Dietrich Paulus, Christian Fuchs, Detlev Droege (Eds.)

Proceedings of the

# OGRW2014

**December 1-5, 2014 Koblenz, Germany**

## 9th Open German-Russian Workshop on Pattern Recognition and Image Understanding

AGAS

COMPUTERVISUALISTIK

UNIVERSITÄT KOBLENZ · LANDAU

The 9th Open German-Russian Workshop continues the successful series of international workshops on "Pattern Recognition and Image Understanding" held since 1990 in Berlin, St.-Petersburg, Erlangen, Valday, Herrsching, Katunj, Ettlingen, and Nizhny Novgorod. The general goal of the OGRWs is to establish and to improve the scientific contacts and collaboration between researchers from Germany, the Russian Federation and other countries in order to advance the field of pattern recognition and image understanding. The scientific programs of OGRWs traditionally reflect the advances and state-of-the-art of the field.

OGRW-9-2014 took place from December 1st-5th, 2014, at the University of Koblenz-Landau in Koblenz, Germany.

## Workshop Co-Chairmen

Professor Dr. Dr. h. c.
Heinrich Niemann
Friedrich-Alexander-University of
Erlangen-Nuremberg, Erlangen, Germany

Professor, Full Member of the RAS
Yuri Zhuravlev
Dorodnicyn Computing Centre of the
Russian Academy of Sciences, Moscow, RF

## Workshop Vice Chairmen

Dr.-Eng. Igor Gurevich
Dorodnicyn Computing Centre of the Russian Academy of Sciences, Moscow, RF

Professor Dr.-Ing. Dietrich Paulus
Computational Visualistics, University of Koblenz- Landau, Koblenz, Germany

Professor Dr. Bernd Radig
Technical University of Munich, Germany

## Local Organization

Professor Dr.-Ing. Dietrich Paulus
Christian Fuchs
Computational Visualistics
University of Koblenz- Landau
Koblenz, Germany
agas@uni-koblenz.de

Conference Website:
http://ogrw2014.uni-koblenz.de

## Proceedings

These proceedings contain all contributed, accepted and presented contributions, provided the author(s) granted the appropriate rights. Published online via the University Koblenz-Landau's OPUS document server (http://kola.opus.hbz-nrw.de/).

URN: urn:nbn:de:hbz:kob7-2015051206
URL: http://nbn-resolving.de/urn:nbn:de:hbz:kob7-2015051206

**Editors**

Dietrich Paulus, Christian Fuchs, and Detlev Droege

Contact: agas@uni-koblenz.de

# Contents

A new Approach for Time-Frequency Features Extraction of Electrical Brain Activity: Aplication to Quantitative Diagnostics of Early Stage Parrkinson's Desease ..... 1
*Yury Obukhov, Ivan Kershner, Michael Korolev, Konstantin Obukhov, Olga Sushkova, Alexandra Gabova, Razina Nigmatullina, Galina Kuznetsova, Michael Ugrumov*

A Survey of Deep Learning Methods and Software for Image Classification and Object Detection ..... 5
*Valentina Kustikova, Pavel Druzhkov*

A Technique for Comparing Images of Paintings for Attribution ..... 10
*Dmitry Murashov*

Adaptivity of Conditional Random Field Based Outdoor Point Cloud Classification ..... 14
*Dagmar Lang, Susanne Friedmann, Dietrich Paulus*

Advanced 3-D Pose Estimation for Articulated Vehicles ..... 20
*Christian Fuchs, Frank Neuhaus, Dietrich Paulus*

An Automatic Initialization of Interactive Segmentation Methods Using Shortest Path Basins 23
*Tomáš Ryba, Milos Zelezny*

Analysis of Ionospheric Parameters and Geomagnetic Field Variations by the Wavelet-Transform and Multicomponent Models ..... 29
*O. V. Mandrikova, Yu. A. Polozov, I. S. Solovjev, N. V. Fetisova, M. S. Kupriyanov, A Dmitriev*

Analysis of Near-Field Diffraction Patterns of Gaussian Beams for Surface Defects Detection 34
*Dmitry Savelyev, Svetlana Khonina*

Application of Mixed Models to Solve the Problems of Restoration and Estimation of Image Parameters ..... 38
*K. K. Vasiliev, V. E. Dementiev, N. A Andriyanov*

Automatic Image Analysis Algorithm for Quantitative Assessment of Breast Cancer Estrogen Receptor Status in Immunocytochemistry ..... 43
*Daria A. Dobrolyubova, Tatiana A. Kravtsova, Olga A. Artyukhova, Andrey V. Samorodov*

Blood Pressure Rhythm Estimation Based on Shape Patterns for Analytic Spectra ..... 47
*V. E. Antsiperov, K. G. Mansurov, V. V. Bonch-Bruevich*

Centroid-Based Ensemble Clustering: Algorithms for Hyperspectral Images Segmentation... 50
*Vladimir Berikov, Igor Pestunov, Gillaume Gonzalez, Pavel Melnikov*

Combinatorial Optimization Problems Related to Machine Learning Techniques ..... 54
*Michael Khachay*

Compression Algorithm for Indexed Images With the use of Context-Based Modeling ..... 60
*Alexandr Borusyak, Yuri Vasin*

Conjugacy Indicator for Hyperspectral Image Thematic Classification ..... 63
*Vladimir Fursov, Sergey Bibikov, Oksana Bajda*

Construction of the Hybrid Intelligent System of Express-Diagnostics of Information Security Attackers Based on Synergy of Several Sciences and Scientific Directions..... 67
*A. Yankovskaya, A. Shelupanov, V. Mironova*

Deconvolution and Correction Based Approach to Restore Images Captured Using Simple Diffractive Lenses ..... 71
*Artem Nikonorov, Sergey Bibikov, Maksim Petrov*

Design and Implementation of the Alida Framework to Ease the Development of Image Analysis Algorithms ..... 75
*Stefan Posch, Birgit Möller*

Development of the Logic Programming Approach to the Intelligent Monitoring of Anomalous Human Behaviour ..... 82
*Alexei A. Morozov, Alexander F. Polupanov*

Efficient Multi-Temporal Hyperspectral Signatures Classification Using a Gaussian-Bernoulli RBM Based Approach ..... 86
*Selim Hemissi, Imed Riadh Farah*

Evaluation of Established Line Segment Distance Functions ..... 89
*Stefan Wirtz, Dietrich Paulus*

Experiments With Automatic Segmentation of Liver Parenchym Using Texture Description. 94
*Miroslav Jiřík, Petr Neduchal*

Fast Implementation of the Niblack Binarization Algorithm for Microscope Image Segmentation of Cell Cultures Infected With Chlamydia ..... 97
*Olga A. Artyukhova, Andrey V. Samorodov*

Fast Model Based Object Recognition and Pose Estimation Using Local Deviation Grids .... 101
*Benjamin Hohnhäuser, Stephan Brodkorb, André Moltmann, Frank Püschel*

Filters Based on Aggregation Operators ..... 103
*V. Labunets, E. Osthaimer*

Formation and Recognition of Metabolic Profile of Cancer on the Basis of Chromatography-Mass Spectrogram of Urine Volatile Metabolite Image Analysis ..... 109
*A. A. Rozhentsov, N. N. Mitrakova, K. A. Lychagin, R. R. Furina, V. L Ryzhkov*

Genetic Algorithm Application in Image Segmentation ..... 113
*Pavel Jedlička, Tomáš Ryba*

Hand-Eye Calibration of SCARA Robots ..... 117
*Markus Ulrich, Andreas Heider, Carsten Steger*

Hierarchical Ensemble Clustering Algorithm for Multispectral Image Segmentation ..... 123
*Igor Pestunov, Sergei Rylov, Vladimir Berikov*

Human Body Part Classification in Monocular Soccer Images ..... 128
*Andreas Bigontina, Michael Herrmann, Martin Hoernig, Bernd Radig*

Image Warping as an Image Enhancement Post-Processing Tool ..... 132
*Andrey S. Krylov, Alexandra A. Nasonova, Andrey V. Nasonov*

Image-Based Characterization of the Pulp Flows ..... 136
*Mikhail Sorokin, Nataliya Strokina, Tuomas Eerola, Lasse Lensu, Heikki Kälviäinen*

Joint Analysis of Electroencephalogram, Electromyogram, and Tremor in the Early Stage of Parkinson's Disease ..... 140
*Olga Sushkova, Alexandra Gabova, Alexey Karabanov, Ivan Kershner, Konstantin Obukhov, Yury Obukhov*

