . We apply the first order geometric distance measure  $d(p, C)$  [3] to obtain the Jacobian matrix

$$J(C_i, \Pi_j) = (J_{kl}(C_i, \Pi_j)) = \frac{\partial d(p_k, C_i(\Pi_j))}{\partial \Pi_{j,l}} \quad (4)$$

where  $k$  denotes the index of the point on the profile and  $\Pi_{j,l}$  is the  $l$ -th component of the  $j$ -th plane. Assuming additive Gaussian noise deviations of the laser line detector that are small w.r.t. the position of the line intersection, the vector comprising the standard deviations of the plane components is given by

$$\Sigma_{\Pi} = \sigma_d \sqrt{\text{diag}(J^T J)^{-1}} \quad (5)$$

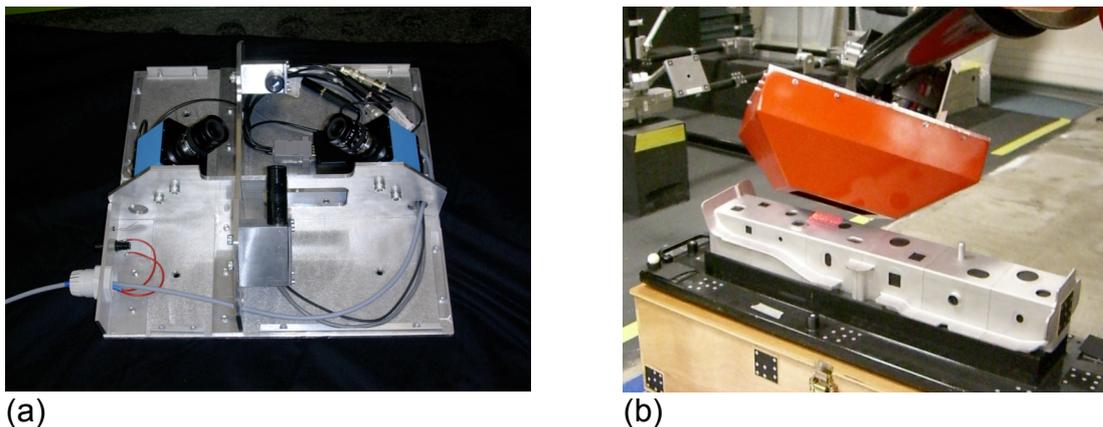
where  $\sigma_d$  denotes the standard deviation of the first order geometric distance  $d$ . Based on these results the following section reports on measurements performed with the proposed sensor geometry using an aeronautic part and a robot-mounted sensor.



**Fig. 4:** Geometric modelling of the laser profiles. (a) Surfaces projected by the laser. (b) Polynomial approximation of the laser profiles.

## Experiments

The geometry outlined in the previous section was used to obtain measurement results on an aeronautic part. Figure 5a outlines the optical sensor comprising two 1 MPixel cameras and two laser projectors. This sensor prototype is further equipped with a laser controller. The task of this component is to synchronise the laser illumination (both on/off and illumination power) to the image acquisition. 16 different illumination patterns each consisting of individually selectable laser powers can be stored on the controller and triggered using a remote command. Image acquisition is performed synchronously using the trigger inputs of the cameras. Prior to the measurement acquisition the sensor has been calibrated using a coordinate measurement machine. The calibration procedure includes the estimation of the intrinsic parameters of each camera, the stereo camera pair extrinsics, the parameters of the polynomials of each laser profile, and the laser extrinsics. Planar calibration targets are used to improve the quality of the parameter set of the whole sensor in a final non-linear optimisation step.

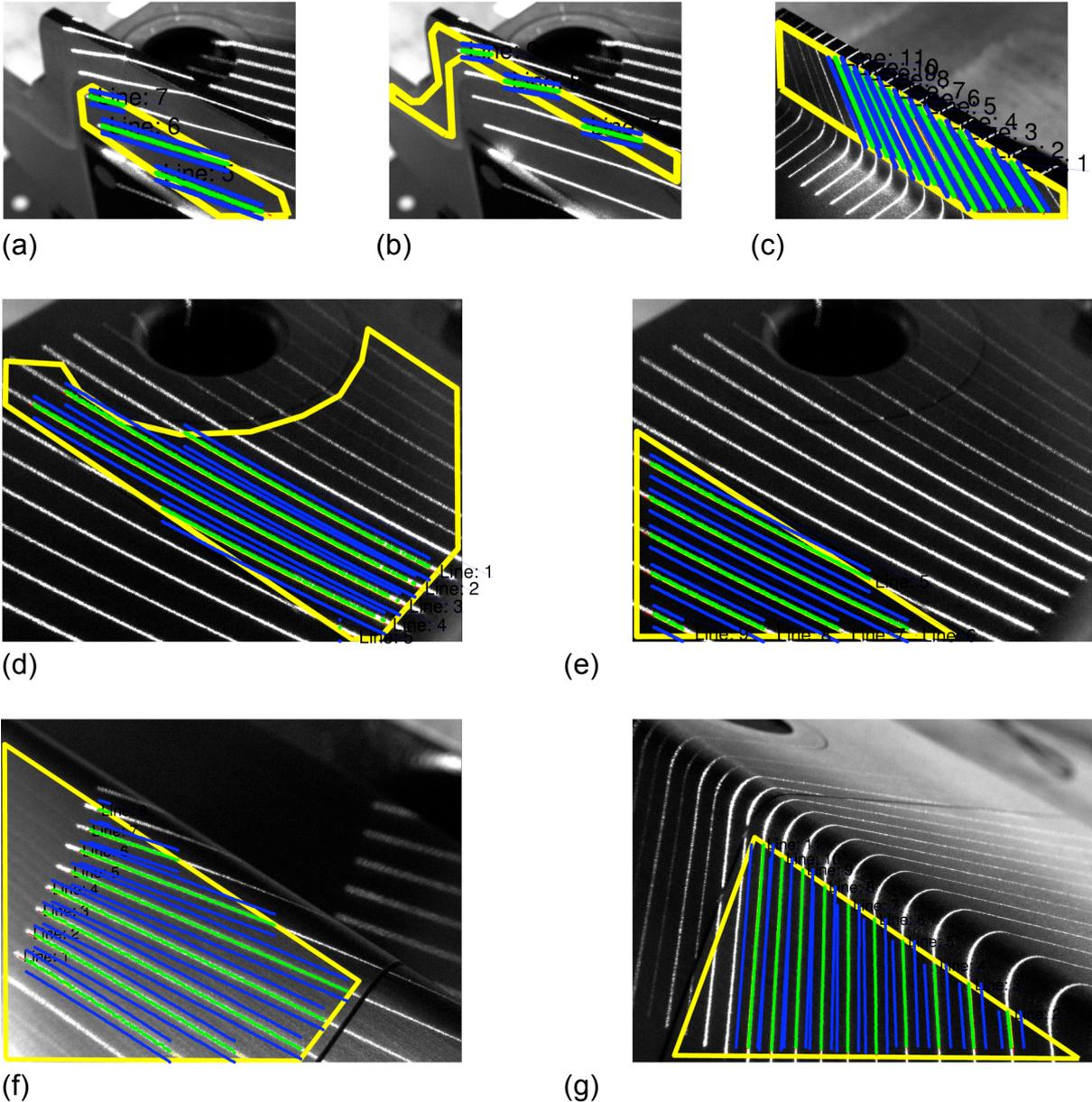


**Fig. 5:** Sensor prototype used for the experiments. (a) Two laser projectors (middle) and two cameras (left and right) are combined. (b) Sensor mounted on the robot arm.

Figure 5b shows the sensor mounted on a robot head. In this experiment the CAD model of the part is known. The position and orientation (i.e. the pose) of the sensor head is related to the robot parameters by means of a hand-eye calibration. Based on the CAD model and the sensor pose the regions of interest (ROI) are determined for both cameras. Measurements performed in this experiment include the thickness of a wall (measurement A), the step between two adjacent planar patches (measurement B), and the angle between two adjacent planes separated by an edge with chamfer (measurement

C). The results of the measurements are summarised in the following table. Note that the expanded uncertainty is computed based on  $k=3$ :

|                          |                  |
|--------------------------|------------------|
| A - Wall thickness       | 3.02+/-0.06 mm   |
| B - Normal distance      | 0.025+/-0.017 mm |
| C - Angle between planes | 91.29+/-0.18°    |



**Fig. 6:** Measurements taken on an aeronautic part. The yellow polygons depict the regions of interest used to restrict the correspondence search in each image. The green profiles are the detected laser profiles. (a-c) Three planar patches acquired during the measurement of wall thickness. (d-e) Measurement of the normal distance between two planes. (f-g) Angle between adjacent planes with chamfer.

## Summary

This paper presents the development of a close-range photogrammetric sensor for vision-based inspection of aeronautic parts. The sensor design is based on two active stereo sets each comprising a camera and a laser projector. The use of the active stereo system allows to robustly estimate plane parameters and derived measurands such as wall thickness of highly reflective surfaces even under partial occlusion. A geometric model for surfaces projected by the laser source is presented which allows for accurate calibration of standard laser projectors and line lenses.

Measurements performed with the calibrated sensor show the applicability of the proposed sensor design to vision-based inspection of aeronautic parts.

## Acknowledgements

The work presented in this paper was supported by the EC FP6 Project IPROMES (*Image PROcessing as METrological Solution*, AST3-CT-2004-502905). The contributing members of the project team were: Harald Ganster, Gert Holler, Daniel Hrach, Miguel Ribo, and Gerald Schweighofer.

## References:

- [1] M. Brandner and Th. Thurner, Uncertainty in Optical Measurement Applications: A Case Study, in IEEE Transactions on Instrumentation and Measurement 55 (2006), pp. 713—720
- [2] R. Hartley and A. Zisserman, Multiple View Geometry, 2<sup>nd</sup> Edition, Cambridge Press 2004
- [3] M. Harker and P. O’Leary, First Order Geometric Distance (The Myth of Sampsonus), in BMVC2006, pp. 87—97

## Author:

Markus Brandner  
Institute of Electrical Measurement and Measurement Signal Processing  
Graz University of Technology  
Kopernikusgasse 24/IV  
A8010 Graz, Austria  
Phone: +43 316 873 7773  
Fax: +43 316 873 7266  
E-mail: brandner@ieee.org

Dr. H. Lettenbauer / Dr. D. Weiss

## **X-ray image acquisition, processing and evaluation for CT-based dimensional metrology**

### **ABSTRACT**

X-ray computed tomography (CT) reconstructs an object from X-ray projection images and has long been used for qualitative investigation of internal structures in industrial applications. Recently, cone-beam CT has been adapted to the task of high-precision dimensional metrology of machined parts, providing a method of rapidly acquiring comprehensive and quantitative data on parts of arbitrary complexity.

High-power micro-focus X-ray tubes, large-size flat panel X-ray detectors, very accurate linear and rotary drives and a high-performance reconstruction solution are combined to image and measure a wide spectrum of parts manufactured from polymer and metal alloys according to established metrology standards.

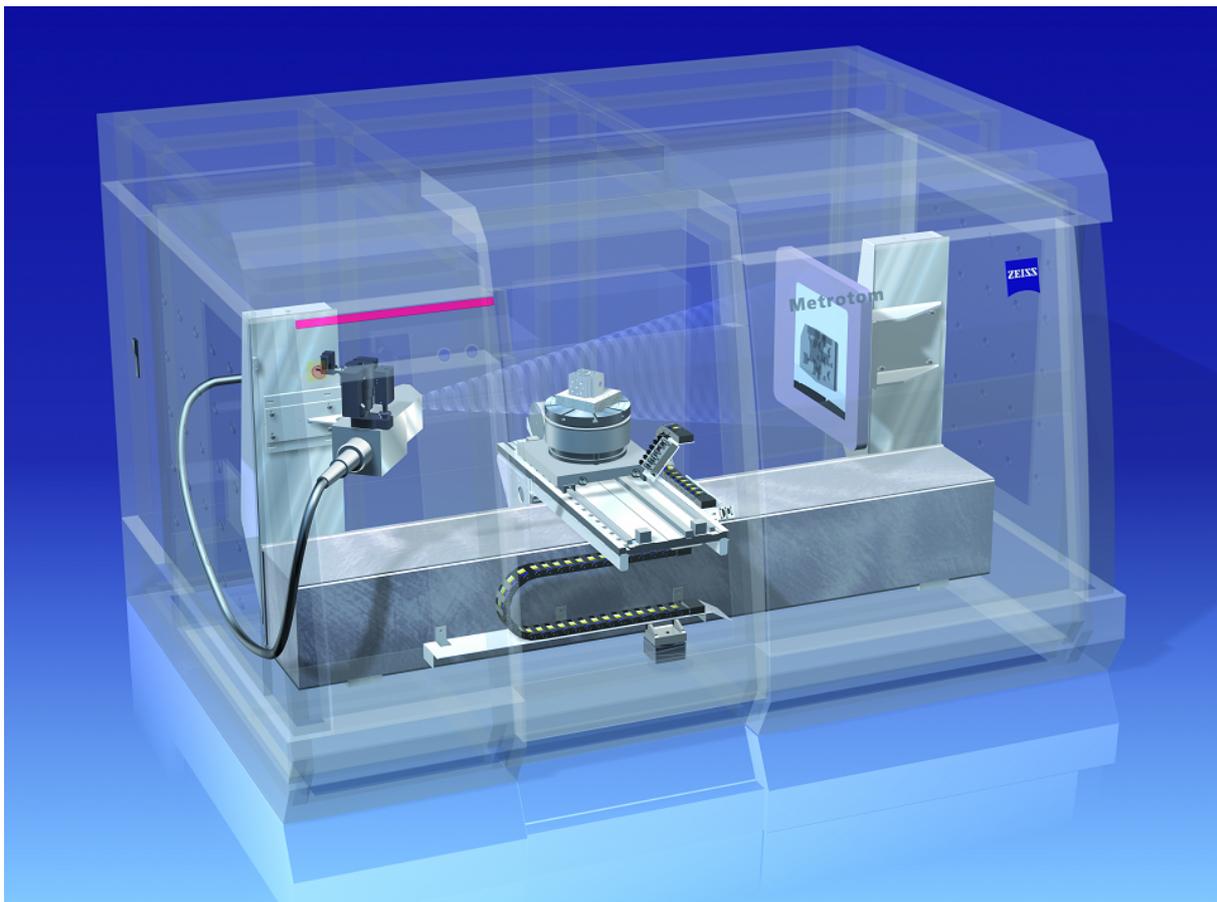
Various calibration techniques are used at the several stages of the measurement process to e.g. characterize the behavior of X-ray tube and detector and precisely define the imaging geometry.

To verify the measurement accuracy, suitable objects such as geometrical elements are calibrated according to international standards and then measured using the CT. Dimensional metrology is performed on the reconstructed objects, either directly or using an intermediate surface-extraction step. The results of these measurements are compared to the reference values of the calibrated objects; the level of agreement of the results defines the accuracy of the CT measurement. Using established methods to define measurement uncertainty ensures a high level of acceptance with dimensional metrology users.

## INTRODUCTION

From an image-processing point of view, computed tomography is based on exploiting the known geometrical configuration of source, object and detector to obtain a numerical representation of the object. For cone-beam X-ray CT, image formation is based on a perspective projection of the object density (which is encoded as X-ray intensity). For metrology purposes, the resulting object representation needs to reproduce the object surface with an accuracy that is much better than the voxel resolution. This accuracy cannot be assessed from the volumetric data with the naked eye. It is in this regard of quantitative evaluation with sub-voxel precision that metrology-capable computer tomographs differ most importantly from their predecessors that were aimed at non-destructive testing (NDT) and a mostly qualitative evaluation of volumetric data.

## COMPONENTS



A metrology-enabled computer tomograph such as the METROTOM (Carl Zeiss) employs the same components as a classical NDT CT: X-ray source, object

manipulator, and detector. A micro-focus X-ray tube is used to provide sharp X-ray projection images even at high magnification, enabling a sharp reconstruction showing maximum detail at a given voxel resolution. At the same time, the X-ray source needs to provide sufficient intensity to enable short image acquisition times (0.5 – 1 s) and thus acceptable CT measurement times. A source size much below the voxel resolution is usually not desirable, since the corresponding high resolution X-ray image cannot be adequately captured (sampled) and the image resolution is therefore not transferred to the volumetric data. Instead it is much more important to maintain a stable source position during the CT measurement, or if thermal, mechanical or electrical drift prevents this, to accurately track the source position and compensate for the changes. Otherwise even small changes of the source position of a few dozen micrometers are sufficient to invalidate high-precision metrology results.

Another key component is a manipulator capable of positioning the object very exactly with regard to translational and rotational movement. It employs the methods of computer-aided accuracy that have been developed for coordinate measurement machines (CMMs). A rotary table with an axial run-out of below 0.1  $\mu\text{m}$  ensures a highly stable and reproducible rotation.

For the flat-panel area detector, size and number of pixels are the key properties determining what size objects can be measured and what voxel resolution can be reached. For metrology, it is also very important to use a detector that shows maximum geometric fidelity, since image distortions are transferred directly to the volumetric data. Reconstructing the object from the projection images must be done with an accuracy permitting metrology at below 1/10 of the voxel resolution, and the reconstruction process should be finished concurrent with the acquisition of the projection images. The time-consuming “backprojection step” of the reconstruction can be executed using a cluster of PCs operating in parallel, or it can be accomplished with dedicated hardware.

## **CALIBRATION**

Large-area flat-panel X-ray detectors are produced in small volumes and suffer from some defects that are not known, or occur to a much smaller degree, in technologically similar products such as TFT monitors. E.g., it is usual to accept several complete rows or columns of defect detector pixels per detector, in addition to a large number of isolated defect pixels (up to 2% of the total number of pixels), which is unthinkable for a

display device or a CCD or CMOS imager. Due to the high radiation exposure of the detector electronics during operation, new defect pixels are spontaneously created and must be accounted for.

It is therefore established practice to regularly perform automatic searches for defect detector pixels using several combinations of source-detector parameters and suitable image processing. In addition, the detector pixels exhibit individual dark currents (signal acquired in the absence of X-ray intensity) and gain factors that must also be compensated when using the images for computed tomography.

The combination of point-like X-ray source and area detector is called cone-beam geometry. For this geometry, a method of processing the X-ray images in order to reconstruct the object published by Feldkamp et al [1] has been adopted by most users. This method is very sensitive to a horizontal misalignment of source, object rotation axis and detector, therefore a regular calibration procedure is executed, usually involving a wire made from high-absorption material such as tungsten, in order to determine the horizontal alignment state.

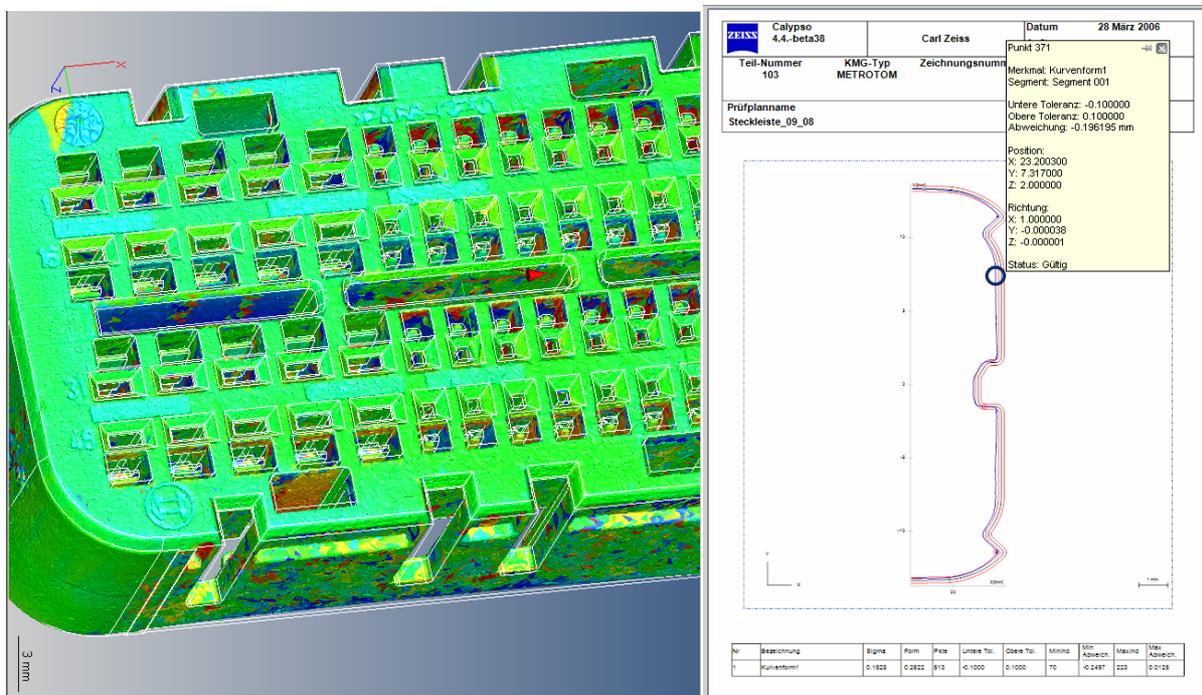
To enable a CT setup for metrology, additional calibrations are necessary. The combination of X-ray source, object and detector can be completely characterized using 9 parameters: the distance from point source to detector plane, the pixel coordinates of the normal projection of the source onto the projection, and the 6 parameters (3 rotation + 3 translation) giving the current object pose. All 9 parameters must be known with a high degree of accuracy to enable metrology substantially below the voxel resolution. For example, the voxel pitch of the volumetric data is usually chosen as a function of detector pixel pitch and magnification:

$$pitch_{\text{voxel}} = pitch_{\text{pixel}} \times \frac{d_{\text{source-rotation\_axis}}}{d_{\text{source-detector}}}$$

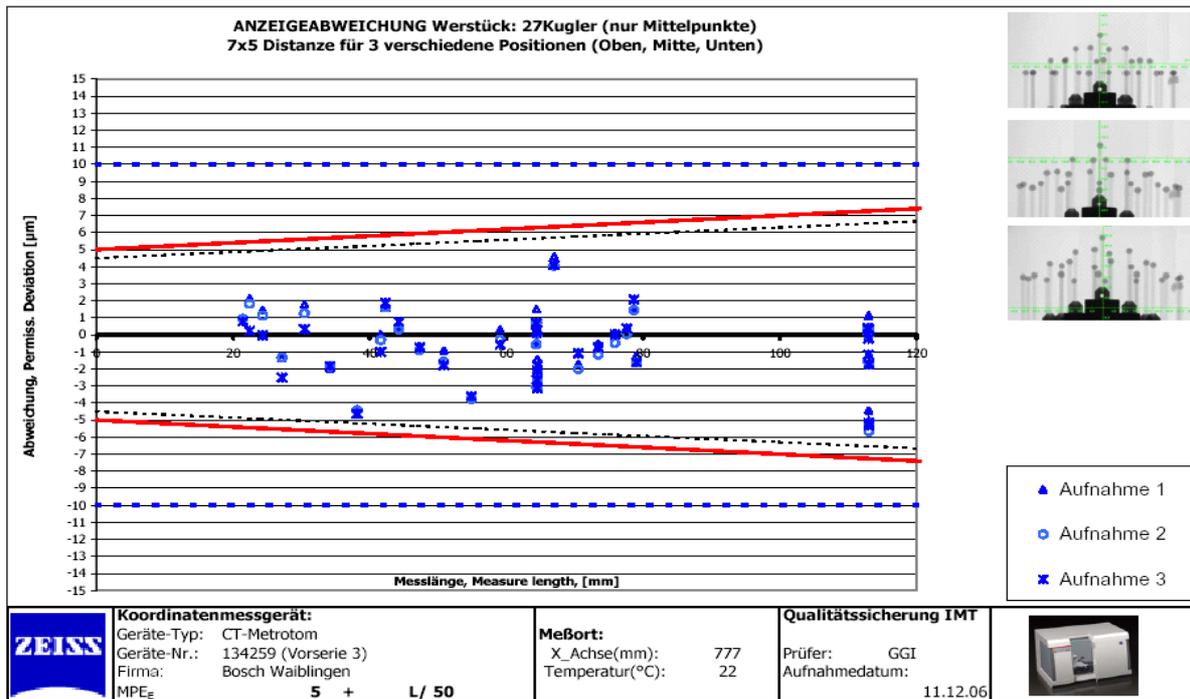
As can be seen, if the true values of the distances between source and object rotation axis resp. detector differ from the values used in the reconstruction, a wrong voxel pitch results and leads to a scaling error of the volumetric data. In the METROTOM, the two distances are known with a relative error of at most  $5 \cdot 10^{-5}$  to enable high-precision metrology.

## METROLOGY AND VERIFICATION

Metrological evaluation of the volumetric data created by the CT is performed with the CALYPSO software. Because of the wide-band energy spectrum of the X-ray source, the volumetric data may display a locally varying level of contrast, and simple segmentation using a global gray-value threshold value is not sufficient for precision metrology. Therefore, CALYPSO employs advanced edge-finding algorithms to extract the object surface locally. Based on the CAD model of a part, the user creates an inspection plan by defining all geometric elements of interest. From this plan the software determines locations and directions of virtual probing in the CAD model, transforms them to the volumetric data, and extracts the surface points.



Once the surface points have been extracted from the volumetric data, CALYPSO offers many possibilities of visualizing the measurement results, such as color-coded model views or curve plots comparing nominal and actual surface positions.



The figure shows the CT measurement errors for a calibration object consisting of several balls, thus realizing a large number of possible measurement distances at different orientations (see inset at upper right with X-ray images). Distances between ball centers are determined from the volumetric data using the CALYPSO software and compared to values determined with a high-precision tactile-probe CMM. The difference is plotted against the nominal ball center distance and shows a deviation of a few microns between the two measurement machines, satisfying a maximum permissible error (MPE) specification for the CT of  $(5 + L/(50 \text{ mm}))$  microns, where L is the measurement length. The fact that the difference plot only shows a very small trend indicates that the X-ray source location, and thus the scale of the volumetric data, has been determined very precisely.

**References:**

[1] L. A. Feldkamp, L. C. Davis, J. W. Kress, Practical cone-beam algorithm, J. Opt. Soc. Am. A 1, 612 - 619 (1984).

**Authors:**

Dr. Hubert Lettenbauer  
 Dr. Daniel Weiss  
 Carl Zeiss IMT GmbH, Carl-Zeiss-Str. 22  
 73447, Oberkochen  
 Phone: +49 7364 20 8996  
 Fax: +49 7364 20 9111  
 E-mail: [h.lettenbauer@zeiss.de](mailto:h.lettenbauer@zeiss.de)

K. Sickel / V. Daum / J. Hornegger

## **Shortest Path Search with Constraints on Surface Models of in-the-ear Hearing Aids**

### **Abstract**

One part of the production of modern hearing aids is the integration of a ventilation tube. This is usually done by applying a shortest path search algorithm on a non-convex ear impression. For the functionality of this so-called vent the path search algorithm has to fulfill several constraints like smoothness and space consumption of the vent. The proposed solution is based upon a modified version of Dijkstra's algorithm to address the constraints and to increase accuracy. The algorithm was evaluated by experts and validated to perform better for the application than a shortest path solution.

### **1. Introduction**

Approximately eight percent of the world population suffers from hearing loss. This handicap impairs the life quality of persons concerned and can lead to social isolation. Hearing aids improve the life quality of these approximately 560 million people in the world [1]. Modern hearing aids can be categorized into two major classes: behind-the-ear (BTE) devices and in-the-ear (ITE) devices. The ITE devices, especially the completely in the canal (CIC) devices, are almost invisible to other people [2]. The major disadvantage of the ITE devices is that the production is complex, expensive and difficult. The production process starts with an ear impression of the patient made by an audiologist. This impression is scanned and the resulting surface mesh is smoothed and trimmed to fit the requirements of an ITE device. The physical impression is then built by a stereolithography machine and outfitted with the necessary components such as receiver and microphone. After some finishing tasks, like varnishing and polishing, the device can be sent to the customer. Due to the fact that an ITE closes the ear no pressure equalization is possible. To avoid this disagreeable effect a ventilation tube is integrated into the device. For a good air flow it is required that the vent is preferably short and smooth. The vent placement is currently done during the virtual editing of the

scanned impression (shell) by using a shortest path algorithm.

Figure 1(a) shows an example of the ventilation tube. The computed path is not always favorable for the manufacturing process. In an ITE only a small amount of space can be used to fit in the rigid electronic components such as receiver, microphone and battery. To ensure the tiniest possible device the vent should be placed in a way that there is enough space for the other components. The usable space is approximated by circles fitted into slices of the impression which corresponds approximately to the shapes of the electronic components, see Figure 2. Altogether the production process enforces the following constraints on a new path search algorithm.

1. The vent shall be short.
2. The vent shall consume only a small amount of usable space inside the shell.
3. The vent shall avoid sharp bends to ensure a good air flow.
4. The computation time of the vent path shall be real time (< 2 sec).

## 2. Preliminaries

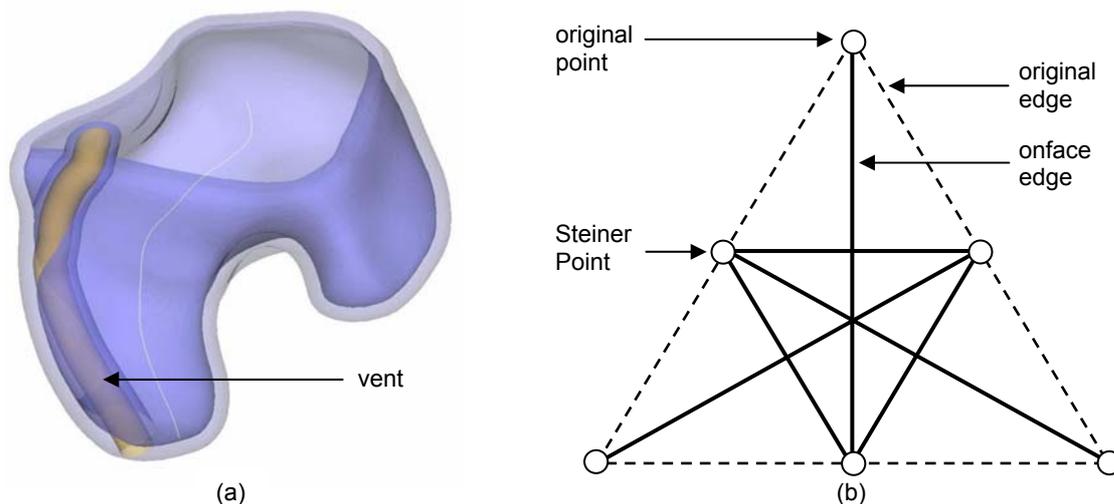
The shortest path problem is well-known in computational geometry. Accordingly there exist a large amount of algorithms to solve this particular problem. These algorithms can be grouped into exact and approximate algorithms, see e.g. [3] for an extensive overview. Due to run-time constraints it is favorable to concentrate on approximate algorithms, which are in general faster than the exact ones. We used Dijkstra's algorithm (DA) [4] as starting point for further investigations, because it has an  $O(n \log n)$  complexity and can handle non-convex data. Additionally its objective function, which is the Euclidean length between the path nodes, can be easily adapted to additional requirements. A drawback of the DA is that the path length depends on the mesh granularity. To avoid this problem the Selective Refinement (SR) of Kanai and Suzuki [5] was employed and also modified. The SR introduces so-called Steiner Points into the mesh, see Figure 1(b). These additional points allow the path search to go across triangles by using so-called onface edges. Kanai and Suzuki showed in [5] that the computed path is on average only 0.4% longer than the exact path. The algorithm by Kanai and Suzuki uses the DA iteratively. In each iteration step Steiner Points are introduced and the search graph is reduced to the vertices in the current path and the neighbors of these vertices. In the first iteration the whole mesh is used. To improve the run-time this was modified such that in the first iteration the DA is used and starting with the second iteration Steiner Points are introduced. The algorithm terminates if either the

difference between current and last path is smaller than a threshold or a fixed maximum number of iterations is reached. The number  $s$  of Steiner Points per edge is chosen to depend on the edge length  $|e|$  and an user given value  $\gamma$  according to the formula:

$$s = \lfloor |e|/\gamma \rfloor - 1 \quad (2.1)$$

### 3. Extension of the basis algorithm

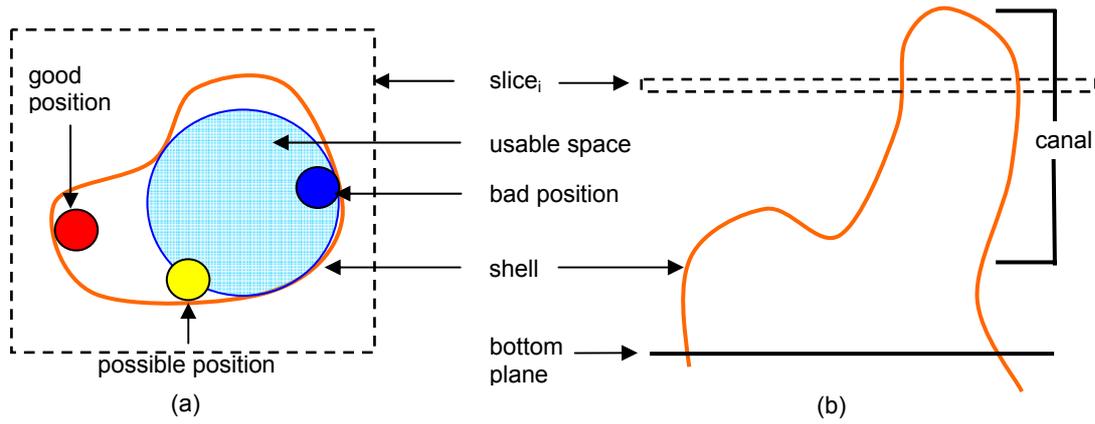
The basis algorithm (DA+SR) fulfills the first and the last requirements of the given problem, but does not consider the space consumption and strong bends in the path. Especially the consumption of usable space is crucial for the production of small hearing devices and avoiding strong bends assures the correct functionality of the vent. To take these two points into account the so-called Volume Measure (VM) and Curvature Measure (CM) are introduced.



**Figure 1.** (a) Ear impression with ventilation tube in editing software. (b) Introducing of Steiner Points into a triangle mesh.

#### 3.1 Volume Measure

The aim of the VM is to optimize the space consumption of the vent. To approximate the usable space in the shell we assume that a circle is the ideal representation for the space consumption of the electronic components. These circles are fitted into slices which are approximately perpendicular to the canal of the shell, see Figure 2. The number of slices depends on the number of vertices in the mesh. To incorporate the space usage of this circular region by the vent a weight  $\omega_{VM}$  is introduced. To compute  $\omega_{VM,i}$  at each vertex  $v_i$  a virtual vent is positioned, see Figure 2(a). If the vent cannot be



**Figure 2.** (a) Slice of the shell with example vents in a good, possible and bad position. (b) The impression is sliced approximately perpendicular to the centerline. The number of slices depends on the number of vertices in the impression.

fitted, a penalty term is added to  $\omega_{VM,i}$ . Second part of  $\omega_{VM,i}$  is the overlapping distance between virtual vent and inner circle. A visual depiction of  $\omega_{VM}$  can be seen in Figure 3, red is  $\omega_{VM,i} = 0.0$  and blue is  $\omega_{VM,i} \geq 1.0$ .

$$f_{i,j} = \omega_{L,i,j} + \beta \cdot \omega_{L,i,j} \cdot \omega_{VM,j} \quad (3.1)$$

The modified object function  $f_{i,j}$  combines the Euclidean length  $\omega_{L,i,j}$  between vertex  $v_i$  and vertex  $v_j$  with the VM. The VM is scaled with the edge length  $\omega_{L,i,j}$ . The variable  $\beta$  allows to control the influence of each part. We calculated the VM in a preprocessing step. Thus it can be examined and optimized independently of the other algorithms. The additional weights for the added Steiner Points are calculated via linear interpolation.

### 3.2 Curvature Measure

Due to the shape of the ear impressions it can happen that the vent has strong bends. It is also possible that the VM introduces bends in order to avoid penalty regions. To address these problems the curvature of the path is taken into account. Since DA is based on graph nodes (vertices) the incorporation of the CM requires the extension of the nodes to tuples of vertices. Each vertex can be combined with each neighbor thus one vertex with  $N$  neighbors results in  $N(N-1)/2$  tuples. The curvature is approximated by an angle  $\delta$  enclosed by three neighboring vertices. To avoid scaling artifacts it is, similar to the VM, scaled with the Euclidean length of the tuple. During the graph search only tuples with a sharing edge can be connected to a path. The additional weight is calculated as  $\omega_{CM,i} = |180-\delta_i|$  and integrated into the objective function.

$$f_{i,j} = \omega_{L,i,j} + \beta \cdot \omega_{L,i,j} \cdot \omega_{VM,j} + \mu \cdot \omega_{L,i,j} \cdot \omega_{CM,j} \quad (3.2)$$

The Euclidean length  $\omega_{L,i,j}$  is in this case the length of the tuple.

#### 4. Results

The algorithm was tested on an Intel Core 2 Duo 2GHz CPU with 2GB RAM. The test data had a range of 2500 – 20000 vertices and 5000 – 40000 faces. Table 1 shows the results for the length of the path, the VM costs  $c_{VM}$  and the CM costs  $c_{CM}$ .

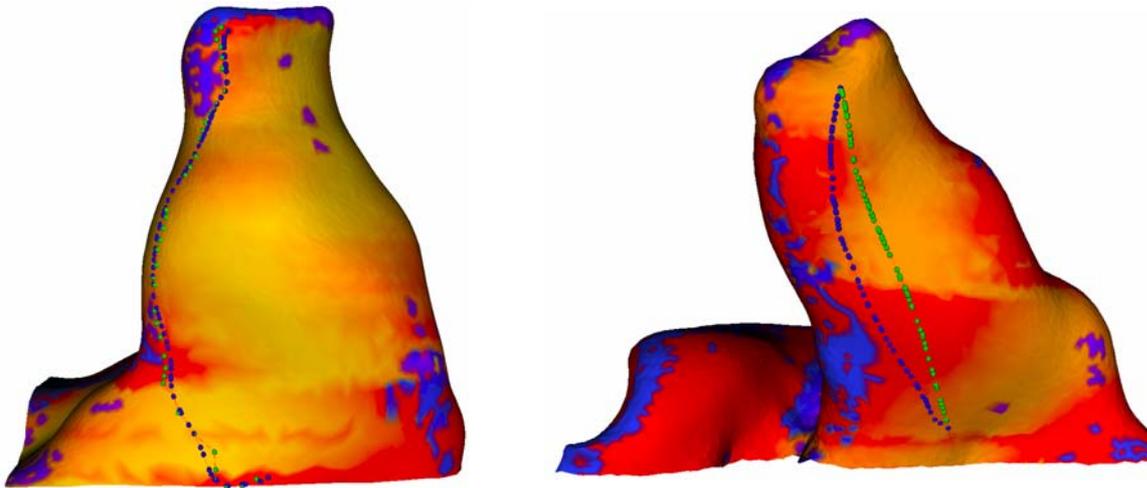
$$c_{VM} = \sum_{i=0}^L \omega_{VM,i} \quad (4.1)$$

$$c_{CM} = \sum_{i=0}^L \omega_{L,i} \cdot \omega_{CM,i} \quad (4.2)$$

L denotes the number of vertices respectively tuples of the path.

| Algorithm combination | $c_{VM}$ | $c_{CM}$ | Path length |
|-----------------------|----------|----------|-------------|
| DA and SR             | 36.5     | 5.7      | 0.833       |
| DA, SR and VM         | 22.6     | 8.2      | 0.837       |
| DA, SR and CM         | 46.5     | 4.0      | 0.930       |
| DA, SR, VM and CM     | 37.5     | 3.9      | 0.92        |

**Table 1.** Results of the different algorithm combinations. The used impression had about 6000 vertices.



**Figure 3.** Example paths for (left) DA, VM, CM (green) in comparison with DA, SR, VM (blue) and (right) DA, SR (green) in comparison with DA, SR, VM (blue).

The costs in table 1 indicate that the VM significantly reduces the intersection between vent and inner circle and so optimizes the available space for the electronic components. The needed preprocessing time for the VM computation is less than a second for meshes with less than 10000 vertices. Also visible is the bad influence of the VM to the

CM costs which can be strongly reduced by using both extensions. In Figure 3(a) it can be seen that the paths favors the red regions of the surface and thus maximize the remaining space. Figure 3(b) shows the positive influence of the CM. The path is longer than a shortest path but optimizes the VM costs without introducing unfavorable bends. Table 2 compares the run times of the algorithm combinations, which reveals that the VM computation is fast. The CM computation with its tuple creation, however, needs further optimization.

| Algorithm combination   | Run time (6000 vertices) | Run time (12000 vertices) |
|-------------------------|--------------------------|---------------------------|
| Dijkstra, SR and VM     | 1.4 sec                  | 2,5 sec                   |
| Dijkstra, SR and CM     | 86.1 sec                 | 269,1 sec                 |
| Dijkstra, SR, VM and CM | 86.5 sec                 | 283,3 sec                 |

**Table 2.** Results of the different algorithm combinations.

## 5. Conclusions

The developed extensions helped to address the open requirements to optimize the usable space consumption and to avoid sharp bends. Especially the VM works well and does not require that much overhead, because it can be done fast in a pre-processing step. The CM achieves an improvement for the path but violates the real-time requirement. It needs further research on how to prune the number of tuples and optimize the tuple-generation algorithm. The resulting paths, which were validated by experts, are superior to a simple shortest path solution.

### References:

- [1] hear-it.org. Es gibt immer mehr Hörgeschädigte. <http://www.german.hear-it.org/page.dsp?forside=yes&area=134>, 1<sup>st</sup> April 2007.
- [2] National Institute on Deafness and Other Communication Disorders. Hearing Aids [NIDCD Health Information]. <http://www.nidcd.nih.gov/health/hearing/hearingaid.asp>, 1<sup>st</sup> April 2007.
- [3] J.S.B Mitchell. Geometric shortest paths and network optimization. In J.-R. Sack and J. Urrutia, editors, Handbook of computational geometry, pages 633-702. Elsevier Science, Amsterdam, 2000.
- [4] E. W. Dijkstra. A note on two problems in connexion with graphs. Numerische Mathematik, 1:269-271, 1959.
- [5] T. Kanai and H. Suzuki. Approximate shortest path on polyhedral surface based on selective refinement of the discrete graph and its applications. In GMP '00: Proceedings of the Geometric Modeling and Processing 2000, page 241, Washington, DC, USA, 2000, IEEE Computer Society.

### Authors:

Konrad Sickel  
 Volker Daum  
 Joachim Hornegger  
 Friedrich-Alexander-Universität Erlangen-Nürnberg, Lehrstuhl für Informatik 5 (Mustererkennung).  
 Martensstraße 3  
 91058 Erlangen  
 Phone: +49 9131 85 27775  
 Fax: +49 9131 303811  
 E-mail: konrad.sickel@informatik.uni-erlangen.de

S. Husung / G. Höhne / C. Weber

## **Efficient Use of Stereoscopic Projection for the Interactive Visualisation of Technical Products and Processes**

### **Introduction**

Today, product development dominates manufacturing time and cost, which are often also contradictory to high standards of quality. This is particularly a challenge for the development of micro-mechanical components and systems, where tight dimensional and form tolerances are important. The use of computer based tools enables the representation of product and process properties and supports their optimisation. One of the goals for the representation of the properties is the use of the *Virtual Reality* technology - a multimodal human-computer-interface with stereoscopic projection for spatial visualisation of 3D-objects. Therewith an immersion in the virtual scene becomes possible. Furthermore this enables an easy comprehension and understanding of complex contexts and relations, which are usually only evident for experts. By using intuitive interaction tools the user can navigate in the VR scene, manipulate and investigate the scene content.

### **VR system at Competence Centre “Virtual Reality”**

In 2006 a new VR system was built at Competence Centre “Virtual Reality” at TU Ilmenau (Fig. 1). This is a flexible 3-side-projection system with the combination of stereoscopic projection and acoustic wave-field-synthesis [2] for creating a realistic sound impression in real-time along with the visualisation.

You can configure the system in three different setups, depending on the application, the size of the user group and the necessary immersion (Fig. 2). The setup with the highest immersion is the 90° CAVE-setup (Fig. 2a). Here the user can turn his head and always see the stereoscopic picture. But this setup is only good for one or two users because of the picture distortion. The two other setups enable an interdisciplinary discussion about the VR scene in larger groups. By using INFITEC [1], the system realises a full channel separation and supports a very good stereoscopic picture.



Figure 1. Audio-visual VR-system at Competence Centre Virtual Reality

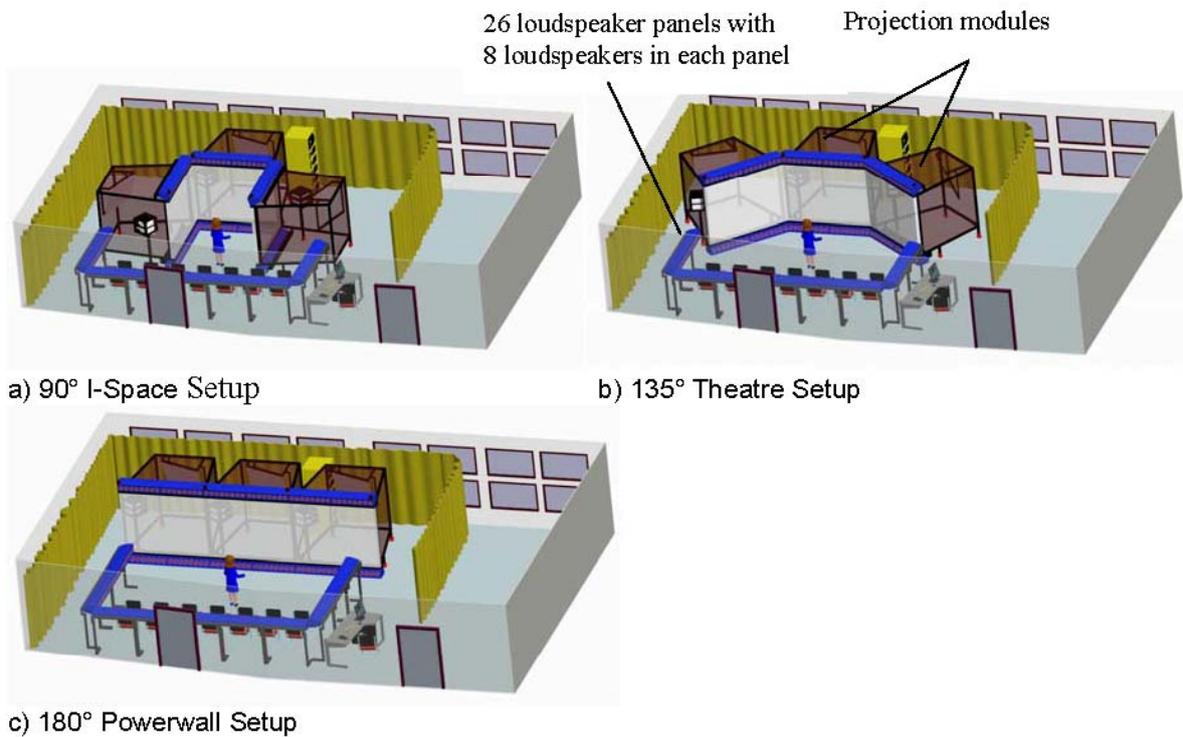


Figure 2. Setups of the flexible VR system

### VR Applications

An important aim with the VR system is the assistance of several research areas for interactive visualisation of complex data. For example, there are many applications for

product development [3]. They become particularly interesting when the virtual environment offers more than visualisation, e.g. the simulation of acoustic behaviour like in the installation at TU Ilmenau (see previous section).

In this contribution, however, another valuable application area is discussed: the interactive visualisation of measurement data, for example from the nano-measuring and nano-positioning machines which are the focus of other research activities of TU Ilmenau (Fig. 3). Here the aim is to get a VR model based on the measurement data automatically and quickly. For this several converters had to be developed.

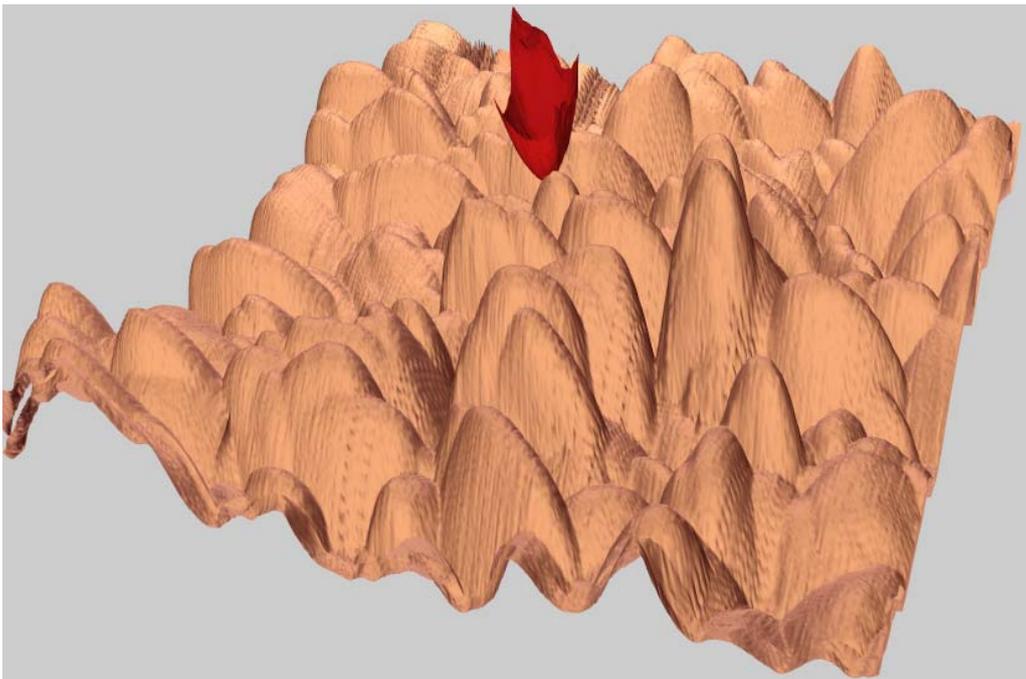


Figure 3. VR model of a measured nano-surface with a measuring tip

The measured data always has an array format with a determined edge length (see Fig.5  $x_n$  and  $y_n$ ). The content of each field is the value of the relative measured height. With this you have 3 coordinates defining each point, which can be used for the triangulation. By additional use of substitute colour representation the monoscopic depth information can be enhanced (Fig. 4). Because of the high resolution of the measured data and the accurate transfer to the VR model the generated representation is a good base for interdisciplinary discussion about the real surface. By using a scale over 10000:1 all details of the surface can be discussed efficiently.

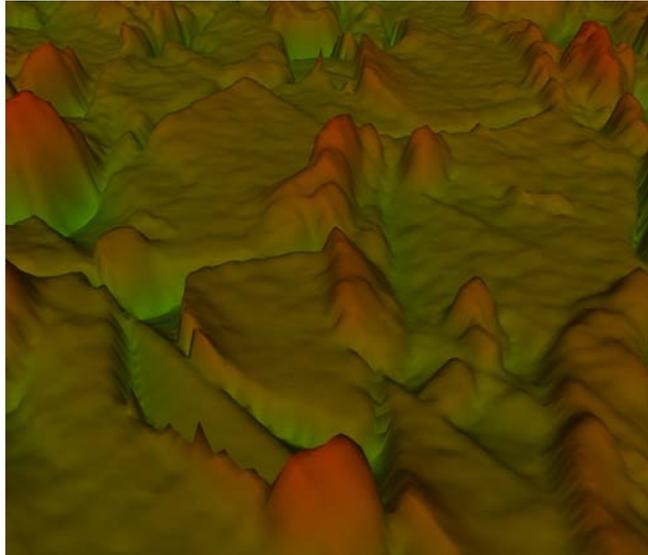


Figure 4. Coloured VR model of a measured nano-surface

In order to analyse the surface with people who do not know the measuring process, an animation of the process can support the understanding. So based on the known geometry of the measuring tip (Fig. 6) and the surface geometry an animation path can be created automatically. The animation path contains a key sequence and an appropriate transformation of the tip model.

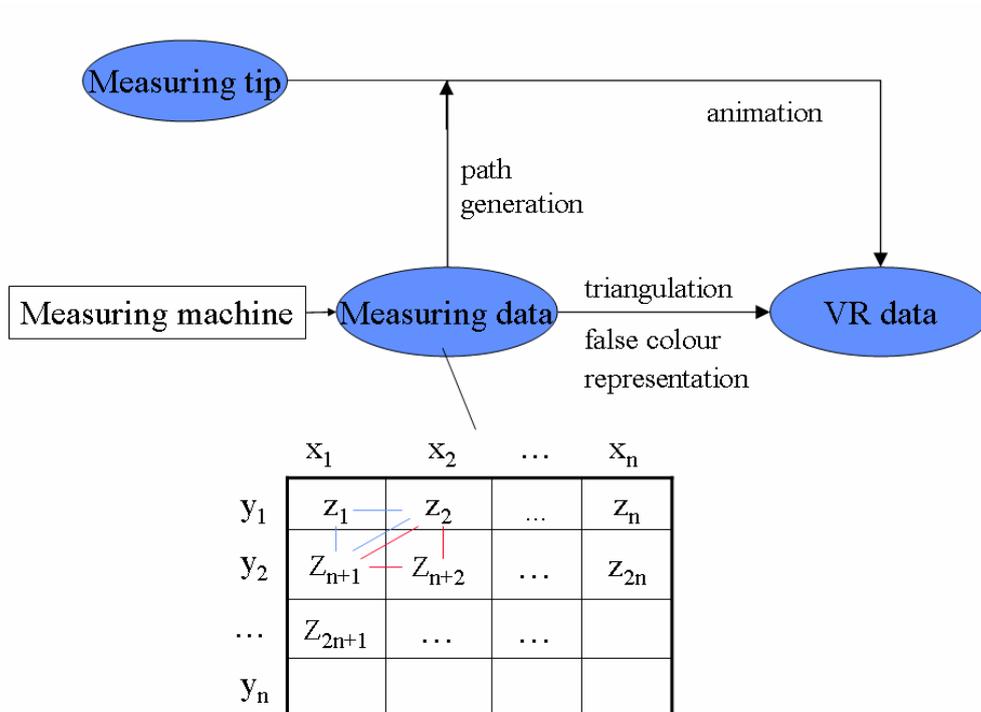


Figure 5. Method of VR data preparation based on measured data

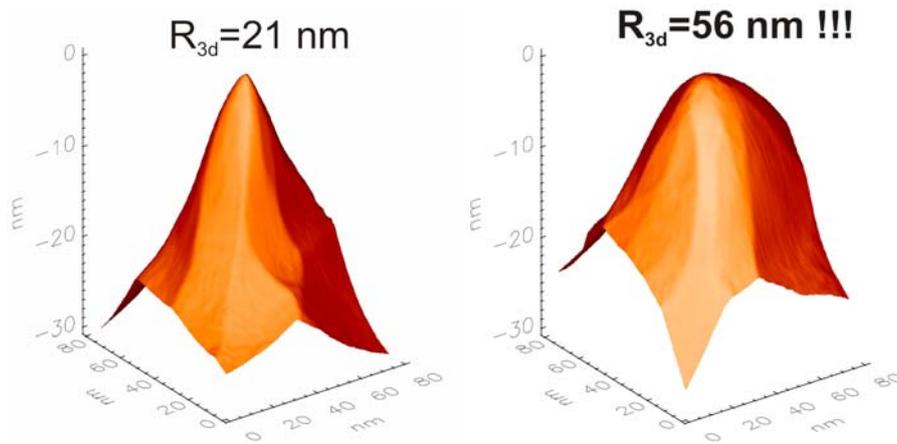


Figure 6. Cantilever tip before and after the use [4]

### Conclusion

The technology Virtual Reality enhances the interdisciplinary discussion about complex contexts by stereoscopic visualisation and interactive manipulation of virtual scenarios of real objects with arbitrary magnification. Automatic converters reduce the expenses for data preparation and assist the integration of VR in other research areas. So VR becomes an important tool for several application areas as shown for Nanotechnology.

### References:

- [1] Infitec Stereodisplay-Technik, [www.infitec.net](http://www.infitec.net), letzter Abruf 09.06.07
- [2] Brandenburg, K.; Brix, S. and Sporer, T.: "Wave-Field-Synthesis: new possibilities for large-scale immersive sound reinforcement". In 18th International Congress on Acoustics (ICA) - Acoustical Science and Technology for Quality of Life, volume 1, pages 507 – 508, April 2004.
- [3] Husung, S.; Höhne, G.; Lotter, E.: "Extended Virtual Prototyping". Proceedings of CIRP conference „The Future of Product Development“, Berlin (Germany), March 2007.
- [4] Frank, K.-H.; Machleidt, T.: Sensornahe Messdatenerfassung und –verarbeitung. Forschungsbericht Teilprojekt C2, DFG-Sonderforschungsbereich 62, TU Ilmenau 2005

### Authors:

Prof. Dr.-Ing. habil. Dr. h.c. G. Höhne, Prof. Dr.-Ing. Christian Weber, Dipl.-Ing. Stephan Husung  
 FG Konstruktionstechnik, Technische Universität Ilmenau  
 98693 Ilmenau  
 Tel.: 03677/69 5078, Fax: 03677/69 5052, E-mail: [stephan.husung@tu-ilmenau.de](mailto:stephan.husung@tu-ilmenau.de)



N. Schuster

## Measurement with subpixel-accuracy: Requirements and reality

### 9.4 Image Based 2D- and 3D-Measurement

The accuracy of an image based machine vision installation determines the choice of camera, measuring lens and illumination. Very often, the accuracy is expressed in parts of pixels because the imaging device forms the reference of the measuring arrangement. The unit "pixel" (here PX) guaranties a quick connection to other parameters of the camera like frame rate, required space and process control.

The digital image is the result of a sophisticated process chain which contains 1. the illuminated piece under test, 2. the image formation by the lens, 3. the transfer of light energy in electrical signals by the imaging device, 4. the charge transfer to analog-digital transducer, 5. the analysis of the grey-value image, 6. the subpixel edge detection algorithm. Serious publications name a maximum of accuracy of 1/5 PX ... 1/10 PX [1]. Much smaller values in advertising publications like 1/20 PX ... 1/40 PX ignore the influence of the steps 1-4.

In the following we present investigations in lab environment about the subpixel-accuracy what is referred to a mechanical material measure. This accuracy  $\delta N$  is composed of two parts: the edge detection error  $\Delta N$  and the calibration error  $\Delta c$ .

#### 1. Method to measure the edge detection error

The distance of two gauge blocks serves as mechanical material measure. The medium distance between gauge blocks is 30 mm, it can be varied in steps of 1/15 PX. Fig. 1 shows this installation.



Fig. 1: Material measure to measure the subpixel-accuracy

The gauge-bloc on the right hand side is fixed, the second gauge bloc is mounted on a piezo-moved table. The two gauge blocs have a thickness of 9 mm. They can be driven

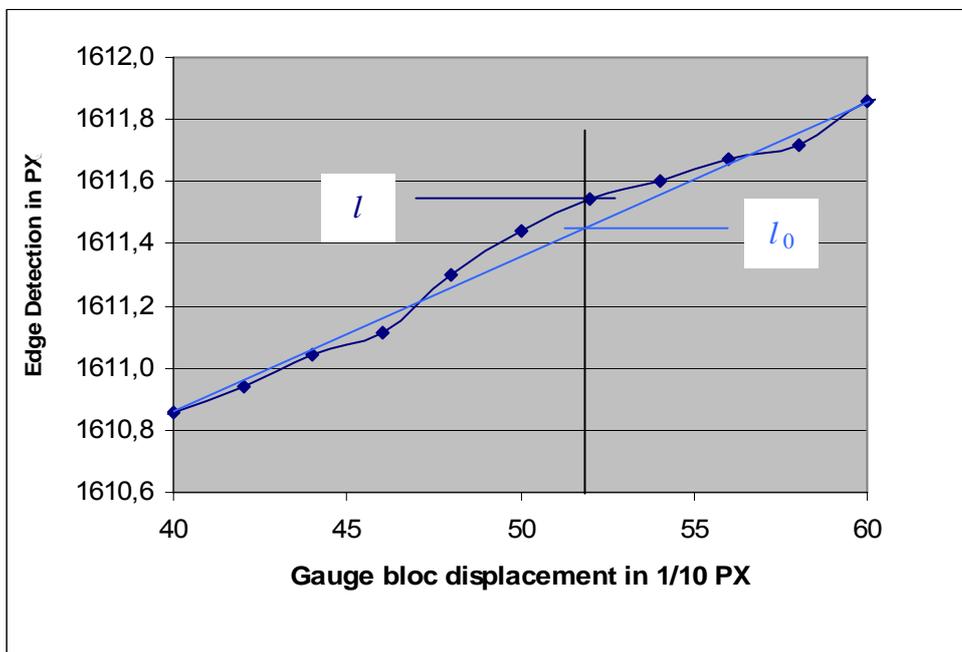
in steps of 1  $\mu\text{m}$  with a repeat accuracy of 75 nm.

The window on the bottom side of the stage permits to install different lighting units working in transmitting light. These units illuminate the piece under test in different manner: diffuse or telecentric, monochrome or white.

A telecentric lens T240/0,13 having a maximum field of view of 68 x 51 mm<sup>2</sup> views from the top side on the two gauge blocks. An entocentric lens would deliver false results in this constellation due to the perspective error.

The principle of measurement of the edge detection error is to compare the translational displacement of gauge blocs with the results of the subpixel edge detection algorithm in the digital image. The gauge bloc distance will be varied in 1/15 parts of one pixel.

Fig. 2 shows typical results: We don't obtain an ideal straight line but rather a lot of line pieces with varying slope. Here, each point represents the average of 7 measurement runs in the corresponding gauge bloc position. The subpixel accuracy follows by interpretation of these curves.



**Fig. 2:** Edge detection and their linear regression

The edge detection error is the averaged deviation  $\Delta N$  what is defined by the modulus of differences between the ideal linear regression and the measured value (see Fig. 2):

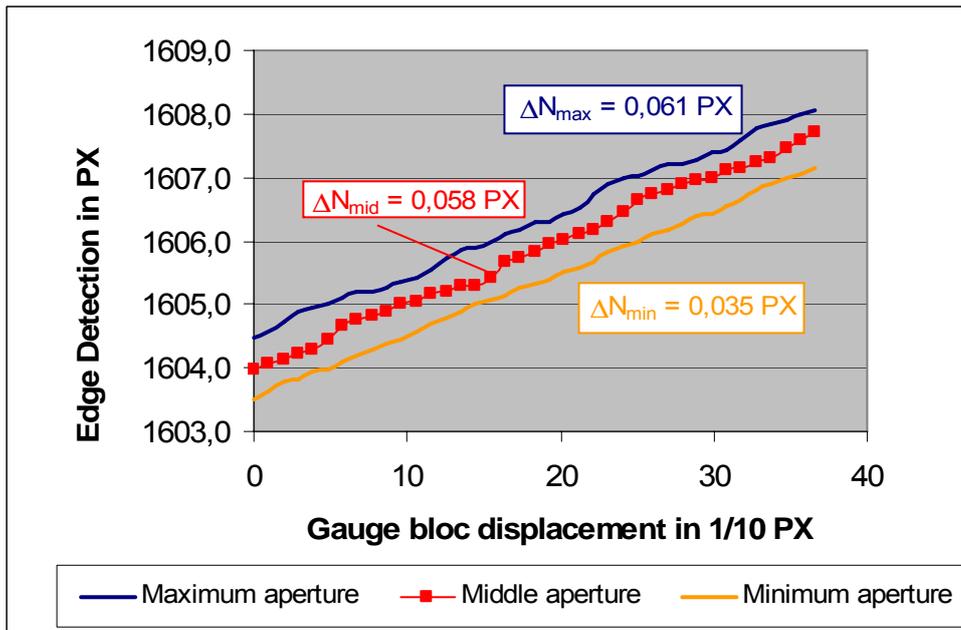
$$\Delta N = \frac{1}{m} \sum_{i=1}^m l - l_0 \quad (1)$$

while  $m$  is the number of gauge bloc positions,  $l$  is the result of the subpixel edge detection and  $l_0$  is the ideal length onto the linear regressed straight line.

In the following, we try to find recommendations to save a maximum in subpixel accuracy. For this, we analyze the results of edge detection under different circumstances as direction of edge detection, pixel-pitch and kind of illumination.

## 2. Aperture of telecentric lens

Fig. 3 shows the aggravating influence of the lens aperture. The gauge bloc is projected by the telecentric lens T240/0,13 on the 2/3"-CCD-matrix of the camera Oscar F810-C at maximum, middle and minimum aperture. Resulting curves disagree.



**Fig. 3:** Edge detection subpixel results versus gauge bloc displacement

Fig. 3 exemplifies the influence of the lens aperture: one pixel difference between the maximum and minimum. The conclusion for the 2D-measurement praxis is evident: Don't change the aperture after calibration.

The smallest deviation from the ideal straight shows the curve at the minimum aperture where the diffraction diffuses the light distribution on the imaging device. This slowly diffused image generates a good condition for the subpixel edge detection. The sharpest and brightest image at the maximum aperture offers the maximum of straight deviation. The average slope of three curves is identical.

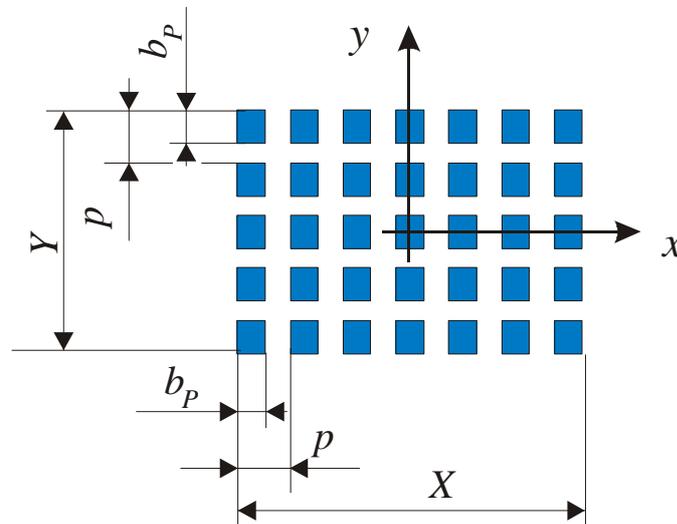
### 3. Illumination color and direction of edge detection

This often discussed problem depends mainly on the geometry of the matrix and of the number of activated light receiving cells. For measuring purposes it is exigent recommended to use an imaging device featuring an equal pixel pitch  $p'$  in  $x'$ - and  $y'$ -direction while  $x'$  and  $y'$  indicate the axis's of the light sensitive area (see Fig.4).

In the following investigation, we combine the color camera Oscar F810-C having a pixel pitch of  $2,8 \mu\text{m}$  in both directions with a telecentric lighting unit TZB60 in different colors. The matrix of the camera F810-C is covered with the Bayer pattern where the half of the cell number is sensible to green, the fourth to red and the other fourth to blue.

White, green, red and blue telecentric lighting units TZB60 are installed on the bottom side of Fig. 1. The quantitative analyze of the edge detection error at different colors corresponds to the number of activated pixel cells:  $\Delta N_{white} = 0,052$  ,  $\Delta N_{green} = 0,04$  ,  $\Delta N_{red} = 0,0$  ,  $\Delta N_{blue} = 0,01$  . The white light assures the smallest edge detection error, the blue light the biggest. Due to the color interpolation into the camera, differences of accuracy between colors are diminished. The conclusion for the 2D-measurement praxis is evident: A color camera have to be combined with a white illumination unit.

The often discussed choice of the direction in the edge detection process will be investigated using the white telecentric lighting unit TZB60-W. The digital image is captured by the camera Oscar F810-C. The analyze of 7 test runs in the two axis directions and at 45°.



**Fig. 4:** Preferable matrix geometry for measuring purposes

The edge detection error in different directions is found to  $\Delta N_x = 0.040$  ,  $\Delta N_y = 0.04$  ,  $\Delta N_{45^\circ} = 0.051$  . The best accuracy follows in x-direction. In y-direction, the edge detection error is 15% higher. Perhaps, it's due to the higher noise level in the column reading than the line reading of the camera CCD. A significant difference is found at the 45° edge detection. Here, the edge detection error is by the factor  $\sqrt[4]{2}$  higher than the average of both axial errors. This relation corresponds to the superposition of two no correlated processes. They are the edge detection in both axes.

The recommendation for the 2D-measurement praxis with CCD-imaging devices is to orient the critical dimension to measure in x'-direction (if it possible).

#### 4. Appropriate resolution of the telecentric lens

Very often, the image quality of a telecentric lens has to be “very good”. In the case of the distortion, the specification can be transformed in mm, what gives a clear reference to the required accuracy of the 2D-measuring system.

The specification of appropriate resolution is more complex. It depends on many factors like color of illumination, pixel pitch, interpolation procedures into the camera and edge detection algorithm.

Fig. 5 shows unwelcome results of edge detection what often remain unperceived in practical applications. Only the comparison with an adapted mechanical material measure discloses this kind of irregularity, while the used material measure must correspond to the piece under test in form, surface roughness and color.

These unwelcome results of edge detection in Fig. 5 come from the too high resolution of the telecentric lens. The lens T240/0,13 produces in combination with the Oscar F810-C camera (pixel pitch 2,8  $\mu\text{m}$ ) and white telecentric lighting unit likely linear edge detection results having an error  $\Delta N \leq 0.01$  (see Fig. 3). Now, we present the combination of the same lens with the monochrome 2/3"-CCD-camera Pictor® M1018 (pixel pitch 6,7  $\mu\text{m}$ ) and the telecentric lighting unit TZB60-B. The blue light generated by a LED has a bandwidth of 25 nm. By this means, all residual chromatical aberrations are eliminated.

Beyond of this, the Pictor® M1018 works without an interpolation algorithm between neighbor pixels.

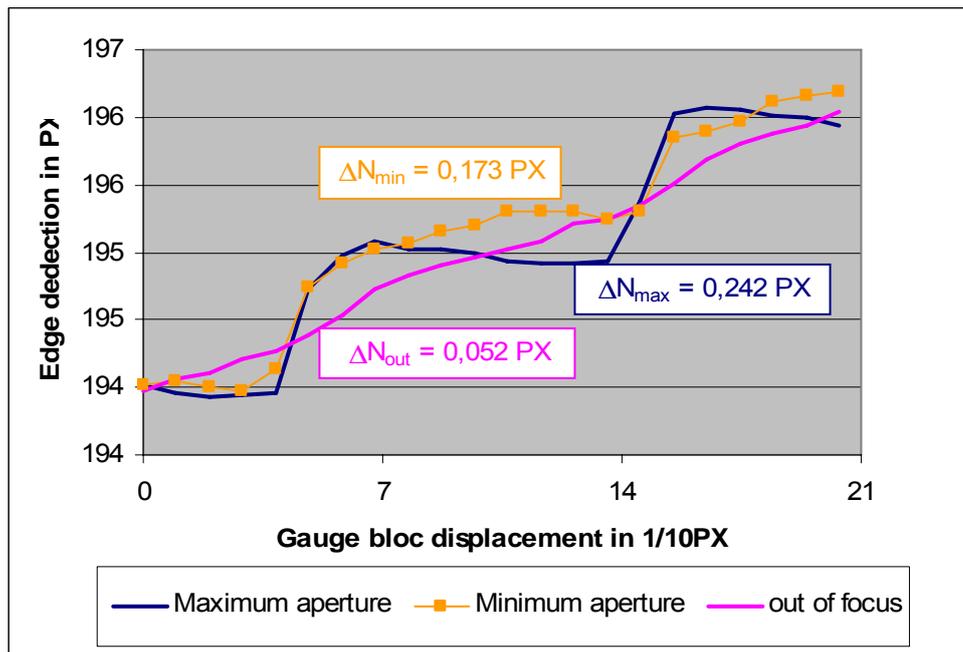


Fig. 5: Edge detection at “to sharp image”

The edge detection error at maximum aperture is four times higher than in the Oscar combination. Respecting the 2,4 times larger pixel pitch in the M1018, the error of accuracy in mm is 10 times larger! At the minimum aperture, this relation increases to twelve. The third curve in Fig.5 confirms the thesis about the “to sharp image”: an object side defocus of 30 mm delivers a likely linear behavior at maximum aperture. The edge detection error of this third curve corresponds to results of all investigated Oscar combinations:  $\Delta N_{out} = 0,052$ . The investigation of the combination M1018 with the TZB60-W at sharp object position delivers nearly the same edge detection error:  $\Delta N_{white} = 0,04$ .

For practical applications, the effect of the “to sharp image” is especially malicious because their disclosure demands expensive test arrangements. Beside of this, a traditional engineer of metrology has to accept the effect of a “to sharp image”. A white illuminating unit helps to avoid the “to sharp image” effect.

## 5. Conclusion

The edge detection error is only one of causes in the 2D-metrology what diminish the accuracy. Their amount depends on the balance of optical resolution, pixel pitch and interpolation runs in the camera. Our investigation under lab conditions delivers a typical value of  $\Delta N \approx 1,20$ . Adverse effects like the “to sharp image” deteriorate this coefficient up to  $\frac{1}{4}$  PX.

The accuracy of the complete 2D-measurement installation  $\delta N$  has to respect the supplement effect of the calibration error  $\Delta c$ . One proposal to calculate the accuracy is

$$\delta N = \Delta N + N \cdot \Delta c = \Delta N + N \cdot \frac{\Delta \beta}{\beta} \quad (2)$$

While  $N$  is the size of the piece under test in PX. The last term in Eq.(2) describes the variation of the lateral magnification  $\beta$  during the measuring process while  $\Delta \beta$  presents the standard deviation over there.  $\Delta c$  represents e.g. the perspective error of entocentric lenses and the “residual distortion”. They are no correlated errors.

The “residual distortion” implicates that the lens distortion can be electronically compensated. The compensation by symmetrical polynomial approximation in the field yields a residual distortion error between 1/10...1/20 PX [2]. The bilinear approximation in parts of the field of view respects the asymmetrical distortion due to the assembling of lens and camera. Here, the share in  $\Delta c$  depends strongly on the concrete measurement installation.

A method to measure the calibration error  $\Delta c$  is to change the position of the calibration piece in the field of view. Each of typical positions delivers one value of the lateral magnification, and  $\Delta c$  is calculated by (2).

The investigation of the combination T240-lens with Oscar camera in white light permits to quantify the effect of telecentric illumination versus the classical diffuse back light. The measured calibration error at telecentric illumination is  $1000\Delta c_{TZB} = 0.152$ , the diffuse back light installation delivers  $1000\Delta c_{DBL} = 0.20$  what is poorer. On the other side, the difference of both installations in the edge detection error is smaller:  $\Delta N_{TZB} = 0.052$  and  $\Delta N_{DBL} = 0.05$ . The resultant accuracy is dominated by the calibration error of the diffuse back light:  $\delta N_{TZB} = 0.25$  and  $\delta N_{DBL} = 0.32$ . Practically, a subpixel result of 1/10 PX with the diffuse back light make a false show of accuracy.

The attainable accuracy in the contact less 2D-metrology depends on the actual measurement installation. Expensive test installations based on a mechanical material measure disclose the relations in accuracy. Moreover, the test installation has to present the equivalent condition as the piece under test.

The presented approach distinguishes the edge detection error and the calibration error. Test runs are proposed. The quantitative evaluation suggests that a stable accuracy value of 1/10 PX demand expensive conditions. Values of 1/20 PX or more fine ignore the imponderability between the illuminated piece under test and the digital image to evaluate.

#### References:

[1] C. Demand, B. Streicher-Abel, P. Waskewitz, *Industrielle Bildverarbeitung*, Springer, Heidelberg 1998, ISBN 3-540-63877-6

[2] N. Schuster, A. Schmidt, “Legende und Wirklichkeit: Messgenauigkeit mittels Subpixel-Interpolation”, in: *Automatisierungsatlas 2006/07*, S. 536-542, TeDo-Verlag Marburg.

#### Author:

Norbert Schuster  
Vision&Control GmbH  
Pfüttschbergstr. 14  
D-98527 Suhl  
Phone: 49/3681/797425  
Fax: 49/3681/797465  
E-mail: dr.schuster@vision-control.com

P. Brückner/ S.C.N. Töpfer/ M. Correns/ J. Schnee

## Position- and colour-accurate probing of edges in colour images with subpixel resolution

### ABSTRACT

This paper presents a novel method for probing of edges in colour images with subpixel resolution. The enhanced method based on [1] enables the highly precise determination of the edge position at edges with a change in colour as well as with a change in intensity or a combination of both. The attained precision is similar to the precision of subpixel-accurate edge position determination in greyvalue images.

### INTRODUCTION

Subpixeling methods enable probing of edges in greyvalue images with a resolution larger than the pixel centre distance of the sensors. Depending on the quality of the image data an increase of resolution of  $1/10^{\text{th}}$  to  $1/100^{\text{th}}$  pixels is attainable [2]. Typically, methods for subpixel-accurate edge position determination presume that only one function value, id est the intensity respectively the greyvalue, varies within the image. Colour images do not fulfil this requirement. They contain three function values, which may vary with the position within the image, for example the primary colours red  $R(x, y)$ , green  $G(x, y)$  and blue  $B(x, y)$ .

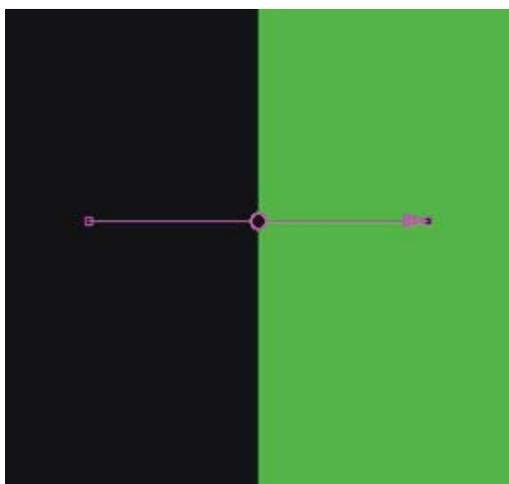


Fig. 1: Colour edge with search line

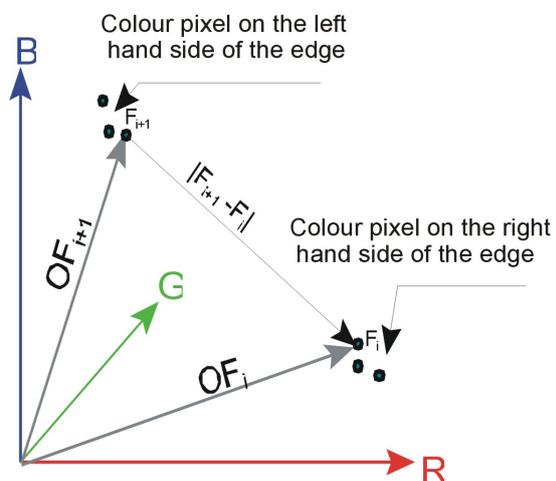


Fig. 2: Definition of the position vectors  $OF$

The solution consists of a vectorial approach. Each pixel of the colour image is considered as a vector. Similar to measurements in greyvalue images a search line is defined (Fig. 1). The search line comprises RGB colour values, which can be represented as vectors in the RGB colour space (Fig. 2). Appropriate vector operations are deployed on each colour vector of two neighbouring pixels. Thus, a one-dimensional function scaled between 0 and 255 is derived for the whole search line. Thus, the edge position is determined by subpixeling as in greyvalue images.

## STATE OF THE ART

The known deployed methods for edge detection in colour images yield the pixel-accurate position only. They are mainly utilised for image segmentation and object recognition. If edges are determined for the individual colour channels R, G and B, as in [3], or for only one colour channel as in [4], image information is lost. This applies also if colour images are converted into greyvalue images, as in [5]. A better method is shown in [6] and [7], which convert image data into the HSI colour space. This enables the detection of changes in hue and in intensity with varied Sobel operators. Vector based methods are showing the best performance for edge detection in colour images. They treat each pixel as a vector in a usually three dimensional colour space. The edge detection is done with a variety of operators. The works [8], [9], [10], [11], and [12] deal with the vector gradient operator. In [8] and especially in [9] vector order statistic operators and the entropy operator are examined. In [8], [11], [12] and [13] the difference vector operator or comparable distance metrics are used. Furthermore, operators based on the second derivatives are introduced in [8]. The variety of these operators for colour edge detection proves that no perfect colour edge detection operator exists. Consequently, there are different colour edge detection operators most suitable for each specific task [8]. Basically, the aim of this paper is not edge detection but the highly precise edge position determination. Only few papers are dealing with this issue, for example [14] and [1]. The paper at hand develops these ideas further.

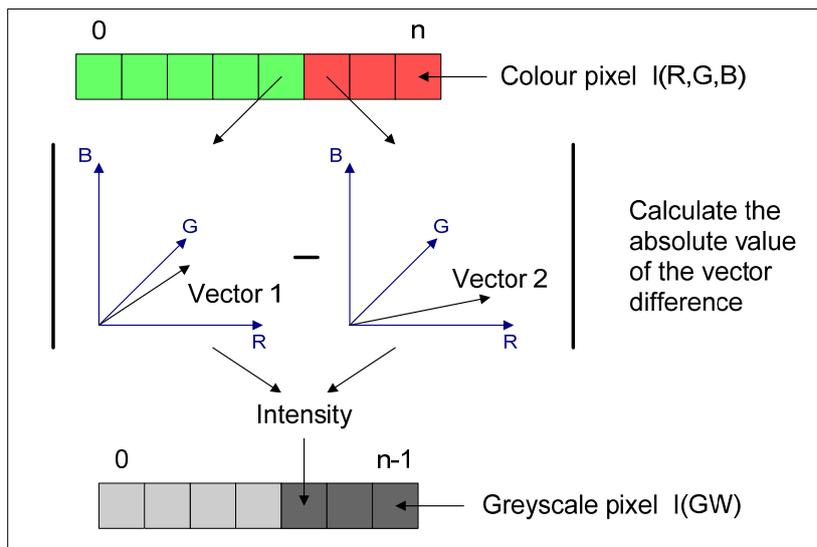
In optical coordinate metrology well known subpixeling methods for probing of edges with subpixel resolution in greyvalue images exist. In [2] an overview of different methods such as threshold  $\sim$ , differential  $\sim$  and integral methods (e.g. photometric mean) as well as correlation and parameter identification is given. Combined with interpolation, approximation and reconstruction subpixel resolution is attained.

## PROCEDURE FOR SUBPIXELING IN COLOUR IMAGES

The absolute value of the vector difference  $|\vec{u}-\vec{v}|$  of the two three dimensional vectors  $\vec{u}$  and  $\vec{v}$  is calculated as:

$$\vec{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}, \quad \vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}, \quad |\vec{u}-\vec{v}| = \begin{pmatrix} u_1 - v_1 \\ u_2 - v_2 \\ u_3 - v_3 \end{pmatrix}.$$

The conversion of the coloured search line into a function with one intensity value only is done via vector difference. Therefore, the absolute value of the vector differences of two adjacent colour vectors is calculated for the whole search line. The absolute value of the vector differences is maximal at the edge because of the changes in intensity and colour. Thus, the calculated one-dimensional function has low intensities at homogeneous areas and an intensity peak at the edge.



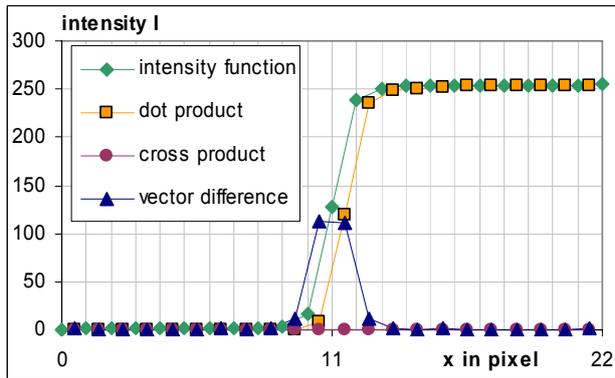
**Fig. 3:** Conversion of a search line through the vector difference

The edge position is at the centre of gravity of the intensity peak. Since the vector difference is calculated from two neighbouring vectors or pixels,  $n$  colour pixels yield only  $n-1$  vector differences (Fig. 3). Consequently, the converted search line has one pixel less

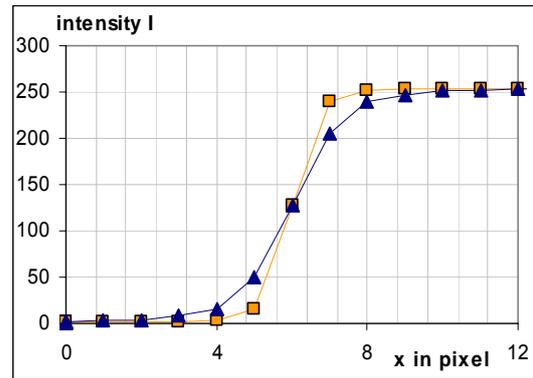
than the coloured original. This also entails a systematic shift of the calculated edge position by 0.5 pixels which has to be included. For the precise determination of the edge position with subpixel resolution the outlined methods from [2] are applied on the one-dimensional function. Due to the characteristic peak of the intensity function marking the edge position a correlation method has been chosen. It determines the maximum correlation between the intensity function and a Gaussian-shaped reference pattern. The subpixel-accurate edge position equals the centre of gravity of the calculated curve of the correlation coefficient. Finally, the mentioned 0.5 pixel shift is charged in order to determine the accurate edge position in the colour image.

## SIMULATIONS

Several simulations have been performed. One simulation demonstrated that the vector operations cross product and dot product are not suitable for the conversion of the coloured search line. The cross product does not yield a change in its characteristics if the edge is represented by a pure intensity change (Fig. 4). The dot product is equally unsuitable for dimensional measurements. Depending on the absolute value of the change in the original intensity function in the colour image its signal flank marking the edge is shifted relative to the actual edge position.



**Fig. 4:** Vector operations at an edge in a colour image (intensity function) with pure intensity change



**Fig. 5:** Simulated intensity function for a thin edge (square) and for a broad edge (triangle)

An other set of simulations was performed in order to evaluate the performance of the outlined method for subpixel-accurate edge probing in colour images independently from the actual lens and the imaging sensor. Therefore artificially generated edges with known edge position were utilised (Fig. 5). The intensity functions have been saved as 8 bit greyvalue images and as 24 bit colour images where each colour channel contained the identical intensity function. The results from the greyvalue subpixeling and from the proposed colour subpixeling have been compared.

**Table 1:** Average deviation from the known edge position in pixels for 24 bit colour images and for 8 bit greyvalue images at different intensity functions with 8 search lines at 4 angles between 45° and 90° relative to the edge for both probing directions.

| Deployed intensity function (see Fig. 5)  | Deviation of edge position in greyvalue image | Deviation of edge position in colour image |
|---|---|--|
| Thin edge with int. transition 0 to 254   | 0,0122  | 0,0150                                     |
| Thin edge with int. transition 30 to 200  | 0,0134  | 0,0165                                     |
| Thin edge with int. transition 50 to 170  | 0,0096  | 0,0140                                     |
| Broad edge with int. transition 0 to 254  | 0,0091  | 0,0126                                     |
| Broad edge with int. transition 30 to 200 | 0,0055  | 0,0163                                     |
| Broad edge with int. transition 50 to 170 | 0,0159  | 0,0201                                     |

The deviation of the determined edge position amounts to the same order of magnitude for greyscale  $\sim$  and for colour subpixeling (Tab. 1). However, the results of the greyscale subpixeling are slightly better. The attained subpixel resolution is approximately  $1/100^{\text{th}}$  pixel.

## EXPERIMENTAL RESULTS

In order to determine the performance of colour subpixeling under real measuring conditions a circular calibration target has been captured with a colour and a greyscale camera. The latter was an ADIMEC MX12P/8X23 CCD camera with  $1024 \times 1024$  pixels with a pixel size of  $7,5 \mu\text{m} \times 7,5 \mu\text{m}$ . The colour camera was a BASLER A113CP equipped with a Bayer filter mosaic with  $1300 \times 1030$  pixels with a pixel size of  $6,7 \mu\text{m} \times 6,7 \mu\text{m}$ . For comparability both cameras were utilised with the same optical reproduction system (one fold magnification) and the same cold light source (colour temperature approx. 5000 Kelvin) as transmitted light illumination.

As Table 2 shows, the experimental results validate the simulation results. The small differences between greyscale  $\sim$  and colour subpixeling are due to the different cameras. The colour camera has a much larger noise and a reduced spatial resolution due to the Bayer filter mosaic. Therefore, the 7<sup>th</sup> and 8<sup>th</sup> column contain the measured values for greyscale subpixeling which has been deployed on the colour images after they have been converted into greyscale images based on the well known luminance equation. Thus, it is proven that the larger part of the deviation is indeed due to the camera characteristics and not due to the subpixeling method.

**Table 1:** Circle radii measured at the same calibration target and captured with greyscale  $\sim$  and colour camera (average from 50 measurements, pixel factor calibrated at 5th circular ring, all values in  $\mu\text{m}$ )

| Circular ring | Calibration value, radius [ $\mu\text{m}$ ] | Measurement in greyscale image |           | Measurement in colour image |           |                                |           |
|---------------|---|--------------------------------|-----------|-----------------------------|-----------|--------------------------------|-----------|
|               |   |                                |           | Original colour image       |           | Converted into greyscale image |           |
|               |   | Radius                         | Deviation | Radius                      | Deviation | Radius                         | Deviation |
| 3             | 249,60                                      | 249,52                         | -0,08     | 249,37                      | -0,23     | 249,32                         | -0,28     |
| 4             | 500,39                                      | 500,35                         | -0,04     | 500,13                      | -0,26     | 500,08                         | -0,31     |
| 5             | 999,72                                      | 999,72                         | 0,00      | 999,72                      | 0,00      | 999,72                         | 0,00      |
| 6             | 2000,40                                     | 2000,31                        | -0,09     | 2000,38                     | -0,02     | 2000,29                        | -0,11     |

The performed experiments prove the suitability of the proposed colour subpixeling for high-resolution dimensional measurements. Nevertheless, most decisive for the achievable accuracy is the quality of the image data as well as the characteristic of the actual measuring object, e.g. colour distribution.

## SUMMARY

A method for the highly accurate determination of the edge position in colour images has been outlined. Key principle is the vectorial approach which allows the subsequent application of established subpixel methods for greylevel images. The originality of the paper lies with the detailed investigation of the applicability of the proposed methods for determining the edge position with subpixel accuracy. There are three basic types of edges in colour images. It must be differed between edges represented by a pure intensity transition and edges represented by a pure colour change and edges represented by any combination of the previous two edge types. The vector difference is the only method that enables the determination of the edge position with subpixel accuracy in colour images for all three edge types. Based on simulated and measured data it is demonstrated that a resolution of approximately  $1/100^{\text{th}}$  pixel can be attained. Prerequisite is the utilisation of a colour camera with low noise.

### References:

- [1] C. Usbeck und P. Brückner. Verfahren zur Bestimmung der Kantenposition in Farbbildern, insbesondere für Farb- und Intensitätsübergänge. Nr.DE10020067. Patent, April 2000.
- [2] Olaf Kühn. Ein Beitrag zur hochauflösenden zweidimensionalen Geometriemessung mit CCD-Zeilensensoren. PHD-Thesis, Technische Universität Ilmenau, 1997.
- [3] Mark Hedley und Hong Yan. Segmentation of color images using spatial and color space information. SPIE Journal of Electronic Imaging Volume 01(04), October 1992, S. 374–380.
- [4] Imao Kaoru und Ouchi Satoshi. SYSTEM FOR DETECTING EDGE PART OF COLOR IMAGE, April 1990. Patent.
- [5] Totsuka Takushi und Mitsunaga Tomoo. Method and device for detecting edge. Nr. US5995662. Patent, November 1999.
- [6] Takahashi Eiji. Image recognition method and apparatus utilizing edge detection based on magnitudes of color vectors expressing color attributes of respective pixels of color image, October 2000. Patent.
- [7] Thierry Carron und Patrick Lambert. Color edge detector using jointly hue, saturation and intensity. IEEE International Conference on Image Processing Volume 3, November 1994, S. 977–981.
- [8] Shu-Yu Zhu, K.N. Plataniotis und A.N. Venetsanopoulos. Comprehensive analysis of edge detection in color image processing. SPIE Optical Engineering Volume 38(4), April 1999, S. 612–625.
- [9] P. Trahanias und A.N. Venetsanopoulos. Color edge detection using vector order statistics. IEEE Transactions on Image Processing Band 2, April 1993, S. 259–264.
- [10] Rafael C. Gonzalez und Richard E. Woods. Digital Image Processing. Prentice-Hall, Inc. 2. Edition, 2002.
- [11] R.D. Dony und S. Wesolkowski. Edge detection on color images using RGB vector angles. IEEE Canadian Conference on Electrical and Computer Engineering Volume 2, May 1999, S. 687–692.
- [12] S. Wesolkowski, M.E. Jernigan und R.D. Dony. Comparison of color image edge detectors in multiple color spaces. IEEE International Conference on Image Processing Volume 2, September 2000, S. 796–799.
- [13] L. Shafarenko, M. Petrou und J. Kittler. Automatic watershed segmentation of randomly textured color images. IEEE Transactions on Image Processing Volume 6, November 1997, S. 1530–1544.
- [14] P. Brückner, C. Usbeck und A. Eichhorn. Aufnahme und Auswertung von Farbbildern in der Koordinatenmesstechnik. 47. Internationales Wissenschaftliches Kolloquium der TU Ilmenau, September 2002.

### Authors:

Dr.-Ing. Peter Brückner, Dipl.-Wirtsch.-Ing. Susanne C.N. Töpfer,  
Dipl.-Ing. Martin Correns, cand. Dipl.-Ing. Jost Schnee

Technische Universität Ilmenau  
Quality Assurance Department  
P.O. Box 100 565  
98684 Ilmenau, Germany  
Phone: +49 3677 693941  
Fax: +49 3677 693823  
E-mail: peter.brueckner@tu-ilmenau.de

E. Sparrer / T. Machleidt / R.Nestler / K.-H. Franke / M. Niebelschütz

## **Deconvolution of Atomic Force measurements in special modes – methodology and application**

### **Abstract**

The signal measured by Kelvin Force Microscopy is a convolution of an effective surface potential and a microscope intrinsic point spread function, which allows the restoration of the measured data by linear deconvolution. An analytical way is demonstrated by gaining the point spread function of the microscope. The linear shift invariant channel is introduced as a signal formation model and a wiener filter supported deconvolution algorithm is applied to the measured data.

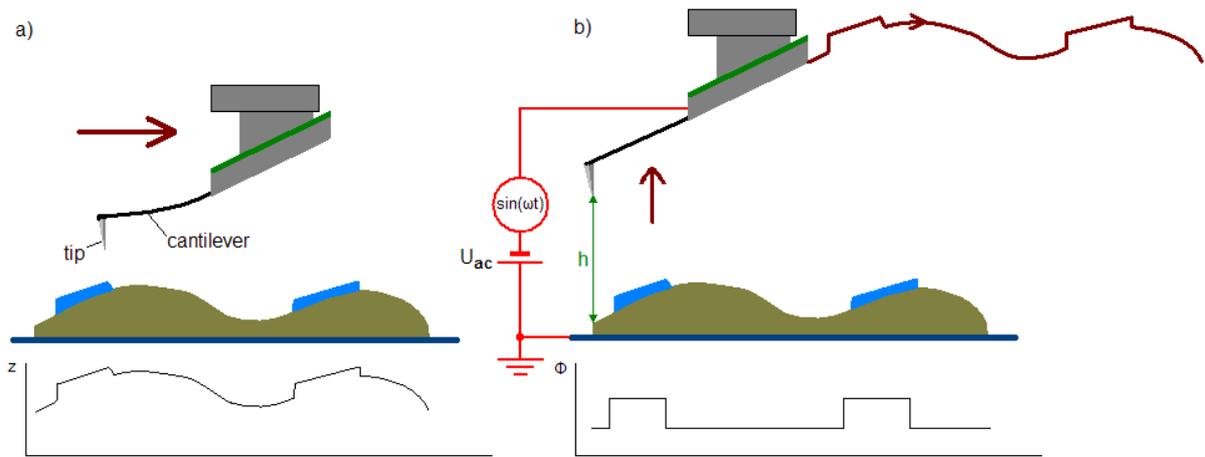
### **Motivation**

Since the early 1970ies semiconductor industry confirms the postulation of Gordon Moore known as Moores Law with remarkable continuity. Year by year the dimensions of semiconductor elements are shrinking. As per current International Technology Roadmap for Semiconductors [1] the size of 20nm is planed to be undershooting by the year 2017. Present, on light microscopy based, analysis systems are not capable to resolve these small dimensions, so new technologies have to be deployed. The well-known techniques of atomic force microscopy (AFM) provide an opportunity to overcome this drawback. In this paper the AFM special mode of Kelvin Force Microscopy (KFM) is investigated. By A. Born and T. Heuer [2,3] the KFM resolution is estimated by 50nm to 100nm. Because common atomic force microscopes are capable of a higher resolution, the measured data seams blurred. Investigation of the signal formation process enables the restoration of KFM data.

### **Kelvin Force Microscopy**

The KFM is a method to detect the surface potential of micro- and nanostructured samples using a common atomic force microscope. To detect the surface potential all

surface forces beside the electrostatic force must be eliminated. While the electrostatic force has a very long range compared to other surface forces, these forces are disposed by moving the cantilever tip to a height  $h$  of at least 100nm above the sample surface. Owing to the indirect dependence of the measured electrostatic force to the quadratic distance  $h^2$  of the charged objects this distance must be kept constant to provide a proper measuring. Therefore a KFM measurement is done in two passes, which are demonstrated in Figure 1.



**Figure 1 operating mode of the KFM in two passes**  
**a) determination of the sample topography**  
**b) measurement of the surface potential with constant altitude  $h$**

In the first pass the surface topography is detected by applying a standard AFM method. To provide a constant altitude  $h$  between tip and surface in the second pass the cantilever is driven at the desired distance  $h$  along the detected surface trajectory. During the measurement the cantilever is oscillated by applying a voltage  $U_{ac}$  between the cantilever and the sample surface, which is altering at the resonance frequency  $f_0$  of the cantilever. According to H. O. Jacobs et. al [4] in case of resonance the acting electrostatic force at the cantilever tip can be expressed by equation ( 1 )

$$F_{c,\omega_0}(x, y) = \frac{dC}{dz} \cdot (U_{dc} - \Phi(x, y)) \cdot U_{ac} \cdot \sin(2\pi f_0 t) \quad (1)$$

The controller of the KFM system adjusts an additional direct voltage  $U_{dc}$  until the acting force vanishes as described by equation ( 2 )

$$0 = C'(x, y) \cdot (\Phi(x, y) - U_{dc}(x, y)), \text{ with } C' = \frac{dC}{dz} \quad (2)$$

The absolute value of the measured potential  $\Phi_{dc} = U_{dc}$  is equal to the surface potential  $\Phi$  at the point of measurement. The tip to surface capacity  $C$  can be described as a distributed capacity as described in Figure 2 and equation ( 3 ).

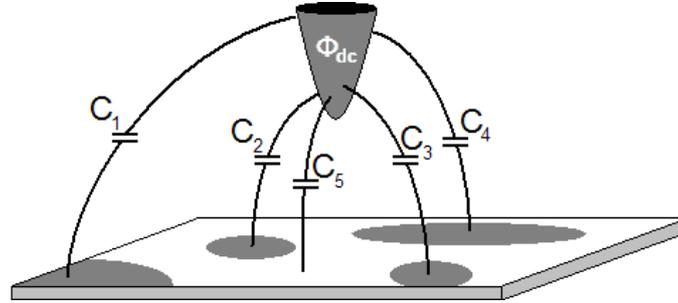


Figure 2 model of distributed capacity between tip and sample surface

$$C(x, y) = \iint_{i,j} C_{ij}(x - x_i, y - y_j) didj \quad (3)$$

Equation ( 2 ) solved for the measured potential  $\Phi_{dc}$  and the capacity substituted by equation ( 3 ) results in the linear convolution integral displayed in equation ( 4 )

$$\Phi_{dc}(x, y) = \iint_{i,j} \frac{C'_{ij}(x - x_i, y - y_j)}{C'(x, y)} \cdot \Phi(x, y) didj \quad (4)$$

The fraction described by the distributed capacity derivations  $C'_{ij}/C'$  can be realized as tip dependent point spread function (PSF) of the KFM system. Proofing the linear dependence of the measured signal  $\Phi_{dc}$  to the actual surface potential  $\Phi$  the linear shift invariant channel can be introduced as a model to describe the signal formation process. The linear shift invariant channel (Figure 3) consists of a linear shift invariant system and additive noise at the channels exit as described by Kreß and Irmer in [5].

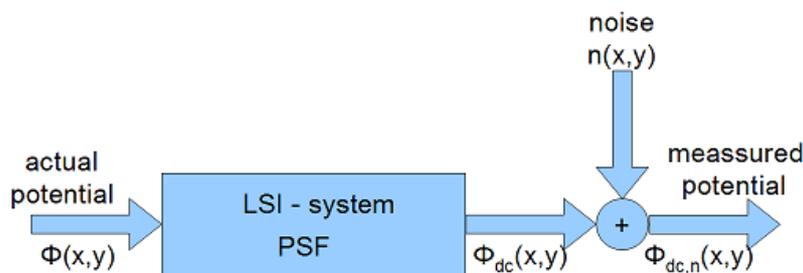


Figure 3 the linear shift invariant channel

This model allows the restoration of the KFM data by a linear deconvolution. To deconvolve the data, the PSF of the KFM system must be known. Therefore a deduction based on the physical interrelations has been applied.

## Estimation of the KFM systems PSF

In R. Königs dissertation [6] a method to calculate the electric field based on a hyperbolic coordinate system is presented. Utilization of this coordinate system allows a straightforward description of the electrostatic field lines between the charged tip apex and its mirror charge [7]. According to S. Gómez-Mónivas et. al [8] only the apex of the KFM tip is responsible for the shape of the PSF and the apex can be well approximated by a hyperboloid. Estimating the apex shape by blind tip estimation [9] respectively measuring it at suitable standards the PSF can be derived by transforming the apex surface to polar coordinates and fitting a hyperbola as shown in equation ( 5 ) for every angle  $\alpha$ .

$$u^2 = C_1(\varphi)z + C_2(\varphi)z^2 \quad (5)$$

By means of the parameters  $\eta_s, \xi$  and  $r_{1,2}$  equation ( 6 ) the electrostatic force acting between a plane surface and the tip apex is given by equation ( 7 ).

$$\eta_s(\varphi) = \sqrt{\frac{1}{1+C_2(\varphi)}}, \quad \xi(\varphi) = \frac{r_1(\varphi) + r_2(\varphi)}{2a(\varphi)} \quad (6)$$

$$\text{with } r_{1,2}(\varphi) = \sqrt{a^2(\varphi) + u^2(\varphi)} \quad \text{and} \quad a(\varphi) = \frac{C_1(\varphi)}{2C_2(\varphi) \cdot \eta_s(\varphi)}$$

$$F_{el}(u, \varphi) = Q \cdot \frac{2\eta_s(\varphi)}{\ln\left(\frac{1+\eta_s(\varphi)}{1-\eta_s(\varphi)}\right) \sqrt{(\xi^2(\varphi) - \eta^2(\varphi))(1-\eta^2(\varphi))}} \frac{U_{Bias}}{h} \quad (7)$$

The PSF is finally calculated by a normalizing equation ( 7 ) to its integral. PSF estimations obtained by the method are shown in Figure 4.

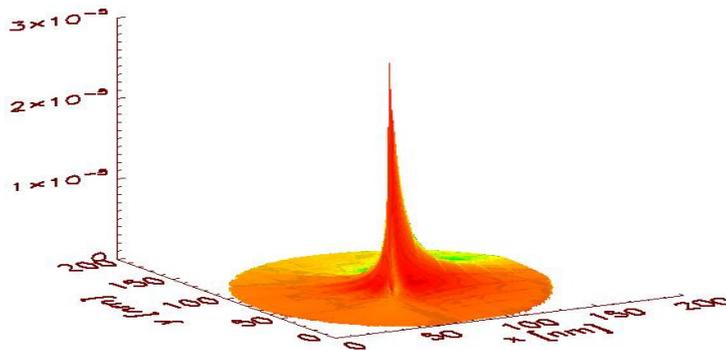
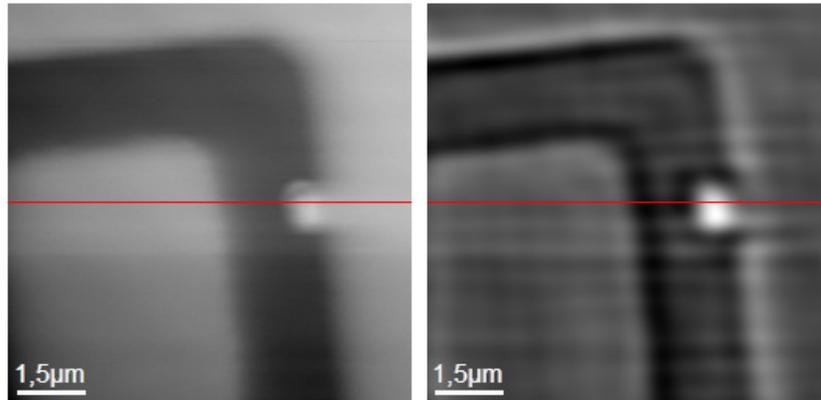


Figure 4 PSF gained analytical deduction based on the physics of electrostatic fields

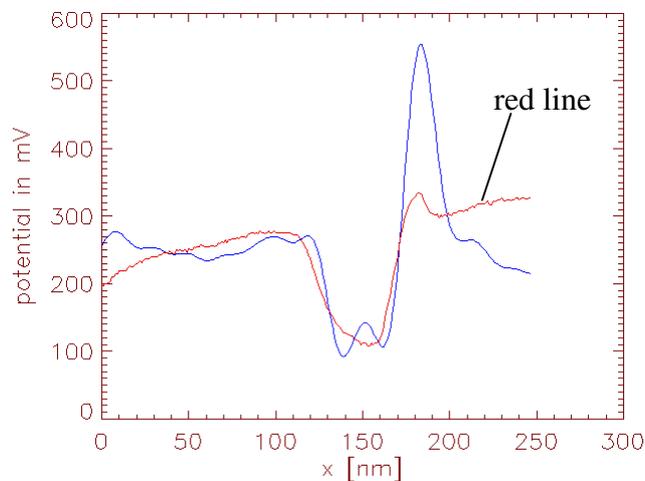
## Deconvolution of the KFM data

To deconvolute the data an algorithm based on a PSF inversion and supported by a wiener filter as a regularization approach was implemented. In the measured KFM data and the deconvolved estimate of the actual surface potential is shown.



**Figure 5 left: measured KFM data – right: deconvolved data**

The red lines in Figure 5 represents the location of the plots to be seen in Figure 6



**Figure 6 plot of the deconvoluted (blue) and measured (red) data along the marked line in Figure 4**

Reasons for a wider basis of the deconvoluted peak and a lower potential to its right side shown in Figure 6, is the peaks topography. The KFM peak is caused by a deposited particle with three times extend than the average topography of the surrounding area. For this reason the formation of the peak can not be described by the linear channel model, which causes failures while applying it for deconvolution.

## Conclusion

Assuming the surface of a sample as plane, the measured KFM data can be understood as a convolution of a microscope inherent transfer function with the actual potential distribution on the sample surface. Therefore the linear shift invariant channel can be introduced as model to describe the signal formation process. By analyzing the physical interrelations the transfer function can be determined and the measured KFM data can be deconvoluted. The restoration can be well done by a direct inversion of the transfer function supported by a wiener filter approach and the deconvoluted data allow a better interpretation of the actual potential distribution.

## Acknowledgments

This work was supported by the German Science Foundation (DFG, SFB 622). The authors wish to thank all those colleagues at the Technische Universität Ilmenau and the ZBS Ilmenau e. V., who have contributed to these developments.

## References:

- [1] Arden, W.; Copez, P.; Ishiuchi, H.; Osada, T.; Moon, J. T.; et. al: "*International Technologie Roadmap For Semiconductors 2006 Update – Overview and Working Group Summaries*" <http://www.itrs.net/Links/2006Update/2006UpdateFinal.htm>
- [2] Born, A.: "*Nanotechnologische Anwendung der Rasterkapazitätsmikroskopie und verwandter Rastersondenmethoden*" Dissertation; Universität Hamburg, Hamburg; 2000
- [3] Heuer, T.: "*Lateral aufgelöste Messung des Oberflächenpotentials mit der Kelvin-Methode*" Dissertation; Universität der Bundeswehr München, München; 2004
- [4] Jacobs, H.O.; Leuchtmann, P.; Homan, O.J.; Stemmer, A.: "*Resolution and Contrast in Kelvin probe force microscopy*" Journal of Applied Physics, issue.84, no.3; 1998; pp. 1168-1173
- [5] Kreß, D. Irmer R.: „Angewandte Systemtheorie, Kontinuierliche und zeitdiskrete Signalverarbeitung“, VEB Verlag der Technik, Berlin; 198
- [6] König, R.: "*Nanostrukturierung mit dem Rastertunnelmikroskop*" Dissertation; TU Braunschweig, Braunschweig; 1997
- [7] Philippow, E.: "*Grundlagen der Elektrotechnik*" Akademische Verlagsgesellschaft Geest & Portig KG, Leipzig, 1976
- [8] Gómez-Mónivas, S.; Froufe, L.S; Carminati, R.; Sáenz, J.J.: "*Tip-shape effects on electrostatic force microscopy resolution*" Journal Nanotechnology, issue.: 12, no.: 4, 2001, pp. 496 – 499
- [9] Machleidt T., Franke K.-H.: "Methods for Reconstruction of Atomic Force Microscope Data Based on Morphological Image Processing", MicroNano Integration, Knobloch H., Kaminorz Y., Springer-Verlag Berlin Heidelberg, 2004

## Authors:

Dipl.-Ing. Erik Sparrer  
Dipl.-Ing. Torsten Machleidt  
PD Dr.-Ing. habil. Karl-Heinz Franke  
TU Ilmenau / Computer Graphics Group / SFB 622 C2  
Ehrenbergstraße 29 (Ernst-Abbe-Zentrum), Postfach 100565  
D-98693, Ilmenau  
Phone: +049 / 3677 / 695068  
Fax: +049 / 3677 / 695071  
E-mail: [erik.sparrer@tu-ilmenau.de](mailto:erik.sparrer@tu-ilmenau.de)

Dipl.-Ing. Rico Nestler  
Zentrum für Bild- und Signalverarbeitung e.V.  
Gustav-Kirchhoff-Str. 5  
D-98693, Ilmenau  
Phone: +049 / 3677 / 2010305  
Fax: +049 / 3677 / 2010302  
E-mail: [rico.nestler@zbs-ilmenau.de](mailto:rico.nestler@zbs-ilmenau.de)

Dipl. Wirtsch. Ing. Merten Niebelschütz  
TU Ilmenau / FG Nanotechnologie / SFB 622 A8  
Postfach 100565  
D-98684 Ilmenau  
Phone: +049 / 3677 / 693408  
Fax: +049 / 3677 / 693709  
E-mail: [merten.niebelschuetz@tu-ilmenau.de](mailto:merten.niebelschuetz@tu-ilmenau.de)

D. Kapusi / T. Machleidt / K.-H. Franke / E. Manske / R. Jahn

## Measuring large areas by white light interferometry at the nanopositioning and nanomeasuring machine (NPMM)

### Abstract

White light interferometry is a new application for the nanopositioning and nanomeasuring machine (NPMM). The NPMM was developed under the leadership of the Institute of Process Measurement and Sensor Technology at the Technische Universität Ilmenau (Germany) and allows highly exact dimensional and traceable positioning with a resolution of 0.1 nm within a volume of 25 mm x 25 mm x 5 mm.

White light interferometry can profit from these features and can take over the device's very high precision and large effective range.

Individual effective areas can be exactly positioned on the lateral level so that they become adjacent to each other and can be stitched together to a common height data map without the use of CPU-intensive registration algorithms, which effectiveness in general are depended on the topology of the measuring object. The article deals with therefore necessary calibration methods to determine and to consider the alignment of the white light sensor according to the machine coordinate system.

### Motivation

At the Technische Universität Ilmenau (TU Ilmenau) a nanopositioning and nanomeasuring machine (NPMM) has been continuously developed within the scope of the collaborative research centre SFB 622. The fundamental and innovative concept of the NPMM is the realization of the Abbe comparator principle in all three measuring axes [1] – that means that the measuring probe and the measuring beams of the machine must be aligned. This has the effect that systematic and random tilting of the guiding elements, also called as first-order tilt errors, are avoided.

The consequent observance of the principle has made possible a state-of-the-art nanopositioning and nanomeasuring machine with a measuring volume of 25 mm x 25 mm x 5 mm and a resolution of 0.1 nm [1, 2]. Besides the large effective range and the

high resolution a significant advantage of the device is, that several different measuring methods can be applied. The different types of sensors are quickly and easily changeable because of their modular conception.

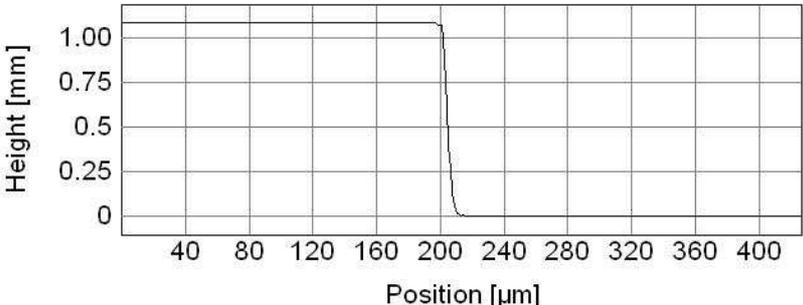
So far atomic force microscopes, focus sensors as well as capacitive and inductive contact systems are applicable sensor types for the NPMM [3]. However, data capture for these sensors is limited to one dimension. This also means that the measurement time requirements on these sensors cannot ever be met.

The integration of white light interferometry into the NPMM [4] is one of the current research and development topics of the Computer Graphics Group (TU Ilmenau).

White light interferometry is a powerful optical measuring method that allows the capturing of a whole surface with very high precision within a measurement time of a few seconds, depending on the topology of the measuring probe.

White light interferometry can profit from the high positioning resolution and large measuring volume of the NPMM. Hence, the main disturbing influence to the measurement's accuracy, the positioning noise, almost can be avoided.

In contrast to modern conventional white light interferometers, which perpendicular pass trough ranges are limited in general to 100  $\mu\text{m}$ , the combination with the NPMM allows the measurement of height differences up to 5 mm. The measuring time can be accelerated significantly by skipping the height regions where no fringes occur with a coarse speed (up to 50 mm/s [2]) during the perpendicular scan (e. g. the range between the top and the bottom of a step height). Fig. 1 shows an average profile of an in this manner measured structure with step height over 1 mm. This measuring has taken only a few seconds.



**Fig. 1 Profile of a 1.089 mm step height standard, measured with each a pass trough range on the bottom and on the top of the sample**

The nanometre-precision lateral positioning of the NPMM permits accurate stitching for the analysis of large areas. The most white light interferometry applications, who provide

the stitching technique, are using matching algorithms to register height data maps, which have to be overlapped to each other. In the overlaid sections, the observed structure must show significant features, like edges, otherwise the registration don't works correctly. All uncertainties from single registration steps have an adding up influence to the total lateral positioning inaccuracy. These disadvantages can be avoided by an exact positioning of adjacent individual effective areas without matching algorithms.

### **Adjustment and consideration of the perpendicular sensor alignment**

An important condition for measuring large ranges in vertical direction by white light interferometry is, that the sensor has to be aligned orthogonal in relation to the machine coordinate system. If the sensor, which contains a CCD camera, is tilted, than the lateral observed position shifts away proportional to the vertical moving distance.

The whole measuring head can be aligned manually by adjusting screws in consideration of fringes (a flat mirror is suggested as a measuring object for this adjustment).

It is improbable that a perfect alignment can be reached by a manually adjustment, the tilt can only be minimized.

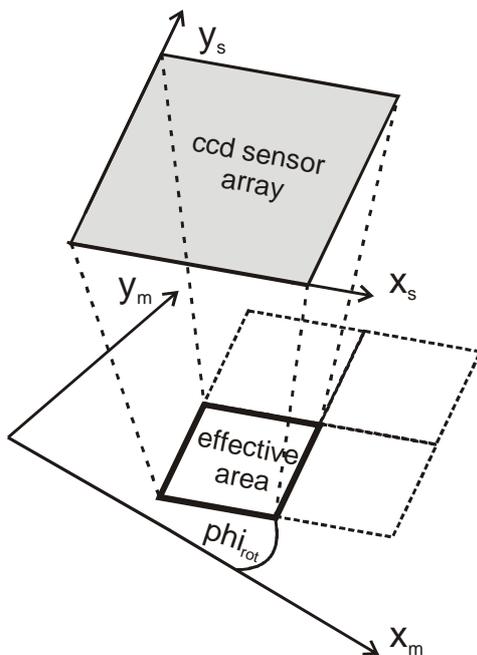
If the remaining tilt of the sensor in accordance to the machine coordinates is not considered, than the resulting height data from each individual measurement are displaced in the height direction to each other by a constant offset depended from the lateral position of the individual effective area. These offsets have to be corrected, if the results from adjacent measurements should be stitched.

### **Calibration of the lateral sensor alignment**

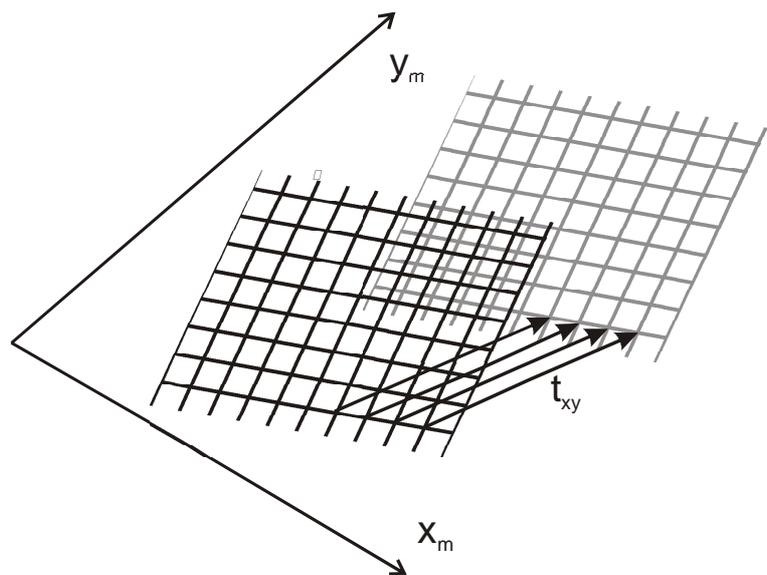
In order to allow an exact probe placement in respect to the lateral position of the sensor's effective area the transformation parameters from the sensor to the machine coordinates have to be determined.

Therefore a calibration procedure is performed in the run-up to the measurements for the estimation of the alignment and the scale factor. Under the assumption, that the sensor is justified perpendicular to the machine coordinate system (due to the manual adjustment), the alignment can be simplified characterized by a rotation about the angle  $\phi_{i_{rot}}$  (Fig. 2) on the lateral level.

The idea of the calibration method is to register a list of pass points between two images of the same target object, which have been captured at different lateral positions. An average translation vector is calculated over all corresponding point pairs. Between the known translation vector of the machine coordinates and the translation vector between the pass points from the images the rotation angle and the scale factor can be determined.



**Fig. 2** Relation between the sensor- ( $x_s$ ,  $y_s$ ) and the machine coordinates ( $x_m$ ,  $y_m$ )

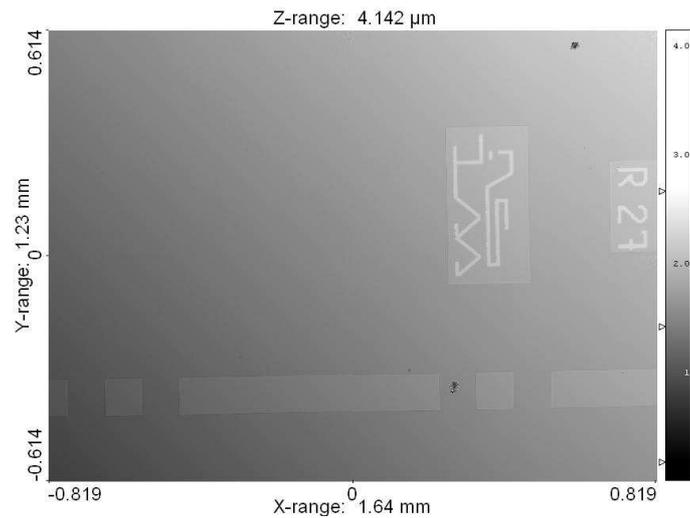


**Fig. 3** Lateral translation of the calibration grid by vector  $t_{xy}$

As a calibration target a grid structure is used, which crossing points are threaded as the points of interest. The gridlines are robust detectable by means of the Hough-transformation. The Hough-transformation is significantly speeded up, because the plausible search interval in the line parameter domain can be limited under the consideration of the previously rough estimated gradient direction at each pixel position. The assignment of the pass points between the two images is followed by the Nearest-Neighbor-principle, which assumes that the transformation parameters are a priori roughly known. The necessary precision of the a priori transformation parameters is depended on the line spacing of the grid and the translation distance between the two images.

## Exemplary results

A layer thickness standard was measured. The results of four individual measurements of adjacent effective areas were stitched to a common height data map with a lateral effective area of  $1.64 \times 1.23 \text{ mm}^2$  and a lateral resolution of  $1.25 \text{ Pixel}/\mu\text{m}$ , which is shown below in Fig. 4. The remaining tilt of the sensor was corrected for each measurement by adding a constant height offset.



**Fig. 4** Stitched height image from a layer thickness standard

## Conclusion

A calibration procedure for determining the lateral alignment and scale of the white light interferometry measuring sensor have been put into practice by means of a robust and sub pixel precise pass point assignment between the images of a grid at different lateral positions. Therewith the conditions for an exact lateral positioning in accordance to the sensor coordinates are complied.

Therefore, in contrast to conventional stitching methods, the results from individual measurements at adjacent lateral positions can be linked without matching algorithms, which precision is depending in general from the topology of the measuring object. The developments were able to be exemplary proven by stitching of the adjacent measurement results from a layer thickness standard to a common height data map (Fig. 4).

## Acknowledgments

The project white light interferometry for nanopositioning and nanomeasuring machines has been sponsored by the ministry of education and arts of the Free State of Thuringia (Germany) under the sign B 514-06 007. The authors wish to thank all those colleagues at the Technische Universität Ilmenau and the ZBS Ilmenau e. V., who have contributed to these developments.

### References:

- [1] Nanopositioning and Measuring Machine, Proceedings of the 2nd euspen International Conference (pages 290-293), European Society for Precision Engineering and Nanotechnology, 2001.
- [2] Nanopositionier- und Nanomessmaschine, data sheet 10/ 2005, SIOS Messtechnik GmbH, Ilmenau, 2005.
- [3] G. Jäger, E. Manske and T. Hausotte, Neue Anwendungen der Nanomessmaschine (NPM-Maschine) durch die Entwicklung nanoskaliger optischer und taktile Tastsensoren, Technisches Messen 73 (2006) 9, Oldenburg Verlag, 2006.
- [4] D. Kapusi, T. Machleidt, K.-H. Franke, R. Jahn, White light interferometry in combination with a nanopositioning and nanomeasuring machine (NPM), Proceedings SPIE Optical Metrology, Paper Number: 6616-4, Munich, 2007.

### Authors:

Dipl.-Ing. Daniel Kapusi  
TU-Ilmenau, EAZ Ehrenbergstr. 29, Room 0322  
98693, Ilmenau  
Phone: (0)493677 695073  
Fax: (0)493677 695071  
E-mail: [daniel.kapusi@tu-ilmenau.de](mailto:daniel.kapusi@tu-ilmenau.de)

Dipl.-Ing. Torsten Machleidt  
TU-Ilmenau, EAZ Ehrenbergstr. 29, Room 0321  
98693, Ilmenau  
Phone: (0)493677 695068  
Fax: (0)493677 695071  
E-mail: [torsten.machleidt@tu-ilmenau.de](mailto:torsten.machleidt@tu-ilmenau.de)

PD Dr.-Ing. habil. Karl-Heinz Franke  
TU-Ilmenau, APZ Gustav-Kirchhoff-Str. 5  
98693, Ilmenau  
Phone: (0)493677 2010300  
Fax: (0)493677 2010302  
E-mail: [karl-heinz.franke@tu-ilmenau.de](mailto:karl-heinz.franke@tu-ilmenau.de)

Dr.-Ing. habil. Eberhard Manske  
TU-Ilmenau, Gustav-Kirchhoff-Str. 1, Room K2004  
98693, Ilmenau  
Phone: (0)493677 691250  
Fax: (0)493677 691412  
E-mail: [eberhard.manske@tu-ilmenau.de](mailto:eberhard.manske@tu-ilmenau.de)

Dr.-Ing. Rainer Jahn  
ZBS Ilmenau e. V., APZ Gustav-Kirchhoff-Str. 5  
98693, Ilmenau  
Phone: (0)493677 2010310  
Fax: (0)493677 2010302  
E-mail: [rainer.jahn@zbs-ilmenau.de](mailto:rainer.jahn@zbs-ilmenau.de)

D. Kapusi / T. Machleidt / K.-H. Franke / R. Jahn

## Measuring large areas by white light interferometry at the nanopositioning and nanomeasuring machine (NPMM)

### Abstract

White light interferometry is a new application for the nanopositioning and nanomeasuring machine (NPMM). The NPMM was developed under the leadership of the Institute of Process Measurement and Sensor Technology at the Technische Universität Ilmenau (Germany) and allows highly exact dimensional and traceable positioning with a resolution of 0.1 nm within a volume of 25 mm x 25 mm x 5 mm.

White light interferometry can profit from these features and can take over the device's very high precision and large effective range.

Individual effective areas can be exactly positioned on the lateral level so that they become adjacent to each other and can be stitched together to a common height data map without the use of CPU-intensive registration algorithms, which effectiveness in general are depended on the topology of the measuring object. The article deals with therefore necessary calibration methods to determine and to consider the alignment of the white light sensor according to the machine coordinate system.

### Motivation

At the Technische Universität Ilmenau (TU Ilmenau) a nanopositioning and nanomeasuring machine (NPMM) has been continuously developed within the scope of the collaborative research centre SFB 622. The fundamental and innovative concept of the NPMM is the realization of the Abbe comparator principle in all three measuring axes [1] – that means that the measuring probe and the measuring beams of the machine must be aligned. This has the effect that systematic and random tilting of the guiding elements, also called as first-order tilt errors, are avoided.

The consequent observance of the principle has made possible a state-of-the-art nanopositioning and nanomeasuring machine with a measuring volume of 25 mm x 25 mm x 5 mm and a resolution of 0.1 nm [1, 2]. Besides the large effective range and the

high resolution a significant advantage of the device is, that several different measuring methods can be applied. The different types of sensors are quickly and easily changeable because of their modular conception.

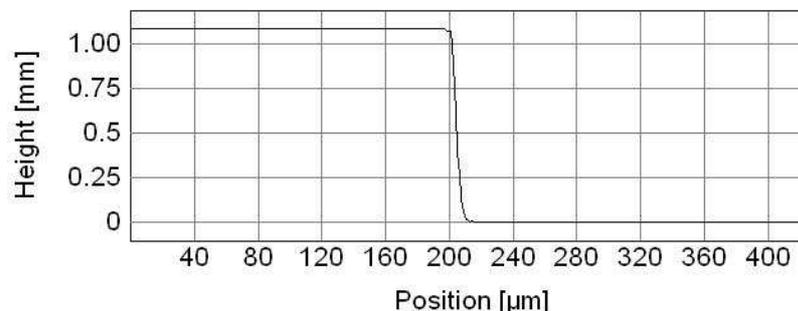
So far atomic force microscopes, focus sensors as well as capacitive and inductive contact systems are applicable sensor types for the NPMM [3]. However, data capture for these sensors is limited to one dimension. This also means that the measurement time requirements on these sensors cannot ever be met.

The integration of white light interferometry into the NPMM [4] is one of the current research and development topics of the Computer Graphics Group (TU Ilmenau).

White light interferometry is a powerful optical measuring method that allows the capturing of a whole surface with very high precision within a measurement time of a few seconds, depending on the topology of the measuring probe.

White light interferometry can profit from the high positioning resolution and large measuring volume of the NPMM. Hence, the main disturbing influence to the measurement's accuracy, the positioning noise, almost can be avoided.

In contrast to modern conventional white light interferometers, which perpendicular pass trough ranges are limited in general to 100  $\mu\text{m}$ , the combination with the NPMM allows the measurement of height differences up to 5 mm. The measuring time can be accelerated significantly by skipping the height regions where no fringes occur with a coarse speed (up to 50 mm/s [2]) during the perpendicular scan (e. g. the range between the top and the bottom of a step height). Fig. 1 shows an average profile of an in this manner measured structure with step height over 1 mm. This measuring has taken only a few seconds.



**Fig. 1 Profile of a 1.089 mm step height standard, measured with each a pass trough range on the bottom and on the top of the sample**

The nanometre-precision lateral positioning of the NPMM permits accurate stitching for the analysis of large areas. The most white light interferometry applications, who provide

the stitching technique, are using matching algorithms to register height data maps, which have to be overlapped to each other. In the overlaid sections, the observed structure must show significant features, like edges, otherwise the registration don't works correctly. All uncertainties from single registration steps have an adding up influence to the total lateral positioning inaccuracy. These disadvantages can be avoided by an exact positioning of adjacent individual effective areas without matching algorithms.

### **Adjustment and consideration of the perpendicular sensor alignment**

An important condition for measuring large ranges in vertical direction by white light interferometry is, that the sensor has to be aligned orthogonal in relation to the machine coordinate system. If the sensor, which contains a CCD camera, is tilted, than the lateral observed position shifts away proportional to the vertical moving distance.

The whole measuring head can be aligned manually by adjusting screws in consideration of fringes (a flat mirror is suggested as a measuring object for this adjustment).

It is improbable that a perfect alignment can be reached by a manually adjustment, the tilt can only be minimized.

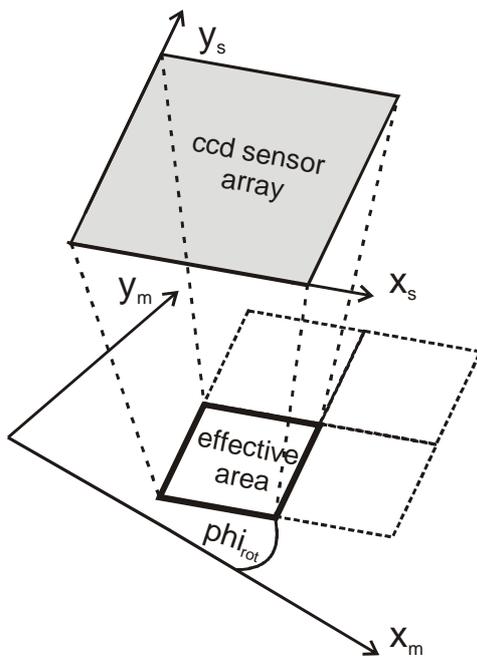
If the remaining tilt of the sensor in accordance to the machine coordinates is not considered, than the resulting height data from each individual measurement are displaced in the height direction to each other by a constant offset depended from the lateral position of the individual effective area. These offsets have to be corrected, if the results from adjacent measurements should be stitched.

### **Calibration of the lateral sensor alignment**

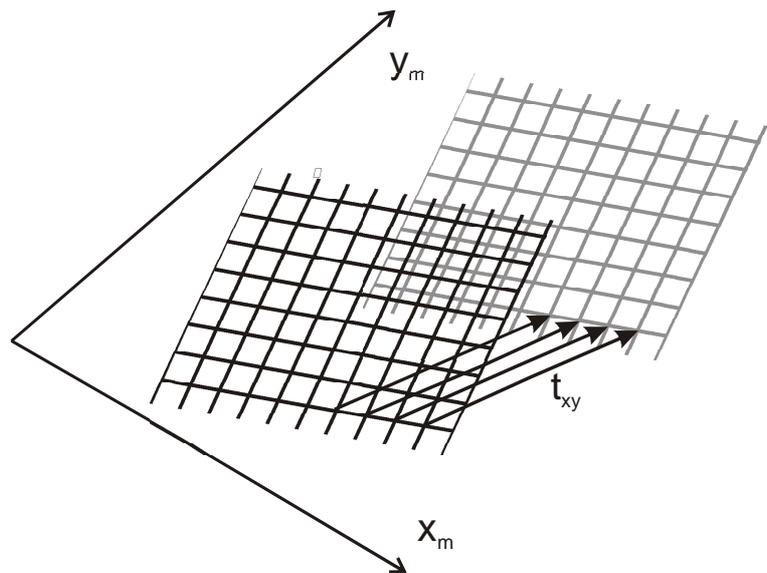
In order to allow an exact probe placement in respect to the lateral position of the sensor's effective area the transformation parameters from the sensor to the machine coordinates have to be determined.

Therefore a calibration procedure is performed in the run-up to the measurements for the estimation of the alignment and the scale factor. Under the assumption, that the sensor is justified perpendicular to the machine coordinate system (due to the manual adjustment), the alignment can be simplified characterized by a rotation about the angle  $\phi_{i_{rot}}$  (Fig. 2) on the lateral level.

The idea of the calibration method is to register a list of pass points between two images of the same target object, which have been captured at different lateral positions. An average translation vector is calculated over all corresponding point pairs. Between the known translation vector of the machine coordinates and the translation vector between the pass points from the images the rotation angle and the scale factor can be determined.



**Fig. 2** Relation between the sensor- ( $x_s$ ,  $y_s$ ) and the machine coordinates ( $x_m$ ,  $y_m$ )

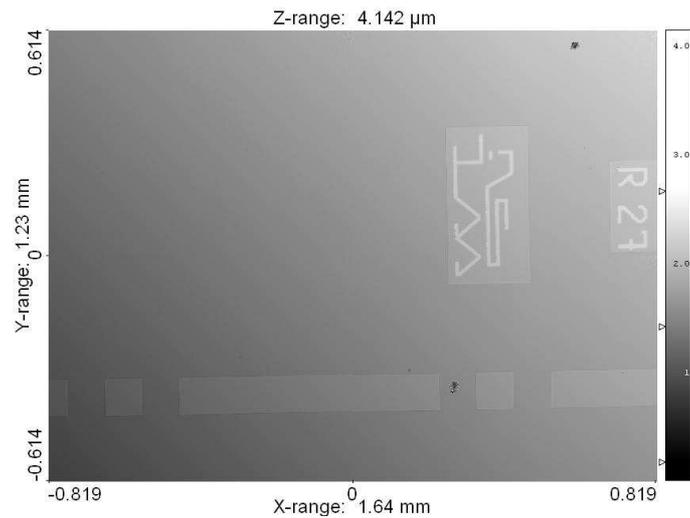


**Fig. 3** Lateral translation of the calibration grid by vector  $t_{xy}$

As a calibration target a grid structure is used, which crossing points are threaded as the points of interest. The gridlines are robust detectable by means of the Hough-transformation. The Hough-transformation is significantly speeded up, because the plausible search interval in the line parameter domain can be limited under the consideration of the previously rough estimated gradient direction at each pixel position. The assignment of the pass points between the two images is followed by the Nearest-Neighbor-principle, which assumes that the transformation parameters are a priori roughly known. The necessary precision of the a priori transformation parameters is depended on the line spacing of the grid and the translation distance between the two images.

## Exemplary results

A layer thickness standard was measured. The results of four individual measurements of adjacent effective areas were stitched to a common height data map with a lateral effective area of  $1.64 \times 1.23 \text{ mm}^2$  and a lateral resolution of  $1.25 \text{ Pixel}/\mu\text{m}$ , which is shown below in Fig. 4. The remaining tilt of the sensor was corrected for each measurement by adding a constant height offset.



**Fig. 4** Stitched height image from a layer thickness standard

## Conclusion

A calibration procedure for determining the lateral alignment and scale of the white light interferometry measuring sensor have been put into practice by means of a robust and sub pixel precise pass point assignment between the images of a grid at different lateral positions. Therewith the conditions for an exact lateral positioning in accordance to the sensor coordinates are complied.

Therefore, in contrast to conventional stitching methods, the results from individual measurements at adjacent lateral positions can be linked without matching algorithms, which precision is depending in general from the topology of the measuring object. The developments were able to be exemplary proven by stitching of the adjacent measurement results from a layer thickness standard to a common height data map (Fig. 4).

## Acknowledgments

The project white light interferometry for nanopositioning and nanomeasuring machines has been sponsored by the ministry of education and arts of the Free State of Thuringia (Germany) under the sign B 514-06 007. The authors wish to thank all those colleagues at the Technische Universität Ilmenau and the ZBS Ilmenau e. V., who have contributed to these developments.

### References:

- [1] T. Hausotte, Nanopositionier- und Nanomessmaschine, dissertation, TU Ilmenau, 2002.
- [2] Nanopositionier- und Nanomessmaschine, data sheet 10/ 2005, SIOS Messtechnik GmbH, Ilmenau, 2005.
- [3] G. Jäger, E. Manske and T. Hausotte, Neue Anwendungen der Nanomessmaschine (NPM-Maschine) durch die Entwicklung nanoskaliger optischer und taktile Tastsensoren, Technisches Messen 73 (2006) 9, Oldenburg Verlag, 2006.
- [4] D. Kapusi, T. Machleidt, K.-H. Franke, R. Jahn, White light interferometry in combination with a nanopositioning and nanomeasuring machine (NPM), Proceedings SPIE Optical Metrology, Paper Number: 6616-4, Munich, 2007.

### Authors:

Dipl.-Ing. Daniel Kapusi  
TU-Ilmenau, EAZ Ehrenbergstr. 29, Room 0322  
98693, Ilmenau  
Phone: (0)493677 695073  
Fax: (0)493677 695071  
E-mail: [daniel.kapusi@tu-ilmenau.de](mailto:daniel.kapusi@tu-ilmenau.de)

Dipl.-Ing. Torsten Machleidt  
TU-Ilmenau, EAZ Ehrenbergstr. 29, Room 0321  
98693, Ilmenau  
Phone: (0)493677 695068  
Fax: (0)493677 695071  
E-mail: [torsten.machleidt@tu-ilmenau.de](mailto:torsten.machleidt@tu-ilmenau.de)

PD Dr.-Ing. Karl-Heinz Franke  
TU-Ilmenau, APZ Gustav-Kirchhoff-Str. 5  
98693, Ilmenau  
Phone: (0)493677 2010300  
Fax: (0)493677 2010302  
E-mail: [karl-heinz.franke@tu-ilmenau.de](mailto:karl-heinz.franke@tu-ilmenau.de)

Dr.-Ing. Rainer Jahn  
ZBS Ilmenau e. V., APZ Gustav-Kirchhoff-Str. 5  
98693, Ilmenau  
Phone: (0)493677 2010310  
Fax: (0)493677 2010302  
E-mail: [rainer.jahn@zbs-ilmenau.de](mailto:rainer.jahn@zbs-ilmenau.de)

R. Burdick / T. Lorenz / K. Bobey

## **Characteristics of High Power LEDs and one example application in white-light-interferometry**

### **ABSTRACT**

Even with binned LEDs there is a strong spread of characteristics. Therefore the selection, seasoning or homogenization of LEDs are important tasks to be solved in measurement applications. Characteristics of LEDs like uniformity and long-term stability for LED measurement applications are discussed in this paper. Furthermore their use in the measurement process, in particular the influences of some LED characteristics on a white-light-interferometer (WLI) line scan sensor are discussed.

### **INTRODUCTION**

LEDs have found their application in general lighting and illumination applications today. Examples are LED spot lights for surgery or LED headlamps in the automotive environment, where they were employed for the first time during the past two years. This strongly drives the development of white and warm white high power LEDs. At the beginning of this year Osram announced that for the first time in the Migros supermarket in Eschenbach (Switzerland), LEDs are used not only for extraordinary effects but also to provide basic lighting. Considering the characteristics and working conditions of this type of LEDs they are very well suited for the use in measurement devices like a WLI. Devices with high luminous efficiency are for example the Seoul Semiconductor P4 and the Cree XLamp-XR LEDs with approx. 70 - 80 lm/W. Depending on the heat sink these single emitter LEDs allowed an absolute maximum electrical power dissipation of 4 W. A long term stable LED Light source, developed by the authors, enabled - together with image-taking colorimeters - quick and reproducible measurements of LED lamp distributions [1].

## LED CHARACTERISTICS

LEDs are solid state light sources whose lighting characteristics are completely different from those of conventional incandescent lamps. The tungsten filament of an incandescent lamp used in projection systems is usually precisely positioned at the stage of gluing the glass bulb into its socket. This, however, is not possible with LEDs, where a misalignment of the LED-dye in its plastic encapsulation can lead to non-axial symmetric radiation patterns, as can be seen in Fig. 1 for the example of a T-1 $\frac{3}{4}$  housing.

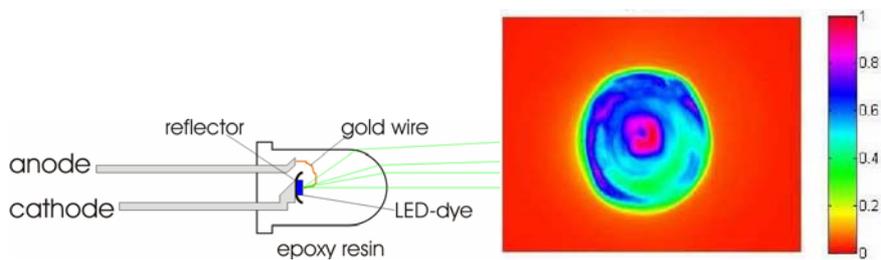


Figure 1: T-1 $\frac{3}{4}$  LED housing and a typical radiation pattern.

The spectral distribution and the intensity are temperature-dependent, with the intensity being subject to degradation processes. P-n heterojunctions are commonly employed in LEDs where the carriers are confined by the heterojunction barriers. This increases the carrier concentration, because the active region is by orders of magnitude smaller than the diffusion length. Therefore the radiative recombination rate increases and leads to devices with high external efficiencies as can be seen in Fig. 2.

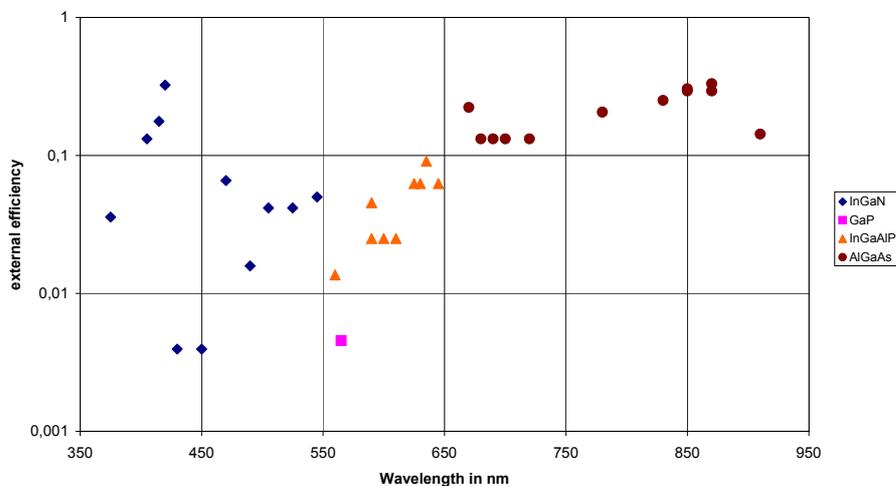


Figure 2: External efficiencies of different III-V-LED materials from Epitex Inc.

Unfortunately some problems can arise with heterostructures. One of them is the heating of the active region caused by heterostructure resistance of the abrupt heterointerface, especially in high-power devices. This can be avoided by a complex grading process. Another problem is the carrier loss beyond the active region, called carrier leakage effect. The energy distribution of carriers is given by the Fermi-Dirac distribution. Thus, the fraction of carriers residing in the active region which have a higher energy than the energy of the barrier is temperature-dependent. With increasing temperature the exponentially temperature-dependent carrier leakage effect leads to a decreasing internal efficiency. With large-band gap materials like InGaN this effect is minor, so that these LEDs can be used at higher temperatures. The emission spectrum of LEDs is also temperature dependent. With increasing temperature the peak wavelength decreases for InGaN-LEDs while it increases for GaAlInP-LEDs.

The generation of white light with LEDs can be achieved in three different ways. The light from three LEDs with the colors red, green and blue mixed in a given fraction of intensity is perceived as white. This is also possible with one UV-LED and three phosphors. The most common way to generate white light is to combine one blue LED with one  $Y_3Al_5O_{12}$  (yttrium aluminium garnet or YAG) phosphor of the complementary wavelength or more phosphors as can be seen in the spectral distribution in fig. 3.

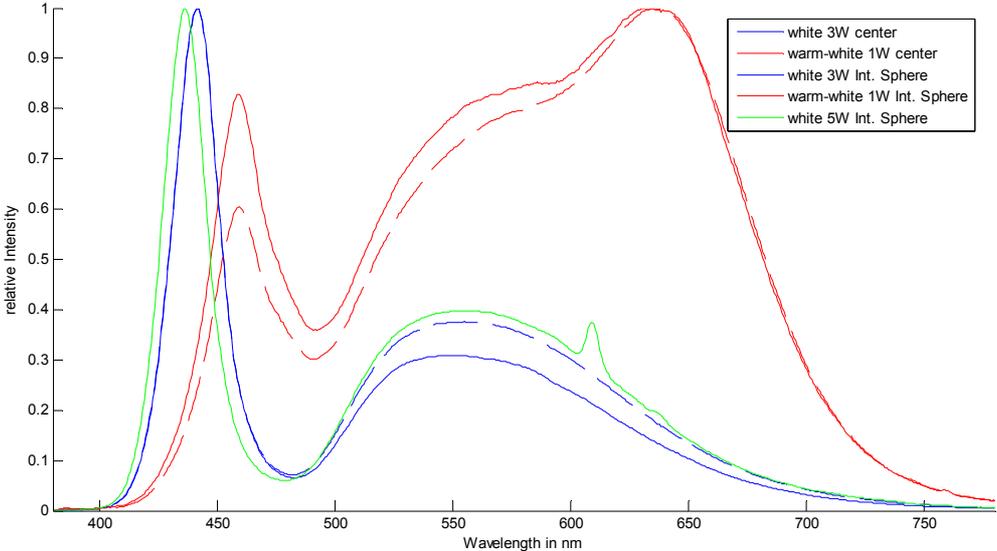


Figure 3: Spectral distribution of various Luxeon star LEDs in the center axis and averaged with an integrating sphere.

Presently, the blue LED chip with a yellow phosphor provides the highest luminous

efficacy. However, the ability to render all the colors of an illuminated object is poor. The radiation pattern of phosphor is, depending on the deposition process, nearly a lambertian distribution while the blue LED has a distribution under a small angle, which leads to more or less spatial color changes (see fig. 3).

Generally, the life time of mature technology LEDs like GaAlInP is very long, lying around some ten thousand hours. The ageing of an LED describes the loss of intensity over the whole operation time, as can be seen in fig. 4. The lifetime is usually defined as a loss of rated luminous flux of 30 % or 50 %. The measurement conditions for lifetime estimation have not been standardized yet. To understand the mechanisms behind ageing it is helpful to differentiate between internal and optical efficiency. The ageing of the optical efficiency is a function of ambient conditions such as moisture and mechanical stress. The ageing of the internal efficiency is at least a two-step mechanism.

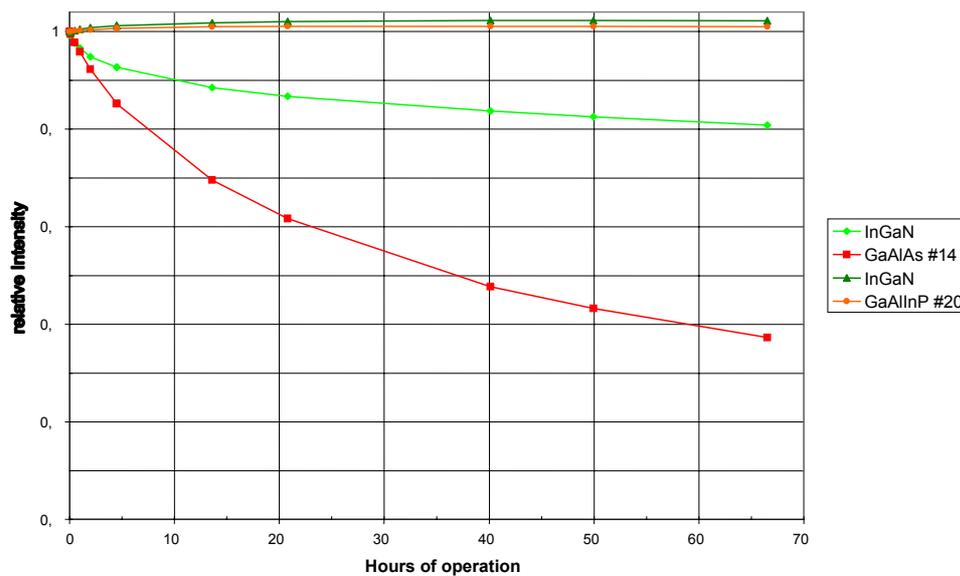


Figure 4: Ageing of different LED materials.

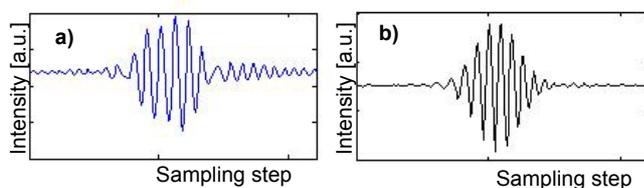
During the first hours of operation, an increase in intensity can sometimes be observed. This behavior depends on the forward current. For GaAlInP-compounds it might be related to a reduced carrier leakage[2]. For InGaN-compounds, a connection with the activation of magnesium acceptors due to the destruction of residual Mg-H complexes has been found [3]. After this initial process, the long-term behavior leads to a continuous decrease in the intensity, which slows down over time. The degradation is a function of the temperature and of the forward current in the active region, mainly caused by an increasing rate of nonradiative recombinations and carrier leakage effects.

As shown in a publication by OSRAM [4] for GaAlInP-compounds, the degradation is also color dependent.

### WLI application

WLI is a key technology for highly accurate inspection of nearly any kind of surfaces regardless of their textures. This means that smooth as well as rough surfaces are accessible to WLI. A white-light-interferometer mainly based on a Michelson arrangement uses a short coherent light source and thus, interference fringes can only be observed if the optical path difference between the two arms is less than the coherence length.

Replacing one mirror by the object and scanning its surface in depth i.e. by moving the reference mirror along the optical axis leads to a cosinusoidal intensity modulation weighted by the fringe visibility function at each point in the image plane [5]. The scan position at which the fringe visibility has its maximum yields the height of the corresponding point on the surface. Therefore, the main task to obtain height information is to find the maximum value of the envelope and its location with respect to the scanning axis for each interference pattern. As the coherence length depends on the spectral bandwidth, the use of a broad bandwidth illumination which results in a narrow envelope is necessary for high resolution measurement of surface heights. Furthermore, the shape of the envelope which is obtained from the spectral distribution by means of a Fourier transform should come close to a Gaussian distribution which can be easily fitted and hence simplifies the detection process. To archive these recommendations with a white High Power LED the blue peak in the spectral distribution of the LED must be suppressed. Otherwise, the superposition of two interference patterns one formed by the blue spectral range and the other one formed by the remaining spectral range give rise to a slightly distorted fringe pattern which lets increase uncertainty in finding the peak of the envelope (see Fig.5 (a)).



*Figure 5: White-light interference pattern formed by a white High Power LED without (a) and after suppression of the blue peak by use of a yellow 475nm longpass-filter (b)*

As shown in Fig.5 (b), the suppression of the blue peak by use of a yellow 475nm long-pass filter results in a symmetrical interference pattern with an envelope close to a Gaussian distribution. Another advantage of filtering out the blue peak is an improved spatial homogenous spectral distribution. Otherwise, it can be observed that more blue light is residing close to the optical axis than at the edge of the field of view. As a result, the center wavelength which forms the white-light interference pattern differs slightly in dependence of the location within the field of view. This can cause problems in WLI when measuring smooth surfaces and using phase information in conjunction with coherence information to increase the resolution.

As a conclusion one can say that with a simple filter technique spectral distribution can be improved to make white High Power LEDs capable in WLI. A white Seoul Semiconductor P4 is actually in use in a demonstrator of a line-profiling white-light interferometer developed at HAWK in cooperation with Mahr inc. Alternatively, a warm white P4 will be tested soon promising a larger amount of optical power in the long wavelength range due to further suppression of the blue peak.

#### **References:**

- [1] Burdick R, Krüger U, Schmidt F, Schanda J. Properties and Applications of Stable LED Light Sources in Automotive Measurement. International Symposium on Automotive Lighting, (2007)
- [2] Alteri P, Jaeger A, Windisch R, Lindner N, Stauss P, Oberschmid R, Streubel K. Internal quantum efficiency of high-brightness AlGaInP light-emitting devices. Journal of Applied Physics 98, 086101 (2005)
- [3] Manyakhin F, Kovalev A, Yunovich A.E. Aging Mechanisms of InGaN/AlGaIn/GaN Light-Emitting Diodes Operating at High Currents. MIJ-NSR Volume 3, Article 53. 1998
- [4] Alteri P, Jaeger A, Windisch R, Lindner N, Stauss P, Oberschmid R, Streubel K. Color-dependent degradation of high-brightness AlGaInP LEDs, SPIE 5349-416, 2004
- [5] Hariharan P. Optical Interferometry. Elsevier Academic Press, 2nd edition, 2003

#### **Authors:**

MSc. Robert Burdick  
Dipl.-Ing. Timo Lorenz  
Prof. Dr.-Ing. Klaus Bobey  
HAWK FH HHG, Von-Ossietzky-Str. 99  
37085, Göttingen  
Phone: (49)551-3705238  
Fax: (49)551-3705101  
E-mail: burdick@hawk-hhg.de

T. Koch / K.-H. Franke

## **Aspekte der strukturbasierten Fusion multimodaler Satellitendaten und der Segmentierung fusionierter Bilder**

### **Einführung**

Im Rahmen verschiedener Missionen zur Fernerkundung der Erdoberfläche entsteht multimodales und multitemporales Bildmaterial sowohl aus dem optischen Spektralbereich als auch Radar-Daten unterschiedlicher Polarität und Auflösung.

Diese Daten können zur Lösung vielfältiger Problemstellungen, wie z.B. die automatische Kartierung von Feldfruchtarten, die Ermittlung biophysikalischer Kenngrößen und die Erfassung der Änderungsdynamik, eingesetzt werden (s. [5],[6]).

Im Rahmen des Verbundprojektes ENVILAND<sup>1</sup> entstand ein Software-Prototyp, der diese Eingangsdaten durch eine Kette mehrerer Verarbeitungsstufen zu einem klassifizierten Labelbild verarbeitet. Wesentliche Teilaspekte<sup>2</sup> sind hierbei die Fusion des multimodalen Bildmaterials und dessen anschließende Segmentierung.

Dieser Beitrag soll einige Ergebnisse darstellen, die im Rahmen des Verbundprojektes ENVILAND bei dem *Zentrum für Bild- und Signalverarbeitung e.V. (ZBS)* entstanden sind.

### **Problemstellung**

Das von den Herstellern (ESA,NASA,...) gelieferte Datenmaterial kann zwar (meist manuell) geocodiert bezogen werden, jedoch ist diese Geocodierung in der Regel nicht so genau, dass damit eine pixelgenaue Registrierung durchgeführt werden kann. Dieses Problem verschärft sich insbesondere bei Verwendung von Bildmaterial verschiedener Hersteller oder unterschiedlicher Modalität (SAR, VIS, IR, ...), da eine jeweils verschiedene Grundlage<sup>3</sup> zur Geocodierung eingesetzt wird.

Die Fusion ist in der vorgesehenen Verarbeitungspipeline nach einer Geocodierung

---

<sup>1</sup> gefördert mit Mitteln des *Bundesministeriums für Wirtschaft und Technologie* (BMW) durch das *Deutsche Zentrum für Luft- und Raumfahrt* (DLR) unter der Fördernummer 50 EE 0406;

weitere Partner: *Jena-Optronik AG, FSU-Jena, ZFL-Bonn und Desotron GmbH*

<sup>2</sup> die Verarbeitungsstufen *Kantenextraktion, Fusion und Segmentierung* wurden durch das ZBS erstellt

<sup>3</sup> bezüglich Vorverarbeitungs-Algorithmen, digitales Geländemodell, manuell gesetzte Passpunkte, ...

und anschließender Kantenextraktion eingegliedert. Sie soll das geocodierte (und orthorektifizierte<sup>4</sup>) Bildmaterial pixelgenau registrieren (Bestimmung von Translation, Skalierung und Rotation eines affinen Modellansatzes) und dadurch einen Bildstapel für die anschließende Segmentierung liefern.

Aufgabe dieser Segmentierung ist die Erzeugung einer Karte homogener Regionen, die dazu dient, einer anschließenden segmentbasierten Klassifikation Regionenmerkmale (inkl. abgeleiteter Texturmerkmale) zu liefern. Mit Hilfe dieser Klassifikation werden die gesuchten Klassenzugehörigkeiten (z.Z. vier Grundklassen) ermittelt.

Hauptproblem der Registrierung von Bildmaterial unterschiedlicher Modalität ist die Tatsache, dass die physikalische Ursache der Bilder und die verwendeten Sensoren völlig verschieden sind. Daraus resultiert eine geringe Korrelation zwischen Bildern unterschiedlicher Modalität. Aus diesem Grund sind herkömmliche grauwertbasierte Korrelationsverfahren ungeeignet und fehleranfällig (Abbildung 4, links).

Weiterhin sind unterschiedliche Auflösungen sowie Rauscheinflüsse (insb. bei SAR) zu bewältigen. Typische zu prozessierende Bildgrößen sind 12000x12000 Pixel! Daraus ergeben sich spezielle Erfordernisse sowohl an die Komplexität der eingesetzten Algorithmen als auch an das Speicherhandling.

Prinzip-bedingte (und durch die Geocodierung nur teilweise kompensierte) lokale Verzerrungen sollen möglichst nur geringen störenden Einfluss zeigen. Diese lokalen Verzerrungen resultieren aus abbildungstypischen Eigenschaften der eingesetzten Sensorik (z.B. bei Radar: foreshortening und perspektivische Verzerrungen bei optischen Abbildungssystemen) die im Zusammenhang mit unterschiedlichen Geländehöhen entstehen.

Weitere spezielle sensortypische Störungen sind Wolken (VIS) und Radar-Schatten (SAR). Diese sollen während der Geocodierung erkannt und ausmaskiert werden.

Eine spezielle Herausforderung sind urbane Regionen. Diese sind gekennzeichnet durch eine hohe Variabilität, was zu Problemen in der Fusion und Segmentierung<sup>5</sup> führt.

### **Lösungsansatz zur Bildfusion**

Um das oben erwähnte Problem der geringen Korrelation zu lösen, müssen Strukturinformationen verwendet werden, welche in allen Modalitäten extrahierbar

---

<sup>4</sup> Herstellung eines homogenen Bezugs zwischen Pixelposition und geografischer Koordinate

<sup>5</sup> macht Einsatz geeigneter Texturfilter (wie z.B. NGLD [5]) oder spezieller Nachbearbeitung erforderlich

sind und in dem jeweils anderen Bildkanal mit hoher Wahrscheinlichkeit wiedergefunden werden können. Als eine solche gemeinsame Strukturinformation können die Übergangsbereiche (Kanten) an Segmentgrenzen benutzt werden. Um diese zu extrahieren, wird eine Filterung mit einem Deriche-Gradienten-Filter [4] durchgeführt. Eine *Non-Maximum*-Unterdrückung und die anschließende *Scan-Along*-Konturverfolgung (s. [1]) führen zur Kanteninformation.

In [3] wird ein Verfahren beschrieben, das eine kantenbasierte Korrelation durchführt. Es korreliert einzelne Kantenpunkte eines Bildkanals mit einem Invers-Distanz-Bild des jeweils anderen Kanals. Durch Variation der Transformationsparameter und anschließender Korrelation wird das beste Transformationsmodell ermittelt. Der Autor selbst schildert aber die „anziehende“ Wirkung stark strukturierter Regionen, die durch sehr hohe Korrelationswerte zur ungerechtfertigten Bevorzugung eben dieser Regionen führt. Die Ursache hierfür ist die glättende Eigenschaft der Invers-Distanz-Funktion, die über die Regionen verschmierte Korrelationswerte verursacht. Außerdem ist das Verfahren nicht determiniert, da nur zufällig gewählte Einzelpixel für die Korrelation Verwendung finden.

Um diese Probleme zu umgehen, wurde eine Alternative entwickelt, die auf dem in [3] beschriebenen Verfahren aufbaut.

Ausgangspunkt ist die Überlegung, dass Fehlkorrelationen hauptsächlich durch einander kreuzende statt durch deckungsgleich liegende Kanten gekennzeichnet sind. Deshalb wird im Gegensatz zu [3] nicht der Kantenort, sondern eine *kombinierte Betrags- und Richtungskorrelation* der jeweiligen Gradienten an den Kantenorten benutzt, die den Korrelationsbeitrag der Kanten minimiert, welche mit unterschiedlicher Orientierung überlagert würden.

Weiterhin werden jetzt alle Konturpixel in beiden Bildern berücksichtigt.

Das Neue an dem hier entwickelten Verfahren ist die mathematische Realisierung der Richtungs- bzw. Winkelkorrelation im Fourierbereich. Unter Zuhilfenahme einer schnellen Fouriertransformation (z.B. Radix-2, FFTW, o.ä.) wird dadurch eine Komplexität des Algorithmus von  $O(N \cdot \log N)$  erreicht (s. [2]). (Bisherige Verfahren berechnen eine Winkelkorrelation im Ortsbereich und erzielen dadurch eine Komplexität von  $O(N^2)$ !) Erst dadurch wird es möglich, die geforderten Bildgrößen in angemessener Zeit zu verarbeiten.

Konkret wird der Umstand ausgenutzt, dass eine Kreuz-Korrelation (KKF) als Faltung eines Bildes mit einem rotierten Kern aufgefasst werden kann. Diese stellt im

Fourierbereich eine Multiplikation des fouriertransformierten Bildes mit dem konjugiert-komplexen, fouriertransformierten Kern dar (Gleichung 1; [1], ff. ME5, ME15).

$$C(dx, dy) = \sum_{\forall x} \sum_{\forall y} A(x, y) B(x + dx, y + dy) = A(x, y) * B(-x, -y) \quad Q(x, y) = \frac{C(x, y) - \text{Im}_C(x, y)}{\min_{\forall x, y} C(x, y)}$$

$$C(dx, dy) = A(x, y) * B(-x, -y) \quad \text{O} \bullet F\{A(x, y)\} \cdot F^*\{B(x, y)\}$$

**Gleichung 1** Äquivalenz der KKF im Fourierbereich; mit  $F\{\bullet\}$  als Fouriertransformation; A,B als Quellbilder; C als Korrelationsbild

**Gleichung 2** Gütemaß Q ( $\text{Im}_C$  ist der Mittelwert von C in einer lokalen Umgebung um (x,y))

Die Translationsparameter lassen sich durch Suche nach dem Optimum (hinsichtlich eines Gütemaßes (Gleichung 2)) im Korrelationsbild bestimmen. Die Skalierungs- und Rotationsparameter werden durch wiederholte Ausführung der Korrelation, bei sukzessiver Variation derselben Parameter und gezielter Auswahl der besten Korrelationsvariante ermittelt.

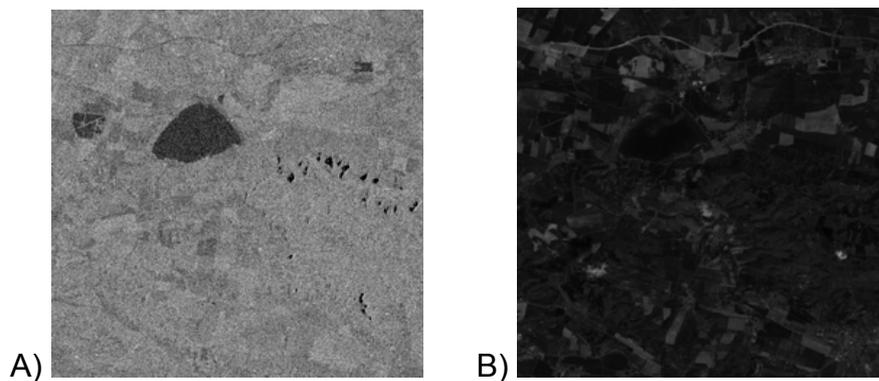
Wenn bereits eine initiale (aber ungenaue) Pre-Registrierung vorhanden ist, ist es meist nicht nötig, das komplette Korrelationsbild zu berechnen. Vielmehr ist ein kleiner Ausschnitt um die erwartete Position des Maximums von Interesse. Dieser kann bestimmt werden, indem beide Eingangsbilder in Kacheln zerlegt, jeweils zugehörige Kacheln der beiden Kanäle korreliert und die Korrelationsergebnisse anschließend aufakkumuliert werden. Hierbei ist aber zu beachten, dass eine einfache (nicht-überlappende) Kachelung eine unvollständige Korrelation in dem Sinne darstellt, dass Informationen in Randbereichen von Kacheln nicht mit denen der jeweiligen Nachbarkacheln in Beziehung gesetzt werden. Die Folge ist eine vom Zentrum des Korrelationsbildes zum Rand hin abnehmende Signifikanz der Korrelationswerte. Um dieses Problem zu lösen, wurde eine überlappende Kachelung eingesetzt, die in einer definierten Umgebung (um die Mitte des Korrelationsbildes) korrekte Werte ermittelt. Hierbei bestimmt die Breite der Überlappung die Größe des Signifikanzbereiches um das Zentrum des Korrelationsbildes. Die dabei erforderliche Kachelgröße kann cache-optimal realisiert werden.

### Bewertung

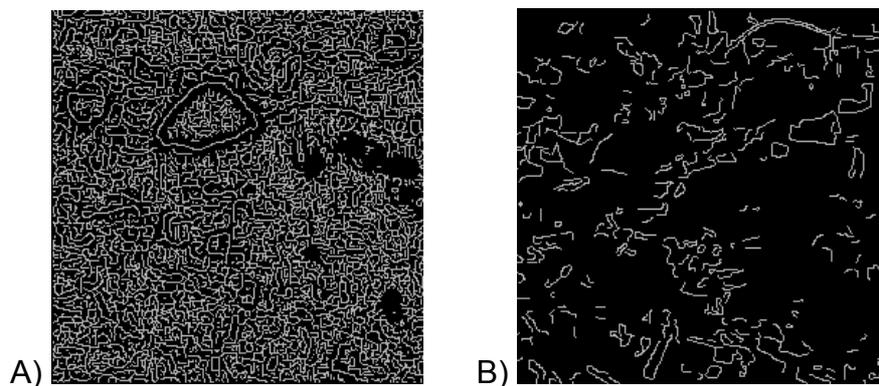
In Abbildung 4 sind die Ergebnisse verschiedener Verfahren dargestellt. Der hier prozessierte Ausschnitt (Abbildung 1) hat eine Bildgröße von 512×512 Pixel.

Allgemein gilt, je ausgeprägter der Peak im Korrelationsbild ist, um so sicherer kann eine korrekte Korrelation bestimmt werden. Es ist deutlich zu erkennen, dass die grauwertbasierte Korrelation (Abbildung 4, links) keine sichere Bestimmung der

Translationsparameter zulässt. Das Verfahren nach [3] liefert dagegen schon ein deutlich besseres Ergebnis (Abbildung 4, mitte). Allerdings wurde dieses, nicht wie in [3] beschrieben, auf Basis einer Korrelation von Einzelpixeln, sondern durch Korrelation aller Kantenpixel berechnet. Nach Meinung des Autors stellt dies somit den optimalen Grenzfall dieses Verfahrens dar.



**Abbildung 1** Beispiel zweier zu registrierender Kanäle (Ausschnitt aus einer Szene bei Nordhausen (A: ASAR<sup>6</sup> VV-Polarisation B: Landsat-5, Kanal 3))



**Abbildung 2** aus Originalbildern extrahierte Kantenbilder (links: Konturbild A, rechts: Konturbild B)

Das im Rahmen dieses Projektes entwickelte Korrelationsverfahren zeigt ein noch deutlicher ausgeprägtes Maximum (Abbildung 4, rechts), was auf eine generell sicherere Ableitung der Translationsparameter hindeutet.

Die Ausführungszeit<sup>7</sup> für die Bestimmung dieses Korrelationsbildes betrug weniger als 2 Sekunden. Wird die komplette Szene (korrelierter Ausschnitt: 4112×3292 Pixel) fusioniert, wird eine Ausführungszeit von ca. 45 Sekunden benötigt.

<sup>6</sup> die ENVISAT ASAR- und ERS-2-Daten wurden mit freundlicher Genehmigung der Europäischen Raumfahrt-Agentur (ESA) zur Verfügung gestellt (Category-1 Project C1P 3115)

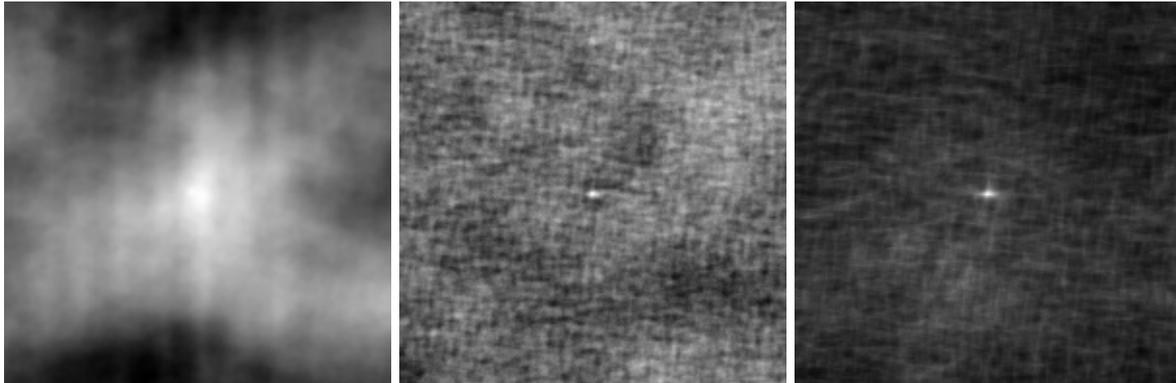
<sup>7</sup> System: Pentium IV; 2,53 GHz; 1GB RAM / Kachelgröße jeweils 512x512 Pixel



**Abbildung 3** Ausschnitt aus Konturbild A nach Bewertung mit Invers-Distanz-Funktion

$$D^{-1}(x,y) = \frac{1}{\left(1 + \frac{r(x,y)}{a}\right)^2}$$

**Gleichung 3** Inverse-Distanz  $D^{-1}$ ; mit  $r(x,y)$  als euklidischer Abstand von Pixel $(x,y)$  zu nächsten Konturpixel ;  $a$  ist Steuer-Parameter



**Abbildung 4** Korrelations-Ergebnisse (links: klassische Grauwert-Korrelation; mitte: Verfahren nach [3]; rechts: kombinierte Betrags- und Richtungskorrelation)

#### Literatur:

- [1] K.-H. Franke: „Grundlagen der digitalen Bildverarbeitung und Mustererkennung“, Schriftenreihe des ZBS e.V., ISSN 1432-3346, Ilmenau, 2004
- [2] Dieter Kreß, Ralf Irmer: „Angewandte Systemtheorie – Kontinuierliche und zeitdiskrete Verarbeitung“, Verlag Technik Berlin; 1989; ISBN 3-341-00693-1
- [3] Axel Pinz, Manfred Prantl, Harald Ganster: „A Robust Affine Matching Algorithm Using an Exponentially Decreasing Distance Function“, Journal of Universal Computer Science, vol. 1. no. 8, pp. 614-631;1995
- [4] R. Deriche, "Using Canny's criteria to derive an optimal edge detector recursively implemented," The International Journal on Computer Vision, Vol. 1, No. 2, April 1987, pp 167-187
- [5] Riedel, T., C. Thiel, C. Schmullius & K.-H. Franke (2006): "Extraktion von Landbedeckungsinformationen - ein automatisierter Ansatz mittels multitemporalen C-Band Daten"; 26. Wissenschaftlich-Technische Jahrestagung der DGPF, Berlin-Adlershof, 11. -- 13 September, CD-ROM
- [6] V. Heinzl, B. Waske, M. Braun, G. Menz (2005): "The potential of multitemporal and multisensoral remote sensing data for the extraction of biophysical parameters of wheat", SPIE Europe Bruges, Belgium, September, 2005, Proc. SPIE Vol. 5976, pp. 404-412

#### Autoren:

Torsten Koch  
 Karl-Heinz Franke  
 Zentrum für Bild- und Signalverarbeitung e.V.(ZBS), Gustav-Kirchhoff-Str. 5  
 98693 Ilmenau (Germany)  
 Phone: +49-3677-201030- 4 (Koch) / 0 (Franke)  
 Fax: +49-3677-2010301  
 E-mail: torsten.koch(at)zbs-ilmenau.de / karl-heinz.franke(at)zbs-ilmenau.de

T. Riedel / C. Thiel / C. Schmallius

## **A reliable and transferable classification approach towards operational land cover mapping combining optical and SAR data**

### **ABSTRACT**

The availability of up-to-date and reliable land cover maps is of great importance for many earth science applications. For the generation of operational land cover products the development of semi- and fully-automated classification strategies is essential. Aim of this paper is to make a contribution towards this problem. The study area is located in northern Thuringia, Germany, including mainly forested regions like the eastern part of the low mountain range Harz as well as intensively used agricultural areas. An object-based hierarchical classification scheme for the derivation of basic land cover classes using both optical and SAR data is presented. By the integration of textural features in the classification process the classification accuracy could be improved significantly, especially in the context of urban areas. In this study the power of the neighboring grey level dependency matrix – rarely used in the field of remote sensing – is demonstrated.

### **INTRODUCTION**

Earth observation represents a unique method for large area land cover mapping providing spatially consistent and multitemporal information. The availability of reliable and up-to-date land cover information is required for a multitude of regional to global applications such as land cover change studies, ecological monitoring, map updating or the control of national and international treaties [1]. For operational applications the development of robust, transferable, semi-automated and automated approaches is of special interest. In regions with frequent cloud cover such as Central Europe the number of suitable optical data is often limited. The all-weather capability is one major advantage of SAR data beyond optical systems. Furthermore, radar sensors provide information complementary to those contained in visible-infrared imagery. The benefit of combining optical and SAR data for land cover mapping was demonstrated in [2, 3, 4].

Emphasis of this study was to develop a robust and transferable methodology for the generation of basic land cover products including a limited number of optical data only, but exploiting the information content of multitemporal SAR data. Regarding operational applications the processing chain should comprise a high potential for automation.

## **EXPERIMENTAL DATA AND METHODOLOGY**

The study site is located in northern Thuringia, Germany including mainly forested areas characterized by rough topography like the low mountain range Harz as well as intensively used agricultural areas like the “Goldene Aue” east of Nordhausen. From April to December 2005 optical and SAR data were acquired continuously over the test site building up a comprehensive time series. HH/HV-polarized ASAR APP and ERS-2 data were recorded nearly simultaneously providing C-band data at all polarizations.

Emphasis of this study was to develop a robust and transferable classification scheme for the derivation of basic land cover maps. The proposed processing chain is composed of three main stages (Fig. 1). First, all EO-data were pre-processed on base of widely used standard techniques. As parts of the test site are characterized by significant topography, the normalization procedure introduced by Stussi et al. [5] was applied to all SAR data. Image segments were delineated on base of the optical EO-data using the multiresolution segmentation approach [6] implemented in the eCognition software. In the next processing step, potential training sites were selected automatically for each land cover class on base of a decision tree. As the application of fixed thresholds sometimes will fail, the thresholds values specified in the nodes of the decision tree will be adapted to each EO-scene separately. To achieve this, for each land cover type an optimal set of characteristic image parameters was defined by analyzing the time series available in a systematic manner. Additionally, information reported in literature and libraries (e.g. European RADar-Optical Research Assemblage library - ERA-ORA) were considered. By the combination of this expert know-ledge about typical target characteristics (e.g. low reflectance of water bodies in the near infrared) and histogram analyses, it is possible to assess scene-specific threshold values. In the third stage of the proposed classification scheme the identified trainings sites will be used as input for a supervised maximum-likelihood classification. The final land cover category assigned to each image object corresponds to the most frequent class per image segment.

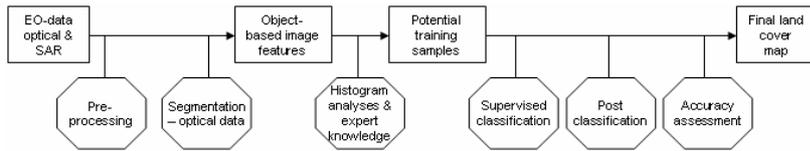


Figure 1: Proposed processing chain

For the detection of urban areas texture parameters are a valuable tool. In this study the potential of various texture measures such as standard deviation and parameters derived on base of the grey-level co-occurrence matrix were investigated. Additionally, the neighboring grey level dependency matrix (NGLD) was calculated [7]. The relation  $r$  used for the computation of the NGLD-matrix was defined as a difference of 0.1 between neighboring pixels at a distance  $d$  of two pixels.  $R$  was selected on base of a statistical analysis of various test areas indicating a proportion of > 40% of neighboring pixels with a difference of at least 0.1 in urban areas. Besides this, urban areas are characterized by a high radar backscatter at C-HH. To account for this behavior following parameter, hereinafter called UADP, was calculated on base of the NGLD matrix:

$$UADP = \sum_{a=0}^{gz-1} \sum_{n_r=0}^N Q_d(a, n_r) * n_r * a \quad (1)$$

where  $a$  = radar backscatter at HH-polarization  
 $n_r$  = number of pixels satisfying the relation  $r$   
 $Q_d$  = NGLD matrix value - number of occurrences of combination  $a$  and  $n_r$

The potential of the textural parameters investigated was evaluated by visual interpretation, separability analyses as well as by a comparison of the achieved map accuracies.

The quality of the final land cover products was assessed by calculating the confusion matrix and the kappa coefficient for fifty randomly distributed reference points per land cover category. The class membership of each target was specified on base of official land information GIS layers, high resolution optical and field data.

## RESULTS

As outlined in the previous chapter textural features in conjunction with spectral and backscatter information were used to map urban area extent. The power of texture information for such applications using medium resolution earth observation data was demonstrated by [8, 9, 10]. In this study several textural measures extracted from multitemporal optical and SAR data were investigated. Figure 2 illustrates eclectic

texture parameters derived from HH-polarized ASAR APP data acquired on April 22, 2005. Best overall performance and relatively stable results over time were found for the NGLD matrix approach, which was therefore used for the mapping of urban areas.

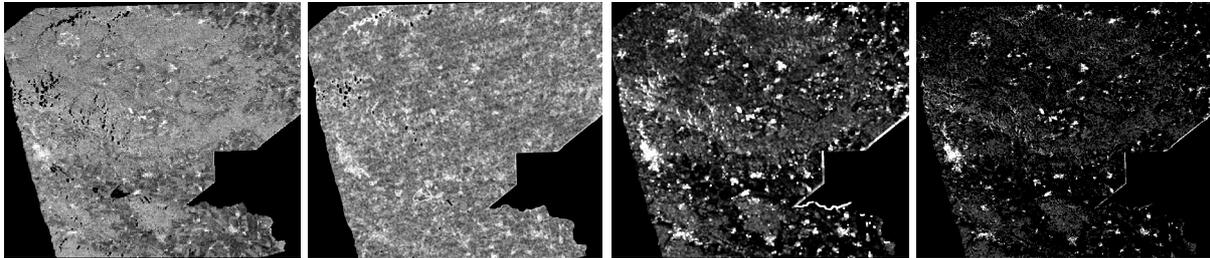


Figure 2: Eclectic texture features derived from HH-polarized ASAR APP data acquired on April 22, 2005 – from left: original image, co-occurrence entropy, standard deviation, UADP

As outlined above, the second step of the proposed processing strategy comprises the automatic selection of potential training samples on base of a decision tree. The absolute threshold values at each node of the decision tree are estimated for each optical scene separately making use of expert knowledge in conjunction with histogram analyses. For example, in agreement with the general knowledge of reflectance properties, the analyses of the time series available and the EO-libraries showed that coniferous forest areas are characterized by a low near (NIR) and middle infrared reflectance (MIR). Figure 3 shows two histograms representing the object mean values in NIR for all segments characterized by a low reflectance in MIR. Mainly coniferous forest and – depending on growth stage – agricultural crops are selected, thereby the lower peak represents coniferous forest segments and the higher one agricultural fields. The corresponding threshold values are assessed by a histogram clustering graph approach [13], whereby the final threshold is defined as the mean between the local maximum and the adjacent local minima (minimum frequency of 5 segments or 10% of the maximum histogram value). Table lists 1 the characteristic image parameters used to compute the thresholds values for each land cover class. The remaining category open land was subdivided into sub-classes with similar spectral characteristics by histogram analysis.

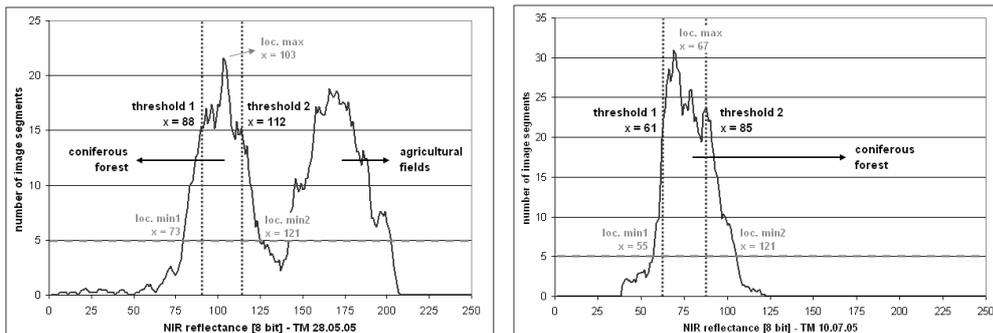


Figure 3: Threshold estimation by histogram analysis – example coniferous forest

| Class                         | Image Parameter                         |
|-------------------------------|---|
| Water bodies                  | NIR                                     |
| Coniferous forest             | NIR, MIR, NDVI                          |
| Dec. / mixed forest - leafoff | Green, NDVI, UADP                       |
| Dec. / mixed forest - leafon  | Green, NIR, MIR                         |
| Urban areas                   | NIR, MIR, UADP, minimum HV-polarisation |

Table 1: Characteristic image parameter used for the automatic selection of training samples

By the proposed methodology a large number of potential trainings sites will be selected. The quality of the obtained training samples was checked by a visual inspection and the calculation of the confusion matrix on base of the reference data used for the accuracy assessment of the final land cover map. For all sets of input data the user accuracy, i.e. the probability that a training sample is in agreement with the reference data, exceeds 90%. In the next processing step the training samples selected were used as input for a supervised, pixel-based maximum likelihood classification. The final land cover category assigned to each image object corresponds to the most frequent class per image segment. For validation issues the algorithm was tested for different sets of input data acquired over the Nordhausen study area in 2003 and 2005 as well as for EO-data acquired over a test site near Bonn (Table 2). The achieved classification accuracies are listed in Table 3 demonstrating the power of the proposed methodology for basic land cover mapping purposes. Regarding the SAR data the usage of ASAR APP scenes acquired at the early beginning of the growing season and after the main harvest and tillage period is highly recommended, as bare fields and fields covered by crops in an early growing stage are characterized by a low radar backscatter value. The multitemporal minimum in HV-polarisation is a useful parameter to reduce misclassifications between open land and urban or forested areas.

|       | Test site Nordhausen |                      |          |          | Test site near Bonn |                      |          |          |
|-------|----------------------|----------------------|----------|----------|---------------------|----------------------|----------|----------|
|       | Landsat-TM/ETM       | ASAR APP IS2 – HH/HV |          |          | Landsat TM / Spot*  | ASAR APP IS2 – HH/HV |          |          |
|       | Date 1               | Date2                | Date3    |          | Date 1              | Date2                | Date3    |          |
| Set 1 | 21.04.05             | 22.04.05             | 10.07.05 | 24.08.05 | 03.04.05            | 12.03.06             | 02.05.06 | 30.07.06 |
| Set 2 | 10.07.05             | 22.04.05             | 14.08.05 | 18.09.05 | 28.05.05            | 12.03.06             | 06.06.06 | 30.07.06 |
| Set 3 | 06.08.03             | 11.06.03             | 20.08.03 | 14.09.03 | 24.06.06*           | 16.04.06             | 25.06.06 | 30.07.06 |
| Set 4 | 22.04.03             | 16.07.03             | 20.08.03 | 14.09.03 |                     |                      |          |          |

Table 2: Input data sets used

|                  | Test site Nordhausen |      |       |       |       |       |       |      | Test site near Bonn |       |       |       |       |      |
|------------------|----------------------|------|-------|-------|-------|-------|-------|------|---------------------|-------|-------|-------|-------|------|
|                  | Set1                 |      | Set2  |       | Set3  |       | Set4  |      | Set1                |       | Set2  |       | Set3  |      |
|                  | PA                   | UA   | PA    | UA    | PA    | UA    | PA    | UA   | PA                  | UA    | PA    | UA    | PA    | UA   |
| water            | 96.0                 | 96.0 | 94.1  | 100.0 | 98.0  | 100.0 | 96.0  | 98.0 | 94.6                | 100.0 | 94.6  | 100.0 | 94.6  | 92.1 |
| forest           | 100.0                | 99.0 | 99.0  | 89.2  | 100.0 | 99.0  | 100.0 | 99.0 | 94.5                | 92.0  | 97.2  | 84.5  | 96.4  | 90.8 |
| urban areas      | 78.0                 | 86.7 | 81.3  | 73.6  | 79.6  | 81.3  | 84.8  | 78.0 | 77.1                | 92.3  | 88.4  | 88.3  | 87.0  | 95.6 |
| open land        | 94.6                 | 92.1 | 83.1  | 91.1  | 93.9  | 93.3  | 91.9  | 94.4 | 97.3                | 92.9  | 91.5  | 96.3  | 96.2  | 95.4 |
| overall accuracy | 93.9                 |      | 89.0  |       | 94.2  |       | 93.9  |      | 92.7                |       | 92.0  |       | 94.4  |      |
| kappa            | 0.909                |      | 0.855 |       | 0.916 |       | 0.918 |      | 0.869               |       | 0.862 |       | 0.901 |      |

Table 3: Achieved classification accuracies – PA: producer accuracy / UA: user accuracy

## CONCLUSIONS AND OUTLOOK

A combined object-/pixel-based classification scheme for the generation of basic land cover maps providing a high potential for automation has been presented. The training samples used as input for a supervised maximum-likelihood classification were selected on base of an object-based decision tree with scene-specific, flexible thresholds. The absolute threshold values were calculated by the combination of expert knowledge and histogram analyses. In order to improve the mapping accuracy for urban areas textural features extracted from HH-polarized SAR data were incorporated in the classification procedure. For validation issues the methodology was applied to different sets of input data acquired over the Nordhausen test site in 2003 and 2005 as well as near Bonn.

With the availability of polarimetric and high-resolution spaceborne X- and L-band SAR data as provided by the TerraSAR-X and PALSAR ALOS mission a further improvement is expected for both the automatic selection of potential training samples as well as the final map accuracy. The generation of more detailed land cover maps is planned in near future, especially for urban and agricultural used areas.

### Acknowledgement:

The ENVISAT ASAR and ERS-2 data were provided courtesy of the European Space Agency (Category-1 Project C1P 3115). The Enviland project – subproject scale integration - is funded by the German Ministry Economy and Technology (MW) and the German Aerospace Centre (DLR) (FKZ 50EE0405). We would like to thank the Centrum for Remote Sensing on Land Applications (ZFL) for providing pre-processed EO-data acquired near Bonn.

### References:

- [1] Jensen, J. R., 2000. Remote sensing of the environment – an earth resource perspective. Prentice Hall, New Jersey.
- [2] Alparone, L., Baronti, S., Garzelli, A. & Nencini, F. (2004). Landsat ETM+ and SAR image fusion based on generalized intensity modulation. - IEEE Transactions on Geoscience and Remote Sensing, 42(12): 2832 – 2839.
- [3] Amarsaikhan, D. & Douglas, T. (2004). Data fusion and multisource image classification. Int. J. Remote Sensing, 25/17: 3529 – 3539.
- [4] Hegarat-Masclé, S. L., Quesney, A., Vidal-Madjar, D., Taconet, O., Normand, M. & Loumagne (2000). Land cover discrimination from multitemporal ERS images and multispectral Landsat images: a study case in an agricultural area in France. Int. J. RS, 21: 435 – 456..
- [5] Stussi N., A. Beaudoin, T. Castel & P. Gigord (1995). Radiometric correction of multi-configuration spaceborne SAR data over hilly terrain. In: Proceedings LandSAR, Toulouse, France, 10-13 October 1995, pp. 469-478.
- [6] Baatz, M. & Schäpe, A. (2000). Multiresolution Segmentation – an optimization approach for high quality multi-scale image segmentation. In: Strobl J et al. (Eds.): XII. Beiträge zum AGIT-Symposium Salzburg, Herbert Wichmann Verlag: 12–23.
- [7] Franke, K.-H. (2006). Grundlagen der digitalen Bildverarbeitung und Mustererkennung. - Schriftenreihe des ZBS e.V., ISSN 1432-3346, Ilmenau.
- [8] Dekker R J (2003). Texture analysis and classification of ERS SAR images for map updating of urban areas in The Netherlands. - IEEE Transactions on Geoscience and Remote Sensing, 41(9):1950 – 1958.
- [9] Dell' Acqua F., Gamba P (2003). Texture-based characterization of urban environments on satellite SAR images. - IEEE Transactions on Geoscience and Remote Sensing, 41(1): 153 – 159.
- [10] Nyoungui A D, Tonye E, Akono A (2002). Evaluation of speckle filtering and texture analysis methods for land cover classification from SAR images. - International Journal of Remote Sensing, 23(9): 1895 – 1925.
- [11] Franke, K.-H. (2004). Vorlesungsskript Farbbildverarbeitung. - Schriftenreihe des ZBS e.V., ISSN 1432-3346, Ilmenau.

### Authors:

Tanja Riedel, Christian Thiel, Christiane Schmuilius  
Friedrich-Schiller-University, Department of Geography, Earth Observation, Loebdergraben 32  
D-07737 Jena  
Phone: +49 – 3641 – 948875 / Fax: +49 – 3641 – 948882  
E-mail: tanja.riedel@uni-jena.de

B. Waske / V. Heinzel / M. Braun / G. Menz

## **Classification of Segmented SAR and Multispectral Satellite Imagery using Support Vector Machines**

### **INTRODUCTION**

The detailed knowledge of land cover is an important input parameter in several environmental monitoring and decision support systems. Earth-observation (EO) systems have the potential to provide spatially distributed, near-real time information on land cover and its environmental state at various spatial scales.