Latent-Space Gaussian Process Gaze-Tracking ..... 144
*Nicolai Wojke, Jens Hedrich, Detlev Droege*

Method of Weak Classifiers Fuzzy Boosting ..... 150
*Andrey V. Samorodov*

Modeling Interactions of Objects in Video Sequences..... 156
*Ali Al-Raziqi, Mahesh Venkata Krishna, Joachim Denzler*

New Approach to the Geoacoustic Emission Signals Analysis..... 162
*Alexander B. Tristanov, Yuriy V. Marapulets, Olga O. Lukovenkova, Alina A. Kim*

On Coordination of Contour Descriptions for the Equivalence Class With a Group of Affine Transformations ..... 165
*Leonid Lebedev, Yuri Vasin*

On the False Rejection Ratio of Face Recognition Based on Automatic Detected Feature Points ..... 168
*Kazuo Ohzeki, Masahiro Takatsuka, Masaaki Kajihara, Yutaka Hirakawa, Kiyotsugu Sato*

Optical Signal Processing to Analyze Fluid Absorption Inside the Skin Using Point by Point Photon Counting ..... 172
*Bushra Jalil, Ovidio Salvetti, Marco Righi, Luca Poti, Antonio L'Abbate*

Optimal Facial Areas for Webcam-Based Photoplethysmography ..... 176
*Mikhail Kopeliovich, Mikhail Petrushan*

Optimization of Mutual Information-Based Stochastic Gradient Ascend Algorithm for Image Registration ..... 180
*Sergey Voronov, Alexander Tashlinskii, Iliya Voronov*

Parallel Implementation of Roadmap Construction for Mobile Robots Using RGB-D Cameras 184
*Marco Negrete, Jesús Savage, Jesús Cruz, Jaime Márquez*

PRIAR (Pattern Recognition Image Augmented Resolution) - a Tool to Combine Pattern-Recognition With Super-Resolution ..... 188
*Marco Righi, Mario D'Acunto, Ovidio Salvetti*

Problems of an Image Reducing to a Recognizable Representation ..... 192
*Igor Gurevich, Vera Yashina*

Real-Time Hand Detection Using Continuous Skeletons ..... 200
*Victor Chernyshov, Leonid Mestetskiy*

Real-Time Texture Error Detection on Textured Surfaces With Compressed Sensing ..... 205
*Tobias Böttger, Markus Ulrich*

Robust Dynamic Facial Expressions Recognition Using LBP-TOP Descriptors and Bag-Of-Words Classification Model ..... 211
*Alexey Spizhevoy*

Searching for Rotational Symmetries Based on the Gestalt Algebra Operation Π..... 216
*Eckart Michaelsen*

Semantic Volume Segmentation With Iterative Context Integration ..... 220
*Sven Sickert, Erik Rodner, Joachim Denzler*

Semi-Automatic Liver Segmentation Using TV-L1 Denoising and Region Growing With Constraints ..... 226
*Artem Nikonorov, Pavel Yakimov, Sergey Chaplygin, Alexander Ivaschenko, Yuriy Yuzifovich*

Semiautomatic Quantitative Evaluation of Micro-CT Data ..... 230
*Miroslav Jiřík, Jiri Kunes, Milos Zelezny*

Solving Problems of Clustering and Classification of Cancer Diseases Based on DNA Methylation Data ..... 234
*Alexey Polovinkin, Ilya Krylov, Pavel Druzhkov, Mikhail Ivanchenko, Iosif Meyerov, Alexey Zaikin, Nikolai Zolotykh*

Stereo EKF Pose-Based SLAM for AUVs ..... 238
*Markus Solbach, Francisco Bonin Font, Antoni Burguera, Gabriel Oliver, Dietrich Paulus*

Synthesis of Quasi-Invariant Regulator for Vibration Damping Multi-Storey Buildings Using Methods of Pattern Recognition ..... 245
*Dmitry Balandin, Igor Kotelnikov, Larisa Teklina*

The Algebraic Approaches and Techniques in Image Analysis ..... 249
*Igor Gurevich, Vera Yashina*

The Development and Research the Digital Image Processing Algorithm on Television Picture for Indoor Positioning ..... 265
*Alexander Tyukin, Ilya Lebedev*

The Method for Effective Clustering the Dendrite Crystallogram Images ..... 271
*Rustam Paringer, Alexander Kupriyanov*

The Study of Features of Expert Signature for Left Ventricle on Ultrasound Images ..... 274
*Vasiliy Zyuzin, Sergey Porshnev, Anastasia Bobkova, Vladimir Bobkov*

Traffic Sign Detection and Recognition Using Modified Generalised Hough Transform ..... 277
*Pavel Yakimov*

Using Bit Representation for Generalized Precedents ..... 281
*Alexander Vinogradov, Yuriy Laptin*

Vehicle Video Detection and Tracking Quality Analysis ..... 284
*Valentina Kustikova*

# A NEW APPROACH FOR TIME-FREQUENCY FEATURES EXTRACTION OF ELECTRICAL BRAIN ACTIVITY: APLICATION TO QUANTITATIVE DIAGNOSTICS OF EARLY STAGE PARRKINSON'S DESEASE

Yu. V. Obukhov[1], I. A. Kershner[2], M. S. Korolev[1], K. Yu. Obukhov[2], O. S. Sushkova[1],
A. V. Gabova[3], G. D. Kuznetsova[3], M. V. Ugrumov[4],
[1] Kotel'nikov Institute of Radio Engineering and Electronics of RAS, Mokhovaya 11-7, Moscow, 125009, Russia,
[2] Moscow Institute of Physics fnd Technology, Institutski per., 9, Dolgoprudny, 141700 Moscow Region,
[3] Institute of Higher Nervous Activity and Neurophysiology of RAS, 5A Butlerova St., Moscow 117485, Russia,
[4] Koltzov Institute of Developmental Biology of RAS, ul. Vavilova 26, Moscow, 119334, Russia, obukhov@cplire.ru,
razinar@mail.ru, agabova@gmail.com, mugrumov@mail.ru

***Abstract* — New time-frequency EEG features of early stages Parkinson's were investigated. There are three main differences of EEG wavelet scalograms were considered as early stage PD features. The first is a disordering ridge of PD patient scalogram in frequency range more then ~6 Hz in comparison with that of normal volunteer. The second is a more powerful PD patient cortex electrical activity in frequency range ~4-6 Hz. And the third feature is a scalograms asymmetry of the left and right brain semi spheres.**

***Keywords— wavelet spectra, electroencephalogram, electromyogram, accelerometer, frequency synchronization, Parkinson's disease.***

## I. INTRODUCTION (*HEADING 1*)

Parkinson's disease (PD) belongs to a wide range class of neurodegenerative diseases caused by the death of dopaminergic neurons of the brain. Particular attention was paid to the mechanisms of the brain plasticity serving to compensate functional insufficiency of the degenerating neurons [1]. From this point of view, the authors consider the dynamic of neurodegenerative diseases, stating the necessity of the development of preclinical diagnostics and the preventive therapy [2]. The main problem of diagnostics PD is to find out markers of disease at pre clinical and early clinical stages [3].

Electroencephalography (EEG) and magnetoencephalography (EMG) are the typical investigations of patient brain electrical activity and diseases. Earlier the decreasing of the dominant rhythm frequency and changes of relative Fourier spectral power of different frequency bands were found with the help of EEG and EMG spectral analysis [4-9].

Disorders of different organism systems, such as movement disorders, vegetative, emotional, psychical and so on, are features of PD. It is assumed that such disorders reflect in electrical activity of brain.

Due to such approach the time-frequency features of spontaneous EEG of early stage PD were investigated with the help of wavelet Morlet transform. Particular attention was paid to EEG theta rhythm (~4-6 Hz), and disordering of alpha rhythm (~8-12 Hz) in brain cortical motor zones.