Land cover classifications are one of the main applications in the field of remote sensing and in several applications different data sources are combined, e.g., synthetic aperture radar (SAR) and optical imagery. In doing so the overall classification accuracy can be significantly increased compared to the quality of a single-source application [1]-[4].

Today the number of various passive and active EO-systems is increasing (e.g. multispectral and SAR), operating at different wavelengths ranging from the visible to microwave. Consequently complementary land cover information can be acquired and multisensor approaches become even more attractive. Hence, an adequate data fusion and classification of multisource imagery is an important ongoing research topic in the field of remote sensing.

Widely used statistical approaches seem inefficient, because often such imagery cannot be modeled by an adequate multivariate statistical model. Hence non-parametric methods as for example support vector machines (SVM), artificial neural networks and decision trees seem more applicable. In several studies multisource data have been classified by *decision fusion* [1], [2], [5], which is a strategy of combining different information, after each individual source has been classified previously. In Waske and Benediktsson [2] a SVM-based decision fusion was applied to SAR and multispectral data. Instead of applying the classifier directly to a multisource data set, each image source was classified separately. The outputs of the SVM were combined by different concepts to create the final classification map. Besides different voting schemes the decision fusion was performed by another SVM, which was trained on the outputs of the two individual SVM classifiers. This strategy outperforms conventional voting methods

as well as other parametric and nonparametric classifiers as a single SVM, which was applied directly to the full multisource data set.

A conceptual approach in classifying remote sensing imagery is that of segment-based classifications, where usually adjacent pixels with similar properties are merged into segments, before the classification is applied. In doing so, additional information as segments' mean value as well as neighborhood relationships can be derived and used during the classification process. Segment-based methods are particularly interesting for agricultural areas that are characterized by typical spatial patterns of planted crops. Moreover such concepts can solve two common problems, which are often arising in pixel based classifications [6], [7]: Site-internal variations in spectral reflectance or backscatter intensity, which might be caused due to within-field heterogeneities (e.g., soil moisture, plant infections etc.), will be eliminated during image segmentation. In addition the negative impact of mixed pixels, which occurs along the boundaries of two objects, can be reduced by averaging the pixel values within the segment. In regard to SAR imagery and the data inherent speckle, image segmentation has the positive effect that the speckle is leveled out [8].

In the presented study, we evaluate the concept of fusing individual outputs of SVM [2] on multitemporal SAR and multispectral data that have been previously segmented. The fused classification is compared to single-source results and that of a single SVM, which is applied on the whole multisource data set.

## **DATA SET AND PREPROCESSING**

The test site is located near Bonn in the German state North Rhine-Westphalia. The area is primarily used for agriculture. In this study a multisensor data set was available, containing multitemporal SAR data and a multispectral Landsat 5 TM image from May 28, 2005. The SAR data set includes 9 ENVISAT ASAR and ERS-2 acquisitions with different polarizations over the period April to September, 2005. The remote sensing imagery was pre-processed following common procedures. Finally image segmentation was applied separately for the multitemporal SAR data and the Landsat image (see Figure 1). In this study a commonly available approach was used that is based on a region-growing algorithm [9] and the pixel value only was used as controlling criterion for the segmentation procedure. To omit errors due to segmentation, relatively small segments were generated (~10 pixel). The segment mean values were derived and used in the subsequent classification process.

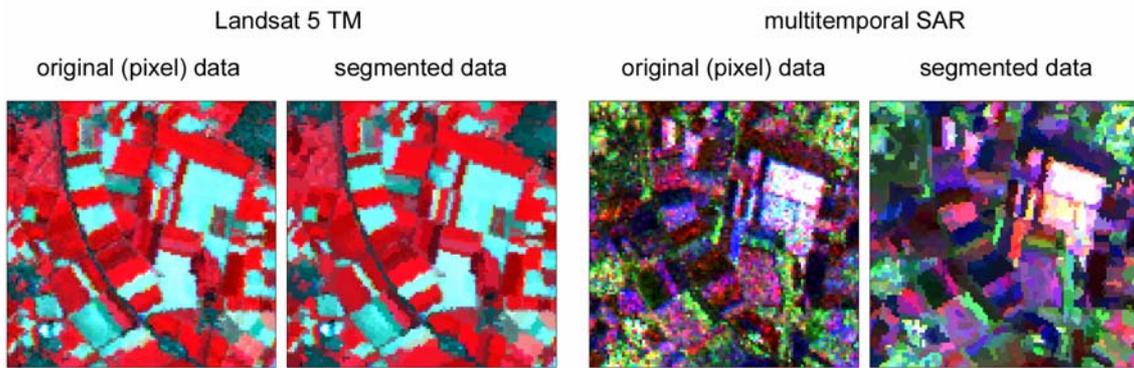


Figure 1: original and segmented Landsat 5 TM (band 4-3-2) and SAR data (21-Apr/26-May/30-Jun.)

## METHODS

### **Support Vector Machines**

SVM differentiate two classes by fitting an optimal linear separating hyperplane (OSH) to the samples in a multi-dimensional feature space. The optimization problem is based on structural risk minimization and aims to maximize the margins between the OSH and the closest samples (i.e., the support vectors) [10]. In a linearly not separable case, the input data are mapped into a higher dimensional feature space. Thus, a linear hyperplane can be fit to the newly distributed samples. In contrast to other classifiers (e.g. decision trees), which directly assign a class label to an unknown pixel, the output images of SVM contain the distance between each pixel and the hyperplane of the binary classification problem, from now on referred to as *rule images*. These rule images are used to determine the final class membership. An introduction on the general concept of SVM is given by Schölkopf and Smola [11].

SVM have originally been developed for binary classification problems, which normally do not exist in the context of remote sensing image classifications. In the literature, different approaches have been introduced to solve multi-class problems. Usually a  $n$ -class classification problem is split into several binary sub-problems and the individual binary SVM are combined into a classifier ensemble. Two main multi-class strategies exist: the one-against-one strategy (OAO) and the one-against-all strategy (OAA): For the OAA approach,  $n$  binary classifiers are trained to differentiate each individual class from the remaining classes. The maximum distance to the hyperplane determines the final class membership. In case of the OAO concept one SVM for each possible pairwise classification problem is applied, resulting in a set of  $n(n-1)/2$  binary classifiers. The final class membership is predicted by a simple majority vote, using the simple sign of the distance to the hyperplane. Even the OAO concepts results in more binary SVM classifiers, the classification problem is divided in many problems that are simpler.

### **Decision Fusion and Classification**

Eight land cover classes were considered in the classification experiment: *Arable crops*, *Cereals*, *Forest*, *Grassland*, *Orchards*, *Rapeseed*, *Root crops* and *Urban*. For each land cover class 150 samples were randomly selected from the ground truth information, by equalized random sampling. In addition an independent validation set with 500 samples per class was generated. Using the same training samples the SVM were trained on three data sets: (1) Landsat image, (2) multitemporal SAR data, and (3) a multisensor data set.

The OAO strategy was used to solve the multi-class problem, using a set of SVM classifiers with optimized parameters for  $\gamma$  and  $C$ . In this study a Gaussian kernel was used [10], which is a common kernel function in context of remote sensing. The training of the SVM and the generation of the rule images were performed using *imageSVM* [12]. *imageSVM* is a freely available IDL/ENVI implementation that is based on the LIBSVM approach by Chen and Lee [13] for the classifier training. Ideal values for  $\gamma$  and  $C$  were selected from a user defined range of possible parameters based on 10-fold cross validation. Beside the classification of the whole multisensor data set, the proposed technique is applied, fusing the individual rule images by another SVM.

Using the independent validation set, the accuracy assessment is performed on a pixel-basis, calculating the overall and class-specific accuracies.

## **RESULTS**

The experiments results clearly demonstrate the positive impact of multisensor imagery. Whereas the single-source classifiers achieve overall accuracies of 73.3% and 73.9% respectively, the application of a SVM on a multisource data set improves the results up to 79.1%. The multisource results can be further increased by the proposed fusion methods. By a separate training of the SVM on the two image source and a subsequent fusion of the outputs by another SVM the overall accuracy can be increased up to 80.9%.

Compared to pixel-based classification results (not presented in detail) the image segmentation increase the overall accuracies at least by up to 3% and even up to 11% when compared to the accuracy achieved with the SAR imagery alone.

The assessment of the class-specific accuracies (i.e., producer and user accuracies) illustrates the different nature of the image types. Some classes are better differentiated by the SAR imagery (e.g., *grassland*, *orchards*), whereas other classes are classified

more accurate by the multispectral data (e.g., *arable crops, urban*). In addition the efficiency of multisensor application is more than confirmed (Table 2).

Table 1: Overall accuracy [%]

| Data set / Method | Overall accuracy |
|-------------------|------------------|
| SAR               | 73.3             |
| TM                | 73.9             |
| SAR + TM          | 79.1             |
| SVM fusion        | 80.9             |

Table 2: Class-specific accuracy, using a single SVM and the proposed fusion strategy

| Land cover class | Producer accuracy |      |            | User accuracy |      |            |
|------------------|-------------------|------|------------|---------------|------|------------|
|                  | SAR               | TM   | SVM Fusion | SAR           | TM   | SVM Fusion |
| Arable crops     | 65.2              | 67.2 | 73.8       | 70.7          | 75.0 | 81.8       |
| Cereals          | 81.0              | 79.2 | 81.4       | 64.4          | 70.6 | 76.1       |
| Forest           | 88.0              | 92.2 | 95.0       | 82.2          | 94.6 | 95.6       |
| Grassland        | 71.6              | 61.4 | 75.0       | 72.9          | 63.9 | 73.3       |
| Orchard          | 72.4              | 58.2 | 77.4       | 68.8          | 59.6 | 73.3       |
| Rapeseed         | 69.2              | 72.0 | 82.2       | 78.1          | 79.6 | 82.7       |
| Root crops       | 71.0              | 77.0 | 78.6       | 68.3          | 64.1 | 76.6       |
| Urban            | 68.0              | 84.0 | 82.6       | 86.1          | 86.9 | 88.4       |

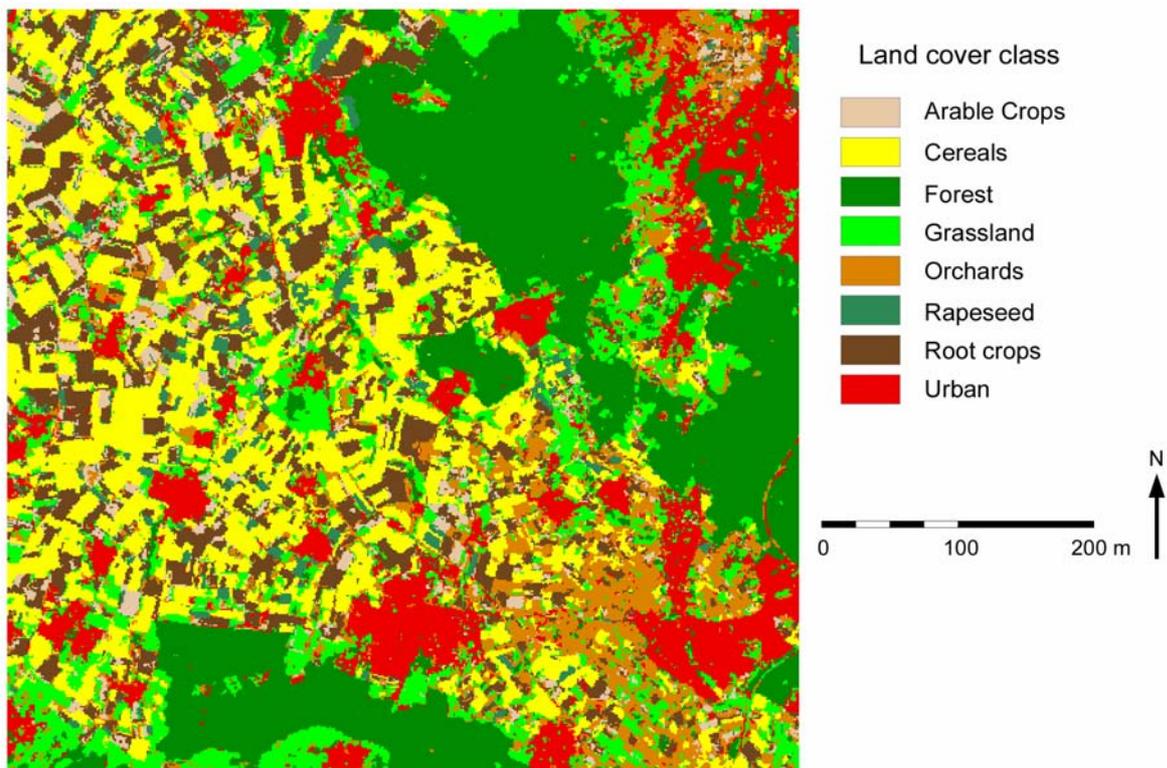


Figure 2: Classification results, using the proposed fusion strategy

## CONCLUSION

A concept of fusing multisensor imagery was evaluated. It has been shown that the concept, which was originally developed on pixel-based data, is appropriate for segmented imagery. The different nature of the data, i.e., the multispectral image and the SAR data, is one reason for the success of the proposed classifier strategy.

Both sources provide different information and may not be equally reliable. Instead of using one specific kernel function for the whole multisource data set, the definition of separate kernel functions for each data source seems more adequate.

In general the results demonstrate the positive effect of multisensor imagery, and the classification accuracy is significantly increased by such data sets. This is particularly relevant with respect to recent and upcoming EO missions like, ALOS, TerraSAR-X and Radarsat-2.

## ACKNOWLEDGMENTS

The authors wish to thank the European Space Agency for providing Envisat ASAR and ERS-2 data through a CAT 1 proposal (C1.3115). The data were acquired within the ENVILAND research project (FKZ 50EE0404), funded by the German Aerospace Center (DLR) and the Federal Ministry of Economics and Technology (BMW).

## References:

- [1] J.A. Benediktsson and I. Kanellopoulos (1999): Classification of Multisource and Hyperspectral Data Based on Decision Fusion. *IEEE Trans. Geosci. and Remote Sens.*, vol. 37, pp. 1367-1377.
- [2] B. Waske and J. A. Benediktsson (2007): Fusion of Support Vector Machines for Classification of Multisensor Data. *IEEE Trans. Geosci. and Remote Sens.*, in press.
- [3] X. Blaes, L. Vanhalle, and P. Defourny (2005): Efficiency of crop identification based on optical and SAR image time series. *Remote Sens. of Env.*, vol. 96, pp. 352-365.
- [4] G. Chust, D. Ducrot, and J.L. Pretus (2004): Land cover discrimination potential of radar multitemporal series and optical multispectral images in a Mediterranean cultural landscape. *Int. J. of Remote Sens.*, vol. 25, pp. 3513-3528.
- [5] J.A. Benediktsson, J. R. Sveinsson, and P. H. Swain (1997): Hybrid Consensus Theoretic Classification. *IEEE Trans. Geosci. and Remote Sens.*, vol. 35, pp. 833-843.
- [6] G.M. Smith and R.M. Fuller (2001): An integrated approach to land cover classification: an example in the Island of Jersey. *Int. J. of Rem. Sens.*, vol. 22, pp. 3123-3142.
- [7] J.W. De Wit and J. Clevers (2004): Efficiency and accuracy of per-filed classification for operational crop mapping. *Int. J. of Rem. Sens.*, vol. 20, pp. 4091-4112.
- [8] B. Tso and R.M. Mather (1999): Crop discrimination using multi-temporal SAR imagery. *Int. J. of Remote. Sens.*, vol. 20, pp. 2443-2460.
- [9] M. Baatz and A. Schäpe (2000): *Multiresolution Segmentation - an optimization approach for high quality multi-scale segmentation*. In: J. Strobel et al. (Eds.): *Angewandte Geogr. Informationsverarbeitung*.
- [10] V.N. Vapnik (1998): *Statistical Learning Theory*, New York: Wiley.
- [11] B. Schölkopf and A. Smola (2002): *Learning with Kernels*, MIT Press, Cambridge, MA.
- [12] Janz, S. Schiefer, B. Waske, and P. Hostert (2007): A user-oriented tool for advanced classification of hyperspectral data using support vector machines. *5th Works. EARSeL Special Interest Group on Imaging Spectroscopy*, Bruges, Belgium.
- [13] C.-C. Chen, and C.-J. Lin (2001): *LIBSVM: a library for support vector machines*. Software available at: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

## Authors:

Björn Waske, V. Heinzl, M. Braun, and G. Menz  
Center for Remote Sensing of Land Surfaces (ZFL)  
University of Bonn, Walter-Flex-Str. 3, 53113, Bonn  
Phone: +49-228-734941  
Fax: +49-228-736857  
E-mail: [bwask@uni-bonn.de](mailto:bwask@uni-bonn.de)

V. Heinzel / J. Franke / G. Menz

# Assessment of differences in multi-sensoral remote sensing imageries caused by discrepancies in the relative spectral response functions

## INTRODUCTION

Spectral vegetation indices derived from satellite observations in the near infrared and visible wavelengths, are widely used within the remote sensing community. Most commonly applied for analysing temporal and spatial vegetation dynamics is the Normalized Difference Vegetation Index (NDVI) [1], defined as:

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \quad (1)$$

where RED and NIR denote the spectral reflectance measurements acquired in the red and near-infrared spectrum. Vital green plants absorb solar radiation in the photosynthetically active radiation (PAR) spectral region, which is their source of energy for the photosynthesis process. On the other hand leaf cells scatter (e.g., reflect and transmit) solar radiation in the near-infrared spectral region. The energy level per photon in that domain would result in over-heating the plant and possibly damage the tissues when absorbed. Hence, vital green plants exhibit rather high NDVI values, while diseased vegetation or non-vegetated areas feature rather low or even negative NDVI values (e.g., water).

For multi-temporal vegetation monitoring or change analysis, a combination of multi-sensoral NDVI is often necessary. However, due to different sensor characteristics (e.g., sensor geometry, spatial or radiometric resolution and relative spectral response functions (RSR)) the NDVI can vary. Within this study the focus will be laid on the relative spectral response functions. Whereby, signature variations are introduced because the sensors receive slightly different components of the reflectance spectra of the illuminated target. Several studies have analysed the offset in data products caused by these spectral

characteristics and introduced approaches to minimize those variations [2]-[6].

For the analysis multispectral bands of the satellites Landsat 5 TM, SPOT 5, Aster and QuickBird were simulated by the use of hyperspectral bands from the airborne HyMap sensor. Variations to original satellite data caused by different spatial resolution or other effects are not considered by the sensor simulation and will not be taken into account for the intercalibration process.

After generating each simulated image the NDVI was calculated and the resulting NDVI-varieties were analysed. An empirical cross-calibration method was finally chosen for the intercalibration process.

## **METHODS AND DATA**

### **Data for sensor simulation**

The flight campaign with the airborne Hyperspectral Mapper (HyMap) took place on 28<sup>th</sup> May 2005 (12.00). HyMap acquires data in 128 bands, with a bandwidth of 15nm in the VIS and NIR region by a geometric resolution of 4m. HyVista corp. and the DLR (German Aerospace Center) carried out the orthorectification and atmospheric correction of the data.

### **Sensor characteristics of the simulated satellites**

Spectral bands are characterized by their spectral range, bandwidth, center wavelength and full width at half maximum (FWHM). The relative spectral response function takes all these features into account and is defined by the effective spectral quantum efficiency (QE) of the detector, including features like the type-dependent sensitivity of the CCD, losses due to light reflecting and transmitting components of the detector (e.g. optics, mirrors, filters, coatings etc.) [7].

Figure1 illustrates the RSR functions of the different used satellite sensors (Landsat 5 TM, SPOT 5, Aster, QuickBird). The curves differ in shape, central wavelength location and the degree of overlap between the bands. Especially in the region of the red edge (red-NIR translation) (680~800 nm) the sensors vary from each other.

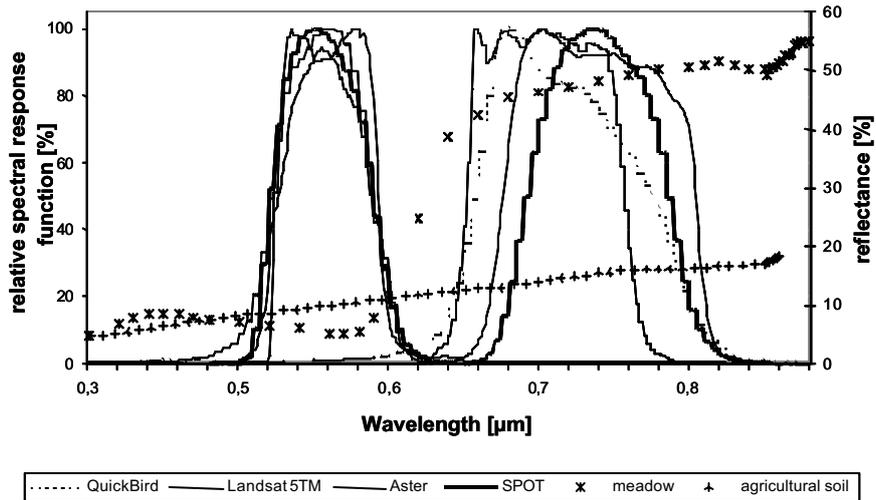


Figure 1. Variations in the red and near-infrared RSR functions among the used earth observing satellites and the spectral profile of two different agricultural targets as reference.

## SENSOR SIMULATION

For the differences assessment within the imageries caused by variable RSR-functions the four satellite sensors were simulated using the hyperspectral image. In the first step each HyMap center wavelength was assigned to the mean RSR-value (in the range of FWHM of the hyperspectral band) of the simulating band. In a second step the hyperspectral reflectance values of each pixel were multiplied by the corresponding RSR-values of the simulating band. The sum of these products was then divided by the sum of the band-specific RSR-values. For the multispectral sensor simulation, each band was simulated according to the following equation:

$$R_{sim_b} = \frac{\sum_{i=1}^n R_i * rSR_{b,i}}{\sum_{i=1}^n rSR_{b,i}} \quad 1 \leq n \leq 126 \quad (2)$$

with  $R_{sim_b}$  as the simulated pixel reflectance value of the simulated band,  $R_i$  pixel reflectance value of the HyMap band and  $rSR_{b,i}$  relative spectral response value of the simulating band at each HyMap corresponding wavelength [7].

The validation of the simulation performance was analysed on the basis of a pre-processed Landsat 5 TM scene (28.05.05; 10.30 am) and the simulated Landsat 5 TM image based on an airborne HyMap scene (28.05.05; 12.00 am) [7].

The calculated reflectance and NDVI differences are very marginal with an absolute NDVI difference of 0.626% (Tab. 1). Regarding the real differences between original sensors, they range from 1 ~ 4% for e.g. Landsat 5 TM and Landsat 7 ETM+ [8]. With the marginal differences achieved here, this accurate simulation method is appropriate for analyzing the impact of different RSR functions on the NDVI.

Table 1. NDVI differences between the simulated and the original imagery.

|      | SIMULATED |       | ORIGINAL |       | DIFFERENCE |            |
|------|-----------|-------|----------|-------|------------|------------|
|      | MEAN      | STDEV | MEAN     | STDEV | absolute   | in percent |
| NDVI | 0.635     | 0.277 | 0.639    | 0.258 | -0.004     | -0.626     |

### NDVI-INTERCALIBRATION

In Table 2 variations between the NDVI of the four simulated sensors are displayed. Differences between SPOT5 and the other sensors feature the widest divergences. Another obvious feature is that the differences between Aster and QuickBird are the smallest both having similar RSR curves.

The general relationships between all sensors can be described as followed. SPOT5 features the highest NDVI values, then Landsat 5 TM, Aster, and the lowest NDVI values exhibits QuickBird.

Table 2. MIN, MAX and MEAN NDVI differences and % difference between the simulated sensors.

|                       | Min NDVI differences | Max NDVI differences | Mean NDVI differences | Differences (%) |
|-----------------------|----------------------|----------------------|-----------------------|-----------------|
| SPOT5-Aster           | -0.116               | 0.104                | 0.012                 | 1.818           |
| SPOT5-Landsat 5TM     | -0.157               | 0.126                | 0.009                 | 1.364           |
| SPOT5-QuickBird       | -0.064               | 0.099                | 0.045                 | 7.087           |
| Landsat 5TM-Aster     | -0.088               | 0.091                | 0.003                 | 0.472           |
| Landsat 5TM-QuickBird | -0.074               | 0.083                | 0.013                 | 1.970           |
| QuickBird-Aster       | -0.058               | 0.040                | -0.002                | -0.322          |

In general a similar result was found by [4] when trying to model the NDVI inter-sensor relationship. They proposed to model the relationship with a higher polynomial order. The regression coefficients [ $R^2$ ] for polynomials of different order vary between  $R^2= 0.73$  and  $R^2= 0.98$  depending on the sensor pairs compared and the order of polynomial. The best correlation results were found for a polynomial of the sixth order.

## VALIDATION OF THE NDVI-INTERCALIBRATION

In Table 3 the differences between the sensors after the cross-calibration are displayed. The results for the intercalibration feature a great enhancement in regard to the non-calibrated sensors (Tab. 2). Thus the differences between SPOT5 and QuickBird, which were the biggest difference before intercalibration, decreased from 7.09 % to  $-0.15$  % or  $0.16$  %, depending on which sensor was taken as the target. Overall, the best results with the smallest bias errors were achieved for translating SPOT5 into Aster and QuickBird. In general the magnitudes of error from translating the sensors into each other lie between  $-0.15$  and  $0.61$  %, which are good result when comparing it with the results from [3], who reached a precision of 1-2% or [4] with an accuracy of  $\sim 2\%$ .

When comparing the results of the sixth order intercalibration with the ones of a second order approach [9] it becomes obvious that the sixth order is able to model the differences in a more accurate way. For the second order modeling the NDVI difference after the intercalibration were in the range of  $-0.91$  to  $0.80$  %, being now significantly smaller.

Table 3. MIN, MAX, MEAN differences and % difference between the original NDVI imagery and the cross-calibrated image. Differences were taken between the original sensor and the cross-calibrated sensor, which is still named after its origin.

| original minus cross-calibrated NDVI image | Min NDVI differences | Max NDVI differences | Mean NDVI differences | Mean NDVI differences (%) |
|--|----------------------|----------------------|-----------------------|---------------------------|
| SPOT5-Aster                                | -0.445               | 0.068                | -0.001                | -0.152                    |
| SPOT5-Landsat 5TM                          | -0.797               | 0.106                | 0.004                 | 0.606                     |
| SPOT5-QuickBird                            | -0.3307              | 0.052                | -0.001                | -0.152                    |
| Landsat 5TM-Aster                          | -0.100               | 0.964                | -0.002                | -0.315                    |
| Landsat 5TM-QuickBird                      | -0.084               | 0.910                | -0.004                | -0.630                    |
| Landsat 5TM-SPOT5                          | -0.107               | 0.954                | -0.003                | -0.472                    |
| QuickBird-Aster                            | -0.057               | 0.031                | 0.001                 | 0.161                     |
| QuickBird-Landsat 5TM                      | -0.609               | 0.0784               | 0.004                 | 0.483                     |
| QuickBird-SPOT5                            | -0.049               | 0.221                | 0.001                 | 0.161                     |
| Aster-Landsat 5TM                          | -0.675               | 0.096                | 0.002                 | 0.320                     |
| Aster-QuickBird                            | -0.030               | 0.058                | -0.001                | -0.160                    |
| Aster-SPOT5                                | -0.067               | 0.225                | -0.001                | -0.160                    |

## DISCUSSION AND CONCLUSION

The sensor simulation method using airborne hyperspectral data performed well. The residual NDVI differences between the simulated and an original Landsat 5 TM image, added up to only 0.62%. When comparing the simulation result with actually proven discrepancies between different satellite sensors [8] the differences between the simulated

image and an original image were significantly smaller than the actually discovered NDVI differences of 1% to 4%.

The analysed differences in NDVI between the source and target sensors were found to have rather complex patterns, which could be best modelled by sixth order polynomials as supposed by [4] and [9]. The chosen empirical NDVI correction method then performed well, reducing the NDVI differences by 98% for the comparison SPOT5 vs. QuickBird (best case) and by 50% for the relationship QuickBird vs. Aster (worst case). When comparing these results with a second order intercalibration, reducing the differences round 94% for SPOT5 vs. QuickBird or by 5% for QuickBird vs. Aster the sixth order method, performed significantly better [9]. Also when comparing it with the results from [4], who reduced the differences by 80% and 65%, respectively the higher order polynomial performed better. Generally, the results indicate that the NDVI intercalibration is a reasonable first step of a processing chain for multi-sensoral satellite data to ensure the comparability of achieved results.

#### **Acknowledgment**

The authors would like to thank Dr. M. Braun from the ZFL for organizing the HyMap flight campaign in Bonn and A. Moll (ZFL) for helping with the IDL programming of the sensor simulation program. The study was realized in the framework of the project ENVILAND (FKZ 50EE0404) funded by the German Aerospace Centre (DLR) and the DFG research training group 722.

#### **References:**

- [1] Rouse, J.W., Haas, R.H., Schell, J.A. and Deering, D.W. (1973), Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation, Prog. Rep. RSC 1978-1, Remote Sensing Center, Texas A&M Univ..
- [2] Trishchenko, A. P., Cihlar, J. and Li, Z. (2002), Effects of spectral response function on surface reflectance and NDVI measured with moderate resolution satellite sensors. *Remote Sensing of Environment*, 81, 1-18.
- [3] Steven, M.D., Malthus, T.J., Baret, F., Xu, H. and Chopping, M. J. (2003), Intercalibration of vegetation indices from different sensor systems. *Remote Sensing of Environment*, 88, 412-422.
- [4] Miura, T., Huete, A. and Yoshioka, H. (2006), An empirical investigation of cross-sensor relationships of NDVI and red/near-infrared reflectance using EO-1 Hyperion data. *Remote Sensing of Environment*, 100, 223-236.
- [5] van Leeuwen, W.J.D., Orr, B.J., Marsh, S.E. and Herrmann, S.M. (2006), Multi-sensor NDVI data continuity: Uncertainties and implications for vegetation monitoring applications. *Remote Sensing of Environment*, 100, 67-81.
- [6] Franke, J. and Menz, G. (2004), Sensor intercalibration- adjustment of MODISNDVI to AVHRR NDVI data. *International Geoscience and Remote Sensing Symposia*.
- [7] Franke, J., Heinzl, V. and Menz, G. (2006), Assessment of NDVI-differences caused by sensor-specific relative spectral response functions. *International Geoscience and Remote Sensing Symposia*.
- [8] Teillet, P.M., Barker, J.L., Markham, B.L., Irish, R.R., Fedosejevs, G. and Storey, J.C. (2001), Radiometric cross-calibration of the landsat-7 ETM+ and landsat-5 TM sensors based on tandem data sets. *Remote Sensing of Environment*, 78, 39-54.
- [9] Heinzl, V., Franke, J. and Menz, G. (2006), Assessment of cross-sensor NDVI-variations caused by spectral band characteristics. *Proceedings SPIE Vol. 5980, SPIE Remote Sensing*.

#### **Authors:**

Dipl. Geogr. Vanessa Heinzl

Dipl. Geogr. Jonas Franke

Prof. Dr. Gunter Menz

ZFL- Center for Remote Sensing of Land Surfaces; Walter-Flex-Str. 3

53113, Bonn, Germany

Phone: +49228-734941

Fax: +49228-736857

E-mail: vheinzel@uni-bonn.de

I. Aksit, K. Büngrer, A. Fassbender, D. Frekers, C. Götze, J. Kemenas,

## **An ultra-fast on-line microscopic optical quality assurance concept for small structures in an environment of mass production**

Today's Printed Circuit Boards (PCB) for sensors meet structure sizes in the  $\mu\text{m}$  or even nm range at high circuit complexity. These structures are usually printed in thin layer technology and coated in thin film technology. Those PCBs are being used in a large variety of applications ranging from cars to space stations, from private security devices to power plant control facilities, from toys and domestic appliances to consumer electronics, from robots to information devices. With increasing complexity, the demands on quality assurance and quality evaluation are quickly increasing. PCBs usually have to be investigated for surface and coating defects by optical techniques. So far, it has been impossible or prohibitively expensive to implement automated quality assurance procedures for each individual PCB by optical measures and at high throughput. Quality assessment is therefore mostly limited to statistical sampling techniques employing specially trained personnel performing the evaluations. This means that performance is frequently subjective and non-reproducible, let aside the often stress-prone conditions facing the personnel. Past attempts to automate optical inspections for each individual PCB and allowing on-line keep-or-reject decisions have largely been unsuccessful because the structures to be inspected are microscopically small, therefore requiring high magnifications and advanced image resolutions, which usually prevents large area inspections in one working step. Clearly, a novel approach is needed in the area of hardware and software, both of them being capable of acquiring and accurately analysing these structures at high resolution and within a time slot that allows on-line implementation in a mass production.

In the present pilot project, the authors have demonstrated a successful implementation of a system providing these features. Each partner company contributed special core competencies and integrated those into a single product. In a first step, the Eclipse optical microscope of the **MedXP GmbH** equipped with the patented AMBIS™

technique acquires a series of images from the entire area of test samples, each of them being about 10mm in diameter. Typical sustained acquisition speeds are about 25 – 30 frames/s at full image resolution in continuous scanning mode. The resulting image sizes depend on the optical system used, and in the present application perfect results were achieved with images of about 18.000x14.000 pixels per sample. In this case, the overall image of a single sample is separated into 14x14 partial images, each of 1280x1024 pixels in size and joined together without the need of a seaming algorithm for any of the edges. The samples can optionally be supplied to the microscope in a magazine of about 100 pieces by an automatic robot system. For acquiring the image of a single sample of the typical size of 10mm in diameter, the microscope needs about 10s, which includes the overhead for the focusing to the object surface. As there are no start-stop sequences involved during image acquisition, the microscope system operates without generating any extra noise and without creating any inertial forces due to acceleration or deceleration or any other frictional forces.

Together with position information for each single image the acquired images are then handed over to the *arivis ImageCore* software developed and distributed by **arivis – Multiple Image Tools GmbH**. This software package is capable of handling extremely large, multi-dimensional image data even on common, budget-priced hardware and manages these image areas for visual inspection at any zoom level or for an automated analysis. The software combines all individual image parts to a single large image for each sample and stores the image for subsequent ultra-fast access into a single file using proprietary formatting software.

Next, the image data are transferred to the *alfaVis* analysis software developed and distributed by **alfa vision systems GmbH**. This software package is developed for the detection of various types of structural defects on the original sample, like missing circuit structures, short circuits or any other surface defects or surface pollution. The analysis provides a quality certificate for every single sample in the line of the production.

The integrated pilot system needs about 20 minutes for the inspection of a magazine containing about 100 samples. Every sample is treated in an identical way making the evaluation independent of time, human subjectivity or human error. The

system is flexible and can be adopted quickly and easily to a change of sample layout requiring only minutes for the adjustment to the new layout and to the optimization process.

**Authors:**

Dipl. Phys. Ishak Aksit,  
MedXP GmbH, Munscheidstraße 14,  
45886 Gelsenkirchen  
Phone: 0251 83 36216  
Fax: 0251 83 34962  
E-mail: [aksit@medxp.de](mailto:aksit@medxp.de)

Dr. Kirsten Bünger,  
MedXP GmbH, Munscheidstraße 14,  
45886 Gelsenkirchen  
Phone: 0209 167 1050  
Fax: 0209 167 1051  
E-mail: [buenger@medxp.de](mailto:buenger@medxp.de)

Alfred Fassbender,  
alfa vision systems GmbH,  
Altenlinde 51a,  
51789 Lindlar  
Phone: 02266 901206  
Fax: 02266 901331  
E-mail: [fassbender@alfavisionsystems.com](mailto:fassbender@alfavisionsystems.com)

Prof. Dr. Dieter Frekers,  
MedXP GmbH, Munscheidstraße 14,  
45886 Gelsenkirchen  
Phone: 0251 83 34996  
Fax: 0251 83 34962  
E-mail: [frekers@medxp.de](mailto:frekers@medxp.de), [frekers@uni-muenster.de](mailto:frekers@uni-muenster.de)

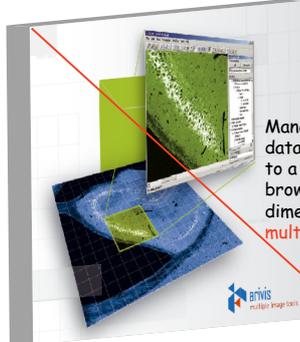
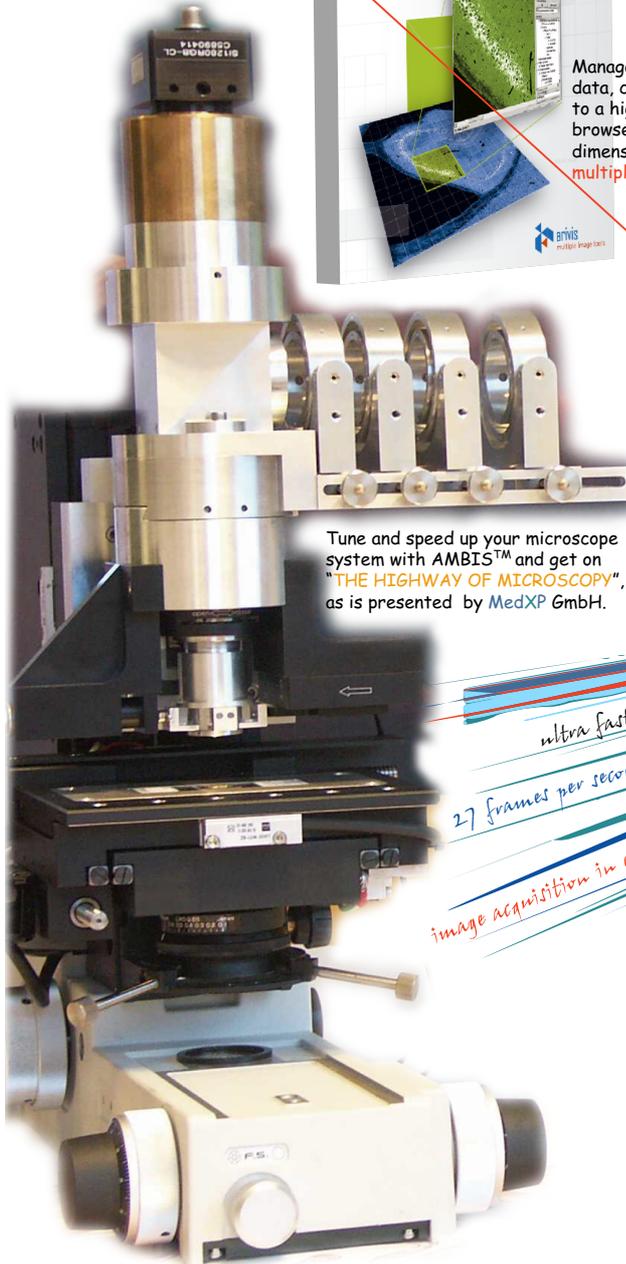
Dr. Christian Götze,  
arivis – Multiple Image Tools GmbH,  
Schwaansche Straße 1  
18055 Rostock  
Phone: 0381 461393-0  
Fax: 0381 461393-99  
E-mail: [christian.goetze@arivis.com](mailto:christian.goetze@arivis.com)

Jürgen Kemenas,  
SMD-Production-Technology,  
Inrather Straße 11-15  
47798 Krefeld  
Phone: 02151 5696573  
Fax: 02151 5696575  
E-mail: [jk@smd-pt.de](mailto:jk@smd-pt.de)

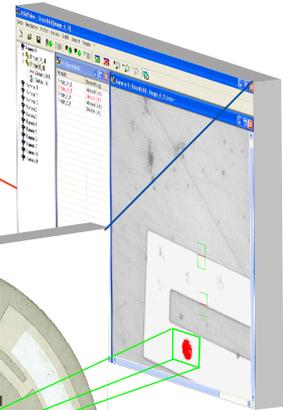


# 52. IWK

Internationales Wissenschaftliches Kolloquium  
International Scientific Colloquium



Manage multi dimensional image data, combine the digitised images to a high resolution single image file, browse and navigate in unexpected dimensions offered by **arivis multiple image tools GmbH**.



Make the quality management with **alfaVis analysis software** by **alfa vision systems GmbH**, the experts in machine vision.

Tune and speed up your microscope system with **AMBIS™** and get on **"THE HIGHWAY OF MICROSCOPY"**, as is presented by **MedXP GmbH**.

*ultra fast high-throughput scanning with AMBIS™*  
*27 frames per second at full resolution*  
*image acquisition in continuous motion at velocity of 29 mm/s*

3 in 1

Three companies  
One project

**MEDXP** GmbH has a patented Anti-Motion-Blurring-Imaging System, called **AMBIS™**. The PCT, patented technology makes the microscope ready for ultra-fast, high-throughput scanning in continuous motion, i.e. no start-stop sequences are needed for image acquisition. Within **13 seconds** an area of **10mm x 10mm** will be completely scanned and stored on disk. Objects will be digitised at **27 images per second** and a velocity of **29 mm/s**. Each image has **(1280 H x 1024 V) pixels** and represents **(850 x 680)µm²** of the object. The higher the field of view the higher the scanning speed.



D. Hofmann / G. Linss

# Application of Innovative Image Sensors for Quality Control

## 1. INTRODUCTION

Modern enabling technologies are working ever faster and in ever smaller scales. Main tasks of objective visual quality control with digital imaging are dimensional and color measurements [1]. In earlier days visual quality assurance with analog imaging have been expensive and the working area of specialists. Innovative vision sensors are more convenient, reliable and affordable (Table 1-1). Their bulk production and application is forthcoming.

Table 1-1 Vision Sensors from Cognex USA, Wenglor Germany and OMRON Japan

|  |   |   |
|--|---|---|
|  <p>A yellow rectangular vision sensor with a lens and a red sensor array visible through a transparent window. The text 'COGNEX DVT 535 VISION SENSOR' is printed on the front.</p> |  <p>A blue, cube-shaped vision sensor with a lens on the front face.</p> |  <p>A silver vision sensor system consisting of a camera unit and a separate processing unit with a color display showing a grid of colored dots.</p> |
| <p>DVT 535 [2]</p>   | <p>BS40V101 [3]</p>   | <p>ZFV Color [4]</p>  |

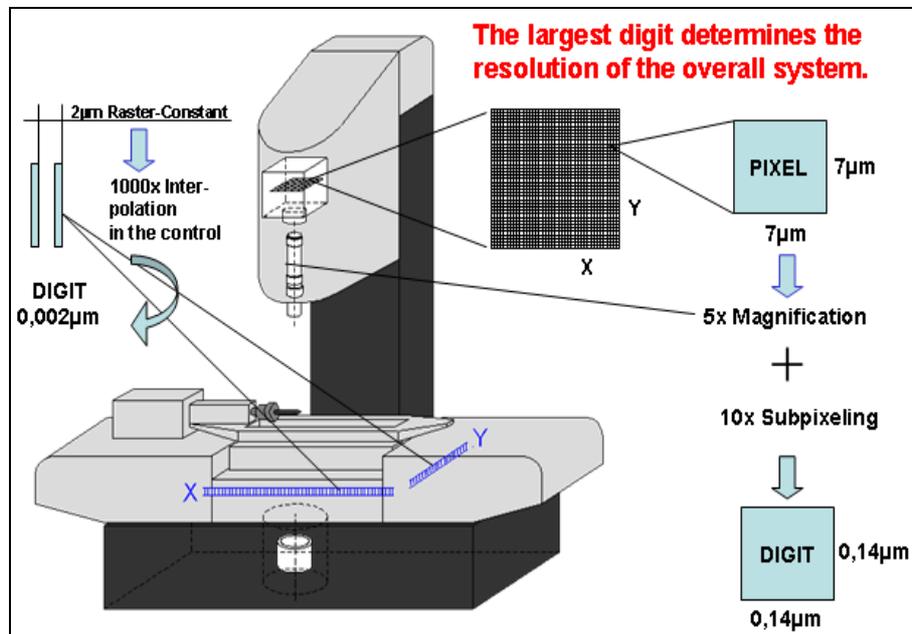
The future of visual quality assurance with imaging is digital, mobile, colored, miniaturized, standardized, networked and convergent. All these features support its fast growth in practical applications. But operation errors are risky. Aim of the paper is to show typical errors of visual quality measurements with digital imaging in micro scale and to give recommendations how to prevent them.

## 2. THE MEASUREMENT CHAIN WITH IMAGE SENSORS

Objective measurement is the comparison of a measurement object with a metrological standard. Reliable measurements are simple, if appropriate metrological standards, measurement methods, measurement instruments and measurement software are available and a qualified measurement person is preparing, performing and evaluating the measurement.

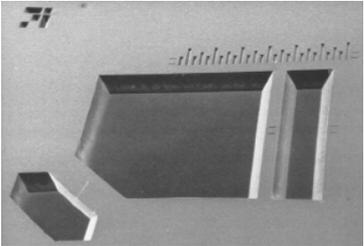
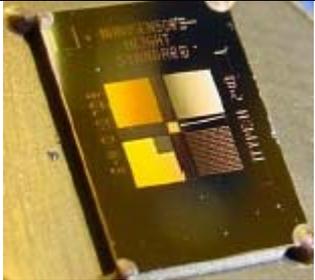
The accuracy, reproducibility and traceability of optical length measurements are dependent on a number of influencing parameters in the measurement chain (Table 2-1) [5]

Table 2-1 Measurement chain for quality assurance with digital imaging



### 3. INFLUENCES OF METROLOGICAL STANDARDS ON IMAGING

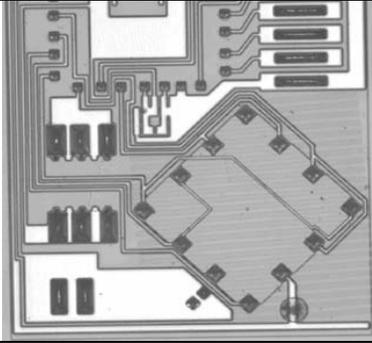
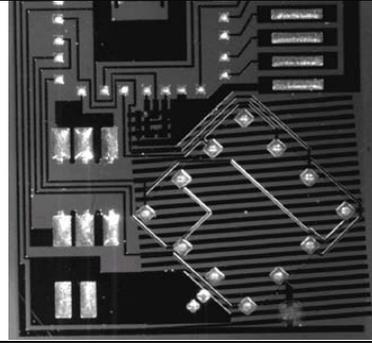
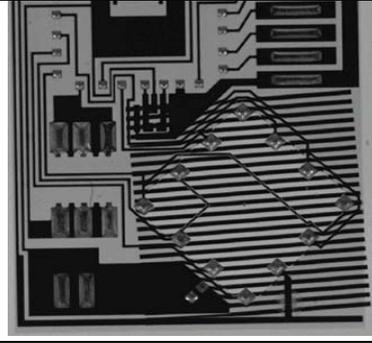
Metrological standards for the metre have a long history. They started in 1799 with metallic rules. Nowadays they are realized with modified laser interferometers [6]. The application of test standards for dimensional measurements in the micrometer and nanometre scale are still under investigation (Table 3-1 thru Table 3-3) [7].

| Table 3-1 Single Step Depth Setting Standard  | Table 3-2 Periodic Step Setting Standards   | Table 3-3 Step Grating Standards  |
|---|---|---|
|  |  |  |
| 50 nanometre ...<br>1000 micrometre   | 20 nanometre ...<br>1500 nanometre  | 8, 24, 80, 240, 800,<br>2 400 nanometre   |

#### 4. INFLUENCES OF LIGHTING ON IMAGING

In optical measurements the image of the object and not the object itself is measured. It is said that lighting influences up to 60 % the success of optical measurements. Practical examples are given in Table 4-1 upper line [8] and lower line [9]

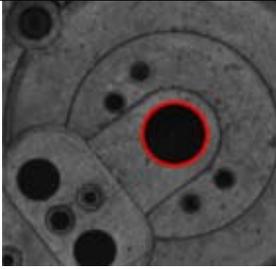
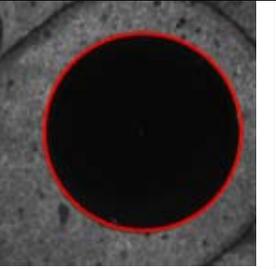
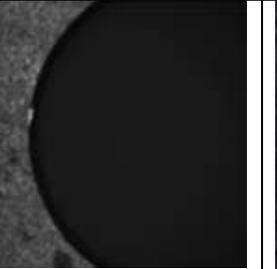
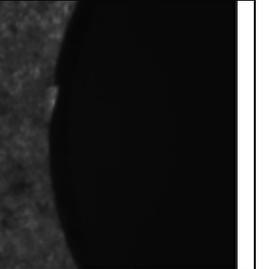
Table 4-1 Influences of lighting on the image of a measurement object

|  |  |  |  |   |
|--|--|--|--|---|
|   |   |     |  |  |
| normal<br>day light  | coaxial<br>light   | low angle<br>ring light  | low angle<br>LED array   | one sided<br>line light   |
|  |  |  |  |   |
| incident light<br>brightfield  | incident light<br>darkfield  | transfer light<br>brightfield  |  |   |

#### 5. INFLUENCES OF MAGNIFICATION ON IMAGING

Strong influence on the accuracy of vision sensors has the magnification of the objective (Table 5-1) [10]

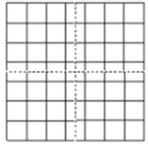
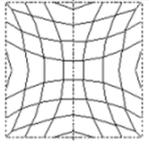
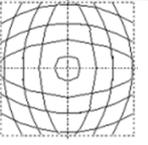
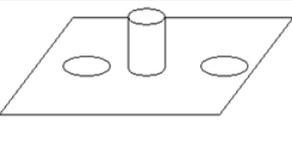
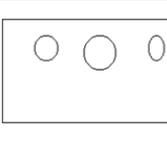
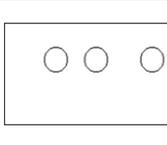
Table 5-1 GO and NO-GO magnifications for the optical measurement of a circle

|   |   |  |   |
|---|---|--|---|
|  |  |  |  |
| objective 1,0x<br>poor  | objective 3,0x<br>good  | objective 5,0x<br>poor   | objective 10,0x<br>very poor  |

## 6. INFLUENCES OF LENSES ON IMAGING

Influences on the accuracy of quality measurements with vision sensors have the optical quality of the lenses and the construction of the objectives (Table 6-1) [11]

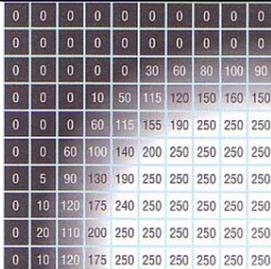
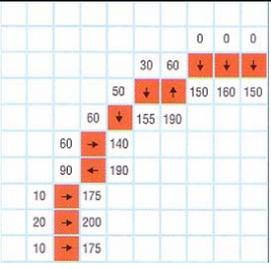
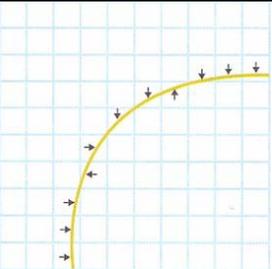
Table 6-1 Image errors caused by lens errors and construction of the objective

|   |   |   |   |   |   |
|---|---|---|---|---|---|
|  |  |  |  |  |  |
| original image  | cushion distortion  | barrel distortion   | measurement object  | entocentric objective   | telecentric objective   |

## 7. INFLUENCES OF SENSORS AND SOFTWARE ON IMAGING

Optical digital precision measurements are knowing no sharp edges. The reasons are optical interferences at edges and pixeling of the sensors. Fundamental influences on the final results have the used algorithms for data processing with the embedded computers (Table 6-1) [12].

Table 7-1 Examples for interference and pixeling

| measurement object  | measurement values  | measurement processing   | measurement result  |
|---|---|--|---|
|  |  |  |  |
| image of a hole   | pixeling  | subpixeling  | edge  |

For real-time education and self-learning comprehensive e-papers and e-lectures are provided for example by [1] and [13].

### References:

- [1] [www.visquanet.de](http://www.visquanet.de)
- [2] [www.cognex.com](http://www.cognex.com)
- [3] [www.wenglor.de](http://www.wenglor.de)
- [4] [www.omron-industrial.com](http://www.omron-industrial.com)
- [5] [www.visquanet.de](http://www.visquanet.de) > Vorlesungen visuelle QS > V002 > page 22
- [6] [www.bipm.org/en/si/history-si/evolution\\_metre.html](http://www.bipm.org/en/si/history-si/evolution_metre.html) 070114
- [7] [www.nanoscale.de/standards.htm](http://www.nanoscale.de/standards.htm) 070114.
- [8] [www.visquanet.de](http://www.visquanet.de) > Vorlesungen visuelle QS > V004 > page 24
- [9] [www.visquanet.de](http://www.visquanet.de) > Vorlesungen visuelle QS > V007 > page 23
- [10] [www.visquanet.de](http://www.visquanet.de) > Vorlesungen visuelle QS > V002 > page 18
- [11] [www.visquanet.de](http://www.visquanet.de) > Vorlesungen visuelle QS > V004 > pages 13, 20

[12] [www.visquanet.de](http://www.visquanet.de) > Vorlesungen visuelle QS > V004 > pages 34, 35, 36

[13] [www.keyence.de](http://www.keyence.de) (German) and [www.keyence.com](http://www.keyence.com) (English)

**Authors:**

Prof. Dr. Dietrich Hofmann, THÜRINGEN innovativ GmbH, Mainzerhofstr. 10, D-99084 Erfurt,  
Phone: +49 3641 448735, Fax: +49 3641 829311, E-mail: [dietrich-hofmann@t-online.de](mailto:dietrich-hofmann@t-online.de)

Prof. Dr. Gerhard Linß, Technische Universität Ilmenau, Gustav-Kirchhoff-Platz 2, D-98693 Ilmenau  
Phone: +49 3677 693822, Fax: +49 3677 693823, E-mail: [gerhard.linss@tu-ilmenau.de](mailto:gerhard.linss@tu-ilmenau.de)



A. Jablonski / K. Kohrt / M. Böhm

## **Automatic quality grading of raw leather hides**

### **ABSTRACT**

The quality assessment of leather hides is to a large extent a tedious manual process. Although in the past many attempts have been started, up to now no working automatic system is available on the market. This is due to the large complexity of the task and the time constraints the system has to meet to become a reasonable alternative for human inspectors. We present a prototype system, which is currently in development but already achieves promising results in the rough production environment. We describe the challenges of the automatic grading of leather hides and explain the ideas behind the efficient and robust defect detection algorithms.

### **INTRODUCTION**

Raw leather hides undergo a sophisticated sequence of processing steps: After removing remains of tissue, the hides are treated with salt and acid. Then they pass several tanning and dyeing steps. The complexity of this treatment varies depending on the quality of the raw leather: The better the quality, the fewer steps have to be passed, and the more cost-saving is the production. The quality also influences the market value of the leather, because the final usage ranges from high-value products like car seats, sofas or upper leather for shoes to small parts like leather belts.

Several attempts have been made to find a way to automatize the quality control of tanned leather, but until now no practical solution has been found. The Leather Research Association (Forschungsgemeinschaft Leder e.V.) stated that there is no quality inspection system for industrial use available on the market.

The Fraunhofer Institute for Industrial Mathematics (Fraunhofer-Institut für Techno- und Wirtschaftsmathematik) in Kaiserslautern is working in a project together with a tannery in Sweden and the Swedish Fraunhofer-Chalmers Research Centre for Industrial Mathematics in Gothenburg. The objective is to find a way to grade hides automatically in the production process with the help of optical sensors and novel image processing

methods. For this purpose, a prototype system has been integrated in the production line. The conditions at the tannery are very rough, humidity and dust are present everywhere, so all system components must be very robust or have to be enclosed in a protective case. Lighting and sensors have to be arranged carefully to deal with the low-contrast structures in the leather in the best possible way. Last but not least, additional challenges arise from the great demands on the defect detection algorithms that form the basis for the grading of the hides. The existing fast segmentation methods are simply unable to detect the relevant regions with reasonable false positives rates. The main difficulty is that many true defects are very similar to natural structures of the leather (Figure 1). Furthermore, the contrast of the defects can be weak and is not always proportional to their severity. The image processing algorithms have to deal with these complexities, and they have to do it fast, because approximately 100 megapixels of raw data have to be processed in 20 seconds.

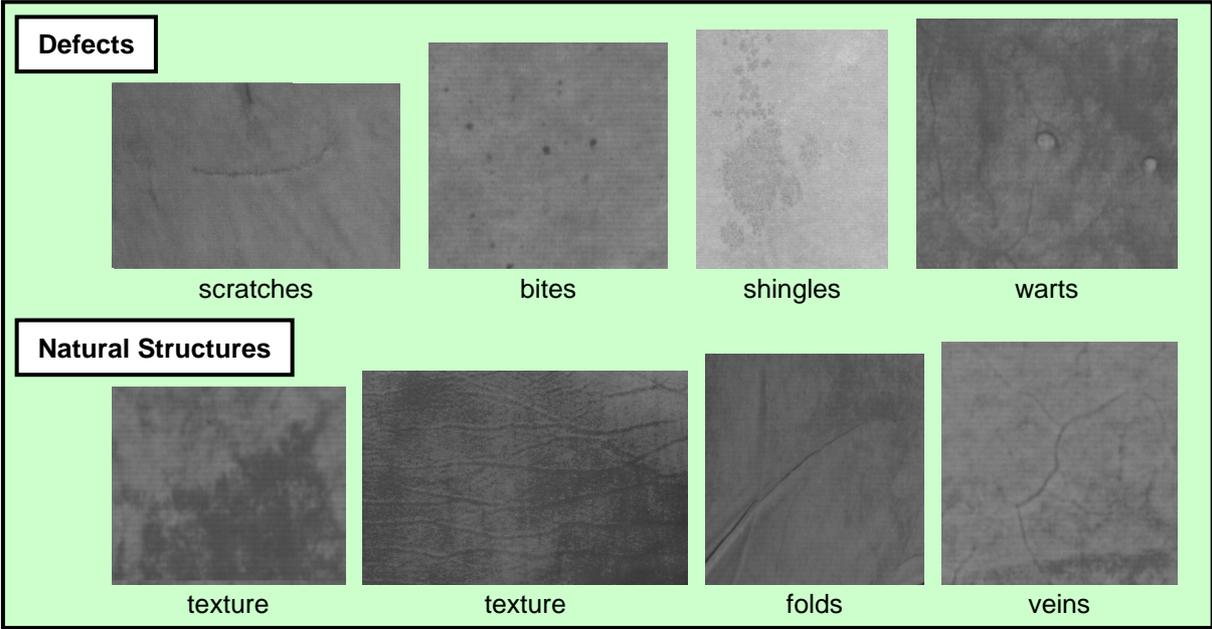


Figure 1: Unwanted (top row) and Natural (bottom row) Structures on Leather

**FROM IMAGE TO GRADING**

Starting from the acquired grayscale images, image analysis methods are applied to attain the grading result. The grading itself depends on the strength, the spatial distribution and the kind of defects present on the hide. A reliable and robust defect detection step is thus the basis for an accurate automatic grading.

The appearance of possible defects can be roughly characterized as follows: **Bites** are small and point-like dark spots, sometimes accumulated in clusters. **Scratches** appear

as line segments and can be exceedingly thin. Their contrast ranges from very good to very poor and the intensity from dark to bright. Locally, they are similar to veins. **Warts** turn out to be bright shiny spots in the images. **Shingles** are agglomerations of circular spots with a complex geometry and usually of extremely poor contrast.

Different kinds of defects require different detection algorithms. The detection of warts and bites is the simplest and can be achieved with standard image processing methods. The detection of shingles is very difficult, the current prototype is not able to detect them reliably. In this article we concentrate on the scratch detection as scratches appear most frequently and the usual standard algorithms are not sufficient to detect them.

### SCRATCH DETECTION ALGORITHM

The scratch detection algorithm consists of several processing steps (Figure 2): The preprocessing step (A) generates a binary image, the detection step (B) enhances and segments line like binary structures and the additional classification step (C) further suppresses the false positive detections.



Figure 2: Image Processing Steps

To segment line like structures (i.e. potential scratches) in a highly noisy binary image two strategies are used. The first one relies on an adaptive anisotropic smoothing (**AAS**) and works well for images with sufficient contrast. The second one is more involved and consists of several iterations of adaptive erosion/dilation and reconstruction steps (**IEDR**). Both approaches try to obtain each candidate scratch as one connected region.

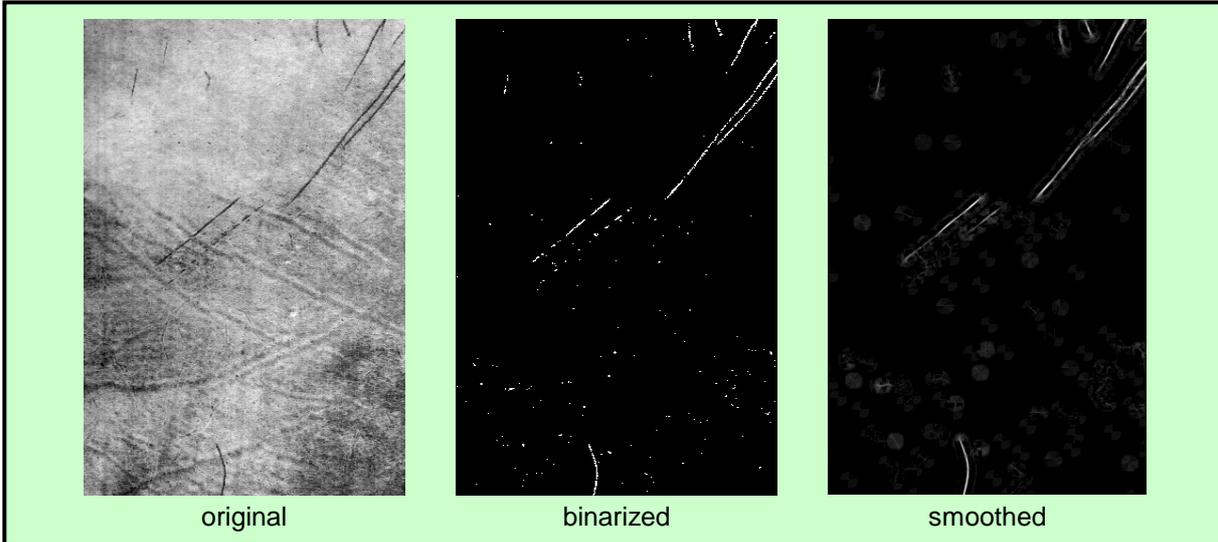
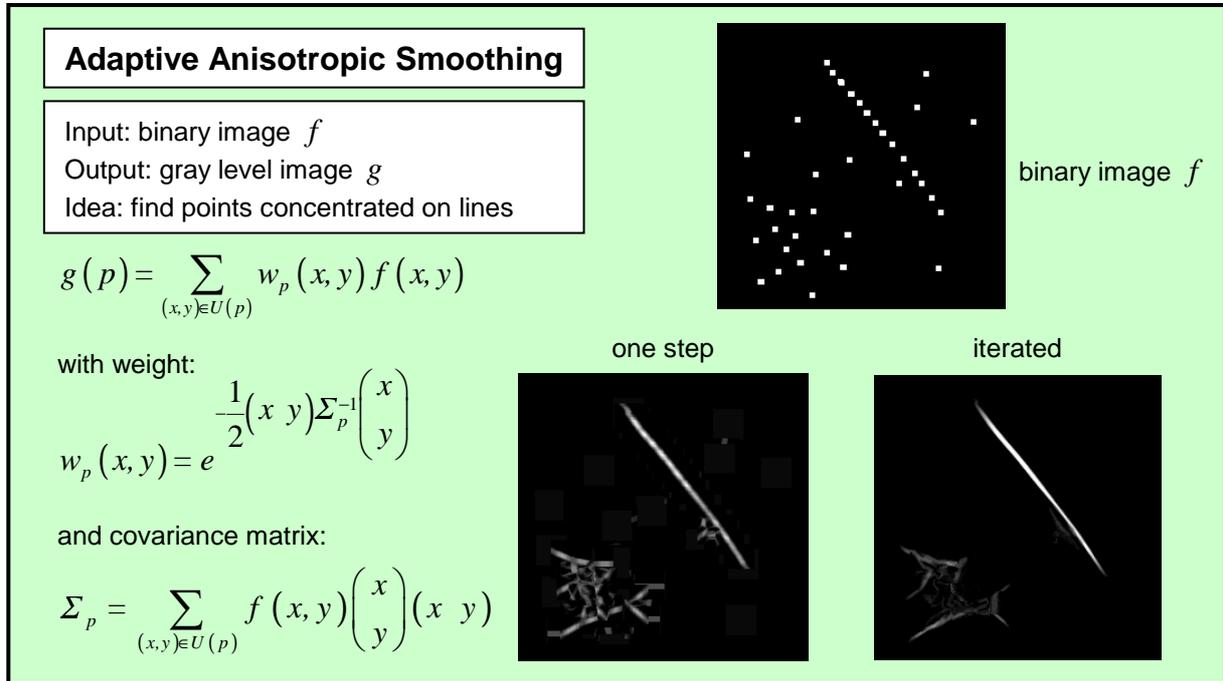


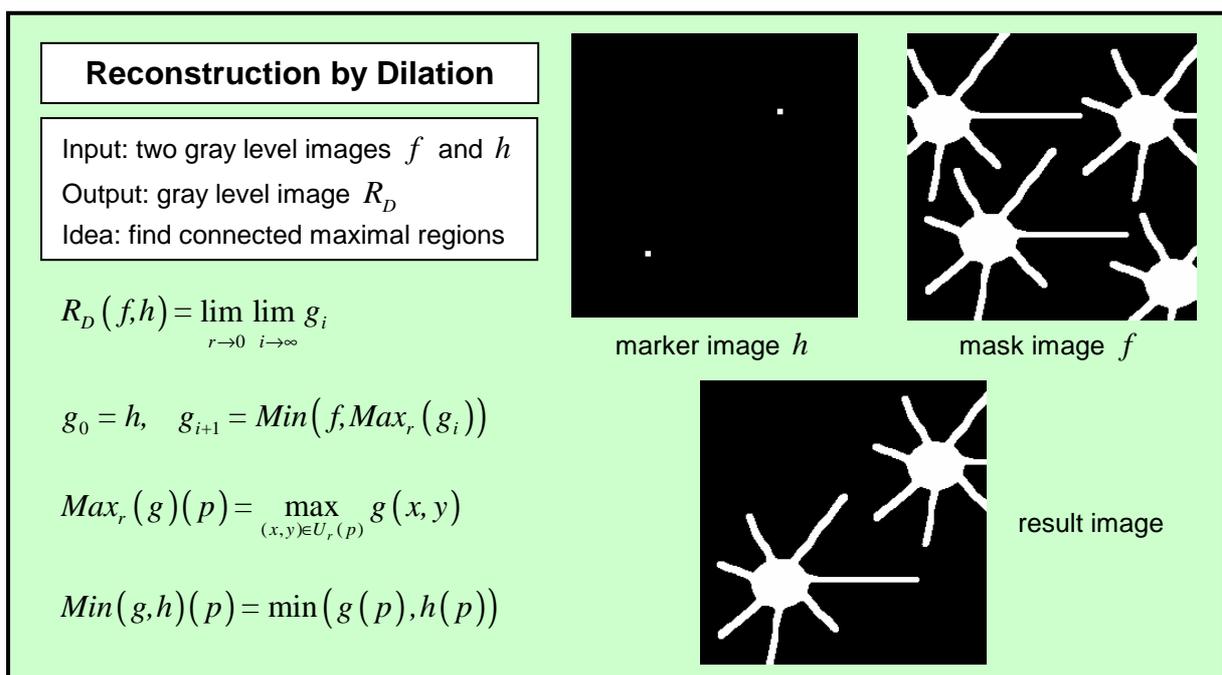
Figure 3: Scratch Detection using AAS

**AAS** (Figure 4) estimates the local covariance matrix from the data and smooths the data in the direction of the largest variance. This effectively removes the binary noise and at the same time improves the contrast of candidate scratches (Figure 3).



**Figure 4: Adaptive Anisotropic Smoothing (AAS)**

The **IEDR** approach (Figure 5, Figure 6) aims to connect the highly fragmented scratch candidates to obtain regions with large aspect ratio. The noise pixels are either removed or become connected components with low aspect ratio. After labeling with the aspect ratio and thresholding, only the correct regions remain (Figure 7).



**Figure 5: Reconstruction by Dilation**

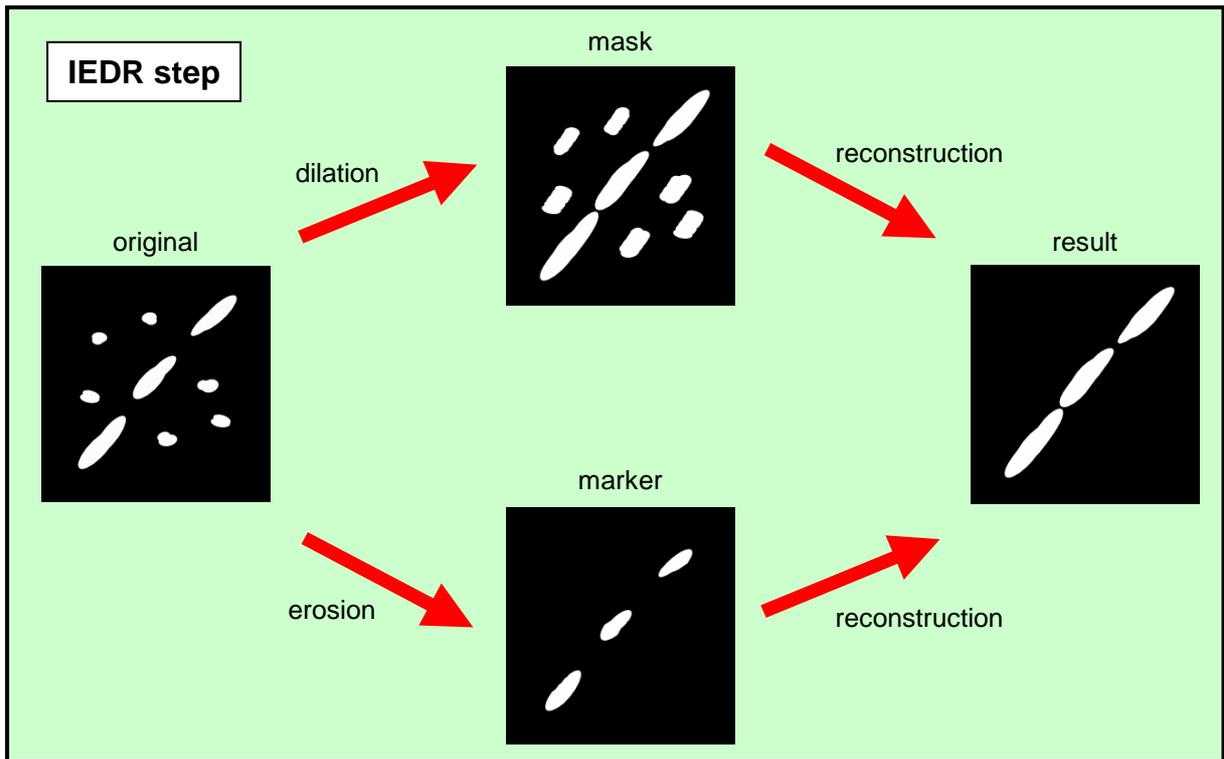


Figure 6: One IEDR Step

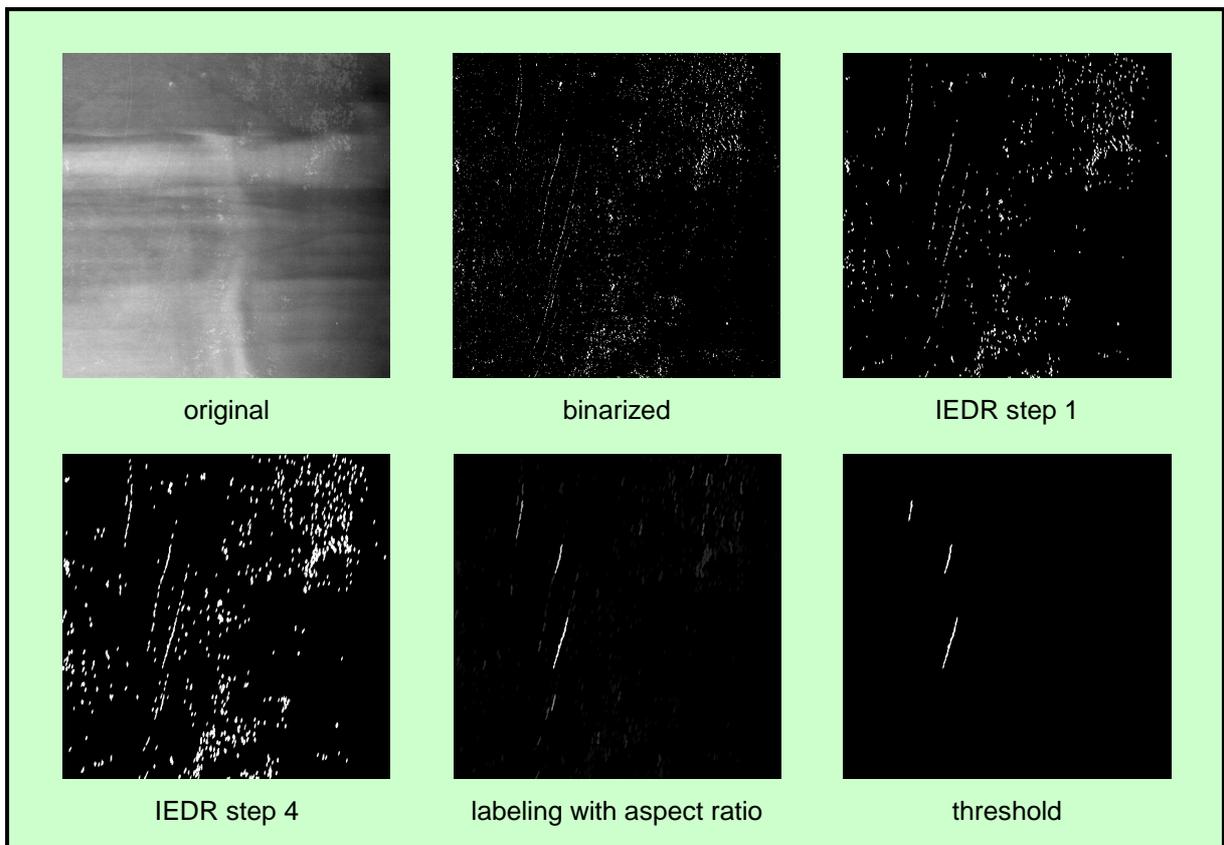


Figure 7: Scratch Detection using IEDR

## SUMMARY

In the course of the mentioned cooperative project, a prototype system has been successfully integrated in the production process of a Swedish tannery. Image processing algorithms have been developed and implemented which already yield satisfying defect detection results. A reliable shingle detection algorithm is still in development, a method based on gray level geodesic reconstruction performs currently the best.

The prototype system achieves a processing speed of approximately 1 hide per minute instead of the aimed 3 hides per minute. Further improvements in the performance of the algorithms are thus still necessary.

### References:

- [1] N. Bauer, M. Maasland: "Oberflächen zuverlässig prüfen", wt Werkstatttechnik online, Springer VDI Verlag, 6: 402-404, 2006
- [2] T. Decker: "Datenklassifikation mittels Bayestechniken", Diploma Thesis, TFH Berlin, Department of Mathematics, 2005
- [3] GFal Berlin: "Automatische Lederfehlererkennung im Produktionsprozess", Abschlussbericht zum AiF FV-Nr.: 13801 BR, Berlin, May 2006
- [4] S. Halim: "Spatially Adaptive Detection of Local Disturbances in Time Series and Stochastic Processes on the Integer Lattice  $\mathbb{Z}^2$ ", Dissertation, University of Kaiserslautern, Department of Mathematics, 2005
- [5] K. Kohrt: "Automatische Qualitätskontrolle – Bildverarbeitung in der Industrie", Keramische Zeitschrift, 2: 90-92, 2005
- [6] M. Rauhut: "Konzeption und Aufbau eines Online-Oberflächeninspektionssystems", Leitfaden zur Inspektion von Oberflächen mit Bildverarbeitung, Fraunhofer Vision, 9: 5-8, 2006
- [7] H. Spies: "Kostenfunktionen und Scheduling von (2D-)Bildverarbeitungsalgorithmen", Diploma Thesis, University of Kaiserslautern, Department of Computer Sciences, 2005
- [8] C. Wolf: "Erstellung eines webbasierten Leitstandes für die industrielle Oberflächeninspektion", Diploma Thesis, FH Zweibrücken, Department of Computer Sciences, 2005
- [9] L. Zedler, G. Stanke, A. Trommer: "Fehlererkennung an Halbfabrikaten und Fertigledern mittels digitaler Bildverarbeitung", Tagungsband 3. Freiburger Kollagensymposium, September 2004
- [10] L. Zedler, G. Stanke, A. Trommer, M. Meyer: "Identification of Surface Leather Defects in Wet Blue State by Digital Image Processing", Poster, 28<sup>th</sup> IULTCS Congress, Firenze, Italy, March 2005
- [11] Plutarch. Der Geist ist kein Schiff, das man beladen kann, sondern ein Feuer das man entfachen muss.

### Authors:

Dipl.-Phys. Andreas Jablonski

Dipl.-Math. Kristina Kohrt

Prof. Dr. Martin Böhm

Fraunhofer-Institut für Techno- und Wirtschaftsmathematik, Fraunhofer-Platz 1

D 67663 Kaiserslautern

Phone: +49 (0) 631 / 31600 – 4385 (Andreas Jablonski)

+49 (0) 631 / 31600 – 4485 (Kristina Kohrt)

+49 (0) 631 / 31600 – 4687 (Martin Böhm)

Fax: +49 (0) 631 / 31600 – 1099

E-mail: andreas.jablonski@itwm.fraunhofer.de

kristina.kohrt@itwm.fraunhofer.de

martin.boehm@itwm.fraunhofer.de

M. Rosenberger/ M. Schellhorn/ P. Brückner/ G. Linß

## **Uncompressed digital image data transfer for measurement techniques using a two wire signal line**

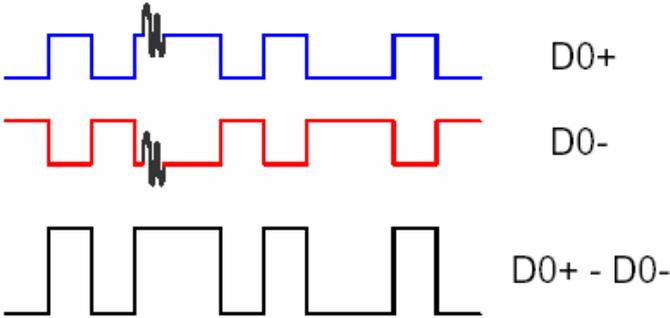
### **Abstract**

Image processing technologies attain an ever higher value in the industrial measuring. Presently growth amounts to approx. 10-15% [1] and according to current studies this trend will continue. The transition from analog interfaces to digital pixelsynchronous systems is in many applications already realized. A disadvantage in the purchase to analog systems is the increased number of necessary wires. Interfaces from the consumer sector as for example Universal Serial Bus (USB) or Fire Wire usually utilizing a lesser number of wires and would be applicable here. Disadvantageously, these interfaces are limited by transmission length and/or transmission rate. The new Firewire standard IEEE 1394b for example, permits transmissions to maximally 100 Mbit/s over Category 5 unshielded twisted pair (UTP) cable. Higher data transmission rates limit the transmission length or require special cables such as plastics optical for fiber (POF) or hard polymere clad fiber (HPCF) known as glass optical fiber (GOF). [2] However, in complicated superstructures respectively old systems often only existing cable systems are usable. In most cases only an analog two wire system is available. There is therefore a demand for a broadband signal transmission on existing line systems. Current developments of the Department for Quality Assurance and Image Processing are dealing with the construction of a low voltage differential signal (LVDS) transmission line for transmitting digital video data over two wires. Based on a parallel digital interface of a CCD camera a possible serialization of the data was analysed. In a prototype application video data with a resolution of 782 pixel (px) to 538 px could be transferred over a distance of 20 m. Furthermore a signal repeater was developed to emphasize the transmission distance. A specification and first test results of the application shows the following paper.

### **Introduction**

As described in the abstract there is a need for image processing technologies in industrial measurement and university research. Mostly some standardized interfaces like USB or Fire Wire were deployed to transmit the digital data from a Charge Coupled Device (CCD) or Complementary Metal Oxide Semiconductor (CMOS) camera. The

disadvantage of these interfaces is the limited cable length. Furthermore actual there exist some measurement setups using analog image signals. For the signal transmission often a coaxial cable or a balanced two wire cable is used. The devices working very well, but a digital transfer gives the possibility of a higher accuracy and more transition safety. That is one of the reasons to research for the possibility to transmit digital images trough existing two wire signal cables, where actually analog signal transfer is used. For a significant range of cameras with a pixel count under 1Megapixel were used. Bridging the distance over 10 m or more is another reason for the actual development. Famous serial digital interfaces like USB, Fire Wire or others use a differential signaling for transmitting the data trough the line. The advantage of use is the high failure safety. Another important technique is the Low Voltage Differential Signaling (LVDS) technique. Therefore a lot of transducers were developed by several companies. The driver generates two signals mostly so called D+ and D- without a ground connection. These signals are inverse. The receiver subtracts the both signals. If there were any interference during the cable length, then it was subtracted too. The principle of work is also shown in figure 1.



*Figure 1: principle of differential signal processing [3]*

The need for more resolution in anyway, for example High Density Tele Vision, drives the industry to develop LVDS driver-receiver systems with large bandwidth up to 1Gigabit per second (Gbits). The advantage of these systems is the need of only one cable pair. Two conditions must be attended: the cable impedance of 100 Ohm for a twisted pair line and a impedance controlled layout for the LVDS Lines on the Printed Circuit Board (PCB). Historically the LVDS standard was a joint development of Texas Instruments and National Semiconductors.

## Theory of the setup

The LVDS standard is standardized by the American National Standards Institute ANSI and the Telecommunication Industry Association (TIA) / Electronic Industries Association (EIA). The standard describes the electrical interface. The signal transition is realized by a current loop. A switchable current source drives the current loop. At the end of the line a termination resistor (cable specific) converts the current signal into a voltage. Under optimal conditions the voltage level reaches  $\pm 350\text{mV}$ . The Receiver interprets a negative voltage level as a logical 0 and a positive level as logical 1.

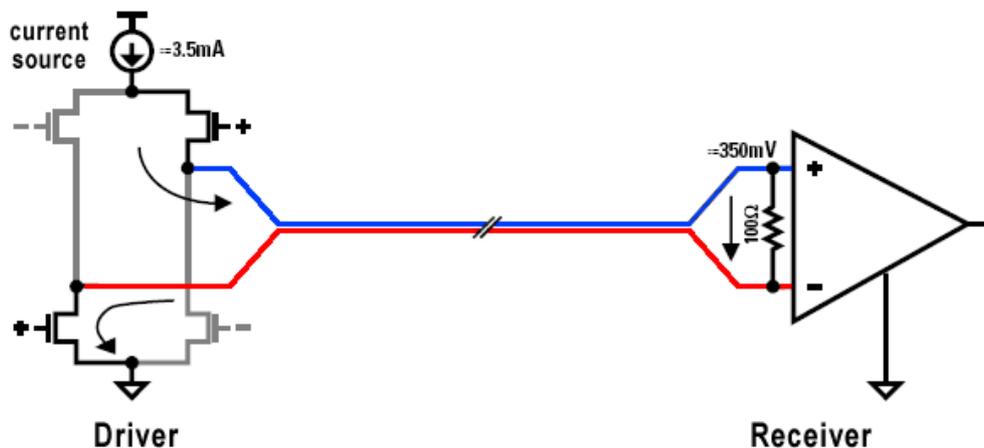


Figure 2: principle of the electrical interface [4]

The low voltage level gives the chance of a higher switching frequency. Therefore a high data rate can be transmitted. Actually it is possible to reach data rates up to 1 Gbits [5]. But it depends on the cable length and the signal frequency. An approximation of reachable data rates and cable length is given in [5]. For the planned setup it is possible to bridge a distance until 10 meters at 0,5 Gbits data rate. The output of a digital interface of a CCD Sensor is mostly parallel. Sometimes it is directly the interface from the analog to digital converter. Often there is a digital filter or a image buffer between CCD Sensor output and digital interface. To transmit these data over LVDS it is necessary to convert these signals in a serial data stream. It is beneficial to convert the output in a LVDS conform signal. This offers the possibility to connect the output directly on a LVDS line driver. The principle of serialization is shown in figure 3. It works like a inverse shifting register. After every clock cycle the actual bit state is shifted into serial data. For imaging systems normal the pixel clock is used to drive the serializer. The clock recovery in the deserializer is realized by sending a special start and stop pattern with the bit stream. For example a 10 bit Information is covered with start bit at the beginning and a stop bit at the end.

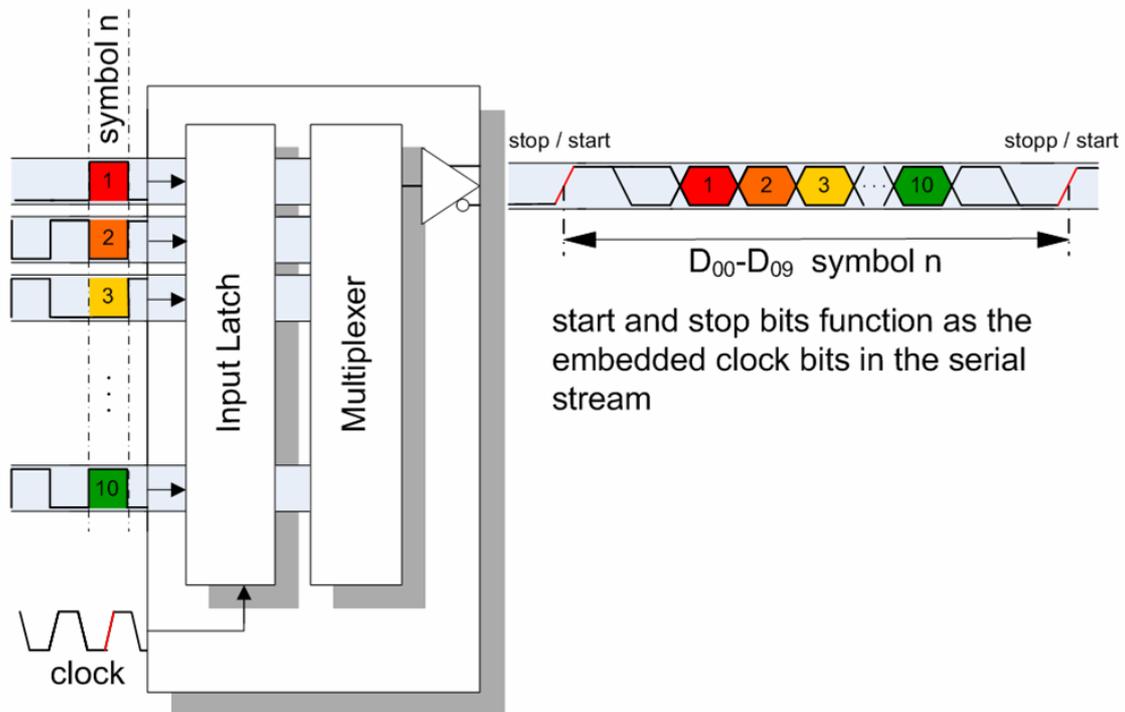


Figure 3: serializing a parallel sensor output [3]

The start condition (bit) is generated by driving both differential lines into the high level. Holding both lines in the low level state the stop condition (bit) is generated. The deserializer has the possibility to lock on these patterns with a PLL. The advantage of this kind of serializing is the need of only two lines. Therefore the data stream will be expanded until  $n+2$  Bit. So the signal frequency on the signal lines increase rapidly. Actual there is a convention between number of input bits and input clock frequency.

### Experimental setup

As described in the abstract, a setup with two signal lines transmitting a digital image was realized. Figure 4 gives an overview about the whole setup. The CCD Sensor streams with a pixel clock of 14,1875 Mhz and a picture size of 768x538 px. The image reference signal, the start-stop conditions and the eight bit image information per each pixel leads to a serial signal frequency of 170,25 MBit.

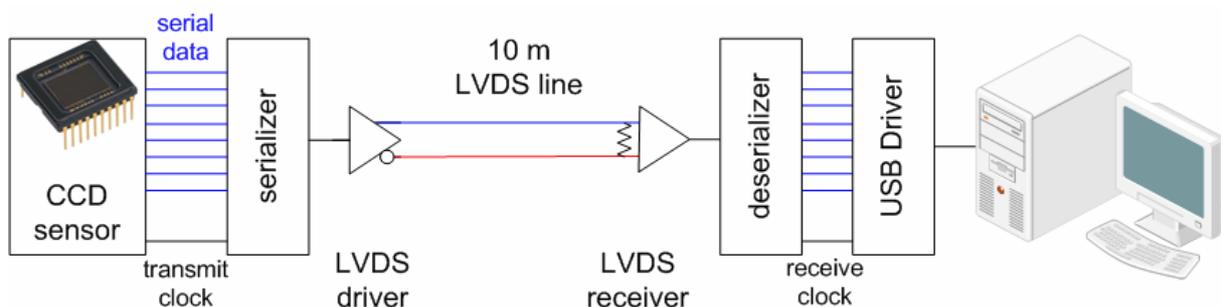
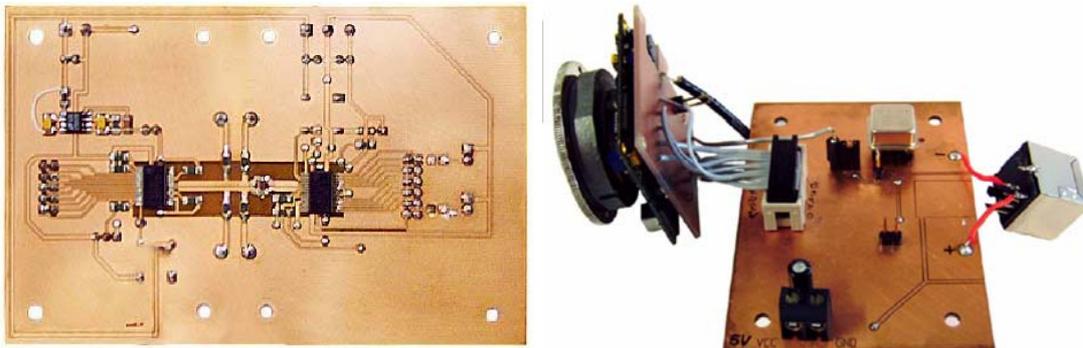


Figure 4: serializing a parallel sensor output

For transmitting of the image data a serializer/deserializer couple from the Texas Instruments company was chosen. The front end, getting the picture into the PC, was a self developed USB interface. For testing the setup, a special printed circuit board was developed. Firstly, both the driver and the receiver were connected on the same PCB. Special attention was paid for the correct differential signaling. So it is necessary to support a differential impedance of 100 Ohms for the LVDS Lines. The simplified dual strip line model was used to estimate the correct impedance on the PCB. The designed board offers the possibility to cut it in the middle. So a separate driver and receiver unit can be generated extreme easily. After the first general test it was cut. Then some different cable lengths were connected between the two units. Looking forward to reaching higher transmit distances a LVDS repeater based on the line driver couple was also developed.

### Experiments

After the general function test with the test setup the board was cut into the driver and the receiver. The picture on the right hand side contains the whole test board.



*Figure 5: test board for LVDS streaming (left), sender unit after cutting (right) [3]*

To get some information about the locked receiver, an error Light Emitting Diode (LED) was connected on a monitor port. A normal RJ 45 Connector was used as the interface for the differential lines. This gives the possibility to utilize standard network cables (normally with  $Z=100$  Ohms). The picture on the right hand side shows the driver unit. A comparable device was build as a receiver. In accordance to the standard the test cable length was 10 meter. For this distance the setup works very well and stable. To get some information about the boundaries the cable length was increased by 5 meter steps. The system works stable until 15 meter cable length. At higher cable length the synchronization was lost accidentally. Therefore a signal repeater was developed to gain the signal. A special adapted driver receiver circuit was discovered at the Texas Instruments Company for this purpose. The repeater gives the capability to bridge higher

distances. Theoretically it is possible to connect every 10 meters a repeater. For the test only one repeater was used.

### **Results**

The setup demonstrates that LVDS signaling gives the possibility to transmit uncompressed digital image data over a balanced two wire cable. A distance until 10 meter can be bridged with this setup. Using a repeater helps bridging higher transition distances. So it gives the chance to exchange some old analog systems through digital ones or bridging distances that standard solutions can not bridge.

### **Conclusion and Outlook**

The actual setup is in an experimental state. So an optimized layout will bring better working conditions for the data transfer. The error LED on the receiver can be used for error recognition in the data transfer. Only the electrical signal must be transferred and analyzed in the Host PC. Furthermore the next step should be the integration of a redesigned layout in an application, where analog signal transfer is actually used. After that, it is possible to get information about the better measurement conditions with the digital data transfer. In the industrial development some LVDS devices were developed. These devices deal with higher transfer distances and give new possibilities for application.

### **References:**

- [1] Branchenportrait Industrielle Bildverarbeitung 2005, VDMA, Frankfurt am Main;
- [2] [www.1394ta.org](http://www.1394ta.org);
- [3] Mathias Schellhorn, Untersuchungen zur digitalen differentiellen Signalübertragung in der Bildverarbeitung, TU-Ilmenau 2006;
- [4] National Semiconductor, LVDS Owner's Manual Low-Voltage Differential Signaling, 3rd Edition, 2004;
- [5] Texas Instruments; LVDS Application and Data Handbook, 2002;

### **Authors:**

Dipl.-Ing. Maik Rosenberger  
cand.-Ing. Mathias Schellhorn  
Dr.-Ing. Peter Brückner  
Prof. Dr.-Ing. habil. Gerhard Linß

Technische Universität Ilmenau  
Quality Assurance Department  
P.O. Box 100 565  
98684 Ilmenau  
Germany  
Phone: +49 3677 693961  
Fax: +49 3677 693823  
E-mail: maik.rosenberger@tu-ilmenau.de, mathias.schellhorn@stud.tu-ilmenau.de

R. Blaschek / B. Meffert

## Feature Point Matching for Stereo Image Processing using Nonlinear Filters

### ABSTRACT

*For evaluating stereo image pairs of two cameras the calculation of the so-called fundamental matrix is an appropriate method. The matching algorithm requires a large number of proper reference points. If the cameras have an almost parallel alignment, typically very simple and fast correlation algorithms (e.g., sum of absolute differences) estimate the similarity between feature point candidates. But in practice such algorithms will fail because the alignment of the cameras is often less restrictive.*

*This article describes a method that uses nonlinear filters, derived from Gabor wavelet filter sets, to extract information about the texture of the reference points. These filters can help to overcome usual matching problems caused by intensity differences or rotation and scaling of surrounding gray values of the reference point candidates.*

### NONLINEAR FILTERS

Corner detectors, like Harris corner detector or SUSAN, provide a sufficient large number of points of interest. For each point a spectral signature can be calculated using filter sets of modified Gabor wavelets. The modifications are based on an adaptation that satisfies the wavelet theory and the neurophysiological constraints for so-called *simple cells* in the visual cortex [1]:

$$\psi(x, y, \omega_0, \theta, \kappa) = \frac{\omega_0}{\sqrt{2\pi\kappa}} e^{-\frac{\omega_0^2}{8\kappa^2}(4(x \cos \theta + y \sin \theta)^2 + (-x \sin \theta + y \cos \theta)^2)} \cdot \left[ e^{i(\omega_0 x \cos \theta + \omega_0 y \sin \theta)} - e^{-\frac{\kappa^2}{2}} \right] \quad (1)$$

The modifications of the Gabor wavelets adjust the elliptical Gaussian that envelopes the complex plane wave such that each of the wavelets has the same number of extremal values. Scale and direction of the wavelet can be chosen by  $\omega_0$  and  $\theta$ , whereas  $\kappa$  allows to specify the bandwidth of the filter ( $\kappa = \pi$  corresponds to one octave). Since  $\psi$  provides a complex-valued 2D Gabor function one can obtain an even-symmetric

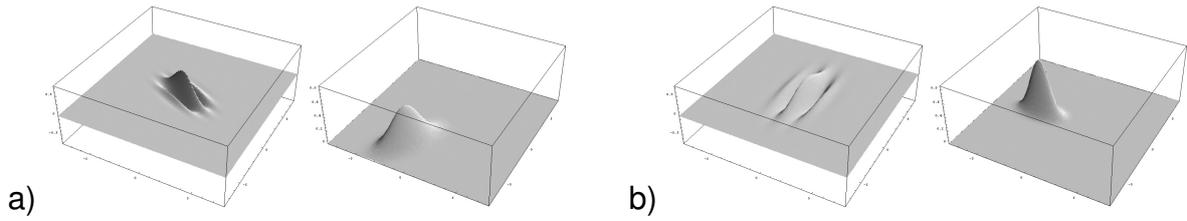


Figure 1: Real part of two Gabor wavelets for orientation angles a)  $\theta = 65^\circ$  and b)  $\theta = 0^\circ$  and the corresponding power spectra.

cosine and an odd-symmetric sine component by splitting the function into its real and imaginary part. Figure 1 shows an even-symmetric wavelet for two orientations.