## II. METHODS

Continuous wavelet transform (1) with mother function Morlet (2) was used to get EEG signal x(t) time-frequency power density scalogram [1]:

$$S_x(\tau,f) = |W(\tau,f)|^2 , \quad (1)$$

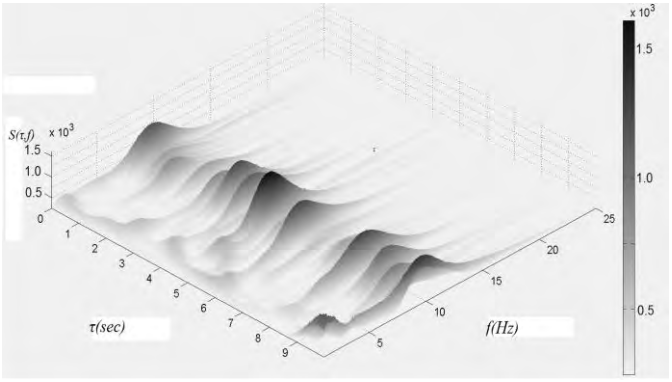
$$W(\tau,T) = \frac{1}{\sqrt{T}} \int x(t)\psi^*\left(\frac{t-\tau}{T}\right)dt , \quad (2)$$

$$\psi(\eta) = \frac{1}{\sqrt{\pi F_b}} e^{2i\pi F_c \eta} e^{-\frac{\eta^2}{F_b}} , \quad (3)$$

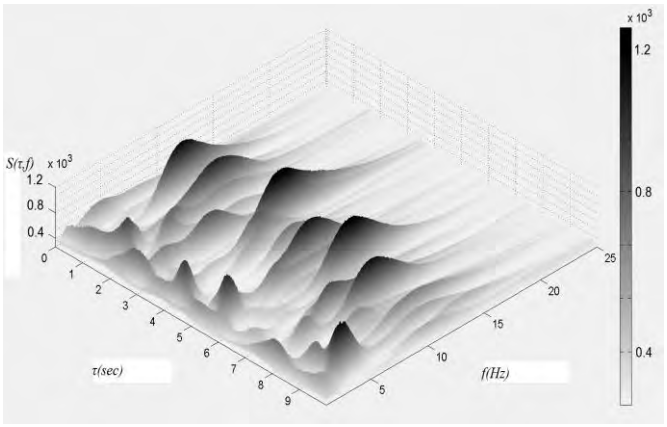
where $\tau$ and $f = 1/T$ are time and frequency of scalogram, $F_b = F_c = 1$.

Figure 1 illustrates the difference in $S(\tau)$ in the brain motor zone C3 (according to the standard 10x20 scheme of electrodes layout) of brain of the normal volunteer (a) and (b) the patient at the first PD according the qualitative stages of PD described by Hoen-Yahr [2]. Below we will take into account two of them. The first takes into account the arising more powerful activity in theta frequency range (4-6 Hz), and the second one

deals with the disorder (non stationary) of electrical activity in the frequency range more then 4-6 Hz.




(a)




(b)

Fig. 1 Time-frequency power density scalograms of normal volunteer (a), and of the 1st stage PD patient (b) of the EEG signals in motor cortex zone C3.

To analyze those features we can consider the scalograms extreme time-frequency distribution. The method of scalograms extreme extraction is written in [10].

Figure 2 illustrates frequency synchronization of C4 EEG, left hand electromyogram (EMG) and measured with the help of accelerometer left wrist tremor of scalograms extreme of the 1st Hoehn-Yahr [11] stage PD patient. It can be considered as an evidence of the role of 4-6 Hz scalograms peaks in movement disorders.

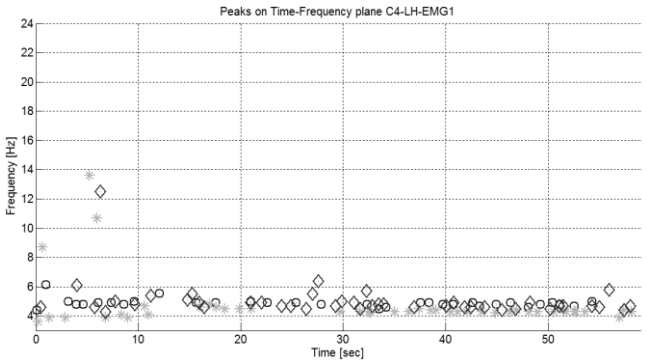
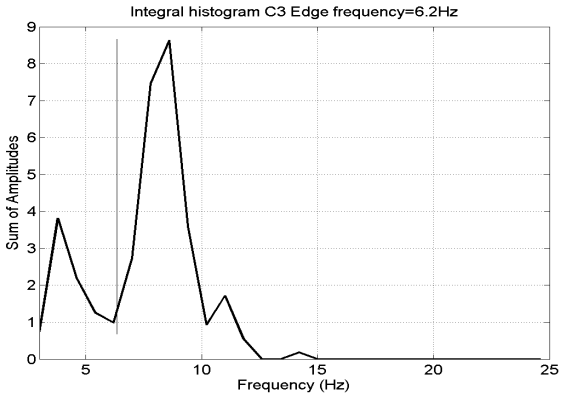
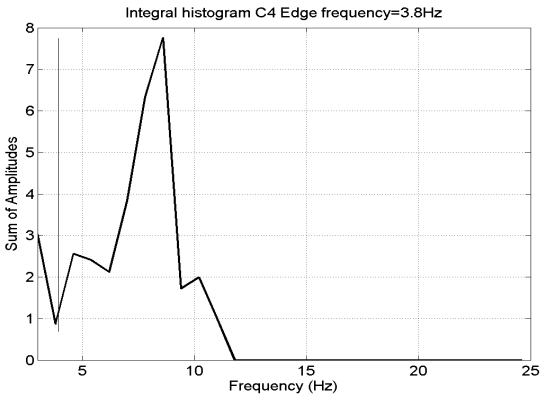



Fig. 2 C4 EEG (circles), left hand electromyogram (rhombs), and left wrist tremor (stars) of scalograms extreme at the 1st stage PD patient

To analyze the scalograms peaks time-frequency distribution we consider the histograms of extreme power sums at ($\Delta f$, $\Delta t$) rectangles.

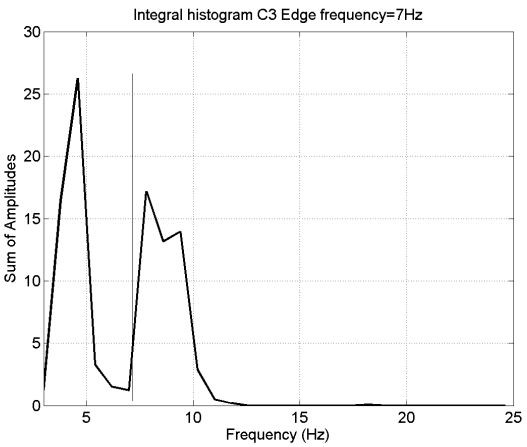



(a)

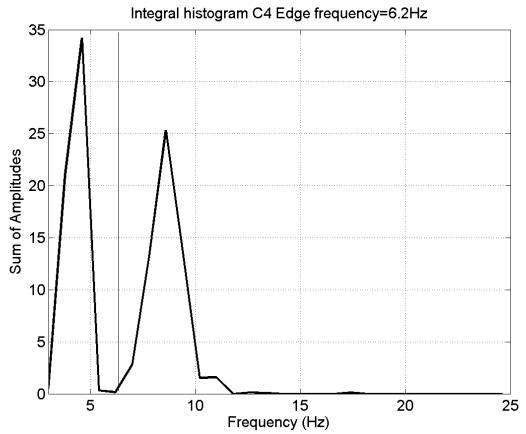



(b)

Fig. 3 Sum of extreme histograms at (0.7 Hz, 180 sec) rectangles for symmetrical C3 (a), and C4(b) EEG electrodes of 1st stage PD patient



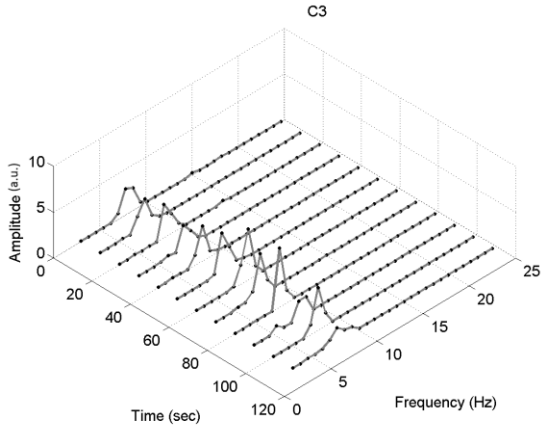

(a)



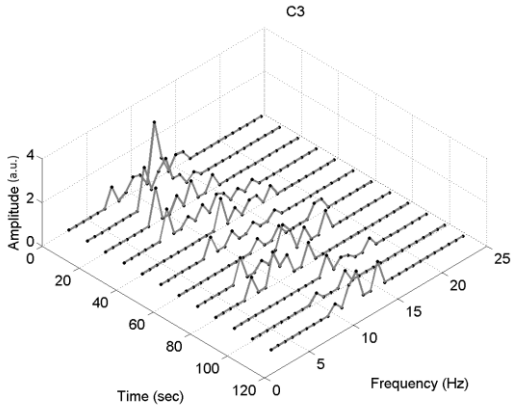

(b)

Fig 4 Sum of extrema histograms at (0.7 Hz, 180 sec) rectangles for symmetrical C3 (a), and C4 (b) EEG electrodes of the 3rd stage PD patient

Figure 3 shows asymmetry of histograms in 4-5 Hz region of the 1st stage PD patient - the existence of theta rhythm in C3 and the absence of such rhythm in C4 electrodes. For the 3d stage PD patient the power of histograms theta rhythms grows in comparison with alpha rhythm (8 Hz), and asymmetry of theta rhythms disappear (fig. 4).




(a)



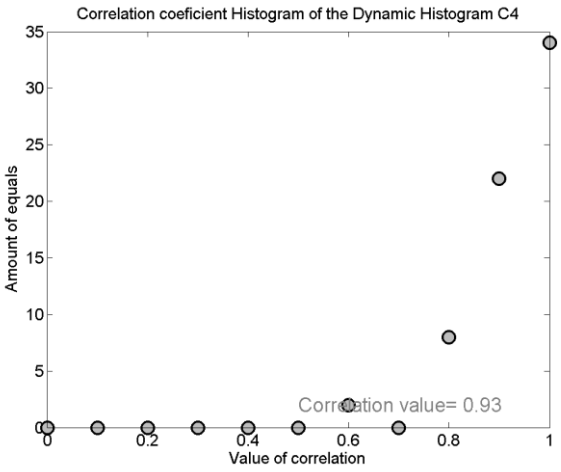

(b)

Fig. 5 Histogram dynamics for the normal volunteer (a), and for the 1st stage PD patient. Histograms was calculated for (0.7 Hz, 10 sec) rectangles

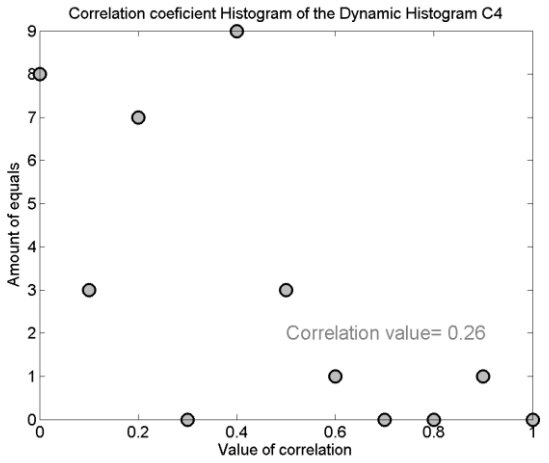
The dynamical histograms calculated at (0.7 Hz, 10 sec) rectangles show the disordering of electrical activity in more then 6 Hz for the 1st stage PD (see fig. 5). Such disordering can be evaluated with the help of dynamical histograms correlation matrixes. This evaluation can be done by histograms of correlation values. Figure 6 shows the difference of such histograms for the normal volunteer and the 2nd PD patient.

The quantitative feature P can be considered as a weighted sum:

$$P = \frac{\sum_{R_i} R_i \cdot N(R_i)}{\sum_{R_i} N(R_i)}, \qquad (4)$$




(a)




(b)

Fig. 6 Histograms of correlation values for the normal volunteer (a), and the 2nd stage PD patient (b).

where $N(R_i)$ is a quantity of correlation values $R_i$ in correlation matrix. The average correlation value $P$ indicated in fig. 6.

## III. CLINICAL RESULTS

EEG investigations of 24 volunteers from control group, 25 non treated PD patients of the 1st Hoehn-Yahr stage and 11 PD patients of the 2nd stage were processed. The mean age of the disease onset was 61.3±7.4, and current age of PD patients was 61.5±9.7 year. The mean current age of healthy control was 61.6±10.9 year. So, two cohorts were age-matched. UPDRS(III) was used to assess the severity of clinical symptoms PD. Among patients of the 1st stage, the average score was 10.29±6.2. The UPDRS (III) score of PD patients of the 2nd stage was significantly higher - 23,5±10,1.

Also 12 patients of the 1st stage were simultaneously investigated with the help of 19 channel EEG, 2 channel EMG, and 2 accelerometers. We evaluated theta rhythm and its asymmetry particularly in motor zone C3 and C4, and non stationary (disordering) properties of alpha rhythms. The diagnosis made with the help of described features gave 80% compatibility with the clinical diagnosis.

Figure 7 shows the decreasing feature *P* or increasing alpha rhythm disorganization with increasing PD stage.

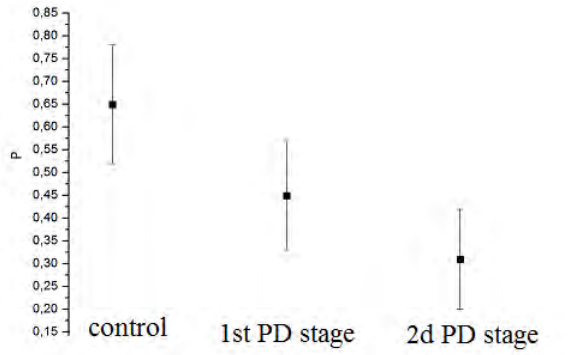



Fig. 7 Average correlation value *P* for investigated normal volunteers, 1st, and 2nd Hoehn-Yahr stages PD patient groups

## IV. CONCLUSION

New time-frequency EEG features of early stages Parkinson's were investigated. There are three main differences of EEG wavelet scalograms were considered as early stage PD features. The first is a disordering ridge of PD patient scalograms in frequency range more then ~6 Hz in comparison with that of normal volunteer. The second is a more powerful PD patient cortex electrical activity in frequency range ~4-6 Hz. And the third feature is a scalograms asymmetry of the left and right brain semi spheres.

This research was supported by Russian Foundation for Basic Research, the project №12-02-00611-a, and by the Program of the Presidium RAS "Fundamental sciences – for medicine".

## REFERENCES

[1] H. Bernheimer, W. Birkmayer, O. Hornykiewicz, K. Jellinger, F. Seitelberger, "Brain dopamine and the syndromes of Parkinson and Hundington. Clinical, neurological and neurochemical correlations". J. Neurol. Sci. 1973. v. 20. № 4. p. 415-455.

[2] "Neurodegenerative Diseases: Fundamental and Applied Issues" / ed. by. M.V. Ugrumov. – Moscow: Nauka, 2010. – ISBN 978-5-02036710-4 (in Russian).

[3] E. Bezard, C.E. Gross, "Compensatory mechanisms in experimental and human parkinsonism: towards a dynamic approach". Prog. Neurobiol. 1998. v. 55. № 2. p. 93-116.

[4] A.C. England, R.S. Schwab, E. Peterson,"The electroencephalogram in Parkinson's syndromee". EEG Clin. Neurophysiol. 1959. v. 11. № 4. p. 723–731.

[5] R. Soikkeli, J. Partenen, H. Soininen, A. Paakkonen, Sr P. Riekkinen, "Slowing of EEG in Parkinson's disease". EEG Clin. Neurophysiol.. 1991. v. 79. № 3. p. 159–165.

[6] D. Stoffers, J. Bosboom, J.B. Deijen, E.C. Wolters, H.W. Berendse, C.J. Stam, "Slowing of oscillatory brain activity is a stable characteristic of Parkinson's disease without dementia". Brain 2007. v. 130. № 7. p. 1847-1860.

[7] K. Sarizawa, S. Kamei, A. Morita, M. Hara, T. Mazutani, H. Yoshihashi, M. Yamaguchi, J. Takeshida, K. Hiravanagi, "Comparison of quantitative EEG between Parkinson disease and age adjusted normal controls". J. Clin. Neurophis. 2008. v. 25. p. 361-366.

[8] M. Moazami-Goudarz, J. Sarnthein, L. Michels, R. Moukhtieva, D. Jeanmonod, "Enhanced frontal low and high frequency power and synchronization in the resting EEG of parkinsonian patients". Neuroimage. 2008. v. 41. № 3. p. 985–997.

[9] H.W. Berendse, C.J. Stam, "Stage-dependent patterns of disturbed neural synchrony in Parkinson's disease". Parkinsonism and Related Disorders. 2007. v. 13, Suppl. 3. p. 440–445.

[10] Yu. V. Obukhov, M.S. Korolev, A.V. Gabova, G.D. Kuznetsova, M.V. Ugrumov, "Method of early stage Parkinson's disease electroencephalography diagnostics" // RF patent. - 2484766, 20.06.2013, (in Russian).