The spectral signature for the reference points is calculated by applying an ensemble of Gabor wavelets on these points and their surrounding gray values. Typically the filter set consists of eight to sixteen orientations at four to six scales. Each filter mask is convolved with the actual image in the surrounding window of the reference points, i.e.  $16 \times 16$  pixel, and the result is accumulated. This leads to a signature with an approximation of the signal energy for certain scale and direction.

Because of the high computational complexity of the convolution operation, this way of calculating the spectral signature is quite ineffective for large numbers of reference points. Heintz and Schäfer [3] calculate the filtering in the frequency domain. To do this the Gabor wavelets have to be converted by Fourier transform (Figure 2):

$$\Psi(\omega, \phi, \omega_0, \theta, \kappa) = -\sqrt{\frac{2}{\pi}} \frac{\kappa}{\omega_0} e^{-\frac{\kappa^2((\omega_0 + \omega \cos \theta + \phi \sin \theta)^2 + 4(\phi \cos \theta - \omega \sin \theta)^2)}{2\omega_0^2}} \cdot \left[ -1 + e^{\frac{\kappa^2(\omega \cos \theta + \phi \sin \theta)}{\omega_0}} \right] \quad (2)$$

Now the windows around the reference points have to be transformed too, but the computational expense is noticeable lower, because there are fast algorithms available for the Fourier transform. Because the signatures require just an approximation of the

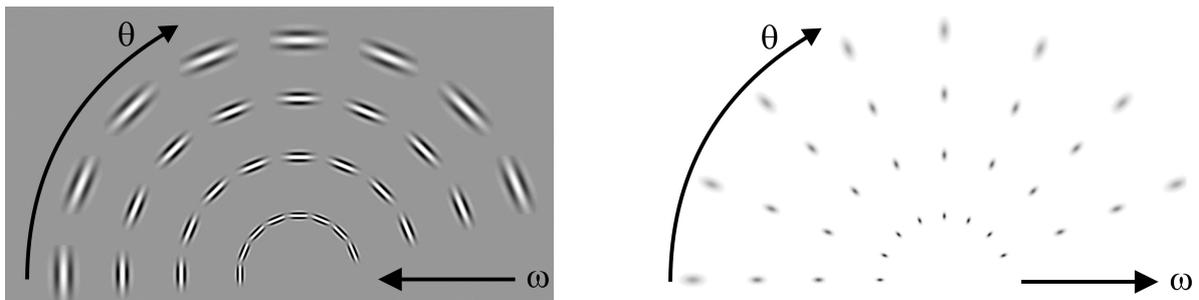


Figure 2: Ensemble of Gabor wavelets and Gabor filters to create a signature with eight orientations in four scales.

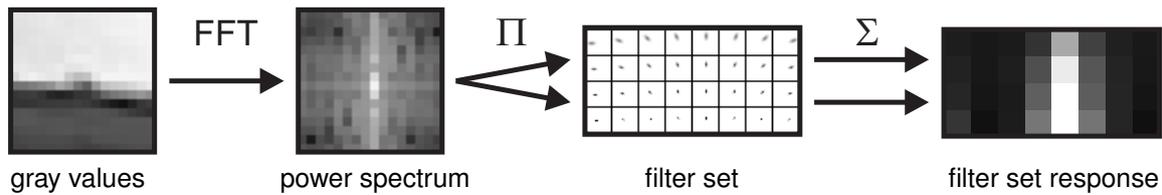


Figure 3: Processing chain for calculating the spectral signature of a reference point.

signal energy, the complex-valued result of the Fourier transform can be substituted by the real-valued power spectrum. Since  $\Psi$  provides a small Gaussian in the scale space, most parts of these Gabor filters will be zero and thus the multiplication and accumulation with the power spectra of the reference points is reduced to just a few elements (Figure 3).

### DESCRIPTORS FOR MATCHING

To accomplish the stereo matching successfully, the matching algorithms need meaningful descriptors that characterise the reference point candidates. The spectral signature described above is such a descriptor:

- It is invariant to changes in scene lighting because the constant component of the Fourier spectrum is never used by the Gabor filter.
- It provides the ability to a rotation invariant comparison because circular shifting the signature along the  $\theta$ -axis is equivalent to rotating the pattern in the image.
- It provides the ability to a scale invariant comparison because shifting the signature along the  $\omega$ -axis is equivalent to scaling the pattern in the image.

First tests with the signatures have shown, that small windows around the reference points – which are typically 16 or 32 pixel wide – interfere with the theoretical rotation invariance. The closer the angular steps  $\Delta\theta$  the less structure is in the signature. This is not surprising, because a  $16 \times 16$ -spectrum has only 137 different spectral coefficients. Using an angular step of e.g.  $\Delta\theta = 10^\circ$  leads to an unwanted oversampling at which some coefficients are used up to 7 times. This is a displeasing property, because little changes in the image as well as noise will be fairly distributed over the signature. One can decide to use larger angular steps, but this is contrary to the ability to compare stereo image pairs with small camera rotations.

Another problem arises from the Fourier transform itself. Because of the discrete transformation and the small window size, there is (nearly always) a noticeable strong concentration of the spectrum along the both axis. Regarding the rotation invariance this part of the signal doesn't behave like expected – it is stationary, but still signal

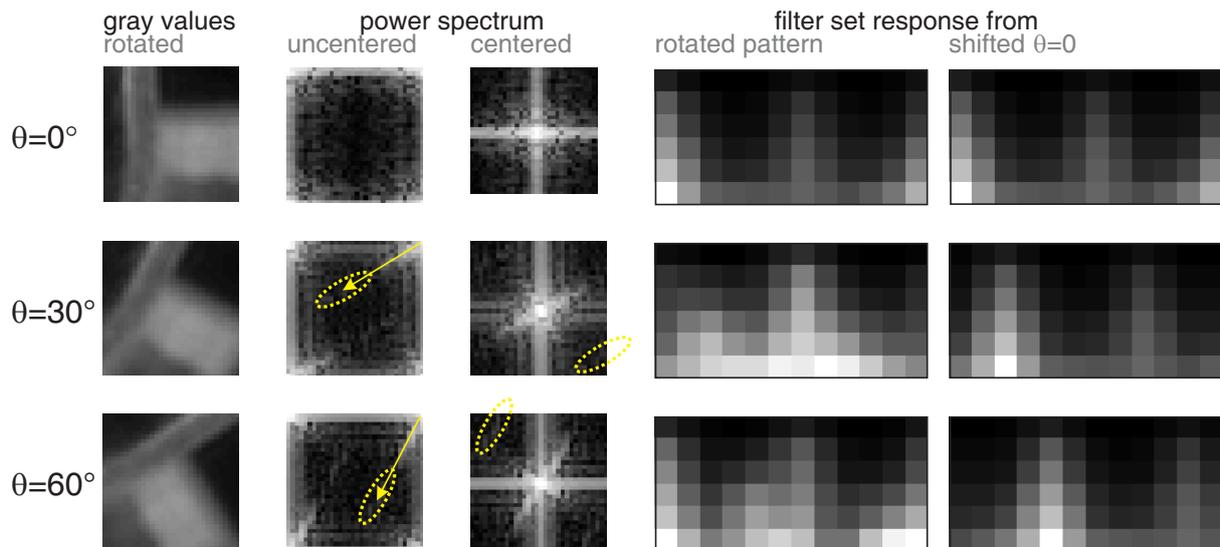


Figure 4: Rotation artifacts on high frequencies and dominant axis for three angles  $\theta$ .

dependent. In addition Figure 4 shows some artifacts in the Fourier spectrum caused by high frequencies in the gray values.

To overcome this problem we have found, that alignment of the windows before the Fourier transform leads to a more stable matching result. According to the approach used in SIFT (Scale Invariant Feature Transform) by Lowe [2], the orientation of each reference point is calculated from the gradient. Subsequently, each window is rotated by its orientation angle – which also simplifies the comparison because the signature can be assumed to be aligned. Furthermore, the image should be processed by a low-pass filter to avoid the superposition of signals in different quadrants. Gabor filters on the diagonals of the spectrum (at  $45^\circ$  or  $135^\circ$ ) may prevail the other filters because they accumulate more coefficients. Therefore all spectral coefficients that are outside the radius of the half window length are set to zero.

Using dense reference point sets makes it very difficult to distinguish between similar looking points. Often the signatures of those points have the same mixture of scales and directions. That is why the descriptors need another extension that not only detects the scales and orientations, but also contains some information about their distribution. In our approach we decided to partition the window around the reference point into four sub-windows, each with the half length of the original window. In doing so the main signature is calculated from a  $32 \times 32$  window and the  $16 \times 16$  sub-windows are used to calculate four sub-signatures. Figure 5 shows the chosen layout for the placement.

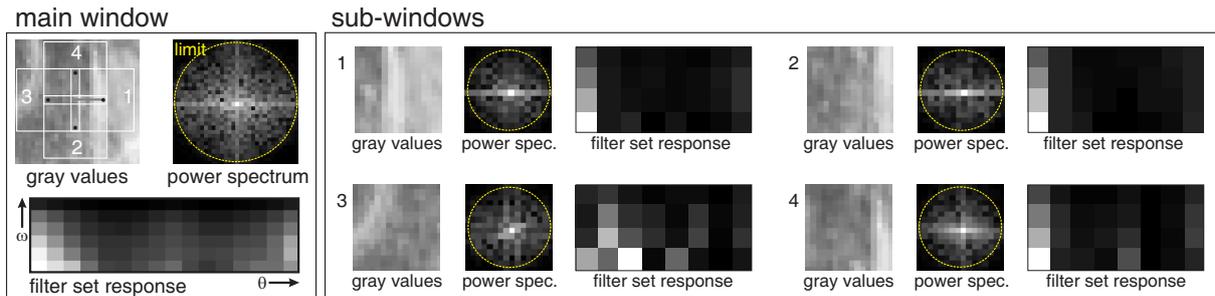


Figure 5: Placement of the four sub-windows and results of the Gabor filtering on the limited spectra.

## RESULTS AND CONCLUSIONS

In this paper we propose a new descriptor for characterising reference points for stereo matching. This descriptor uses nonlinear Gabor filters to extract illumination and rotation invariant features from the Fourier spectrum. Tests with different image pairs have shown, that the descriptors are well suited for the following classes of images:

- random pattern surface with low depth – Figure 6, top (*pothole*)
- TV-lens scene of poor images with noticeable lens distortion and differences in exposure – Figure 6, center (*office room*)
- well structured surface with large depth levels – Figure 6, bottom (*amphora*).

The reference points are taken from a corner detector. For the right images the number of points and their density is three times higher than for the left images to rise the chance that the corresponding points are within these point sets. The extracted descriptors are compared and the candidate with the minimal distance is chosen iff there is no other candidate with a distance smaller than 1.5 times the minimal distance.

Even though the descriptor works very well for ‘unscaled’ images, it should be further improved to profit from the ability of the Gabor filters to a scale invariant comparison.

### References:

- [1] T. S. Lee: *Image Representation Using 2D Gabor Wavelets*. Carnegie Mellon University of Pittsburgh, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, No. 10, October 1996.
- [2] D. G. Lowe: *Distinctive Image Features from Scale-Invariant Keypoints*. University of British Columbia Vancouver, Computer Science Department, International Journal of Computer Vision, 60, 2, pp. 91-110, 2004.
- [3] R. Heintz und G. Schäfer: *Lokale invariante Objektlokalisierung mittels Gaborfiltern*. FH Karlsruhe, GMA-Kongress, VDI-Bericht 1883, S. 441-448, 2005.

### Authors:

Dipl.-Inf. Roman Blaschek  
 Prof. Dr. Beate Meffert  
 Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099 Berlin  
 Institute of Computer Science, signal processing and pattern recognition group  
 E-mail: {blaschek,meffert}@informatik.hu-berlin.de

| image pair         | size     | used points (l/r) | matches      |
|--------------------|----------|-------------------|--------------|
| <i>pothole</i>     | 848×1280 | 1500 / 4500       | 1285 (85.6%) |
| <i>office room</i> | 288×352  | 627 / 808         | 285 (45.4%)  |
| <i>amphora</i>     | 640×480  | 1500 / 4500       | 669 (44.6%)  |

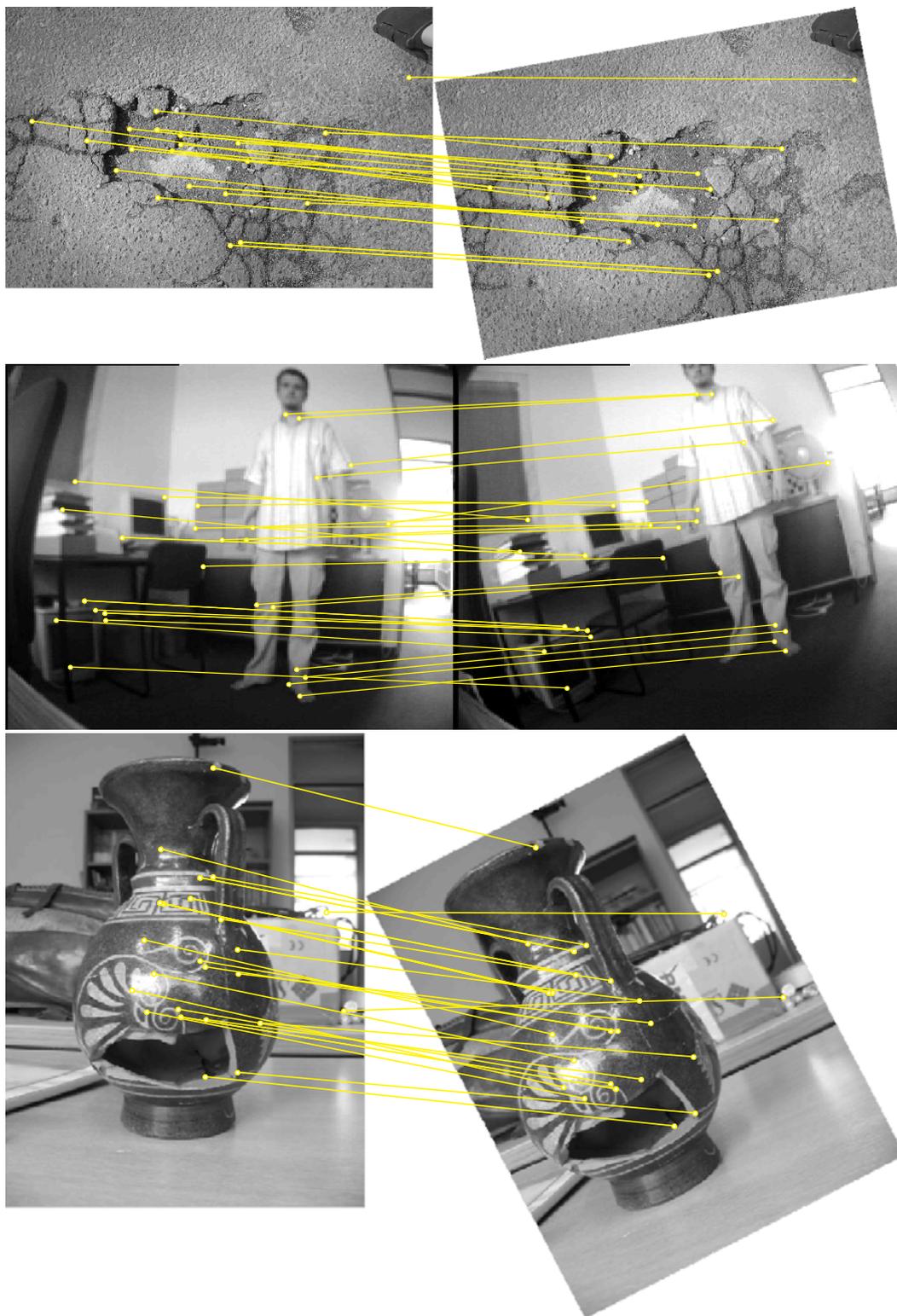


Figure 6: Sample image pairs for three different scene classes with randomly chosen matches and some details of the image pairs.

A. Mitsiukhin /V. Pachynin/ E. Petrovskaya

## **Hartley Discrete Transform Image Coding**

### **IMAGE PROCESSING, IMAGE ANALYSIS AND COMPUTER VISION**

The prevailing development vector in modern radio-electronic information-measuring systems (complexes) is the introduction of digital methods and processing, transformations and signaling. However the use of digital methods leads to an increase in a frequency bandwidth, reduction of the processing rate and the volume of transferred data. Traditional problems for digital systems are their speed and utilized algorithms. In part these problems can be resolved by the development of effective digital coding methods (compression) of information.

Information compression methods eliminate the redundancy of digital representation of the information. They represent a procedure of linear digital signal coding on the basis of discrete unitary transforms (discrete Fourier (DFT), Walsh-Hadamard (WHT), Haar, cosine (DCT), etc.).

This paper addresses the compression of video information (images) that is to be transmitted using two-dimensional digital signals, using the Hartley discrete orthogonal transform (DHT). The compression efficiency is estimated and compared with the presently used DCT and WHT.

In general the transformation coding is realized in two operations. The first operation is the linear transformation of the original data. The second operation reduces (selects) the transformation factors (transformants). The number of these transformants can be reduced by fixing some threshold level of values or by allocating the most informative zones in a floor of spatial frequencies of digital images. Below we describe the procedures of zone selection (filtration) of transformation factors with the use of DHT. The procedure of an efficient coding of two-dimensional digital signals is as follows:

1. A direct two-dimensional DHT of the original  $N \times N$  image fragment is calculated. It can be represented the matrix form as

$$[G(k_1, k_2)] = \frac{1}{N} [v(k, n)] [g(n_1, n_2)] [v(k, n)]^T, k_1, k_2, n_1, n_2 \in \{0, 1, \dots, N-1\}, \quad (1)$$

where  $[g(n_1, n_2)]$  is the readout matrix of the image of size  $N \times N$ ;  $[G(k_1, k_2)]$  - the  $N \times N$  matrix of DHT factors;  $[v(k, n)]$  - an  $N \times N$  kernel of the DHT transform;

$$[v(k, n)] = \begin{bmatrix} \cos \frac{2\pi kn}{N} \\ \sin \frac{2\pi kn}{N} \end{bmatrix}, k, n \in \{0, 1, \dots, N-1\}$$

where  $\cos\left(\frac{2\pi kn}{N}\right) = \cos\left(\frac{2\pi kn}{N}\right) + \sin\left(\frac{2\pi kn}{N}\right)$ . (2)

The 2D reverse Hartley transform has the form of

$$[g(n_1, n_2)] = \frac{1}{N} [v(k, n)]^T [G(k_1, k_2)] [v(k, n)]. \quad (3)$$

Transformation matrices of the direct and reverse DHT are identical, as

$$[v(k, n)] = [v(k, n)]^T.$$

2. Zone filtration of transformant is being performed, which requires prior knowledge of the distribution function for a two-dimensional dispersion of transformation factors

$$diag[\sigma^2] = diag[\tilde{K}_C] \otimes diag[\tilde{K}_R], \quad (4)$$

where  $\otimes$  is the Kronecker product of matrices;  $diag[\tilde{K}_C]$  и  $diag[\tilde{K}_R]$  - diagonal covariances matrices of transformation factors columns and rows respectively. Covariance matrices  $K_C$  and  $K_R$  in the field of originals and in the area of images are connected by transformation of similarity. Then  $[\tilde{K}_C]$  и  $[\tilde{K}_R]$  are defined from expressions

$$[\tilde{K}_C] = [v(k, n)] [K_C] [v(k, n)]^T, \quad (5)$$

$$[\tilde{K}_R] = [v(k, n)] [K_R] [v(k, n)]^T. \quad (6)$$

After lexicographic transformation of a matrix  $diag[\sigma^2]$  the matrix  $[L]$ , is formed, which defines the zone of transformant selection. Recovery of the original image fragments is performed by the reverse spectral transformation according to the dispersive criterion, where the kept transformants possess the greatest dispersions chosen by the filtration zone. The following example shows the estimation of quality of 8x8 fragment compression for the three transformations: DCT, DWAT, and investigated DHT.

*Example.* An input image fragment has the size of 8x8. The image is described by the Toeplitza distribution. The correlation factor  $p$  for adjacent readouts is set to 0.9.

The covariance matrix for the fragment's columns and rows and the DHT kernel (2) is as follows:

$$K_C = K_R = \begin{pmatrix} 1 & 0.9 & 0.81 & 0.729 & 0.656 & 0.59 & 0.531 & 0.478 \\ 0.9 & 1 & 0.9 & 0.81 & 0.729 & 0.656 & 0.59 & 0.531 \\ 0.81 & 0.9 & 1 & 0.9 & 0.81 & 0.729 & 0.656 & 0.59 \\ 0.729 & 0.81 & 0.9 & 1 & 0.9 & 0.81 & 0.729 & 0.656 \\ 0.656 & 0.729 & 0.81 & 0.9 & 1 & 0.9 & 0.81 & 0.729 \\ 0.59 & 0.656 & 0.729 & 0.81 & 0.9 & 1 & 0.9 & 0.81 \\ 0.531 & 0.59 & 0.656 & 0.729 & 0.81 & 0.9 & 1 & 0.9 \\ 0.478 & 0.531 & 0.59 & 0.656 & 0.729 & 0.81 & 0.9 & 1 \end{pmatrix}$$

$$v(k, n) := \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1.414 & 1 & 0 & -1 & -1.414 & -1 & 0 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 0 & -1 & 1.414 & -1 & 0 & 1 & -1.414 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -1.414 & 1 & 0 & -1 & 1.414 & -1 & 0 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 0 & -1 & -1.414 & -1 & 0 & 1 & 1.414 \end{pmatrix}$$

The covariance matrix in the area of images for columns and rows (5), (6) is equal to:

$$\tilde{K}_C = \tilde{K}_R = \begin{pmatrix} 49.478 & -0.956 & 0 & 0.069 & 0 & -0.166 & -0.566 & -2.307 \\ -0.956 & 6.03 & 1.907 & 1.351 & 0.956 & 0.558 & -4.414 \times 10^{-3} & -1.351 \\ 0 & 1.907 & 1.97 & 0.8 & 0.566 & 0.331 & 0 & -0.79 \\ 0.069 & 1.351 & 0.8 & 1.06 & 0.402 & 0.235 & 6.56 \times 10^{-4} & -0.558 \\ 0 & 0.956 & 0.566 & 0.402 & 0.706 & 0.166 & 0 & -0.396 \\ -0.166 & 0.558 & 0.331 & 0.235 & 0.166 & 0.589 & -1.586 \times 10^{-3} & -0.235 \\ -0.566 & -4.414 \times 10^{-3} & 0 & 6.56 \times 10^{-4} & 0 & -1.586 \times 10^{-3} & 0.838 & -0.011 \\ -2.307 & -1.351 & -0.79 & -0.558 & -0.396 & -0.235 & -0.011 & 3.328 \end{pmatrix}$$

The diagonal covariance matrix for columns and rows is

$$diag[\tilde{K}_C] = diag[\tilde{K}_R] = \begin{pmatrix} 49.478 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 6.03 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.97 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.06 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.706 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.589 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.838 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3.328 \end{pmatrix}$$

From formula (4), a matrix of values can be obtained:

$$diag(\sigma^2) = \begin{bmatrix} 38.254 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 4.663 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.522 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.816 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.544 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.458 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.649 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2.573 & 0 & 0 & 0 \\ * & * & * & * & * & * & * & * & * & * & * \\ * & * & * & * & * & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.031 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.044 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.173 \end{bmatrix}$$

After the lexicographic transform of the matrix  $diag(\sigma^2)$  the matrix [L] is formed as:

$$L := \begin{pmatrix} 2448 & 298.352 & 97.472 & 52.447 & 34.931 & 29.143 & 41.463 & 164.663 \\ 298.352 & 36.361 & 11.879 & 6.392 & 4.257 & 3.552 & 5.053 & 20.068 \\ 97.472 & 11.879 & 3.881 & 2.088 & 1.391 & 1.16 & 1.651 & 6.556 \\ 52.447 & 6.392 & 2.088 & 1.124 & 0.748 & 0.624 & 0.888 & 3.528 \\ 34.931 & 4.257 & 1.391 & 0.748 & 0.498 & 0.416 & 0.592 & 2.35 \\ 29.143 & 3.552 & 1.16 & 0.624 & 0.416 & 0.347 & 0.494 & 1.96 \\ 41.463 & 5.053 & 1.651 & 0.888 & 0.592 & 0.494 & 0.702 & 2.789 \\ 164.663 & 20.068 & 6.556 & 3.528 & 2.35 & 1.96 & 2.789 & 11.076 \end{pmatrix}.$$

From the matrix L it is possible to define the zone of the transformant filtration. For example, for the compression factor of 1.3 the zone has the following shape (the highlighted part of a matrix):

$$L = \begin{bmatrix} 2448 & 298.352 & 97.472 & 52.447 & 34.931 & 29.143 & 41.463 & 164.663 \\ 298.352 & 36.361 & 11.879 & 6.392 & 4.257 & 3.552 & 5.053 & 20.068 \\ 97.472 & 11.879 & 3.881 & 2.088 & 1.391 & 1.16 & 1.651 & 6.556 \\ 52.447 & 6.392 & 2.088 & 1.124 & 0.748 & 0.624 & 0.888 & 3.528 \\ 34.931 & 4.257 & 1.391 & 0.748 & 0.498 & 0.416 & 0.592 & 2.35 \\ 29.143 & 3.552 & 1.16 & 0.624 & 0.416 & 0.347 & 0.494 & 1.96 \\ 41.463 & 5.053 & 1.651 & 0.888 & 0.592 & 0.494 & 0.702 & 2.789 \\ 164.663 & 20.068 & 6.556 & 3.528 & 2.35 & 1.96 & 2.789 & 11.076 \end{bmatrix}$$

Distortions, resulting from the compression procedure, can be estimated by the mean-square error (MSE) of one signal readout, defined using the formula:

$$\sigma^2 = \frac{1}{64} \left[ \sum_{i=0}^7 \sum_{j=0}^7 \sigma_{i,j} \right],$$

where  $i, j$  belong to the chosen zone of filtration. The coding efficiency estimate using a dispersive criterion is illustrated in Figure 1 and 2.

Table 1. Mean-square error of DCT, DHT and DWAT transforms

| k    | 1.1       | 1.3      | 1.63     | 2        | 4     |
|------|-----------|----------|----------|----------|-------|
| DCT  | 0,0003422 | 0,001208 | 0,003208 | 0,005973 | 0,035 |
| DHT  | 0,0006529 | 0,002231 | 0,006028 | 0,011    | 0,043 |
| DWAT | 0,0008346 | 0,002354 | 0,006042 | 0,009432 | 0,04  |

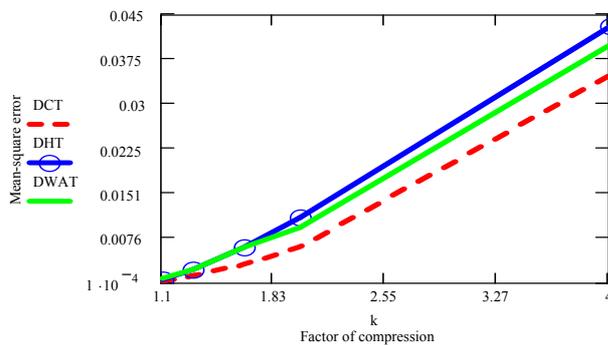


Fig. 1: MSE as a function of on compression factors with the correlation factor  $\rho = 0.9$ , image size of fragments is  $8 \times 8$ .

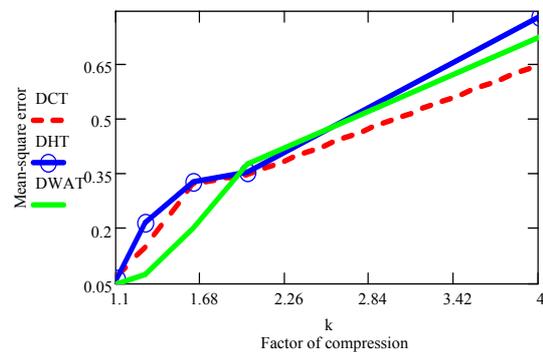


Fig. 2: MSE as a function of compression coefficient for a real fragment of the size of  $8 \times 8$

Conclusion: Computing complexity of the fast DHT algorithm is lower than that of the fast DCT algorithm. Therefore, DHT compression might prove to be valuable for practical purposes.

#### References:

- [1] Mitsiukhin A.I., Petrovskaya E.S. Coding of the information by means of Hartley transform. Materials international scientifically-practical conference.: Mn.: MGVRK, 2006
- [2] Ahmed N., Rao K.R. Orthogonal Transforms for Digital Signals Processing, Springer –Verlag, New York, 1975.

#### Authors:

Dr.- Ing. A. Mitsiukhin  
 Dean.Ph.D. V. Pachynin  
 Dipl.-Ing. (FH) E. Petrovskaya  
 Company, street, P.O.B. Belarussian State University of Informatics & Radioelectronics, Brovka Str. 6  
 Zip code, city BY-220013, Minsk  
 Phone: 00375+172+938991  
 Fax: 00375+172+913908  
 E-mail:mityuhin @bsuir. by



S. Hellbach / B. Lau / J. P. Eggert / E. Körner / H.-M. Gross

# Multi-Cue Motion Segmentation

## 1 Introduction

This paper presents an approach to multi-object image segmentation based on object motion using Markov Random Fields<sup>1</sup>. To support the information gained from motion and to achieve robustness, several additional visual cues extracted from the image data are integrated. Depth information gained from stereo disparity is included to maintain segmentation in case an segmented object stops moving. Motion is estimated with a correspondence matching scheme. The approach differs from regular optical flow in the way that rich matching results are used for segmentation rather than only the best matches. The representation of segmented regions is realized implicitly as labeling on a 2D lattice.

Motion segmentation is a key to many modern image processing applications. In video compression algorithms, the analysis of motion and regions with coherent motion helps to drastically reduce the amount of information that has to be stored and transmitted for each frame [11]. Motion segmentation and motion understanding, for example, plays an essential role in detecting and/or avoiding obstacles in vehicles or with a mobile robot.

The rest of this paper is organized as follows: while Sect. 2 provides a short overview of the work done in the field of image segmentation and relations to our approach, Sect. 3 describes the architecture of the proposed system. Its evaluation is presented in Sect. 4. Finally, the paper concludes with Sect. 5.

## 2 State of the Art

Recent research in the area of motion segmentation focuses on feature-based motion estimation, approaches using level set methods [4], and on multi-cue segmentation on Markov Random Fields [6]. Optimization on Markov Random Fields is an established approach to segmentation [9]. It permits to let motion estimation be part of the segmentation process in one single optimization framework.

When Markov Random Fields are used in combination with optimization methods like iterated conditional modes (ICM) [2], the initialization of the field configuration and the label parameters has a strong influence on the time needed for convergence as well as on the quality of the result. In multi-frame applications, results from the last frame can be used for initialization [3]. Our approach uses motion estimates that fulfill certain quality criteria, as well as results from the previous frame.

---

<sup>1</sup>This paper reflects the results of Boris Lau's diploma thesis, available online at <http://www.borislau.de/computerscience/publications/>

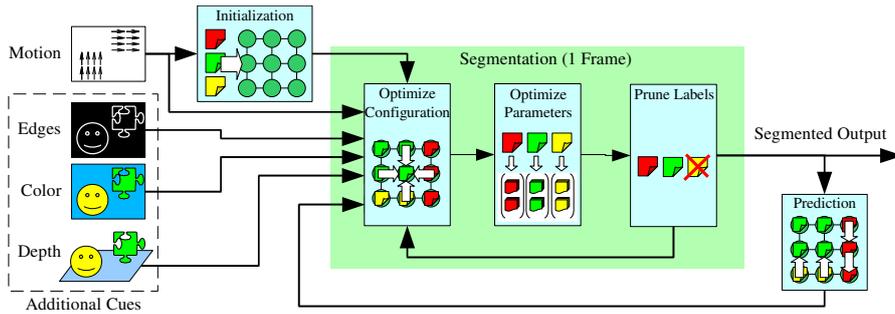


Figure 1: Schematic overview of the framework’s architecture. Multiple visual cues are incorporated by the segmentation framework. The data resulting from the segmentation in the current time step is used as initialization for the following by applying the motion information for each segment.

In state-of-the-art optimization-based motion segmentation systems with an implicit representation, motion information is included either in the form of spatio-temporal image gradients [9], or as optical flow [12]. Our system transfers the approach to use sum of squared differences (SSD) surfaces instead of just the best matches [8] to motion segmentation by directly integrating SSD surfaces into a cost function.

Our system as well as other approaches (e.g. [5]) alternately estimates the labeling and the motion parameters of the labels using least squares fitting as in [3].

Correct correspondence matches cannot be found for occlusions, i.e. areas that are freshly covered or uncovered by a moving object. Stiller [10] proposed to use the displacement field of the previous image frame. The estimation of a binary (dis)occlusion field has been done using the residual *motion-compensation error* as a marker [12].

Whenever pure motion information is not sufficient for robust segmentation, the integration of supplementary visual modalities is appealing. In most cases the multiple cues are combined in a sequential way (e.g. [1]), but this does not exploit the supplementary nature of the cues [6]. However, some work has been done that follows the concept of combining multiple cues in one cost function, for example by using edges [12] to enhance segment boundaries, by evaluating uniformity of both color and motion with adaptive weights [6]. In contrast to [7], who utilizes disparity information for occlusion detection, our system uses it directly for segmentation.

### 3 System Architecture

The algorithm presented in this paper is designed following the concept of Fig. 1. Motion, edges, color, and depth information are the visual cues used in this paper. For the actual segmentation algorithm we have chosen a Markov Random Field (MRF) framework as fundament. As done by other authors, we are alternating two different optimization steps during the iterative optimization for one image frame.

**Motion estimation:** Motion is estimated by rating correspondences, using the sum of squared differences metric  $SSD(x, y, \Delta x, \Delta y)$ . Classic motion estimation approaches select the displacement with the minimum SSD value from the SSD surface as the motion estimate and discard the rest. In many cases this is not appropriate, for example, if no or only low contrast is present at all or only in one direction in the image patch, also referred to as “aperture problem”. Instead, the whole SSD surface is used as input data for the optimization procedure during segmentation. This way, information from ambiguous matching surfaces is used in a sensible way.

To detect areas of occlusion where the image changes cannot be explained by motion, we use the minimum of an SSD surface as a measure for its validity: only if there is a low minimum in an SSD surface, the respective motion can account for the brightness changes in the correlation window at that particular location. If the minimum is higher than a certain threshold  $\tau_{\text{occl}}$ , the matching is assumed to be bogus. Such SSD surfaces are completely set to zero so they do not affect motion segmentation.

**Motion segmentation:** The segmentation of an image is specified by a labeling (configuration) of a Markov Random Field with one site (vertex in the Markov Random field graph)  $s(x, y)$  for each pixel. The value assigned to a site is its label  $l \in \{1, \dots, L\}$ . The number of labels  $L$  can change during optimization and from frame to frame. Each label  $l$  is associated with a displacement  $\Delta x_l, \Delta y_l$  and other optional features like a normalized color histogram  $c_l(h, s)$  of hue and saturation values or the average depth  $d_l$  of all sites labeled with  $l$ . Thus, sites with the same label are considered to show the same translational motion, and if using the optional descriptive features, similar coloring and depth. So, the framework bootstraps the object knowledge from the presented scene.

**Optimization of site configuration:** How well a label  $l$  fits the local image properties is expressed with the *fidelity* term  $FID$ . Good correspondence matches have a low SSD value, and similar colors have a high value in the histogram. The weights  $\alpha_{\text{col}} \geq 0$  and  $\alpha_{\text{dep}} \geq 0$  control the influence of the color cue and the depth cue respectively:

$$FID(x, y, l) = SSD(x, y, \Delta x_l, \Delta y_l) - \alpha_{\text{col}} \cdot c_l(H(x, y), S(x, y)) + \alpha_{\text{dep}} \cdot (d(x, y) - d_l)^2 \quad (1)$$

Furthermore, to introduce smoothness constraints, a regularization term (2) is formulated for each site  $s$  that assigns a penalty for each adjacent site  $s'$  in a 4-neighborhood that has a different label than  $s$ . We use the inverted Kronecker-Delta-function  $\bar{\delta}(a, b)$  which is 0 if  $a = b$  and 1 otherwise. The penalty for a neighboring site  $s'$  is reduced if a brightness edge goes through exactly one of the two sites in a pair  $(s, s')$  with image coordinates  $(x, y)$  and  $(x', y')$  respectively. With  $\mathcal{N}_4$  being the 4-connected neighborhood, the regularization is defined as follows:

$$REG(x, y, l) = \sum_{(x', y') \in \mathcal{N}_4} \bar{\delta}(l, s(x', y')) \cdot (1 - \alpha_{\text{edg}} \cdot \bar{\delta}(e(x, y), e(x', y'))) \quad (2)$$

The optimization of the labeling of the sites is done with greedy local optimization called *Iterated Conditional Modes* (ICM), which is identical to *Simulated Annealing* with a minimal temperature from the very beginning [2]. The update in each iteration is done for all sites in a random order, always selecting the label for a site which minimizes the sum of fidelity and regularity for that site.

**Optimization of labels:** The motion parameters of the labels as well as the color information have to be updated along with the configuration in each iteration step. For each label  $l$  the displacement values are chosen that belong to the minimum in the sum of all SSD surfaces of sites with the same label  $l$ . Labels that are not assigned to any site are deleted from the list. Labels with identical motion parameters are unified. This way the total number of labels  $L$  can decrease during one iteration.

For color representation, the normalized 2D ( $10 \times 10$ ) hue-saturation color histogram  $c_l(h, s)$  of a label  $l$  is computed by accumulating the hue and saturation values  $H(x, y)$  and  $S(x, y)$  from all sites with label  $l$ . The depth information  $d_l$  is represented as an average over all pixels with the same label  $l$ , i.e. pixels that belong to the same segment.

After this optimization step is done, the information describing the segmented object is represented by the label parameters. Hence, the system is able to segment unknown objects. Furthermore, objects with smoothly changing occurrence can be tracked.

**Initialization and motion propagation:** Besides a standard (background) label with  $\Delta x = \Delta y = 0$ , the initial labels are determined by searching for “good” motion estimates in the SSD surfaces. These have to meet all of the following criteria: (a) Motion needs to be clearly present, which is indicated by bad matching results for  $\Delta x = \Delta y = 0$ . (b) Occlusion areas are excluded, as defined for motion estimation. (c) The correspondence matches have to be unambiguous, which corresponds to a peaked minimum in an SSD surface.

When segmenting a sequence of images, the label parameters are carried over to the next frame. To account for dramatic changes or new occurrences of motion, new additional labels are included. The configuration is also carried over to the next frame as a prediction for the positions and shape of the segments: the motion estimates from the labels are applied to each site, yielding a linear motion model.

## 4 Experiments and Results

This section presents experiments to demonstrate the performance of our system. The evaluation is done on synthetic as well as on real world data. The synthetic data (Fig. 2a), showing two textured objects moving in front of cluttered background, is generated using POV-Ray<sup>2</sup>. The real world data contains standard MPEG evaluation sequences, as well as videos recorded with a VidereDesign STOC<sup>3</sup> stereo camera device. Due to missing

---

<sup>2</sup><http://www.povray.org>

<sup>3</sup><http://www.videredesign.com>

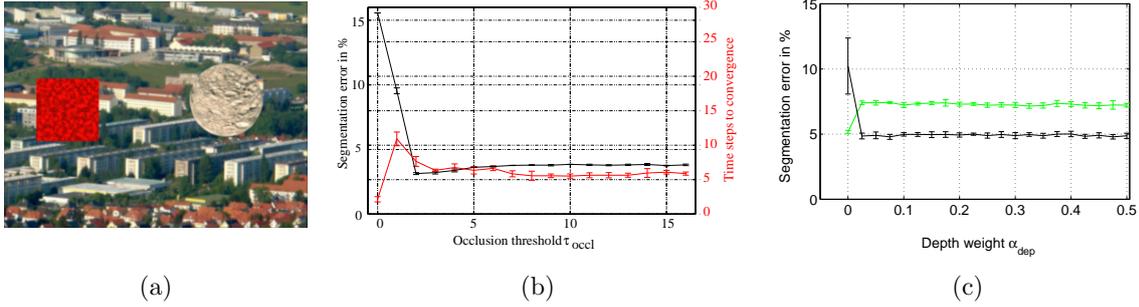


Figure 2: Experimental results using a synthetic test sequence (a) with two objects moving in front of cluttered background. The diagram (b) show the segmentation error (black) and the needed time steps (red) while changing the the occlusion threshold  $\tau_{occl}$ . The plot for the depth weight  $\alpha_{dep}$  (c) shows the results for different types of sequences (black and green) . Values are averaged over 10 trials with 10 frames each. The error bars show the standard deviation over the trials.

ground truth data the quantitative tests are performed only on the synthetic data set. Both segmentation error and number of iterations are evaluated by changing one of the system’s parameters while keeping the others fixed. The segmentation error is calculated taking the percentage of pixels which were labeled the wrong segment, according to [13]. Results considering real world images are not discussed within this paper, due to a lack of space. The SSD surfaces are obtained using a  $5 \times 5$  pixel correlation window, and a  $15 \times 15$  pixel search window. The parameter  $\alpha_{reg} = 0.5$  is determined by experimental evaluation.

In this experiment, segmentation performance is tested with different settings for the occlusion detection threshold  $\tau_{occl}$ . When occlusions are not treated in particular, false segmentation occurs at frontal boundaries of moving objects. The results of this experiment are presented and discussed in Fig. 2b. If the threshold  $\tau_{occl}$  is too high to be reached, the occlusion handling does not work. With adequate settings the false segments at the front border of the moving objects are suppressed. If  $\tau_{occl}$  is too low, good motion estimates are discarded and the segmentation deteriorates. Further experiments use  $\tau_{occl} = 4$ .

To analyze the influence of the depth cue, two different scenes are regarded. The first one is our standard POV-Ray generated one. In the second one, one of the objects slows down below a detectable velocity. The results can be seen in Fig. 2c. The green plot is evaluated with the sequence containing continuous movement, while the black one shows the results with absent movement. It can be seen in the black curve, that the segmentation result becomes better with an enabled depth cue  $\alpha_{dep} > 0$ . Our framework is able to compensate the absent motion with the help of the depth cue. The green plot shows a slight disimprovement with the stereo cue enabled. This is due to the fact, that the depth information we gain from the stereo camera device contains fuzzy borders around each

object. So, the position of the border caused by the motion differs from the one of the depth estimation, and the system is unable to find the right one.

## 5 Conclusion

We have presented a motion segmentation system that integrates additional cues like edges, color, and depth information in one optimization scheme. Correspondence-based motion information is incorporated with full SSD surfaces instead of best matches only. This way, data from evenly good matches in ambiguous cases is not discarded.

Results have been presented for segmentation on rendered scenes. We have shown the advantage of multi-cue segmentation over approaches with pure motion, and demonstrated the importance of occlusion handling. Our system converges in less than 30 iterations.

Despite the good results, our current system is limited in some ways. Motion is represented as local translational motion. The use of 2D affine or 3D motion models would improve the performance in cases where large rotations or changes of distance take place. Our general approach with its special characteristics however is suited for other motion models. The analysis of ways to handle the additional complexity in the optimization could also be part of future work.

## References

- [1] Yucel Altunbasak, P. Erhan Eren, and A. Murat Tekalp. Region-based parametric motion segmentation using color information. *Graphical Models and Image Processing*, 60(1):13–23, January 1998.
- [2] Julian Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(3):259–302, 1986.
- [3] Patrick Bouthemy and Edouard Francois. Motion segmentation and qualitative dynamic scene analysis from an image sequence. *Intl. Journal of Computer Vision*, 10(2):157–182, April 1993.
- [4] D. Cremers and S. Soatto. Motion competition: A variational framework for piecewise parametric motion segmentation. *IJCV*, 62(3):249–265, May 2005.
- [5] A. Jepson and M.J. Black. Mixture models for optical flow computation. In *Proc. of CVPR*, pages 760–761, New York, NY, USA, 1993.
- [6] S. Khan and M. Shah. Object based segmentation of video using color, motion and spatial information. In *Proc. of CVPR*, volume 2, pages II-746 – II-751, 2001.
- [7] Vladimir Kolmogorov, Antonio Criminisi, Andrew Blake, Geoffrey Cross, and Carsten Rother. Probabilistic fusion of stereo with color and contrast for bilayer segmentation. *IEEE Transactions on PAMI*, 28(9):1480–1492, September 2006.
- [8] Shang-Hong Lai and Baba C. Vemuri. Reliable and efficient computation of optical flow. *IJCV*, 29(2):87–105, 1998.
- [9] D.W. Murray and B.F. Buxton. Scene segmentation from visual motion using global optimization. *IEEE Transactions on PAMI*, 9(2):220–228, March 1987.
- [10] C. Stiller. Object-based estimation of dense motion fields. *IEEE Transactions on Image Processing*, 6(2):234–250, February 1997.
- [11] L. Torres, M. Kunt, and F. Pereira. Second generation video coding schemes and their role in mpeg-4. In *ECMAST*, pages 799–824, May 1996.
- [12] Jun Zhang and G.G. Hanauer. The application of mean field theory to image motion estimation. *IEEE Transactions on Image Processing*, 4(1):19–33, January 1995.
- [13] Y.J. Zhang. A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29(8):1335–1346, 1996.

### Author Information:

Dipl.-Inf. Sven Hellbach<sup>1</sup>

Dipl.-Inf Boris Lau<sup>1</sup>

Dr.-Ing. Julian P. Eggert<sup>2</sup>

Prof. Dr. Edgar Körner<sup>2</sup>

Prof. Dr. Horst-Michael Gross<sup>1</sup>

<sup>1</sup>Ilmenau Technical University, Neuroinformatics and Cognitive Robotics Lab, POB 10 05 65, 98684 Ilmenau

<sup>2</sup>Honda Research Institute Europe GmbH, Carl-Legien-Strasse 30, 63073 Offenbach/Main

Phone: +49 3677 69 1306

Fax: +49 3677 69 1665

E-mail: sven.hellbach@tu-ilmenau.de

Rozbeh R. Alavi / Klaus Brieß

## **Image Processing Algorithms for Using a Moon Camera as Secondary Sensor for a Satellite Attitude Control System**

### **Idea**

Today satellites play a major roll in our life. They are not only used for communication and navigation, they also have a share in better understanding our planet and the universe. The increasing complexity of scientific questions causes higher demands to the satellites. On this reason new technical solutions are needed, which are aligned to the satellite mission.

The moon is our immediate cosmic neighbour and it still arouses interest in the astronomers. To observe the moon by a satellite from an earth orbit is a task with increasing importance. For this kind of mission objective the satellite can use a camera, which takes pictures of the moon and sends them to a ground station. One of the subsystems of a satellite is the attitude control system. It provides for stability of the satellite on its orbit. For this job several sensors are necessary to determine the actual attitude of the satellite. The most well-known sensors are: gyroscopes, magnetic field sensors, star sensors, sun sensors and earth sensors [6].

This paper examines the use of a moon camera as a secondary sensor in the attitude control loop of a satellite. The moon circle can be recognized by image processing algorithms. Then the software calculates the deviation between the centre of the moon disk and the centre of the image sensor and return two angles to the attitude control system. The emphasis of this investigation is on the application of Hough transformation for finding the parameter of the moon circle.

### **Difficulties**

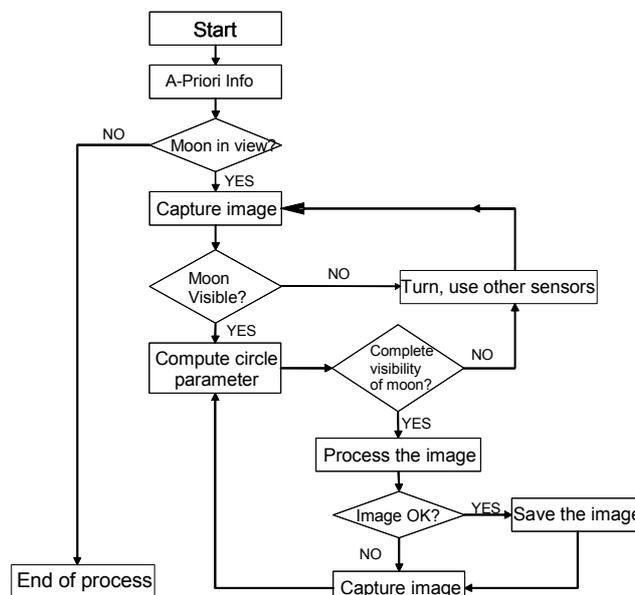
There are several difficulties to master, before the camera can be used as a sensor. The first problem is the visibility of the moon. How long the moon is visible for the satellite depends on the orbit parameters. Also before the camera can be applied as a sensor, the moon must be partially in its focus. This is one of the reasons, why the moon camera can not be used as a primary sensor. The satellite has to be brought into position with the help of other sensors before the camera is able to return reasonable values. The brightness of the moon is variable during a month. It also depends on the moon phases. For observation new moon represents a great challenge. The moon liberation also represent a difficulty for finding the parameters of the moon circle. Further difficulties are the large distance and the limited computer capacity.

## Operation principles

This chapter describes the operation principles of the moon camera as a sensor of the attitude control system.

The visibility of the moon depends on the orbit parameters. The orbit parameters have to be selected for maximum visibility of the moon. Also the brightness of the moon is changing by reason of moon phases and liberation. Consequently the demands to the camera, image processing algorithms and camera computer are rising. To return useful values, the camera software needs a-priori information. For example the software can draw conclusions on the brightness and diameter of the moon disk based on current date and the actual position of the satellite on its orbit. So the necessary computer capacity drops enormous.

The attitude control system of the satellite works only if it is activated. Otherwise it is switched off to save power. In this mode the satellite rotates arbitrarily on its axes. If the attitude control system is switched on again, it is possible, that the moon is not in the field of view of the camera. In this situation the satellite has to be rotated with help of the other sensors until the camera can see a fraction of the moon. Now the camera will be integrated into the control loop (Figure 1).



**Figure 1** attitude control loop with moon camera

On a low earth orbit the distance to the moon is 400,000 km. The optics of the moon camera has to reproduce an error free picture of the moon. It is also very difficult to construct such optics space-saving.

## Process model

This chapter describes the process model for the image processing of the moon camera as a secondary sensor of the attitude control system. Image processing applies in a lot of scientific fields such as medicine, astronomy, security and so on. It also can be used in a moon camera as image sensor.

An image is an information source. The information of an image is distributed in several layers. For using this information some mathematical algorithms have to be applied. For our application there are four types, which help to find the parameters - diameter and centre of the disk - of the moon circle. The figure 2 shows the stepwise processing of a moon image.

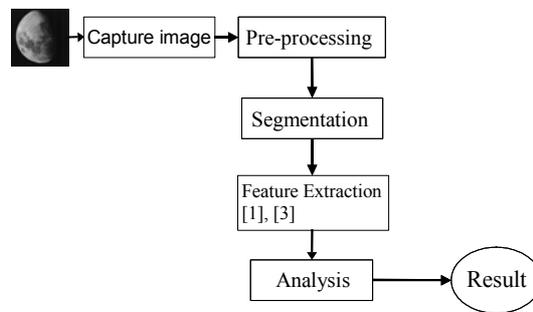


Figure 2 process model

### *Pre-processing*

The first step after capturing a picture is the pre-processing. In this phase of operation the information content of the image has to be reduced. This step is very important, because of the limited storage and processing resources of a satellite. Also the quality of the image can be improved by special filtering operations.

### *Segmentation*

Segmentation is the next step. In this phase interesting segments of an image have to be extracted. So it is possible to cluster the objects of an image. For solving this problem several filter operations also morphological operations are needed [1]. At this point the information content of the picture is reduced to the most important objects on it.

### *Feature Extraction*

After finding and clustering the objects of interest, their characteristics have to be determined. This phase of the process model is often very complex and needs more computer capacity. Hough transformation is used in our application to extract features.

## Analysis

In the last phase of the process model the computed feature parameters have to be analysed. In our application the actual attitude of the satellite can be computed based on the parameters of the moon circle.

## Implementation of the process model

This chapter describes the implementation of the process model. The intention here is to find the moon disk in a taken picture and calculate its parameter such as diameter and its centre. The actual attitude of the satellite can now be computed based on the deviation between the centre of the moon disk and the centre of the image sensor. The figure 3 shows the process model for the moon camera.

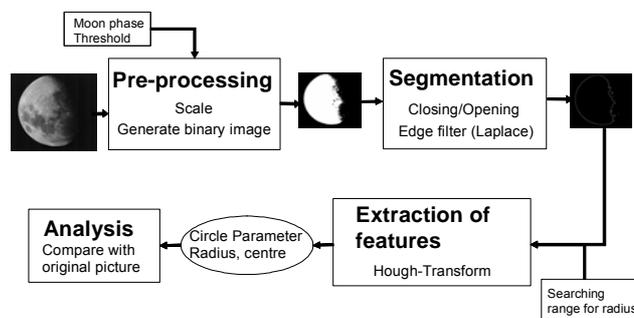


Figure 3 process model applied on the moon image

After capturing a picture the system begins the pre-processing. At first the picture can be scaled. This allows lossless reduction of relevant information, because the moon is a relatively homogenous structure. In this phase the system needs a-priori information, such as moon phase to define a threshold for generating a binary picture. The output of this operation step is a small binary picture of the moon.

In the next phase the moon disk has to be extracted. Dependent on the moon phase and on its brightness, there might be structures visible, which complicate the calculation of the disc parameters. This problem can be mastered by morphological operators, such as closing or opening. The result of this step is a homogenous moon area. The edge of this area can be detected by edge operators. Because the attitude of the camera is unknown, it is helpful to use an operator with second derivative, such as Laplace [2], [3]. Also the relative position especially the rotation between the moon and the sensor becomes neglectable.

$$\Delta^2 f(x, y) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

**Equation 1 Laplace matrix**

After the segmentation phase the information of the moon image is reduced to a circle or a circle fraction.

In the next step the Hough Transform will be used to find the circle parameters. P.V.C. Hough developed this Transform to detect straight lines in 1962 [4], [5]. During this transformation a discrete parameter space will be defined, named accumulator. Each point in the original image builds a curve in the parameter space. A representation of a straight line is given by:

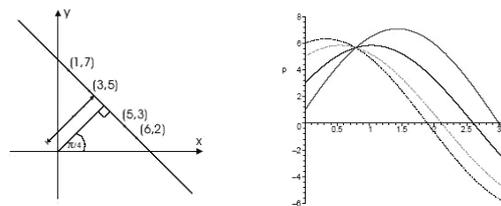
$$\vec{n} * (\vec{x} - \vec{a}) = \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} * \begin{pmatrix} x \\ y \end{pmatrix} - \rho \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix} = x \cos \theta + y \sin \theta - \rho = 0$$

$$\Rightarrow$$

$$\rho = x \cos \theta + y \sin \theta$$

**Equation 2 Parametric equation for a straight line**

Where  $\rho$  is the distance of the line from the origin and  $\theta$  is the angle between this perpendicular and the x-axis. Our parameter space is now in  $\rho$  and  $\theta$ .  $\theta$  can take values between 0 and  $2\pi$  and  $\rho$  is limited by the image size. For each point with  $(x, y)$  coordinates and for each angle the distance  $\rho$  will be calculated. The peak in the accumulator array defines the equation of the line. The figure 4 shows an example.



**Figure 4 Example for Hough Transform**

The Hough Transform works for other geometrical figures too. A representation of a circle is given by:

$$r^2 = (x - a)^2 + (y - b)^2$$

**Equation 3 Parametric equation of a circle**

Where  $r$  is the radius of the circle and  $a$  and  $b$  are the relative distance of the centre of the circle from the origin. As is evident from the equation the parameter space has three dimensions. This is the reason, why the computing time can be enormous. Well chosen pre-processing algorithms and segmentation drop the computing capacity and computing

time.

At last the finding parameters can be compared with the original image to control the results. If the calculated values are acceptable, the angular deviation of the satellite can be defined by these values and the optic parameters.

### Discussion

In this chapter the results of this application will be discussed. For testing the functionality of this method software was written. It grabs a picture from a CCD camera and applies the described process model on it. Technical specifications of the camera: Plunix TM-6EX, Sensor: 1/2" CCD, Resolution: 560(H) X 420(V), Pixel: 8.6 $\mu$ m X 8.3 $\mu$ m.

The accuracy of the process model was tested by using a picture of a circle. The table below represents the optic parameters. The accuracy depends on the camera resolution and size of the sensor pixels. Also the camera optics plays a major roll. With our test stand we could measure a deviation of 0.2°

| Parameters         |      |         |     |         |
|--------------------|------|---------|-----|---------|
| Pixel (HxV)        | 8.6  | $\mu$ m | 8.3 | $\mu$ m |
| Pixel average      | 8.4  | $\mu$ m |     |         |
| Focal distance f   | 12   | mm      |     |         |
| f-number           | 2.8  |         |     |         |
| Image distance     | 12.3 | mm      |     |         |
| Object distance    | 458  | mm      |     |         |
| Reproduction scale | 0.03 |         |     |         |

Table 1 Optic parameters

The process model was applied on a moon image too. The picture represents a nine days old moon. The image distance was 633mm. After grabbing the picture it was scaled with a factor of 1/6. After scaling a binary image was generated. For eliminating troublesome structure morphological operators - closing and opening - were applied on the picture. With help of Laplace operator the moon disc was reduced to its edge. After that the Hough Transform calculated the circle parameters. The figure 6 shows the operation steps.



Figure 5 Result of image processing steps

The cycle takes ca. 0.7s on an ordinary PC with 2.4GHz Intel Processor. The radius of the moon was detected with an accuracy of 3 pixels.

## References:

- [1] **Prof. Dr.-Ing. Wolfgang Förstner (03/04):** Ausgewählte Kapitel der Bildverarbeitung, Skriptum zur Vorlesung, Universität Bonn, Lehrstuhl für Informatik
- [2] **Carsten Köhn (1996):** Bildanalyse und Bilddatenkompression, Carl Hanser Verlag München Wien, 1996
- [3] **Prof. Bernd Radig (1993):** Verarbeiten und verstehen von Bildern, R. Oldenbourg Verlag München Wien, 1993
- [4] **Dipl.-Ing. Andre Trettin(2003):** Die Hough Transformation
- [5] **Sabrina Skibak (2002):** Proseminar Computervision: Geradenbestimmung mit der Hough-Transformation
- [6] **Udo Renner, Joachim Mauck, Nikolaos Balteas (1988):** Satellitentechnik, eine Einführung, Springer-Verlag Berlin, Heidelberg 1988
- [7] **Rolf Isernhagen (2000):** Softwaretechnik in C und C++, Carl Hanser Verlag München Wien, 2000
- [8] **Linos-Katalog(2004):** Optische Systeme, 2004, S. 80, Internetlink: [http://www.linos-katalog.de/pdf/de04/02\\_079-084\\_d04.pdf](http://www.linos-katalog.de/pdf/de04/02_079-084_d04.pdf)
- [9] **Intel(1997-2000):** Intel Image Processing Library Reference Manual, Veröffentlicht von Intel(1997-2000)
- [10] **Denton Woods (2002):** Developer's Image Library Manual (DevII), Abysmal Software(2002)
- [11] **o.V.(2004):** mvAcquireControl-SDK, Vision application development kit, <http://www.matrix-vision.de>, Matrix Vision GmbH
- [12] **o.V.(2000):** Dokumentation zum MV Tutorial, Matrix Vision GmbH
- [13] **Rozbeh R. Alavi (2004):** Konzept zur Mondbeobachtung mit Pico-Satelliten, Studienarbeit, 2004

## Authors:

Dipl. Ing Rozbeh R. Alavi

Prof. Dr. Ing. Klaus Briess

Institute of Aeronautics and Astronautics, Marchstr. 12

D-10587, Berlin

Phone: +49(030) 314-21305

Fax: +49(030) 314-21306

E-mail: [rozbeh.alavi@ilr.tu-berlin.de](mailto:rozbeh.alavi@ilr.tu-berlin.de)

[klaus.briess@ilr.tu-berlin.de](mailto:klaus.briess@ilr.tu-berlin.de)



S. Bauer / T. Döring / F. Meysel / R. Reulke

## Traffic Surveillance using Video Image Detection Systems

### ABSTRACT

The use of non-intrusive video-detection for traffic flow observation and surveillance is the primary alternative to the conventional inductive loop detector. A Video Image Detection System (VIDS) can derive traffic parameters by means of image processing and pattern recognition methods. Classic systems like inductive (double) loops or microwave radar detectors are able to measure presence and length of a vehicle as well as speed and time gap to the preceding vehicle. Existing VIDS emulate the inductive loop by virtual loops and derive the traffic parameters in a similar way. Additional benefits of the VIDS, e.g. large area observation of traffic flow or detection of special behaviors of single vehicles can not be exploited in this way. To expand the common approach for new types of information (queue length or erratic movements), we use a trajectory based recognition algorithm, which is related to the detection and tracking of individual vehicles in an image sequence. As an essential application for this algorithm it will be demonstrated how the usual characteristic traffic parameters can be derived from the measured trajectories. This facilitates a direct comparison between the video based and the classic inductive loop measurement of the traffic parameters.

### INTRODUCTION

An intelligent traffic management is based on exact knowledge of the traffic situation. Traffic monitoring of roads and intersections is therefore an essential prerequisite for the implementation of an Intelligent Transportation System (ITS). The most common detection and surveillance systems to measure traffic flow on public roads are inductive loops and microwave radar systems. The analysis and comparison of different sensors is found in [17]. In the last 15 years VIDS using real time image processing [1, 2, 3, 14, 15] became more attractive. Besides traditional parameters like presence, vehicle length, speed and time gap between two vehicles they can also determine congestion length, source-destination matrices, blockage or accidents and estimate travel times [6, 7, 8, 9]. This paper is organized as follows: After an overview of existing VIDS, the

approach is introduced. An example installation is then described and the results for this installation are presented. The article concludes with a summary and an outlook.

**COMMERCIAL VIDEO DETECTION SYSTEMS OVERVIEW**

There is already a variety of commercial VIDS available. Details of selected commercial VIDS can be found in Table1.

| <b>Supplier</b>  | <b>Product</b>          | <b>Detection Features</b>   |
|--|-------------------------|---|
| Traficon<br><br><a href="http://www.traficon.com/">http://www.traficon.com/</a>                      | VIP/P                   | vehicle count, queue length measurement   |
|  | VIP/D                   | vehicle count, speed, occupancy, classification, gap time, headway  |
|  | VIP/I                   | stopped vehicle, wrong-way drive, pedestrians, lost cargo, smoke / fire, queue, speed reduction                             |
| Image Sensing Systems<br><br><a href="http://www.imagesensing.com/">http://www.imagesensing.com/</a> | Autoscope<br>RackVision | vehicle count, vehicle length classification, speed, occupancy time, gap time, congestion, accident, wrong way drive        |
| Aimetic<br><br><a href="http://www.aimetis.com/">http://www.aimetis.com/</a>                         | AIRA 2005               | automatic detection and tracking, object classification, automatic pan / tilt / zoom, unauthorized movement in target zones |

Table 1: Commercial VIDS for traffic detection and surveillance

These systems usually provide reliable traffic parameters according to the specifications but also lack the versatility for changing requirements on the surveillance system. For instance, limitations from measurement principles in the spectral or spatial range can be compensated, if complementary sensor systems are combined [10]. Examples are the combination of sensors for visible and infrared radiation or the use of several cameras with different views to cope with occlusions caused by buildings, traffic signs, trees, or cars [16]. Another trend is to combine the traffic object data with image information and geo-coded object description to get a user-friendly traffic characterization with GIS tools [18].

**APPROACH**

The approach presented here can be separated into four steps (Figure 1). Firstly, all moving objects have to be extracted from each frame of the surveillance video sequence. Then these traffic objects have to be projected onto a geo-referenced world plane, tracked in the sequence and thus will be associated with trajectories. Next, those have to be evaluated to compute the predefined traffic parameters. The four steps shall be described more precisely in the following.

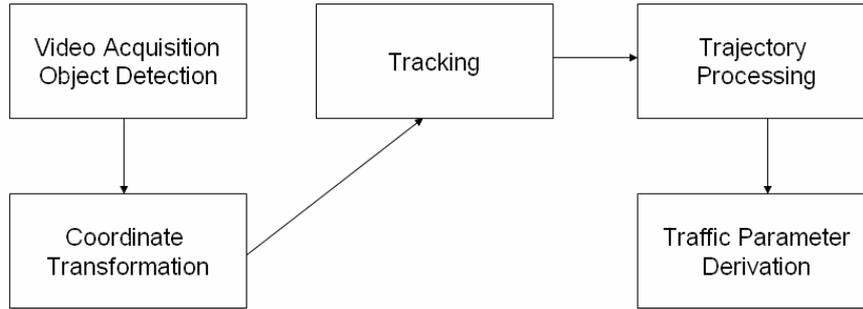


Figure 1: Process chain

### 1. Video Acquisition and Object Detection

Only digital camera systems are considered. To receive a reliable and reproducible result, only compact industry cameras are used. In addition, only widespread interfaces and protocols (e.g. IEEE1394, Ethernet) are used for the data transmission.

In order to extract the traffic objects from an image sequence, different image processing libraries or programs (e.g. OpenCV or HALCON) can be utilized. Here the development environment HALCON was used. HALCON provides easy access to different image sources and a large software library for standard computer vision tasks. An already implemented Kalman based background estimator adapts to the variable background and generates a difference image with the searched traffic objects. The extracted objects (Figure 2) in the image are then grouped using a cluster analysis combined with additional filters that incorporate scene knowledge. Such knowledge can be the maximum or minimum traffic object size or the possible position of occurrence. Typical errors of such an approach are e.g. ghosts and shadows [11,12,13].

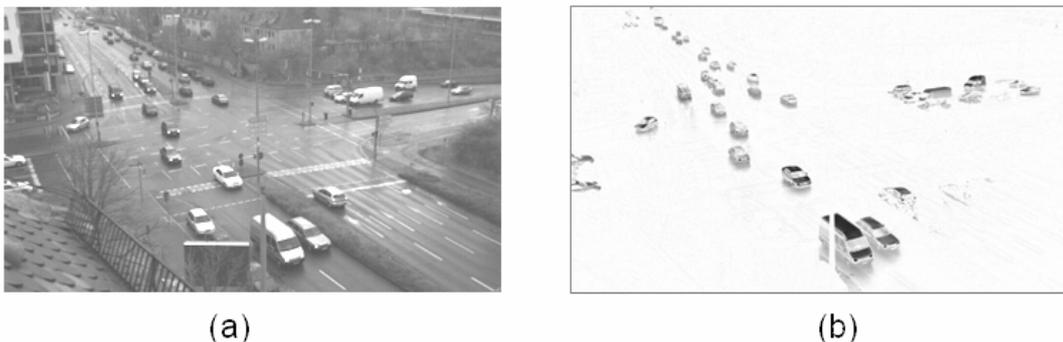


Figure 2: (a) Grabbed Image (b) Extracted objects

### 2. Coordinate Transformation and Camera Calibration

The existing tracking concept is based on extracted objects, which are geo-referenced to a world coordinate system. Therefore, a transformation between coordinates of a camera-image to the world coordinate system is necessary. Using collinearity equations the world coordinates can be derived from image coordinates. Additional necessary

input parameters are interior and exterior orientation of the camera. The interior orientation (principal point, focal length and additional camera distortion) can be determined by using a well known lab test field. The 10 parameter Brown camera model [19] was used for describing interior orientation. Calculating the exterior orientation of a camera in a well known world coordinate system is based on previously measured ground control points (GCPs) with differential GPS. The accuracy of the points is within the range of less than 5 cm. With these coordinates an approximate orientation can be deduced using DLT [20]. Then the exterior orientation is calculated with the spatial resection algorithm. The lateral error in X- and Y-direction achieved by this approach is 20 cm in 100m distance from projection centre.

### **3. Tracking and Trajectory Creation**

The tracking algorithm is supposed to provide object data information combined in a so-called state vector with respect to time. The state of an object can be described with position, velocity and acceleration in X-, Y- and Z-direction. Features like form, size and color can be added. The first task is the object identification in a video sequence by its predicted state vector. This is done by observation-object-association [14,15]. The tracking of a single object was realized with a Kalman-filter. It estimates the state of an object for the time stamp of the following picture, thus allowing the estimated state and the observed object data to be compared. If both are within a certain feature space distance they can be associated to the same object. An important problem is the initialization of the Kalman-filter. The trajectories are then submitted to the analysis module as soon as they are created for the derivation of traffic parameters (TP).

### **4. Traffic Parameters Derivation and Data Structures**

Incoming trajectories are processed and used for the computation of traffic parameters by a separated module in this approach. The procedure for handling the trajectories is essentially organized in two nested loops, an inner loop for the analysis of each submitted trajectory, and an outer loop that aggregates and updates traffic parameters for a time cycle with the analysis results of each trajectory. In the inner loop the trajectories are compared to detector structures, defined in world coordinates. Here these structures are line detectors or area detectors, placed in the scene. For each trajectory, the set of detector structures is tested for trajectory-detector interaction and the results are saved. The testing is done for each trajectory by interpolating the object positions between discrete moments and finding intersection of detectors and trajectory

interpolation between pairs of points in space. For a positive test, the detector stores the interpolated spatial coordinates of the intersection and the interpolated moment of activation. It will hold this data until it is queried from within the inner loop. At the end of the inner loop, detector activations are queried and sorted by time. For multiple activations, trajectories are detected, that passed multiple detectors. After a certain time interval has elapsed, the outer loop reads out the advanced traffic parameter data structures. Also all detectors' long-time memories are queried and interpreted for the derivation of integrated cycle parameters.

## FIRST RESULTS

The described approach has been implemented and tested on a traffic intersection in Nurnberg, Germany. While the observation conditions were often not agreeable, the object detection still provided enough objects for the processing chain. The coordinate transformation, multi-object-tracking and trajectory creation worked together on a designated PC. Trajectories have been sent to remote computer for the analysis and computation of traffic parameters. The results are promising. Incoming trajectories were evaluated and traffic parameters computed. The trajectories could be visualized in real-time (Figure 3 left) and the current situation could be described by means of the derived traffic parameters. The update cycle for advanced parameters was chosen as one minute. In each interval, activation counts combined with speed information and new source destination matrices have been filled and evaluated. For example, Figure 3 (right) shows a comparison of computed and manually obtained vehicle counts.

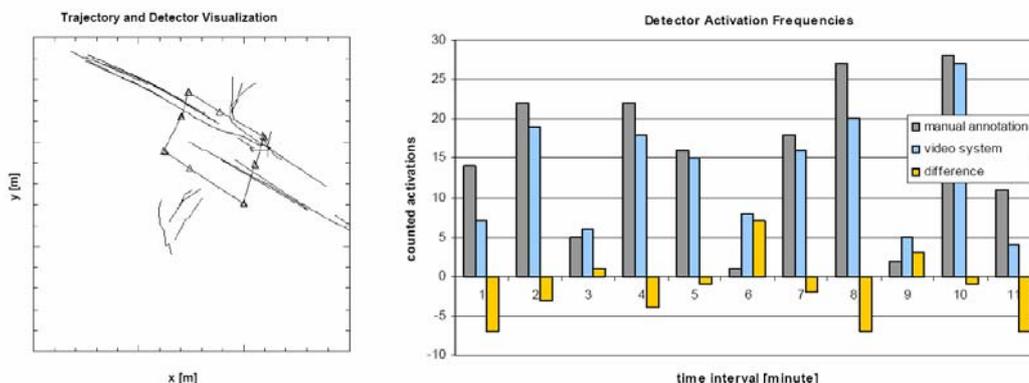


Figure 3: (left) example trajectories in world coordinate system, (right) example count results of one detector

These first results are already very successful. Nevertheless, due to occlusions caused by the observation position the precision is not convincing.

## CONCLUSION AND OUTLOOK

The presented approach for a traffic surveillance system has been implemented and tested. Thus, it could be shown that standard traffic parameters can be derived based on video detection, tracking and trajectory analysis. This step is essential on the way to future traffic surveillance systems. Still, the robustness of the processing chain needs to be improved. Accumulated data is not very much affected by small errors. However, detection errors and tracking problems can deteriorate the trajectory data and lead to less usable trajectories for analysis or less reliable traffic parameters. Methods to recognize object detection errors and deteriorated trajectories to stitch them together correctly as well as deriving new traffic parameters are key factors for current and future work.

### References:

- [1] Michalopoulos, P.G., (1991). Vehicle Detection Video through Image-Processing - the Autoscope System, *IEEE Transactions on Vehicular Technology*, vol. 40(1), pp. 21-29.
- [2] Wigan, M.R., (1992). Image-Processing Techniques Applied to Road Problems, *Journal of Transportation Engineering-Asce*, vol. 118(1), pp. 62-83.
- [3] Kastrinaki, V., Zervakis, M., and Kalaitzakis, K., (2003). A survey of video processing techniques for traffic applications, *Image and Vision Computing*, vol. 21(4), pp. 359-381.
- [4] Kan, W.Y., Krogmeier, J.V., and Doerschuk, P.C., (1996). Model-based vehicle tracking from image sequences with an application to road surveillance, *Optical Engineering*, vol. 35(6), pp. 1723-1729.
- [5] Morita, T., (1999). Tracking vision system for real-time motion analysis, *Advanced Robotics*, vol. 12(6), pp. 609-617.
- [6] Datta, T.K., Schattler, K., and Datta, S., (2000). Red light violations and crashes at urban intersections, *Highway and Traffic Safety: Engineering, Evaluation, and Enforcement; Trucking and Motorcycles*, (1734), pp. 52-58.
- [7] Harlow, C. and Wang, Y., (2001). Automated accident detection system, *Highway Safety: Modeling, Analysis, Management, Statistical Methods, and Crash Location*, (1746), pp. 90-93.
- [8] Setchell, C. and Dagless, E.L., (2001). Vision-based road-traffic monitoring sensor, *See Proceedings-Vision Image and Signal Processing*, vol. 148(1), pp. 78-84.
- [9] Yung, N.H.C. and Lai, A.H.S., (2001). An effective video analysis method for detecting red light runners, *IEEE Transactions on Vehicular Technology*, vol. 50(4), pp. 1074-1084.
- [10] Harlow, C. and Wang, Y., (2002). Acoustic accident detection system, *Its Journal*, vol. 7(1), pp. 43-56.
- [11] Cucchiara, R., Grana, C., Piccardi, M., and Prati, A., (2003). Detecting moving objects, ghosts, and shadows in video streams, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25(10), pp. 1337-1342.
- [12] Prati, A., Mikic, I., Trivedi, M.M., and Cucchiara, R., (2003). Detecting moving shadows: Algorithms and evaluation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25(7), pp. 918-923.
- [13] Cho, J.H., Kwon, T.G., Jang, D.G., and Hwang, C.S., (2005). Moving cast shadow detection and removal for visual traffic surveillance, *Ai 2005: Advances in Artificial Intelligence*, vol. 3809, pp. 746-755.
- [14] Kumar, P., Ranganath, S., Huang, W.M., and Sengupta, K., (2005). Framework for real-time behavior interpretation from traffic video, *IEEE Transactions on Intelligent Transportation Systems*, vol. 6(1), pp. 43-53.
- [15] Luo, X.Z. and Bhandarkar, S.M., (2005). Real-time and robust background updating for video surveillance and monitoring, *Image Analysis and Recognition*, vol. 3656, pp. 1226-1233.
- [16] Reulke, R., Boerner, A., Hetzheim, H., Schischmanow, A., and Venus, H. (2002). A sensor web for road-traffic observation. *Image and Vision Computing, New Zealand 2002*, pp. 293-298.
- [17] Klein, L.A., M.R. Kelley, and M.K. Mills, Evaluation of overhead and in-ground vehicle detector technologies for traffic flow measurement. *Journal of Testing and Evaluation*, 1997. 25(2): p. 205-214.
- [18] Ernst, I., et al., New approaches for real time traffic data acquisition with airborne systems, in *TRB Transportation Research Board. 2005: Washington*.
- [19] Light, D.L., The New Camera Calibration System at the Us-Geological-Survey. *Photogrammetric Engineering and Remote Sensing*, 1992. 58(2): p. 185-188.
- [20] Grussenmeyer, P. and O. Al Khalil, Solutions for exterior orientation in photogrammetry: A review. *Photogrammetric Record*, 2002. 17(100): p. 615-634.

### Authors:

Sascha Bauer

Thomas Döring

Frederik Meysel

Prof. Dr. Ralf Reulke

Deutsches Zentrum für Luft- und Raumfahrt, Institut für Verkehrsführung und Fahrzeugsteuerung  
Rutherfordstraße 2, 12489 Berlin, Phone: +49 30 67055 (409, 232, 285, 518), Fax: +49 30 67055 283  
E-mail: (Sascha.Bauer, Thomas.Döring, Frederik.Meysel, Ralf.Reulke) at dlr.de

M. A-Megeed Salem / B. Meffert

# Wavelet-based Image Segmentation for Traffic Monitoring Systems

## ABSTRACT

*Multiresolution representation of images has gathered significant attention in recent years. In part this was due to its representation of the information encoded in an image in different scales and the simplification of the later processing. However, it has not been widely used for segmentation of time sequence images. In this work we introduce different variants of a multiresolution algorithm based on the 3D wavelet transform to extract the regions of moving objects. The 3D wavelet transform gives the advantage of considering the relevant spatial as well as temporal information of the movement.*

## 1 INTRODUCTION

Since the appearance of the wavelet transform before two decades it has been used in many applications in signal processing. An early use of the wavelet was in image compression [2], feature extraction for the purpose of image indexing [7], and for medical image segmentation [5]. The use in video processing was in video coding and compression as well as in video transition [4]. However, its use for segmentation of frames in a video is scarce.

Image segmentation in traffic monitoring systems is the detection and isolation of the moving objects that take part in the current traffic situation. A reliable and efficient segmentation is needed to meet the requirements of the subsequent tracking and interpretation tasks. Because it has to work on a great volume of data, the guarantee of a high-speed response is a must for real time applications. In the case of traffic monitoring image segmentation is applied to a sequence of images that encode the traffic parameters and temporal information during a definite time period. The robust segmentation algorithm must utilize such information.

Toreyin *et al.* [8] proposed to use the 2D wavelet analysis to extract the region of interest (ROI). The work is based on a background subtraction, where generally

a mask is created from the difference between the current traffic scene image and the estimated background image. The background is estimated based on the wavelet transform coefficients at the 3<sup>rd</sup> level. This work was chosen as a reference to compare our results with.

## 2 3D WAVELET-BASED SEGMENTATION

### 2.1 3D Wavelet Transform

The fast wavelet transform as introduced by Mallat [3] uses separable orthogonal basis functions, and therefore the multidimensional transform can be decomposed into a tensor product of orthogonal sub-spaces. The 3D scaling function and the 3D wavelet functions can be each expressed as a product of three one-dimensional functions. The smallest input to the 3D wavelet analysis is an octal data element which may have a cubic shape. Eight coefficients are the result. One coefficient contains the approximation  $a$  of the data in the input cube and other 7 detail coefficients  $d^i, i = 1..7$ . For the extraction of the ROI two detail coefficients are relevant, namely  $d^4$  and  $d^7$ . The sub-band  $d^7$  is computed by applying the wavelet function  $\psi$  on all the three axes  $x, y$  and  $z$ -axis. So it contains information about the possible changes along all the axes. The other sub-band  $d^4$  is computed by applying the scaling function  $\phi$  on the  $x$  and  $y$  axes and the wavelet  $\psi$  to the  $z$ -axis. It represents the change in time between the approximated spatial domains. In [6] we have proposed a 3D wavelet-based algorithm for the detection of moving objects in a traffic surveillance video. Moving object detection is based on extracting a mask for a group of frames. This mask represents the ROI in this group. The number of frames in a group depends on the level of the wavelet analysis. Previously we have used only the Haar wavelet. In this work we study the use of other Daubechies mother wavelets and we introduce an enhancement on the proposed algorithm in order to adapt the generated masks to the scene and the aimed application.

### 2.2 The Proposed Algorithm

The proposed algorithm consists of three parts. The first and main part is the analysis of the input image sequence by the 3D wavelet transform. As explained before,  $d^7$  contains a great part of the motion information. But the extracted regions show only the borders of the motion. On the other hand  $d^4$  shows the area where the movement occurred clearly but noisy. The combination of both sub-bands gave the best results for this purpose. The output of this step is a primary segmentation. The aim of the

second part of the algorithm is to create a binary mask. This is done by thresholding the output of the wavelet analysis followed by a smoothing step using the median filter. For the following region growing step a morphological dilation is used. This part of the algorithm can be considered as a segmentation improving step. The masks are then used to extract the area of the active traffic for each frame of the sequence.

### 2.3 Testing Different Mother Wavelets

In this section we introduce the use of different mother wavelets among the family of Daubechies Wavelets. Daubechies discovered the first wavelet family of scale functions that are orthogonal and have compact support [1].

Haar wavelet  $\psi_{haar}$ , also known as the 1<sup>st</sup> Daubechies wavelet, is the basis of the simplest wavelet transforms. The associated filter is of length two. This means that the resulting approximation and details are all eighth the number of columns, rows and slices. For higher order Daubechies wavelets  $\psi_{dbN}$  the  $N$  denotes the order of the wavelet. The support length of  $\psi_{dbN}$  and  $\phi_{dbN}$  is  $2N - 1$  and the length of the associated filter is double as the number of the vanishing moments, *i.e.*  $2N$ . If  $n = length(s)$  where  $s$  is the given signal, then the approximation and detail coefficients are of length  $floor(\frac{n-1}{2}) + N$ . Because of the length of the filters the analysis is done overlapping with a stepwise shift. The detection of an event by higher order wavelets takes longer than that by lower order. Due to the overlapping the sharp edges are detected as wide events. This affects the segmentation negative by producing enlarged moving area. Results of  $\psi_{haar}$ ,  $\psi_{db4}$  and  $\psi_{db8}$  are presented here.

### 2.4 Masks From Different Resolutions

To improve the results of the proposed algorithm, the generated masks from the 3 different levels are combined together. Five different combination strings are tested. The 1<sup>st</sup> combination string assigns a pixel to a ROI if the corresponding pixel in the 1<sup>st</sup> level's mask belongs to the ROI and if any of the corresponding pixels in the 2<sup>nd</sup> or 3<sup>rd</sup> levels' masks belongs to the ROI too. This string gives the results of the 1<sup>st</sup> level of the analysis a dominant rule in the combination with the other levels. The 2<sup>nd</sup> and 3<sup>rd</sup> combination strings give the results of the 2<sup>nd</sup> and 3<sup>rd</sup> levels the dominant rules, respectively. The 4<sup>th</sup> combination string combines simply all the masks by an or-operator. The 5<sup>th</sup> combination string uses the and-operator for all masks.

### 3 EXPERIMENTAL RESULTS

Many data sets were captured using a stationary video camera. The data sets can be categorized into three groups: 1) a front view with small camera observation angle to a street, 2) varying lighting conditions 3) wide view of a crossing, with many types of traffic. The results have a lower resolution than the original images, depending on the algorithm and the tested data set. In all cases a projection of the segmented images to a higher resolution image is done after the running of the algorithms. So, the extracted ROI is shown in the original resolution of the input image.

For statistics and a better visualization the results are given in form of bounding boxes. The error can be measured as 1) ratio of the boxes that contain no objects related to the total number of extracted boxes (false alarm). 2) number of objects that are not contained in any box (missed objects) and 3) number of objects that are delayed to be detected related to the total number of objects in the sequence.

Table 1 shows the results at each level for the 3D wavelet-based algorithms in terms of extracted boxes against the results of the 2D wavelet-based segmentation proposed by [8]. The main features of the results of [8]: The algorithm is sensitive to changing in light conditions, was not able to detect a new object as soon as it appears, it needs much time for a stable estimation of the background, and tends to miss more objects for a scene with slow motion. Figure 1(a) shows a sample of the results.

For some data set the results of the 3D wavelet-based algorithm show perfect detection of the objects from the first appearance. Generally, the extracted ROI is larger than the object if the level of the analysis increases. Therefore, the error measures decrease as the level of the analysis increases. However, the results are much better than those obtained by the 2D wavelet-based algorithm specially for the problem of the movement of the trees in the background as shown in Figure 1(b).

Two data sets were used to test the performance of the different mother wavelets. This is done because of the stable lighting conditions, the homogeneity of the background, and the simplicity of the scenes in this category. As displayed in Table 2 the

Table 1: Results of all methods

| Method        | False alarms | Bounding boxes | Missed objects | Delay detections |
|---------------|--------------|----------------|----------------|------------------|
| 2D Seg        | 521 43,5%    | 1198           | 23,5%          | 7,9%             |
| 3D Seg Msk1   | 27 3,8%      | 711            | 10,3%          | 5,9%             |
| 3D Seg Msk2   | 153 17,2%    | 890            | 3,0%           | 0,3%             |
| 3D Seg Msk3   | 185 24,3%    | 762            | 2,8%           | 1,4%             |
| CombMsk1      | 49 7,1%      | 690            | 10,7%          | 7,2%             |
| CombMsk2      | 112 16,5%    | 679            | 5,9%           | 3,5%             |
| CombMsk3      | 112 18,0%    | 622            | 9,5%           | 6,2%             |
| CombMsk (OR)  | 242 34,7%    | 697            | 0,6%           | 0,4%             |
| CombMsk (AND) | 40 6,5%      | 614            | 17,0%          | 13,3%            |



Figure 1: Results of the 2D, 3D wavelet-based algorithm and the proposed enhancements

number of the extracted bounding boxes as well as the number of false alarms are increased dramatically. We refer that to the increase of the length of the filters and the overlapped analysis. Moreover, there were delayed detected objects by the results of DB4 3<sup>rd</sup> level for both test sets. Computationally, as the length of the filters increases, the complexity of the analysis also increases.

The highest rate of false alarms of 35% is found to be generated by the results of the 4<sup>th</sup> combination string. However, it gives the best results in terms of missed objects and delayed detection. The main problem with this combination scheme is that the extracted ROI is too big considering the moving objects themselves as shown in Figure 1(c). In contrast are the results of the 5<sup>th</sup> combination string. The false alarm rate is very low but the missed objects and delayed detections score very high.

The results of the other combination strings are found to enhance the results obtained by single level analysis. They are all better in terms of false alarms compared to the results of the 2<sup>nd</sup> and 3<sup>rd</sup> levels of the 3D wavelet-based algorithm. Figure 1(d) shows the same frame as in Figure 1(b), one false alarm was removed.

Table 2: Results of DB1, DB4 and DB8 for the first and second data sets in terms of extracted boxes

| No. | Method   | Bounding boxes | False alarms | Missed objects | Delay detections |
|-----|----------|----------------|--------------|----------------|------------------|
| 1   | DB1-Lvl1 | 14             | 2 14,3%      | 0 0,0%         | 0 0,0%           |
|     | DB1-Lvl2 | 20             | 5 25,0%      | 0 0,0%         | 0 0,0%           |
|     | DB4-Lvl1 | 24             | 11 45,8%     | 0 0,0%         | 0 0,0%           |
|     | DB4-Lvl2 | 36             | 25 69,4%     | 2 15,4%        | 2 15,4%          |
|     | DB8-Lvl1 | 22             | 9 40,9%      | 0 0,0%         | 0 0,0%           |
|     | DB8-Lvl2 | 24             | 11 45,8%     | 0 0,0%         | 0 0,0%           |
| 2   | DB1-Lvl1 | 12             | 1 8,3%       | 0 0,0%         | 0 0,0%           |
|     | DB1-Lvl2 | 16             | 5 31,3%      | 0 0,0%         | 0 0,0%           |
|     | DB1-Lvl3 | 16             | 5 31,3%      | 0 0,0%         | 0 0,0%           |
|     | DB4-Lvl1 | 22             | 10 45,5%     | 0 0,0%         | 0 0,0%           |
|     | DB4-Lvl2 | 36             | 24 66,7%     | 0 0,0%         | 0 0,0%           |
|     | DB4-Lvl3 | 32             | 24 75,0%     | 3 27,3%        | 3 27,3%          |
|     | DB8-Lvl1 | 22             | 11 50,0%     | 0 0,0%         | 0 0,0%           |
|     | DB8-Lvl2 | 28             | 19 67,9%     | 0 0,0%         | 0 0,0%           |
|     | DB8-Lvl3 | 32             | 21 65,6%     | 0 0,0%         | 0 0,0%           |

## 4 DISCUSSIONS AND CONCLUSION

The 3D wavelet-based algorithm is more reliable to detect objects entering the scene than the 2D wavelet-based algorithm. The detection of the moving objects takes place from the 1<sup>st</sup> group of frames. In contrast to the conventional background update algorithms, the proposed algorithm needs no time to adapt itself to the image sequence. Because of the longer observation interval it can be supposed that the use of lower resolution levels helps in the detection of objects that stop for a short time and then move again.

The use of Haar wavelet gave better results than those of DB4 and DB8. It is computationally very simple and can be implemented efficiently on hardware.

Generally the results obtained from the 2<sup>nd</sup> combination string gave the best compromise considering all error measures.

A main advantage of the multiresolution algorithms is that the processing is done in a lower spatial resolution than that of the input images. The size of the low resolution images is much smaller and the computational complexity is reduced. With a simple projection or mapping step the results return to the resolution of the input images.

### References:

- [1] Ingrid Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, 1992.
- [2] Gerald Kaiser. The fast haar transform, gateway to wavelets. *IEEE potentials*, 1998.
- [3] S. G. Mallat. A theory for multiresolution signal decomposition, the wavelet representation. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2(7):674–693, 1989.
- [4] E. Moyano, F.J. Quiles, A. Garrido, J. Duato, and L. Orozco-Barbosa. Efficient 3d wavelet transform decomposition for video compression. In *DCV '01: Proceedings of the Second International Workshop on Digital and Computational Video*, page 118, Los Alamitos, CA, USA, 2001. IEEE Computer Society.
- [5] Mohammed Saeed, Hamid R. Rabiee, W.C. Karl, and T.Q. Nguyen. A new multi-resolution algorithm for image segmentation. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP98*, volume 5, pages 2753 – 2756. IEEE Computer Society, May 12-15 1998.
- [6] Mohammed A-Megeed Salem and Beate Meffert. A comparison between 2d and 3d wavelet based segmentation for traffic monitoring systems. In *3rd International conference on Intelligent Computing and Information Systems*, Cairo, Egypt, March 15-18 2007. Faculty of Computer and Information Sciences.
- [7] Mohamed A. Tahoun, Khaled A. Nagaty, Taha I. El-Arief, and Mohammed A-Megeed Salem. A robust content-based image retrieval system using multiple features representations. In *IEEE International Conference on Networking, Sensing and Control(ICNS'05)*, Arizona, USA, March 19-22 2005. IEEE Computer Society.
- [8] B. Ugur Toreyin, A. Enis Cetin, Anil Aksay, and M. Bilgay Akhan. Moving object detection in wavelet compressed video. *Signal Processing: Image Communication*, 20:255–264, March 2005.

### Authors:

M.Sc. Mohammed A-Megeed Salem

Prof. Dr. Beate Meffert

Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099 Berlin

Institute of Computer Science, signal processing and pattern recognition group

E-mail: {salem,meffert}@informatik.hu-berlin.de

E. Einhorn / C. Schröter / H.-J. Böhme / H.-M. Gross

# A Hybrid Kalman Filter Based Algorithm for Real-time Visual Obstacle Detection

## 1 Introduction and Related Work

Obstacle detection and collision avoidance must be considered important capabilities of mobile robots. Vision-based approaches provide a large field of view and supply a large amount of information about the structure of the local surroundings.

In this paper we present a sparse feature-based “shape-from-motion” approach for mobile robots which is applicable for collision avoidance, online map building and scene reconstruction in real-time. Our method processes a sequence of images which are taken by a single camera mounted on a mobile robot. In contrast to similar monocular shape-from-motion algorithms we combine two different approaches: A traditional motion stereo approach and a Kalman filter based algorithm for scene reconstruction. We show that the disadvantages of the traditional stereo approach are compensated by the Kalman filter and vice versa. Our special method of initializing the Kalman filter leads to a faster convergence compared to other plain Kalman based approaches. Moreover, we present a feature matching algorithm which is faster and more reliable than the widely-used KLT-Tracker in the domain of scene reconstruction. These image features are extracted using the “FAST” high-speed corner detector [8].

As we intend to use the reconstructed scene for obstacle detection and collision avoidance our camera is mounted in front of the mobile robot and tilted towards the ground. This results in two major problems we have to deal with:

**1. The camera is moving along its optical axis.** In a sensitivity analysis Matthies and Kanade [7] proved that when using forward motion shape-from-motion leads to higher uncertainties in the depth estimates. Compared to the ideal lateral camera translation parallel to the image plane - which is used in standard binocular approaches - the estimation must be applied over a long base distance in order to achieve the same accuracy.

**2. Many objects are visible during a few frames of the captured image sequence only,** while the robot is approaching these obstacles. Hence, most image features cannot be tracked over a large number of frames and the scene reconstruction algorithm must be able to provide a reliable estimate by using a few image measurements only.

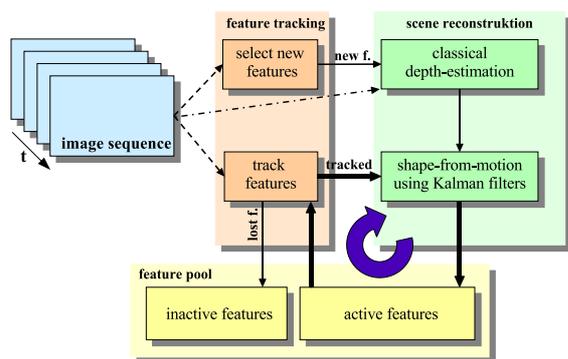
To overcome the first problem Matthies et al. [7] suggest scene reconstruction using Kalman filters, since they can integrate the depth and scene information over a long base distance. Consequently, many shape-from-motion solutions that have been researched and published in recent years are based on Kalman filtering [10, 5, 1, 7]. Since Kalman filter based methods

solve the reconstruction problem in an iterative manner the speed of convergence depends on the choice of the initial estimate which is used for the initialization of the Kalman filter. If unfavourable initial estimates are used, Kalman filter based approaches tend to suffer from a low speed of convergence, i.e. several iterations must be processed to get a reliable estimation of the obstacle positions. Unfortunately, as stated above, most image features cannot be tracked over many frames and it is not possible to compute enough iterations. To prevent this problem our hybrid algorithm combines two completely different approaches for scene reconstruction: A Kalman filter based method for scene reconstruction and a “classical” correlation-based depth estimation approach.