[11] M.M. Hoehn, M.D. Yahr, "Parkinsonism: onset, progression and mortality"., // Neurology. - 1967, V. 17, pp. 427-442, PMID 6067254.

# A Survey of Deep Learning Methods and Software for Image Classification and Object Detection*

V.D. Kustikova, P.N. Druzhkov
Computational Mathematics and Cybernetics Department
Lobachevsky State University of Nizhni Novgorod
Nizhni Novgorod, Russian Federation
itlab.ml@cs.vmk.unn.ru

***Abstract*—Deep learning methods for image classification and object detection are overviewed. Existing software packages for deep learning tasks are compared.**

***Keywords—deep learning; image classification; object detection; sparse coding; autoencoder; restricted Boltzmann machine; convolutional neural networks.***

## I. INTRODUCTION

Since the first works on artificial neural networks they have experienced lots of ups and downs, but have always been of special interest for researchers. Neural networks-based methods have been successfully applied to classification, clustering, forecasting, approximation and recognition tasks in medicine, biology, commerce, robotics etc. The latest advance in the field has been caused by the invention of deep learning methods [1 – 3], induced by the progress of parallel computing hardware and software. The key component of deep learning is the multilayered hierarchical data representation typically in the form of a neural network with more than two layers. Such methods allow automatically synthesizing data descriptions (features) of a higher level based on the lower ones. In terms of image analysis hierarchy levels can correspond to "pixels → edges → combinations of edges" chain. Though deep learning has been expired by neural networks there are some attempts to apply its ideas to other types of models.

Here a survey of deep learning methods aimed at image classification and object detection in images is represented. The applicability of the methods under consideration to these tasks is confirmed by the latest results of well-known competitions such as ImageNet [5] and PASCAL Visual Object Challenge [6], in the context of which the breakthrough in image classification task has been recently made [15, 64].

## II. IMAGE CLASSIFICATION METHODS

### A. Image Classification Problem

Image classification task requires determining the category (class) that it belongs to. The problem is considerably complicated with the growth of categories' count, if several objects of different classes are present at the image and if the semantic classes' hierarchy is of interest, because image can belong to several categories simultaneously. Fuzzy classes place another difficulty of probabilistic categories' assignment.

### B. Sparse Coding

Bag-of-words methods [7] have been among the most popular and successive approaches for solving classification problems before the expansion of deep learning. One of the most advanced methods of this type is *sparse coding* [8]. It maps the initial image description $x \in \mathbb{R}^d$ (with reference to image classification it can be a vector of pixels' intensities, dense SIFT-descriptor [9, 10] etc.) to a vector $s \in \mathbb{R}^m$ with lots of zero components, such that $x \approx D \cdot s$, where $D \in \mathbb{R}^{d \times m}$ and $m$ can be much greater than $d$. $D$ is called a dictionary and $s$ is a code. With a fixed dictionary the task of new representation (i.e. code) computing is stated as $s^* = \arg\min_s (Err(x, D \cdot s) + \lambda \Psi(s))$, where $Err(x, D \cdot s)$ component is responsible for the coding error, and, generally equals to $(\|x - D \cdot s\|_2)^2$ and $\Psi(s)$ defines a code sparseness constraint, for example, as $\|s\|_0$ or $\|s\|_1$. Dictionary can be learned from data $X$, solving $D^* = \arg\min_D \sum_{x \in X} \min_s (Err(x, D \cdot s) + \lambda \Psi(s))$.

The common sparse coding classification pipeline is as follows. The dictionary is learned from a set of unlabeled images, code is computed for each labeled image at hand, and a classifier is trained trying to predict the class by the code. As labeled data is not required at the dictionary learning stage, this approach is advantageous in scarce labeled data situations. In opposite case dictionary learning can be considered in a supervised manner, providing additional information to improve features' quality [11 – 13]. Sparse coding as it described here is not capable of building features' hierarchies and it is not straightforward to simply stack one coding model on top of the other [63]. Though rather successful attempts to make sparse coding deep exist [10, 14] but there is still room for improvement. It should be also mentioned that sparse coding is not the only algorithm that has been attempted to make it deep [65, 66].

* This work is supported by Russian Foundation for Basic Research (project No 14-07-31269) and has been done in the "Information Technology" laboratory at Computational Mathematics and Cybernetics Department of Lobachevsky State University of Nizhni Novgorod.

### C. *Deep Learning Models: Autoencoders, Restricted Boltzmann Machines, Convolutional Neural Networks*

During the past years deep learning approaches that are heavily based on neural networks have been of special interest. Each node of a network (i.e. artificial neuron) is associated with a feature, and neurons of the subsequent layer generalize essential features from the previous one. The big amount of trainable parameters leads to necessity of network topology and activation functions constraining and development of highly parallel algorithms for training. Therefore considerable emphasis is placed on searching of networks' topologies that better suit for image classification and effective methods for their training. One of the common training techniques is to use the unsupervised pre-training stage that allows a preliminary fitting using only unlabeled data. It is proved to be a good starting point for a subsequent supervised fine-tuning. In this context two main types of models can be distinguished:

- *Autoencoder* (AE) [20]. AE performs coding with loss of information so that the result of subsequent decoding is as close to the original data as possible (fig. 1a). In general coding function can be represented as $h = f(x) = s_f(Wx+b_h)$, and decoding one as $y = g(h) = s_g(W^T h+b_y)$. Code $h$ is nothing more than a feature vector of the current level of hierarchy. AE is designed to find such values of parameters $\{W, b_h, b_y\}$ that lead to minimal deviation of $y$ from $x$, defined by the loss function. Different optimization techniques for parameters fitting exist and the most popular one is a back propagation (BP) method. Depending on a nonlinear coding function part $s_f$ and a loss type AEs are subdivided into sparse (SAE) [21], contractive (CAE) [22] and denoising (DAE) [23]. Using an AE to some extent allows filtering out insignificant details for visual object modeling. AEs can be stacked (StAE) (fig. 1b) to produce the hierarchy of features [23].

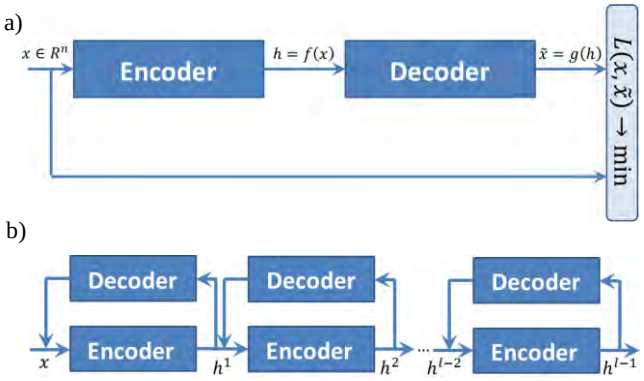



Fig. 1. Scheme of autoencoder and stacked autoencoder.

- *Restricted Boltzmann machine* (RBM) [25, 33, 62]. As opposed to AEs RBMs are stochastic neural models. An RBM is a neural network with two layers corresponding to visible and hidden states of a probabilistic system. Each node of one layer is connected to every node of the other layer (fig. 2). Visible neurons correspond to a given initial feature description and hidden ones – to features derived as functions of visible variables. RBM defines the probability distribution on the set of its visible states and training objective is to maximize the likelihood of a training sample. Effective methods for RBM training have been designed. Among them are contrastive divergence (CD) and its k-step (kCD) and persistent (PCD) [33] variations. RBMs can be used in deep models such as deep belief networks (DBN) [26], deep Boltzmann machines (DBM) [34] and be applied for deep AEs [35] and convolutional neural networks [26] pre-training.

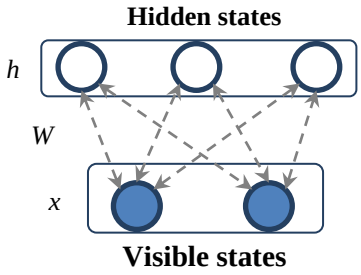



Fig. 2. Restricted Boltzmann Machine.

Pre-training stage can be avoided if using *convolutional neural networks* (CNNs) [36]. Each layer of CNN represents a feature map. Feature map of an input layer is a matrix of pixel intensities (a separate matrix for each color channel). And feature map of any internal layer is an induced multi-channel image, where every "pixel" corresponds to a specific feature. Every neuron is connected with a small portion of adjacent neurons from the previous layer (receptive field). It is typical to interleave the layers doing different types of transformations [15, 24, 37, 38] on feature maps, such as filtering and pooling (fig. 3). Filtering function computes a discrete convolution of filter-matrix with a receptive field neurons' values followed by a non-linear transform such as sigmoid and pooling is a possibly non-linear transformation that allows summarizing a receptive filed by one value making feature descriptions more robust. Max, average, $L_2$-polling are among the most popular choices. Local contrast normalization [41] is another operation that has proved to be useful in CNNs.