The depth estimation algorithm is used to compute a reliable initial estimate for the Kalman filter, which then will refine the estimate and recover the three-dimensional model. We show that this novel kind of initialization leads to a significantly better convergence of the filter.

## 2 Scene-Reconstruction

Since depth estimation and scene reconstruction using Kalman filtering are common techniques in computer vision they will not be described in detail here. Further information can be found in [3, 10, 5, 1, 9].



The figure on the left illustrates the complete architecture of our algorithm. Our motion stereo approach is inspired by the publication of Bunschoten and Kröse [3], where a multi-baseline depth estimation algorithm for panoramic image data is presented. Based on their work we have developed a similar correlation-based algorithm for projective cameras.

To obtain the image data we work with a single projective camera mounted on our mobile robot to capture not only a sequence of images - each taken from a different pose (i.e. position and orientation) during the robot’s locomotion - but also the corresponding odometry data measured by the robot drive. Hence, for each image of the sequence the approximate position of the camera is known, including uncertainty in odometry measurements from systematic and non-systematic errors.

To correct these errors we use correspondences our feature tracker has found over the frames of the image sequence to estimate the pose of the camera. Since the translation vector of the camera movement can only be computed up to a scale, we are content with estimating the angle of roll and the pitch of the camera, since inaccuracies in the orientation of the camera cause the largest error in the scene reconstruction. Starting with values provided from the robots odometry both angles are varied using Gauss-Newton iteration in order to minimize the Sampson error, which is defined by the used image point correspondences and

the fundamental matrix.

The estimated depth is then used to compute the approximate 3D position of the feature in the real scene. This position is used as a reliable initial estimate for the Kalman filtering, which then will refine the estimate and recover the three-dimensional model. In contrast to [5, 1] where one single Kalman filter with a large state vector is used to recover the 3D-positions of all features (model points), we use a separate filter for each feature point. According to [10] this leads to a linear space and time complexity in terms of the number of features while the loss in accuracy is small. Similar to [10] we choose the 3D position of the feature point as state vector  $\mathbf{X} \in \mathbb{R}^3$  which is to be estimated. Using this estimated 3D position of each scene point an a-priori estimate of its image position can be computed by projecting it onto the image surface of the camera. The observed measurement is the position of the real image point in the current image which is provided by a feature tracker, that tracks each image point over consecutive frames. With each new frame the tracked features will pass through this Kalman filter cycle and their 3D positions will be estimated more precisely in each iteration.

### 3 Feature Tracking

In order to track the image features over several frames, we apply a feature matching algorithm. First we select the image features independently in each frame using the FAST corner detector [8]. Similar to the IPAN feature tracker [4] corresponding features are matched in subsequent frames then. While the IPAN tracker solves a pure motion correspondence problem by using three consecutive frames and solely kinematic constraints, we only use two frames. To eliminate the resulting ambiguities we additionally take the image similarity into account.

Let  $I_{t-1}$  and  $I_t$  be two consecutive frames of the image sequence. In order to find the correspondences  $\mathbf{x}_{t-1}^{(i)} \leftrightarrow \mathbf{x}_t^{(i)}$  between the previously selected image features of both frames, possible hypotheses of matching points are chosen first. Each hypothesis  $h = (\mathbf{x}_{t-1}^{(i)}, \mathbf{x}_t^{(j)})$  consists of a pair of two potentially matching points  $\mathbf{x}_{t-1}^{(i)}$  and  $\mathbf{x}_t^{(j)}$  of the frames  $I_{t-1}$  and  $I_t$ . To reduce the number of hypotheses we use a maximum speed constraint, i.e. we only choose pairs of image points that satisfy  $\left\| \mathbf{x}_{t-1}^{(i)} - \mathbf{x}_t^{(j)} \right\|_2 \leq r_{max}$ , where  $r_{max}$  defines the maximum speed, at which a point can cross the frame within the image sequence.

For each hypothesis  $(\mathbf{x}_{t-1}^{(i)}, \mathbf{x}_t^{(j)})$  we compute a cost function which is defined by the following weighted sum:  $\text{cost}(\mathbf{x}_{t-1}^{(i)}, \mathbf{x}_t^{(j)}) = w_1 c_1 + w_2 c_2 + w_3 c_3$ , where  $c_1 = \left\| \mathbf{x}_{t-1}^{(i)} - \mathbf{x}_t^{(j)} \right\|_2^2$  is the squared euclidean distance between the image point  $\mathbf{x}_{t-1}^{(i)}$  and the predicted feature position  $\mathbf{x}^{*(i)}$ . The latter can be computed as  $\tilde{\mathbf{x}}^{*(i)} = \mathbf{P}_t \tilde{\mathbf{X}}_t^*$  using the reconstructed 3D position  $\tilde{\mathbf{X}}_t^*$  of the feature which was estimated so far and the corresponding camera projection matrix  $\mathbf{P}_t$ , which is computed from the corrected odometry data of the robot. Since we perform an initial depth estimation as described in the previous section, an estimate of the 3D position is already

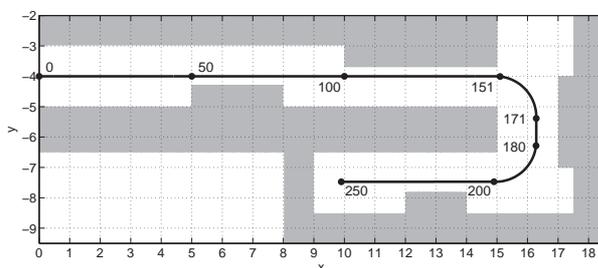
available for newly selected features. For features that have been tracked over several frames more accurate estimations of the 3D positions were computed by the Kalman filters and their location in the current frame can be predicted more precisely.

Additionally, corresponding image points must satisfy the epipolar constraint, hence an image point  $\mathbf{x}_t^{(j)}$  that corresponds to  $x_{t-1}^{(i)}$  is located on or near the epipolar line that is induced by  $x_{t-1}^{(i)}$ . The distance of the image point  $\mathbf{x}_t^{(j)}$  from that epipolar line can be computed as follows:  $c_2 = \frac{|\bar{\mathbf{x}}_t^{(j)\top} \mathbf{F} \bar{\mathbf{x}}_{t-1}^{(i)}|}{\sqrt{(\mathbf{F} \bar{\mathbf{x}}_{t-1}^{(i)})_1^2 + (\mathbf{F} \bar{\mathbf{x}}_{t-1}^{(i)})_2^2}}$ , where  $\mathbf{F}$  is the corresponding fundamental matrix which again is computed using the robot's odometry. Alternatively, the Sampson distance [6] could be used which is, however, computationally more complex.

As stated above we also use a similarity constraint to eliminate ambiguous matchings. For each pair of potentially matching points  $\mathbf{x}_{t-1}^{(i)}$  and  $\mathbf{x}_t^{(j)}$  we compute the similarity of their neighborhood patterns. Again we use the SAD as measure of correlation:  $c_3 = \text{SAD}_W(\mathbf{x}_{t-1}^{(i)}, \mathbf{x}_t^{(j)})$ . The weights of the above cost function must be chosen empirically. We use  $w_1 = 1$ ,  $w_2 = 3$  and  $w_3 = 20$ . From all hypotheses those with minimal matching costs are chosen by a greedy algorithm. Hypotheses whose costs are larger than a certain threshold are rejected. An appropriate threshold depends on the image data. Finally, all chosen hypotheses represent the corresponding image points.

## 4 Results

In order to make a quantitative analysis and to be able to compare our hybrid approach with others we have rendered a sequence of a synthetic scene consisting of 250 frames and their corresponding ground truth depth images using the raytracer POV-Ray<sup>1</sup>. The ground truth depth images are used to measure the tracking error and the error of the reconstructed 3D model. We use realistic textures and add some gaussian image noise. To simulate odometry errors and the sway of the camera we add gaussian noise to the camera position and orientation while rendering the images.



The figure on the left shows a top view of the synthetic scene where the camera trajectory is plotted and its position at certain frames is marked. In figure 1 and 2 the guided feature matching algorithm we have proposed in this paper and Birchfield's implementation of the

KLT feature tracker [2] are compared. Because of the guided matching, our greedy feature linking algorithm has a smaller tracking error. Although the tracking error of the KLT tracker can also be reduced if guided tracking is used and the tracker is provided with the predicted feature locations as described in the previous section, the runtime of the KLT

<sup>1</sup><http://www.povray.org/>

tracker remains a problem for realtime applications. Using the proposed feature matching algorithm we were able to reduce the runtime that is needed for feature tracking dramatically as shown in the second diagram of figure 2. Using 200 features per image the hybrid scene reconstruction and the feature linking can be computed in just 20 ms per frame. Hence we can process up to 50 frames per second.

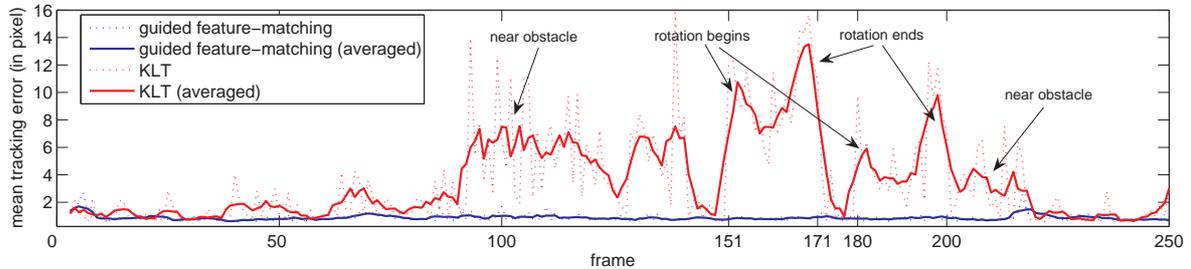


Figure 1: Mean tracking error for each frame of the synthetic image sequence. Due to the large optical flow while the camera is rotating and approaching near obstacles the tracking error of the KLT tracker becomes larger while with guided matching it remains small.

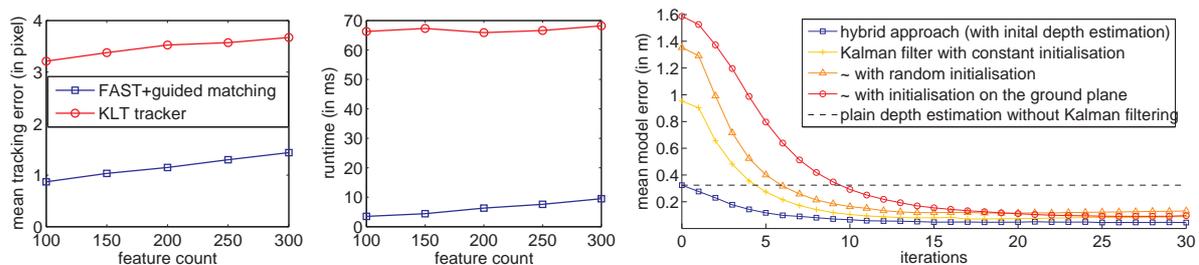
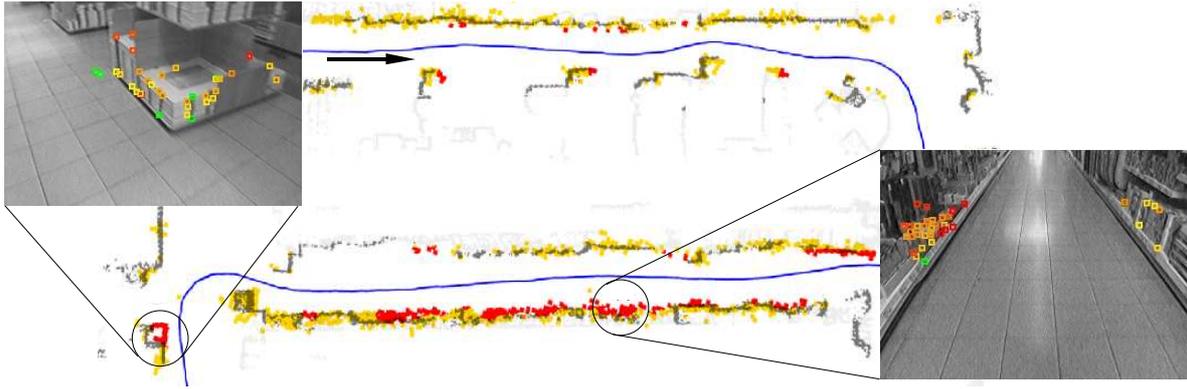


Figure 2: **left:** Mean tracking error averaged over all frames of the synthetic sequence for different feature counts. **middle:** Runtime that is needed for feature selection and feature tracking depending on the number of features that are selected in each frame. The time was measured on a Pentium 4 with 3.4 GHz. **right:** comparison of different methods for initializing the Kalman filters.

In the right diagram of figure 2 it can easily be seen that the hybrid algorithm presented in this paper converges faster than plain Kalman filter based approaches which use simple heuristics for choosing the initial estimates. Less iterations and therefore less images are necessary to obtain a reliable scene reconstruction.

The above figure shows a map which was created while the robot was moving through a real indoor environment. The estimated positions of the features are visualized using red and orange dots. The estimated z-coordinate is used only to determine if a point belongs to an obstacle or if it lies on the floor, i.e. if the z-coordinate of a point is less than a certain threshold of 0.1 m it is regarded as belonging to the ground plane and not included in the map. The gray map in the background was built using a laser range finder and is used as reference. The accuracy of the map which was built using our approach is similar to the laser-built reference map. Moreover, our visual method is able to detect some obstacles which are not “visible” to the laser because they are too small and lie beneath the laser



range finder. Those obstacles were labeled manually and are highlighted by the red color. Additionally, the corresponding camera images are shown for two of these obstacles. It can easily be seen that one part of the left obstacle is not included in the laser map, since it is too small and located below the laser plane. This would have led to a collision if solely laser based navigation had been used. Using our hybrid approach for visual obstacle detection instead, this obstacle can be detected very well.

### References

- [1] A. J. Azarbayejani, T. Galyean, B. Horowitz, and A. Pentland. Recursive estimation of CAD model recovery. In *Proceedings of the 2nd CAD-Based Vision Workshop*, Champion, PA, 1994.
- [2] Stan Birchfield. Derivation of kanade-lucas-tomasi tracking equation. January 20 1997.
- [3] R. Bunschoten and B. Kröse. Robust scene reconstruction from an omnidirectional vision system. *IEEE Transactions on Robotics and Automation*, 19(2):351–357, 2003.
- [4] Dmitry Chetverikov and Judit Veresty. Tracking feature points: a new algorithm. In *Proc. of 14th International Conference on Pattern Recognition*, pages 1436–1438, Brisbane, Australia, 1998.
- [5] A. Chiuso, P. Favaro, H. Jin, and S. Soatto. Structure from motion causally integrated over time. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 4, Apr. 2002.
- [6] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0-521-54051-8, second edition, 2006.
- [7] Larry Matthies, Takeo Kanade, and Richard Szeliski. Kalman filter-based algorithms for estimating depth from image sequences. *International Journal of Computer Vision*, 3(3):209–238, 1989.
- [8] Edward Rosten and Tom Drummond. Machine learning for high-speed corner detection. In *European Conference on Computer Vision*, volume 1, pages 430–443, May 2006.
- [9] Greg Welch and Gary Bishop. An introduction to the kalman filter. Technical report, Chapel Hill, NC, USA, 1995.
- [10] Y. Yu, K. Wong, and M. Chang. Recursive 3d model reconstruction based on kalman filtering. *SMC-B*, 2004.

### Author Information:

Erik Einhorn, Dipl. Inf. Christof Schröter, Dr. Hans-Joachim Böhme, Dr. Horst-Michael Groß  
 Neuroinformatics and Cognitive Robotics Lab,  
 Faculty of Computer Science and Automation  
 Ilmenau Technical University,  
 POB 10 05 65, 98694 Ilmenau  
 Tel: +49 3677 69 1306  
 E-mail: Erik.Einhorn@t-online.de

U. Knauer / R. Stein / B. Meffert

## Detection of Opened Brood Cells at an Early Stage

### ABSTRACT

We present an algorithm for an early detection of the uncapping process of honeybee brood cells. It is a challenging problem, because the appearance of openings varies highly and the combs surface is crowded by bees.

The approach consists of two steps, segmentation of the image and analyzing the segments to give a report about openings.

### INTRODUCTION

One of the biggest threats for the native honeybee *Apis mellifera* is the mite *Varroa destructor* [1]. A promising approach to block the mites is the rearing of resistant bees. Therefore, current research in the field of apiculture focuses on the genetic selection of hygienic bees [2]. The selection of hygienic bees requires a time consuming observation of the combs. Processing all the material that is typically recorded for a period of one week (24 hours a day) requires at least twice the time for analysis by a human expert. Details on this procedure and proposals for acceleration by image processing techniques can be found in [3,4].

To summarize previous work, the identification of hygienic bees is always guided by the search for brood cells which have been uncapped by such bees. Only if openings can be detected as early as possible (see Fig. 1) the identification of hygienic bees will be improved. In this paper we present an algorithm to achieve a reliable early detection in cluttered images.

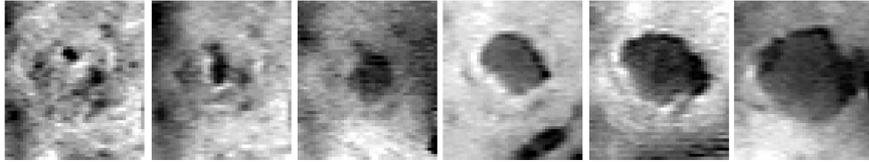


Fig. 1. Contrast-enhanced images of a brood cell. Early detection (state at the most left picture) is required.

The paper is structured as follows. First, we describe the preprocessing and segmentation of images. Next, a classifier for the detection of openings and first results are presented. Finally, we show the limitations of the approach and outline future work.

### PREPROCESSING OF IMAGES

Fig. 2 gives an impression how the recordings from the observational beehive look like. A two step prescreening is applied to extract non-occluded images of each brood cell (i.e. as shown in Fig.1) for analysis. A survey on the performance of different algorithms on this problem and our final choice of the classifiers are provided in [4].



Fig. 2. Typical image of the comb. In our approach bees are treated as clutter which makes an early detection of uncapping difficult.

The preprocessing step results in sets of images which show the surface of the brood cell. Each set contains images of just a single brood cell. As an option the images can be transformed to have an equal mean gray value. This kind of normalization improves the evaluation of the image sets by a human expert because it helps to focus on more relevant changes of the surface.

## SEGMENTATION OF IMAGES

Ideally, at this point of processing the image sets only consist of images without bees, but typically, a couple of such images passes the mentioned prescreening step. Therefore, extraction and evaluation of potential openings must be performed. We decided to rely on two features of the small defects on the cap of a cell:

- dark and compact appearance compared to their neighborhoods and
- their fixed positions.

We assume, that the first feature allows the segmentation into foreground (the opening) and background regions (the rest). The second feature allows to exclude isolated bee eyes or extremities which are typical sources of error. A contrast-limited adaptive histogram equalization is applied to improve the first feature.

In [5], Balthasar et al. proposed an efficient technique for analyzing of 1D histograms. Originally used for the extraction of typical color values, this technique can also be applied for the required adaptive segmentation of images. The method is based upon the search for peak values (local maxima) in the distribution. Additionally, the label of the nearest peak value is assigned to each histogram bin. This method can be used for an adaptive binarization of cell images as follows:

- 
- FIND the number  $n$  of local maxima  $M_i$  of the gray value histogram
  - DIVIDE the histogram into segments  $S_i$ , one for each maximum  $M_i$
  - SORT list of segments by  $M_i$
  - INITIALIZE a binary image  $B$  of same size as the original image (all pixels  $B_{x,y}$  are set to zero)
  - FOR EACH segment
    - GENERATE a binary image mask  $K$  for all pixels of the original image that have a gray value which belongs to the segment
    - IF the number of non zero pixels in  $K$  is lower than a threshold  $T$ 
      - Update  $B$ , such that  $B_{x,y} = \begin{cases} K_{x,y}, & \text{if } K_{x,y} = 1 \\ B_{x,y}, & \text{if } K_{x,y} = 0 \end{cases}$
    - ELSE
      - BREAK (skip all remaining segments)
-

This procedure excludes all homogenous regions larger than a certain threshold  $T$ . Fig. 3 shows the segmentation results for a single image.

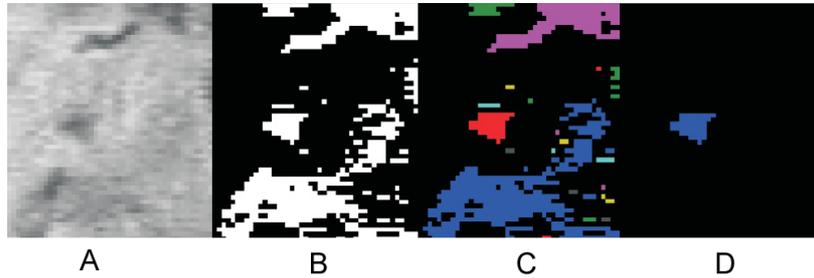


Fig. 3. (A) original image, (B) foreground image, (C) labeled connected components, (D) classified as opening

Binarization of the images in terms of background-foreground estimation can also be based on models of the appearance of the brood cells. So called *intrinsic images* [6,7] seems to be an adequate and robust method. Other techniques like Gaussian mixture models can also be applied [8]. A disadvantage of these approaches is the need for a sufficient large number of images to create a reliable model.

A sequence of masks can be combined to increase the confidence, because an opening persists for a certain time until the brood cell is closed again or has been fully uncapped. Hence we recursively calculate an average image  $\bar{B}$ . The parameter  $\alpha$  controls the update rate.

$$\bar{B}_t = (1 - \alpha) \cdot \bar{B}_{t-1} + \alpha \cdot B_t$$

For binarization of the image  $\bar{B}$  a fixed threshold is recommended. The threshold value depends on  $\alpha$  and the difference  $\Delta t = t_i - (t_{i-1})$ . Due to the nature of prescreening (which eliminates occlusion, see [4])  $\Delta t$  is not constant.

Fig. 4 shows the input images, the segmentation results, and the evolution of the average image.

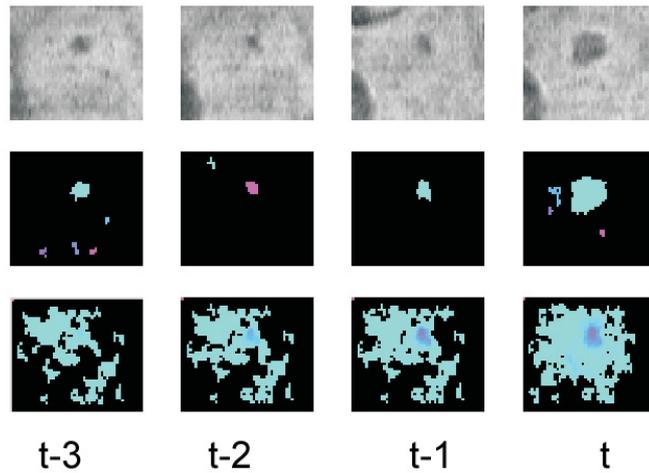


Fig. 4. Sequence of images (top row), extracted regions (middle row), color coded average mask  $\bar{B}$  (bottom row)

### CANDIDATE SELECTION

An evaluation of the candidate regions is required to reduce or eliminate false alerts. A set of criteria has been checked for 1400 images of openings (taken from 28 cells at different times) and approximately 142,000 images with an intact cell cap.

Among the criteria are:

- Minimum distance from image boundary
- Size of region
- Different measures for the compactness of a region
- The ratio between the update rate  $\alpha$  and the threshold for binarization of  $\bar{B}$

If an image contains a connected component which meets the criteria, the detection of an opening is reported. We minimized a cost function in order to find suitable parameters for all criteria. The cost function was defined as the weighted number of misclassifications. A false negative classification was penalized with a cost factor of 10,000 to compensate the higher number of negative examples and to emphasize the need to detect the openings as early as possible. A false positive classification is penalized with a cost of only 1. By varying the parameters systematically within reasonable intervals we obtained a set of optimal parameters for the defined cost functions.

With this approach 61 % of the openings are reported. If we count the reported brood cells then the detection rate increases to 79 %. Reducing the cost of a false negative classification results in a more sensitive classifier. On the other hand, this would lead to a further loss of specificity.

## CONCLUSIONS

We have presented a robust method for the segmentation of openings of honeybee brood cells. The segmentation can be improved by averaging.

The described approach for the automatic detection of (small) openings marks still a beginning. The sensitivity of the detector is already acceptable to improve the process for genetic selection of bees. However, the number of false alerts exceeds the true ones by factor 36. Hence future work must concentrate on more specific classifiers. Despite this limitation current experiments at the *Länderinstitut für Bienenkunde Hohen-Neuendorf* already benefit from the detector. The presented method complements the generation and analysis of reports by further reducing the number of irrelevant images.

### References:

- [1] Jong, D. D.: Mites: Varroa and other parasites of brood. In R.A. Morse and R. Nowogrodski, editors, *Honeybee Pests*, pages 200-218. Cornell University Press, Ithaca, 1990.
- [2] Bienefeld, K. and Arnold, G.: Studies on the genetic determination of uncapping of varroa-infested brood cells. In First European Conference of Apidology, EurBee, pages 103-104, 2004.
- [3] Knauer, U., Himmelsbach, M., Winkler, F., Zautke, F., Bienefeld, K., and Meffert, B.: Application of an adaptive background model for monitoring honeybees. In IASTED International Conference on Visualization & Image Processing, pages 46-50, 2005.
- [4] Knauer, U., Bienefeld, K., and Meffert, B.: A Comparison of Classifiers for Prescreening of Honeybee Brood Cells. Proc. International Conference on Computer Vision Systems, ICVS, 2007.
- [5] Balthasar, D., Rehrmann, V.: Robustes histogrammbasiertes Farbmatching. In Farbbildverarbeitungsworkshop, FarbBV, 1999.
- [6] Weiss, Y.: Deriving intrinsic images from image sequences. In International Conference on Computer Vision, ICCV, 2001.
- [7] Porikli, F.: Multiplicative Background-Foreground Estimation Under Uncontrolled Illumination using Intrinsic Images. In Workshop on Motion and Video Computing, WACV/MOTION, Vol. 2, pages 20-27, 2005.
- [8] Stauffer, C. and Grimson, W. E. L.: Adaptive background mixture models for real-time tracking. In Computer Vision and Pattern Recognition, CVPR, 1998.

### Authors:

Dipl.-Inf. Uwe Knauer

Robert Stein

Prof. Beate Meffert

Humboldt-Universität zu Berlin, Institut für Informatik, Unter den Linden 6, D-10099 Berlin

E-mail: {knauer, rstein, meffert}@informatik.hu-berlin.de





K. Ghanem, N. Zamin-Khan, M. A. A. Kalil and A. Mitschele-Thiel

## **Dynamic Reconfiguration for Distributing the Traffic Load in the Mobile Networks**

### **ABSTRACT**

The widespread deployment of mobile networks drives the necessity for a novel dynamic scenario to balance the network load. Various load balancing means have been proposed to manage the traffic load of mobile networks. Building a dynamic architecture can be an optimal solution to create networks which are economic, flexible with high performance. This work suggests a new scenario (Multilink's stations scenario) based upon the idea of redistributing the peak traffic load between network elements to enhance the network performance, reconfigure the busy traffic paths and reduce the infrastructure and operation costs. Multilink's stations scenario which uses the multiple paths between Node-Bs and RNCs can be applied as a good solution to balance the load in the mobile network also it can be used to decrease the on-peak load and allows new users to join the network. The presented scenario offers a flexible solution to balance the load in the mobile networks also it can be used to decrease the on-peak load and allows new users to join the network.

### **INTRODUCTION**

Mobile networks are the most common of all public wireless communication systems. Over the last decade the mobile communication networks have become increasingly widespread worldwide. The migration of telecommunication towards the wireless networks has become a significant phenomenon in the recent communications. Mobile networks such as GPRS, UMTS and HSDPA could provide the user demands [1] [2]. However, the increasing demands on the multimedia and data services in the recent mobile networks create a significant challenge. This challenge for the network designers is how to balance the requirements of the user and improve the operated network [3] [4]. One of the basic principles is to reuse radio resource after a certain distance. The whole area is divided up into a number of small areas called cells, with one base station giving radio coverage for each cell by its associated antenna [5][6].

Mobile operators have designed the radio access networks to cover traffic demand of the planned services in a static approach, considering the busy hour traffic in each geographic zone. That means the operator builds base stations depending on the requirements of the traffic forecasted in each area. It is worth to mention that not all the network spots have the same traffic load. The traffic load depends upon the area and it differs from rural and urban area [7]. Moreover, the peak time load appears within some delimited periods when the network resources are heavily used. So increasing the network capacity is not necessary at any time, that is because of some periods the network are empty and others are overloaded. The idea is to redistribute the usage of the existed network resources.

A significant need for a novel dynamic scenario that provides a flexible and a cheap solution to improve overall utilization and performance is required. Dynamic reconfiguration scenario for distributing the traffic load can be a mean to control the demand of the mobile network resources. Reconfigurability creates many challenges, but also provides opportunities, ranging from the scale growth to harmonization and inter-working among radio access technologies both is in the radio networks and core network. Reconfigurability offers the flexibility to adapt the settings to the actual demands, however to determine the implement the most efficient system configuration, mechanisms for advanced planning and management are required. This paper deals with these issues. An overview and related work, network traffic load, load balancing using cell breathing and Multilink's Station's Scenario are presented in the subsequent sections. Conclusion and future work have presented later on at the end of this paper.

## **OVERVIEW AND RELATED WORK**

The increased demand for network services has led to substantial investments in the network infrastructure. The radio frequency band allocated to GSM and UMTS is limited, the obvious solution to increase available capacity is through cell splitting and frequency re-use, resulting in smaller cell clusters. Hence, urban areas now have far more cells per square kilometer than rural ones [8] [3]. The trend towards Pico-cells and nano-cells to serve a one building can increase the costs of the operating and infrastructure [9].

Previous researches on mobile cellular networks have led to many schemes to balance the traffic load [11]. Traffic load balancing in mobile cellular network has been well-studied since the first generation of mobile communication systems. Many methods have been proposed to-address this problem, such as cell splitting [12], channel borrowing [10] [4], channel sharing [13], dynamical channel allocation [8] [13], new soft handover schemes [12]. However, most of the research related to traffic load balancing only focuses on different radio channel allocation schemes.

The current solutions in the mobile networks don't offer any means for the stations to switch between the RNCs. In this work, a dynamic reconfiguration for distributing the traffic load in the mobile networks is achieved by using multilink's stations. These stations can switch from loaded RNC to unloaded one and enables more users to use the network in order to reduce the peak load. Studies on dynamic sectorization [14], and use tilted antennas [13] have shown that the system performance can be improved. However, switching the links between the node\_Bs and the RNCs depending on the RNC load has not studied well so far. This work looks at a method that can utilise the existing network infrastructure efficiently. A decision has to be made of which connection should be while the network has multiple paths [15]. Real measurements can help the operator to decide which Node\_Bs should provide with multilinks. Our scenario can be applied after having such these measurements and can be applied to existing networks.

As different wireless network technologies such as 3G mobile networks have been deployed at increasing rates, inter-working of these various technologies has become a significant issue. When multiple access networks are available it is possible to perform a seamless handovers between these different access technologies. Moreover, IEEE 802.21 standard Media Independent Handover (MIH) supports the network to have a seamless handovers across heterogeneous networks. Our mechanism/scenario focuses on switching between multi-links' Node\_Bs, which is based on the network load.

## NETWORK TRAFFIC LOAD

As the demand for cellular service increases within a particular area, the network must be re-engineered and more base-stations must be installed to meet the demand, which can be costly and time consuming. However, the dynamic nature of traffic capacity demand makes it difficult for the current cellular networks to operate efficiently and to optimize both cost and quality of service. Analysis of traffic performance implies capturing the three-way relation between capacity, demand and performance. Analysis of real network traffic can help the operator to determine the optimal solution for the best utilization of the network resources.

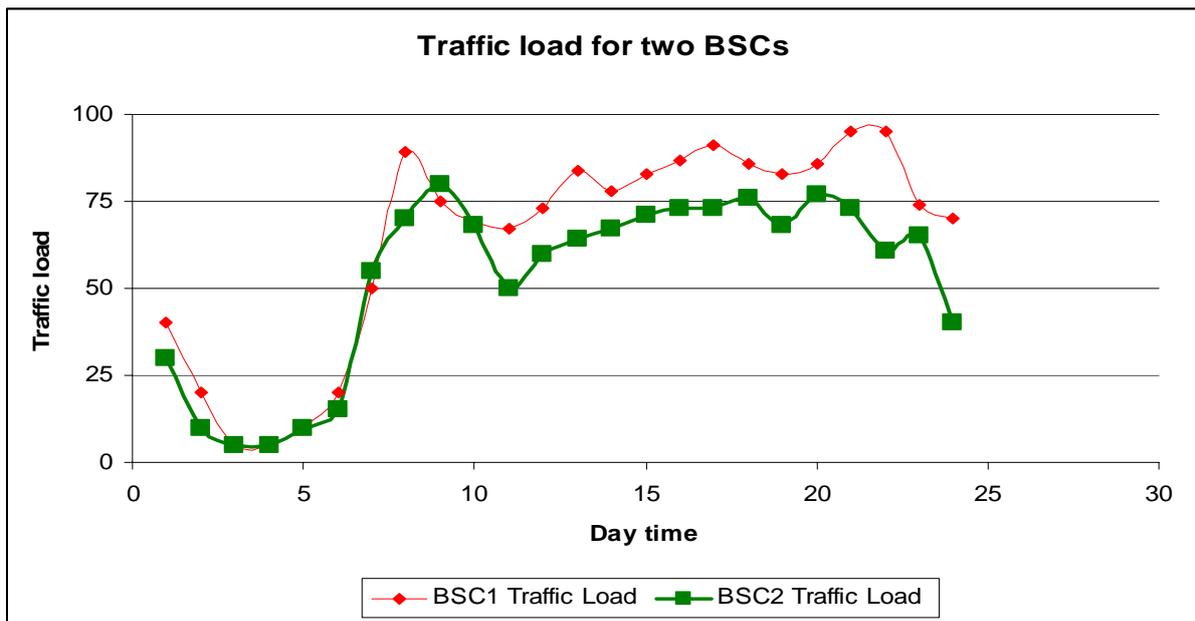


Figure 1 shows traffic load of a time-varying network for two different BSCs

Hourly time-scale variations in a customer market are shown in Figure 1 (where the x-axis refers to the measurement of time in a 24-hour format and the y-axis shows the resulting time-average traffic). It is worth to mention that the daily load differs from network to another, and it varies from time to time. Moreover, rural area peak load differs from urban area peak load. They fluctuate according to the seasons. Multilink's stations scenario utilizes from the previous idea and implements it to manage and balance the network load.

## LOAD BALANCING USING CELL BREATHING

Cell breathing is a mechanism that attempts to keep the forward and reverse link handoff boundaries balanced by changing the forward link coverage according to the changes in the reverse link interference level. This technique is used to balance the radio resources between the cells [15]. Applying our scenario can enable this technique to be more efficient as we will explain later on this paper.

In the overlapping region, users are assigned to use the strongest-serving cell. Figure 4 shows that cell 1 is more heavily loaded than cell 2. When cell 1 and cell 2 belong to different RNCs, a lot of traffic can be lost to serve the users in the overlapping area. Our

scenario can be used to achieve the optimal usage of the radio resources. The primary advantage of our scenario is that it enables the mobile network to dynamically change the links between different RNCs and optimally utilize its capacity resources without having to change its hardware design. Another advantage of the proposed scenario is to serve the increased demands in a particular area. Cell breathing technique can be used to distribute the traffic after applying our scenario.

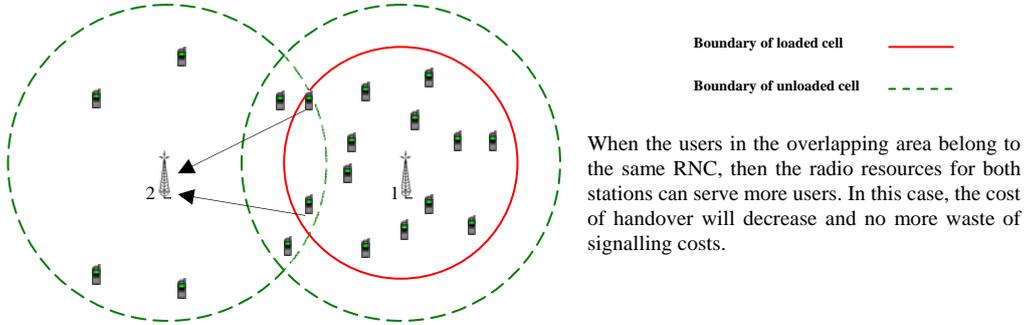


Figure 2 presents cell breathing technique

**MUTILINK’S STATIONS SCENARIO**

Mutilink’s stations scenario uses the multiple paths between Node-Bs and RNCs in the UMTS networks. The main idea here is to reconfigure the network upon the load. The traffic load differs from RNC to another one, so to achieve an optimal usage for the network; Mutilink’s stations scenario can be applied. The presented scenario can reconfigure the links between the Node\_Bs and the RNCs according to the network load. The decision can be taken according to the availability in the network. Calculations and measurements for the load in the different RNCs can lead to select which links should rebuild. This reconfiguration can help the operator to allow more users to use the network in the peak load. The radio resources in this case can be used efficiently.

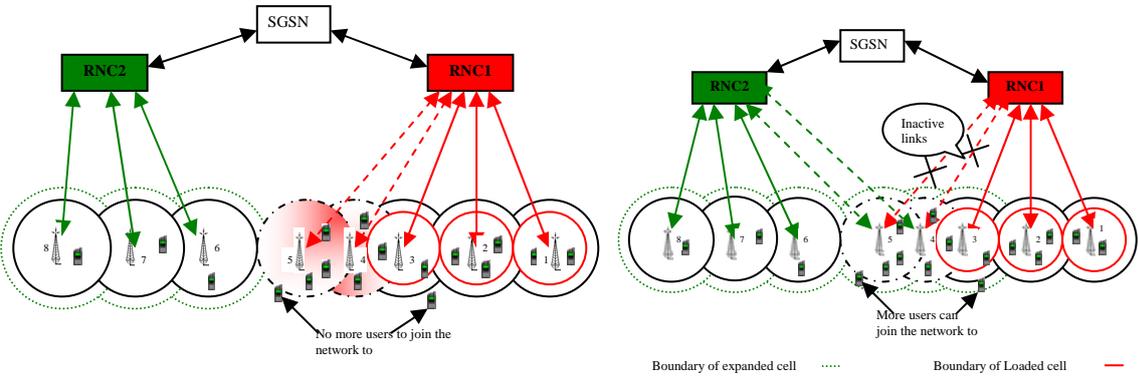


Figure 3 illustrates mutilink’s stations scenario which applied on an UMTS network. Before applying the scenario (Left), after applying mutilink’s stations scenario (right)

In UMTS networks, the network load is not regular. The traffic curve load of the network differs from cell to cell and it varies from RNC to another. Figure 3 shows the case when the network has two RNCs, one of them is loaded while the other one is not. In this case and by using a dynamic links between Node-Bs and the RNCs, the operator can switch between the links to connect the Node-Bs with the unloaded RNC2. More-download power control can be provided to each Node\_B in the loaded RNC1 after switching decision. This enables the cells to expand, share the load and redistribute the load with the other network elements. As it is appear from the Figure4 both stations (4, 5) can switch their connection and move from loaded RNC1 to unloaded RNC2. After switching the links, the unloaded station 6 can share the load with loaded station 5(new neighbor). Station 5 will become unloaded cell, and then it can share the load with the overloaded station 4. In this case the network can serve more new users with good QoS. All the other loaded stations in our scenario such as stations number (1, 2 and 3) will be affected. The new load of the network can be reduced. Here, the loaded RNC1 can achieve a significant decreasing in the load which enables new users to join the network as it is shown in Figure 5. Moreover, for the unloaded RNC2, nothing will affect while the network already has the capability to serve a new user. The load between the cells themselves can be shared using the already mechanisms such as cell breathing. Applying this scenario the SGSN should guide RNC to take the suitable decision about the switching time.

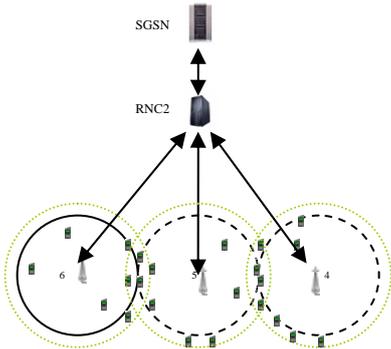


Figure 4 shows sharing the load between the cells after switching between the links in the loaded RNC1.

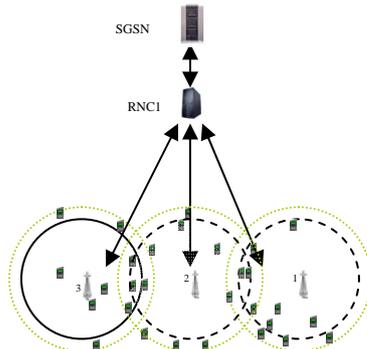


Figure 5 shows sharing the load between the cells after switching between the links in the unloaded RNC2.

It is important to notice that in this scenario we are not going to enable all the Node\_Bs to have multilinks. Real measurement can help the engineers to select which Node\_B has to have this capability. Under heavy load conditions at an RNC, large delays which can cause a serious degradation in the QoS for video and voice applications may be happened. Redistribute the load on the network among the existing network elements can reduce delays and collisions in order to improve the QoS perceived by users. Multilink’s stations scenario can decrease especially for the mobile users in the geographic area who they are under the ping-pong handover phenomena. Our scenario can be very helpful when the ping-pong handover is happening between two BTSs related to different RNCs, and it can reduce the handover cost.

When several mobile users move in a geographic area which can be served by two stations related to different SGSNs, our scenario can be applied to reduce the signaling cost and to balance RNC and SGSN loads. The multilink’s scenario can be used to balance the load among the coverage areas, reduce the signaling of the system and minimize the inter\_SGSNs traffic cost.

## CONCLUSION

Multilink's stations scenario to distribute the peak traffic load between network elements is presented. The study shows that our scenario can be applied as a good solution to balance the load in the existing mobile networks. It can decrease the on-peak load and allows new users to join the network in order to achieve significant increase in the overall network throughput and improve the network performance during the congestion periods. Future work will present a novel dynamic scenario to balance the network load which considers user's geographical footprints and user's movement and adjust the network up on that.

## References

- [1] M. Mouly and M.-B. Pautet, "The GSM System for Mobile Communications" Palai, France 1992.
- [2] P Demestichas , G Dimitrakopoulos , J.Luo , R.Agusti , E Mohyledin , O.Sallent , D Grandblaise , R Pintenet , P Leaves , K "Moessner Radio resource management and network planning in a reconfigurability context"Proc. 2004 IST Mobile Summit, Lyon June 2004.
- [3] W.C.Y.Lee " Mobile communications Design Fundamentals " John Wiley, New York, NY,1993.
- [4] S. Borst, G. Hampel, I. Saniee, and P. Whiting, "A Shadow-Priced-Based Distributed Approach to Dynamic Load Balancing in CDMA Networks," Handbook of Optimization in Telecommunications Springer Verlag, Norwell, MA, forthcoming 2005.
- [5] A. Bedekar , S.C. Borst ,K. Ramanan , P.A. Whiting, E.M. Yeh , " Downlink scheduling in CDMA data networks" Proc. Globecom 1999 (Rio de Janiero, Braz.,1999),PP.2653-2657.
- [6] Hector Velayos, Victor Aleo and Gunnar Karlsson, "Load balancing in overlapping wireless LAN cells", in Proc. of IEEE ICC 2004, Paris, France, June 2004.
- [7] K. Ghanem, and R. Abbas, " Coverage Roles in GSM networks" Journal for Studies and Scientific Research- Engineering Sciences Series. Tishreen University, Lattakia, Syria Vol (26) No (1) 2004
- [8] W.I. Kim and C.S. Kang, "A new traffic-load shedding scheme in the WCDMA mobile communication systems," IEEE Veh. Technol. Conf., vol.4, pp.2405–2409, Sept. 2002.
- [9] D. wan Tcha, S. yon Kang, and G. whan Jin, "Load analysis of the soft handoff scheme in a cdma cellular system," IEEE J. Select. Areas Commun., vol. 19, no. 6, pp. 1147–1152, June 2001.
- [10] G. Tsoulos, M. Beach, and S. Swales, "Performance enhancement of ds-cdma pcs cellular networks with smart antennas," in Proc. IEEE Global Telecommunications Conf., GLOBECOM '95, vol. 1, Singapore, Nov. 1995, pp. 213–217.
- [11] D. Everitt and D. Manfield, "Performance analysis of cellular mobile communication systems with dynamic channel assignment," IEEE J. Select. Areas Commun., vol. 7, pp. 1172–1180, Oct. 1989.
- [12] M. Feuerstein, "Applications of smart antennas in cellular networks," in Proc. IEEE Int. Symp. Antennas and Propagation Soc. , vol. 2, Orlando, FL, Aug. 1999, pp. 1096–1099.
- [13] C. Lee, H. Kang, and T. Park, "Dynamic sectorization of microcells for balanced traffic in cdma: genetic algorithms approach," IEEE Trans. Veh. Technol., vol. 51, pp. 63–72, Jan 2002.
- [14] Y. Argyropoulos, S. Jordan, and S. Kumar, "Dynamic channel allocation in interference-limited cellular systems with uneven traffic distribution," IEEE Trans. Veh. Technol., vol. 48, pp. 224–232, Jan. 1999.
- [15] Ylitalo, J. Jokikyyny, T. Kauppinen, T. Tuominen, A.J. Laine, J. 'Dynamic network interface selection in multihomed mobile hosts' System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on system science 6-9 Jan. 2003
- [16] F. Delli Priscoli, N. Magnani, V. Palestini, and F. Sestini, "Application of dynamic channel allocation strategies to the gsm cellular network," IEEE J. Select. Areas Commun., vol. 15, pp. 1558–1567, Oct. 1997.

## Authors:

Mr. Kinan Ghanem

Mr. Nadir Zamin-Khan,

Mr. Mohamed Abdrabou Kalil

Prof. Andreas Mitschele-Thiel

Technical University of Ilmenau

Faculty of computer science and automation

Integrated Hardware and Software Systems Group

D-98693 Ilmenau

Phone: +49 (0)3677 69 1696

Fax: +49 (0)3677 69 1220

E-mail: {Kinan.Ghanem /nadir.zamin-khan/mohamed.abdrabou/ Mitsch}@tu-ilmenau.de

Nadir Z. Khan / M. A. A. Kalil / K. Ghanem / A. Mitschele-Thiel

## **Generic Autonomic Architecture for Self-Management in Future Heterogeneous Networks**

### **ABSTRACT**

Nowadays, self-management capabilities, in any system, are provided by developing some independent modules. The modules performs different targeted operations. As the element of heterogeneity will increase in future networks, the interactions among these modules will become complexer and will be harder to manage. Thus the autonomy in the system will decrease. We proposed a generic five layered architecture which focuses in increased self-management capabilities for heterogeneous networks. This paper presents the idea of a generic five layered architecture for self-management in heterogeneous networks with an overview of interactions and interfaces.

### **INTRODUCTION**

Self-management capabilities are considered very essential for large heterogeneous telecommunication networks. A common practical approach is to provide different self-management capabilities with the help of one or more external modules. Usually these modules are developed having certain targets, in mind, for specific requirements. Usual targets of these modules are to monitor a specific system behavior, adapt to changes at run time, to configure the system, to improve performance or to recover from different faults. It is obvious that the combination of different modules, working together with each other, will increase as the heterogeneity in the telecommunication system will increase. Sometimes, it is desired to integrate different modules in a telecommunication system either in order to increase performance of the combined modules or to have new functionalities in the management system of the heterogeneous network. Since the combination of the modules increases with the number of desired functions, obviously the extent of autonomy in the management system of any heterogeneous network will decrease. That means, the automatic functions in management system will be more complex and difficult. Firstly, without any generic architecture of management system for heterogeneous system it would nearly be impossible for all the independent modules to work in an integrated way. Secondly, if in any case it becomes possible it would be only due to a large number of different types of independent interfaces among different

modules.

### **RELATED WORK**

Different architectures have been proposed to develop specific self-management functionalities in networks. In [1], policy based management architecture is developed. This architecture is related to solving complex issues with the help of integrated mechanism of knowledge-based reasoning and business policies. The work, in [2], is to have better co-ordination for global configuration and fault repairing. This work is focusing on the self-management video conferencing system. This work is not a solution for a generic architecture of autonomic self management system. Authors in [3], [4] proposed autonomic service architectures ensuring the service level agreements. In [3], a layered architecture is proposed for autonomic networks. These two proposed architecture are not talking about the issues of reducing complexity in integration of independent management modules. The work, in [5], have described some design pattern and recommended some interaction towards an autonomic architecture.

### **FIVE LAYER ARCHITECTURE**

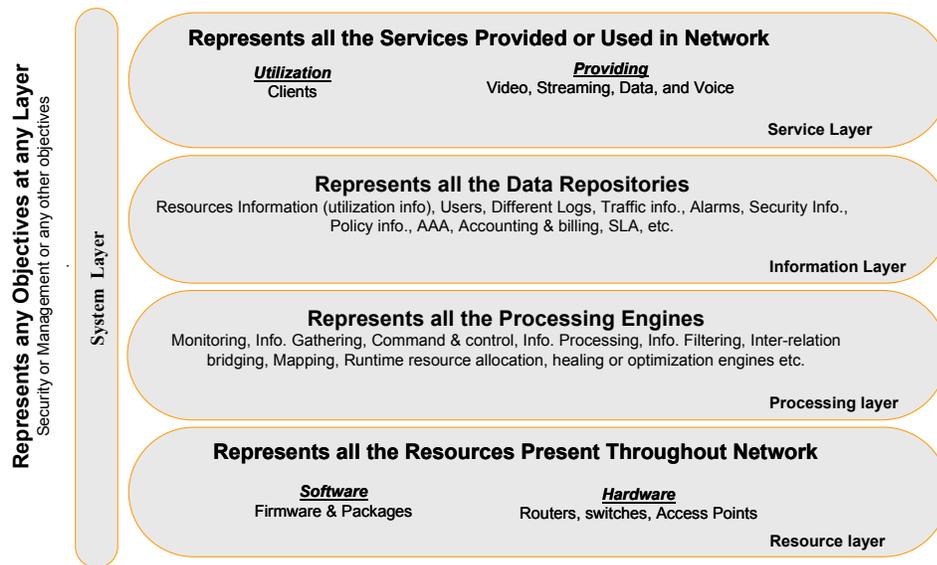
To have a proper way to achieve autonomy in management of any heterogeneous system, we realized the need of defining a generic architecture for management of any heterogeneous system. We proposed a layered architectural approach to have efficient integration of different self-management modules in heterogeneous networks. This architecture focuses on the theories and architectural definition of behavioral requirements for self-management components in heterogeneous networks. It discusses the minimum interactions among components and recommends design patterns to introduce system-level properties of self-configuration, self-optimization, self-healing and self-protection.

The main objective is to define an architecture which introduces the dynamic re-configurability in the future network elements as well as the capability to share and distribute the services between the different providers. The automation of services management has a significant importance in the future management systems. Thus the proposed generic architecture is designed to provide the capabilities to get a dynamic, reusable, scalable and reliable infrastructure that can integrate with the existing systems too.

The proposed architecture is divided in different logical layers according to their purposes and functions. It is a two dimensional layered architecture. The architecture is composed of four horizontal layers and a single vertical layer.

## Vertical Logical Layer

In figure 1, **System Layer** is vertically placed in the architecture. It represents different global objectives of the network systems, for example service authorization & access management, configuration management, fault management, security management, performance management, bandwidth management, and accounting management. Each management system can have a lot of processes, information flow mechanisms, resources and services to carry on the specific management tasks. To perform the management of a network system, the vertical layer can have some sub-systems that may include systems like monitoring systems, security systems, network administration systems, resource management systems. Each system at the vertical layer will have a lot of elements that are categorized in different horizontal layers. The elements in horizontal layer combine together to make up a system that is represented in the vertical layer.



**Figure 1: Logical view of different layers in generic autonomic architecture for self-management in future heterogeneous networks**

## Horizontal Logical Layers

The horizontal logical layers are viewed logically in any system. The idea is to introduce the flexibility by logically dividing all the elements for future network systems. This is the key to enable flexible integration of different modules in any system. **Service Layer** is divided into service provisioning and service utilization. This layer differentiates into service provider and service utilization entities. Service provisioning could include video, audio, data or any composite service. **Information Layer** represents the knowledge base of the system. In simple words it is the data resource of all the system. This layer distributes all information among the system such as policy information, authentications, accounting and billing, SLA, security alerts, and alarms.

**Processing Layer** deals with all the processes and engines that make a system. This layer represents any process that is present in the system as a separate and integrable entity. This can for example include the monitoring engine, the command & control interface, and the resource allocation interface. **Resources Layer** is further divided into hardware and software resources. The software resource could be firmware or any software package that is being utilized or might be needed by any other piece of software and hardware or by any service providing elements. Hardware represents any resources that are present in the network system such as access points, routers, and switches. In order to have autonomous capability in the system the resources, that might be a software and hardware resource, are divided into different categories according to their ability and functionality.

### **Categorization of Resources**

Different resources (i.e. hardware or software) are divided, into different categories, according to their ability and functionalities. A resource is categorized as Cat-0A assuming it has very less negotiation power and a dedicated information delivery. A resource belongs to the Cat-1A category when it has reconfigurable service capability, by changing firmware or what so ever. However it will lack for good coordination capabilities with different resources. A Cat-1B categorized resource is supposed to be a resource that could be reconfigurable and would have strong coordination and collaboration capabilities. The highest category, considered so far, Cat-1C is considered as a powerful resource with good storage and processing capabilities. It can provide multiple reconfigurable services at a time. It also may act as a leader and information source for similar and for lower categorized resources. The table1 summarizes the categories described above.

**Table1: The categorization of resources**

| <b>Resource Category</b> | <b>Capabilities and Functionalities</b>   |
|--------------------------|---|
| <b>Cat-0A</b>            | Dedicated Information delivery and Limited Negotiations   |
| <b>Cat-1A</b>            | Reconfigurable Service. One at a time. Less Co-ordination Capability.   |
| <b>Cat-1B</b>            | Strong Co-ordination & Collaborative Capabilities   |
| <b>Cat-1C</b>            | Reconfigurable Multi-Services at a time. In case of hardware resource: Can share firmware. Can be acting as a source too. |
| ....                     | ....  |
| <b>Cat-xx</b>            | Categories to be used in future.  |

### **Interactions & Interfaces Among Different Layers**

The interactions among different layers are considered in order to clarify the required interfaces among different layers. Figure 2, shows the alternate view of logical architecture. All the autonomous management system is shown at the top of vertical System layer in figure 2. Different horizontal layers are lying over it. System Layer has

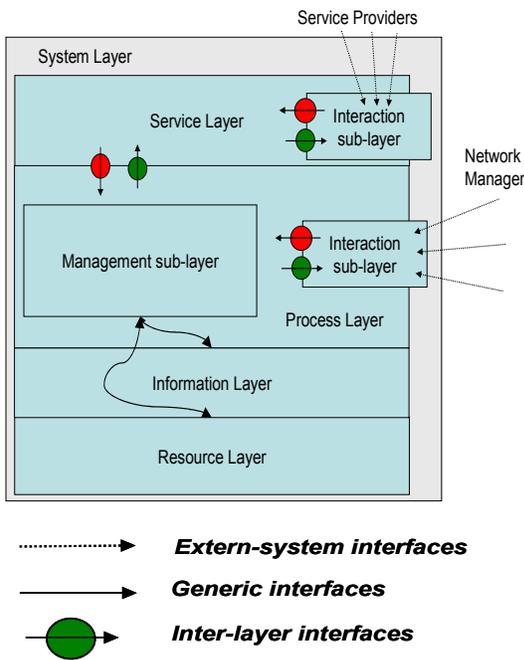


Figure 2: Alternative views of logical layers with interactions among different layers.

direct interactions to all other layer. The Process Layer transforms the decisions and information of the management sub-layer into instructions understandable by the Resource Layer. The Process Layer will have dedicated processes related to management and managers interaction sub-layer. So different processes performing management, interactions or security tasks, may lie on Process Layer.

There are three different types of interfaces. One type is used to provide the interaction between the system with external entities such as service providers or network

manager. These are called **extern-system interfaces**. Second types of interfaces are shown with circle and an arrow head. These interfaces are inter-layer interfaces. These are used as internal interfaces to provide special features along with the interaction between different layers. For example, **inter-layer interfaces** can provide secure transactions of information among different layers making sure that only the concern processes and resources are involved even within the same system that has several others non-concern processes and resources. The third type of interfaces are rather involved in normal interactions among different layers.

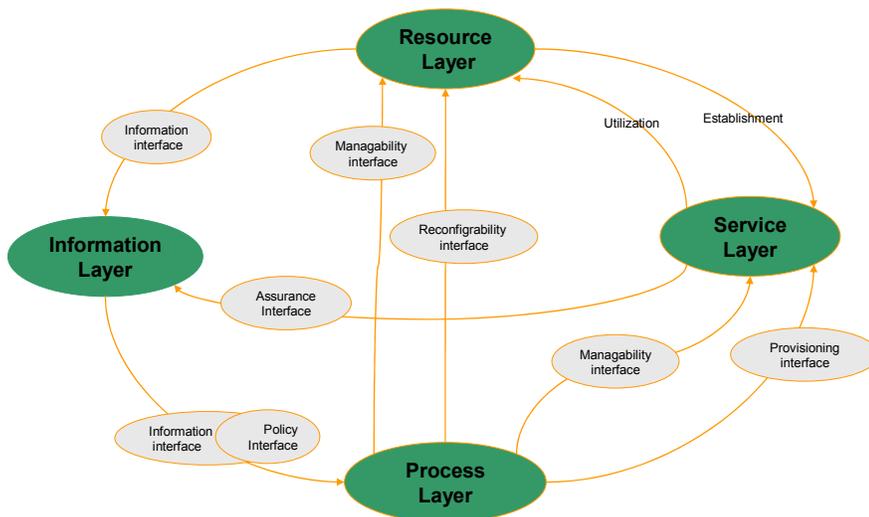


Figure 3: Example of possible interfaces and work flow among logical layers

In figure 3, different possible interactions with required interfaces in shown. In a self-management system resource will send some information to the information base. At the same time different processes at Process Layer will process the data from Information Layer such as policies and information to manage and provide different services at the Service Layer. Also Process Layer will use information from Information Layer to reconfigure and manage the resources according to defined policies. The reconfigured resources at the resource layer will establish different services at the Service Layer. Service Layer will provide information in order to ensure the service level agreements.

### **Summary and Conclusion**

In this paper, we introduced the generic five layered architecture for future heterogeneous networks which is focused to enable the development of systems with autonomous self-management capabilities. The presented architecture is divided into one vertical layer and four horizontal layers. A prototype of a future management system with different modules applications is developed according to core idea of this architecture. Many advantages such as increased flexibility, availability and mobility. have been observed during the development of these applications and modules. A complex future management system for heterogeneous network, with the idea of proposed architecture, is under development. In future, the benefits and weakness of developed systems with other system will be observed and compared.

### **References:**

- [1] Davy, S.; Barrett, K.; Balasubramaniam, S.; van der Meer, S.; Jennings, B.; Strassner, J. "Policy-based architecture to enable autonomic communications - a position paper," 3rd IEEE Consumer Communications and Networking Conference, Vol 1, pp. 590 – 594, Jan. 2006.
- [2] Shang-Wen Cheng; An-Cheng Huang; Garlan, D.; Schmerl, B.; Steenkiste, P. "An architecture for coordinating multiple self-management systems" Proceedings in Fourth Working IEEE/IFIP Conference on Software Architecture (WICSA), pp. 243 – 252, June 2004.
- [3] Gracanin, D.; Bohner, S.A.; Hinchey, M. "Towards a model-driven architecture for autonomic systems" Proceedings. 11th IEEE International Conference and Workshop on the Engineering of Computer-Based Systems, pp. 500 – 505, May 2004
- [4] R. Farha1, M. Sup Kim, A. Leon-Garcia, and J. W. Hong "Towards an Autonomic Service Architecture,". in Lecture Notes in Computer Science, Volume 3751, pp. 58–67, Oct. 2005.
- [5] White, S.R.; Hanson, J.E.; Whalley, I.; Chess, D.M.; Kephart, J.O. "An architectural approach to autonomic computing", Proceedings. International Conference on Autonomic Computing, pp. 2 – 9, May 2004

### **Authors:**

Mr. Nadir Zamin Khan

Mr. Mohamed Abdrabou

Mr. Kinan Ghanem

Prof. Andreas Mitschele-Thiel

Technical University of Ilmenau

Faculty of computer science and automation

Integrated Hardware and Software Systems Group

D-98693 Ilmenau

Phone: +49 (0)3677 69 1696

Fax: +49 (0)3677 69 1220

E-mail: { nadir.zamin-khan/ Kinan.Ghanem/ mohamed.abdrabou/ Mitsch}@tu-ilmenau.de

N. Z. Khan / K. Ghanem / S. Leistritz / F. Liers / M. A. A. Kalil / H. Kärst / R. Böringer

## **Network Management of Future Access Networks**

### **ABSTRACT**

The future heterogeneous networks will be more sophisticated and complex. The access network in future will mostly depend on wireless medium of communication with a demand of high bandwidth for the subscribers. The management of such a heterogeneous system is a big challenge. In this paper, we present the overview of a study for future access network management. General ideas involved in the development of simulated and emulated future networks are also presented. Different tools those are developed during the study are also discussed in this paper. To provide the higher bandwidth we used **Multi Protocol Label Switching (MPLS)** networks and reconfigurable wireless devices such as multi party access gateways to represent the wireless access networks are used in this work. The paper discusses the idea of management in future heterogeneous networks.

### **INTRODUCTION**

The future heterogeneous networks are supposed to be more sophisticated and complex than today's networks. The management of such networks is becoming a greater challenge. On one hand the access networks are becoming more dependent on wireless access networks and on the other hand the high speed packet forwarding technologies such as **Multi Protocol Label Switching (MPLS)** are expected to be a part of future access networks. There are lots of different issues to manage when one looks into an integrated view of such future access networks. Management issues related to wireless access to client are on one hand important. On the other hand the management of faults and traffic at the packet level is becoming more important too. In order to study these problems, a work was done keeping in mind the demands of future access networks that involves MPLS technology with wireless access to the clients.

### **MULTI PROTOCOL LABEL SWITCHING**

MPLS is a technology that attempts to improve the speed of packet forwarding in IP networks. MPLS-capable routers known as **Label Switching Routers (LSRs)** can use link-level forwarding to provide a simple and fast packet forwarding capability [1-4]. MPLS combines the simplicity of IP routing with the high-speed switching capability of **Asynchronous Transfer Mode (ATM)**. It has significant benefits which can allow future systems to use the MPLS not only in the backbone but also in access networks. It is necessary to use a label distribution protocol to setup the **Label Switching Path (LSP)** [5][6].

### **MULTI PARTY ACCESS GATEWAYS**

The multi party access gateway is a wireless access point, which could be used by more than one wireless access networks as well as a service provider. As an example, two service providers could offer their services over one physical device. One provider could offer VoIP and the other could offer internet web browsing without interfering each other. This new technology is provided by IDEO Laboratories GmbH and was developed in cooperation with the TU Ilmenau. The usage of this technology causes a higher administration work load than normal networks. Due to the existence of independent environments on one single access gateway, the configuration work load for an administrator is higher than in a usual network without these virtual environments. Manual configuration of all devices of a large scale networks with a high number of involved parties is very time consuming and therefore expensive. The delay between the arrival of a request for a network setup and the finishing of the installation is high and would prevent fast changes due to variable user traffic load. For an efficient usage in a large scale network an automatic configuration and management system is needed.

### **DYNAMIC MONITORING OF NETWORK ELEMENTS**

Future networks have to control and optimize themselves in wide ranges. The operator is not longer able to take decisions as it is required to ensure that services are not interrupted. Therefore, a smart monitoring system has to gather information about the network status and failures in order to enable a decision engine to react on the problem in the right way. It is easy to imagine that this would lead to the issues of high signaling load and high demand for computing resources for examining the monitoring data (e.g. determining trends and reducing the alarm rates). Two main approaches can be used to overcome such problems:

- Reducing the information per monitored network element or
- Decreasing the number of monitoring points in the network.

This work has examined these approaches. The aim is to observe an MPLS-based network and to detect if a network element (nodes or links) fails. Therefore a network generator was built in co-operation with IDEO & TU-Ilmenau that provides large network structures which are close to real network structures. Emulation and simulation of MPLS networks has been developed and used to study different monitor placement mechanisms. One of the targets of this work is to develop some approaches to reduce the number of monitors. The number of monitoring nodes can be dramatically reduced if a smart placement mechanism is used.

### **IMPLEMENTATIONS**

The TU Ilmenau has developed such a system, which provides a fully automatic management of the multi party access gateways as well as the optimum placement of monitors to find the faults in the access network. It supports the whole workflow for configuring the devices from the moment of power on until the full integration in the network. This workflow consists of three phases, which are separated in different steps. The first phase is the information collection phase with the steps of hardware integration, discovery and description. The new hardware device had to be integrated in the network physically. In special there had to be power supply of the device and a wire or wireless connection to the network backbone. Now the device had to be detected by the network and a description of the device and its features had to be determined. The second phase is the information evaluation and decision making phase. Based on the collected data of the device and the configuration of the other network elements a reconfiguration decision is made. Therefore the required services for the network have to be mapped to the available resources. This service to device mapping must take into account several constrains. First of all the hardware requirements of all services must be fulfilled. A device, which should run a service, must provide enough memory, disk capacity and CPU speed. Second a device should not be overloaded with services. This is especially critical for services with resource consumption depending on the number of user and requests, respectively. Therefore the load of a service and its resource needs must be forecast. Due to the uncertainty of this forecast a part of the resources of a device should be reserved for peak load and finally the quality of service for services must be considered for the placement of the service entities. The main focus here is on the delay

and communication relationships between the services and the service entities. If services must communicate with each other and if they are placed on different devices, a communication delay occurs. In case of a defined maximum delay for the whole service, the placement must respect the delay constraints. When the decision phase is over, the last phase starts. This reconfiguration phase consists of a building and delivering the device and service configuration and execution of services on the devices. The configuration does not consist of some plain configuration files only. For most devices complete hardware depended operation systems with special device drivers and application programs are needed. They must be compiled with cross compiler, which supports the hardware architecture and CPU instruction set.

With this management system it is possible to detect and configure multi party wireless access gateways without manual interactions on one hand and on the other hand the same system manages the dynamic placement of monitors to reduce the monitoring cost of access networks. This work has resulted in the development of a more efficient and integrated management system for future access networks.

## CONCLUSIONS

The future heterogeneous networks will mostly be using wireless medium and a high bandwidth will be demanded by subscriber. To have an uninterrupted higher bandwidth, the networks should be managed by using less costly and dynamic monitoring mechanisms. Advanced wireless access gateways are configurable and a step further towards a management of future access networks. TU Ilmenau along with IDEO laboratories has researched management of future access networks. The research resulted in development of a management system prototype which dynamically configures devices according to demands and also places different types of monitors according to demand. A prototype of this management system is currently working in Wireless lab of TU Ilmenau. In future, more sophisticated fault detection and smarter placement of monitors for several different services will be considered.

### References:

- [1] X. Xiao, A. Hannan, B. Bailey, and L. Ni, "Traffic Engineering with MPLS in the Internet," IEEE Network, March/April 2000.
- [2] Rosen, E., Viswanathan, A. and R. Callon, "A Proposed Architecture for MPLS", Work in Progress.
- [3] Menth, M.; Milbrandt, J.; Reifert, A.: Self-Protection Multipaths - A Simple and Resource-Efficient Protection Switching Mechanism for MPLS Networks, Technischer Bericht Nr. 321, Universität Würzburg, Februar 2004
- [4] Scharma, V.; Hellstrand, F.: RFC 3469 - Framework for Multi-Protocol Label Switching (MPLS)-based Recovery, Februar 2003
- [5] Stoller, S.: Leader Election in Asynchronous Distributed Systems, In: IEEE Transactions on Computers Ausgabe 49 Heft 3,

S. 283-284, März 2004

[6] Subramanyan, R.; Miguel-Alonso, J.; Fortes, J.; u.a.: A Scalable SNMP-based Distributed Monitoring System for Heterogeneous Network Computing, In: Proceedings of Supercomputing, S. 52, 2000

[7] Thottan, M.; Li, L.; Bin, Y.; u.a.: Distributed network monitoring for evolving IP networks, In: International Conference on Distributed Computing Systems Nr. 24, S. 712 - 719, Tokyo, 2004

**Authors:**

Mr. Nadir Zamin Khan

Mr. Kinan Ghanem

Mr. Florian Liers

Mr. Mohamed Abdrabou

Mr. Holger Kärst

Technical University of Ilmenau

Faculty of computer science and automation

Integrated Hardware and Software Systems Group

D-98693 Ilmenau

Phone: +49 (0)3677 69 2821

Fax: +49 (0)3677 69 1220

E-mail: {nadir.zamin-khan/ Kinan.Ghanem/ Florian.Liers/ mohamed.abdrabou}@tu-ilmenau.de

Mr. Stefan Leistritz

Mr. Rene Böringer

IDEO Laboratories GmbH

Ehrenbergstr. 11

D-98693 Ilmenau

Phone: +49 (0)3677 66 8354

Fax: +49 (0)3677 69 8359

E-mail: { s.leistritz / r.boeringer}@ideo-labs.com



Steffen Schmidt, Holger Kärst, Andreas Mitschele-Thiel

## **Towards Cost-Effective Area-wide Wi-Fi Provisioning**

### **ABSTRACT**

Wi-Fi provides small cells resulting in a huge amount of Access Points necessary to obtain an adequate coverage. This directly leads to enormous network cost.

This paper proposes a revolutionary platform, referred to as ZONOS, which allows for a cost effective installation and operation of a carrier grade Wi-Fi network, offering public wireless access points.

This is achieved by CUCULUS, using private Access Points, whereby a part (meaning resources like CPU cycles, RAM, Flash or Interfaces) of each Access Point is virtually given back to a large provider. Those Access Points can be shared between private users and an Internet Service Provider. A sharing process may be initiated on-demand and is as simple as ordering any product in an online-shop. The process includes a self-configuration of the device meaning a logical but secure separation of so called zones within the Access Point. Internal security is provided by an integrated certificate management system and an independent control unit. One of the zones of the so configured Access Point is then automatically registered with the provider. Every zone behaves like a physical device. The control unit supports negotiation processes between the parties of the shared device. No party of the Access Point has the ability to influence, may be due to configuration processes, the functionality of any other zone without the acceptance of the affected party. Thus, the platform ZONOS enables various parties to use the same physical device securely.

A provider can use its own monitoring and control tools to manage the new part of its network. A secure user communication by means of a ZONOS enabled Access Point is realized both by WPA encryption on the wireless and SSL tunnelling on the wired part of the path.

From the management point of view CUCULUS means that new approaches have to be invented and studied. Currently known and often centralized solutions cannot be applied to networks of the resulting size. Today very large networks consist of several 10s of thousands of devices. If private devices are included the size will grow up to several 100 thousand or even million devices. In this paper we will additionally present the challenges for network management that have to be faced in these networks foreseen in the future.

The uniform area-wide low-cost access using Wi-Fi networks will revolutionize the entire mobile and stationary communication. New services and use cases will open up new markets for the applying providers.

## INTRODUCTION

While almost 300 million people worldwide use wired fast broadband connections for accessing the internet, wireless broadband access everywhere is still a dream. The reason is, that establishing and operating a public Wi-Fi infrastructure supporting area-wide coverage nowadays causes enormous network cost. Wi-Fi provides small radio cells so that it requires a huge amount of Access Points for obtaining a sufficient coverage. For those Access Points the operator of a Wi-Fi network has capital expenditure (capex) for the Access Points themselves, the installation and sites of the Access Points. Also, operating expenditure (opex) for power, connecting the Access Point to a wired broadband access, rental fee for the site of the Access Point and more lead to high costs. A brief overview of capex and opex is provided in Tables 1 and 2.

| Item                       | Unit Cost     | Total Cost          |
|----------------------------|---------------|---------------------|
| Hardware                   | \$504         | \$3.528.000         |
| Software                   |               | \$3.000.000         |
| Site Survey                | \$450         | \$2.250.000         |
| Installation/Configuration | \$544         | \$2.720.000         |
| <b>Total</b>               | <b>\$1498</b> | <b>\$11.498.000</b> |

*Table 1: Capex for providing a public Wi-Fi with 7000 Access Points ([2])*

| Item                      | Unit Cost    | Total Cost          |
|---------------------------|--------------|---------------------|
| Broadband Access Backbone | \$185        | \$11.100.000        |
| Backoffice                |              | \$13.573.800        |
| Support/Maintenance       |              | \$2.250.000         |
| <b>Total</b>              | <b>\$185</b> | <b>\$21.373.800</b> |

*Table 2: Opex per year for providing a public Wi-Fi with 7000 Access Points ([2])*

Furthermore, every country has its own special laws and regularities for providing Wi-Fi networks. Restrictions to transmission power, the usage of radio frequencies or laws for securing Wi-Fi-networks are examples.

Therefore a huge amount of customers is necessary for providing a profitable Wi-Fi network. Customers, however, cannot be attained only by providing to them a simple transport service over a wireless broadband access. Customers want to use innovative services which can be offered having an already established wireless broadband access. However, due to the high capex and opex nowadays a Wi-Fi network can only be cost-efficient if no changes for the infrastructure are necessary for a long period of time. Following these restrictions, providers of networks are limited in innovation and flexibility, meaning the rollout of new services for customers is difficult, too expensive and therefore too slow for increasing the group of wireless users significantly.

To break this loop this paper presents a new approach, referred to as CUCULUS, which allows for a cost effective installation and operation of a carrier grade Wi-Fi network. It will support to achieve the target of area wide Wi-Fi provisioning and innovation in wireless networks. The concept of CUCULUS is based on using infrastructure multiple times by several parties in parallel, whereas for every party its zone appears to be stand-alone. As an example, the paper describes the CUCULUS principle by sharing private Access Points between their owners and providers of public Wi-Fi networks. The potential of parallel usage of Access Points providing a private as well as a public Wi-Fi will be supported by the novel platform ZONOS. ZONOS allows for sharing of infrastructure by partitioning the physical devices into logical encapsulated zones controlled by a secure and independent control unit.

## THE CUCULUS TECHNOLOGY AND ZONOS®

The most common and largely preferred broadband internet access technology today is DSL. As stated in [1], the number of DSL connections has risen over 200 million and still increases by approximately 30% per year. In Germany alone more than 15 million DSL users are registered, most of whom are customers of one of the major DSL providers.

Commonly nowadays, a DSL connection comes with a wireless DSL router, to supply the household with a wireless connection (Wi-Fi / WLAN) to the internet. This leads to a high Wi-Fi coverage, whereas the wireless medium may only be used by the owner of the DSL router. Using the already available private Wi-Fi infrastructure connected by broadband access to the internet (mostly DSL) to provide public Wi-Fi hotspots in parallel, the dream of wireless access everywhere will take one step further to reality.

Built up and maintenance for public networks, supported by private equipment will be significantly more cost efficient than today's wireless networks are. The high operational expenditure will be distributed between all parties participating in the network. Different organizations have taken up this idea forming so-called Wi-Fi communities. Those approaches rely on the agreement of private DSL users to give other people free access to their Wi-Fi network, thereby use their DSL account and providing a community hotspot. This issue, however, directly leads to significant deficiencies concerning legal issues, privacy and security. When opening their DSL / Wi-Fi to other, commonly unknown users, the owner of the DSL account has to take responsibility of all the actions taken using this account. Furthermore, the private network of the DSL user is exposed to foreigners, that might very likely access, retrieve or manipulate private data. Even if the people accessing the hotspot are registered within the community, there is no guarantee provided, that the registration information is valid. For those and several other reasons community hotspots are limited to a specific user group and not suitable for the mass market resulting in very low coverage and acceptance.

To exclude all the disadvantages of community hotspots and thereby providing a reliable, area-wide public Wi-Fi network, CUCULUS has been developed. This technology allows for the secure usage of hardware devices and existing infrastructure by several independent parties without influencing one another. This creates the possibility of using private Wi-Fi router devices in parallel by their private owner and an Internet Service Provider. Both parties possess a certain amount of the device (a so-called zone) and are able to manage and configure their dedicated zone on the hardware device independently. One physical device behaves as several completely separated devices. This enables the Internet Service Provider to establish a public Wi-Fi hotspot encapsulated in one zone within the private wireless router. This mechanism releases the private DSL user of all the responsibilities connected to the public Wi-Fi access, which are now transferred to the Internet Service Provider. Additionally it strongly secures the private network against intrusion, building up a highly reliable security frame, isolating both zones and thereby separating the resulting networks from each other. Table 3 summarizes the most important advantages of the CUCULUS approach over Wi-Fi communities.

| CUCULUS Technology (ZONOS)                                     | Community Approach                                |
|--|---|
| Every party controls its own part of the device independently. | Everything is under the control of the community. |
| Real device sharing  | Only different SSIDs                              |
| Security by design   | Open by concept                                   |
| A user has to be identified and authorized                     | No user identification                            |
| Risks are bound to parts                                       | Risks are shared by concept                       |
| Hardware independent   | Dedicated hardware                                |
| Efficient management by providers                              | Managed by the community.                         |

Table 3: Comparison CUCULUS vs. Community Approach

The core technology building the foundation of CUCULUS is formed by ZONOS, the ZONed Operation System. A schematic overview of ZONOS is given in Figure 1. ZONOS is a Linux-based operating system allowing for the secure, logical separation of hardware devices into several independent zones. All zones may be configured and managed separately, without influencing one another. No user of a zone can gain root access to the whole device, only for the zone itself. For management and control functions, ZONOS provides a Control Unit that autonomously handles the coordination between the different zones, distributes device resources and ensures the secure separation and independence of the zones. The hardware device may be separated / shared in terms of bandwidth, CPU cycles, RAM, File System and interfaces.

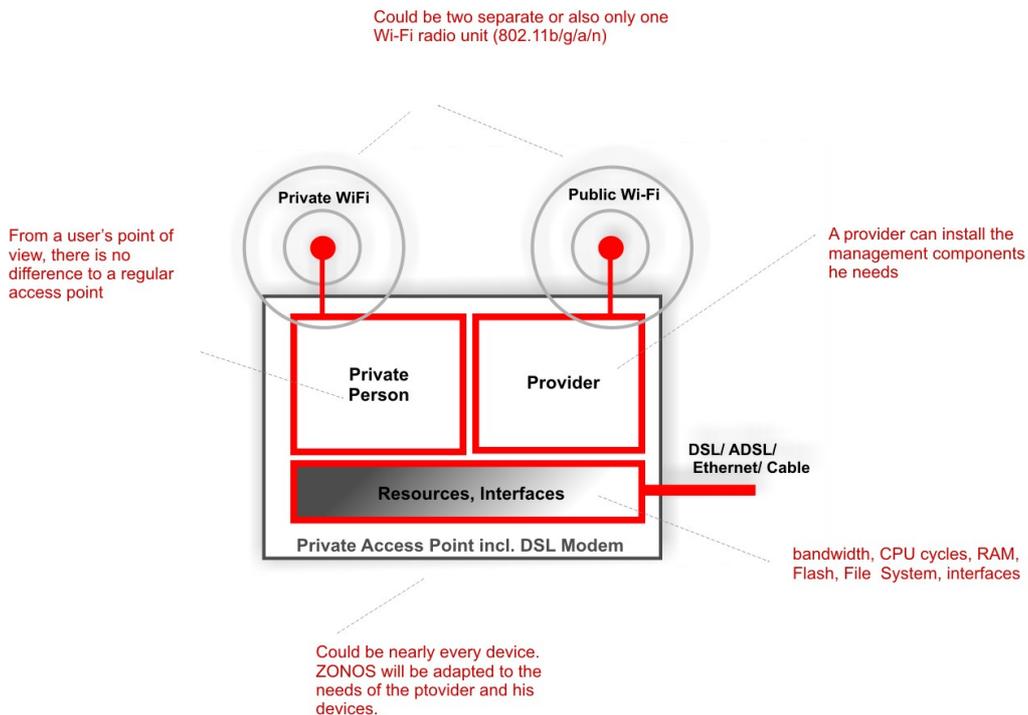


Figure 1: ZONOS - Overview

ZONOS requires very limited hardware resources and is widely portable to a large variety of hardware platforms for embedded and non-embedded devices. Additionally, ZONOS strictly follows a service oriented approach. Because of the independent manageability and configurability of all zones contained within a ZONOS device, ZONOS supports and features the installation of different services

within a zone, as to adapt the device to certain requirements. This feature creates the capability for service providers to migrate services directly to the public access point, enabling for mobile or location-based services, as well as the presentation of context-sensitive information to mobile users of the hot spot.

In every zone a different service set can be installed and configured independently. New services and service architectures can be developed and distributed, leading to unknown possibilities. The CUCULUS Technology, based on ZONOS, thereby leads to the creation of provider-driven, carrier-grade, encapsulated hotspots.

## **MANAGEMENT CHALLENGES**

From the CUCULUS approach two major challenges arise. The first challenge concerns the sheer size of the resulting network structure managed by a single provider. Current wireless access network structures include up to a number several thousands of access points. Applying CUCULUS this number increases within a very short time to several 100s of thousands or even millions of devices to be managed. The numbers directly correlate with the earlier described number of DSL accounts belonging to a provider.

This fact demands the re-design of management structures, interfaces, mechanisms and facilities to achieve a reliable and secure approach to managing and configuring the Wi-Fi devices. This effort contains AAA (Authentication, Authorization and Accounting) mechanisms, fault and availability management to present reliable network and service structures, as well as configuration management, guaranteeing a minimal influence of Wi-Fi devices on each other and a broad coverage of services within the Wi-Fi network. To decrease the management effort the separation of three types of providers was proposed in [3]. The Infrastructure providers offer their physical network resources to management providers. Management providers integrate the offered infrastructure in their management domain whereas their customers, the service providers use the managed network for providing services like VoIP or IPTV.

The second challenge in management will arise due to the possibility of ZONOS allowing the installation of services on Access Points directly. Thus, services can be moved from the core network to the network edges. The Access Points will become a kind of service platform, whereas the provided service set can be adapted every time easily and flexible. On one side it will enhance the reliability and QoS for services, because the service provisioning can be fast and flexible adapted to the current service utilization. New options for load balancing as well as moving of services between locations will be possible due to the huge amount of available interconnected service platforms. On the other side new challenges for managing the distributed services arise. The complexity for management will increase enormously. Optimization problems have to be solved determining an optimal placement for services so that the service level agreements are fulfilled whereas the service provisioning is as cost effective as possible. An optimal set-up of infrastructure has to be found by selecting shared resources corresponding to the demand for service provisioning. This set-up varies between locations. The complexity will arise again if services haven't got a fix location any more so that during the runtime of services a replacement can occur. The service resources as well as the state of the service could be moved whereas for the user of the service the replacement must be transparent.

Furthermore the configuration management becomes more complex. For any service platform at any time this service configuration has to be known so that in case of

failures the service set can be installed on a fallback platform with the same configuration. Reconfiguration of this interconnected service platforms is quite difficult because of dependencies between them. The dependencies arise on one side due to the service distribution within service chains in case of one service needing another service to run correctly. On the other side services could also disturb each other. For example, neighbouring Access Points configured to communicate over the same radio frequency will generate interferences.

Summarized, the new point of interest for research is that an infrastructure based on CUCULUS is highly dynamic in availability and interconnection of network elements, as well as in service provisioning. At any time new network elements could appear whereas other ones leave the network. Services change location dynamically adapting the network to the current demand.

Since 2001 several activities in research are observable engaging with complexity degradation proposing self- or autonomous network management [4, 5]. These activities should be strengthened, so that future networks with expected enhanced functionalities and size will be manageable.

## CONCLUSION

The CUCULUS technology, based on ZONOS, offers the ability for the creation of an area-wide broad-band internet access and thereby a wide range of possibilities. It, by design, includes security mechanisms and creates capabilities for yet unknown new service structures and usage scenarios. CUCULUS seamlessly integrates into existing provider systems and is therefore a great step forward to the internet of the future. However, it requires new approaches for fault-, configuration-, performance and service management taking into account the new kind of infrastructure and service provisioning. It must be determined, whether centralized management approaches are still able to handle the new challenges or decentralized solutions would be able to solve this complexity in a better way.

### References:

- [1] DSL Forum, News Release, "More than 200 Million Customers Chose DSL", Beijing China, June 2007
- [2] Richard Dean, Shiv K. Bakhish, "An Analysis of Total Cost of Ownership for Service Providers Deploying Public WLANs", March 2005
- [3] Holger Kärst, René Böringer, Andreas Mitschele-Thiel "Reconfigurability as a Key for new Network Services", Würzburg Germany, July 2006
- [4] Jeffrey O. Kephart, "David M. Chess "The Vision of Autonomic Computing", Journal IEEE Computer, January 2003
- [5] Shen, C.; Pesch, D.; Irvine, J. "A framework for self-management of hybrid wireless networks using autonomic computing principles" Communication Networks and Services Research Conference, Nova Scotia Canada, May 2005

### Authors:

Dipl.-Inf. Steffen Schmidt  
Dipl.-Inf. Holger Kärst  
Cuculus GmbH, Ehrenbergstraße 11  
98693 Ilmenau  
Phone: 03677/668 53-0  
Fax: 03677/668 53-9  
E-mail: {s.schmidt, h.kaerst}@cuculus.net

Prof. Dr.-Ing. habil Andreas Mitschele-Thiel  
Technische Universität Ilmenau  
Gustav-Kirchhoff-Str. 1  
98693 Ilmenau  
Phone: 03677/692819  
Fax: 03677/691220  
E-mail: mitsch@tu-ilmenau.de

Ausama Yousef, Mohamed Abd rabou Kalil

## **A New Algorithm for an Efficient Stateful Address Auto-configuration Protocol in Ad hoc Networks**

### **ABSTRACT**

A mobile ad hoc network (MANET) is an autonomous system of mobile routers that are self-organizing and decentralized without any need for infrastructure support. A dynamic topology of MANETs makes any configuration process difficult to be handled with standard methods as known in the statical one. Therefore, the challenge is to make MANETs able to configure themselves automatically. Address assignment of nodes is the main issue of auto-configuration. Although many protocols have been created to handle address auto-configuration in MANETs, neither of them can handle this issue efficiently. In this paper we present a solution for an efficient auto-configuration protocol designed to work in MANET. We are developing Logical Hierarchical Addressing protocol (LHA) which is classified in a stateful protocol. The LHA protocol allows the nodes to get a unique address fast and dynamically. Also it efficiently manages and handles some scenarios such as merging and partitioning of networks.

### **I- INTRODUCTION**

Wireless communication networks are developing rapidly. An Ad hoc network [1] is one of these networks, which is an autonomous system of mobile devices. These devices work as both hosts and routers. This enables the mobile nodes to interconnect each other through multi-hops without any need for a predefined communication infrastructure. The cost of infrastructure networks in uncovered areas is one of the main issues so that using MANETs can be a well solution. Because of the dynamics of MANETs and in order to support spontaneous networking, it is important to achieve the configuration process automatically. Address assignment of nodes is the main issue of auto-configuration. However, address auto-configuration protocols known from the infrastructure network, e.g. Dynamic Host Configuration Protocol (DHCP) [2], are not

sufficient for Ad hoc networks for several reasons. One of them is the need for a centralized management. Although many protocols have been created to handle address auto-configuration in MANETs but neither of them can handle this issue efficiently. There are several limitations of these protocols. They suffer from low efficiency caused by large protocol overhead, resources limitations and potential address conflicts during merging operations. Thus, new protocols have to be developed to overcome these issues.

This work proposes LHA protocol for a distributed address auto-configuration process in MANETs. The LHA provides a fast mechanism to configure a joining node with a unique address. It generates almost no protocol overhead since it uses routing protocol messages to update and distribute the required information.

A significant challenge for auto-configuration protocols is the frequent partitioning and merging of networks, which may cause address conflicts. The LHA protocol provides a good method to solve these address conflicts.

The rest of this paper is organized as follows: an overview of related research efforts is given in section 2. The basic algorithm of the LHA protocol is presented in section 3. Section 4 concludes the paper and gives an overview about the next steps and future research issues.

## II- RELATED WORKS

Various auto-configuration approaches have been developed to meet the requirements of Ad hoc networks. These approaches can be classified as described in [3] into stateful, stateless and hybrid approaches, focusing on the way to maintain address allocation tables in the nodes.

Protocols following the stateless approach, assign a new address to a joining node faster than protocols of the stateful approach. Stateless auto-configuration protocols allow the new node to construct its addresses by itself. These addresses are typically based on a hardware ID or on a random number given by a random number generator. Of course there is no guaranty that this address is unique in the network. This means that a Duplicate Address Detection (DAD) mechanism is needed to verify the uniqueness of the address in the network. Thus the focus in this area has been on designing and optimizing the DAD mechanisms. The use of a DAD mechanism can slow down the attachment process of a mobile node. This is especially the case when

the DAD is used before the address configuration is finished as in [4]. Even when the DAD mechanism is applied after the address has been assigned as is the case with Weak DAD (WDAD) [5] and Passive DAD (PDAD) [6], it can not meet the requirements of real time applications. This may occur if the connected node changes its address which may interrupt the transport layer connection.

In contrast, by using a stateful approach the uniqueness of the address can be ensured by concept. However, this has to be paid by a slower assignment process which does not suite the interworking aspects in hybrid networks, such as fast handoff. As presented in [7] the MANETconf protocol distributes the available address space to the nodes in network. It prevents the assignment of the same address to more than one node by maintaining additional allocation tables called pending allocation table. However, this approach needs a synchronisation procedure to ensure that the allocation tables of the nodes are always up to date. Hence an additional and reliable message exchange is needed among all nodes. These messages increase the overhead in the network. Also they cause difficulties in case of network partitioning and merging.

Another approach is used in [8]. Here, the authors propose a new protocol utilizing multiple disjoint allocation tables, which means that they split the global allocation table among all nodes. Here the already nodes can fast assign a unique address to the requester one without asking other nodes for permission. Problems can arise if a node crashes, because this crash would also destroy a part of the address space. A synchronisation procedure is needed to detect the resulting holes in the address space. This again increases the protocol overhead.

Summarizing, to integrate Ad hoc networks with infrastructure ones, a fast assignment of a unique address should be achieved, an address change due to an address conflict should be minimized and a high overhead of the already proposed protocols should be reduced.

### **III- LHA PROTOCOL**

The Logical Hierarchical Addressing (LHA) protocol is a novel stateful approach. It enables every node in the network to quickly assign a unique address to new nodes. It synchronizes important parameters among nodes and implements a DAD mechanism by using the routing protocol messages. It can solve address conflicts in case of partitioning and merging of the network.

### 3.1 BASIC IDEA

In our approach there are two roles for a node in the network, predecessor and successor. Every node has a number  $N$  of free addresses. It can assign one of its free addresses to a requester node. This requester node chooses its Address Agent (AA) node from its neighbours, which will provide the address to it. Per definition the AA node is the predecessor of the requester which is called the successor. However, every predecessor could have  $N$  direct successors, whereas every successor has only one predecessor. Thus, an address hierarchy is built in the network. Every node maintains a table containing information on addresses and parameters of its successors and its predecessor, this table is called hierarchical table. The parameters are as follows:

- $HI$  is the number of levels in the hierarchy of the node.
- $AfA$  is the number of currently available addresses at this node.
- $CAPS$  is the number of all (direct and indirect) successors of this node.
- $Seq$  is the sequence number of the current requester. This number will be calculated by AA node from the equation in (3) and will be saved by the successor.
- $MS$  and  $MP$  are the number of lost successors and predecessors respectively.

### 3.2 ALGORITHM DESCRIPTION

To understand the algorithm we will describe it by starting with a single node initiating the configuration process. Then, other nodes can join and leave the network.

In the initial case, when the node is the first one in the network, it observes the medium for beacon messages of other nodes. When a specific timer expires without receiving any messages, it configures itself with the first address from the address space of the network. Then, the node sets its parameters as in equation (1):

$$\{HI = 1, AfA = N, CAPS=0, Seq=0, MS=0, MP=0\} \quad (1)$$

When a new node enters the network and receives the beacon message, it broadcasts an Address Agent solicitation ( $AA\_sol$ ) message to its neighbours. Upon receiving this message by the neighbour nodes each of them responses by sending an Address Agent replay ( $AA\_rep$ ) message including its  $AfA$  parameter. This parameter indicates that the node can assign directly a unique address to the requester. As soon as the new node receives the  $AA\_rep$  messages, it selects its AA node depending on some conditions. If the  $AfA$  of some or all sender nodes is bigger than 0, then the AA node address is the smallest one of them. Elsewhere it chooses the smallest node address from all the

sender nodes. After the selection process the new node sends to the selected sender an Address Agent selection (*AA\_sel*) message. This mechanism solves the case when several *AA\_rep* messages are received from several neighbours.

Upon receiving the *AA\_sel* message by the selected node, two cases are possible. If the AA node has free addresses, then it uses equation (2) to calculate the new address (*New\_Ad*) of the requester node. In the equation, the term *Root\_ad* is the address of the root node in the network. *Agent\_ad* is the address of the AA node. In the same time the AA node updates the parameters of requester node depending on its own parameters as depicted in equations 3.

$$New\_Ad = Root\_Ad + N*(Agent\_Ad - (Root\_Ad - 1)) - AfA + 1 \quad (2)$$

$$\{HI (successor) = HI (AA) + 1, AfA (successor) = N, CAPS (successor) = 0, \quad (3)$$

$$Seq (successor) = N - AfA (AA), MS(successor) = 0, MP(successor) = 0\}$$

At last, the AA node updates its own parameters as the equations (4). Then the AA node sends an Address Agent confirm (*AA\_conf*) message including the unique address and all information needed to the new node. From this information the new node knows the address of its predecessor, which is the AA node in this case. The *AA\_conf* message enables the requester to build its hierarchical table which describes the node hierarchy in the network. The new node is now able to start to communicate.

$$\{HI (new) = HI (old), AfA (new) = AfA (old) - 1, CAPS (new) = CAPS (old) + 1, \quad (4)$$

$$Seq (new) = Seq (old), MS (new) = MS (old), MP (new) = MP (old)\}$$

In the second case, if the AA node has no free address available, it sends an Address Agent Address request (*AA\_A\_req*) message to all its neighbours. Every node which receives this message checks its *AfA* parameter. If it has a free address, then it responds by sending an Address Agent replay (*AA\_A\_rep*) message. If the node has no free addresses, it will forward the *AA\_A\_req* message to its neighbours. The AA node selects one of the responding nodes according the described conditions above. Then it sends an Address Agent Address selection (*AA\_A\_sel*) message to the chosen node. This node proceeds with the same steps as described before to calculate the address and to set the parameters derived from the equations (2), (3) and (4). Then it sends a regular *AA\_conf* to the AA node, which in turn forwards the message to the new node. From the information of this message the new node knows the address of its predecessor, which is not the AA node in this case.

The described algorithm solves the problems of partitioning and merging of networks. In case of a partitioning, the nodes continue to assign the unique addresses. Subsequently, if the two parts merge again, there are no duplicated addresses. Because all nodes in one part assign a unique addresses, which will not be duplicated in the other part according to the conditions and equations described above.

LHA protocol benefits from ongoing research in routing protocols. Every node using the control messages of these protocols can detect crashed nodes. Then, notification messages will be broadcast to all nodes in MANET to update their hierarchical tables. Thus the protocol overhead is reduced at minimum level.