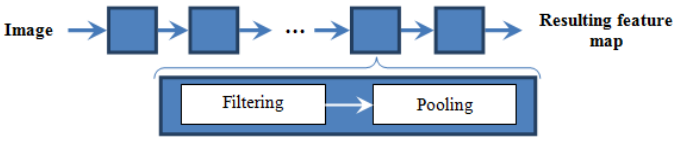



Fig. 3. Structure of a convolutional neural network.

As soon as initial features hierarchy is constructed it can be fine-tuned in a supervised manner [4]. Final layer is added to the network such that each output neuron gives a conditional probability that the input image belongs to a specific class. Sigmoid and softmax are typically used as activation functions for this layer [15] and mean squared error $MSE = 1/n \sum_{1 \le i \le n}(y_i - y^*_i)^2$ or cross-entropy loss $L = -\sum_{1 \le i \le n} \log (P(y_i = y^*_i))$ is minimized via the stochastic gradient descent (SGD) method. If fine-tuning of features is not required any type of classifier such as SVM [16] can be simply applied to the output of the final layer.

TABLE I. COMPARISON OF DEEP LEARNING SOFTWARE TOOLS

| Name | Language | OS | FC NNs | CNNs | AE | RBMs | Learning method | Loss |
|---|---|---|---|---|---|---|---|---|
| DeepLearn-Toolbox [45] | Matlab | Win, Lin | + (Dropout) | + | + (StAE,DAE) | + (DBN) | BP | MSE |
| Teano [53] | Python | Win, Lin, Mac | + | + | + (DAE) | + (DBN) | SGD | 1. Negative Log-Likelihood 2. Zero-One |
| Pylearn2[1] [55] | Python | Lin, Vagrand | + (Maxout Networks) | + | + (StAE, CAE, DAE) | + (GRBM[2], ssRBM[3]) | SGD, BGD[4], NCG[5] (line search)[6] | 1. Cross-entropy 2. Log-likelihood |
| Deepnet[7] [56] | Python | Lin | + (Dropout) | + | + | + (DBN, DBM) | GD, LBFGS[8], CD, PCD | 1. Squared 2. Cross-entropy 3. Hinge |
| Deepmat [49] | Matlab | ? | + (Dropout) | + | + (DAE,StDAE[9]) | + (Gaussian RBM, DBN, DBM) | SBP[10], Adagrad, CD, PCD | ? |
| Darch [50] | R | Win, Lin | + | – | + | + (DBN) | BP, CG[11], CD | 1. MSE 2. Cross-entropy 3. Quadratic error |
| Torch [52] | Lua, C | Lin, iOS, Android | + | + | + | – | SGD, LBFGS,CG | 1. MSE 2. Binary cross-entropy 3. Hinge 4. Others |
| Caffe [43] | C++, Python, Matlab | Lin | + (Dropout) | + | – | – | SGD | 1. Euclidean 2. Hinge 3. Info gain 4. Logistic 5. Sigmoid cross entropy 6. Softmax |
| nnForge [44] | C++ | Lin | + (Dropout, MaxoutNetworks) | + | – | – | SGD, SDLMA[12] | 1. MSE 2. Squared Hinge 3. Negative Log-Likelihood. 4. Cross-Entropy. |
| CXXNET [60] | C++ | Lin | + (Dropout, Drop Connection) | + | – | – | SGD | ? |
| Cuda-convnet [46] | C++ | Win, Lin | + | + | – | – | ? | 1. Logistic regression 2. Sum-of-squares |
| Cuda CNN [48] | Matlab | Win, Lin | + | + | – | – | SGD, SDLMA | ? |
| EBLearn [47] | C++ | Lin, Mac | + | + | – | – | BP | Energy-based functions |
| Hebel [59] | Python | Win, Lin | + (Dropout) | in development | in development | in development | SGD | ? |
| Crino[13] [61] | Python | ? | + | – through [53] | + | – through [53] | SGD, BP | 1. MSE 2. Cross-entropy 3. Mean Absolute Error |
| Lush [54] | Lush | Win (Cygwin), Lin | + | + | – | – | SGD, BGD, SLMA, BCG[14] | 1. MSE 2. ? |
| R-CNN[15] [42] | Matlab | Lin | – through [43] | + | – | – | – through [43] | – through [43] |

[1]Based on Teano, uses Cuda-convnet, implements differentiable sparse coding and spike-and-slab sparse coding.
[2]Gaussian RBM (GRBM).
[3]The spike-and-slab RBM (ssRBM) [[62]].
[4]Batch Gradient Descent (BGD).
[5]Nonlinear conjugate gradient descent (NCG).
[6]BGD and NCG with searching for an optimal local solution at every step.
[7]Uses cudamat and cuda-convnet.
[8]Limited-memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithm.
[9]Stacked Denoising Auto-Encoders (StDAE).
[10]Stochastic Back Propagation (SBP).
[11]Conjugate gradient method (CG).
[12]Stochastic Diagonal Levenberg-Marquardt algorithm (SDLMA).
[13]Based on Teano, implements MLP, DNN, IODA.
[14]Batch Conjugate Gradient (BCG).
[15]Based on Caffe and provides an interface to it, implements a special type of CNNs [19].

Resulting classifier can be applied to an arbitrary image of the same size, as those from the training set. If the image has a different resolution it can be scaled and cropped as in [15].

## III. Object Detection Methods

Detection problem is more general in sense that it requires not only to determine whether the object of interest is present in image but also tell where all of its instances are located. Object detection is still a challenging problem due to a big amount of factors that must be handled: variety of possible objects' forms and colors, occlusions, lighting conditions, perspective etc.

Detection methods can be distinguished into three groups [27]: approaches, based on features extraction [28, 29]; template searching methods [30, 31]; movement detection [32]. Deep learning methods fall into the first group. The most straightforward way is to apply a deep classifier, trained as considered above, to regions of interest generated by the extensive *sliding window approach* [18, 17, 37], or some more cost efficient method [19, 39, 40]. Nevertheless determination on object position, size and scale can be embedded into the neural network [17, 18, 25] by adding layers of a special type.

Experimental results have shown that deep detectors are among the best-performing modern models, but no extreme improvement of the state-of-the-art has been performed yet.

## IV. Implementation

Software tools implementing deep learning methods include libraries and packages [44 – 49, 53, 55, 56, 58 – 61], programming language extensions [52, 57] and self-contained languages [54]. Provided functionality varies a lot (tab. 1) raising the problem of choosing an appropriate tool.

A great batch of tools support a broad series of models including fully connected neural networks (FCNNs), CNNs, AEs and RBMs and implement popular training methods and loss functions. Deep Learn Toolbox [45], Teano [53], Pylearn2 [55], Deepnet [56] and DeepMat [49] are among them. Darch [50] package for R [51] also falls into this category, but it doesn't support CNNs. It is worth mentioning that Deep Learn Toolbox is solely implemented in MATLAB. It leads to some performance issues and forces to be cautious with the use of big data. Torch [52], Caffe [43], nnForge [44], CXXNET [60], Cuda-convnet [46] and Cuda CNN [48] represent the group of deep learning tools aimed at high-performance training of CNNs using GPUs through CUDA. Some of them are used inside the tools mentioned above. The rest of tools mostly implement varieties of deep neural networks using other libraries. For example, Crino [61] and R-CNN [42] are based on Teano and Caffe respectively. Lush [54] programming language is also an interface to Torch library.

This overview doesn't cover all of the available tools because of their vast and steadily increasing amount.

## References

[1] G.E. Hinton, "Learning Multiple Layers of Representation," Trends in Cognitive Sciences, vol. 11, pp. 428-434, 2007.

[2] J. Schmidhuber, "Deep Learning in Neural Networks: An Overview," [http://arxiv.org/abs/1404.7828].

[3] Resources and pointers to information about Deep Learning. [http://deeplearning.net]. 07.08.2014.

[4] D.P. Vetrov, "Machine Learning: Current State and Perspectives," In Proc. of RCDL, vol. 1, pp. 21-28, 2013. (In Russian).

[5] ImageNet [http://www.image-net.org].

[6] PASCAL Visual Object Challenge [http://pascallin.ecs.soton.ac.uk/challenges/VOC].

[7] C. Dance, J. Willamowski, L. Fan, C. Bray, G. Csurka, "Visual categorization with bags of keypoints," In Proc. ECCV Int. Workshop on Statistical Learning in CV. 2004.

[8] H. Lee, A. Battle, R. Raina, A.Y. Ng, "Efficient sparse coding algorithms," In Proc. of NIPS, pp. 801–808, 2006.

[9] D. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

[10] Y. He, K. Kavukcuoglu, Y. Wang, A. Szlam, Y. Qi, "Unsupervised Feature Learning by Deep Sparse Coding," In Proc. of SIAM Int. Conf. on Data Mining, pp. 902-910, 2014.

[11] J. Yang, K. Yu, T. Huang, "Supervised translation-invariant sparse coding," In Proc. of CVPR, pp. 3517-3524, 2010.

[12] Q. Zhang, B. Li, "Discriminative k-svd for dictionary learning in face recognition," In Proc. of CVPR, pp. 2691-2698, 2010.

[13] Z. Jiang, Z. Lin, L.S. Davis, "Learning a discriminative dictionary for sparse coding via label consistent k-svd," In Proc. of CVPR, pp. 1697-1704, 2011.

[14] A. Coates, H. Lee, A.Y. Ng, "An analysis of single-layer networks in unsupervised feature learning," In Proc. of Artificial Intelligence and Statistics, vol. 15, pp. 215-223, 2011.

[15] A. Krizhevsky, I. Sutskever, G.E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," In Proc. of NIPS, pp. 1097–1105, 2012.

[16] K. Simonyan, A. Vedaldi, A. Zisserman, "Deep Fisher Networks for Large-Scale Image Classification," In Proc. of NIPS, pp. 163–171, 2013.

[17] C. Szegedy, A. Toshev, D. Erhan, "Deep Neural Networks for Object Detection," In Proc. of NIPS, pp. 2553–2561, 2013.

[18] D. Erhan, C. Szegedy, A. Toshev, D. Anguelov, "Scalable Object Detection using Deep Neural Networks," In Proc. of CVPR. 2014.

[19] R. Girshick, J. Donahue, T. Darrell, J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," In Proc. of CVPR, pp. 580–587, 2014.

[20] M. Hayat, M. Bennamoun, S. An, "Learning Non-Linear Reconstruction Models for Image Set Classification," In Proc. of CVPR, 2014.

[21] M. Ranzato, C. Poultney, S. Chopra, "Efficient Learning of Sparse Representations with an Energy-Based Model," In Proc. of NIPS, pp. 1137–1144, 2006.

[22] S. Rifai, P. Vincent, X. Muller, X. Glorot, Y. Bengio, "Contractive Auto-Encoders: Explicit Invariance during Feature Extraction," In Proc. of ICML, pp. 833–840, 2011.

[23] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion," Journal of Machine Learning Research, vol. 11, pp. 3371–3408, 2010.

[24] K. Kavukcuoglu, P. Sermanet, Y-lan Boureau, K. Gregor, M. Mathieu, Y.L. Cun, "Learning Convolutional Feature Hierarchies for Visual Recognition," In Proc. of NIPS, pp. 1090–1098, 2010.

[25] P. Luo, Y. Tian, X. Wang, X. Tan, "Switchable Deep Network for Pedestrian Detection," In Proc. of CVPR, 2014.

[26] H. Lee, R. Grosse, R. Ranganath, A.Y. Ng, "Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations," In Proc. of ICML, pp. 609-616, 2009.

[27] V.D. Kustikova, N.Yu. Zolotykh, I.B. Meyerov, "A review of vehicle detection and tracking methods in video," Vestnik of Lobachevsky State University of Nizhni Novgorod, no. 5(2), pp. 347–357, 2012. (In Russian).

[28] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, D. Ramanan, "Object Detection with Discriminatively Trained Part Based Models," IEEE Trans. on PAMI'10, vol. 32, no. 9, pp. 1627–1645, 2010.

[29] J. Shotton, A. Blake, R. Cipolla, "Contour-based Learning for Object Detection," In Proc. of ICCV, vol. 1, pp. 503-510, 2005.

[30] C.H. Hilario, J.M. Collado, J.M. Armingol, A. de la Escalera, "Pyramidal Image Analysis for Vehicle Detection," In Proc. of Intelligent Vehicles Symposium, pp. 88–93, 2005.

[31] Y. Amit, "2D Object Detection and Recognition: models, algorithms and networks," The MIT Press. 2002.

[32] M. Sonka, V. Hlavac, R. Boyle, "Image Processing, Analysis and Machine Vision," Thomson. 2008.

[33] Restricted Boltzmann Machines (RBMs) [http://www.deeplearning.net/tutorial/rbm.html]. 07.08.2014.

[34] R. Salakhutdinov, G. Hinton, "Deep Boltzmann Machines, DBMs," [http://www.cs.toronto.edu/~fritz/absps/dbm.pdf]. 07.08.2014.

[35] Q. Le, M. Ranzato, R. Monga, M. Devin, K. Chen, G. Corrado, J. Dean, A. Ng, "Building high-level features using large scale unsupervised learning," In Proc. of ICML. 2012.

[36] Y. LeCun, K. Kavukcuoglu, C. Farabet, "Convolutional networks and applications in vision," In Proc. of ISCAS, pp. 253–256, 2010.

[37] M. Oquab, L. Bottou, I. Laptev, J. Sivic, "Weakly supervised object recognition with convolutional neural networks," In Proc. of NIPS, 2014.

[38] M. Oquab, L. Bottou, I. Laptev, J. Sivic, "Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks," 2013. [http://hal.inria.fr/docs/00/91/11/79/PDF/paper.pdf].

[39] J.R.R. Uijlings, K.E.A. van de Sande, T. Gevers, A.W.M. Smeulders, "Selective Search for Object Recognition," In International Journal of Computer Vision, vol. 104, no. 2, pp. 154–171, 2013.

[40] X. Wang, M. Yang, S. Zhu, Y. Lin, "Regionletsfor generic object detection," In Proc. of ICCV, 2013.

[41] K. Kavukcuoglu, M. Ranzato, R. Fergus, Y. LeCun, "Learning invariant features through topographic filter maps," In Proc. of CVPR, pp. 1605–1612, 2009.

[42] R-CNN – a visual object detection system [https://github.com/rbgirshick/rcnn].

[43] Caffe – a deep learning framework [http://caffe.berkeleyvision.org].

[44] nnForgeLibrary [http://milakov.github.io/nnForge].

[45] DeapLearnToolbox [https://github.com/rasmusbergpalm/DeepLearnToolbox].

[46] Cuda-convnet – high-performance C++/CUDA implementation of convolutional neural networks [http://code.google.com/p/cuda-convnet].

[47] EBLearn – a machine learning library [http://eblearn.sourceforge.net].

[48] Cuda CNN Library [http://www.mathworks.com/matlabcentral/fileexchange/24291-cnn-convolutional-neural-network-class], [https://bitbucket.org/intelligenceagent/cudacnn-public/wiki/Home].

[49] DeepMat Library [https://github.com/kyunghyuncho/deepmat].

[50] Package Darch [http://cran.r-project.org/web/packages/darch/index.html].

[51] Software Environment R [http://www.r-project.org].

[52] Torch – a scientific computing framework [http://www.torch.ch].

[53] Teano Library [https://github.com/Theano/Theano], [http://deeplearning.net/software/theano].

[54] Lush programming language [http://lush.sourceforge.net].

[55] Pylearn2 – a machine learning library [http://deeplearning.net/software/pylearn2].

[56] Deepnet Library [https://github.com/nitishsrivastava/deepnet].

[57] DeCAFFramework [https://github.com/UCB-ICSI-Vision-Group/decaf-release].

[58] Cuda-convnetот NYU [http://cs.nyu.edu/~wanli/dropc].

[59] Hebel – GPU-accelerated deep learning library [https://github.com/hannes-brt/hebel].

[60] CXXNET – a neural network toolkit [https://github.com/antinucleon/cxxnet].

[61] Crino – a neural network library [https://github.com/jlerouge/crino].

[62] A. Courville, J. Bergstra, Y. Bengio, "A Spike and Slab Restricted Boltzmann Machine," 2011. [http://jmlr.org/proceedings/papers/v31/luo13a.pdf].

[63] Y. He, K. Kavukcuoglu, Y. Wang, A. Szlam, Y. Qi, "Unsupervised Feature Learning by Deep Sparse Coding," In Proc. of ICDM, pp. 902–910, 2014.

[64] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," [http://arxiv.org/abs/1409.0575].

[65] C. Vens, F. Costa, "Random Forest Based Feature Induction," In Proc. of ICDM, pp. 744–753, 2011.