## V- CONCLUSION

We have presented a hybrid address configuration protocol (LHA) for nodes in a MANET. The protocol enables MANET nodes to be configured with an address very rapidly if they join a network. The proposed solution minimizes the control messages by using the ongoing routing protocol. Thus, the protocol is suiting the interworking aspects in hybrid networks.

Currently, we develop a mathematical model to evaluate the performance compared to other proposals. In addition simulations to study our approach and compare it with others are underway.

### References:

- [1] Basagni, S., Conti, M., Giordano, S., Stojmenovic, I.: Mobile Ad Hoc Networking. IEEE, Inc. ISBN 0-471-37313-3, 2004.
- [2] Droms, R., Dynamic host configuration protocol, RFC 1541, October, 1993.
- [3] Weniger, K., Zitterbart, M.: Address Autoconfiguration in Mobile Ad Hoc Networks: Current Approaches and Future Directions. IEEE Network, Vol. 18, No. 4, 2004.
- [4] Perkins, E.C., Royer, E.M., and Das, S.R.:IP Address Auto-configuration for Ad-hoc Networks. Internet Draft, 2001.
- [5] Vaidya, N.H.: Weak Duplicate Address Detection in Mobile Ad Hoc Networks. MOBIHOC'02, EPFL Lausanne, Switzerland, 2002.
- [6] Weniger, K.: Passive Duplicate Address Detection in Mobile Ad Hoc Networks. IEEE WCNC 2003, New Orleans, LA, 2003.
- [7] Sanket, N., Ravi, P.: MANETconf: Configuration of Hosts in a Mobile Ad Hoc Network. IEEE INFOCOM 2002, New York, 2002.
- [8] Mohsin, M., Prakash, R.: IP Address Assignment in a Mobile Ad Hoc Network. IEEE MILCOM 2002, Anaheim, 2002.

### Authors:

Dipl.-Ing. Ausama Yousef  
Dipl.-Ing. Mohamed Abd rabou Kalil  
TU Ilmenau, Faculty for IA, Gustav-Kirchhoff-Str. 1 Informatikgebäude  
D-98693 Ilmenau  
Fax: +49 (0)3677 69 1220  
E-mail: ausama.yousef@tu-ilmenau.de  
E-mail:mohamed.abdrabou@tu-ilmenau.de

M. A. A. Kalil / N. Zamin Khan / H. Al-Mahdi / A. Mitschele-Thiel

## **Evaluation and Improvement of Queuing Management Schemes in Multihop Ad hoc Networks**

### **ABSTRACT**

Recent researches have observed inadequate performance in multihop ad hoc networks. Many factors lead to poor performance. Among those factors is an appropriate queue management scheme which is fair for all types of traffic. In multihop scenario, increasing the number of hops decreases the effective bandwidth. Longer hop flows suffer from unfairness problem as compared to shorter flows. This fact is undesirable in multihop ad hoc networks. In this paper, we evaluate different queueing management schemes and forwarding strategies for multihop networks. We show that queue management scheme, at intermediate nodes, have a great effect in limiting the performance of longer hop flows. Also, we propose an idea to improve the performance of multihop flows by fairly sharing the available buffer at each node.

### **I. INTRODUCTION**

Multihop ad hoc network is a collection of mobile wireless nodes that dynamically form a network without any pre-existing infrastructure. This type of network is established when the source and destination are not in the transmission range of each other. However, in multi-hop networks, increasing the number of hops leads to decrease in bandwidth. This leads to unfairness problem because longer hop flows suffer from low throughputs when compared with shorter hop flows. Many factors let this problem happen. Hidden and exposed terminal problems are one of those factors. Another important factor that causes this problem is the link layer buffer management scheme at the intermediate nodes. Current queueing management schemes do not take into account the number of hops a packet has traversed when inserting it into the link layer queue. This will lead to unfairness of flows spanning multiple hops. According to a lot of routing protocols, when a packet has to be forwarded by an intermediate node, the routing process sends the packet down to the link layer, which inserts this packet into the

interface queue (IFQ). Generally, the link layer uses a drop tail queue management scheme in which newly arriving packets are dropped if the queue is full, regardless of the number of hops the packet has already traversed.

Most existing queue management schemes are not suitable for multihop ad hoc networks. They have been designed for controlling congestion in the Internet routers. For example, RED [1] was designed to control congestion and avoid queue buildups at Internet routers by measuring the average queue length. RED has some drawbacks, for example, RED fails to differentiate shorter flows and longer flows while dropping the packets, thus presenting unfairness. In [2], the authors have presented several queuing schemes at intermediate relay nodes for achieving fairness in multihop wireless networks. They suggested that each individual source should have a separate queue at all relaying nodes. Maintaining a separate queue for each individual source is a very difficult task. Recently, in [3], the authors developed analytical models for hop by hop congestion control in ad hoc networks and proposed layer 2 congestion control mechanisms for controlling the traffic load generated at the source nodes. However, the authors did not mention the role of link layer buffers. In [4], the authors have addressed the unfairness problem in multihop mesh networks by proposing an Inter-tap fairness algorithm in which the nodes exchange channel usage information and decide their maximal channel access times. However, the authors in [4] did not consider the link layer buffer management. In contrast, the main focus of our work is to evaluate the role of queue management schemes in the intermediate node and to present a queueing scheme to enhance the performance of IFQ in multihop ad hoc networks.

The rest of the paper is organized as follows. In section II, overview of queueing management schemes are discussed. Next, a description and analysis of unfairness problem is presented in section III. In section IV, the proposed scheme is presented. We conclude the paper in section V.

## **I. OVERVIEW OF QUEUE MANAGEMENT SCHEMES**

In multihop ad hoc network, nodes may have to buffer messages and in case of congestion decide which messages to drop from its queue. They also have to decide which messages to forward to another node. In this section we describe the different queueing management schemes used in this paper for the evaluation.

DropTail - is the most commonly used algorithm in the current Internet gateways. The main idea of this algorithm is to drop packets from the tail of the full queue buffer. Its

main advantages are simplicity, suitability to heterogeneity and its decentralized nature. Random early drop (RED) - RED [1] is a queue management scheme that monitors and controls buffer occupancy. RED detects congestion by monitoring the average buffer size of the router. When the average buffer size is larger than the first threshold (minth) but lower than the second threshold (maxth), the incoming packets are dropped with probability  $P$ , which increases linearly as the average buffer size increases. When the average buffer size exceeds the second threshold (maxth), the router drops randomly chosen packets from within the buffer with probability one.

Fair queuing (FQ) [5] - The main idea of this algorithm is to divide the router buffer into sub-queues, one for each incoming traffic source (per-flow queuing). Then the router serves packets in the round-robin fashion (packet-by-packet round-robin scheduling). This algorithm, however, assumes the packet size is constant, and thus it fails to provide throughput fairness when packets have different sizes.

Deficit round robin (DRR) [6] - this algorithm was proposed to approximate the performance of FQ using a less complex computational structure. DRR serves sub-queues in the round-robin fashion. For each subqueue, a deficit counter (in bytes) is assigned. In each round of service, the deficit counter is incremented by a quantum (in bytes). Each sub-queue, when served, is allowed to send its packets one by one if the packet size is smaller than the deficit counter. The deficit counter is decremented by the packet size after a packet is sent. When the deficit counter is depleted, the DRR scheduler moves to the next sub-queue.

Virtual Queue Algorithm (VQ) [7] - In this scheme, a virtual queue is added in the link with the same arrival rate as the real queue. However, the capacity of this queue is smaller than the capacity of a real queue. When the virtual queue drops a packet, then all packets already enqueued in the real queue as well as all of the new incoming packets are marked until the virtual queue becomes empty again. The fixed size FIFO virtual queue seems to be a weakness of this algorithm.

Random Exponential Marking Algorithm (REM) [8] - this Algorithm is a technique for congestion control, whose main aim is to achieve a high utilization of link capacity, scalability, negligible loss and delay. It attempts to match user rates to network capacity while clearing buffers (or stabilize queues around a small target), regardless of the number of users.

From the above description of queue management schemes, we can conclude that none of those schemes take into account the number of hops the packet takes to reach

its destination. They have been designed for controlling congestion in the Internet routers not working over multihop networks.

## II. SIMULATION

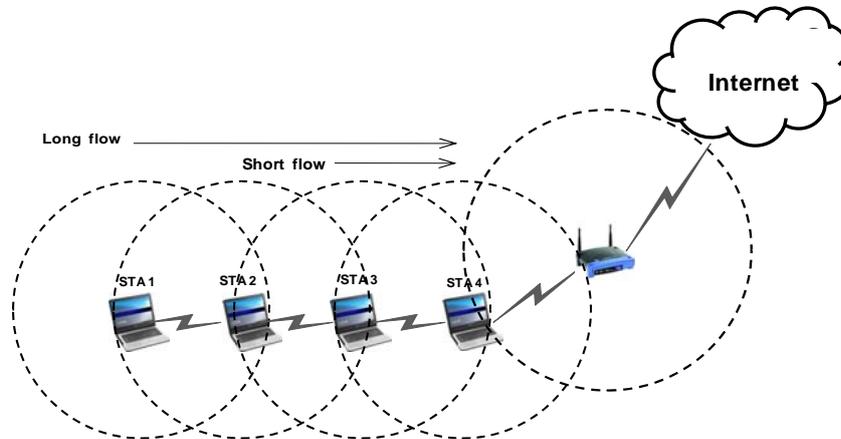


Figure 1. Simple multihop ad hoc scenario

In this section, we will illustrate the existing unfairness problem in multihop ad hoc networks. As we mentioned in section I, many factors let the unfairness problem happened. Our evaluation only focuses in link layer buffer management. To do that, we consider a four station communicate with each other using the legacy IEEE 802.11 based interface. Access Point (AP) serves as a gateway to connect other station to the internet.

In order to see the effect of link layer buffer queue, we use the scenario illustrated in Figure 1. In this scenario, STA1, STA2, STA3 and STA4 try to connect to the internet. But, only STA4 is in the transmission range of the access point. So, the other station will use it as a gateway. We assume that all stations use IEEE 802.11 DCF operating at 2 Mbps with RTS-CTS handshake enabled. By implementing this scenario in Network Simulator (NS2) [9] and monitoring the data packets sent from each station, we conclude that the IFQ plays an important role in the performance of multihop networks and leads to unfairness problem. For example, STA1 transmitted 4576 CBR packet, only 1276 are forwarded by STA2, while the remaining packets are dropped due to lack of buffer space at the link layer. STA2 sent 4148 CBR packet, only 914 are forwarded by STA3. Figure 2, illustrates that. From this figure, we conclude that IFQ plays an important role in the performance of multihop ad hoc networks. Also, we simulate this scenario with different queueing policies such as RED, REM, DRR and VQ and we conclude that the effect of each of those queueing management schemes is approximately the same.

They fail to present fairness to longer hops when it is compared with shorter hops packets.

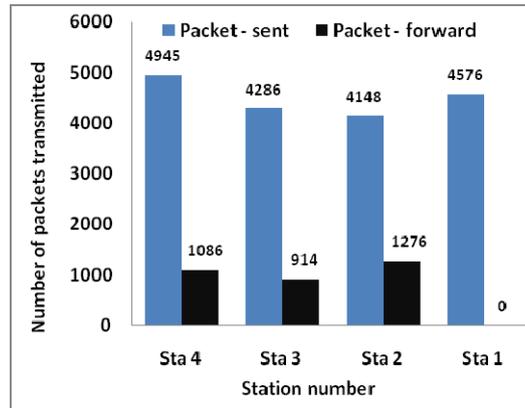


Figure 2. Packet transmitted by each station MAC layer

### III. OUR PROPOSED SCHEME

In this section, the proposed scheme will be presented. The packets received to each node will be classified as follow:

- Packets with a large number of hops to reach its destination (class-1) and
- Others with a small number of hops (class-2).

All types compete for the available finite buffer space at nodes. The number of hops can be determined from the routing table at each node. In this paper, without loss of generality, it is assumed that the class-1 packets are more important than class-2 packets, but this fact may change according to the load of the traffic.

#### Dynamic Buffer

In this scheme, each node has a finite buffer with fixed size of  $C$  places where each one can hold one packet. The  $C$  places are divided into two regions, by threshold  $k$ , where  $k < C$ . As mentioned before, each node receives two different classes: class-1 and class-2. If a class-2 packet arrives at the node and the buffer occupancy is less than  $k$ , then the packet can be accepted. If the buffer occupancy is greater than or equal to  $k$ , then the class-2 packet can be accepted with probability  $\alpha$ . Otherwise, the packet is rejected and lost. If a class-1 packet arrives at the node and the buffer occupancy is less than  $C$ , the class-1 packet is accepted, otherwise, it is rejected. Here, we rank the size buffer for class-1 and class-2 types as follows:

- Rank-1: assigns buffer for both classes.
- Rank-2: assigns buffer for class-1 and assigns buffer for class-2 with probability  $\alpha$ .

For example, a class-2 will first try Rank-1, if unavailable, it tries Rank-2, if unavailable, it will be rejected and lost. The value of  $\alpha$  can be adjusted depending on the

traffic load. For example, if we put  $\alpha=0$ , we give the class-1 packets a good chance to find a free space in the buffer. In other word, if the class-1 traffic is high, we can make the value of probability  $\alpha$  close to zero.

#### IV. CONCLUSION

In this paper, we have evaluated different queueing management schemes in multihop ad hoc networks. From the evaluation presented in this paper, it is evident that the queueing management schemes in IFQ at intermediate relay ad hoc nodes, can greatly impact the performance of longer hops over shorter hops in multihop ad hoc scenario. We proved that by evaluating different queueing management schemes such as droptail, RED, VQ, DRR and REM.

We evaluated the performance of those queueing management schemes through a simulation study using network simulator 2 (NS2) in a multihop ad hoc network scenario. This scenario consists of five nodes in linear topology. Also, we presented a simple and effective solution for improving the performance of multihop flows by fairly sharing the available buffer at each intermediate node among all the active source nodes whose flows are being forwarded. Evaluation of the proposed scheme using ns2 and mathematical methods will be done in the future work.

#### References:

- [1] Floyd, S., and Jacobson, V., "Random Early Detection gateways for Congestion Avoidance," IEEE/ACM Transactions on Networking, vol.1, no.4, August 1993.
- [2] Jun, J. and Sichitiu, M. L., "Fairness and QoS in multihop wireless networks," in Proc. Of the IEEE Vehicular Technology Conference (VTC), vol 5, pp 2936- 2940, October 2003.
- [3] Yung, Yi and Shakkottai, S., "Hop-by-hop congestion control over a wireless multi-hop network," In Proceedings of IEEE INFOCOM '04, vol 4, pp 2548 - 2558, March 2004.
- [4] R., Conti M., and Gregori E., "Mesh Networks: Commodity Multihop Ad Hoc Networks," IEEE Communications Magazine, Vol. 43, No. 3, March 2005.
- [5] Demers, A., Srinivasan, K., and Shenker, S., "Analysis and simulation of a fair queueing algorithm". SIGCOMM Symposium on Communications Architectures and Protocols, pages 1–12, Sep 1989.
- [6] Shreedhar, M. and Varghese, G., "Efficient Fair Queueing Using Deficit Round Robin" IEEE/ACM Transactions on Networking, vol 4, pp375 - 385, Jun 1996.
- [7] Kunniyur, S. and Srikant, R., "Analysis and design of an adaptive virtual queue (AVQ) algorithm for active queue management". ACM SIGCOMM Computer Communication Review, vol 31, pp 123 – 134, October 2001.
- [8] H. Low, S., Athuraliya, S., H. Li, V., and Yin, Q., "REM: Active Queue Management" IEEE Network journal, vol 15, no. 3, pp 46-53, 2001.
- [9] Network Simulator (NS-2), Available at <http://www.isi.edu/nsnam/ns/index.html>.

#### Authors:

Mohamed Abd rabou Ahmed Kalil  
Nadir Zamin khan  
Andreas Mitschele-Thiel  
Fachgebiet Integrierte Hard- und Softwaresysteme  
Technische Universität Ilmenau  
98693, Ilmenau  
Phone: +49 (0)36 77-69 13 79  
Fax: +49 (0)36 77-69 12 20  
Email: {mohamed.abdrabou, nadir.zamin-khan, mitsch}@tu-ilmenau.de  
Hassan Al-Mahdi  
Faculty of Computers and Informatics  
Suez Canal University  
Ismailia, Egypt  
Email: hassanwesf@yahoo.com

Mark Ritzmann

## **Scientific visualisation on mobile devices with limited resources**

### **ABSTRACT**

The article focusses on a base, that makes scientific visualisation possible on mobile devices with limited resources. The aim is to gain access to databases, produce a visual representation of selected content and display it to a viewer on a limited capability device. Hardware addressed are low end appliances, that have a limited radio network connection at their disposal, however in its CPU speed, graphics performance and display size as well as available RAM are restricted (cell phones and PDA). To enable the production and representation of interpretable<sup>1</sup>[1] graphics in this sphere, existing technologies like web applications and the platform Java 2 Micro Edition are combined [2]. The introduced architecture permits mobile access to data and the visualisation on thin-clients (web front end) as well as limited capability devices using MIDP. Especially the J2ME APIs were looked at considering their applicability in the context. Applications arise in geoinformatics, civil engineering, mechanical engineering and medicine.

### **MOTIVATION**

In addition to the application as a communication instrument mobile devices are used increasingly as a "computer system on the Internet". The mobile phone allows access independent of place to data and accessibility with the functionality of computers. Applications for mobile devices profit from these qualities and open thus perspectives for available software (e.g. e-mail, PC-synchronized organizer). Nevertheless, the integration of mobile devices as front end of a client server architecture develops in the scenario of the visualisation difficultly. Because of limited resources (memory, arithmetic achievement and graphics system) and missing programming interfaces (APIs) for this platform the solutions which still allow the access to the data bases and a high-quality visualisation must be found.

---

<sup>1</sup> The viewer should understand the context of the real world from the graphic and connections and parameters in the value distribution can infer.

## 1 J2ME and the MID-Profile - Client side

The platform Java2 Micro edition addresses by configurations (CDC, CLDC) and profiles (MIDP, PDAP) different device sections and devices. Optional packages serve as an interface for functionalities specific for devices and/or specific for manufacturer. Covered by the aimed application scenario J2ME APIs offers for the network communication (generic connection framework) as well as APIs for the realisation of user interfaces and the issue of graphic representations (LCDUI, GameAPI). Nevertheless, the available graphic system is limited to 2D graphic, above all, game development (collision recognition, use of sprites etc.) and it is usually not hardware accelerated. Support of the optional package "mobile 3D-API" or NVIDIA-GPUs on mobile phones don't yet belong to the standard. Therefore, necessary visualisation technologies like vertex, transformation, rasterization, texturing, lighting etc. are not available at all or very much limited.

## 2 System architecture

The use of an application server is necessary for different reasons:

- Mobile devices do not support the communication with databases and require a protocol converter / adaptation layer
- The unsafe network connections of mobile devices (reason for the name "limited connected device") could lead to interrupted connections and the release of requested resources in the back end must be guaranteed by an additional layer.
- Because of the possible client count the resource requirement with the database server would raise needlessly and might become not manageable.
- The direct connection of the client with the database would require in case of changes in the database scheme an adaptation of the client software.
- MIDP 1 capable devices are limited to have at least http-protocol support.

Figure 1 shows the favoured three tier architecture. The application server is realised as a servlet based web application. Servlets can serve as a data source for contents in HTML pages (text, images, SVG objects) and are also localizable by mobile devices [6]. The implementation of the application server as a J2SE component allows the access to available APIs and standard extensions of the platform: database access JDBC [4]; compression and conversion of graphics - ImageIO; visualisation OpenGL [7] or Java3D. To enhance the communication for MIDP 2 compliant devices, a socket connection can be used for a more interactive protocol than stateless http. Selected

images and measuring data are temporary stored server sided too, to avoid unnecessary calculations an database request with repeated or parallel request.

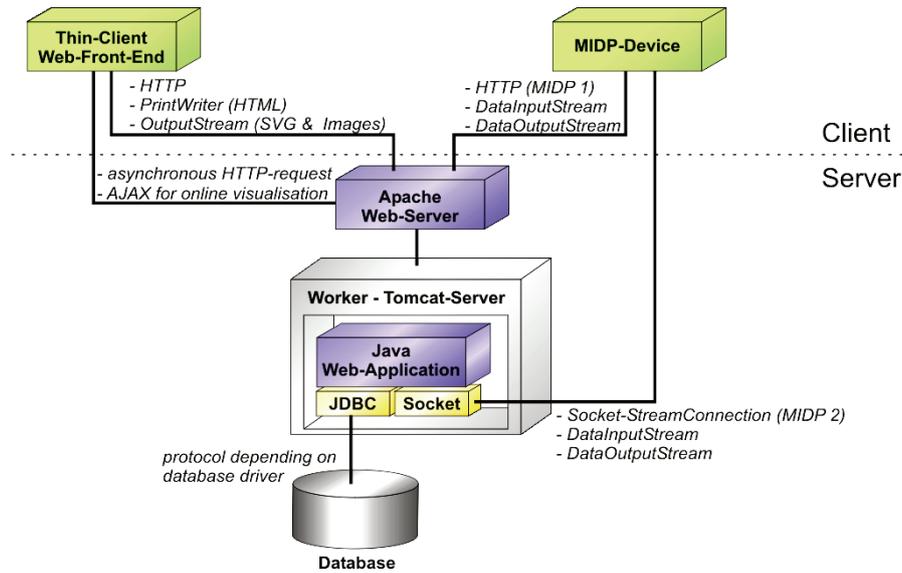


Figure 1: system model and protocols in the tree tier architecture

## 2.1 Web-Front-End

The visualisation within the scope of a web application occurs in form of interactive SVG [5], static images or their integration for compositions with SVG.

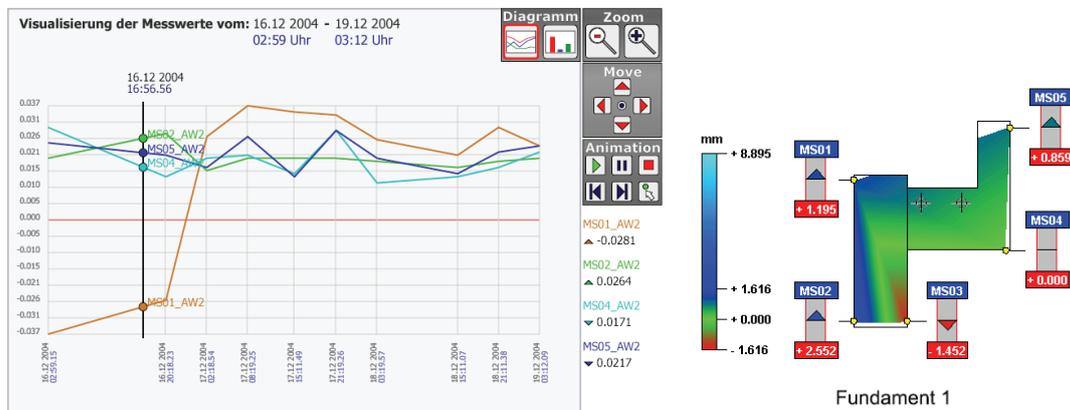


Figure 2: visualisation as interactive SVG (left) and embedded static png-image (right)

For SVG visualisation merely the required records are encoded as variables in JavaScript. The production of the final representation occurs through JavaScript client-sided. Therefore the data volume is reduced for the transfer and the desired visualisation technique can be changed by the user, without renewed communication with the server, interactively. This addresses mobile Internet access through a mobiles radio network. For online visualisation of frequently changing data AJAX/JavaScript is used to retrieve new data using asynchronous http-requests. This results in periodically refreshes without loading the whole page and non cacheable embedded elements. The response are either measuring data which are directly used to update the SVG-DOM or picture

references which are updated in the web page or SVG. Examples of interactions are animations of temporal course (linear interpolation), change of representation, observing of certain measuring times or the overlay of contour plots and ground plans on demand. The production of the static graphics is based on off-screen rendering using a 2D/3D graphic systems. Varied visualisation technologies are applicable through this. The contents of the frame buffer is grabbed and it is converted in a web-compliant format. As source for the picture data (img-tag, src-attribute) the rendering servlets URL is defined. 3D Graphic systems offer functions which would not be possible in SVG graphics or partly only by costly pre-calculations [9]. Besides, the transfer as an image limits the requirements for the client component and forms the interface for the visualisation on mobile devices with even more constrained resources.

## **2.2 Visualisation on mobile devices with limited resources**

Devices capable of J2ME dispose of a Microbrowser for the representation of HTML pages (required for over the air user initiated provisioning too). Because of the low display resolution, the supported media formats and the application model (if necessary service about keypad) the realisation requirements differ from their one thin client web application. For the data selection a combination of dialogues and graphics (LCDUI) is offered [3]. The visualisation occurs in form of server-sided generated graphics. Color reduction, scaling and tiling etc. can be done by filter (-chains) of the web application. The client is reduced to view images.

### **2.2.1 Applicability of the MIDP-API to the scientific visualisation**

**LCDUI:** Static background images like in games or the composition of the scene from returning tiles resisting in the MIDlet-suite (TiledLayer) are not related within the scope of the visualisation. The wanted scenario is insufficiently supported by the available APIs. Depending on the OEM's implementation, large images, or even with a resolution greater than the display, will not be supported. Therefore and for efficient scrolling large images must be divided into tiles. To allow the scroll about the partial pictures, several graphics must exist locally. The segmenting and the reconstruction were examined at the example by CAD graphics, academic visualisation and real pictures of the resolution 512x512. The derived parametres can be used for load balancing in the distributed system and optimisation of the representation quality on the client. Depending on contents and tile size the essential parameters which should be introduced here only at an example could be derived: Data volumes by transfer of the whole image:

CAD: JPEG 1x2 = ~78MB - JPEG 5125x512 = ~28KB - Break Even Point: 106x133

CAD: RGBZIP: 1x2 = ~15 MB – RGBZIP 512x512 = ~9KB – Break Even Point: 86x86

Among the rest, additional ones were looked the RGB difference in the whole image and the number of the individual parts (relevant for the function and efficiency of the cache algorithms, as well as load of the KVM). Statements about the footprint in memory can be derived only partly, because this strongly depends on the respective implementation. The use of the SVG-API was not considered, because this description would be suitable merely with CAD data. The topical MIDlet allows the smooth scrolling about large graphics and missing tiles are reloaded asynchronously (viewport first then borders).

**Generic Connection Framework:** As a protocol for the communication with the application server are available http (MIDP 1) and socket (MIDP 2). In the http protocol the restriction exists in the strict request response model. The client requests here for a number of partial pictures. For the realisation of Session-Tracking the Session-ID is added as a header to the response and encoded in inquiries of the client accordingly [8]. Sockets allow here a higher delicacy in the communication. After every transfer the connection can be used once more and topically required data if necessary be requested. This is especially relevant if the user during the transfer scrolls and earlier required partial pictures lie, in the meantime, beyond the viewport.

**Record Store Management System:** The application uses different caches for interactive visualisation. Sprite cache contains the tiles of the viewport and border areas. Memory-cache and RMS-cache are realised as a ring buffer more determinably of steady size. Not required tiles of the viewport, are filed at first in the RAM and later if necessary and available in the RMS with increasing latency while object reincarnation. The objects or parts of it (compressed image-data with tile localisation in RMS) reside alternatively either in one of the memories or must be loaded on the network. In the RMS entries of the same size whose contents are substituted cyclically are put on. This is conditioned by the fact that, deleted RMS records space is not always released immediately. Replacing the contents solves the problem. The maximum image size in bytes is determined by the server for all tiles and defines the maximum size of a record. The divergence between maximum size and average was likewise examined in the test.

### **3 Application Example**

The introduced system model orientates itself by a concrete application problem from the area of the geoinformatics. The company "Position Control GmbH" works among

other things by order of the "Deutsche Steinkohle AG" and supervises buildings in regions of topical or former conveyor sites. For this sensors are placed in objects which determine periodically transformations and temperatures and feed this into a database.



Figure 3: Possible operating surface for the MIDP application. Representation of a server-sided processed CAD file of the supervised object and overlapping with sprites for the eligible measuring points (on the left). Visualisation by contour plots (in the center) and a view with missing tiles marked as cross (on the right). The engineers can carry out by evaluation of the data an appraisal and initiate countermeasures in border-valued situations. To be able to analyze on site the topical data, the integration of mobile devices was necessary.

## 4 Results

Both attempts of distributed visualisation vary in their interactivity. Especially the pre-calculation of graphics reduces the client requirements and allows such an interpretation of the data also on mobile devices with limited resources. However, the implementation of applicable graphic representation and interaction models requires huge expenditure and is only partly supported by the available APIs.

## References

- [1] Heidrun Schumann, Wolfgang Müller. *Visualisierung*. Springer Verlag, 2000.
- [2] Roger Riggs et.al. *Programming Wireless Devices with the Java 2 Platform Micro Edition*. Addison Wesley, 2003.
- [3] Cynthia Bloch, Annette Wagner. *MIDP 2.0 Style Guides for the Java 2 Platform Micro Edition*. Addison Wesley, 2003.
- [4] Maydene Fisher et.al. *JDBC API Tutorial and Reference*. Addison Wesley, 2003
- [5] SVG - [www.w3.org/TR/SVG](http://www.w3.org/TR/SVG) Scripting - [www.w3.org/TR/SVG/script.html](http://www.w3.org/TR/SVG/script.html)
- [6] Jason Hunter, William Crawford. *Java Servlet Programming*. O'Reilly Associates, 1999.
- [7] Mason Woo, Jackie Neider, Tom Davis, Dave Shreiner. *OpenGL Programming Guide*. Addison Wesley, 2001.
- [8] Michael Junato Yuan, Ju Long. Track wireless sessions with J2ME/MIDP. JavaWorld.com, 2002.
- [9] Dr. Philip A. Mansfield. Common graphical object models and how to translate them to SVG. <http://www.svgopen.org>.

## Author:

Dipl.-Inform. (FH) Mark Ritzmann  
 FH-Schmalkalden, Am Schwimmbad  
 98574 Schmalkalden  
 Phone: 03683 688-4216  
 Fax: 03683 688-4297  
 E-mail: [m.ritzmann@fh-sm.de](mailto:m.ritzmann@fh-sm.de)

R. Brecht / A. Kraus / H. Krömker

## **Entwicklung von Produktionsrichtlinien von Sport-Live-Berichterstattung für Mobile TV Übertragungen**

### **1. Einleitung**

Telefonieren, im Internet surfen, Musik bzw. Videos abspielen, digital fotografieren – es gibt kaum eine Funktion, die nicht von Handys übernommen werden kann. Eine weitere Anwendung ist derzeit stark im Kommen: das mobile Fernsehen. Mobile TV auf dem Handy oder PDA ist ein Markt der Zukunft.

Mobile TV galt schon 2006 in der Fachöffentlichkeit als großer Trend und vor allem langfristige Marktanalysen sind bezüglich des Potentials von Mobile TV für den Massenmarkt optimistisch [1]. Den einschlägigen Studien und Nutzerbefragungen zufolge, werden als Content für Mobile TV hauptsächlich klassische TV-Inhalte erwartet. Vergleicht man den Erfolg der verschiedenen Inhalte im klassischen TV, so zeigt sich, dass vor allem Sportübertragungen stark nachgefragt werden. Dabei hat Fußball eine herausragende Stellung. Bester Beweis dafür ist, dass sich unter den 25 meistgesehenen Fernsehsendungen von 1992 bis 2006 23 Fußball-Übertragungen befinden [2].

Mit Mobile TV steht nun ein weiterer, viel versprechender Kanal für die Fußball-Live-Berichterstattung zur Verfügung. Die beiden hierzulande konkurrierenden Rundfunkstandards DMB und DVB-H weisen jedoch qualitative Unterschiede zu klassischem Fernsehen auf. Deshalb stellt sich die Frage, ob und wie sich die Einschränkungen in der Übertragungsbandbreite und der Bildwiederholfrequenz sowie der Auflösung am Endgerät, auf die Darstellung und Wahrnehmung der TV-Inhalte auswirkt.

Um festzustellen, ob die klassische TV-Produktion von Sport-Live-Berichterstattung auch für Mobile TV geeignet ist, wurde eine Studie an der Technischen Universität Ilmenau in Zusammenarbeit mit der Plazamedia GmbH, einem der größten Produzenten

von Sportsendungen, durchgeführt. Dazu wurde die Wahrnehmbarkeit und Erkennbarkeit von Sport-Live-Berichterstattung von Fussballspielen auf mobilen Endgeräten in einem Nutzertest evaluiert. Die Testclips wurden hinsichtlich der für die Wahrnehmung bestimmenden Produktionsparameter, wie z.B. Kameraeinstellungen, Grafikeinblendungen, Ton variiert.

## 2. Evaluation

### 2.1 Operationalisierung der Usability-Maße

Im Mittelpunkt der Evaluation von Sport-Live-Berichterstattung steht die Wahrnehmbarkeit und Erkennbarkeit der medialen Inhalte auf mobilen Endgeräten. Für das Testdesign müssen die Maße der Usability (Effektivität, Effizienz und Zufriedenheit) entsprechend operationalisiert werden [3].

Die Testclips, die typische Szenen eines Fußballspiels zeigen, werden ohne jegliche Zusatzinformationen und Möglichkeiten zur Interaktion auf dem mobilen Endgerät angeboten. Die Nutzung erfolgt also passiv. Daher haben nur die Effektivität und die Zufriedenstellung Einfluss auf die Usability. Die Effizienz, welche die Effektivität ins Verhältnis zum Aufwand stellt, hat bei einer passiven Nutzung keinen Einfluss. Abbildung 1 zeigt die Operationalisierung der Usability Maße.

| Usability   |                  |   |
|---|------------------|---|
| <b>Effektivität</b> <ul style="list-style-type: none"> <li>• Erkennbarkeit</li> <li>• Klarheit</li> <li>• Lesbarkeit</li> <li>• Verständlichkeit</li> </ul> | <b>Effizienz</b> | <b>Zufriedenstellung</b> <ul style="list-style-type: none"> <li>• Emotionaler Faktor</li> <li>• Grad der Begeisterung</li> <li>• Subjektive Wertschätzung</li> <li>• Optimales Zusammenspiel</li> </ul> |

Abbildung 1: Operationalisierung der Usability Maße

#### Effektivität

Die Maße der Effektivität setzen „...die Ziele oder Teilziele des Benutzers ins Verhältnis zur Genauigkeit und Vollständigkeit, mit der er diese Ziele erreichen kann.“ [3].

Hier entsprechen diese Ziele der Möglichkeit, die dargestellte Information wahrzunehmen. Der Teil 12 der ISO 9241 beschäftigt sich mit der

Informationsdarstellung und definiert die charakteristischen Eigenschaften, die bei der Gestaltung visueller Information beachtet werden müssen [4]:

- Klarheit
- Unterscheidbarkeit
- Kompaktheit
- Konsistenz
- Erkennbarkeit
- Lesbarkeit
- Verständlichkeit

Um die Darstellung von Szenen der Fußballübertragung zu bewerten, wird die Effektivität über die Erkennbarkeit, Klarheit, Lesbarkeit und Verständlichkeit der dargestellten Information operationalisiert:

- Erkennbarkeit: Aufmerksamkeit des Benutzers wird zur benötigten Information gelenkt
- Klarheit: Informationsgehalt wird schnell und genau vermittelt
- Lesbarkeit der in der Sportübertragung üblichen Grafikeinblendungen (Einblendung von Karten und Auswechslungen und Spielstand / -zeit)
- Verständlichkeit: Übersicht über das Spielgeschehen

### **Zufriedenstellung**

Die „Maße der Zufriedenstellung beschreiben das Ausmaß, in dem Benutzer von Beeinträchtigungen frei sind, und ihre Einstellung zur Nutzung des Produkts“ [3].

Die Zufriedenstellung ist also ein emotionaler Faktor, der den Grad der Begeisterung widerspiegelt. Sie zeigt die subjektive Wertschätzung des Produkts, die Zufriedenheit bei der Nutzung und den Grad, in dem bestimmte Ziele der Usability als erfüllt angesehen werden.

Für die Zufriedenstellung wird angenommen, dass Sie dann besonders hoch ist, wenn der Nutzer den dargestellten Inhalt gut erkennen kann. Die Verständlichkeit dagegen wird durch weit gefasste Kameraeinstellungen gefördert, welche möglichst viel vom Spielgeschehen zeigen. In der Summe ist die Zufriedenstellung das optimale Zusammenspiel der sich entgegenstehenden Faktoren Erkennbarkeit und Verständlichkeit.

## **2.2 Testdesign und Ergebnisse**

Zur Evaluation von Sport-Live-Berichterstattung auf mobilen Endgeräten werden zwei aufeinander folgende Tests durchgeführt. Dabei wird das TV-Signal entsprechend den Mobile TV Rundfunkstandards umgewandelt und auf einem mobilen Endgerät bewertet.

Folgende Forschungsfragen bilden die Grundlagen der Evaluation:

- Ist die klassische TV-Produktion von Sport-Live-Berichterstattung auch für Mobile TV geeignet?
- Wie sieht eine mediengerechte Produktion für Mobile TV aus?

### **2.2.1 Test 1 - „Erkennbarkeitstest“**

#### **Ziel**

Der erste Test untersucht die Erkennbarkeit der klassischen TV-Produktion auf einem mobilen Endgerät. Das Ziel des Tests ist es, herauszufinden, ob das entsprechend der Mobile TV Rundfunktechnologie umgewandelte TV-Signal auch auf einem mobilen Endgerät erkennbar ist, bzw. bei welchen Produktionsparametern die Erkennbarkeit nicht mehr gewährleistet ist.

#### **Testdesign**

Die Effektivität (abhängige Variable) wird über die Erkennbarkeit, Klarheit und Lesbarkeit operationalisiert [5]. Die unabhängigen Variablen bilden beim ersten Test die Produktionsparameter Einstellungsgröße (Totale, Halbtotale, Halbnah, Nahe), Grafik und Ton.

Die Textclips bestehen aus einzelnen Szenen eines Fussballspiels, die hinsichtlich der Produktionsparameter Einstellungsgröße (Totale, Halbtotale, Halbnah, Nahe) variieren. Die Einstellungen werden in 5-10 Sekunden langen Clips gezeigt.

Der Test wurde im Usability Labor der Technische Universität Ilmenau mit zehn Testpersonen durchgeführt.

#### **Ergebnisse**

Im Einzelnen zeigt das Ergebnis der Evaluation, dass die Erkennbarkeit mit nahen Einstellungen steigt. Der erste Test macht deutlich, dass die Erkennbarkeit von der nahen Einstellung über die Halbnah und die Halbtotale zur Totalen hin stetig sinkt. Die

User bezeichnen aber allein die Totale als nicht bzw. sehr schlecht erkennbar. Auch die Grafikeinblendungen (z.B. Spielstand) sind nicht geeignet für Mobile TV, da die Erkennbarkeit nicht gegeben ist. Der Produktionsparameter Ton ist dagegen problemlos auf das neue Medium übertragbar.

So beschränkt sich also der Änderungsbedarf der Produktion auf die totale Kameraeinstellung und die Grafikeinblendungen.

### **2.2.2 Test 2 - „Verständlichkeit und Zufriedenstellung“**

#### **Ziel**

Die Ergebnisse des ersten Tests bilden die Grundlage des zweiten Tests. Der zweite Test sucht nach Alternativen für die schlecht oder nicht erkennbaren Produktionsparameter. Dazu kommt neben einem Verständlichkeitstest auch ein Zufriedenstellungstest zum Einsatz.

#### **Testdesign**

Im zweiten Test werden zwei abhängige Variablen untersucht. Einerseits die Effektivität operationalisiert über die Verständlichkeit, andererseits die Zufriedenstellung der Benutzer mit verschiedenen Szenen der Fußballübertragung. Die unabhängige Variable wird, wie im ersten Test, durch die Produktionsparameter Einstellung gebildet. Diese werden hier jedoch nicht isoliert dargestellt, sondern im Kontext von typischen Spielsituationen wie etwa Tor oder Eckstoß.

Um festzustellen, wie eine mediengerechte Produktion für mobile Endgeräte aussehen kann, werden vier verschiedene Varianten der Darstellung für die im ersten Test als schlecht bzw. nicht erkennbar eingestuften Produktionsparameter evaluiert. Dabei werden die Varianten für die Totale jeweils im Spielzusammenhang als komplette Szene gezeigt. Die Schnitte zwischen den Einstellungen werden beibehalten. Die Varianten für die Grafik unterscheiden sich allein in der unterschiedlichen Größe der eingeblendeten Grafik. Alle Varianten werden aus dem TV-Basis-Signal gewonnen, da dieses bereits in der Produktionsumgebung der Sendeabwicklung und des Live-Encodings vorliegt.

Dieser Test wurde ebenfalls im Usability Labor der Technische Universität Ilmenau mit zehn Testpersonen durchgeführt.

### Variante „klassisch“

Die Variante „klassisch“ ist unverändert und dient als Basis für den Vergleich. Hier wird das Material, welches im Seitenverhältnis 16:9 produziert wird, mit dem Letterbox Verfahren in das Seitenverhältnis 4:3 konvertiert. Dabei werden die Seitenränder an die Zielbildgröße angeglichen und der freie Raum oberhalb und unterhalb des Bildes mit Schwarzen Balken versehen.



Abbildung 2: Darstellungsvariante "klassisch"

### Variante „4 zu 3“

Diese Variante entsteht wie die Variante „klassisch“, nur werden hier der obere und der untere Rand des Videos an die Zielbildgröße angeglichen. Dabei verliert man den Bildinhalt des rechten und linken Bildrandes. Dies ist aber zu vertreten, da die Hauptaktion immer in der Bildmitte stattfindet. Diese Bildveränderung entspricht einer Vergrößerung auf 133% der Originalgröße.



Abbildung 3: Darstellungsvariante "4 zu 3"

### Variante „Pan & Scan“

Diese Variante entsteht aus der Variante „klassisch“. Hier werden für alle totalen Einstellungen der Szene Veränderungen vorgenommen. Alle näheren Einstellungen

werden nicht verändert. Die Totale wird auf 150% vergrößert und der Bildausschnitt dem Spielgeschehen nachgeführt.

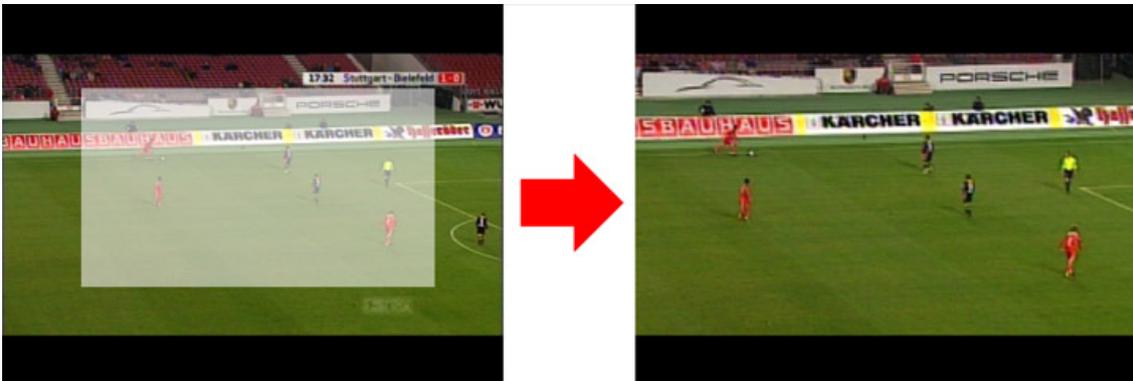


Abbildung 4: Darstellungsvariante "Pan & Scan"

### Variante „Pan & Scan Nah“

Die Variante „Pan & Scan Nah“ wird aus der Variante „4 zu 3“ generiert. Dies geschieht nach demselben Verfahren wie bei „Pan & Scan“, hier mit einer Vergrößerung auf 175% der Originalgröße. Da die Variante „4 zu 3“ schon um den Faktor 1,33 gegenüber dem Original vergrößert ist, entspricht die Gesamtvergrößerung ca. 230%.



Abbildung 5: Darstellungsvariante "Pan & Scan Nah"

### Grafik Varianten

Bei der Grafik wird die Originalgrafik für die Bauchbinden (Einblendung von Karten und Auswechslungen) und Spielstand / -zeit in der Größe variiert. Die Variante 1 ist bei der Grafik die Originalgröße und dient als Referenz zur Einordnung der Ergebnisse. Die Grafik-Variante 2 zeigt die Grafik vergrößert auf 150%. Für die Variante 3 wurde die Größe der Originalgrafik verdoppelt und bei der Variante 4 an den rechten und linken Bildrand angeglichen, was ca. einer Vergrößerung auf 300% entspricht.



Abbildung 6: Darstellungsvarianten der Grafikeinblendungen

## Ergebnisse

Der zweite Test untersucht neben der Zufriedenstellung auch die Verständlichkeit der variierten Produktionsparameter. Dabei zeigt sich, dass die Verständlichkeit situationsabhängig bewertet wird. Generell lässt sich aber feststellen, dass in den Totalen alle an der Situation bzw. dem Spielzug beteiligten Akteure zu sehen sein müssen, damit die Verständlichkeit gegeben ist. Daher sind aus Sicht der Verständlichkeit je nach Spielsituation unterschiedliche Varianten vorteilhaft.

Der eigentlich entscheidende Test jedoch beschäftigt sich mit der Zufriedenstellung. Beim Test bestätigt sich, dass die Zufriedenstellung bei sehr effektiven Varianten

besonders hoch ist. Die Effektivität wird dabei durch die Erkennbarkeit und die Verständlichkeit bestimmt, welche sich teilweise entgegenstehen.

Das ist auch der Grund, weswegen die klassische Variante beim Zufriedenstellungstest durchgehend sehr schlecht abschneidet, was den Änderungsbedarf der Produktion noch einmal bekräftigt. Bevorzugt wird dagegen eine nachgeführte Vergrößerung der Totalen, wie sie Varianten „Pan & Scan“ sowie „Pan & Scan Nah“ darstellen. Der erwünschte Grad der Vergrößerung ist dabei je nach Spielsituation unterschiedlich.

Auch bei den alternativen Grafikdarstellungen schneidet die klassische Variante sehr schlecht ab und führt nicht zur Zufriedenstellung bei den Benutzern. Hier bevorzugen die Testpersonen bei dauerhaft eingeblendeten Grafiken (Spielzeit und Spielstand) eine Vergrößerung, welche die Grafik nur über die Erkennbarkeitsschwelle hebt (Variante 2; Vergrößerung 150%). Bei kurzfristig eingeblendeten wünschen die Probanden eine noch etwas größere Darstellung, was das Lesen deutlich erleichtert (Varianten 3; Vergrößerung 200%).

### **3. Styleguide für die mediengerechte Produktion von Sport-Live-Berichterstattung auf mobilen Endgeräten**

Im Folgenden wird für die einzelnen Produktionsparameter getrennt beschrieben, wie eine mediengerechte Produktion für Mobile TV aussehen muss. Basis hierfür ist das für das klassische Fernsehen produzierte Signal.

#### **Audio**

Alle auditiven Komponenten aus der klassischen TV-Produktion sind auch ohne Änderungen für Mobile TV geeignet und können direkt übernommen werden.

#### **Video**

Der Änderungsbedarf im Bereich Video liegt bei der Darstellung der Totalen. Diese Einstellung wird über 60% der Zeit gezeigt und ist entscheidend für die Spielübersicht. Die Benutzer bevorzugen hier auf dem mobilen Endgerät eine nachgeführte Vergrößerung des Bildes. Das bedeutet, dass die Totale vergrößert wird und immer der Ausschnitt des Bildes gezeigt wird, in welchem das Spiel gerade stattfindet bzw. sich der Ball befindet.



Abbildung 7: Nachgeführte Ausschnittsvergrößerung

Der gewünschte Grad der Vergrößerung ist dabei abhängig von der Spielsituation. Da die Vergrößerung aber konstant ist, muss ein Wert gewählt werden, der bei allen Spielsituationen die Verständlichkeit und Erkennbarkeit gleichermaßen gewährleistet. Dies ist bei einer Vergrößerung der Totalen auf 130% der Originalgröße gegeben. Da Fußball ausschließlich im Seitenverhältnis 16:9 produziert wird, die Displays der mobilen Endgeräte und auch der Übertragungsstandard auf ein Bildseitenverhältnis von 4:3 ausgelegt sind, würde man bei Beibehaltung des Seitenverhältnisses 25% des Displays an schwarze Balken verschenken. Daher wird empfohlen, das Seitenverhältnis von 16:9 auf 4:3 zu ändern. Hierfür wird der obere und der untere Bildrand an die Zielbildgröße angeglichen. Dabei gehen der rechte und linke Bildrand verloren, was aber vertretbar ist, da keine wichtige Information im Bildrand enthalten ist.



Abbildung 8: Bildwandlung von 16:9 auf 4:3

Diese Bildtransformation hat zur Folge, dass der dargestellte Bereich um den Faktor 1,33 vergrößert wird. Durch diese Vergrößerung steigt die Erkennbarkeit aller Einstellungsgrößen. Das resultierende Bildmaterial hat also bei der Totalen  $130 \cdot 1,33 = 172,9$  Prozent der Originalgröße und 133% bei allen anderen Einstellungen.

## **Grafik**

Bei den Grafikeinblendungen wird zwischen der dauerhaft eingeblendeten Grafik Spielstand/ -zeit und den Bauchbinden (bei Karten oder Spielerwechseln) unterschieden.

Die dauerhaft eingeblendete Grafik muss im Vergleich zur klassischen Produktion nur so weit vergrößert werden, bis sie erkennbar wird. Diese Erkennbarkeitsschwelle liegt dem Test zufolge bei ca. 150%. Somit ist die Grafik gut lesbar und es wird nicht all zu viel Bildinhalt verdeckt.

Bei den kurzfristig eingeblendeten Bauchbinden mit Informationen zu Spielerwechseln und Karten ist eine sehr schnelle Lesbarkeit wichtig. Daher wird für diese Grafiken vorgeschlagen, die Größe zu verdoppeln.

## **4. Zusammenfassung**

Als Beispiel für die Sportberichterstattung auf mobilen Endgeräten wurde die Fußball-Live-Berichterstattung gewählt, da Fußball mit zu den populärsten Sportarten zählt. Die Ergebnisse zur Live-Berichterstattung lassen sich auch auf andere Mobile TV Formate, wie „Near-Video-on-Demand“ oder „Video-on-Demand“ übertragen [1].

In zwei Nutzertests wurde der Änderungsbedarf der aktuellen Produktion von Fußball-Live-Berichterstattung ermittelt. Aus den Testergebnissen wurden Richtlinien für die Produktion abgeleitet und ein Styleguide entwickelt.

Diese Testreihe wirft auch Fragestellungen für die weiterführende Forschung und Entwicklung auf. So wäre zum Beispiel eine weitere Testreihe sinnvoll, die den Styleguide prototypisch umsetzt und die Akzeptanz unter möglichst realen Bedingungen über die komplette Länge eines Spiels testet.

Außerdem zeigt das Ergebnis der Arbeit, dass noch ein Mangel an ausgereiften Systemen besteht, welche klassische TV-Inhalte automatisiert für Mobile TV aufbereiten. Die Ausschnittsvergrößerung ist dabei nur ein mögliches Instrument zur Verbesserung.

### **References:**

- [1] Klein, Andreas; Lessig, Michael: Mobile TV 2010. Berlin: Goldmedia GmbH Media Consulting & Research, 2006
- [2] Gerhard, Heinz: Die Fußball-WM als Fernsehesevent. In: Media Perspektiven (2006) Heft 9, S.465-474
- [3] EN-ISO 9241-11, Ausgabe 1998-03. Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten – Teil 11: Anforderungen an die Gebrauchstauglichkeit – Leitsätze
- [4] EN-ISO 9241-12, Ausgabe 1998-12. Ergonomische Anforderungen an Bürotätigkeiten mit Bildschirmgeräten – Teil 12: Informationsdarstellung
- [5] Bortz, Jürgen; Döring, Nicola: Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. 4.Auflage, Heidelberg: Springer Medizin Verlag, 2006

**Authors:**

Alto Kraus

Rike Brecht

Prof. Dr. Heidi Krömker

Technische Universität Ilmenau

Institut für Medientechnik, FG Medienproduktion

Postfach 10 05 65, 98694 Ilmenau

Tel.: +49 3677 691538, Fax: +49 3677 692888

E-Mail: [alto.kraus@googlemail.com](mailto:alto.kraus@googlemail.com); [rike.brecht@tu-ilmenau.de](mailto:rike.brecht@tu-ilmenau.de); [heidi.kroemker@tu-ilmenau.de](mailto:heidi.kroemker@tu-ilmenau.de)

Ngo Ai Tam

## **RCS-M: A Rate Control Scheme to Transport Multimedia Traffic over Satellite Links**

### **INTRODUCTION**

Satellite communication systems [7,8] are being developed very quickly in many countries around the world, promising new capability for wideband Internet and new broadcasting services. Satellite communication systems become a new media in addition to traditional communication systems based on copper cable, optical cable, marine cable and microwaves. Due to the operation at rather high altitudes, satellite communication can cover a rather wide territorial region and surmount the distance limitation of the current fixed and wireless links.

However, due to the signal transmission through the space, satellite communication has to face technical challenges such as high bit error rates, limited bandwidth, interference, etc. The signal transmission over the space can be degraded and disturbed by environment and weather condition, etc. On the other hand, the demand on multimedia services grows very fast. Multimedia services have stringent requirement on quality service. In order to ensure quality service over satellite links, suitable control mechanisms are needed. Recently, new technologies and techniques have been developed including the modification of traditional techniques used in fixed networks, new designs with respect to particular properties of satellite links. In fact, investigation and development of control mechanisms for Quality of Services (QoS) over satellite links remain a technical challenge. Suitable control mechanisms need to consider properties of satellite environment and should ensure QoS as well as be able to congestion control. Current research works [2,3,5] showed that neither traditional TCP and UDP protocols are suitable to transport multimedia traffic over satellite links. The delay and error bit of satellite links have considerable influence on the operation of TCP. Modifications are needed such as: window size increase, slow-start, using congestion control, more effective acknowledgement mechanism, reduction of backward traffic, etc. However, TCP itself is affected by delay and bit error on satellite links and is inappropriate to transport multimedia traffic because of its retransmission mechanism due to packet loss. UDP is also inappropriate because there is no guarantee about the order of packets. Thus, there is a need to develop new transmission protocols which can overcome these difficulties and support multimedia transmission over satellite channel. In recent years, novel protocols, such as RAP (Rate Adaptation Protocol) [10] and RCS (Rate Control Scheme) [2,5] have been proposed to improve data transmission in environment of high bit error rates and high bandwidth-delay product. However, both RAP and RCS still

persist with blemishes. In this paper, we propose a Rate Control Scheme to transport Multimedia traffic over satellite links named RCS-M, which combines the advantages of both RCS and RAP (where M stands for Multimedia).

The paper is organized as follow. In section 2, we introduce RCS-M and present some simulation results in section 3 to show that RCS-M outperforms its predecessors. The experiments indicate that our protocol can be deployed for multimedia transport over satellite links with QoS guarantee in a TCP-friendly fashion. Finally, the section 4 concludes the paper.

## RCS-M: A PROTOCOL FOR MULTIMEDIA TRANSPORT OVER SATELLITE LINKS

### RCS-M Architecture

RCS-M is an end-to-end rate control scheme, which uses additive-increase, multiplicative decrease (AIMD) [13] to transmit multimedia and run on top of UDP. RCS-M uses dummy packet [5] to distinguish packet losses. The cause of packet loss may be network congestion, link error or temporary signal loss. The disconnection might also be due to handoff, signal fading regarding environment obstacles such as mountains, bridges, tunnels, or weather conditions such as storms, rains etc. Thus, the data could not be successfully transmitted to the receiver and hence performance severely degrades.

At the destination, the RCS-M layer sends back an acknowledgement (ACK) for any received packet. These ACKs are used only for flow control. The RCS-M receiver passes the received data packet to the decoder and discards dummy packets.

### Components of RCS-M Architecture

The RCS-M architecture consists of following components:

- + Init: used to set up the connection between RCS-M source and RCS-M destination.
- + Timeout Timer: used to establish the system timeout.
- + RTT-Estimator: used to estimate the suitable Round Trip Time (RTT).
- + ACK Counter: used to count the number of received ACKs.
- + Initial Rate Calculator: calculates the suitable initial sending rate for RCS-M.
- + Rate Regulator: used for transmission rate control.
- + Dummy Packet Generator: generates and transmits dummy packets for checking network state (network resources).
- + Packet Sender / Receiver: for sending and receiving data packets.
- + Packet Loss Detector: used to distinguish the packet losses due to congestion, link error or temporary signal loss.
- + Backoff: this component used to adjust the rate in case of packet loss due to temporary signal loss.

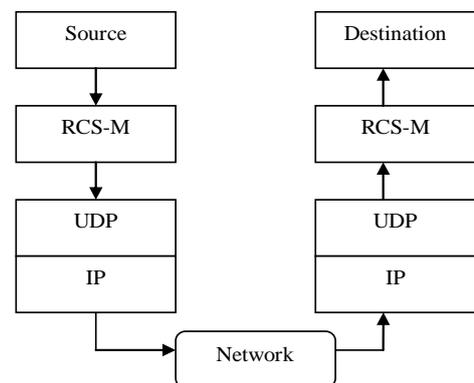


Fig 1. RCS-M Architecture

As indicated above, RCS-M is based on the use of dummy packet. Dummy packets are low priority packets used by the source to probe the availability of network resources. When network is congested, dummy packets will be discarded first. Hence, the transmission of dummy packets does not decrease throughput of data packets. If network is not congested, dummy packets arrive at destination and ACK packets are sent back. These ACKs for dummy packets have low priority. As the source receives ACKs of dummy packets, it indicates that there are unused resources in the network and the source can increase its transmission rate accordingly. RCS-M transmits dummy packets in two cases: at the connection setup and when a data packet loss is detected.

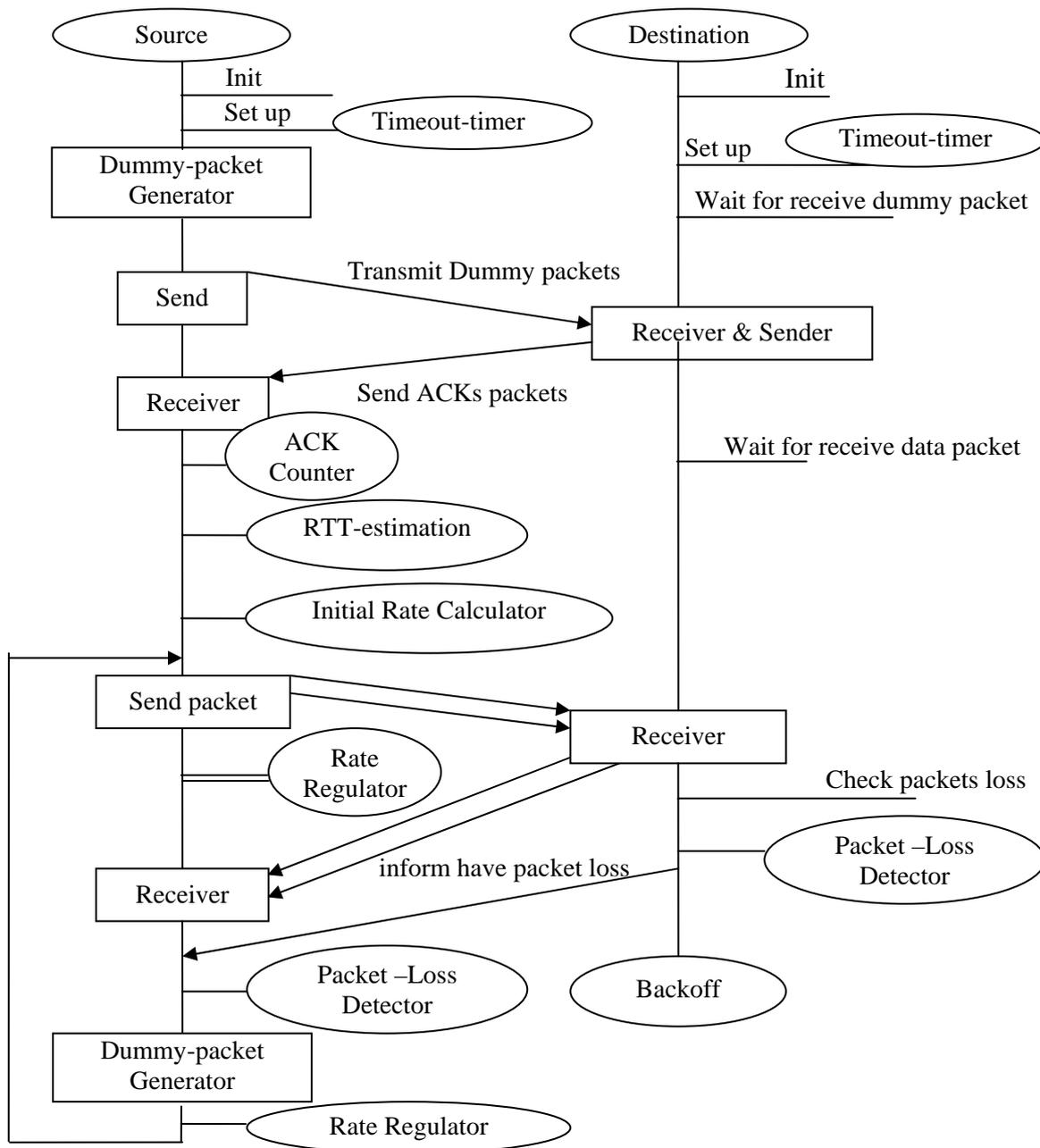


Fig 2. Principle of operation of RCS-M

### Operation of RCS-M Protocol (Figure 2)

First, source and destination will be connected and set up the necessary timeout. The

source sends dummy packets to the destination, and the destination sends back ACKs packets for the source counts. After that, it estimates RTT value and calculates the sending rate. Meanwhile the destination waits for receiving next packets.

When calculating RTT and transmission rate, the source transmits data packets. If detected packet loss, the source sends dummy packets and regulates suitable rate. RCS-M source waits dummy packets for determining the cause of packet loss. Then, RCS-M switches to suitable state.

### State of RCS-M

RCS-M is implemented by a finite state machine model with four states: Initial, Steady, Detected and Backoff (Figure 3). In Initial state, the source sends dummy packets, which are used to probe the network resources, to decide the initial transmission rate that does not cause congestion. In the Steady state, RCS-M source assumes that the network is not congested. Thus, according to the additive-increase scheme [13], it increases its transmission rate in a step-like fashion periodically. When a data packet loss is detected, RCS-M source leaves the Steady state for the Detected state. At start of Detected state, RCS-M doesn't know the reason of packet loss, thus, it assumes that all packet losses are due to network congestion and halves its data transmission rate, simultaneously it also starts

transmitting dummy packets in order to probe availability of network resources and distinguish the cause of lose packets. At the end of the Detected phase, RCS-M source goes back to the Steady or Backoff state. In the Detected phase, data packets are sent with rate  $S$  and two dummy packets are

transmitted for each data packet. , RCS-M source waits for an ACK during the Detected state to assess the actual reason for the packet loss event. If the packet loss is due to the random link error, then RCS-M source receives an ACK during the SRTT period of Detected state. If the packet loss is due to congestion, then all TCP friendly protocol sources along the bottleneck would perform rate throttle for congestion resolution in at most one SRTT. Hence, if no ACK is received until the end of one SRTT, it can be inferred as an indication of temporary signal loss instead of the congestion. If no ACK is received during the Detected phase, then the RCS-M source does not go back to Steady state, instead it goes to Backoff state. In Backoff state RCS-M source sends one dummy packet for each data packet transmitted. During Backoff phase, RCS-M source stops halving its transmission rate and waits for an ACK informing that signal is back. Once RCS-M source receives any ACK for either data or dummy packet, it goes into Steady state. If no ACK is received for a certain timeout period, the source timeouts and moves back to Initial State.

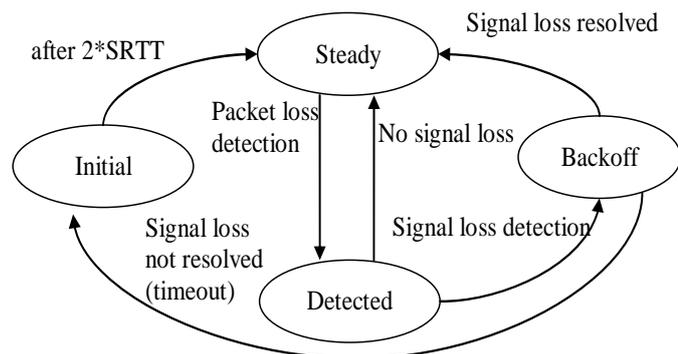
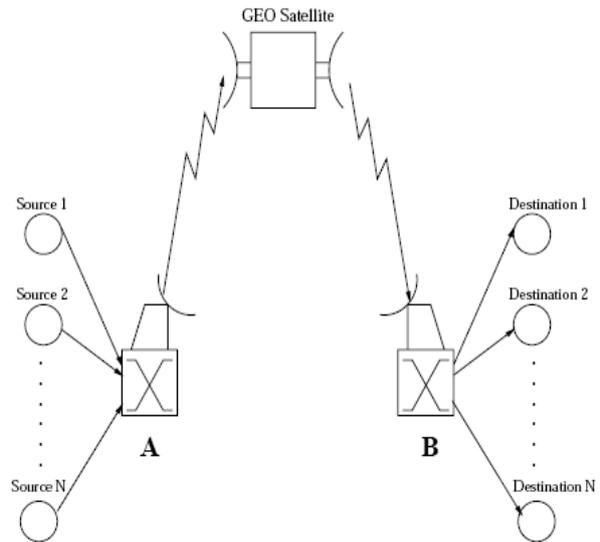


Fig 3. State of RCS-M

## SIMULATION RESULTS

### Simulation scenario

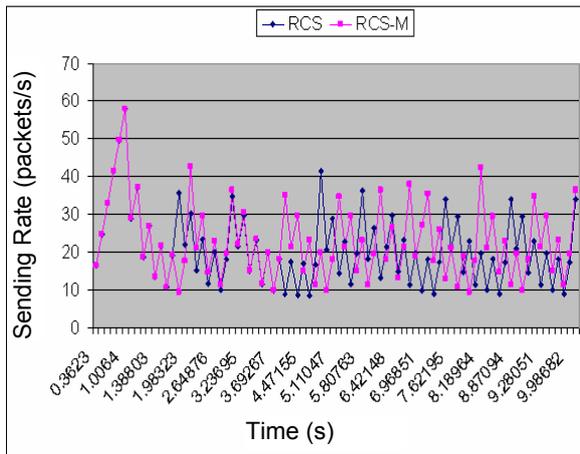
Satellite network is a typical example of network with high bandwidth-delay products and high bit error rates. The simulation system is shown in Figure 4 where N sources transmit data to N destinations. N sources and N destinations are connected with Earth Station A and Earth Station B that receives and transmits satellite signals. We assume that  $N=4$ , packets of length is 1000 bytes.



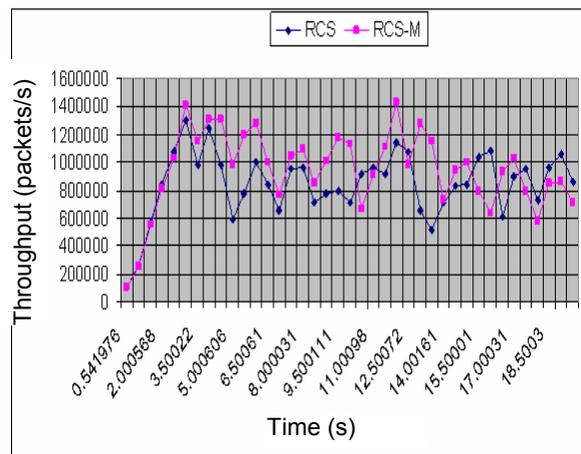
**Fig 4.** Simulation configuration

### Simulation results

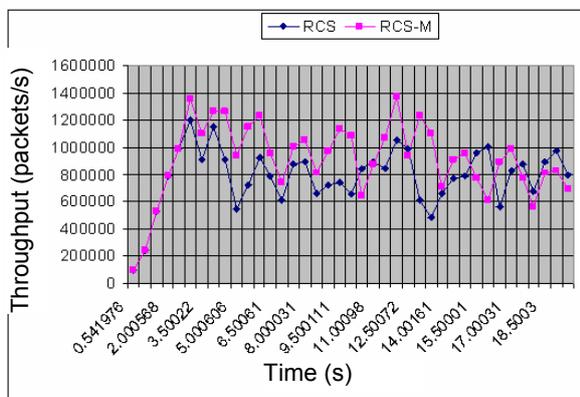
Figures 5a - 5f show the experiment results for sending rate (5a), throughput (5b, 5c, 5d), fairness (5e, 5f) of RCS-M in comparison to the original RCS proposed in [5].



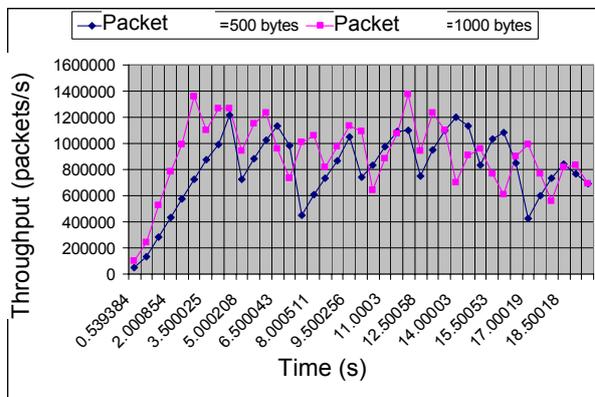
**Fig 5a.** Sending rate of RCS-M in comparison to RCS



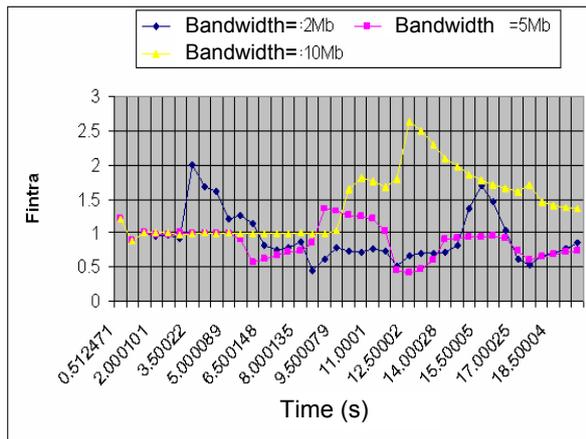
**Fig 5b.** Throughput of RCS and RCS-M with Initial rate = 22 packets/s



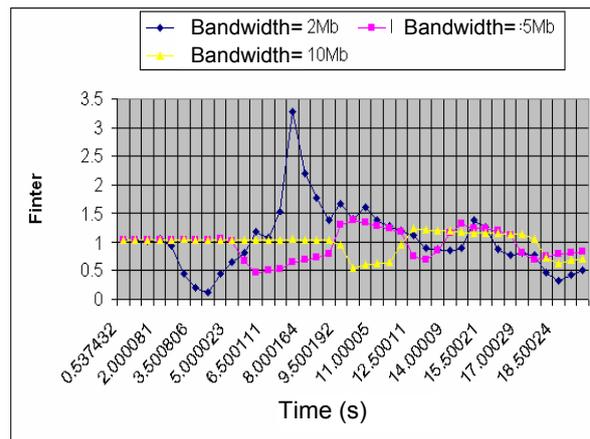
**Fig 5c.** Throughput of RCS and RCS-M with packet length = 1000 bytes



**Fig 5d.** Throughput of RCS-M with different packet sizes (500 bytes, 1000 bytes)



**Fig 5e.** Fairness of RCS-M flows (Intra fairness)



**Fig 5f.** Fairness of RCS and RCS-M (Inter fairness)

The experiment results showed the performance improvements of our RCS-M protocol with respect to the old RCS protocol proposed in [5] in the same simulation conditions.

## CONCLUSION

Multimedia transmission over satellite links faces several technical challenges due to the properties of environments and QoS requirements of multimedia flows. In the paper, we have presented a new protocol based on an end-to-end control scheme called RCS-M using dummy packets and packet loss detection. RCS-M run directly on top of UDP and is TCP-friendly. The simulation results showed that our RCS-M is suitable for multimedia transmission over satellite links and can help to improve transmission throughput in comparison to the old RCS protocol.

## References

- [1] Bradford W. Parkinson. "GPS Eyewitness: The Early Years" GPS World 5, no. 9, 9/1994, 32-45
- [2] Ian F. Akyildiz, Özgür B. Akan, and Giacomo Morabito, "A Rate Control Scheme for Adaptive Real-Time Applications in IP Networks With Lossy Links and Long Round Trip Times", IEEE/ACM TRANSACTIONS ON NETWORKING, VOL. 13, NO. 3, JUNE 2005.
- [3] F. Akyildiz, M. Johnson, G. Morabito, S. Palazzo and J. Tang, "Research Issues for Transport Protocols in Satellite IP Networks, Submitted for publication", November 2000.
- [4] Ivan A. Getting. "The Global Positioning System", IEEE Spectrum, 12/1993, 36-47.
- [5] J. Tang, G. Morabito, I. F. Akyildiz and M. Johnson, "RCS: A Rate Control Scheme for Real-Time Traffic in Networks with High Bandwidth-Delay Products and High Bit Error Rates", IEEE Infocom, Alaska, April 2001.
- [6] Johnson, Nicolas L. "GLONASS Spacecraft", GPS World 5, no. 11, 11/1994, 51-58.
- [7] Joe Montana, "Introduction to Satellite Communications", George Mason University.
- [8] Dennis Roddy, "Satellite Communications", 3rd Edition, McGraw Hill, 2002.
- [9] Higbie, Paul R. and Norman K. Blocker. "Detecting Nuclear Detonations with GPS", GPS World 5, no. 2, 2/1994, 48-50.
- [10] R. Rejaie, M. Handley, and D. Estrin, "RAP: An end-to-end rate-based congestion control mechanism for realtime streams in the internet", Proc. IEEE INFOCOM, vol. 3, Mar. 1999, pp. 1337-1345.
- [11] Ron Frederick, "Experiences with real\_time software video compression", Xerox PARC, July 22
- [12] Wysocki, Joseph, Lt. Col. "GPS and Selective Availability - The Military Perspective" GPS World 2, no. 7, 7&8/1991, 38-44
- [13] D. Chiu and R. Jain, "Analysis of the increase and decrease algorithm or congestion avoidance in computer networks," Computer Network and ISDN system J., vol. 17, no. 1, pp. 1-14, Jun. 1989.

## Authors:

BSc. Ngo Ai Tam

Faculty of Information Technology, University of Civil Engineering

55 Giai Phong Road, Hanoi, Vietnam

Phone: +84 4 8212677

E-mail: [delangthang@gmail.com](mailto:delangthang@gmail.com)

C. Kellner / A. Mitschele-Thiel / A. Diab

## **Performance Evaluation of MIFA, MIP and HAWAII**

### **ABSTRACT**

All-IP networks become increasingly visible. The various communication networks are aimed to be connected with each other through a common IP core, so that the user will stay always online, anytime and anywhere. We believe that these networks will be the popular network in the future. However a lot of challenges remain unsolved until today. One of the major challenges is how to achieve a seamless and fast handoff while moving from one point of attachment to another.

Mobile IP Fast Authentication Protocol (MIFA) is proposed to avoid the problems of MIP and to match the real-time requirements without introducing intermediate nodes and without making any restriction on the network topology.

In this paper we evaluate the performance of MIFA compared to Mobile IP (MIP) and Handoff-Aware Wireless Access Internet Infrastructure (HAWAII). The evaluation is performed by means of network simulator 2 (ns-2). The three protocols are evaluated deploying the same topology.

Our simulation results have shown that MIFA outperforms the other two protocols with respect to handoff latency and the number of dropped packets for uplink and downlink traffic. This is because MIFA needs only to register with the new Foreign Agent (FA) to be able to resume sending of packets, while MIP needs to register with the HA and HAWAII requires updating the new location at the old FA. For downlink traffic HAWAII and MIFA perform comparable to each other.

### **I. INTRODUCTION**

All-IP networks become increasingly visible. The various communication networks are aimed to be connected with each other through a common IP core, so that the user will

stay always online, anytime and anywhere. We believe that these networks will be the popular network in the future. However a lot of challenges remain unsolved until today. One of the major challenges is how to achieve a seamless and fast handoff while moving from one point of attachment to another.

MIP [1], [2] presents the standard protocols used to support mobility in IP based networks. With MIP, the Mobile Node (MN) has to be registered and authenticated by the Home Agent (HA) every time it moves from one subnet to another. This produces communication latency during the movement, especially, if the HA is far a way. This latency makes MIP only suitable for the management of global mobility. Therefore, it is important to develop a new mobility solution able to satisfy the real-time requirements.

The rest of this paper is organized as follows: In section (II) we present the related work. After that, we describe the simulation scenarios and discuss the results in section (III). Lastly, we conclude with the main results and the future work in section (IV).

## II. RELATED WORK

In order to avoid MIP drawbacks, several approaches have been proposed to support local mobility. These approaches can be classified into two main groups [3], Proxy Agent Architectures (PAA) and Localized Enhanced Routing Schemes (LERS). PAA solutions try to extend MIP principle by using intermediate agents to process the movements of MNs inside a certain domain locally. Regional Registration for MIPv4 (RMIP) [4], Hierarchical Mobile IPv6 (HMIPv6) [5] and MIFA [6] are examples of this group. LERS solutions introduce a new dynamic layer3 routing protocol inside a certain localized area. A special path set-up protocol is used to implement per host soft-state forwarding entries. HAWAII [7] is an example of this group.

With RMIP and HMIPv6 the HA is not aware of every change in the point of attachment. This is due to the fact that the handoff procedures are processed locally by a special node, e.g. a Gateway Foreign Agent (GFA) or Mobility Anchor Point (MAP), when the MN moves inside a certain domain. The MN communicates with the HA only if it changes the domain. HAWAII does not try to replace IP. Its domain is controlled by a certain Gateway called Domain Root Router (DRR). There is a set of stations under this DRR. Each station maintains a routing cache, which is updated when the MN moves inside the domain. The packets are forwarded then hop by hop towards the MN. The two main schemes are forwarding (UNF) and non-forwarding path scheme (MSF). Using forwarding path scheme, the MN is assumed to be able to detect the new BS while still connected to the old one. In

this case a path update message is sent to the old FA, which sends it towards the new one. The packets are forwarded after that to the new BS. By non-forwarding path scheme, the MN moves to the new BS and after that the MN sends a path update message to the old one. No packets are forwarded here from the old BS to the new one. All nodes of the domain should be mobility aware. Although HAWAll processes the mobility locally inside a certain domain, it still needs MIP when moving between different domains.

MIFA is developed to eliminate the latency sources of MIP and to match the real-time requirements without requiring intermediate nodes between the Foreign Agent (FA) and the HA. The basic idea is that MIFA deploys a set of neighbouring FAs called Layer3 Frequent Handoff Region (L3-FHR). This is motivated from the fact that the MN moves always to a neighbour subnet. Therefore, after moving to the new FA, the MN performs re-authentication only with this FA, it's not necessary to communicate with the HA. So, MIFA reduces the registration latency and let the MN able to quickly resume transmission on up- and downlink.

### III. SIMULATION RESULTS

#### a) Simulation scenario

In order to evaluate the three protocols we have implemented MIFA in the network simulator ns-2 [8] (version 2.29). For comparison with MIP we use the existing implementation in ns-2. The two protocols are compared to the HAWAll UNF scheme using the "Columbia IP Micro-Mobility Suite (CIMS)" updated to ns2.29 [9].

The network topology is presented in figure 1. We use a set of sixteen FAs. There are ten mobiles in the range of each FA. 60 of these mobiles generate traffic; other mobiles are in idle mode. The mobiles communicate with 6 different correspondent hosts. The distance between each two FAs is 140 meters. Neighbor FAs are overlapped. The MN loses the communication, when it receives an advertisement message from the new FA. The delay on the wired links in our topology is 5 ms with a capacity of 100 Mbit/s. We observe the traffic of one MN, which moves with a speed of 40 km/h from the first to the last FA. This means, that the MN makes 15 handoffs. A UDP-traffic in up- and downlink is generated. We use CBR with an interval of 20 ms and a packet size of 500 bytes. The scenario is re-simulated 10 times to get stable results.

The same network topology with the same parameters is used for the evaluation of the three protocols.

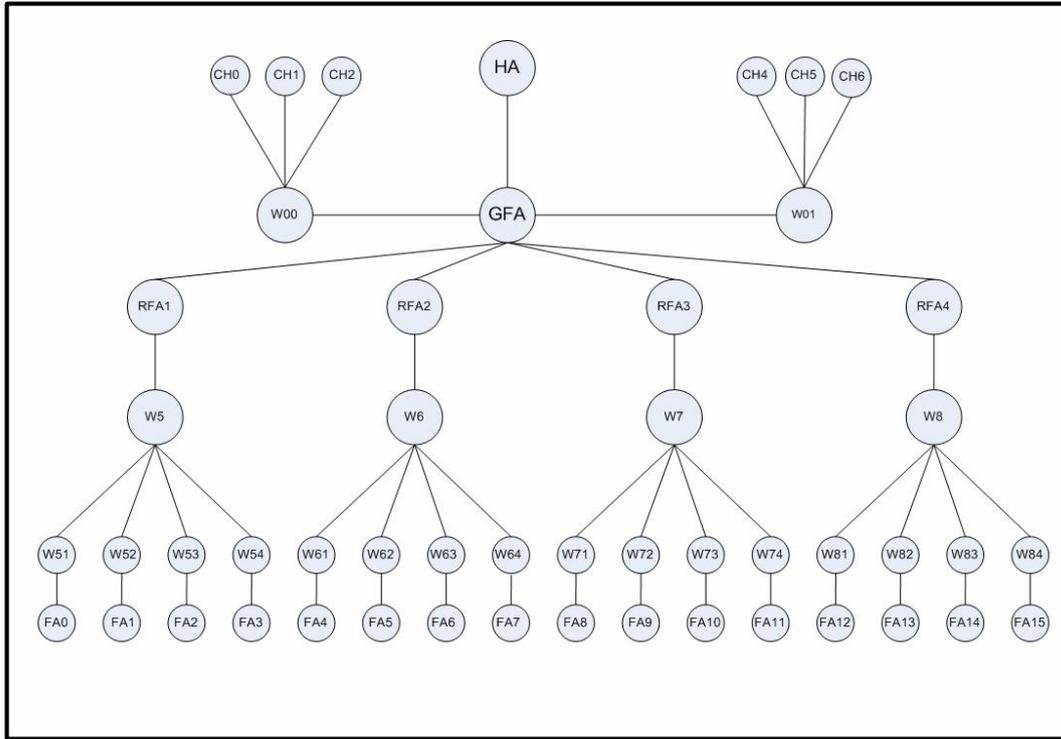


Figure 1. Simulation scenario

## b) Handoff latency

The handoff latency is measured as the duration between sending of registration request message and receiving of the registration reply message. The network load is changed randomly through changing the sending interval of other mobiles from 40 to 90 ms. Figure 2 shows the distribution function of the handoff latency experienced by each protocol. From figure 2 we can notice that MIFA density function is more stable than the other protocols, 90% of the handoffs take a time less than 15 ms on uplink and less than 82 ms on downlink. HAWAII and MIP suffer from higher handoff latency, 90% from the handoffs by MIP takes a time less than 160 ms and by HAWAII less than 205 ms. MIFA is always better than HAWAII and MIP for low and high loads. HAWAII is better than MIP for low loads, while MIP starts to be better while increasing the load after a certain threshold. Taking the average value of the handoff latency over the all measurements, we see that HAWAII is about 0,8% better than MIP, while MIFA is about 64% better than MIP and HAWAII on downlink and about 97% on uplink.

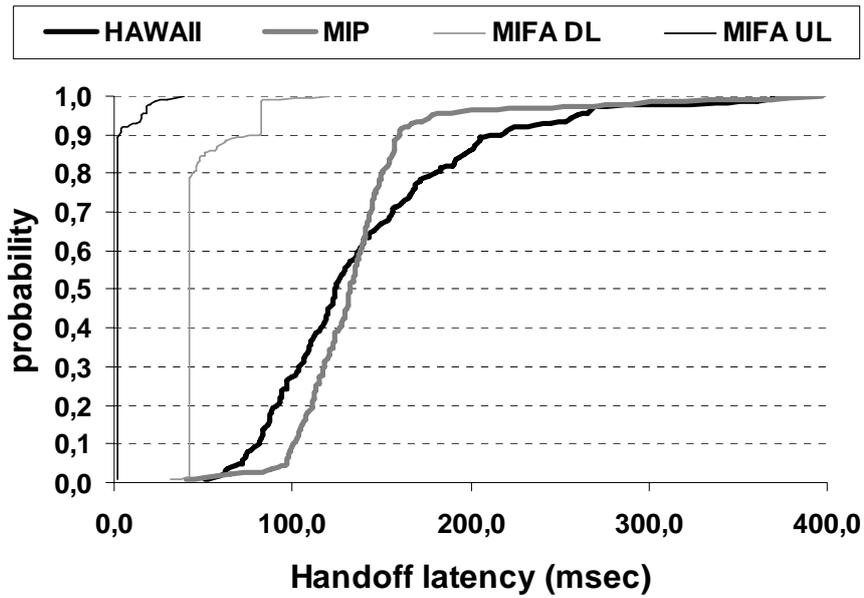


Figure 2. Distribution function of the handoff latency

### c) Packet dropping

Figure 3 shows the distribution function of the expected number of dropped packets. Similar results can be obtained here. MIFA clearly outperforms HAWAII and MIP. The average number of dropped packets by MIFA is about 0,87 and 1,53 on up- and downlink respectively. HAWAII on the other side produces about 2,89 dropped packets, while MIP is the worst and generates about 7, 54 dropped packets on average.

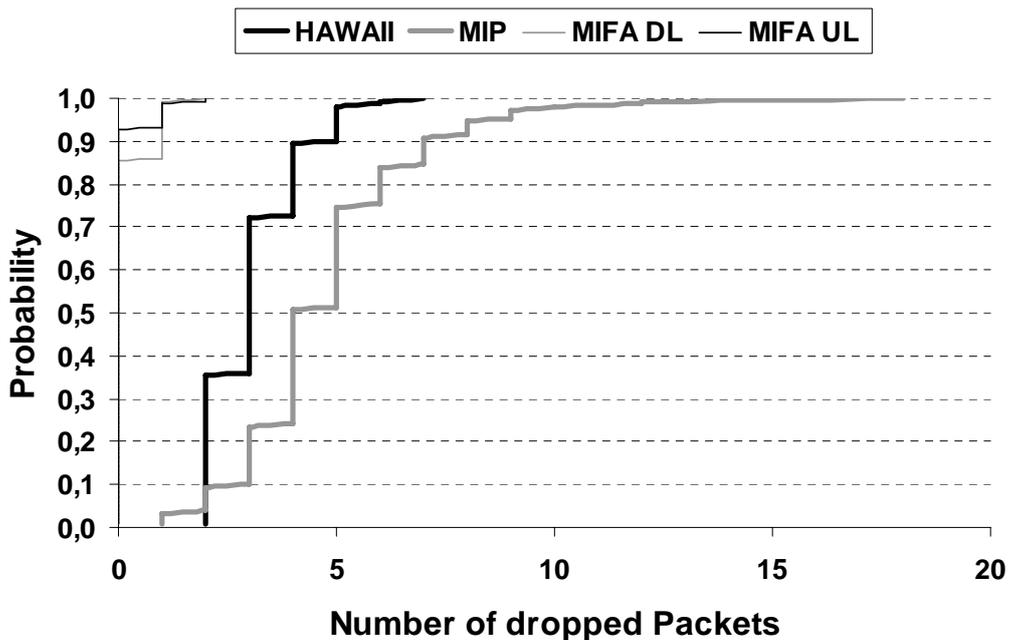


Figure 3. Distribution function of expected dropped packets

## IV. Conclusion for future work

In this paper we have evaluated the performance of MIFA compared to MIP and HAWAII deploying the same topology. Our simulation results have shown that MIFA outperforms the other two protocols with respect to handoff latency and the expected number of dropped packets for uplink and downlink traffic. MIFA achieves a seamless handoff and satisfies the requirements of real-time applications.

Currently we are measuring tcp throughput to evaluate the impact of MIFA on TCP performance. Further we are studying the impact of the MN's speed on the performance.

### References:

- [1] C. Perkins, Ed, " IP Mobility Support for IPv4 ", RFC: 3344, 2002.
- [2] D. Johnson, C. Perkins, J. Arkko, " Mobility Support in IPv6 ", < draft-ietf-mobileip-ipv6-23.txt >, 2003.
- [3] P. Eardley, A. Mihailovic, T. Suihko, " A framework for the evaluation of IP mobility protocols ", 11th PIMRC, 2000.
- [4] D E. Gustafsson, A. Jonsson, C. E. Perkins, " Mobile IPv4 Regional Registration ", < draft-ietf-mobileip-reg-tunnel-08.txt >, 2003.
- [5] H. Soliman, et.al, " Hierarchical Mobile IPv6 mobility management (HMIPv6)", < draft-ietf-mobileip-hmipv6-08.txt >, 2003.
- [6] A. Diab, A. Mitschele-Thiel, J. Xu, " Perfomance Analysis of the Mobile IP Fast Authentication Protocol ", MSWIM04, 2004.
- [7] R. Ramjee, T. La Porta, S. Thuel, K. Varadhan, " IP micro-mobility support using HAWAII ", < draft-ietf-mobileip-hawaii-00 >, 1999.
- [8] Network simulator ns-2, URL: <http://www.isi.edu/nsnam/ns/>
- [9] Columbia IP Micro-Mobility Suite (CIMS) updated to ns-2.29, URL: <http://wcms1.rz.tu-ilmenau.de/fakia/Micro-Mobility.5233.0.html>

### Authors:

Christian Kellner

Dipl.-Ing Ali Diab

Prof. Dr.-Ing. habil. Andreas Mitschele-Thiel

TU - Ilmenau, Faculty for Informatics und 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.kellner@stud.tu-ilmenau.de](mailto:christian.kellner@stud.tu-ilmenau.de), [ali.diab@mitsch}@tu-ilmenau.de](mailto:ali.diab@mitsch}@tu-ilmenau.de)

A. Diab / A. Mitschele-Thiel

## **MIFAv6: A Fast and Smooth Mobility Protocol for IPv6**

### **ABSTRACT**

In this paper we present Mobile IP Fast Authentication protocol for IPv6 (MIFAv6). MIFAv6 is developed to support a fast and smooth mobility in IPv6 networks. The Mobile Node (MN) needs to update its binding only at the new Access Router (AR) to be able to resume sending and receiving of data packets. The latency between the current AR and the Home Agent (HA) / Corresponding Node (CN) is hidden from the application and does not affect the performance. A main advantage of MIFAv6 is that it does not make any restrictions on the network topology.

### **I- INTRODUCTION**

The growing demand for high-speed wireless access to the Internet anywhere and anytime are the driving forces resulting in the current trend to develop All-IP wireless networks. A common IP core is used to connect all wireless radio access networks and technologies. The radio gateways and MNs use IP protocol for signaling and data transport. However, an All-IP wireless network requires efficient and flexible mobility solutions, which can satisfy the requirements of delay sensitive applications.

Mobile IP version 4 (MIPv4) [1] and version 6 (MIPv6) [2] present the well known IETF standards used to support mobility in IP based networks. With MIP, the MN has to be registered and authenticated by the HA every time it moves from one subnet to another. This introduces communication latency when the MN is moving. This latency makes MIP only suitable for the management of global mobility, referred to as macro mobility too. Therefore, it is important to develop a new mobility solution able to overcome the shortcomings of MIP and to satisfy the real-time requirements.

The rest of this paper is organized as follows: The related work is presented in section (II). After that, we describe MIFAv6 in section (III). In section (IV) we conclude with the main results.

## **II- RELATED WORK**

In order to avoid MIP drawback, several approaches have been proposed to support local, termed as micro, mobility. These approaches can be classified into two main groups [3], Proxy Agent Architectures (PAA) and Localized Enhanced Routing Schemes (LERS). PAA solutions try to extend MIP principle by using a hierarchical network architecture with intermediate agents to process the movements of MNs inside a certain domain locally. Regional Registration for MIPv4 [4] and Hierarchical Mobile IPv6 [5] are examples of this group. LERS solutions introduce a new dynamic layer3 routing protocol inside a certain localized area. A special path set-up protocol is used to implement per host soft-state forwarding entries used to forward the packets towards the MNs. Cellular IP [6] and Handoff-Aware Wireless Access Internet Infrastructure (HAWAII) [7] are examples of this group.

In order to accelerate the layer3 handoff in macro or micro mobility solutions, some approaches propose the use of layer2 information to trigger the handoff on layer3. This information is called layer2 triggers. These triggers force the MNs to perform a layer3 handoff in advance and so the layer3 handoff latency can be minimized or even eliminated. The well known protocols that employ this principle are Fast MIPv6 [8], pre- and post-registration methods [9] and seamless MIP [10].

Another trend in developing mobility management solutions is currently obtaining a big interest. The researchers are trying now to realize network-based mobility management frameworks. In these frameworks the MN has a minimal support of mobility or even no support. The network should perform all the tasks related to the mobility on behalf of the MN. The well known example of these approaches is Proxy MIP [11]

## **III- MIFAV6: MOBILE IP FAST AUTHENTICATION PROTOCOL FOR IP VERSION 6**

### **a) Basic idea**

In order to avoid the problems of MIPv6 without introducing any new entities and without making restrictions on the network topology, Mobile IP Fast authentication protocol for IPv6 (MIFAv6) is proposed. The new protocol is a continuous development of MIFA [12], which is developed mainly for IPv4. The network topology of MIFAv6 is the same as by

MIPv6 with the exception that the ARs should support mobility. The topology and the handoff procedure of MIFAv6 are plotted in figure1.

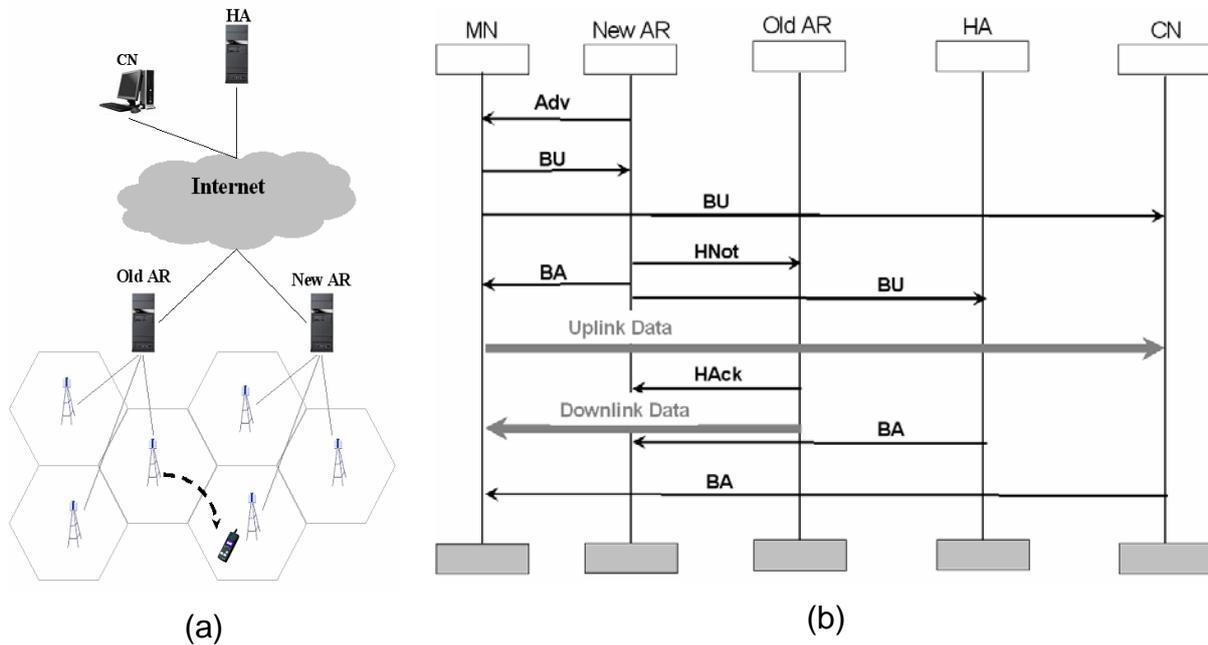


Figure 1: Network topology and handoff procedure of MIFAv6

In reality, the MN moves always to a neighbor AR, which has normally a short distance to the old AR with respect to the number of hops. By utilizing this fact, the HA can delegate the authentication to the ARs. As a result, the MN requires only contacting the new AR to resume the communication on uplink. On downlink, the new AR relies on the old AR to obtain the packets of the MN until the HA or the CN is informed about the new MN's binding.

Each AR builds a Layer 3 Frequent Handoff Region (L3-FHR), which is composed of a set of neighboring ARs to which the MN is likely to move [12]. L3-FHRs can be built statically by means of a certain algorithm such as neighboring graph [13]. L3-FHRs can be built dynamically too using neighbor discovery protocol of IPv6 [14] or by tracking the MNs movements.

We suppose here, that there are Security Associations (SA) between the ARs in one L3-FHR. These SAs can be established using AAA infrastructure [15], IKE [16], or any other key distributing mechanism [17].

## b) Protocol operation

When the MN attaches to the network, it uses the same procedures of MIPv6. Firstly, the MN performs a return routability procedure with the CN and obtains a binding management key (k<sub>bm</sub>) to authenticate the control messages exchanged with the CN.

Afterwards, Binding Updates (BU) are sent to the CN and to the HA. The MN informs the AR and the HA that it wants to use MIPv6 in the next registrations. This is achieved by setting one of the reserved bits in BU message, referred to as *M flag*. Of course, the AR advertises its support of MIPv6 by setting one of the reserved bits in the Router Advertisement message (RA), referred to as *M flag too*. The BU sent to the CN is built according to MIPv6 specifications. The HA responds to the BU by building two SAs. One is between the HA itself and the AR ( $K1_{HA,AR}$ ). The second is between the MN and the AR ( $K1_{MN,AR}$ ). These two SAs are sent with the Binding Acknowledgement (BA) message to the AR, which extracts the SAs, generates two random variables  $R1$ ,  $R2$ , generates another key ( $K2_{MN,AR}$ ) between the MN and the ARs in the current L3-FHR.  $K2_{MN,AR}$  is encrypted with  $K1_{MN,AR}$  and sent to the MN with  $R1$  and  $R2$  in suitable extensions to the BA message. The security between the AR and the MN should be granted through the using of ESP [18] in transport mode. As said above, the SAs can be built through AAA infrastructure or any other key distribution method.

After the MN has been registered and authenticated, Movement Probability Notification (MPN) and Movement Probability Acknowledgement (MPAck) messages are exchanged between the current AR and the HA. MPN contains the random variables  $R1$ ,  $R2$  and a SA,  $K2_{HA,AR}$ , between the HA and the new AR, to which the MN will move in the future.  $K2_{HA,AR}$  is encrypted using  $K1_{HA,AR}$ . MPAck contains all the information required to authenticate the MN during the registration with the new AR. The HA computes two authentication values using  $R1$  and  $R2$ , the first value (*auth1*) present the authentication value the MN should generate, while the other (*auth2*) expresses the value the HA should generate as a response to *auth1*. These two values are sent encrypted to the current AR. For security consideration, it is highly recommended that the messages exchanged between the AR and the HA are authenticated and authorized using ESP in transport mode too.

After the current AR obtains all data required to authenticate the MN, it distributes this data to all ARs locating in the current L3-FHR. This is done by sending another MPN message to each neighbor AR. This data is recorded in soft state and will be used by one AR in the future and deleted from the others. The MPN messages sent to the neighbor ARs should be authenticated and authorized using the existing SA established between the ARs in the L3-FHR.

After the MN moves to another AR, it waits for a RA message. When receiving this

message, it sends a BU message to the new AR and a BU message to the CN. The BU message sent to the CN is sent according to MIPv6 specifications, while the *MI* flag is set in the BU sent to the new AR. This BU contains *auth1* too and authenticated using  $K 2_{MN,AR}$ . After the new AR authenticates the MN, it compares the *auth1* generated by the MN with the value of *auth1* generated by the HA and checks if the HA can satisfy the MN's requirements. This is achieved by checking the data distributed from the old AR with the MPN message. If the check is successful, the new AR sends a Handoff Notification (HNot) message to the old AR and a BA to the MN. The BA contains two new random variables and a new key  $K 3_{MN,AR}$  for the next registration with the next new AR, to which the MN will move in the next registration. Additionally, the value of *auth2*, generated by the HA, is sent with BA message to the MN, which in turns should generate this value using *R1* and *R2* and compare the generated value with the value, which has received from the new AR.

In addition, the new AR sends another BU message to the HA to inform it about the new binding. The new generated random variables and a new key ( $K 3_{HA,AR}$ ), as a SA between the HA and the next new AR, are sent with this BU message. The HA responds by sending a BA message to the new AR and starts forwarding the packets to the new Care of Address (CoA) if the triangular routing is used. This BA message contains the information required to authenticate the MN in the next registration. Upon the new AR receives this BA message it distribute this data to the ARs in the current L3-FHR with MPN messages.

When the old AR receives the HNot message, it responds by sending a Handoff Acknowledgement (HAck) message to the new AR and starts forwarding the MN's packets to the new CoA.

The BU sent to the CN follows the same procedures as by MIPv6. This is motivated from the fact that any update in the CNs should be avoided to make MIFAv6 applicable.

As known from MIPv6, the MNs configure their CoAs in stateful or stateless mode. In case of stateless mode, after the MN configures its CoA, a Duplicated Address Detection procedure (DAD) should be performed to ensure that the address is not used by another MN. Configuration of CoA and DAD procedure slow up the handoff. This is avoided by MIFAv6 because it configures the CoA and executes the DAD procedure in advance. Pre-configuration of the new CoA depends on the used mode. In stateless mode, the MN generates a random value "*host-id*". Host-id can be the MAC address of the MN. This value is sent to the new AR with the BU message. After the new AR obtains the

authentication information required to authenticate the MN during the next registration, it distributes this information with the host-id to the ARs of the current L3-FHR. Each neighbor AR combines the CoA, *net-prefix+host-id*, and executes DAD in advance in each AR in this L3-FHR. In stateful mode, the MN does not generate any host-id. Upon the ARs of the current L3-FHR are informed about the MN, they can configure the CoA in advance.

#### IV- CONCLUSION AND FUTURE WORK

In this paper we have proposed a new mobility protocol for IPv6. From the discussion above we see that MIFAv6 presents a good mobility framework to support fast and smooth handoff in IPv6 backbone. The MN needs to update its binding only at the new AR to be able to resume sending and receiving of packets, It does not have to wait for update the binding at the HA or the CN. The latency between the current AR and the HA / CN is hidden from the application and does not affect the performance. MIFAv6 does not require any intermediate nodes or hierarchical topology. It is a protocol to support macro and micro mobility management. A main advantage of MIFAv6 that the updating is only restricted to the MNs, the ARs and the HA

#### References:

- [1] C. Perkins, Ed, " IP Mobility Support for IPv4 ", RFC: 3344, 2002.
- [2] D. Johnson, C. Perkins, J. Arkko, " Mobility Support in IPv6 ", < draft-ietf-mobileip-ipv6-23.txt >, 2003.
- [3] P. Eardley, A. Mihailovic, T. Suihko, " A framework for the evaluation of IP mobility protocols ", 11th PIMRC, 2000.
- [4] D E. Gustafsson, A. Jonsson, C. E. Perkins, " Mobile IPv4 Regional Registration ", < draft-ietf-mobileip-reg-tunnel-08.txt >, 2003.
- [5] H. Soliman, et.al, " Hierarchical Mobile IPv6 mobility management (HMIPv6)", < draft-ietf-mobileip-hmipv6-08.txt >, 2003.
- [6] A. Campbell, et.al, " Cellular IP ", < draft-ietf-mobileip-cellularip-00.txt > 2000.
- [7] R. Ramjee, T. La Porta, S. Thuel, K. Varadhan, " IP micro-mobility support using HAWAII ", < draft-ietf-mobileip-hawaii-00 >, 1999.
- [8] R. Koodli, " Fast Handovers for Mobile IPv6 ", RFC 4068, 2005.
- [9] K. El Malki, et al. " Low Latency Handoffs in Mobile IPv4 ", <draft-ietf-mobileip-lowlatency-handoffs-v4-04.txt>, 2002
- [10] R. Hsieh, Z.G. Zhou, A. Seneviratne, " S-MIP: A Seamless Handoff Architecture for Mobile IP ", INFOCOM, 2003.
- [11] S. Gundavelli, et.al., "Proxy Mobile IPv6", Internet draft <draft-ietf-netlmm-proxymip6-00.txt>, 2007.
- [12] A. Diab, A. Mitschele-Thiel, J. Xu, " Performance Analysis of the Mobile IP Fast Authentication Protocol ", MSWIM04, 2004.
- [13] S.K. Sen, et al., " A Selective Location Update Strategy for PCS Users ", ACM/Baltzer J. Wireless Networks, 1999.
- [14] T. Narten, E. Nordmark, W. Simpson, " Neighbor Discovery for IP Version 6 (IPv6) ", RFC2461, 1998.
- [15] C. Laot, G. Gross, et al., " Generic AAA Architecture ", RFC 2903, 2000.
- [16] D. Harkins, D. Carrel, " The Internet Key Exchange (IKE) ", RFC2409, 1998
- [17] D. B. Johnson, N. Asokan, " Registration Keys for Route Optimization ", < draft-ietf- mobileip-regkey-03.txt >. 2000.
- [18] S. Kent, R. Atkinson, " IP Encapsulating Security Payload (ESP) ", RFC2406, 1998.

#### Authors:

Dipl.-Ing Ali Diab

Prof. Dr.-Ing. habil. Andreas Mitschele-Thiel

TU - Ilmenau, Faculty for Informatics und Automation, Gustav-Kirchhoff-Str. 1, P.O.B. 10 0565  
98693, Ilmenau

Phone: +49 3677 69 2819

Fax: +49 3677 69 1220

E-mail: [ali.diab@mitsch@tu-ilmenau.de](mailto:ali.diab@mitsch@tu-ilmenau.de)

A. Diab / A. Mitschele-Thiel

## **CAMP: A New Tool to Analyse Mobility Management Protocols**

### **ABSTRACT**

All-IP networks will be as known the future B3G networks. Many Challenges have to be solved in order to make these networks as good as 2 or 3G networks. One of the main challenges is the developing of suitable mobility solutions to support seamless and fast movements in the network. A lot of mobility management protocols have been developed. These protocols should be analyzed and evaluated before the real implementation takes place. The analysis is performed by means of mathematical models, simulations or implementation in test labs. Mathematical analysis can be performed fast and delivers a good approximation of the performance. Simulation and implementation of protocols deliver detailed results. This takes however a long time. Therefore, it is important to build tools to simplify the analysis of mobility management protocols.

In this paper we introduce a new tool for Comparative Analysis of Mobility Management Protocols "CAMP". This tool provides a set of functions that can be used to analyze the performance and the total cost of mobility protocols. The structure of the tools enables simple integrating of new protocols, which results in fast evaluate of the protocols.

### **I- CAMP OVERVIEW**

CAMP, developed by the group of Integrated Hardware and Software systems (IHS) in the technical university of Ilmenau, is a flexible tool for analysis of mobility protocols and the protocols, which support Quality of Service (QoS) simultaneous with mobility. CAMP has of a flexible Graphical User Interface (GUI), shown in figure 1. This GUI is created dynamically based on configuration data stored in a database. The Tool analyzes the performance and the costs resulting from the protocols mentioned above. CAMP is

designed in such a way, that new protocols and algorithms can be integrated simply. This ensures the extendibility of CAMP.

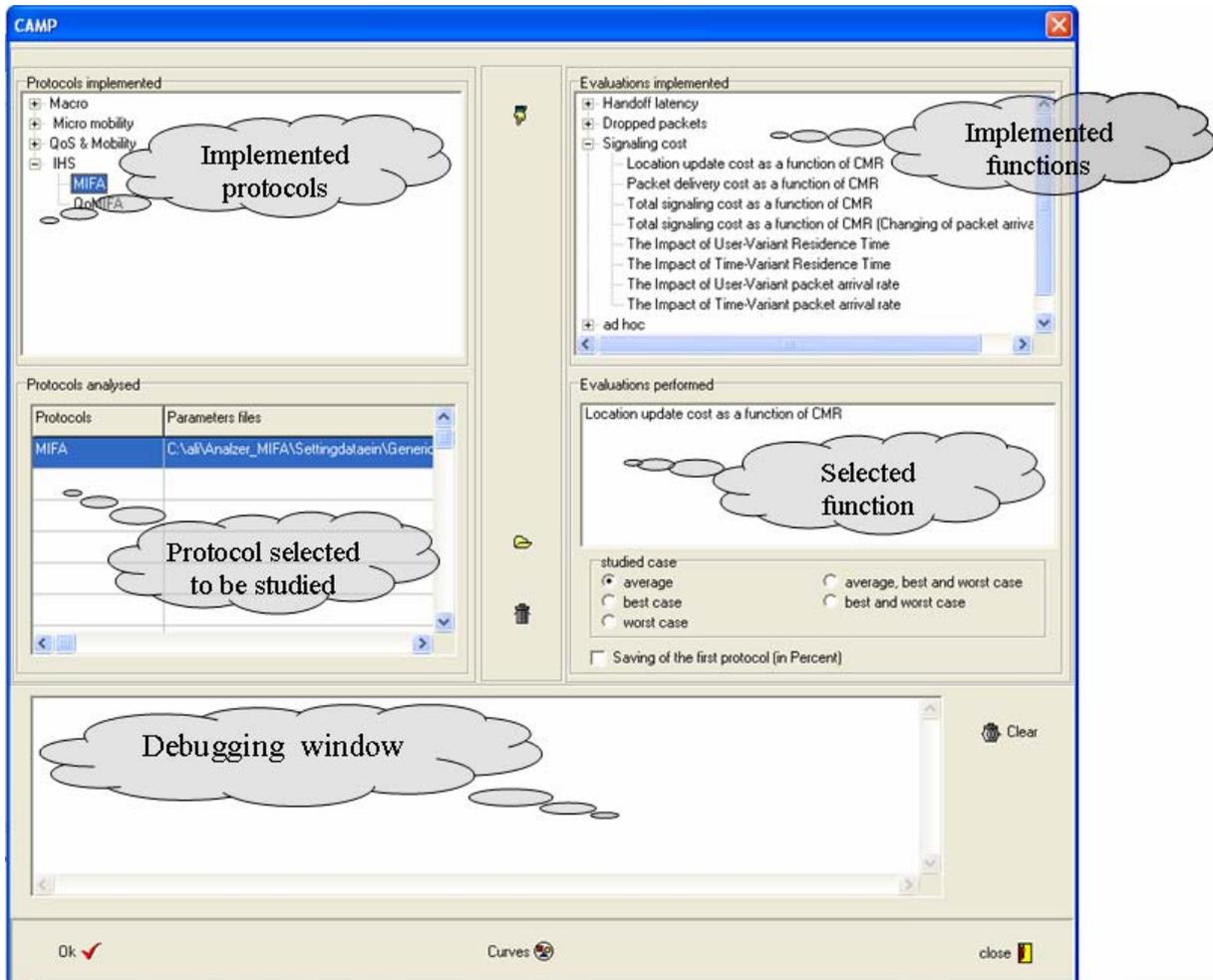


Figure 1: GUI of CAMP

### I- TOOL STRUCTURE

Additional to the GUI, CAMP consists mainly of libraries package and a configuration database, the structure of CAMP is plotted in figure 2. All configuration parameters of CAMP are recorded in the configuration database, e.g. *GUI parameters, the support protocols, the implemented functions...*etc. This makes CAMP flexible and re-configurable. The libraries package contains the libraries required for the analysis. “G-libraries” presents the libraries specified for all implemented protocols, e.g. mobility models, queuing...etc. The “P-libraries” are protocol-specific libraries.

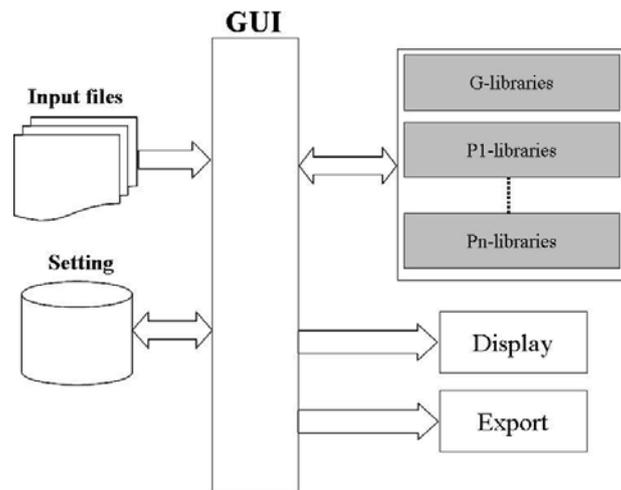


Figure 2: Structure of CAMP

“Input files” contains the parameters of the studied protocols. These parameters are written simply in scripts, an example of a parameter file is presented in figure 3. Our tool defines a simple semantic of the “Input files”. The parameter name should be preceded by “#”, where the value of the parameter value should be written in a new line direct under the parameter name. The comments can be expressed in many ways, see figure 3.

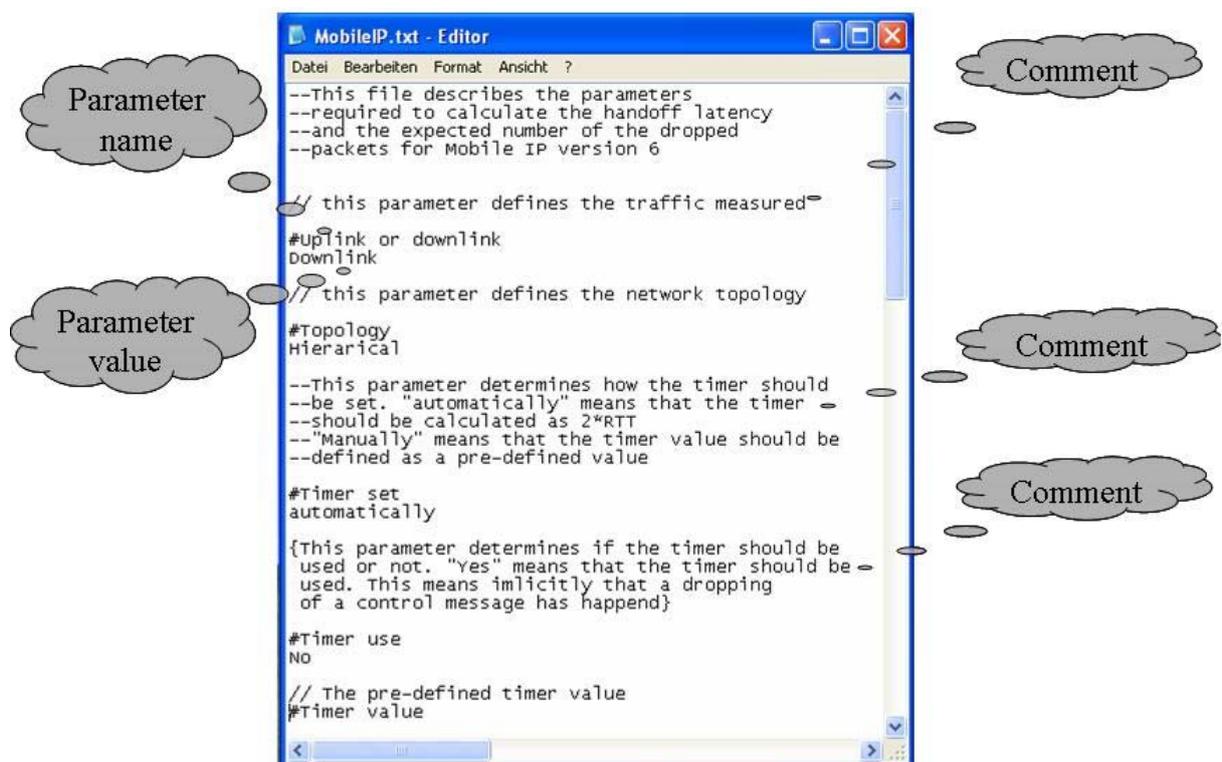


Figure 3: Example of a parameter file

The user should firstly define the protocols, which he wants to analyze. After that, each protocol should be linked to a parameters file. The user should select then the function

he wants to perform. CAMP reads the parameters, selects the adequate libraries and performs the desired analysis. The results are saved in a stream file. These results can be then shown in CAMP itself, as shown in figure 4, or can be exported to Microsoft Excel, where the user can analyze the results himself.

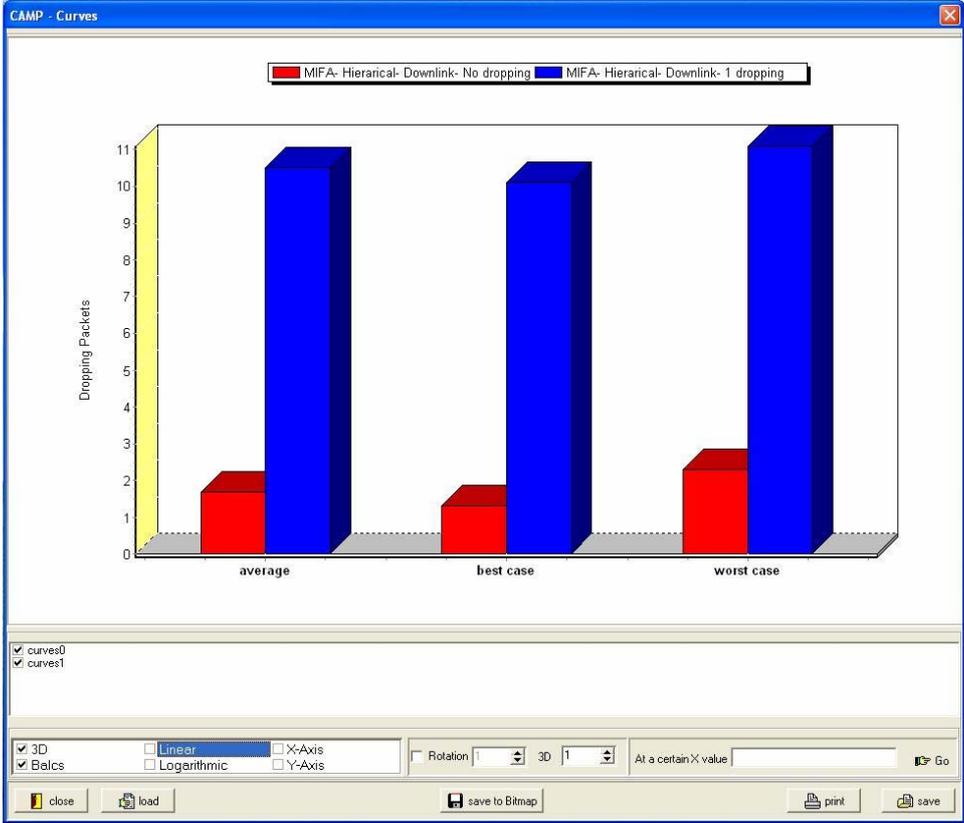


Figure 4: Display unit of CAMP

**II- APPLICATION OF CAMP**

CAMP implements a set of the known mobility management protocols such as Mobile IP (MIP) [1], Regional Mobile IP (RMIP) [2], Hierarchical Mobile IP (HMIP) [3], Cellular IP (CIP) [4], Handoff-Aware Wireless Access Internet Infrastructure (HAWAII) [5], Mobile IP Fast authentication protocol (MIFA) [6], low latency MIFA [7], Fast MIP [8], Proxy MIP [9]...etc. A set of the protocols supporting QoS simultaneously with mobility, e.g. Mobile RSVP [10], Hierarchical Mobile RSVP [11], QoS aware MIFA [12], is implemented too. The analysis comprises the performance and the total costs deploying different network topologies und different mobility models.

**a) Performance analysis**

Performance analysis comprises the estimation of the handoff latency, the expected number of dropped packets, the time required to reserve resources on a certain path

and the expected number of the packets sent as best effort while the reservation is in progress. CAMP analyses the performance taking into account the dropping of control messages, different network topologies and different mobility patterns. Figure 5 presents a sample for hierarchical network topology and the expected number of the dropped packets for MIFA, MIP, HMIPv6, AFA and HAWAII.

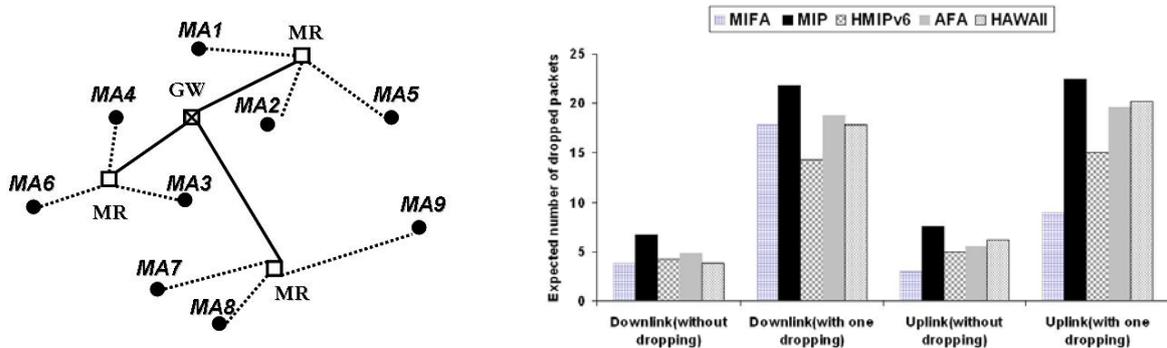


Figure 5: Expected number of dropped packets for MIFA, MIP, HMIPv6, AFA, HAWAII

### b) Total cost

The total cost, produced from a certain protocol, comprises the location update cost and the packets forwarding overhead. Calculating of this cost depends on the model developed in [13]. In addition, the total cost can be studied under user- or time-variant parameters. Figure 6 presents a sample of network topology and the location update cost resulting from CAMP for MIFA, RMIP and HAWAII.

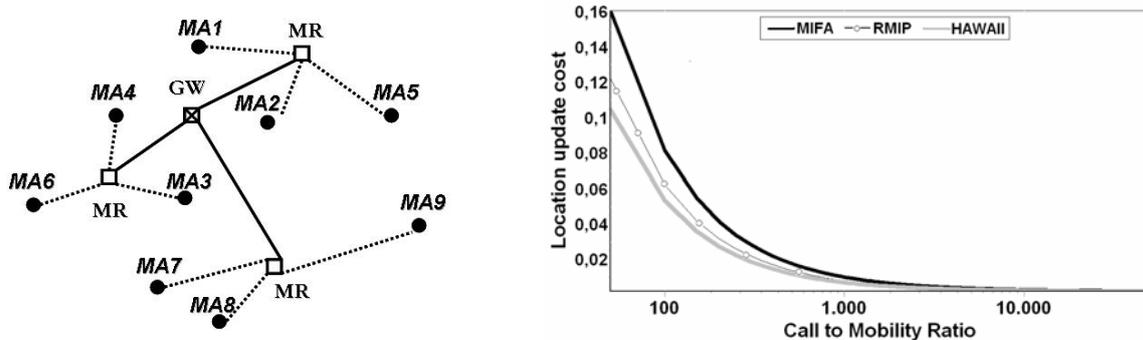


Figure 6: Location update cost for MIFA, RMIP and HAWAII

## IV- CONCLUSION

In this paper we have presented a new tool to analyze the mobility management protocols and the protocols supporting QoS simultaneously with mobility. The analysis comprises the studying of the performance and the costs resulted from a certain protocol. The performance can be studied under various network topologies and under different mobility patterns. The dropping of control messages can be taken into account too.

CAMP is a powerful tool, which implements a large set of protocols and enables a fast estimation of the performance of a certain protocol compared to many others. New protocols and new algorithms can be integrated simply.

Currently, we are implementing other protocols and trying to write the parameters files automatically. In addition, we are working on interconnecting between TRIAS [14] and CAMP to enable more users with specified rights to work remote on this tool.

#### References:

- [1] D. Johnson, C. Perkins, J. Arkko, " Mobility Support in IPv6 ", < draft-ietf-mobileip-ipv6-23.txt >, 2003.
- [2] D.E. Gustafsson, A. Jonsson, C. E. Perkins, " Mobile IPv4 Regional Registration ", < draft-ietf-mobileip-reg-tunnel-08.txt >, 2003.
- [3] H. Soliman, C. Castelluccia, K. El-Malki, L. Bellier, " Hierarchical Mobile IPv6 mobility management (HMIPv6)", < draft-ietf-mobileip-hmipv6-08.txt >, 2003.
- [4] A. Campbell, J. Gomez, C-Y. Wan, S. Kim, Z. Turanyi, A. Valko, " Cellular IP ", < draft-ietf-mobileip-cellularip-00.txt > 2000.
- [5] R. Ramjee, T. La Porta, S. Thuel, K. Varadhan, " IP micro-mobility support using HAWAII ", < draft-ietf-mobileip-hawaii-00 >, 1999.
- [6] A. Diab, A. Mitschele-Thiel, J. Xu, " Performance Analysis of the Mobile IP Fast Authentication Protocol ", Seventh ACM Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWIM 2004), 2004.
- [7] A. Diab, A. Mitschele-Thiel, R. Böringer, " Extension of Mobile IP for Fast Authentication and Low Latency Handoff ", 11th European Wireless, Cyprus, 2005.
- [8] R. Koodli, " Fast Handovers for Mobile IPv6 ", Internet draft, July 2004.
- [9] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury, B. Patil, " Proxy Mobile IPv6 ", < draft-ietf-mobileip-proxy-mip6-02.txt >, 2007.
- [10] A. K. Talukdar, B. R. Badrinarayan, A. Acharya, " MRSVP: A Resource Reservation Protocol for an Integrated Services Network with Mobile Hosts ", Wireless Nets., vol. 7, no. 1, 2001.
- [11] C.-C. Tseng et al, " HMRSVP: A Hierarchical Mobile RSVP Protocol ", Proc.Int'l. Conf. Distrib. Comp. Sys. Wksp., April, 2002.
- [12] E. Alnasouri, A. Mitschele-Thiel, R. Böringer, A. Diab, " QOMIFA: A QoS enabled Mobility Management Framework in ALL-IP Network ". 17th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'06), Helsinki, Finland, 2006
- [13] A. Diab, A. Mitschele-Thiel, R. Böringer, " Evaluation of Mobile IP Fast Authentication Protocol compared to Hierarchical Mobile IP ", IEEE Conference On Wireless and Mobile Computing, Networking and Communications WiMob'2005, Montreal, August 2005.
- [14] R. Böringer, A. Mitschele-Thiel, G. Scharfe, " TRIAS: Optimization of Complex Networks ", 13th GI/ITG Conference on Measurement, Modeling, and Evaluation of Computer and Communication Systems (MMB 2006), Nürnberg, March 2006

#### Authors:

Dipl.-Ing Ali Diab

Prof. Dr.-Ing. habil. Andreas Mitschele-Thiel

TU - Ilmenau, Faculty for Informatics und Automation, Gustav-Kirchhoff-Str. 1, P.O.B. 10 0565  
98693, Ilmenau

Phone: +49 3677 69 2819

Fax: +49 3677 69 1220

E-mail: {ali.diab|mitsch}@tu-ilmenau.de





S. Bräunig / H.-U. Seidel

## **Learning Signal and Pattern Recognition with Virtual Instruments**

### **EINLEITUNG**

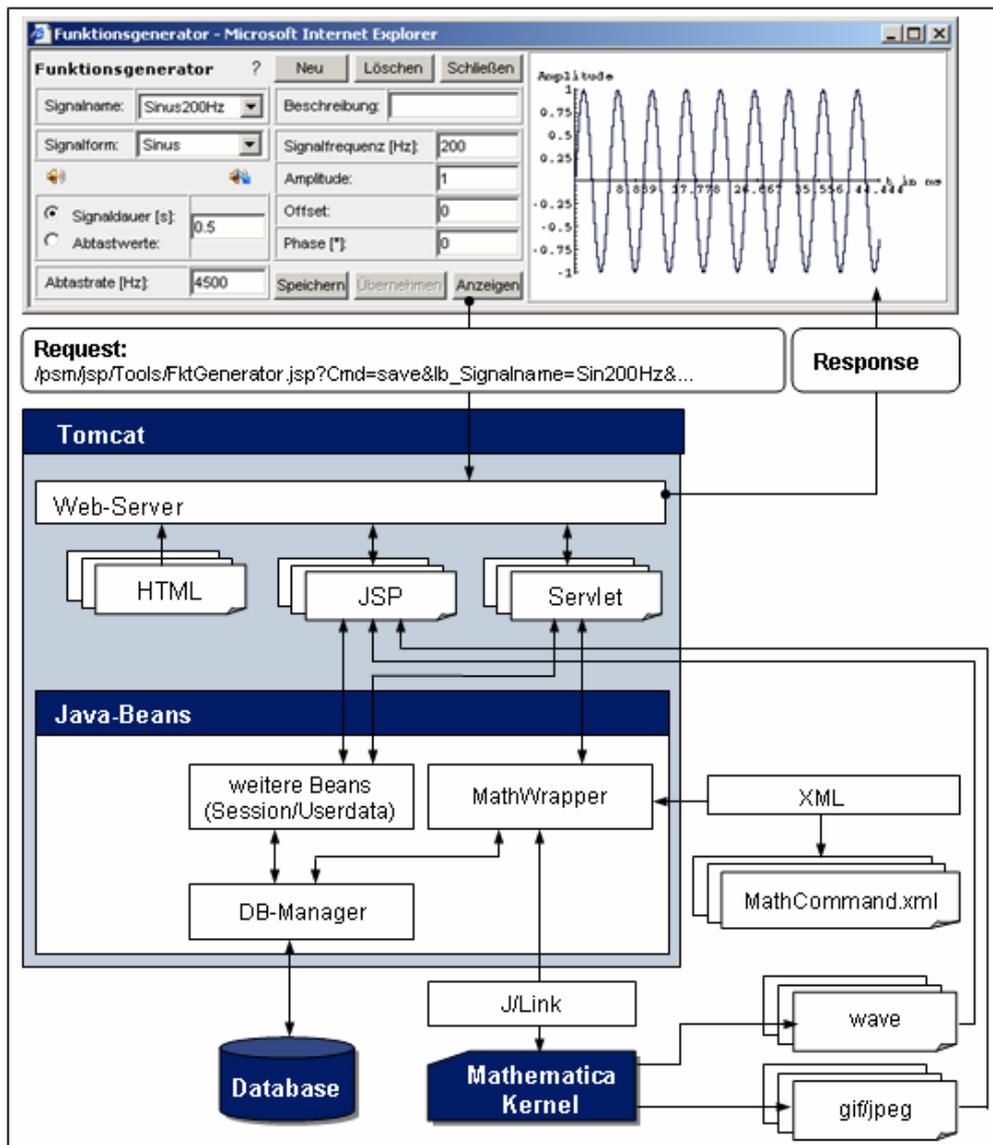
Die Methoden und Algorithmen der Signalanalyse und Mustererkennung erfordern ein vielseitiges und meist sehr komplexes mathematisches Instrumentarium. Gleichzeitig werden in der Regel umfangreiche Datenmengen bzw. hochdimensionale Datenräume betrachtet. Durch den Einsatz des Rechners, des Internets und moderner Softwaretechnologien bei der fachspezifischen Ausbildung eröffnen sich neue Möglichkeiten Lehr- und Lernprozesse zu unterstützen und zu verbessern. Eingegliedert in bewährte Unterrichtsformen sollen durch Animationen, Simulationen und interaktive Elemente folgende Effekte erreicht werden:

- Unterstützung der Lernenden bei der Bewältigung der komplexen signaltheoretischen Zusammenhänge durch Einsatz geeigneter Visualisierungstechniken
- dem Kenntnisstand angepasste Modellierung und Simulation der betrachteten Sachverhalte
- Aktivierung und Motivierung der Lernenden in traditionellen Unterrichtssituationen
- Förderung praktischer Erfahrungen und konstruktivistischen Lernens

Entsprechend dem Konzept des blended-learning sind dabei Methoden und Mittel gefragt, die sich individuell an die jeweilige Lernsituation anpassen lassen und so für den Einsatz in Vorlesungen, Seminaren, Praktika und zum Selbststudium geeignet sind. Der von uns verfolgte Ansatz basiert auf der Entwicklung und dem Einsatz virtueller Geräte, die, eingebunden in eine multimediale webbasierte Lernumgebung, die genannten Einsatzszenarien unterstützen und so die gewünschten Effekte erzielen.

### **TECHNISCHER HINTERGRUND DER VIRTUELLEN GERÄTE**

Die virtuellen Geräte stellen interaktive Softwaremodule dar, in denen jeweils funktionsverwandte Algorithmen zusammengefasst sind. Neben den eigentlichen mathematischen Funktionen stellen sie die Datenbasis bereit, organisieren die für die Funktionen notwendigen Parametereingaben und erzeugen geeignete Grafiken, Animationen und Sounddateien zur auditiven und visuellen Reproduktion der berechneten Ergebnisdaten. Die wesentlichen verwendeten Technologien zur Implementierung der virtuellen Geräte sind in der folgenden Abbildung am Beispiel des Funktionsgenerators dargestellt.



**Abbildung 1** Technisches Konzept zur Realisierung virtueller Geräte am Beispiel *Funktionsgenerator*

- Als Webserver und JSP-Engine wird der Tomcat-Server verwendet.
- Das Kernstück für die Implementierung der mathematischen Funktionen, die Erzeugung von Simulationen und die Berechnung der Ergebnisdaten (Grafiken, Signaldaten, Sounddateien) ist die Nutzung des Computer-Algebra-Systems Mathematica. Über die Java-API J/Link und eine entwickelte XML-Schnittstelle erfolgt die Anbindung zur Webanwendung. Umfangreiche Mathematica-Funktionsbibliotheken stehen zur Verfügung und einmal entwickelter Programmcode kann in verschiedenen Kontexten genutzt werden. Durch die definierten Schnittstellen lassen sich umfangreicher bzw. häufig verwendeter Programmcode elegant auslagern.
- Die Datenbank mit der Anbindung über JDBC hat die Funktion, alle persistenten Daten wie: Nutzerdaten, Rechte, Signaldaten, Aufgaben und Ergebnisse zu speichern und zu verwalten. So lassen sich die von einem Nutzer einmal erzeugten Daten in verschiedenen Kontexten wieder verwenden, verschiedene Einsatzszenarien organisieren bzw. automatische Auswertemechanismen anschließen.

## INHALTE UND KONZEPTION

Die Einsatzmöglichkeiten, didaktischen Konzepte und Erfahrungen der Lernumgebung mit den virtuellen Geräten im Rahmen der Lehrveranstaltungen „Grundlagen der Signal- und Mustererkennung“ sollen am Beispiel des Shannonschen-Abtasttheorems vorgestellt werden. Die Lehrveranstaltungen finden im 5. bzw. 6. Semester mit jeweils 14 Vorlesungen und 7 Seminaren statt. Abgeschlossen wird der Themenkomplex mit einem Praktikum. Ziel ist dabei, den komplexen Prozess der Signalverarbeitung und Mustererkennung angefangen von der problemangepassten Messwerterfassung bis hin zur Klassifikation zu betrachten. Im Vordergrund stehen Untersuchungen zur Geräuschanalyse elektromechanischer Systeme bzw. Sprachsignale.

Die Basis der Signalverarbeitung ist meist eine mittels analoger Messtechnik aufgenommene physikalisch messbare Größe (z.B. Temperatur, Beschleunigung, Schalldruck), die in der Regel durch das Messsystem in ein analoges elektrisches Signal gewandelt wird. Um dieses analoge Signal mit einem Rechner und den Methoden der digitalen Signalverarbeitung weiterverarbeiten zu können, ist es mittels eines Analog-Digital-Wandlers zu digitalisieren. In diesem Zusammenhang und auch bei der Interpretation der daraus resultierenden Ergebnisse stellt sich das Abtasttheorem als grundlegendes Theorem sowohl bei der Signalanalyse, der Nachrichtentechnik als auch der Informationstheorie heraus. Dessen Kenntnis, Verständnis und Anwendbarkeit gilt als wesentliche bei den Studierenden zu entwickelnde Kernkompetenz .

Nach Einführung der fundamentalen theoretischen Zusammenhänge anhand traditioneller Unterrichtsmedien und rechnergestützter Beiträge in der Vorlesung, die vorwiegend auf symbolischer Ebene erfolgt, werden zur Illustration und zur Beschreibung ausgewählter Phänomene einzelne in der Regel einfache bzw. besonders anschauliche Beispiele vorgestellt. Die Quintessenz bei der Einführung des Abtasttheorems ist, dass die Informationen, die ein mit der Grenzfrequenz  $f_g$  begrenztes Basisbandsignal trägt, bei Einhaltung des Abtasttheorems wieder rekonstruiert werden kann. Dies soll nun gezielt im Rahmen der Seminare und Praktika und im Selbststudium untersucht werden.

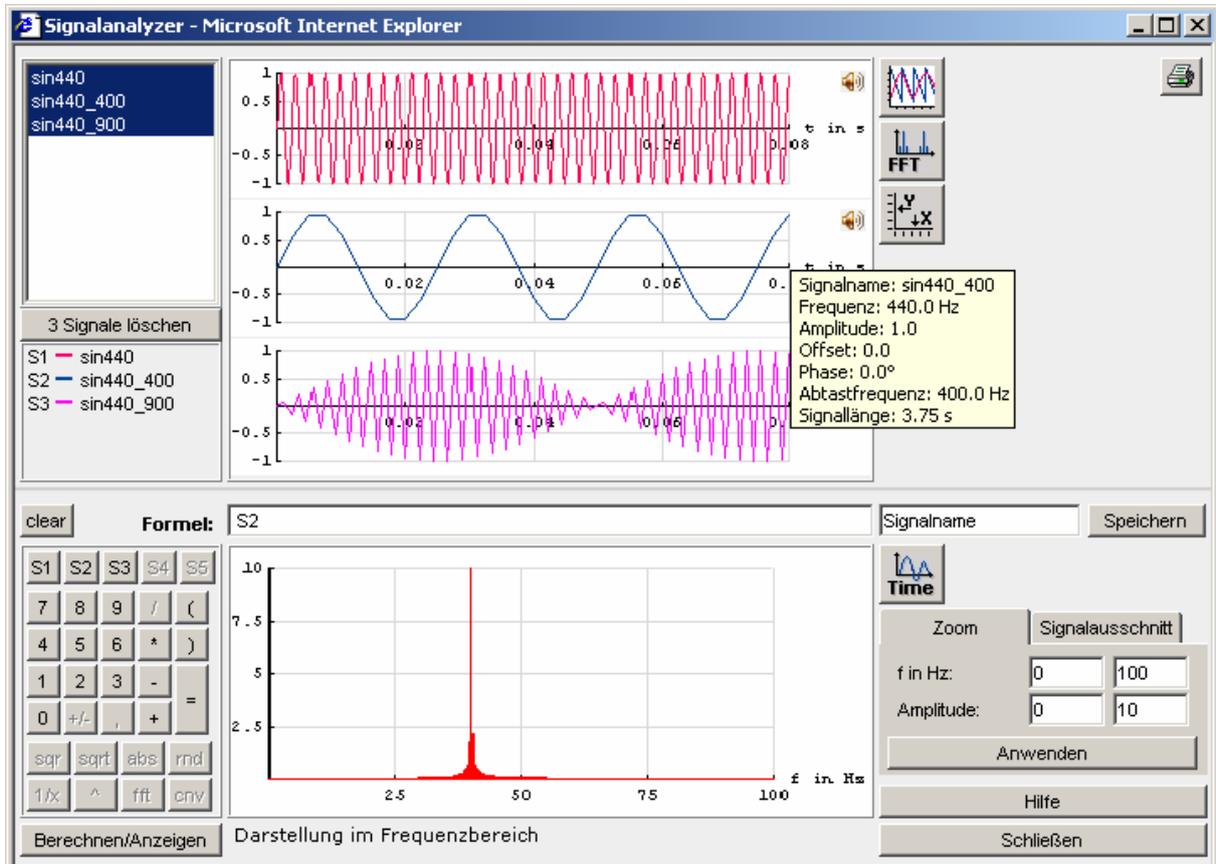
## LERNEN MIT VIRTUELLEN GERÄTEN

Ausgehend von der mathematisch einfach zu formulierenden Forderung des Shannonschen Abtasttheorems

$$f_{\text{Abtast}} > 2 f_{\text{GrenzSignal}}$$

werden in den Seminaren weitere Beispiele und Aufgaben behandelt. Die Seminare finden in einem Computerkabinett statt. Neben dem Einsatz Mathematica-basierter elektronischer Dokumente werden die virtuellen Geräte der Lernumgebung eingesetzt. Der Einsatz dieser interaktiven Medien gestattet den Lernenden, sich aktiv am

Lernprozess zu beteiligen. Anhand vorgegebener Fragestellungen werden z.B. die Effekte der Unter- und Überabtastung anhand einfacher, synthetisch zu erzeugender Sinus- und Multisinusfunktionen untersucht. Die virtuellen Geräte (Siehe Abbildung 2) gestatten neben der visuellen auch die akustische Wiedergabe der Ergebnisse.



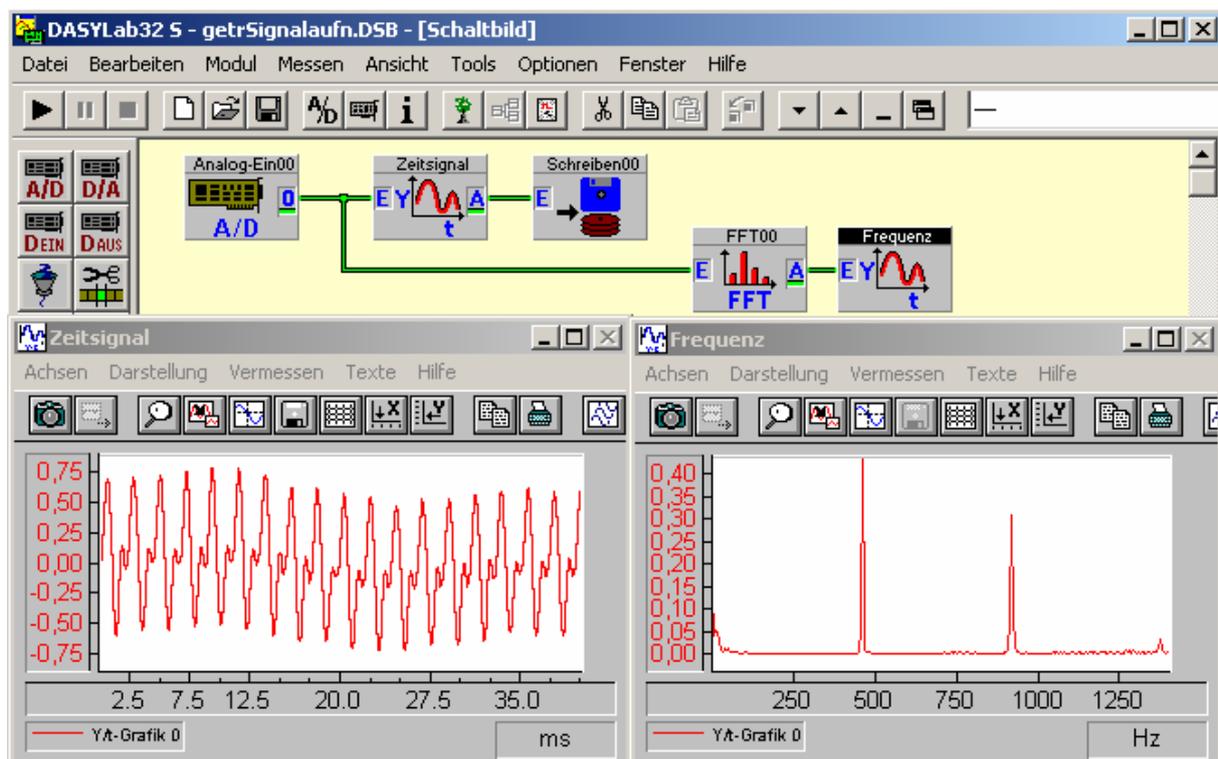
**Abbildung 2** Darstellung der abgetasteten Zeitsignale im virtuellen Gerät *Signalanalyzer*

Des weiteren werden die Studierenden aufgefordert, selbstständig geeignete Abtastraten für vorgegebene komplexere Signale zu finden und die Ergebnisse visuell und akustisch zu überprüfen. Dabei werden weitere Problemstellungen, die unmittelbar mit dem Abtasttheorem zusammenhängen, abgeleitet und diskutiert, z.B. die Behandlung quasistationärer bzw. instationärer Signale, der Einsatz von Anti-Aliasing-Filtern, die Anwendung von Fensterfunktionen und der Kurzzeit-Fourier-Transformation.

Abgerundet und praktisch untermauert werden die Erkenntnisse zum Abtasttheorem im Rahmen des Praktikums. So kann gleichzeitig überprüft werden, in wieweit, der behandelte Stoff verstanden wurde und auf praktische Aufgabenstellungen übertragen werden kann. Das Praktikum findet im wesentlichen auf der Basis der webbasierten Lernumgebung statt. Sie organisiert zudem die Nutzer- und Aufgabenverwaltung. Die Teilnehmer haben sich über ein Passwort zu identifizieren. Jeder Teilnehmer erhält so, die für ihn vorgesehenen Aufgaben und Signaldaten. Die Aufgaben sind in der Regel mehrstufig gestaltet. Zur Lösung sind meist mehrere Arbeitsschritte durchzuführen, die vom

Teilnehmer individuell festzulegen sind. Als Resultat entstehen Ergebnisdatensätze, die wiederum nutzergebunden gespeichert werden und für weitere Arbeitsschritte bzw. Betrachtungen zur Verfügung stehen.

Bei der Behandlung der Thematik des Abtasttheorems spielt besonders der Aspekt der korrekten Messwertaufnahme und Analog-Digital-Wandlung eine Rolle. Dazu werden die Praktikumsteilnehmer aufgefordert, selbstständig Messwertreihen zu erfassen und die korrekte Digitalisierung zu überprüfen und zu begründen. Unterstützt durch die Messwert- und Datenverarbeitungssoftware DasyLab von National Instruments sind zu den bereits vorhandenen (zum größten Teil synthetisch generierten) Signalen nun eigenständig über ein Mikrofon und die Soundkarte akustische Signale aufzunehmen.



**Abbildung 3** Einsatz von DasyLab zur Signalaufnahme

Die in der Abbildung 3 dargestellte Konfiguration zur Messwertaufnahme mit DasyLab lässt sich schnell und problemlos durch das Prinzip der visuellen Programmierung an die gewünschte Problematik anpassen. Auch hier wird das Konzept der virtuellen Geräte eingesetzt. Neben der Auswahl und Parametrisierung der eigentlichen Geräte lässt sich auf graphischer Basis durch Verbinden der einzelnen Funktionsmodule der geplante Signalfluss festlegen. So sind auch komplexe signaltechnische Systeme konfigurierbar. Die mit DasyLab aufgenommenen Daten sind vermittels einer Importfunktion in die Lernumgebung zu importieren und können dort weiterverarbeitet werden.

Soll lässt sich z.B. die Problematik der notwendigen Bandbegrenzung analoger Signale vor der Abtastung zur Vermeidung von Aliasing-Effekten studieren. Nur korrekt erfasste Messwertreihen gestatten auch eine sinnvolle und korrekte Weiterverarbeitung und

Interpretation der Ergebnisse, was auf den Aspekt der problemangepassten Messwertaufnahme aufmerksam machen soll.

### ZUSAMMENFASSUNG

Das Konzept für die Lernumgebung und die virtuellen Geräte ist so gestaltet, dass sich die folgenden Einsatzmöglichkeiten ergeben:

- Nutzung in **Präsenzlehrveranstaltungen**: zur Veranschaulichung des aktuell behandelten Lehrstoffes (Bild, Ton, Simulation) und durch aktive Einbeziehung der Lernenden
- **Unterstützung des Selbststudiums**: Nachvollziehen demonstrierter Beispiele mit der Möglichkeiten weiterer Parametervariationen und Bearbeiten weiterer Aufgaben
- Vorbereitung und Durchführung eines **webbasierten Praktikums** innerhalb der Lernumgebung: komplexe Aufgabenstellungen sind mittels der virtuellen Geräte zu lösen. Bearbeitungsreihenfolge und Parameterwahl ist selbstständig festzulegen und die Geräte anzuwenden.
- Kombination **professioneller und selbst entwickelter Verarbeitungssysteme**: Ergänzung vorhandener Tools mit den zur Lösung spezifischer Fragestellungen notwendigen Eigenentwicklungen. Dies gibt einen Einblick in ingenieurmäßiges Arbeiten.
- Anwendung der virtuellen Geräte auf **eigene Fragestellungen**: Verarbeitung eigener Signaldaten und Verfügbarkeit der Ergebnisdaten

Unterstützt durch den zeit- und ortsunabhängigen Zugriff durch Verwendung der Internet-Technologien lassen sich noch weitere Einsatzszenarien ableiten. Entscheidend ist, dass sich die virtuellen Geräte sowohl kontextgebunden z.B. bei der Lösung von Aufgaben im Praktikum als auch kontextunabhängig verwenden lassen.

### References:

- [1] Seidel, H.-U.; Burger, P.: Computergestützte Lehre in Signalanalyse mit Mathematica; 4. Workshop „Multimedia für Bildung und Wirtschaft“, 28./29.09.2000; TU Ilmenau
- [2] Bräunig, S.: Konzeption und Realisierung einer webbasierten Lernumgebung für die Signal- und Mustererkennung; Diss.; TU Ilmenau; 2006
- [3] Karrenberg, U.: Signale–Prozesse–Systeme: eine multimediale und interaktive Einführung in die Signalverarbeitung; Springer Verlag; 2002; Berlin
- [4] Neaf, O.: Real laboratory, virtual laboratory or remote laboratory: What is the most efficient way?; International Journal of Online Engineering; www.i-joe.org; Vol.2, No. 3; 2006

### Authors:

Dr.-Ing. Sylvia Bräunig  
Univ.-Prof. Dr.-Ing. habil. Heinz-Ulrich Seidel  
TU Ilmenau, Helmholtzplatz 2, PF 100 565  
98693 Ilmenau  
Phone: +49 3677 692628  
Fax: +49 3677 691125  
E-mail: sylvia.braeunig@tu-ilmenau.de

S. Lambeck

## Use of Rapid-Control-Prototyping Methods for the control of a nonlinear MIMO-System

### Abstract

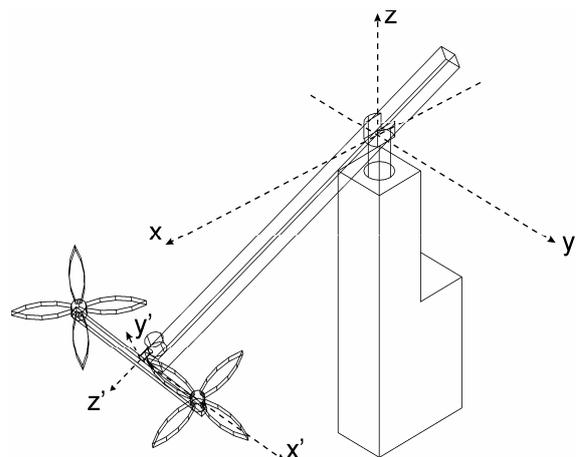
In the last few years different Rapid-Control-Prototyping (RCP) Methods, like “Hardware in the Loop (HiL)”, “Software in the Loop” (SiL) and automatic code generation, became an important part of the control design and implementation for industrial applications. Our aim is to familiarize the students with the knowledge in this special field of control testing and realization. Therefore different experimental set-up’s associated to the juniorprofessorship of automation technology at the technical university of Ilmenau are built up. In this paper one the set up associated to a 3D-model of a helicopter will be described by its equations of motion and the different ways of control design will be shortly summarized. The use of the RCP-Methods related to the specifics of a control education course will be also presented in this paper.

### Introduction

The system treated in this paper is illustrated in Fig.1 and represents a classical example of a Multi-Input Multi-Output (MIMO) System.

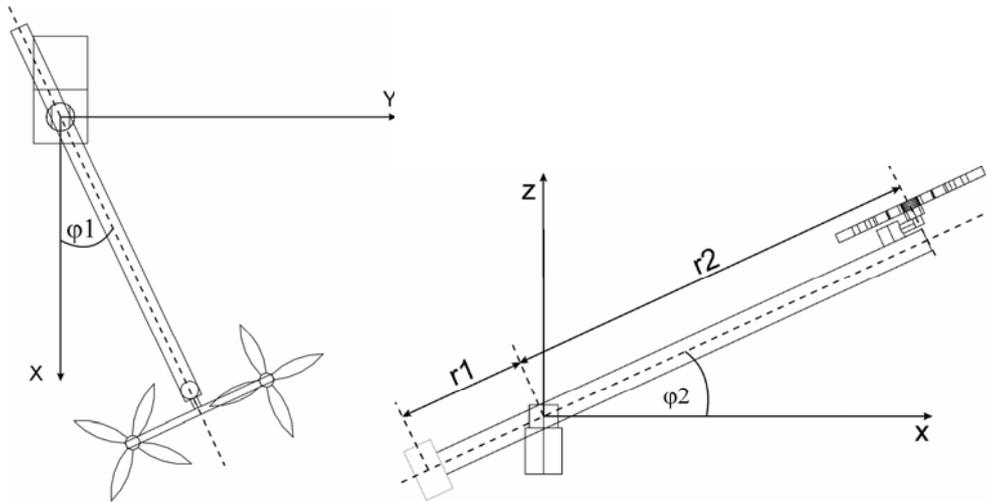


Fig. 1: Experimental set-up



The two inputs of the system are the lift forces engendered by the two rotors, which are related to the control voltage of the driving motors in an approximately linear manner. As

outputs of the system the 2 angles  $\varphi_1$  and  $\varphi_2$  illustrated in Fig.2 can be chosen.



**Fig. 2: Output-Angles of the system**

The equations of motion will be derived in the next section using the Lagrange formalism. We receive a representation in state space form with two coupled subsystems (turning and raising of the helicopter), which complicates the control design. In a first step, the design of simple decentralized controllers for each separate axis can be done using well-known techniques in the frequency domain (e.g. root locus) which are applicable by undergraduates at the beginning of a control theory course. The performance is only satisfactory in the near of a predefined operating point. Advanced control design techniques (shortly summarized in Section 3) like state space controllers based on a linearized model or Input-Output Linearization can be applied in a second step with the result of a better control performance in the case of simultaneous movements of the different axes.

In our set-up's MATLAB / SIMULINK was chosen as a standard tool for simulation and control design. For the realization of the different RCP-Methods described in Section 4, Hardware and Software components developed and distributed by the dSPACE GmbH<sup>1</sup> were used. So a fast and efficiency control design with online testing of the developed algorithms became possible and in connection with a comfortable GUI, where the different control- and manipulated variables are visualized, the whole set-up represents a challenging possibility for the students to consolidate the knowledge from the control theory courses.

### Model Derivation

The equations of motion for the three axes can be derived using the Lagrange formalism

<sup>1</sup> www.dspace.com

based on the kinetic and the potential energy of the system which results in the following relations for  $\varphi_1$  and  $\varphi_2$  [1]

$$\begin{aligned}\ddot{\varphi}_2 &= \frac{M}{J_2} + \frac{g}{J_2} \cos(\varphi_2 + \varphi_0)[m_1 r_1 - m_s s - m_2 r_2] - \frac{1}{2} \frac{J_2 - J_{Stab}}{J_2} \sin(2[\varphi_2 + \varphi_0]) \dot{\varphi}_1^2 \\ \ddot{\varphi}_1 &= \dot{\varphi}_1 \dot{\varphi}_2 \frac{J_2 - J_{Stab}}{J_1} \sin(2[\varphi_2 + \varphi_0]) - \frac{r_2}{J_1} \cos(\varphi_2 + \varphi_0) \sin(\varphi_3)(F_1 + F_2)\end{aligned}\quad (1)$$

$J_i$  represents the moment of inertia in the different movement directions. The corresponding equation for  $\varphi_3$  contains the driving forces from the rotors as inputs of the system.

$$\ddot{\varphi}_3 = \frac{r_3(F_{AR} - F_{AL}) - 2 \cdot m_{Motor} \cdot g \cdot r_{s3} \cdot \cos(\varphi_2) \cdot \sin(\varphi_3)}{2J_3}\quad (2)$$

The nonlinear system can be noted in state-space representation after a linearization near the operating point (0,0,0), which implies different approximations for the trigonometric expressions in (1) and (2).

$$\begin{aligned}\begin{pmatrix} \dot{\varphi}_1 \\ \ddot{\varphi}_1 \\ \dot{\varphi}_2 \\ \ddot{\varphi}_2 \\ \dot{\varphi}_3 \\ \ddot{\varphi}_3 \end{pmatrix} &= \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \frac{-m_{Motor} \cdot g \cdot r_{s3}}{J_3} & 0 \end{bmatrix} \begin{pmatrix} \varphi_1 \\ \dot{\varphi}_1 \\ \varphi_2 \\ \dot{\varphi}_2 \\ \varphi_3 \\ \dot{\varphi}_3 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ -\frac{r_2}{J_1} \varphi_3 & -\frac{r_2}{J_1} \varphi_3 \\ 0 & 0 \\ \frac{r_2}{J_2} & \frac{r_2}{J_2} \\ 0 & 0 \\ \frac{r_3}{2J_3} & -\frac{r_3}{2J_3} \end{pmatrix} \begin{pmatrix} F_{A1} \\ F_{A2} \end{pmatrix} \\ \underline{y} &= \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \dot{\varphi}_1 \\ \varphi_2 \\ \dot{\varphi}_2 \\ \varphi_3 \\ \dot{\varphi}_3 \end{pmatrix}\end{aligned}\quad (3)$$

### Control Design

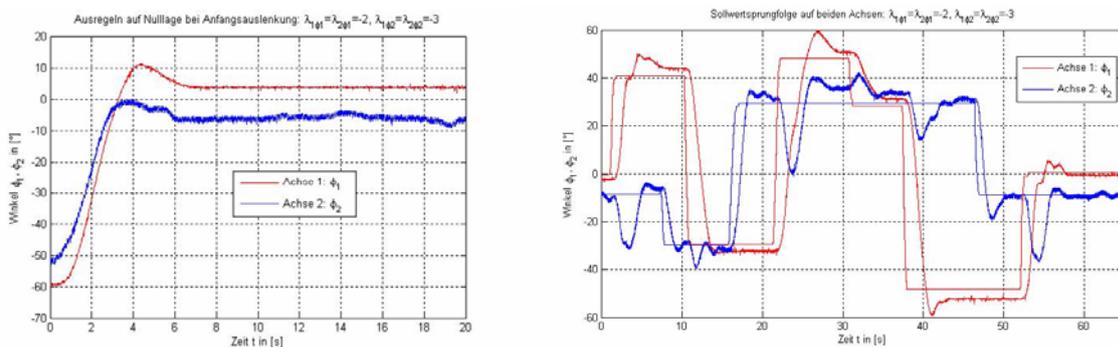
Following the design proposed in [2] a simple control strategy which is based on a linear state-space controller can be chosen. Due to the coupled two axes respectively the coupling of the two manipulated signals with the movement of the two axes, the design of a decoupling controller in connection with a stationary filter seems to be suitable in a first step. The filter can be calculated using (3)

$$\underline{S} = \begin{pmatrix} \underline{c}_1 \underline{A}^{\delta_1-1} \underline{B} \\ \dots \\ \underline{c}_p \underline{A}^{\delta_p-1} \underline{B} \end{pmatrix}^{-1} \begin{pmatrix} \underline{\mu}_1 & \underline{0} \\ \dots & \dots \\ \underline{0} & \underline{\mu}_p \end{pmatrix} \quad (4)$$

The corresponding controller results following [2] in

$$\underline{K} = \begin{pmatrix} \underline{c}_1 \underline{A}^{\delta_1-1} \underline{B} \\ \dots \\ \underline{c}_p \underline{A}^{\delta_p-1} \underline{B} \end{pmatrix}^{-1} \begin{pmatrix} \underline{c}_1 \underline{A}^{\delta_1} + \sum_0^{\delta_1-1} a_{1n}^* \underline{c}_1 \underline{A}^n \\ \dots \\ \underline{c}_p \underline{A}^{\delta_p} + \sum_0^{\delta_p-1} a_{pn}^* \underline{c}_p \underline{A}^n \end{pmatrix} \quad (5)$$

$\delta$  represents the difference order of the system and  $a^*$  are real parameters which can be calculated based on a simple pole placement. Fig. 3 illustrates the control of the two axes from a predefined starting point into the operating point (0,0,0) on the left side and the control behaviour after a sequence of setpoint changes on the right side.



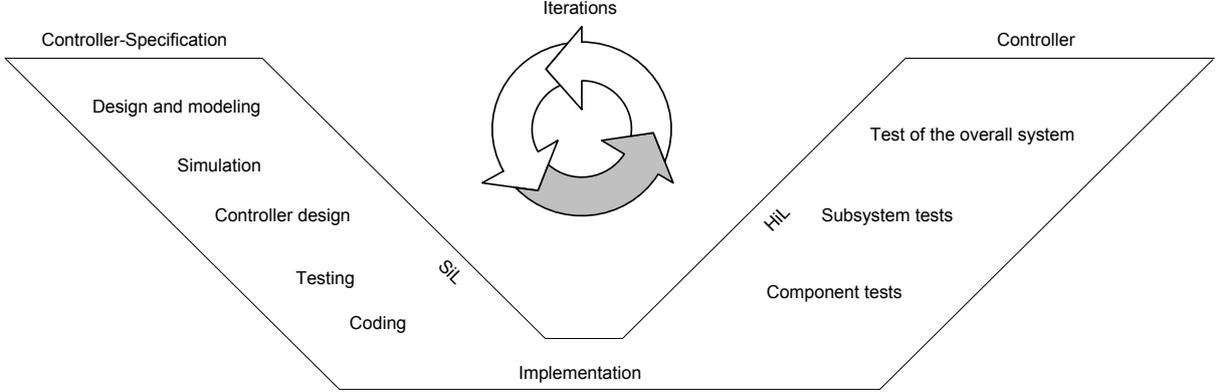
**Fig. 3: Control performance**

The movement of  $\varphi_1$  has a significant influence of  $\varphi_2$  as can be seen on the right side of Fig.3. This effect is caused by the limitations of the design based on a linearized model. We can improve the performance by an alternative control design, e.g. exact input-output linearization [3]. The application of this method for the considered system is described in detail in [1]. The idea is to calculate a linear controller in connection with a matrix, which eliminates the nonlinear coupling between the inputs and the outputs. With the input-output linearization a significant better performance can be reached. The inter-influence of the different movement directions is minor compared to Fig. 3.

### Rapid Control Prototyping

In [4] a systematic procedure for the control design and implementation was presented, which is related to the known V-Model from software engineering. The procedure is

illustrated in Fig. 4. Beginning on the upper left side of the V, the complete system analysis, modeling, simulation and control design takes place on the left side. After coding and implementing the algorithm on the specific target, the tests of the different components up to the test of the overall system takes place on the right side of the V.



**Fig. 4: V-Model**

The classical V-Model contains the iterations at first only in vertical direction, which implies different drawbacks [4]. The idea of RCP was the introduction of horizontal iterations, whereby e.g. the applicability of the developed control algorithm on the specific system can be tested direct after the design. The most known terms in this field are “Software in the Loop” (SiL – the control algorithm will be executed on a high-performance Hardware, which is connected with parts of the real process) and “Hardware in the Loop” (HiL – the control algorithm will be implemented on the target, which is connected with a simulation model of the real process). So, the focus using SiL is the investigation of the suitability of the algorithm and the focus using HiL is the investigation of the implemented algorithm on the target considering the different restrictions concerning the computational power and the environmental conditions.

For an effective use of the mentioned RCP-Methods a continuous tool-chain is a necessary supposition. We use MATLAB/SIMULINK in connection with different tools distributed by the dSPACE GmbH. The two different configurations are illustrated in Fig. 5. For the SiL-tests the configuration on the right-side of the Figure is used. The different control strategies are developed using MATLAB/SIMULINK. After this, real-time code is generated from the block diagrams (using the so-called “Real-time Interface” RTI) and automatically downloaded on a high-performance Power-PC. Using AD- and DA-Channels the process values are visualized on a host PC and the performance of the control algorithm can be online evaluated. As a specific target for the control algorithm the 32bit Controller TC1775 (Infineon) is used in the set-up. So, for the HiL-tests (left

side of Fig. 5) the developed algorithm has to be transferred on this target.

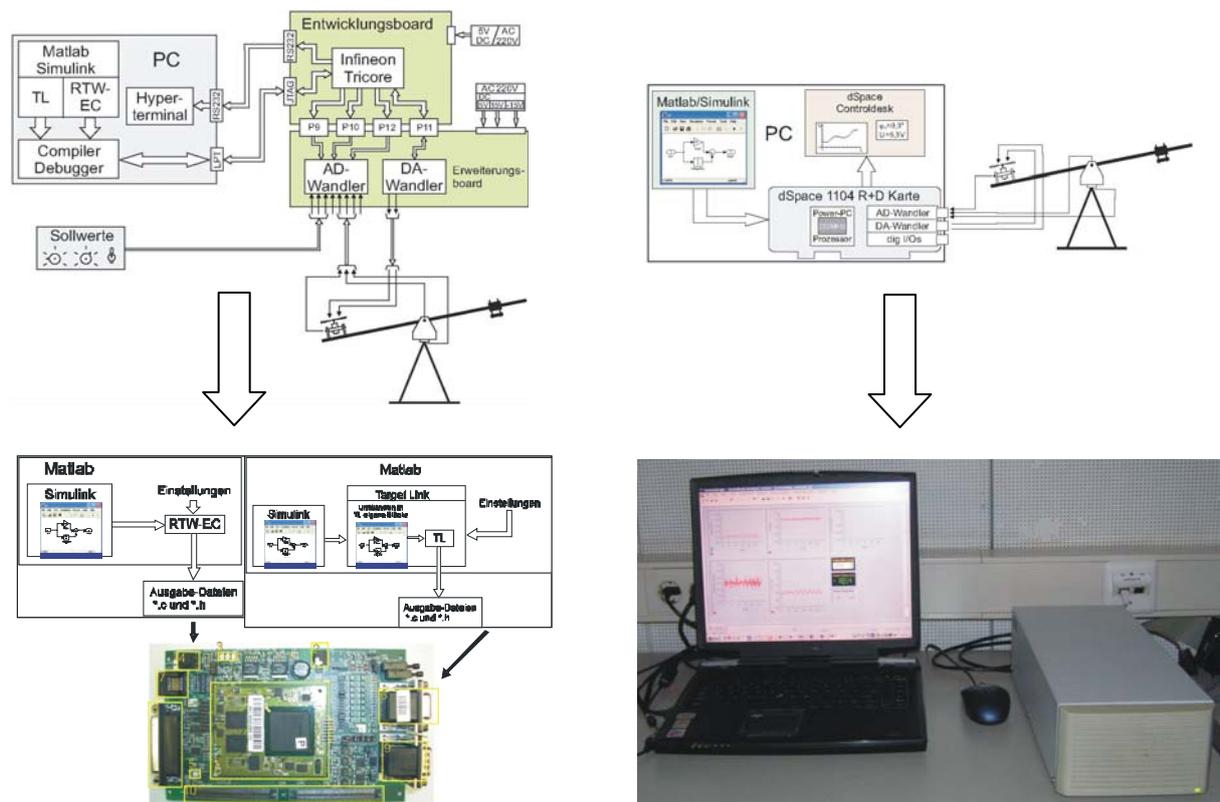


Fig. 5: Different RCP-Configurations for the helicopter set-up

We choose two different Code generators with the capability to generate optimized real-time code for specific targets. The first one is “TargetLink” distributed by the dSPACE GmbH and the second one “RTW-Embedded Coder” developed by “The Mathworks”. A comparison between these two tools was done in a diploma thesis [5].

#### References:

- [1] Busch, Christian:” Entwurf eines linearisierenden Entkopplungsreglers für ein nichtlineares MIMO-System”, Studienarbeit am Institut AS, Juniorprofessur Automatisierungstechnik
- [2] Föllinger, Otto: “Regelungstechnik”, 8. Auflage, 1994, Hüthig Verlag Heidelberg
- [3] Wey, Torsten: „Nichtlineare Regelungssysteme – ein differentialalgebraischer Ansatz“, 2002, Teubner Verlag
- [4] Abel, Dirk; Bollig, Alexander: „Rapid Control Prototyping – Methoden und Anwendungen“, 2006, Springer Verlag Berlin
- [5] Tischer, Felix: “Vergleichende Untersuchungen zur Effizienz automatischer Codegeneratoren und deren Anwendung auf regelungstechnische Problemstellungen”, Diplomarbeit am Institut AS, Juniorprofessur Automatisierungstechnik, Februar 2006

#### Authors:

Jun.Prof. Dr.-Ing. Steven Lambeck  
 TU Ilmenau, Institute of Automation and System Science P.O.B. 100565  
 98684 Ilmenau  
 Phone: +49 3677691432  
 Fax: +493677691415  
 E-mail: steven.lambeck@tu-ilmenau.de

Reinhard Pittschellis

## **Automatisierungstechnische Ausbildung an Gymnasien**

### **Ausgangslage**

Heute gibt es an deutschen Gymnasien selten Unterricht im Fach Technik, sieht man von einigen speziellen technischen Gymnasien ab. Dies ist um so dramatischer, als Deutschland sich derzeit einem immer größer werdenden Mangel an Ingenieur Nachwuchs gegenüber sieht: Es besteht die große Gefahr, dass dieser Mangel unseren Wohlstand bedrohen wird, der ja maßgeblich auf der Entwicklung und Produktion hochwertiger, technischer Güter basiert.

Diese Situation beginnt sich zu ändern. In einigen Bundesländern wird jetzt das Fach Technik in die Gymnasien eingeführt, so z.B. unter dem Begriff Naturwissenschaft und Technik (NWT) in Baden Württemberg oder Natur und Technik (NuT) in Bayern.

Leider sind die Ausgangsbedingungen in Baden Württemberg alles andere als gut:

- Technik wird nicht selbständiges Fach, sondern als Fächerverbund eingeführt.
- In aller Regel ist wenig Budget für die Beschaffung von Lehr- und Lernmaterialien vorhanden.
- Es existieren keine Fachräume.
- Es gibt keine ausgebildeten Techniklehrer an Gymnasien.

Es wird davon ausgegangen, daß die naturwissenschaftlich ausgebildeten Lehrer auch problemlos Technik unterrichten können. Wie alle Techniker wissen, ist Technik aber nicht angewandte Naturwissenschaft, obwohl Technik natürlich auf vielen anderen Wissenschaften basiert. Vielmehr sind die Ingenieurwissenschaften eine eigenständige Disziplin mit eigenen Methoden, die es verdienen, auch als solche wahrgenommen zu werden. So wird z.B. im Lehrplan für das Fach NWT in Klasse 10 in

Baden Württemberg betont, daß immer der Bezug zu den Basiswissenschaften herzustellen sei [1]. Technik als eigenständige Wissenschaft kann aber so nicht von den Schülern wahrgenommen werden. Vielmehr besteht die große Gefahr, daß die Lehrer, die ja alle ausgebildete Naturwissenschaftler sind, das neue Fach im Sinne eines

erweiterten Physik- oder Biologieunterrichtes zu nutzen mit einem kurzen Verweis auf technische Anwendungen. Beispiele aus der Praxis bestätigen diesen Verdacht [2].

### **Anforderungen an ein Lernsystem für allgemeinbildende Schulen**

Zunächst muß betont werden, dass es beim Technikunterricht an allgemeinbildenden Schulen nicht um eine vorgezogene Berufsausbildung, sondern um einen Teil der Allgemeinbildung in einer zunehmend technisierten Welt geht [4].

Unter den möglichen Technologien, die im Unterricht behandelt werden können, nimmt die industrielle Automatisierungstechnik einen besonders wichtigen Stellenwert ein. Die Automatisierungstechnik ist einer der Exportschlager der deutschen Wirtschaft. Hier nimmt die Bundesrepublik einen internationalen Spitzenplatz ein, die Branche wächst stetig [3]. Daher erscheint es sinnvoll, Jugendliche mit dieser Technologie vertraut zu machen. Eine Beschäftigung mit Automatisierungstechnik bietet im Sinne einer ganzheitlichen Bildung auch einen guten Ansatzpunkt für eine Diskussion über die gesellschaftlichen Auswirkungen des Einsatzes technischer Lösungen, da nahezu jeder von automatisierten Dingen umgeben ist.

Wie sollte ein solches Lernsystem nun gestaltet sein? Zunächst sollte es den schöpferischen Aspekt der Technik hervorheben. Bei der Technik steht nicht die Analyse vorhandener technischer Lösungen im Vordergrund, sondern die Methoden, mit denen neue technische Lösungen für konkrete Problemstellungen erarbeitet werden können.

Auch darf bei der Entwicklung eines solchen Lernsystems die konkrete schulische Situation nicht außer acht gelassen werden. Der Unterricht findet in Einheiten von 2 bis 4 Unterrichtsstunden pro Woche statt. Eine Ausstattung mit Werkzeugen ist in der Regel nicht vorhanden. Die Lehrer sind nicht technisch geschult. Es sollte daher möglich sein, auch mit wenig Vorwissen in kurzen Unterrichtseinheiten schnell Erfolge erzielen zu können, um die Schüler zu motivieren. Nicht zuletzt ist die angespannte Budgetsituation vieler Schulen zu beachten.

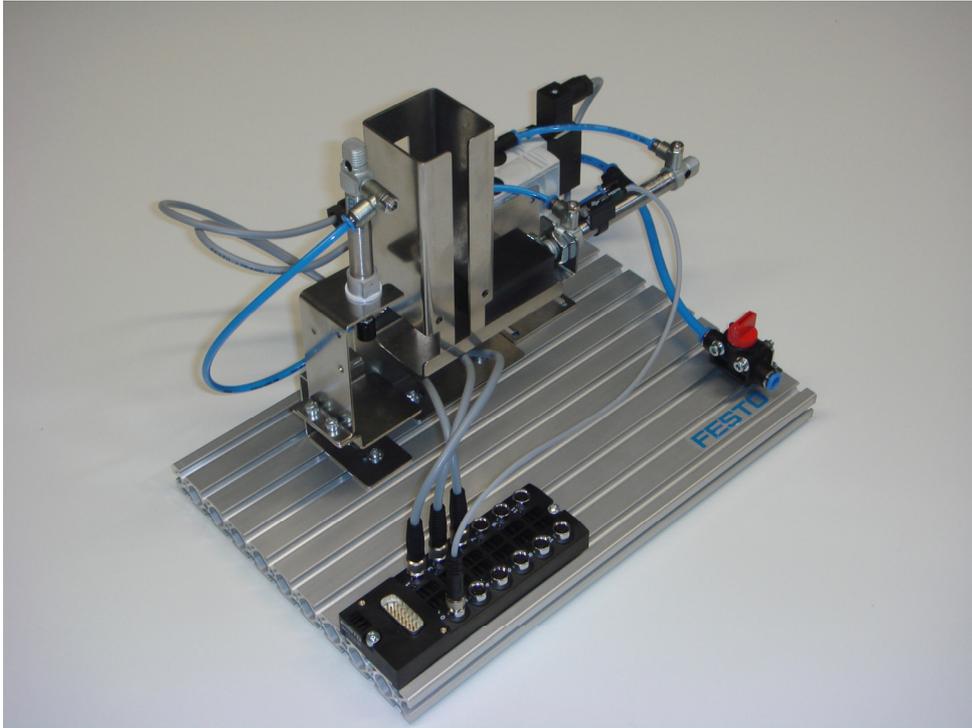


Abbildung 1: Einzelmodul Stapelmagazin

Die Anforderungen an das Lernsystem kann man daher so zusammenfassen:

- Es sollte ein Bezug zur industriellen Wirklichkeit hergestellt werden.
- Das System muß einfach handhabbar und robust sein.
- Das System muß schnell und einfach auf- und abbau sein, damit wenig Unterrichtszeit für die Vorbereitung verloren geht.
- Vorhandene Ressourcen der Schule sollten genutzt werden (z.B. vorhandene PC-Ausstattung).
- Das System darf nur wenig Vorwissen bei den Lehrern (und Schülern) voraussetzen.
- Das Lernsystem muß preiswert sein.
- Das System sollte neben Fachkenntnissen auch die Grundlagen der ingenieurwissenschaftlichen Methodik vermitteln.

### **Aufbau des Lernsystems**

Das vorgeschlagene Lernsystem besteht aus mehreren kompakten Modulen, die jeweils für sich ein abgeschlossenes automatisierungstechnisches System darstellen. Die Module orientieren sich an den in jeder automatisierten Produktion vorhandenen Funktionen wie Transportband, Magazin oder Handhabungsgerät. Realitätsnähe wird durch die Verwendung von industriellen Komponenten sichergestellt (vgl. Abbildung 2).

Durch die kompakten Abmessungen ist eine Nutzung in PC-Räumen leicht möglich. Die einzelnen Module sind so konstruiert, daß sowohl ein Zusammenbau zu einer kleinen „Produktionslinie“ einfach möglich ist (vgl. Abbildung 1) als auch der Umbau der Module. Selbst der Austausch von Einzelteilen zwischen den Modulen (z.B. Sensoren) ist durch einheitliche Schnittstellen leicht möglich. Durch Verwendung von Industriesteckern ist auch das Umverdrahten ohne Werkzeug durchführbar.

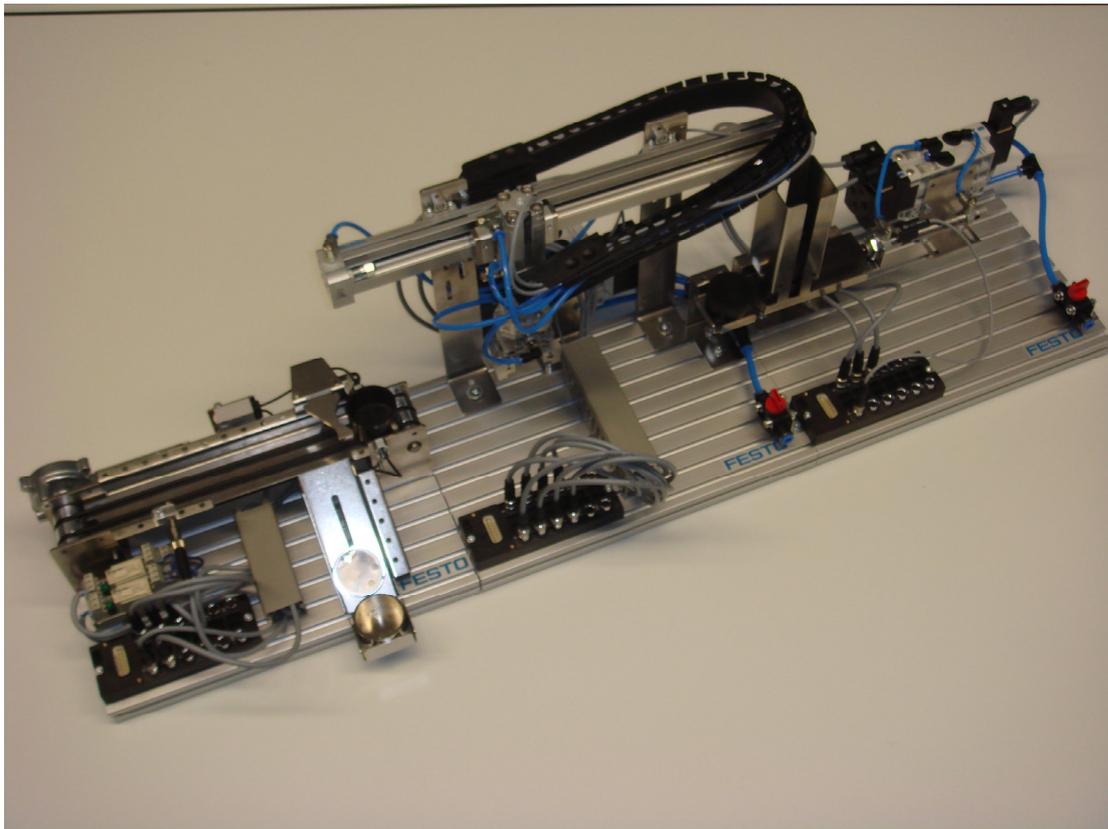


Abbildung 2: Module des Lernsystems kombiniert zu einer Linie

Die Steuerung erfolgt mit einem einfach zu erlernenden Programm, das sowohl die Simulation pneumatischer, elektrischer und logischer Schaltungen erlaubt als auch die Steuerung realer Hardware über ein spezielles Interface (vgl. Abbildung 3). Damit können die Schüler sowohl die Simulation als wichtige Methode kennenlernen, als auch einfache Techniken der Steuerungstechnik (Relaissteuerungen und speicherprogrammierbare Steuerungen).

Nicht zuletzt erlaubt die Software auch die Dokumentation der Schaltungen und Programme in einer normgerechten Art und Weise.

Angeleitet durch Arbeitsblätter erlernen die Schüler Schritt für Schritt die genutzten Technologien und werden auch motiviert, das System zu verändern. Das dazu notwendige Werkzeug liegt bei. Der Lehrer sollte in dieser Phase als „Kunde“ auftreten, der bei den Schülern eine spezielle technische Lösung „bestellt“. Die Schüler

organisieren dann im Team die Entwicklung und die Realisierung der Lösung.



Abbildung 3: Steuerung über PC-Interface und Software FluidSIM

Höhepunkt des Unterrichtes kann dann die Verknüpfung der Module zu einer kleinen Produktionslinie sein (vgl. Abbildung 1). Bei dieser Aufgabe müssen die Schüler im Team Probleme wie die Vernetzung der einzelnen Steuerungen oder die Definition von Übergabepunkten lösen.

Das beschriebene Lernsystem ist in einem Schulversuch unter Beteiligung des Lehrstuhls für Technikdidaktik an allgemeinbildenden Schulen der PH Schwäbisch-Gmünd getestet worden.

### **Erfahrungen aus dem Schulversuch**

Am Schulversuch haben 10 Gymnasien und eine Realschule aus Baden Württemberg , ein Gymnasium aus Sachsen-Anhalt sowie eine Polytechnische Schule in Österreich teilgenommen.

Da das Fach NWT in Baden Württemberg erst 2007 eingeführt wird, wurde das Lernsystem in den meisten Schulen in Form von (freiwilligen) Arbeitsgemeinschaften (Informatik-AG, Sensorik-AG, NWT-AG) eingesetzt. Die Jahrgangsstufen beginnen bei der 8. Klasse und enden in der Klasse 13, wobei der Schwerpunkt in Klasse 10 liegt. Die Anzahl der Schüler pro Kurs lag zwischen 6 und über 20. Bei Gruppen mit bis zu 12 Schülern kamen 3 Module, bei Gruppen darüber bis zu 6 Module zum Einsatz, so daß immer zwei bis drei Schüler mit einem Modul arbeiten konnten.

Die Erfahrungen in den Versuchskursen zeigen, daß die Schüler durch die konsequente Verwendung von Industriekomponenten sehr motiviert sind und mit den teilweise komplexen Aufgabenstellungen gut fertig werden. Die Leistungen der Schüler bei der Arbeit mit dem Lernsystem korrespondiert übrigens nach Erfahrung einiger der am Schulversuch beteiligten Lehrer nicht mit denen im „normalen“ Unterricht.

In den bisherigen Kursen ist eine eher freie Aufgabenstellung gewählt worden, d.h. die Schüler haben sich mit dem System beschäftigt, um dieses „zum Laufen zu bringen“. Dies führt natürlich zu sehr unterschiedlichen Herangehensweisen der Schüler:

- Einige Gruppen haben sich vornehmlich mit der Hardware beschäftigt und diese umgebaut, teilweise durch selbstgebaute Erweiterungen. Zur Steuerung wurden dann nur die mitgelieferten Beispielprogramme genutzt.
- Einige Gruppen ließen die Hardware völlig unverändert und haben sich ausschließlich mit der Programmierung der Module befaßt.

Beide Wege führten bei den Schülern zu Erfolgserlebnisse und zeigen die Flexibilität des gewählten didaktischen Ansatzes.

Wichtig scheint eine manuelle Bedienmöglichkeit zu sein, damit die Schüler die Funktion der Module erkunden können, ohne vorher programmieren zu müssen. Dies ist ein Grund, warum die Schüler die Module mit Pneumatik gern nutzten, weil diese durch die Handhilfsbetätigung der Ventile einfach manuell steuerbar sind.

Weiterhin hat sich gezeigt, daß das Vorwissen der Schüler teilweise so gering ist („was ist ein Relais?“), dass eine Vorschaltung von einführenden Kursmodulen notwendig scheint.

### **Ausblick**

Der gute Erfolg des Lernsystems im Schulversuch zeigt, dass anspruchsvoller Technikunterricht an allgemeinbildenden Schulen möglich ist. Entscheidend wird dabei die Qualifikation der Lehrer sein. Flächendeckende Lehrerfortbildungen müssen organisiert werden. Aber auch die materielle Ausstattung der Schulen muß gewährleistet werden. Technikunterricht setzt das Vorhandensein moderner Technik voraus. Mit der Ausstattung eines Physiklabors allein kann man dem Anspruch, Technikunterricht gestalten zu wollen, nicht gerecht werden.

#### **References:**

- [1] Lehrplan NWT Klasse 10 Baden Württemberg
- [2] Autorengruppe: Naturwissenschaft und Technik (NWT), unter: <http://www.nwt-bw.de>
- [3] Automation wächst auf ganzer Linie, VDI-nachrichten, 20.4.2007
- [4] Hüttner, Andreas: Technik unterrichten, Europa Lehrmittel, S. 37

#### **Authors:**

Dr.- Ing. Reinhard Pittschellis  
Festo Didactic GmbH & CO KG  
Rechbergstr. 3  
73770 Denkendorf  
Phone: +49-711-3467-1415  
E-mail: [pitt@de.festo.com](mailto:pitt@de.festo.com)

A. Diab, H-D. Wuttke, K. Henke, A. Mitschele-Thiel, M. Ruhwedel

## **MAeLE: A Metadata Driven Adaptive eLearning Environment**

### **ABSTRACT**

The topic eLearning is taking on an increasingly important role in the discussion about modern teaching and learning methods. New technologies and in particular the Internet open many new opportunities, however in many cases these have not yet been exhausted. For example the management of large amounts of information and the provision of eLearning show that many demands remain to be met. Designing and Implementing of adaptive eLearning environments is a major point of interest. Currently, many learning environments offer little or no support for adaptivity. This is especially regrettable for personalize the content. In this paper we present a new SCORM compatible Metadata Driven Adaptive eLearning Environment (MAeLE). MAeLE is a framework for personalized adaptive eLearning. The adaptivity depends on the user characteristics and on adaptation metadata, which describe the contents and define how the adaptivity events should take place. The contents themselves do not contain any sequence logic or metadata. MAeLE generates a personalized course with adequate navigation and strategy at run time. A main advantage of MAeLE is its flexibility, extensibility and compatibility to SCORM standard.

### **I- INTRODUCTION**

E-learning was identified as one of the very important areas in the last few years. New technologies and in particular the Internet open many new opportunities, however in many cases these have not yet been exhausted. Using the new technologies can improve the knowledge transfer itself. However, in order to be able to cover the problematic areas that can not be covered by these technologies [1] or the problematic areas of the learner himself like weaknesses and even disabilities of learners, eLearning contents should be personalized and adapted to the learner.

There are two main directions for improving the adaptivity in eLearning [2]. The first one is the learner-centric approach, while the other approach aims to improve the authoring facilities. The first approach describes how the learning contents can be adapted according to the learner's requirements. These requirements can be for example the goals of the user, previous knowledge or a preferred learning style. The data concerning to the preferences of the user and the events resulting from the user's behaviour are stored in the user model. These data can be gathered by explicit user inputs, observing their behavior and by tracking the user's progress [2].

The second approach can be made by providing a more comfortable environment for creating, linking and reusing learning materials. The platforms, currently used in universities, provide a good support for student administration, progress tracking and editing tools for authors. However, the support for individualization and adaptivity is still to be developed [2].

The paper is organized as follows: In section (II) the well known adaptive eLearning environments are discussed. MAeLE will be presented in details in section (III). After that, we conclude with the main results and the future work in section (IV).

## **II- RELATED WORK**

The Sharable Content Object Reference Model (SCORM) [3], is a well known model for eLearning environments. It has get a great acceptance in different fields of eLearning. However, the adaptation abilities of SCORM are very restricted and focused only on allowing the defining of several organizations for the same course and defining of sequencing information, which allows determining a set of rules to select the next activity to be shown. [4] tries to enhance SCORM with adaptivity. Adaptivity on two levels is proposed. The first is adaptivity at the activity level, while the second is adaptivity at the Sharable Content Object (SCO) level. Adaptivity at the activity level depends mainly on a set of sub-activities related to each other according to pre-defined adaptation rules. Adaptivity at the SCO level depends on self-adaptive SCOs. It is mentioned that the SCOs should be able to show different behaviors according to the user's characteristics. Trnkova et. al. [2] present and implement an e-learning environment called Adaptive Learning Environment (ALE). It integrates an intelligent tutoring system, a computer instruction management system and a set of cooperative tools. This generic environment can produce individualized courseware for students, based on their current state of knowledge, their preferences and learning styles according to the chosen learning

strategy. The authors can create contents using pre-defined templates. These templates combine several content elements with different pedagogical functions, e.g. introduction, definition...etc. Metadata are added to each content element too. The user model records the “*interaction history*”, “*tested knowledge*”, and the “*user's readiness*”, which results from comparing the learner object pre-requests with the “*interaction history*” of the user.

Leidig [5] presents a learning environment called L<sup>3</sup> (life-long learning). L<sup>3</sup> is a learning platform combining the functionality of a traditional Learning Management System (LMS) with the power of a Content Management System (CMS). L<sup>3</sup> defines four types of containers of the course materials, namely *learning networks*, *learning objects*, *instructional elements* and *tests*. Instructional elements represent the actual learning contents. Many instructional elements are combined in a learning object. Learning paths are defined in each course depending on the relations between the elements and their content type. L<sup>3</sup> defines two categories of strategies, macro and micro strategy. Macro strategies are responsible for defining the order of the higher-level elements, while micro strategies cover the order of the instructional elements.

Conlan [6] proposes a multi-model metadata driven approach for adaptivity. This approach enhances the eLearning contents with metadata to support adaptivity. This approach separates the content and the sequencing logic into two models namely content and narrative model. The content model contains metadata description of a piece of learning content with a reference to this learning content, while the narrative model stands for the adaptive logic. The adaptivity engine uses these metadata and produces depending on the user model a personalized course at run-time.

### **III- MAeLE: A METADATA DRIVEN ADAPTIVE ELEARNING ENVIRONMENT**

#### **a) Objectives and requirements of MAeLE**

Similar to [6], in order to overcome the shortcomings of the other adaptive eLearning environments, the developed environment should deliver adequate eLearning experience to the learner. Of course, the pedagogical aspects should be taken into account. The developed environment should provide different adaptive effects based on different sets of models. Additionally, it should be extensible and accept new adaptive models and rules. The contents, built in this environment, should be reusable in other eLearning systems. Presentation and navigation adaptivity should be kept separate. This

helps to ensure the reusability. Adaptivity metadata should be added to the contents, navigation methods and strategies. New navigation methods and strategies should be added simply to the environment. A main requirement of the system is that it should react dynamically to any changes in the user's behavior and change the whole or a part of the course. The behavior of the adaptive eLearning environment should be easy configurable through adequate parameters. An important requirement is that the developed eLearning environment should be compatible with SCORM model.

## **b) Structure of MAeLE**

MAeLE extends the functionality of a traditional LMS through developing an adaptivity framework. The main structure of MAeLE is plotted in figure 1.

The user model is used to characterize the user, mainly depending on the current state of knowledge, preferences, learning styles, goal of the course...etc. Some user parameters are entered by the user himself. Some others are obtained through tracking the user's behavior during the use of the system. Keeping the user model in a separate block simplifies the extending of it and even enables the use of a new user model.

It highly recommended that the content should be reusable. Therefore, the eLearning contents should not contain any sequence logic or metadata. The eLearning contents are saved in small meaningful learning objects, *referred to assets in SCORM model*. The metadata for each asset are written in a separate database, called "**Metadata**" in figure 1. These metadata describe the content and contains adaptive data, which will be used to build / alter the adaptive course. A reference to the adequate strategy should exist in these metadata, if the asset is only suitable for one or more strategies.

Metadata about the supported eLearning strategies and navigation methods are recorded in separate databases, called "**Dedactical Strategies**" and "**Navigation**" in figure 1. These metadata contain a description of the strategies / navigation methods and a pointer to the actual implementation of strategies / navigation methods. This enables adding new strategies and / or new navigation methods simply to the system without the need to change the code of other parts of MAeLE.

The adaptation engine is the main component of MAeLE, its task is to determine the structure of the course, depending on the user's characters and on the metadata. In other words, it has to define the adequate eLearning contents as well as the adequate strategy and navigation method for a certain user. The adaptation is relayed on some adaptivity rules. In order to make MAeLE flexible and extensible with respect to new

adaptivity rules, these rules are recorded in a separate database, called “Rules” in figure 1. In this way new adaptation rules can be added easily to the system. The adaptation engine should read only the adaptation metadata and the adaptation rules and generate thereafter the structure of the desired course.

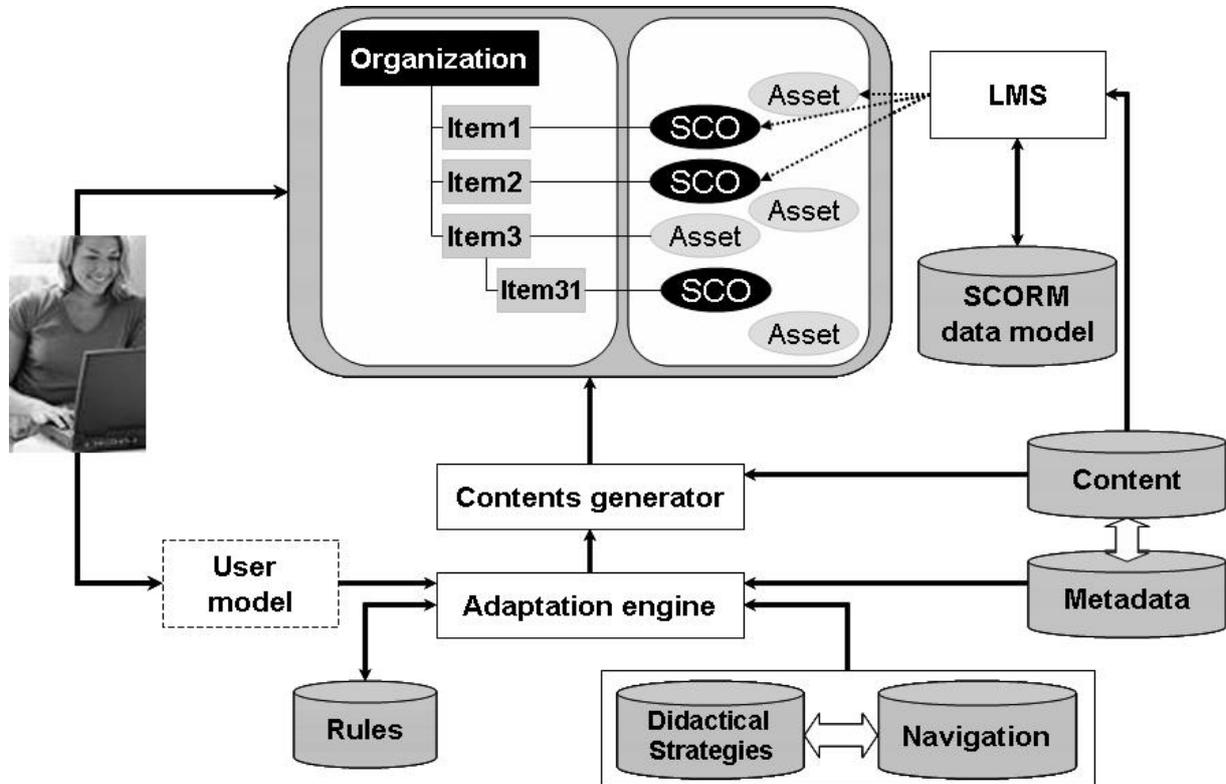


Figure 1: Structure of MAeLE

The output of the adaptation engine is a set of parameters, which defines accurately the desired contents, the adequate strategy and the suitable navigation method. These parameters are fed in the content generator, which in turn generates the required course and build the defined navigation.

At run time, the system records the behavior of the user, e.g., *which links has been visited, which kind of eLearning content is preferred...*etc. The user’s behavior can be tracked using other technologies, e.g., eye tracking [7]. This may change the user’s characters on run-time. MAeLE reacts dynamically on such changes and re-generates the whole or a part of the course. In addition to this, the system may propose changing or altering some rules, depending on the obtained experience. These changes should be performed however by an administrator to ensure that the system stays controllable.

## IV- CONCLUSION AND FUTURE WORK

In this paper we have presented a new metadata driven adaptive eLearning environment (MAeLE). The proposed environment extends the functionality of a traditional LMS through an adaptivity framework. MAeLE generates personalized eLearning courses with an adequate navigation and strategy at run time. The adaptivity in MAeLE depends on the user's characteristics and on adequate metadata. The separation between contents and sequence logic or metadata enables the reusability of the contents, which is a main requirement of an eLearning environment. The proposed framework is flexible and extendable. New rules, strategies, navigation methods and contents can be simply added to the environment. The main advantage of MAeLE is its compatibility with SCORM. This enables it to be integrated in current LMSs.

### References:

- [1] T. Dietinger, " Aspects of E-Learning Environments", Dissertation at Graz University of Technology, Austria, 2003.
- [2] J. Trnkova', G. Roeoßling, M. Muehlhaeuser, " L4 - A Sophisticated Adaptive E-Learning Laboratory ", Workshop Ontologie-basiertes Wissensmanagement (WOW), 2003.
- [3] Advanced Distributed Learning, <http://www.adlnet.gov/scorm/20043ED/Documentation.aspx>, last visit: Juni, 2007
- [4] M. Rey-López et al., " Providing SCORM with adaptivity", WWW 2006
- [5] T. Leidig, " L3 - Towards an Open Learning Environment ", Journal of Educational Resources in Computing (JERIC), 2001.
- [6] O. Conlan, " The Multi-Model, Metadata Driven Approach to Personalised eLearning Services ", PhD thesis, Trinity College, Dublin, 2004.
- [7] V. Garcia Barrios, C. Gütl, A. Preis, K. Andrews, M. Pivec, C. fmoedrit, Trummer " AdELE: A Framework for Adaptive E-Learning through Eye Tracking" , Proceedings of IKNOW 2004, Graz, Austria.

### Authors:

Dipl.-Ing Ali Diab

Dr.-Ing. Heinz- Dietrich Wuttke

Dr.-Ing. Karsten Henke

Prof. Dr.-Ing. habil. Andreas Mitschele-Thiel

Michael Ruhwedel

TU-Ilmenau, Faculty for Informatic and Automation, Gustav-Kirchhoff-Str. 1, P.O.B. 10 0565

98693, Ilmenau

Phone: +49 3677 69 2821

Fax: +49 3677 69 1220

E-mail: {ali.diab|dieter.wuttke| Karsten.Henke|mitsch}@tu-ilmenau.de / ruhwedel@freenet.de

V. Zöppig / O. Radler / M. Beier / N. Neumann / M. Fuchs / T. Ströhla

## **Modular smart systems for motion control teaching**

### **Introduction**

The main professional challenges for engineers in the drive technology are the complexity and heterogeneity of modern mechatronic systems. In order to prepare the future engineers for these tasks project based teaching methods are more and more included in the mechatronics curricula. In order to allow for the solving of high level problems it is crucial to provide a robust, easy-to-learn, and flexible hardware platform for the teaching of motion control and mechatronic system design, e.g. for robotics or mobile and autonomous systems.

The paper presents a modular system which is developed at the Mechatronics department at the TU Ilmenau in co-operation with the STZ Mechatronik Ilmenau.

### **Requirements on the modular teaching platform**

In order to support the main goals of the mechatronics teaching based on the design methodology proposed in the guideline VDI 2206 [1] it is necessary to fulfill the following requirements:

- Since the sensors and actuators of a mechatronic systems are application specific devices the system should concentrate on the signal and information treatment, power supply, and communication blocs.
- These modules must be usable / adaptable for a wide range of different actuators and sensors.
- The hardware modules and software tools should be available at low costs. This supports the creation of the enthusiasm to do “more then expected” during mechatronics projects or contests.
- Obviously, a certain robustness and fault tolerance is required.

The system should be scalable concerning the input/ output complexity as well as concerning the computing power for different application levels.

It is necessary to have a wide commitment to use such a system in different

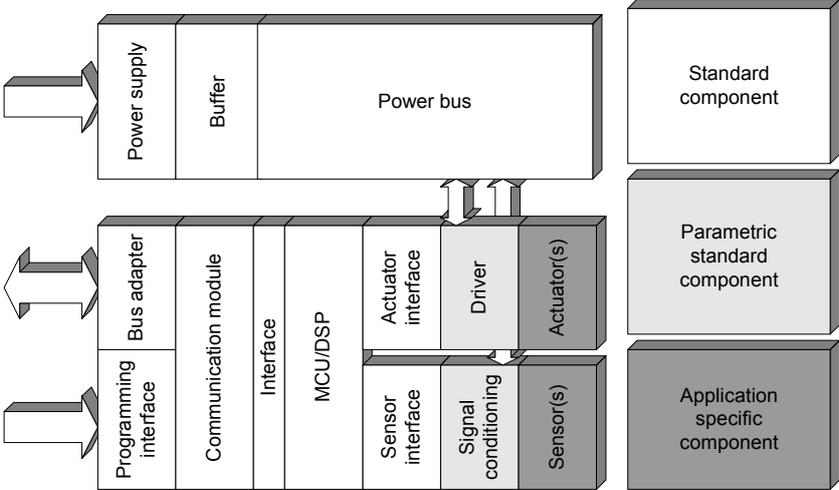
departments and faculties. The platform has to be integrated in the teaching flow. It means that base courses in the early phase of a mechatronics curriculum must provide the knowledge to use the system later.

**Needed elements of a teaching platform**

A typical drive system for distributed architectures consists of the main parts:

- Power supply with EMC circuitry, voltage generation and buffering
- Communication interfaces e.g. field bus or Ethernet connection
- Controller device with peripherals and software
- Interface electronics for sensors e.g. ADC and amplifiers
- Interface electronics for the actuators e.g. power amplifier and safety circuitry

The sensors and actuators of such drive systems are application specific units in most cases. The signal conditioning for sensors and the power amplifiers for the actuators are also application specific or parameterized modules. The other parts of the system can be equal for many different applications (see **Figure 1**).



**Figure 1 Networking drive general structure**

In order to address a wide application field as well as different levels of complexity a modular system is the best choice. The modularity of the system has to be realized in different levels: schematics, functional hardware modules, simulation blocks and software routines. Since the sensors and actuators of a mechatronic systems are application specific devices the system (see **Figure 1**) may concentrate on the signal and information treatment, power supply, and communication blocs.

For all levels adequate software tools must be available. For this system the following tools are proposed:

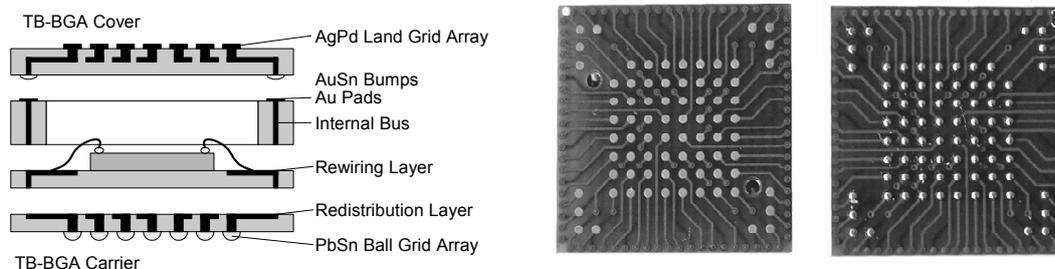
|   |  |
|---|--|
| Electronic design (schematics, layout, CAM processing): | CadSoft EAGLE, Version 4.16 for Windows [6]            |
| Systems simulation, controller design                   | MatLab/SIMULINK [7]                                    |
| Software development                                    | CodeWarrior for 56800/E Digital Signal Controllers [8] |

In addition, project examples and tutorials are provided as PowerPoint presentations.

### The root: Match-X framework

The presented platform bases in parts on the Match-X system. Since the MEMS technology opens a further miniaturization and integration potential for mechatronic drive components which is crucial for the system integration aspects. The European machine tool and plant construction industry is characterized by medium size company structures. Nonetheless there is also enormous cost and innovation pressure in this area. No costly development can be financed due to the small production numbers. One opportunity to solve the cost and production numbers problem while meeting their functional requirement is the modularization of MEMS on the basis of a modular framework.

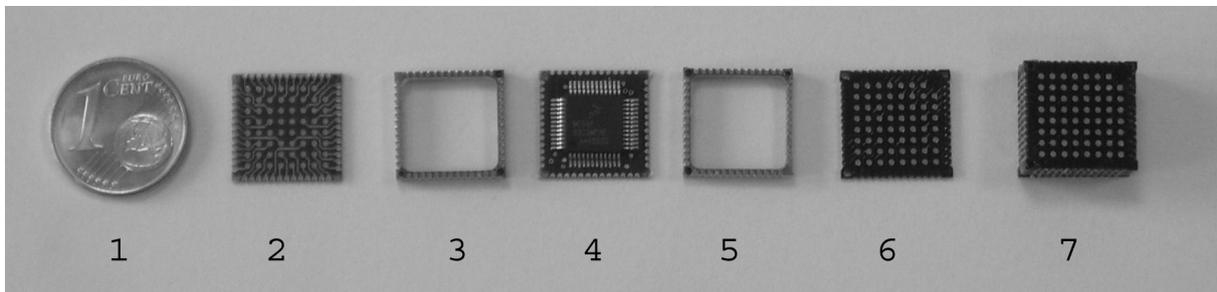
A modular framework called “Match-X” for the implementation of modular microsystems aimed at the requirements of the machine tool and plant construction industry was proposed from the VDMA (German Machinery and Plant Manufacturers Association) in cooperation with German research institutions. Microsystem technology units (modules) are set up with standard packages and standard interfaces (electric, mechanical, fluidic and optical) [3]. These modules are capable of fulfilling specified tasks. Complete modular structured microsystems can be implemented by assembling these modules. The modules available so far were implemented in their majority as FR4- or LTCC TB-BGA (top bottom ball grid arrays, see **Figure 2**).



**Figure 2 TB- BGA Package cross-section [2] and 17.5 mm FR4 module (land grid array left, ball grid array right)**

### Adaptions of the Match-X framework for teaching

Outgoing from this generalized network-ready drive system structure several multi-purpose modules were developed. These are a power supply module (5 V and 3.3 V), a controller module and a switch-mode power amplifier with current feedback. The new modules with a footprint of 12.5 to 12.5 mm<sup>2</sup> are realized in PCB technique with ball/land grid arrays on the top and bottom sides of the modules. In comparison with earlier presented modules with a footprint of 17.5 to 17.5 mm<sup>2</sup> realized in LTCC technique [4] considerable progresses are achieved in the field of volume and cost reduction. **Figure 3** shows the realized controller module with the Digital Signal Controller (MCU with 16 bit fixed point DSP core) MC56F8322 [5].



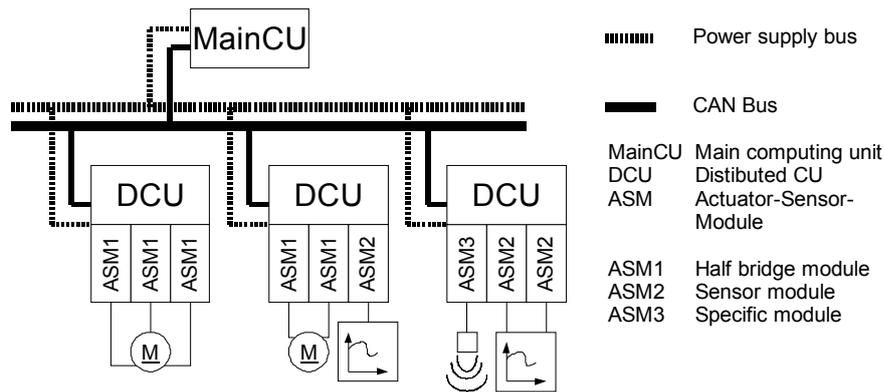
**Figure 3 Controller module (1 ... 7: scale, carrier, frame, controller PCB, frame, cover, and module)**

For these controllers a powerful Integrated Development Environment (CodeWarrior Development Studio for Freescale 56800/E Hybrid Controllers) exists. Thus, an easy-to-learn link between the hardware modules and the software development is provided. This is especially supported by the integrated Processor Expert Tools (see [8] for details). The new modules are compatible to the commercially available didactic kit “efm-Kit” [9].

### **New hardware platform**

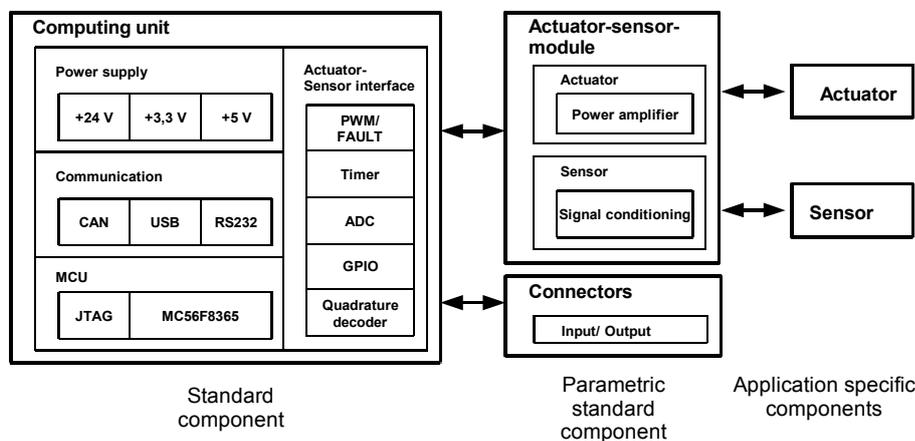
In order to provide a more robust platform for applications with lower miniaturization needs, a new concept is under development. This new systems uses similar schematics and ensures the reuse of software modules in Match-X systems.

This system defines three levels of components: main computing unit, distributed computing unit, and actuator-sensor modules. By combining these units, distributed mechatronic systems can be implemented. Networking is achieved by CAN bus (see **Figure 4**).



**Figure 4 Distributed mechatronic system outline**

All standard components (refer to **Figure 1**) are integrated on one PCB (100 x 100 mm<sup>2</sup> footprint). This board can be equipped with piggy-bag actuator-sensor-modules. The individual actuators or sensors are connected to these modules (Figure 5).

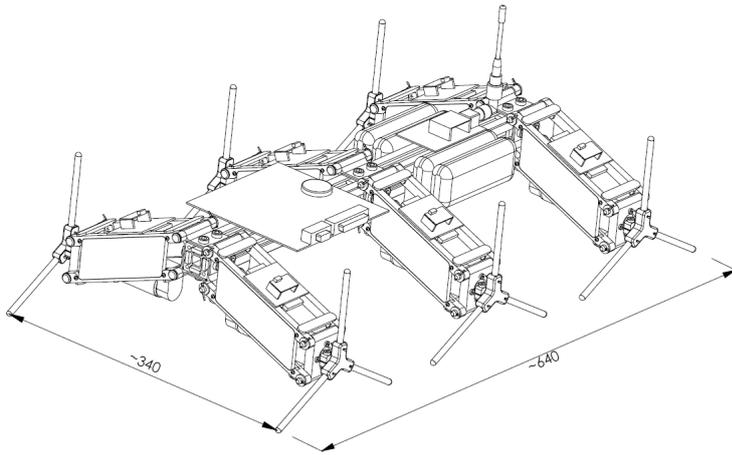


**Figure 5 Distributed computing unit block schema**

In order to allow the networking by CAN bus a special CAN communication protocol was implemented. It allows the networking of up to 8 nodes. Up to 64 global 16-bit variables of the individual nodes can be influenced (read, write) over the bus. An additional master device allows the visualization and the control of the network over an USB interface. This master device also runs communication state machines for the individual nodes. The defined states are NOT\_INITIALIZED, BOOT\_UP\_RECEIVED, COMMUNICATION\_STARTED, COMMUNICATION\_STOPPED, BYE\_RECEIVED, and ERROR\_RECEIVED. A step-by-step instruction exists for the adding of this protocol to a software project.

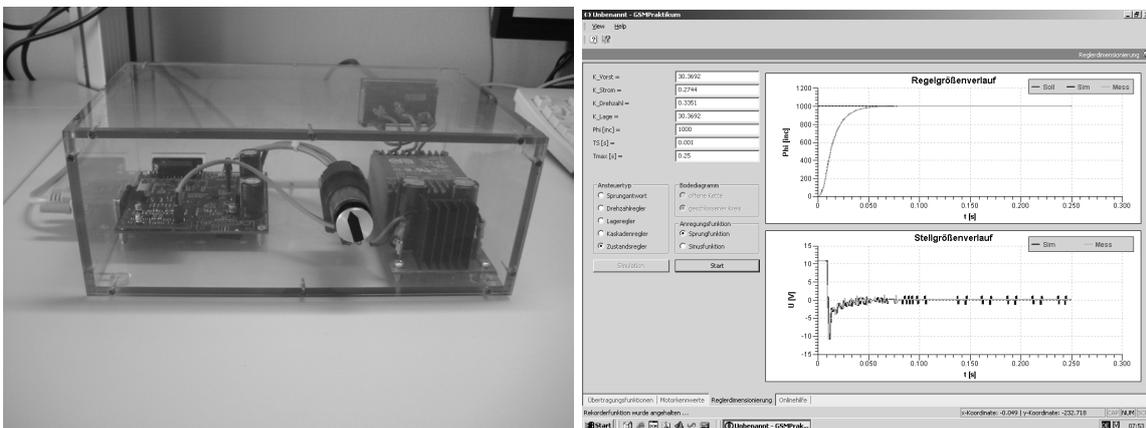
### Platform use

The platform is in use for different teaching purposes. A so called Rotating Leg Vehicle (RoLV) consists of 6 networking DC-motor drives, which control the individual wheel position and velocity (see **Figure 6**).



**Figure 6 CAD of the TU Ilmenau Rotating Leg Vehicle [10]**

Another usage is the DC motor practical training (including state control) at the Mechatronics department of TU Ilmenau (see **Figure 7**).



**Figure 7 DC motor practical training (left: set-up, right: GUI screenshot)**

The platform is subject of further development. This is influenced by the applications.

**References:**

- [1] VDI 2206 Entwicklungsmethodik mechatronischer Systeme. Beuth-Verlag GmbH 2004
- [2] M. Schünemann et al.. MEMS Modular Packaging and Interfaces, in Proc. IEEE 50th Electronic Components Conference (ECTC 2000), Las Vegas, 2000.
- [3] "Bausteine und Schnittstellen der Mikrotechnik", VDMA-Einheitsblatt 66305 (German Standard, in German), Beuth-Verlag, Berlin, 1999.
- [4] V. Zöppig et al. "Rapid development of micromechatronic systems containing magnetic actuators," in Proc. 18th Workshop on high performance magnets & their applications, Annecy (France), 2004.
- [5] 56F8322 Technical Data, Rev. 11.0, Freescale Semiconductor, Inc., 2005.
- [6] CadSoft Computer GmbH. "CadSoft Online: EAGLE Layout Editor," <http://www.cadsoft.de/>, 30.06.2007
- [7] The MathWorks, Inc. "The MathWorks - MATLAB and Simulink for Technical Computing," <http://www.mathworks.com/>, 30.06.2007
- [8] Freescale Semiconductor, Inc. "CodeWarrior for 56800/E Digital Signal Controllers Product Summary Page," [http://www.freescale.com/webapp/sps/site/prod\\_summary.jsp?code=CWS-568-CX](http://www.freescale.com/webapp/sps/site/prod_summary.jsp?code=CWS-568-CX), 30.06.2007
- [9] efm-systems GmbH. "efm-systems Das Mikrosystemhaus: ENTWICKLUNGSWERKZEUGE," <http://www.efm-systems.de/de/produkte/entwicklungswerkzeuge/index.html>, 30.06.2007
- [10] F. Beinersdorf. „Konstruieren eines bionisch inspirierten Laufroboters,“ Diploma thesis, TU Ilmenau:2007

**Authors:**

Dr.-Ing. Veit Zöppig

STZ Mechatronik, W.-v.-Siemens-Str. 12  
 D-98693 Ilmenau  
 Phone: +49 3677 462721  
 Fax: +49 3677 462711  
 E-mail: [veit.zoepfig@stz-mtr.de](mailto:veit.zoepfig@stz-mtr.de)

Jun.-Prof. Dr.-Ing. Tom Ströhma, Dipl.-Ing. Oliver Radler,  
 Dipl.-Ing. Marco Beier, Martin Fuchs, Norman Neumann  
 TU Ilmenau, PF 100565  
 D-98693 Ilmenau  
 +49 3677 692485  
 +49 3677 691801  
[tom.stroehla@tu-ilmenau.de](mailto:tom.stroehla@tu-ilmenau.de)

Nico Pranke, Konrad Froitzheim

## The Media Internet Streaming Toolbox

### Abstract

MIST<sup>1</sup> - the Media Internet Streaming Toolbox - enables the creation of various live streaming applications in a very flexible and simple way. The MIST-server encodes multiple stream formats using different compression algorithms and sends them through a variety of channels with heterogeneous quality of service (QoS) over the internet. The architecture is designed to scale well with the number of clients. MIST provides the infrastructure to create and run complex flowgraphs that process the media data, encode different streams, and send them over the internet. These objectives are achieved by combining Component Encoding Stream Construction (CESC) [1], dynamically loadable functional components (Compresslets) [2], and automated format negotiation.

### 1. Introduction

The design of the Media Internet Streaming Toolbox is based on the following approaches, addressing different objectives:

1. Component Encoding Stream Construction (CESC)
2. Dynamically loadable functional components (Compresslets)
3. Automated format negotiation

The main focus of CESC is to support the creation of multiple video stream formats that use conditional replenishment [3] in a scaleable way. The term *Compresslet* was first introduced in [2] and refers to a functional, dynamically loadable component, which implements a specific part of the compression process. The vertices of a *flowgraph* apply the compresslets to process the media data that flows along their edges. The flowgraph is the core of a MIST-based streaming application. It provides the technical description of media processing and stream construction for different given streams. Flowgraphs are created from logical descriptions, so called *usergraphs*, applying automated format negotiation that uses some definitions and principles from [4].

Due to the flexible and scalable graph-based design the toolbox covers a wide

---

1 <http://ente.informatik.tu-freiberg.de/mist/index.php>

range of potential applications. Examples include: streaming WebCam servers and corresponding clients; video conferencing/chatting scenarios; fast implementation and evaluation of new compression methods and stream formats.

## 2. Component Encoding Stream Construction

The CESC-architecture [5] was designed to encode different stream formats and to handle multiple clients simultaneously. Beyond that, it works efficiently and scales well the number of connected clients by exploiting structural similarities in the encoding process.

In order to support the creation of multiple video stream formats in a scaleable way, only those image parts are encoded and transmitted that have changed since the last frame for a particular client has been sent. Since this frame is, due to the heterogeneity of the internet, usually not the same for different clients, a client-dependent extraction of the changed area must be performed by means of a so called generation-map. It stores for each pixel the sequence number of the frame that modified it and thus keeps track of the history of all changes. It allows to extract the changed area for every client independently.

The generation-map is updated with the clock of the source, while the extraction of the area to encode is done with the clock of the particular client. This temporally decouples the processing of the video and precomputation of components from the actual stream construction. On frame request only the stream construction step has to be done for a particular client using the precomputed components.

## 3. Compresslets

Compresslets implement specific algorithms of video processing and encoding, for instance the discrete cosine or wavelet transforms and their inverses, colorspace conversion or entropy encoding/decoding. They can have several named *input-pins* to receive their data from upstream and *output-pins* to send their data downstream. Compresslets can have different types, depending on the nature of the service they provide. The different compresslet types are listed in table 1 and follow the classification made in [4]. Compresslets provide partially qualified and not necessarily unique media formats (see section 4) they support for their pins, i.e. they specify only the format parameters that are necessary to perform their service. The information about supported formats is mandatory to perform format negotiation.

To facilitate the cost optimal creation of flowgraphs, every compresslet has to provide a cost factor. Since the true costs in terms of run-time are hard to determine, we use

| Type          | Description   | Input-Pins | Output-Pins |
|---------------|---|------------|-------------|
| Source        | produces data (e.g. a WebCam)   | 0          | 1           |
| Sink          | consumes data   | 1          | 0           |
| Converter     | transforms the data from one representation to another but retains the contents | 1          | 1           |
| Filter        | manipulates the contents but retains the representation                         | 1          | 1           |
| Multiplexer   | multiplexes 2 or more data streams to one resulting stream                      | 2..*       | 1           |
| Demultiplexer | demultiplexes an incoming stream into 2 or more resulting streams               | 1          | 2..*        |

Table 1: Compresslet-types.

a qualitative estimation of the true cost.

#### 4. Automated Format Negotiation

We apply a common approach to describe a streaming application by specifying a *media flowgraph* [4, 6], which is a directed, acyclic graph. In our toolbox, it is the technical description of media processing and stream construction for different given streams. The flowgraph is created from so called *usergraphs*, applying an automated format negotiation. Format negotiation and format definition are based on the work done in [4], although we introduced important extensions and modifications to satisfy the constraints of live media streaming applications and CESC.

Usergraphs are directed, acyclic graphs that describe the logic of parts of the streaming application. The elements of the usergraphs are vertices with type source,

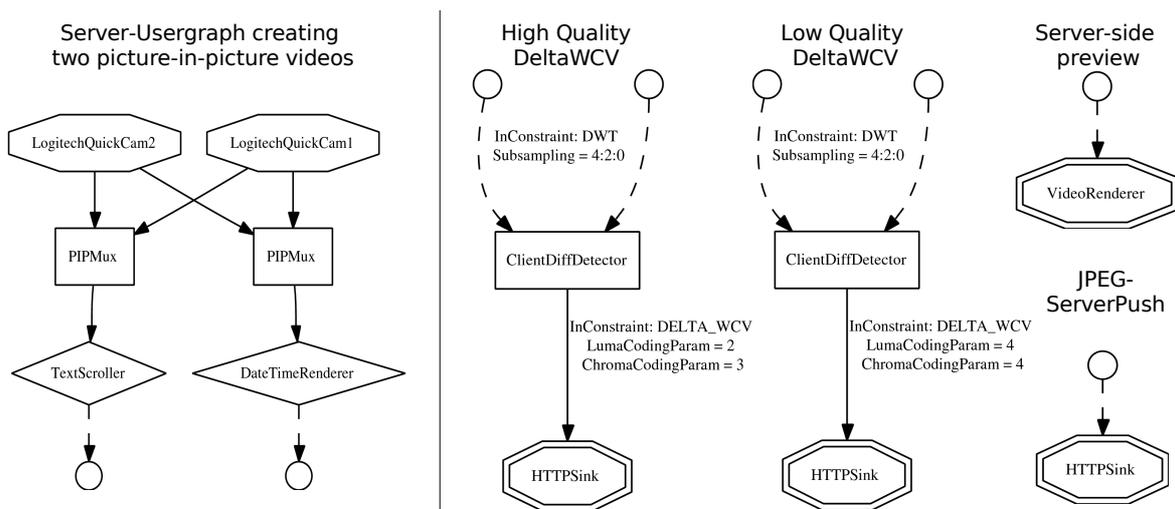


Figure 1: The left part shows a server-usergraph that creates two picture-in-picture videos by combining two sources in butterfly-manner and adding a scrolling text and date/time. The right part shows client-usergraphs that describe the stream construction for particular streams. The desired format and QoS requirements can be specified as a constraint.

sink, filter, multiplexer or demultiplexer. The edge of a usergraph describes a logical connection between an output-pin of its start- and an input-pin of its end-vertex. Their output- and input-formats do not have to be compatible. To technically connect them, format negotiation has to add converter-vertices. There are two possible types of usergraphs. *Server-usergraphs* express how the media preprocessing step has to be done. They contain at least one source- and no sink-vertex. An example of a server-usergraph is shown in figure 1. *Client-usergraphs* describe the client-specific processing of the data (usually the stream-construction). They contain no source- and at least one sink-vertex. Some examples are given in the right part of figure 1. To create a complete, valid flowgraph one server-usergraph and at least one client-usergraph are combined. In doing so, unconnected input-pins of the client-usergraphs are connected with the unconnected output-pins of the server-usergraph.

To perform automated format negotiation it is necessary to use a sophisticated format definition. According to [4], formats are defined by their *classification* and *specification*. The classification divides all possible formats into main categories and consists of *type* and *subtype*<sup>2</sup>. The specification adds parameters<sup>3</sup> that contain a set of values. A format is *unique*, if all its parameters have only one value. The unique formats created from a non-unique format can be obtained by enumerating all combinations of parameters with one value.

If a parameter accepts any value, a wildcard-attribute can be used. Furthermore each value can be attached with an optional quality, which comes into play when more than one format are a valid solution. The quality of a unique format is defined as the mean of the qualities of its parameter values. Contrary to the approach in [4] our primary goal is to find a low cost solution due to the real-time constraints of live-media streaming applications. Only in the case of solutions with equal cost, the quality values are considered to choose the best solution.

To control the negotiation process an additional *constraint format* can be specified for the pins of usergraph-vertices. A format is accepted for the pin only if it matches the given constraint. This enforces the creation of a specific format at the pins of the vertex. To model explicit dependencies between formats of input- and output-pins of a compresslet the concept of an *io-partner* is introduced in [4]: each input format can be associated with an output-format and vice versa. If the format of an input-pin has an io-partner it is accepted only if the io-partner matches the constraint format of its output-pin. We extend this concept to model dependencies

---

2 e.g. type=VIDEO and subtype=DWT, DCT or RGB32

3 e.g. Resolution=320x240, 640x40; Subsampling=4:2:0, 4:2:2, 4:4:4

between different output-pins of demultiplexers (*output-partner*) and different input-pins of multiplexers (*input-partner*).

The underlying mathematical problem of finding a minimum-weight connected subgraph of a subset of vertices of a graph, the so-called *Steiner tree*, is known to be NP-complete and is even hard to approximate [7]. The number of vertices in the graph depends on the number of available converters and the number of values they provide for their parameters and can add up to several thousands. Since we aim to do format negotiation in less than 100 milliseconds, we do not solve the problem exactly, but rather use a heuristic approach. Our heuristic performs the format negotiation for each edge of the usergraphs and for the edges joining client- and server-usergraphs separately. The negotiation process connects adjacent usergraph-vertices by inserting converter-vertices to convert the format of the output-pin of the start-vertex to the format of the input-pin of the end-vertex of the edge. To decide, whether the pins of two vertices can be connected, their formats have to be compared. They *match* if type and subtype are equal and the value sets of all parameters that exist in both formats have a non-empty intersection. Parameters that exist in only one format are added to the resulting *intersection format*.

To actually perform the format negotiation for a usergraph-edge a *negotiation-graph* is defined, which initially contains only the start- and end-vertices of the edge in question. The Dijkstra-algorithm is then applied to find the shortest path between them. In contrast to the negotiation-graph used in [4] our graph is not constructed completely at the beginning, but is rather extended dynamically during the negotiation process. While processing a vertex  $v$  with the Dijkstra-algorithm the negotiation-graph is updated in the following way: For all available converter-compresslets the pairs of unique input- and output-formats are enumerated and a new converter-vertex  $w$  is created for the pair, if the input-format matches the output-format of  $v$ . The vertex  $w$  and a new edge  $e$  connecting  $v$  and  $w$  are added to the negotiation-graph, if the resulting output-format of  $w$  is not already produced by another vertex that has been processed by the algorithm before. The edge  $e$  is annotated with the intersection format resulting from the output format of  $v$  and the input format of  $w$ . Although the complexity of the Dijkstra-algorithm is  $O(n \log n + m)$ , its run-time can be considerable, since the complete negotiation-graph contains a large number of converter-vertices with a high degree that produce a non-disjoint set of formats. The successive extension of the graph leads to paths that create new formats only, keeps the negotiation-graph small and thus greatly reduces the run-

time necessary to create the flowgraph.

Since all output formats of converters on the path originating from the start-vertex of a usergraph-edge represent the same media data just with different formats, these converter-vertices can be used as a potential start-vertex for further usergraph-edges originating from the same vertex. Our heuristic computes the shortest path for a usergraph-edge from each possible start-vertex to the end-vertex of the edge. The start-vertex that results in the shortest path is finally chosen.

The negotiation process is done for each usergraph-edge independently. This leads to the problem of combining all the local solutions to a global one. Two approaches are suggested in [4]. One is to compute all possible solutions for every usergraph-edge and to examine all possible combinations of them to find a valid global solution. This approach is feasible only if the number of solutions per edge is relatively small and the number of valid combinations is small as well. But this can not always be guaranteed. The other suggested approach is to find a “narrowest” edge through which all media data in the graph flows. The parameter values are then propagated up and downstream from this location. But the requirement for the existence of such an edge limits the possible flowgraph layout largely.

Our negotiation process is directed from the source-vertices to the sink-vertices of the usergraphs and is done in breadth first manner. If a solution is found for a usergraph-edge, the resulting format at the input-pin of the end-vertex determines the io-partner format and thus the output-format of the end-vertex. Filter-vertices retain their input-format. So, if the end-vertex is a filter, the io-partner can be manipulated directly by computing the intersection with the input-format to reflect the parameter values, fixed by the input-format. For demultiplexer- and multiplexer-vertices the io-partner format is used as format for the output-pins. The breadth first approach guarantees that all usergraph-edges that influence the output-formats of the start-vertex of an edge have been negotiated before. Parameter values are propagated from the sources to the sinks taking into account the changes they experience by intermediate converters. Thus no unresolvable conflicts can occur and there are no constraints on the layout of the flowgraph.

## **5. Execution of the flowgraph**

One of the most important approaches of CESC is the decoupling of video processing from stream construction. To reflect this, the two-part structure induced by the client-/server-usergraphs is retained for the execution of the flowgraph. The

first part of the flowgraph, corresponding to the server-usergraph is executed with the clock of the sources. The source vertices trigger the execution if they can deliver new data. There are two possibilities for the execution of the flowgraph-parts corresponding to the client-usergraphs:

1. they are executed automatically with the clock of the source
2. they are executed asynchronously for a specific client; triggering has to be done by the sink in the corresponding client-usergraph in this case

The first is appropriate for streams without client dependent states sending all the data to every connected client. The latter is appropriate to support video streams that have client dependent states, e.g. streams that apply conditional replenishment. Then a special multiplexer-compresslet, called *ClientDiffDetector* can be used to determine the changed area for a particular client by means of the generation-map. The information about the area to encode is passed downstream the flowgraph along with the media data. Furthermore manually triggered usergraphs can set a maximum frame rate to support different quality of service requirements. The triggering process allows to group clients according to the sequence number of their preceding frame. The flowgraph part that corresponds to a specific stream and thus the stream construction is executed only once for all clients in the group in this case. It increases scalability of state dependent streams largely. Detailed measurements of scalability are subject for further research.

## 6. Example

As an example we connected the client-usergraphs from figure 1 with each of the two unconnected pins of the server-usergraph from figure 1. The generated flowgraph contains additional converter-vertices to do tasks like scaling, colorspace-conversions, discrete cosine and wavelet transforms and entropy encoding. The resulting streaming server creates for each of the two videos produced by the server-usergraph a JPEG-ServerPush stream and two DeltaWCV [8] streams in different qualities. DeltaWCV is a wavelet based video stream that encodes and transmits only the changed blocks with respect to the last frame sent and is far superior to JPEG-ServerPush regarding rate-distortion. Since the DeltaWCV-streams depend on the state of the client, their construction is decoupled by means of the generation-map contained in the ClientDiffDetector-compresslet as introduced in section 2. For JPEG-ServerPush complete frames are always encoded and thus no decoupling is necessary. Both streams are constructed asynchronously on frame request. Furthermore a synchronous client-usergraph containing a VideoRenderer-

sink is added to offer a server-side preview of the resulting videos. This example is currently used to stream videos of our internet-controllable model railroad<sup>4</sup>.

## 7. Results and Further Research Topics

We combined and modified CESC, the usage of compresslets, and automated format negotiation to flexibly create live streaming applications that are highly scalable with respect to the number of connected clients. We applied our implementation to create video streams with different formats and qualities of service. Flexibility is addressed by the usage of usergraphs that are the building blocks of the flowgraph. Scalability is addressed by the usage of a generation-map, the classification of clients with respect to their state and the asynchronous execution of flowgraph parts that relate to the construction of specific streams.

The toolbox greatly benefits from the existence of a large number of compresslets. An emphasis of our further work is thus the implementation of compresslets not only for video but also for audio to support a wider range of applications. Moreover we plan to introduce dynamic vertex-parameters to control the behaviour of the application at run-time. Further work is required to realize particular applications, e.g. we plan a GUI-application to create, manage, and combine usergraphs interactively.

### References:

- [1] **Kaufman, L.**, *Component Encoding Stream Construction*, Diploma thesis, University of Ulm, 1999.
- [2] **Bönisch, H., Froitzheim K.**, *Server Side 'Compresslets' for Internet Multimedia Streams*, IEEE Multimedia Systems, ICMCS, Vol. 2 1999: 82-86.
- [3] **Mounts, F. W.**, *A video encoding system with conditional picture-element replenishment*, Bell Systems Technical Journal, vol. 48, no. 7, pp. 2545--2554, Sept. 1969.
- [4] **Marco Lohse, Philipp Slusallek, and Patrick Wambach**, *Extended Format Definition and Quality-driven Format Negotiation in Multimedia Systems*, Multimedia 2001 - Proceedings of the EUROGRAPHICS Workshop, Manchester, United Kingdom, September 8th-9th, 2001
- [5] **Bönisch, H., Froitzheim K.**, *Flexible Server Architecture for Optimal Presentation of Internet Multimedia Stream at the Client*, Proc. SPIE Vol. 3969, p. 232-239, Multimedia Computing and Networking 2000, Klara Nahrstedt; Wu-chi Feng; Eds.: Januar 2000.
- [6] **Microsoft DirectShow Architecture**, <http://msdn.microsoft.com>, 2000
- [7] **Chlebik, M. and J.Chlebikova, J.**, *Approximation Hardness of the Steiner Tree Problem on Graphs. Proc. 8th Scandinavian Workshop on Algorithm Theory (SWAT)*. Springer-Verlag, pp. 170-179, 2002.
- [8] **Pranke, N.**, *Ein wavelet-basierter Video-Codec – Algorithmen und Implementierung*, Diploma thesis, Technical University BergakademieFreiberg, 2002

Nico Pranke, Konrad Froitzheim

Technische Universität Bergakademie Freiberg,

Institut für Informatik, Bernhard-von-Cotta-Strasse 2, 09596 Freiberg

Phone: +49 (0) 3731 / 39 – 3016

Fax: +49 (0) 3731 / 39 – 22 98

E-mail: [pranke@informatik.tu-freiberg.de](mailto:pranke@informatik.tu-freiberg.de), [frz@informatik.tu-freiberg.de](mailto:frz@informatik.tu-freiberg.de)

---

<sup>4</sup> <http://rr.informatik.tu-freiberg.de>

A. Fleischer / R. Andreev / Y. Pavlov / V. Terzieva

## **An Approach to Personalized Learning: A Technique of Estimation of Learner's Preferences**

### **INTRODUCTION**

The term “e-learning” presents “technology enhanced learning”. It describes the use of computer-based technologies to support and enhance learning practice. The focus of e-learning engineering is on the design and implementation of e-learning environment. There are two approaches to the development of an e-learning environment that take into consideration two of its aspects. According to the first aspect, e-learning environment is a server that provides learners with learning objects and services. The approach concerning this aspect bases on service-orientated design and implementation models in the development of e-learning environment. Such environment consists of two fundamental types of servers:

- Servers that provide learning content (content providers);
- Servers for provision of learning services - the learning services support various learning styles, which correspond to the existing pedagogical methods.

The second approach considers e-learning environment as environment that supports educators in process of producing of learning resources that are composition of learning objects and services. It ensures the development of *production-oriented (productive) e-learning environments*.

Personalized learning concerns learner-centered adaptation of learning resources. In the web-based e-learning environment the process of personalization of learning objects requires *model of an individual learner* that represents its profile [6]. This model is an element of adaptive e-learning environments that carry out various adaptation methods. Some of them use semantic-based reasoning for achieving automatic adaptation of e-learning environment [1]. Others utilize to the fullest extent reasoning that bases on matching learning material described by metadata with learners' characteristics. A service-based strategy for implementation of an adaptation method of reasoning that uses heuristic rules for deriving recommendations using learner profile information model is presented in [7]. The adaptive servers providing learners with educational

services are personalized, as well. An approach to personalization of adaptive learning services provision is given in [8]. A model of personalization of an adaptive e-learning environment that provides learning objects and services together is presented in [2]. The adaptation of learning resources to individual learners in a productive e-learning environment bases on transformation of generalized learning resources into specific ones. This is an adaptation technique, which bases on *generalization/specification* method of reasoning and determines the e-learning environment as “adaptable”[13].

### **A MODEL-BASED FRAMEWORK FOR PERSONALIZED LEARNING IN PRODUCTION-ORIENTED E-LEARNING ENVIRONMENT**

A model-based framework for supporting of personalized learning in a production-oriented e-learning environment is presented on Figure 1. It integrates various entities that take part in learner-adapted production of this environment. The architecture of this framework consists of the following components: success, learner, product, process and evaluation models. The overriding top-priority success model defines the optimization criterion of the productive e-learning environment and makes some of the other components the primary driver for framework integration [4]. In our case, the success model of productive e-learning environment work is to “Achieve usability of products of e-learning environment”. It requires learner-centered adaptation of learning resources that are products of teacher work and are represented by work product models. According to the success model, the key goal of the productive e-learning environment is to deliver “fit for characteristics and attributes of a learner” learning resources. This requirement to the production-oriented environment qualifies the learner model as primary driver of learning resources production.

The development of personalized learning resources depends on the learner profile, which has multi-layers structure. The bottom layer represents the learner’s preferences, which determine *how* a man learns with pleasure [5]. The upper layer describes learner background that relates to learner’s experience. The top two layers present the attributes of a learner: learner requirements and goals. In the presented framework, the way of transformation of work product models conforms to the model of the teaching process. The top level of the upper loop involves the generalized conceptual model of the initial work product. The lower loop involves elaboration of the conceptual products models until these are reified to the point that can be transformed to a learning resource suitable for a learner.

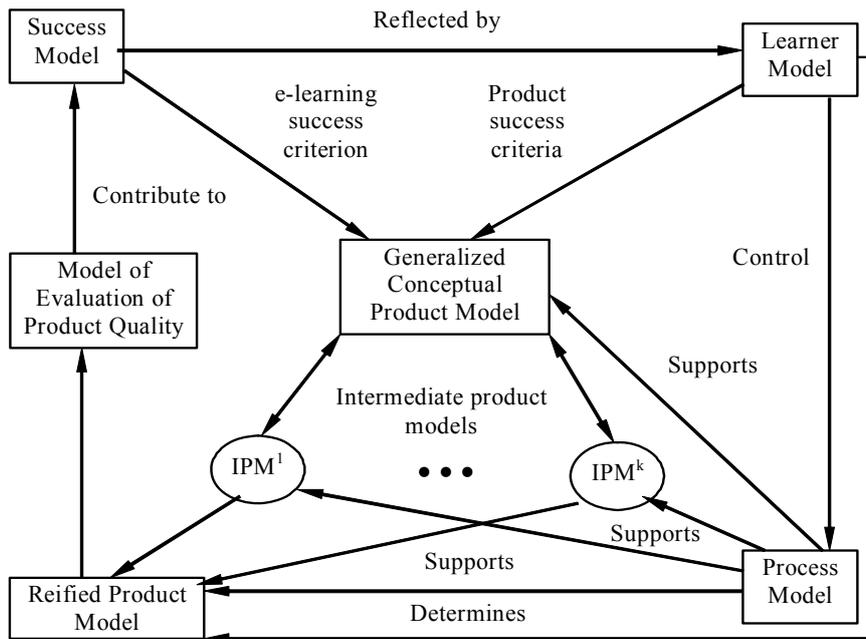


Fig. 1 Framework for personalization of productive e-learning environment

## MATHEMATICAL FORMULATIONS AND METHODS

The preferences of learners are estimated on the basis of a mathematical approach that refers to the utility theory [3]. In this case, the learner is considered as a decision-maker (DM). The explicitly expressed preferences of the DM can serve for evaluation of learning resources (object). The main objective of this evaluation is to keep the correlation between the expressed preferences and the DM's utility function. Standard description of the utility function is presented by formula (1). There are a variety of possible final results ( $x \in \mathbf{X}$ ) that are consequence of a DM activity. A utility function  $U(\cdot)$  assesses each of these results. The DM judgment is measured quantitatively by the

following formula:

$$U(p) = \sum_i P_i U(x_i), \quad p \text{ is probability distribution } \sum_i P_i = 1 \quad (1)$$

We denote with  $P_i$  ( $i = 1 \div n$ ) subjective or objective probabilities, which reflect the uncertainty of the final results. Strong mathematical formulation of the utility function is the next: Let  $\mathbf{X}$  be the set of alternatives and  $\mathbf{P}$  be a subset of the set of probability distributions over  $\mathbf{X}$ . The DM's preferences over  $\mathbf{P}$  are described by the binary "preference" relation ( $\succ$ ) including those over  $\mathbf{X}$  ( $\mathbf{X} \subseteq \mathbf{P}$ ). A utility function is any function  $U(\cdot)$  for which is fulfilled:  $((p \succ q), (p, q) \in \mathbf{P}^2) \Leftrightarrow (\int U(\cdot) dp > \int U(\cdot) dq)$ . Thus, the mathematical expectation of the utility  $U(\cdot)$  is a quantitative measure concerning the DM preferences about the probability distributions over  $\mathbf{X}$ . In practice the set  $\mathbf{P}$  is a set of finite probability distribution. We suppose that the singleton distributions belong to  $\mathbf{P}$ , ( $\mathbf{X} \subseteq \mathbf{P}$ ). There are

quite different evaluation methods of the utility functions [9, 12] that based prevailing on the "lottery" approach. A "lottery" is every discrete probability distribution over  $\mathbf{X}$ . We mark the lottery "x with probability  $\alpha$  and y with probability  $(1-\alpha)$ " as  $\langle x, y, \alpha \rangle$ . There are different systems of axioms (like *Von Neumann and Morgenstern's axioms*) that give satisfaction conditions of utility existence.

We start with the assumption that any convex combination of elements of  $\mathbf{P}$  belongs to  $\mathbf{P}$ :  $(q, p) \in \mathbf{P}^2 \Rightarrow (\alpha q + (1-\alpha)p) \in \mathbf{P}$ , for  $\forall \alpha \in [0, 1]$  [10, 11]. This condition and  $(\mathbf{X} \subseteq \mathbf{P})$  determine the utility function over  $\mathbf{X}$  (when this function exists) with the accuracy of an affine transformation. The most used utility assessment approach is comparisons of the kind:  $(z \sim \langle x, y, \alpha \rangle)$ , where  $(x \succ z \succ y)$ ,  $\alpha \in [0, 1]$ ,  $(x, y, z) \in \mathbf{X}^3$ . Every comparison of this kind defines a "learning point"  $t = (x, y, z, \alpha)$ . With probability  $D_1(x, y, z, \alpha)$  the DM assigns the "learning point" to the set  $A_u$  or with  $D_2(x, y, z, \alpha)$  to  $B_u$ :

$$A_u = \{(x, y, z, \alpha) / (\alpha U(x) + (1-\alpha)U(y)) > U(z)\}, B_u = \{(x, y, z, \alpha) / (\alpha U(x) + (1-\alpha)U(y)) \leq U(z)\}.$$

The DM answers ( $\succ \Leftrightarrow 1$ ;  $\prec \Leftrightarrow -1$ ;  $\sim \Leftrightarrow 0$ ) are with probability and subjective uncertainty.

The main recurrent stochastic procedure in the proposed approach has the form:

$$c_i^{n+1} = c_i^n + \gamma_n \left[ D'(t^{n+1}) - \overline{(c^n, \Psi(t^{n+1}))} + \xi^{n+1} \right] \Psi_i(t^{n+1}), \sum_n \gamma_n = +\infty, \sum_n \gamma_n^2 < +\infty, \forall \gamma_n \geq 0. \quad (2)$$

Here  $(c^n, \Psi(t))$  denotes scalar product and  $D' + \xi$  are the teacher's answers ( $\succ \Leftrightarrow 1$ ;  $\prec \Leftrightarrow -1$ ;  $\sim \Leftrightarrow 0$ ) were  $\xi$  is noise (uncertainty) in the teacher answers with mathematical expectation equal to zero [10, 11]. The scalar product has the form:

$$(c^n, \Psi(t)) = \alpha(c^n, \Phi(x)) + (1-\alpha)(c^n, \Phi(y)) - (c^n, \Phi(z)) = \alpha g^n(x) + (1-\alpha)g^n(y) - g^n(z) = G^n(x, y, z, \alpha). \quad (3)$$

The coefficients  $c_i^n$  take part in the decomposition of  $g^n(x)$  by a chosen family of functions  $(\Phi_i(x))$ :  $g^n(x) = \sum_{i=1}^N c_i^n \Phi_i(x)$ . The line above  $\bar{T} = \overline{(c^n, \Psi(t))}$  means that  $\bar{T} = 1$ , if

$T > 1$ ,  $\bar{T} = -1$  if  $T < (-1)$  and  $\bar{T} = T$  if  $-1 < T < 1$ . It is known that under the procedure (2)

conditions specified above the next integral converges to the (min):

$$J_D(G^n(x, y, z, \alpha)) = M \left( \int_{D'(t)}^{G^n(t)} (\bar{y} - D'(t)) dy \right) = \int \left( \int_{D'(t)}^{G^n(t)} (\bar{y} - D'(t)) dy \right) dF \xrightarrow{\text{p.p.}} \inf_{s(t)} \int \left( \int_{D'(t)}^{s(t)} (\bar{y} - D'(t)) dy \right) dF. \quad (4)$$

Here p.p. denotes "almost sure" and  $s(t)$  denotes  $s(t) = \alpha s(x) + (1-\alpha)s(y) - s(z)$ . After some calculations the following make the convergence clear:

$$\inf_{s(t)} \int \int_{D'(t)}^{s(t)} (\bar{y} - D'(t)) dy dF \geq \lim_n \left( \frac{1}{2} \int \overline{(G^n(t) - D'(t))^2} dF \right) \geq 0 \quad (5)$$

Taking in to account the convergence and the structure of the function  $G^n(x, y, z, \alpha)$  (3) it

is assumed that  $g^n(x)$  is approximation of the empirical utility if ( $n$ ) is sufficiently great.

### A USAGE OF THE TECHNIQUE OF LEARNER'S PREFERENCES ESTIMATION

There are shown examples of usage of this technique in the estimation of learner's preferences with regard to the content attributes of learning resources. Its mathematical description is implemented by means of a decision support system developed in the environment of Visual Studio (Visual Basic 6.0). The final calculations and graphics are performed in MATLAB environment.

We apply the developed evaluation tool for representation of learner's preferences to content exposition of learning resources through a function. It concerns the choice of proportion of theoretical presentation to the example-based presentation of knowledge. Figure 2 presents the preference of a learner for content exposition – % theoretical presentation of the whole content presentation.

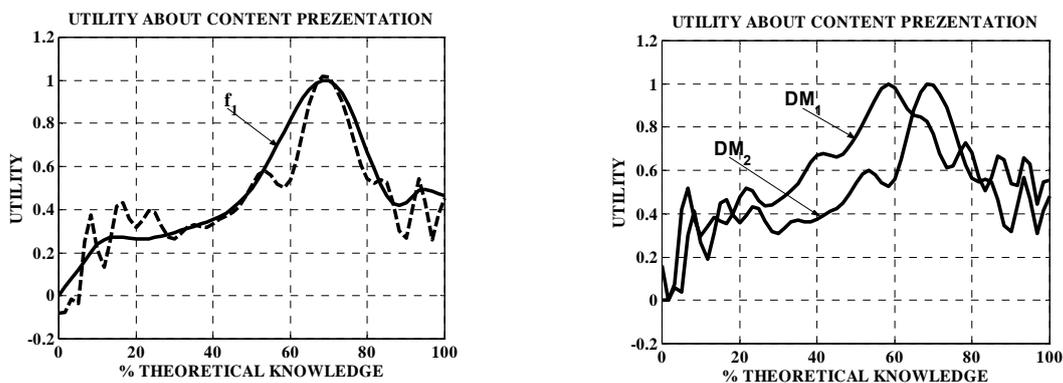


Fig. 2. Learner's preferences to form of content presentation

The seesaw line of the left graphic in the figure recognizes correctly more than 95% of the learner's answers. The stochastic uncertainty makes to see the utility function as a seesaw line. The first evaluation is based on 64 questions (learning points) – utility function  $f_1(x)$ . It is sufficient only for the first raw approximation. This assessment is fast, since the examination takes about 20 minutes. The utility function  $f_1(x)$  clearly reveals the tendency of learner's preference. The right graphic shows estimation of the preferences of two learners ( $DM_1$  and  $DM_2$ ). It is noticeable that learners have different preferences for the percentage of theoretical knowledge presentation given in learning resources ( $DM_1$  chooses lower level of theoretical presentation than  $DM_2$ ). Consequently, learners will need altered learning resources, adapted to their preference. Presented graphics are with stochastic uncertainty.

## CONCLUSIONS

The presented framework for learner-centered *adaptable e-learning environment* has the followings characteristics: the personalized learning is ensured by a learner model-driven, product-focused process of producing learning resources; learner reference model has layered structure; the learner model controls the production of learning resources; learner-centered adaptation is teaching process capability, this framework supports the optimization of teaching process.

### References:

- [1] Aroyo, L., Dicheva, D.: The New Challenges for E-learning: The Educational Semantic Web. Educational Technology and Society Vol. 7 (4) (2004) 59-69.
- [2] Aroyo, L., Dolog, P., Houben, G-J., Kravcik, M., Naeve, A., Nilsson, M., Wild, Fr.: Interoperability in Personalized Adaptive Learning. Educational Technology & Society, Vol. 9 (2) (2006) 4-18.
- [3] Barbera S., P. Hammond, C., Seidl. Handbook of Utility Theory, Volume 1 Principles, Kluwer, Boston, 2001.
- [4] Boehm, B., Port, D.: Conceptual Modeling Challenges for Model-Based Architecture and Software Engineering (MBASE). In: Chen, P.P., Akoka, J., Kangassalo, H., Thalheim, B. (eds): Conceptual Modeling. Lecture Notes in Computer Science, Vol. 1565. Springer-Verlag, Berlin Heidelberg New York (1999) 25-33.
- [5] Brusilovsky, P.: Adaptive Hypermedia. User Modeling and User Adapted Interaction Vol. 11 (1-2) (2001) 87-110.
- [6] Denaux, R., Dimitrova, V., Aroyo, L.: Interactive Ontology-Based User Modeling for Personalized Learning Content Management. In: Proceedings of SW-EL Workshop Applications of Semantic Web Technologies for E-learning, August 23-24 (2004) Eindhoven,
- [7] Dolog, P., Henze, N., Nejdil, W., Sintek, M.: Personalization in Distributed e-Learning Environments. In: Proceeding of the Thirteen International WWW Conference, New York, USA, (2004).
- [8] Kay, J.: Learner Control. User Modeling and User Adapted Interaction Vol. 11 (1-2) (2001) 111-127.
- [9] Keeney, R.L., Raiffa, H.: Decisions with multiple objectives: preferences and value tradeoffs. Cambridge University Press (1993)
- [10] Pavlov, Y.: Subjective Preferences, Values and Decisions: Stochastic Approximation Approach. Comptes rendus de L'Academie bulgare des Sciences Vol. 58 (4) (2005) 367-372.
- [11] Pavlov, Y., Tzonkov, St.: An algorithm for constructing of utility functions. Comptes Rendus de l'Academie Bulgares des Sciences Vol. 52 (1-2) (1999) 21-24.
- [12] Raiffa, H.: Decision Analysis. Addison-Wesley Reading Mass. (1968).
- [13] Stephanidis, C.: Adaptive Techniques for Universal Access. User Modeling and User Adapted Interaction Vol. 11 (1-2) (2001)

### Authors:

Dipl.-Inf. Alexander Fleischer  
Technical University Ilmenau, Faculty of Informatics and Automation  
PSF 98694 Ilmenau, Germany  
Phone: + 49 3677 69 1435  
E-mail: [alexander.fleischer@tu-ilmenau.de](mailto:alexander.fleischer@tu-ilmenau.de)

Assoc. Professor Rumen Andreev PhD  
Assoc. Researcher Valentina Terzieva  
Institute of Computer and Communication Systems, BAS  
Acad. G. Bonchev str. Bl. 2  
1113 Sofia, Bulgaria  
Phone: +358 2 872 01 32  
E-mail: {rumen, valia}@isdip.bas.bg

Assoc. Researcher Yuri Pavlov PhD  
Central Laboratory of Biomedical Engineering, Bulgarian Academy of Sciences,  
Acad. G. Bonchev str. Bl. 105  
1113 Sofia, Bulgaria  
Phone: +359 02 979 36 48  
E-mail: [yupavlov@clbme.bas.bg](mailto:yupavlov@clbme.bas.bg)

N.Tsyrelchuk, E. Ruchaevskaia

## **Innovational pedagogical technologies and the Information educational medium in the training of the specialists**

### **Education in Computer Science and Automation**

The model of the base of the informational-educational environment is created on the base of informational-communicative technologies, a system of the mutual action and its influence on the quality of the preparation of specialists and their informational culture are defined. The need for highly - educated specialists causes innovations in educational system, which is one of the main directions of human activity. That is why it's necessary to put modern information technologies /IT/ into practice today.

Today the information-educational environment occupies a very important place in many countries of the world. In important sense is a constant creative renovation, the development and the perfection of the personality during all life. The world civilization got a powerful instrument in the form of informational - communicative technologies /I.C.T./ with the development of I.C.T. the content of the education are changed greatly. The students and the pupils begin to use widely and actively the possibilities, given by I.C.T. It is expressed in the increase of the time for the independent work, it means that the education begins to be accessible to any person, in any place, in any time and on any speciality [1].

New conditions of the information-educational environment inevitably present new demands to the system of the professional education, and first of all to the quality of the education, new demands. Among them we have: the flexibility of the terms and the process of the education; elaboration and inculcation in the educational process of new pedagogical technologies, new educational systems, new educational plans and programmers on complex integral and scientific -capacious professions. All these requirements are realized in our college. Sometimes we have not enough money on it. But we have the main thing, it is the tendency of the development.

Today we have not only the change of the technological base of the functioning of the educational institutions, but also the change of their institutional essence. As a result today we see many innovational organizational forms of the institutions, which

all use the extended spectrum of new pedagogical methods, new economical and organizational- administrative mechanisms of their functioning [2].

The educational environment is based on the principles of the continuousity, publicity, personal orientation, development of managers on their position, pedagogues, pupils, pedagogics of the collaboration and the partnership. From the point of the main criterion of the estimation of the educational environment we consider the positive valuable attitude of the students.

Today we can notice real dignities of the professional education in the informational-educational environment;

- it gives a constant perfection of the professional knowledge, skills and abilities;
- raises the motivation of the professional activity;
- provides the possibility of the individualization of the education;
- lets realize the succession of professional educational programs of different chaines of the education system;
- gives the possibility to combine the work and the studies.

We believe, that such approach to the educational and pedagogical technologies gives us the harmonic development of the personality, develops the abilities of putting the tasks, modeling, optimization and taking the solutions in the conditions of the uncertainty among our young generation, all this is very important during the preparation of future specialists in different fields of science and technique.

The main particularities of the modern society are:

- constant variety of the scientific and technological process,
- requirement in the rapid work of large volumes of information.

It is obviously, that the necessity in the high level of the preparation of future specialists in the institution gives the conditions for the innovations in the educational system - in one of the main directions of man's activity. That's why it is necessary to create modern IT during this period. Till recently the creation of the IT in the educational sphere had any big innovations. The changes in the technology of the education are not only technical necessities, but it is also the answer on the definite social question. Modern students use the newest carriers of the information's very intensively: TV, PC, E-mail, Internet. Computer technologies are means, which allow to make wider a creative potential; knowledge's appear in a new field; a new view of well known facts and visions appear in order to find the necessary material it's necessary:

- to understand the notions clearly,
- to determine the subject domain, when it is impossible to make asking for the information.

In the process of using the Internet the students receive the qualities:

- autonomy in making decision,
- ability of thinking,
- for the collaboration,
- the capability of giving knowledge's to other people,
- mobility and others.

There is a good connection in MSHRC:

- of the educational-computer center of the college with global telecommunication systems of Internet,
- the students have the possibility:
  - to visit the reading hall of the college,
  - to get acquaintance with new magazines, journals, encyclopedias,
  - to see a system of catalogues, the lists of new documents.
- "to visit" the different libraries of the world - this everything influences at the good study of many topics.

However, different sociological studies show us, that the students are interested in electronic carries of the information very much, in comparison with the traditional paper carries - books, magazines, journals.

In our college we have elaborated ourselves:

- applied programs,
- individual laboratory - practical tasks, which allow us to use the modern informational technologies during the studies of some subjects. The uninterrupted study and using of the computer technique in all the courses in chain are examined in the educational plans:
  - Applied computer science - Computer technology and programming - Information technologies.

In order to use better it in the educational subjects there is a special course in our college which is called "Information technologies" as a basic course for all specialities, including the specification. We have elaborated the packet of practical tasks on the example MS IE, where the students study:

- the necessary parameters and the standart Interface MS IE,
- hypermedia and multimedia of the technology,
- work of different Web-sites.

We pay a great attention to our studies to the computer modulation - the most important component of the educational technology. Excel, Power Point are used in the quality of means. All the documents are made by means of text editor Word. By means of the spreadsheet Excel we make a count of a middle point of studies, the making of diagrams, the sorting of the facts etc. The using of the electronic presentation Power Point allows to create a special booklet, to advertising booklet, to prepare a business plan, the presentation about college.

It is necessary to notice that the proposed practical tasks increase:

- interest of the students to study course,
- allow to produce a creative and individual approaching during the execution some tasks.

While comparing with the lecture on any subject the lecture on any subject the course "Information technologies" :

- more democratic,
- allow to embrace a great quantity of students.

All the teachers known that:

- the students are not always very attentive at the lectures,
- not all of them have time to make notes,
- many of the students cannot hear very well,
- sometimes the student is afraid that he does not understand the material.

The studios on "IT" allow us to make the lectures individual and such studies can regulate the process of getting knowledge's. One can choose an individual rhythm taking into consideration the degree of preparation, the specification of understanding and the needs of each student. We believe, that when we work on an interactive regime every student is able:

- to do the mistakes every time without any shame, when he did not understand some question;
- to work with a rhythm given only for himself;
- to organize a self control;
- to come back to the most difficult questions;
- to lose the situation again, that allows him to find those questions, which he did not understand enough;
- to eliminate the blanks in knowledges;
- to do the some task many times, to ask the some material some times;
- to look through the results of the tests and to analyze them.

In the process of the study of the "IT" course the students make the conclusion about the necessity of having good and profound knowledge's when understanding and using the new types of IT in different fields of the professional activity.

At the 4<sup>th</sup> course our students use the computer technique at the level of the professional user, therefore the introducing the new IT in the educational process helps our students to become more qualified users of new technologies.

When using the computer technologies appears:

- problems of the communicative competition,
- real danger of the transferring of the technical using with the computer towards the relations with the people appears.

We shall notice, that the emotional moments of the acting disappear during the showing the material of studies:

- mimicry, gestures, intonation.

Therefore it is impossible to expel the teacher from the educational process.

The application of the modern IT during the preparation of the students will allow them to adopt themselves easily to the changeable demands of the technical progress.

Computer technologies and the Internet first of all are not the technical means of the educational, but it is a new technology. Therefore in modern conditions the preparation of the professional engineers-teachers, economist, managers etc. cannot be enough successful without using the informational resources and the possibilities of Internet.

Everybody in the world understands, that a new informational culture appears among us. Today the field of the application of modern IT is so great that it is very difficult to find the branch of knowledges where it is impossible to use them with a success of course. A very important place here belongs to the teacher. But it is necessary for this:

- new approaching,
- to refuse from the old schemes of the organization of the educational process,
- to be always ready to learn new fields of knowledge's.

Such approach to the education gives us:

- harmonically development of the personality,
- contribute the possibility to get knowledge's among the students, to develop the putting new tasks, of the modulation, the adoption of the solutions in the conditions of the uncertainty, it is all necessary during the preparation of future specialists [3].

## References:

- [1] Ruchaevskaia E.G. Information technologies in the process of the interaction of the student with the educational environment in Minsk State High Radio-engineering College// The personal-orient education of the professional education: content - forms - problems: Collection of the materials, /Ministry of the education of the republic of Belarus. MSHRC; edited by the professor N.A.Tsyrelchuk.-Minsk, 2005. -244 p.
- [2] Ruchaevskaia E.G. Modern information technologies and the education in the high college// Informatization of the educational process: automatization of the direction, technologies, education from the distance: Collection of the materials, /Ministry of the education of the republic of Belarus. MSHRC; edited by the professor N.A.Tsyrelchuk.-Minsk, 2006. -412 p.
- [3] Pivovarova M.A. Advantages and the lacks of the use of the information technologies. - Minsk, "Secondary professional education", 2006 N2.

## Authors:

Rector, Dr., prof. Nikolai Tsyrelchuk  
Dr. of faculty computer science, Elena Ruchaevskaia  
Minsk State Higher Radioengineering College, 62, Nezavisimosti ave.  
220005, Minsk  
+ 375 17 23172 67  
+ 375 17 231 89 45  
[elena@college.unibel.by](mailto:elena@college.unibel.by), [rector@college.unibel.by](mailto:rector@college.unibel.by)

C. Noack / S. Schwintek / C. Ament

## **Design of a modular mechanical demonstration system for control engineering lectures**

### **1.Introduction**

In lectures concerning Control Engineering methods for controller design and parameter estimation are only shown by simple theoretical examples at the blackboard or in seminars. For Students it is only possible to make practical experience at labs, separated from the lectures.

Experimental Demonstration direct at the lectures would have the advantage, to make the context of the presented methods easier to understand and better to remember, without substitution of the practical experience gained at labs.

The idea was to design a small and light demonstration system, which can be easily transported and installed at the lecture theatre with the ability of modularity to make several different experimental demonstrations directly at the lecture.

It would be possible for the students to take part in the whole controller design process from modelling over controller design up to optimisation, So the student really can see several different practical effects like nonlinearities and noise which doesn't exist in basic theoretical examples and the consequences they have to the linear controller design.

Using methods of automated Code generation and Hardware in the loop testing, it is possible to generate Controllers easily direct in software and let them operate at a practical experiment.

### **2.Mechanical construction**

The Requirements to the mechanical construction were formulated to apply the chosen structure, which can be used for teaching purpose. With change of location special requirements occur. First the structure must be easy to carry and second the sensors must be insensitive to the changing environmental influences. A notebook is used to

visualize the measurements of the sensors and control the system. Due to the limited time before lectures the system must be build up fast. In a list of requirements more details concerning function, structure and economic items are formulated. The most important ones are the length of the distance should be about 300mm and a mass of 250g should be moved with a velocity of 0.3m/s. Due to that, requirements for the actuator element were fixed.

The procedure of construction corresponded to the procedure of Krause [4]. After commitment of requirements in a sketch and list of requirements a sketch of function was developed by dividing the whole systems function in part functions. Next a combination table was made out of technical principles which fulfil the part functions. Combining the part function principles, different solutions for the whole system will be applied. For example the actuator can be an electric, pneumatic, hydraulic or mechanic drive. Every principle has its advantages and disadvantages. The solutions are compared under consideration of the requirements. After this you will have the best solution for your special task. Now the parts can be constructed or chosen.

The actuator is a DC-motor which fits most to the requirements. The linear characteristic curve makes the control simple. The motor needs only an H-bridge to generate turns in both directions. More advantages are the low costs, the electrical connection and the low weight. With the calculated force needed to move the sledge the moment can be calculated. The moment and the revolution allow choosing a fitting motor by searching in datasheets.

To change a rotation in translation a suitable gearing is needed. The best solution of this task is a synchronous belt. A synchronous belt drive provides a linear transmission function, need less space than other gearings and is very quiet. The disadvantage of a synchronous belt drive is the preload force which increases the loadings of the convolution and other elements like the bearings.

The guidance of the sledge is realized by a slide linear guiding. These types of linear guides allow high loadings, are very quiet, cheap and smaller than comparable roller linear guides. With the given mass and velocity of the sledge the slide linear guiding was chosen in a diagram. Slide bearing were also chosen to support the convolutions because of the same advantages that the linear guidance has.

The set up of the sledge serves following tasks. It builds up the connection with the synchronous belt by clamping. Further elements that serve the experimental set up are fixed on the sledge. To change the experimental set up fast it is necessary to construct

a modular platform. The set up for the experiments position control and inverse pendulum consists of two slide bearings that support the convolution. On this convolution the pendulum and the angle sensors are attached.

The position of the sledge is measured by a potentiometer with 10 revolutions. This potentiometer is attached on the driving convolution by a compensating coupling. Potentiometers provide an analogue signal which is useful because the binary inputs of the PC-card are limited. Using an incremental sensor for this task will lead to a complex interpretation circuit which will increase the costs. Other linear measurement systems are often more expensive and need more space. To measure the angles routing plastic potentiometers are used. These potentiometers have besides the general advantages of a potentiometer a smaller work moment, which means that the influence of friction is reduced.

The construction provides a good attempt. The dimensions of the bridge crane fits with the requirement of 500x300x300 mm<sup>3</sup>. The abundance of the dimensions also requested a large amount of parts to be produced because for the bought parts not many components were found. The motor provides the needed power and has reserves for different experimental set ups. Because of the use of slide bearings the system runs quiet but the higher friction can cause higher control complexity. The synchronous belt is the optimal solution for this linear drive and the used preload apparatus is very simple, but not the best solution. The chosen sensors fulfil the requirements on the resolution and they are also cheap. The requirement of low weight isn't really fulfilled. To achieve that the frame elements must be thinner and should be connected by brackets

### **3. Electrical construction and data acquisition**

In order to control the built system a data acquisition device and a current amplifier were necessary. For data acquisition a 6024-E DAQ card from national Instruments was chosen, because it meets the demands of usability at a notebook with a PCMCIA Interface, high resolution with 16 Bit analog input resolution and 12 Bit analog output resolution and also a fast sample time and also the operability with Matlab/Simulink, where the software design was done. The analog output signal was simply amplified by a current amplifier designed and constructed by the electrical workshop of the Institute for Control Engineering at Technische Universität Ilmenau.

#### 4.Linear drive theoretical modelling

For lecture purposes a simple demonstration example was chosen to show the basic principles of controller design on real systems. In order to implement a controller a theoretical modelling was required. A state space design method was used in this case to get a proportional state space controller. In the beginning a theoretical analysis of the DC motor and the Linear Drives after [1] and [2] was made to get models for these Systems. After that these Models were combined to get a model of a DC motor driven linear drive. The resulting model is:

$$\begin{bmatrix} \dot{s} \\ \ddot{s} \\ \dot{i} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -\frac{K_R}{J_{Ges}} & \frac{K_T r}{J_{Ges} n} \\ 0 & -\frac{K_M n}{Lr} & -\frac{R}{L} \end{bmatrix} \begin{bmatrix} s \\ \dot{s} \\ i \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \frac{1}{L} \end{bmatrix} u$$

With  $s$ -position,  $\dot{s}$ -velocity,  $\ddot{s}$ -acceleration,  $i$ -current,  $R$ -Resistance,  $L$ -inductance,  $K_R$ -friction factor,  $K_M$ -revolution constant,  $K_T$ -torque constant,  $n$ -transmission factor,  $r$ -radii,  $J_{Ges}$ -moment of inertia. For this state space model a state controller by pole placement after [3] was designed. The Model with controller was implemented and simulations, presented at Figure 1, were performed.

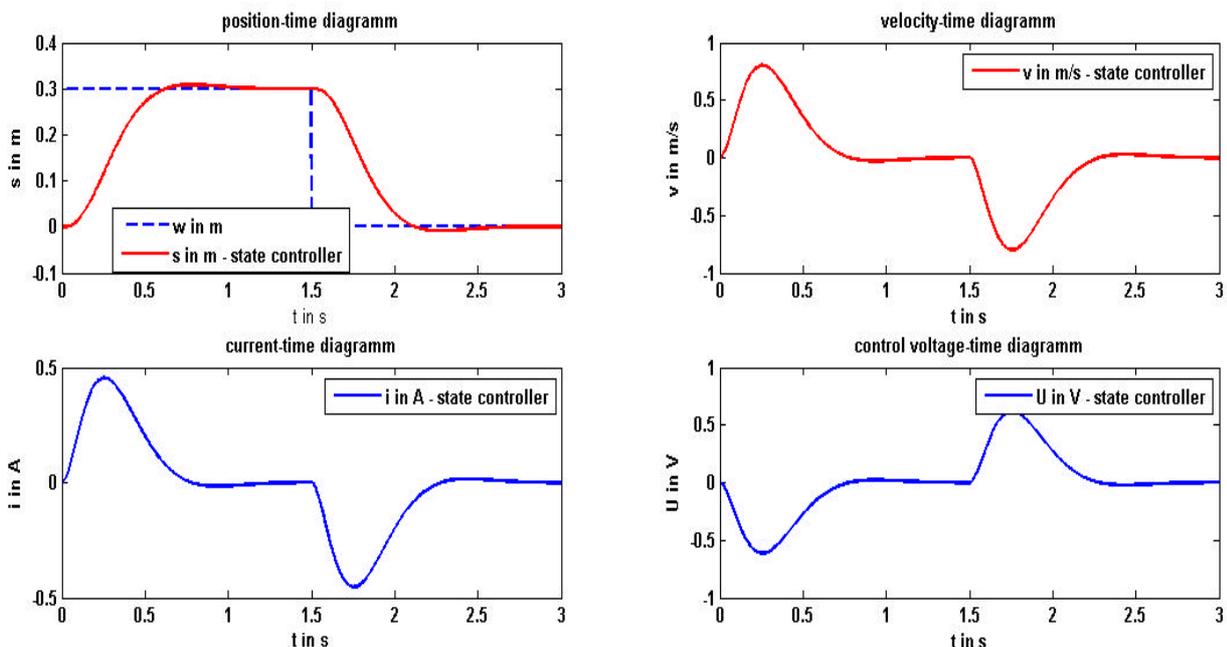


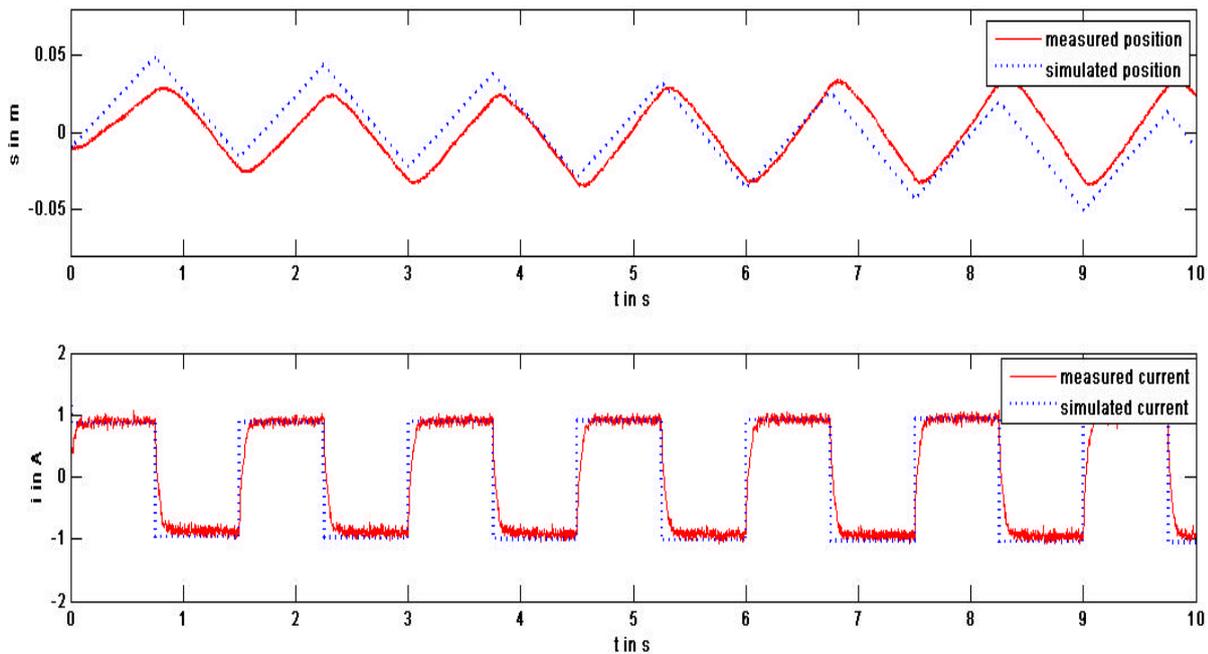
Figure 1: States and controlling voltage over time

## 4. Practical implementation

The Practical implementation was done on another bigger system, because the modular demonstration system was currently at construction. The parameters of this system were measured and estimated by heuristical methods. The system was implemented with these parameters, a controller was designed, implemented and tested.

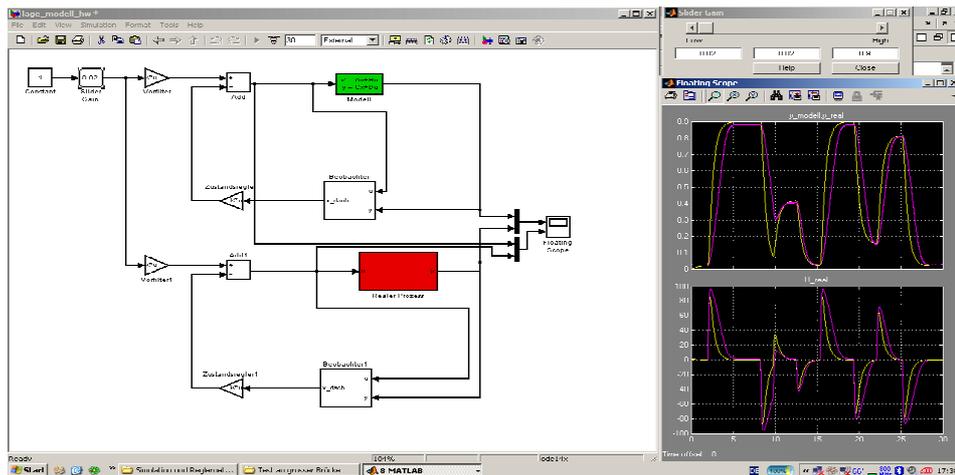
## 5. Practical implementation of the linear drive

The model for the linear drive was implemented and simulated. The results of the Simulation and the measured values for estimation are shown at Figure 2.



**Figure 2: Measured and simulated position and current over time**

The simulated position shown at Figure 2 is slightly different from the measured that can be a reason of nonlinearities in the real system. You can also see a small drift of the measured position, which is a result of different parameters for the different directions of the linear drive. After this simulation a linear state controller was designed, the system was implemented into a Simulink model working with the Real Time Workshop to control the real system via the 6024-E DAQ card and was tested, which is shown at Figure 4.



**Figure 4: graphical user interface to control the demonstration system**

Here you can see the developed graphical user interface to control the demonstration system. In the scope on the right hand side you can see the measured values and the simulated ones there also is a sliding controller to drive the linear drive to a specific position. On the left hand side the simulation model (top) and the interface to the DAQ card (bottom) is shown.

## 6. Future prospects

After all it was shown how to construct and control a small frame modular system for lectures with capabilities of easy useability and installability for the teaching lecturer. It is also possible to create more difficult demonstrations by adding other mechanical components like springs or dampers and also to control them in advanced lectures. In [1] also 3 other demonstrations (loading bridge, inverted pendulum and parameter estimation at the linear drive) are shown.

### References:

- [1]Noack, Christoph: Konzeption und Aufbau eines modularen Didaktik Systems zur Regelung mechatronischer Systeme, Studienarbeit, 2007 TU-Ilmenau, Fachgebiet Systemanalyse
- [2]Luntze, Jan: Vorlesungsunterlagen Rechnergestützte Systemanalyse, 2006, Ruhr Universität Bochum
- [3]Föllinger, Otto: Regelungstechnik, 1994 Hüthig Verlag Heidelberg
- [4]Krause, Werner: Konstruktionselemente der Feinmechanik, 1993 ,Hanser Fachbuchverlag München

### Authors:

Christoph Noack  
 Sebastian Schwintek  
 Univ.-Prof. Dr.-Ing. habil. Christoph Ament  
 Ilmenau Technical University  
 Institute of Automation and Systems Engineering  
 Department of System Analysis  
 98693 Ilmenau  
 Phone: +49 3677 69 2815  
 Fax: +49 3677 69 1434  
 E-mail: christoph-manuel.noack@stud.tu-ilmenau.de





| <b>A</b>            |                    |
|---------------------|--------------------|
| A.-Megeed Salem, M. | 347                |
| Aksit, I.           | 293                |
| Alavi, R. R.        | 333                |
| Alder, J.           | 143                |
| Aley, M.            | 67                 |
| Al-Mahdi, H.        | 397                |
| Ament, Ch.          | 497                |
| Andreev, R.         | 485                |
| Asayenok, I. S.     | 123                |
| <b>B</b>            |                    |
| Batsukh, N.         | 47                 |
| Bauer, S.           | 341                |
| Beier, M.           | 471                |
| Bernard, T.         | 3                  |
| Blaschek, R.        | 315                |
| Bobey, K.           | 263                |
| Bock, R.            | 81                 |
| Böhm, M.            | 303                |
| Böhme, H.-J.        | 353                |
| Boochs, F.          | 203                |
| Böringer, R.        | 379                |
| Brandner, M.        | 209                |
| Braun, M.           | 281                |
| Bräunig, S.         | 447                |
| Brecht, R.          | 409                |
| Brieß, K.           | 333                |
| Brückner, P.        | 239, 309           |
| Bünger, K.          | 293                |
| Burdick, R.         | 263                |
| Burgert, O.         | 97                 |
| Buyankhishing, N.   | 47, 67             |
| <b>C</b>            |                    |
| Chang, N. B.        | 57                 |
| Correns, M.         | 239                |
| <b>D</b>            |                    |
| Däne, B.            | 131                |
| Daum, V.            | 221                |
| Diab, A.            | 427, 433, 439, 465 |
| Döring, T.          | 341                |
| Driesch, D.         | 103, 107           |
| Drobot, S. V.       | 123                |
| <b>E</b>            |                    |
| Earthman, J.        | 167                |
| Eggert, J. P.       | 327                |
| Einhorn, E.         | 353                |
| Ellervee, P.        | 147                |
| Erfurth, F.         | 173                |
| Espiritu, R.        | 167                |
| <b>F</b>            |                    |
| Fassbender, A.      | 293                |
| Faßler, D.          | 173                |
| Fengler, W.         | 137                |
| Fernandes, A. A.    | 113                |
| Ferreira, J. C.     | 113                |

| Fleischer, A.                 | 485                     |
|-------------------------------|-------------------------|
| Franke, J.                    | 287                     |
| Franke, K.-H.                 | 245, 251, 257, 269      |
| Frekers, D.                   | 293                     |
| Froitzheim, K.                | 477                     |
| Fuchs, M.                     | 471                     |
| <b>G</b>                      |                         |
| Gaßmann, F.                   | 185                     |
| Gebhardt, R.                  | 107                     |
| Ghanem, K.                    | 367, 373, 379           |
| Göde, B.                      | 51                      |
| Gonzalez, A.                  | 57                      |
| Götze, C.                     | 293                     |
| Groß, H.-M.                   | 327, 353                |
| Guthke, R.                    | 103, 107                |
| <b>H</b>                      |                         |
| Hamolka, P.                   | 63                      |
| Hamsch, M.                    | 91                      |
| Heinzel, V.                   | 281, 287                |
| Hellbach, S.                  | 327                     |
| Hellrung, L.                  | 85                      |
| Henke, K.                     | 465                     |
| Hofmann, D.                   | 297                     |
| Höhne, G.                     | 227                     |
| Holz Müller-Laue, S.          | 51                      |
| Hopfgarten, S.                | 11                      |
| Hornegger, J.                 | 221                     |
| Huber, R.                     | 103                     |
| Husung, S.                    | 227                     |
| <b>I</b>                      |                         |
| Igney, C. H.                  | 91                      |
| <b>J</b>                      |                         |
| Jablonski, A.                 | 303                     |
| Jacobi, M.                    | 35, 73                  |
| Jahn, R.                      | 185, 257                |
| Jenihhin, M.                  | 147                     |
| Jervan, G.                    | 147                     |
| Jutman, A.                    | 153                     |
| <b>K</b>                      |                         |
| Kalil, M. A. A.               | 367, 373, 379, 391, 397 |
| Kapusi, D.                    | 251, 257                |
| Karimanzira, D.               | 35, 73                  |
| Kärst, H.                     | 379, 385                |
| Kellner, Ch.                  | 427                     |
| Kemenas, J.                   | 293                     |
| Khatanbaatar<br>Altantuul, A. | 11                      |
| Kinne, R. W.                  | 103                     |
| Knauer, U.                    | 359                     |
| Koch, T.                      | 269                     |
| Koczan, D.                    | 103                     |
| Kohrt, K.                     | 303                     |
| Kollhoff, D.                  | 179                     |
| Körner, E.                    | 327                     |
| Kraus, A.                     | 409                     |

|                     |   |
|---------------------|---|
| Krol, O.            | 3   |
| Krömker, H.         | 409                                       |
| <b>L</b>            |   |
| Lahmann, H.-W.      | 191                                       |
| Lambeck, S.         | 453                                       |
| Langner, T.         | 179, 251                                  |
| Lau, B.             | 327                                       |
| Leistritz, St.      | 379                                       |
| Lettenbauer, H.     | 215                                       |
| Li, P.              | 11  |
| Liers, F.           | 379                                       |
| Linke, H.           | 3   |
| Linß, G.            | 159, 297, 309                             |
| Lorenz, T.          | 263                                       |
| Lucht, C.           | 185                                       |
| <b>M</b>            |   |
| Machleidt, T.       | 245, 251, 257                             |
| Masumoto, M.        | 67  |
| Meffert, B.         | 315, 347, 359                             |
| Meier, J.           | 81  |
| Meissner, Ch.       | 143                                       |
| Melitzki, A.        | 197                                       |
| Menz, G.            | 281, 287                                  |
| Meyer, J.           | 167                                       |
| Meysel, F.          | 341                                       |
| Michelson, G.       | 81  |
| Mitschele-Thiel, A. | 367, 373, 385, 397, 427,<br>433, 439, 465 |
| Mitsiukhin, A.      | 321                                       |
| Müller, M.          | 137                                       |
| <b>N</b>            |   |
| Nestler, R.         | 245                                       |
| Neumann, N.         | 471                                       |
| Neumuth, T.         | 97  |
| Niebelschütz, M.    | 245                                       |
| Noack, Ch.          | 497                                       |
| Nyúl, L. G.         | 81  |
| Nyuyki, B.          | 173                                       |
| <b>P</b>            |   |
| Pacholik, A.        | 137                                       |
| Pachynin, V.        | 321                                       |
| Parkoun, V.         | 63  |
| Pavlov, Y.          | 485                                       |
| Petrovskaya, E.     | 321                                       |
| Pfaff, M.           | 103, 107                                  |
| Pfützenreuter, T.   | 23, 29                                    |
| Pittschellis, R.    | 459                                       |
| Pohlens, D.         | 103                                       |
| Pranke, N.          | 477                                       |
| Pretschner, A.      | 97, 143                                   |
| <b>R</b>            |   |
| Raab, Ch.           | 203                                       |
| Radler, O.          | 471                                       |
| Raik, J.            | 147                                       |
| Rauschenbach, T.    | 23, 29                                    |

|                    |          |
|--------------------|----------|
| Reulke, R.         | 341      |
| Riedel, T.         | 275      |
| Rimane, K.         | 51       |
| Ritzmann, M.       | 403      |
| Röll, S.           | 11       |
| Rosenberger, M.    | 159, 309 |
| Ruchaevskaia, E.   | 491      |
| Ruhwedel, M.       | 465      |
| <b>S</b>           |          |
| Saluz, P.          | 173      |
| Santos, A. D.      | 113      |
| Scharaw, B.        | 41       |
| Schaub, M. J.      | 159      |
| Scheibe, A.        | 173      |
| Schellhorn, M.     | 309      |
| Schmidt, St.       | 385      |
| Schmidt, W.-D.     | 173      |
| Schmullius, C.     | 275      |
| Schnee, J.         | 239      |
| Schröter, C.       | 353      |
| Schuster, N.       | 233      |
| Schütze, R.        | 203      |
| Schwintek, S.      | 497      |
| Seidel, H.-U.      | 447      |
| Sergiyenko, T. F.  | 123      |
| Shandiz, H.        | 119      |
| Sickel, K.         | 221      |
| Sokol, V.          | 63       |
| Sparrer, E.        | 245      |
| Stanke, G.         | 197      |
| Stein, R.          | 359      |
| Stöckmann, M.      | 191      |
| Stoll, N.          | 51       |
| Ströhla, T.        | 471      |
| Svirnovskiy, A. I. | 123      |
| <b>T</b>           |          |
| Tam, N. A.         | 421      |
| Terzieva, V.       | 485      |
| Thiel, C.          | 275      |
| Thiesen, H.-J.     | 103      |
| Toepfer, S.        | 103, 107 |
| Töpfer, S. C. N.   | 159, 239 |
| Tong, Z.           | 23       |
| Traiser, J.        | 203      |
| Trost, M.          | 85       |
| Tsyrelchuk, N.     | 491      |
| <b>U</b>           |          |
| Ubar, R.           | 147      |
| <b>V</b>           |          |
| Vauhkonen, M.      | 91       |
| Vrublevsky, I.     | 63       |
| <b>W</b>           |          |
| Waske, B.          | 281      |
| Weber, C.          | 227      |
| Weckend, F.        | 197      |

|                 |                    |
|-----------------|--------------------|
| Weiss, D.       | 215                |
| Westerhoff, Th. | 41                 |
| Wirth, H.       | 203                |
| Woetzel, D.     | 103, 107           |
| Wuttke, H.-D.   | 465                |
| <b>Y</b>        |                    |
| Yousef, A.      | 391                |
| <b>Z</b>        |                    |
| Zacepin, E. N.  | 123                |
| Zahedi, E.      | 119                |
| Zamin-Kahn, N.  | 367, 373, 379, 397 |
| Zellmer, S.     | 107                |
| Zöppig, V.      | 471                |