[66] V.Yu. Martyanov, A.N. Polovinkin, E.V. Tuv, " Image classification with codebook based on decision tree ensembles," In Proc. of Intelligent Information Processing, pp. 480–482, 2012. (In Russian).

# *A Technique for Comparing Images of Paintings for Attribution**

D. Murashov
Dorodnicyn Computing Centre of RAS
Moscow, Russian Federation
d_murashov@mail.ru

***Abstract*— In this work, the problem of comparing images for the purpose of attribution of fine-art paintings is considered. A technique for comparing images of paintings is proposed. Features that are used in this work describe texture of artworks and characterize an artistic style of a painter. A procedure for feature extraction is developed. Paintings are compared using three informative fragments segmented in a particular image. Selected image fragments are compared by information-theoretical dissimilarity measure based on Kullback-Leibler divergence. The technique is tested on images of portraits created in 17-19th centuries. The results of the experiments showed that the difference between portraits painted by the same artist is substantially smaller than one between portraits painted by different authors. The proposed technique may be used as a part of technological description of fine art paintings for attribution.**

***Keywords—images of paintings; attribution; artistic style; image ridges; structure tensor; orientation angle; image difference***

## I. INTRODUCTION

The paper deals with the developing techniques for computer-based image analysis for attribution of fine-art paintings. The idea of applying image analysis in attribution is that to compare images of authentic and studied paintings by features characterizing individuality of an artist. This idea is based on the concepts of Giovanni Morelli, who laid the foundations of the method for comparative analysis in fine arts [1].

In this work, the images of portraits are analyzed. The experts associate individuality of an artist with brushwork features. In accordance to recommendations of art experts [1,2], we use for comparing paintings the groups of brushstrokes that form details of paintings. For example, in portraits such details are lips, chin, nose, forehead, eyes, folds of clothes, etc. In [3] for attribution of portrait miniatures, the homotypic fragments of the human faces were compared. The fragments were segmented using geometric model of face.

In this paper we also use images of the homotypic informative face details: forehead, nose, and cheek. It should be noticed that the size of the images in current research is much larger than in [3] and selected image fragments differ from those in [3]. Three types of informative face fragments used in this work are shown in Fig. 1. The proposed techniques are aimed at extraction of compatible features of an artistic manner from images of paintings.

This work is supported by the RFBR grant No 12-07-00668.

The paper is organized in the following way. In the next sections we give the formal problem formulation, describe the analyzed images of paintings and textural features capturing the artistic manner, and propose feature extraction procedures. Then we introduce a technique for comparing image feature descriptions based on Kullback-Leibler divergence. In two last sections we present the results of computing experiment and make conclusions.

## II. PROBLEM FORMULATION

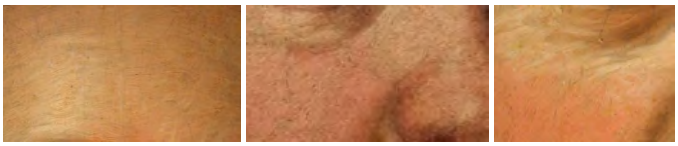
The problem is formulated as follows. Let $U_j$, $j = 1,2,...,J$ be images of paintings by $J$ authors; $U : R^2 \to R$. Let $u_j^i : \Omega \to R,\ \Omega \subset R^2,$ be an informative fragment of type $i$ taken from image $U_j$, $i = 1,2,...,\mathrm{I}$. We suppose that $u_j^i$ is characterized by a feature vector $\mathbf{x}_j^i = \left[ x_j^{i1}, x_j^{i2}, ...x_j^{is}, ..., x_j^{iS} \right]^T$, $x_j^{is} = \gamma_s(u_j^i)$, $\gamma_s : R^2 \times R \to R$, $s = 1,2,...,S$. The difference between two fragments $u_j^i$ and $u_k^i$ of type $i$ of images $U_j$ and $U_k$ we define as

$$D_{jk}^i = \sqrt{\sum_{s=1}^{S} \left( d(x_j^{is}, x_k^{is}) \right)^2}, \quad (1)$$

where $d(x_j^{is}, x_k^{is})$ is a measure of difference between the features $s$ of these two images. The difference $D_{ik}$ between the images $U_j$ and $U_k$ is calculated from differences between corresponding informative fragments as follows:

$$D_{jk} = \sqrt{\sum_{i=1}^{I} \left( D_{jk}^i \right)^2}. \quad (2)$$

Let $U_l$, $l = J+1$ be an image with unknown attribution. It is necessary to find image $U_m$ (and the author of the painting) providing minimum of distance $D_{ml}$.



(a) (b) (c)

Fig. 1. Informative fragments of portraits: (a) – "forehead"; (b) – "nose"; (c) – "cheek".

## III. IMAGES OF PAINTINGS

The data used in current research are the image fragments of artworks painted in 17 – 19th centuries by different authors. The images are fixed by a digital camera. The size of the images is about 4272x2848 pixels. Distortions conditioned by an acquisition process are compensated and images are uniformly oriented. The size of informative fragments varies from 990x814 to 1800x1000 pixels at resolution of 200 dots per cm, that corresponds to the quality of the data used in the analogous studies. For example, Johnson et al. [4], Polatkan et al. [5], and Li et al. [6] analyzed images obtained at resolution of 196 dots per inch. Some of the paintings have retouched and repainted areas. The features should be extracted only from areas with original brushwork. Thereby, retouched and repainted areas should be excluded during feature extraction. A technique developed in [7] is used for localizing damaged paint layer areas.

## IV. IMAGE FEATURES

Taking into account the complexity of brushstroke segmentation, it is preferable to use image features that do not need segmentation of a single brushstroke. The features will be extracted only from the areas containing maximum information about the artistic manner of the painter. For describing individual manner of artists the following textural features have been proposed in [8]: (a) local orientation of grayscale image ridges; (b) simple neighborhood orientation based on the local structure tensor. Histograms of brushstroke ridge orientation and local neighborhood orientation are considered as the features of a brushstroke group that describe the individual artistic manner specific to a particular detail of a painting.

Image ridges are localized by modified technique described in [10]. A set of points forming a ridge of an object in grayscale image is extended by including parabolic and umbilic points of a grayscale image relief.

Let the grayscale image relief be a function $f \in C^2(R^2, R)$. It is assumed that $Df \neq 0$, $Df = \left(f_x, f_y\right)^T$. We denote $N = Df \big/ |Df|$, $T = Df^{\perp} \big/ |Df|$, and $Df^{\perp} = \left(-f_y, f_x\right)^T$, where $N$ is a normal and $T$ is a tangent to level lines (isophotes) of image $f$. The following expression based on Hessian describes the local properties of function $f$:

$$-\frac{1}{|Df|}\begin{bmatrix} N^T D^2 f N & N^T D^2 f T \\ T^T D^2 f N & T^T D^2 f T \end{bmatrix} = \begin{bmatrix} g & \mu \\ \mu & k \end{bmatrix}, \quad (3)$$

where $k = -T^T \left(D^2 f \big/ |Df|\right) T$ is an isophote curvature; $\mu = -T^T (D^2 f \big/ |Df|) N$ is a gradient flowline curvature; $g = -N^T (D^2 f \big/ |Df|) N$ is a measure of gradient variation along the flowlines. At ridge points of $f$ the following conditions are taking place: $\mu = 0$ and $k > \max\{0, g\}$. We also consider the points of $f$ where one or both eigenvalues of matrix (3) are equal to zero. For obtaining ridge directional histogram, the orientation angle of inertia axes of ridge connected components is calculated [12]:

$$\theta = \frac{1}{2} \arctan \frac{2\mu_{1,1}}{\mu_{2,0} - \mu_{0,2}}, \quad (4)$$

where $\mu_{i,j}$ are the elements of inertia tensor:

$$J = \begin{bmatrix} \mu_{2,0} & -\mu_{1,1} \\ -\mu_{1,1} & \mu_{0,2} \end{bmatrix}.$$

For computing direction histogram the length (in pixels) of ridge connected components was taken into account.

Another feature describing local orientation of painting texture is based on the notion of structure tensor, or the second moment matrix at a point $x$ weighted by a window function:

$$\mu_f(x) = \int_{p \in R^2} (Df(p))(Df(p))^T w(x-p)dp, \quad (5)$$

where $w(x-p)$ is a window Gaussian function [11]. The angle of simple neighborhood orientation $\varphi$ is determined as:

$$\varphi = \frac{1}{2} \arctan \frac{2\mu_{f1,1}}{\mu_{f2,0} - \mu_{f0,2}}, \quad (6)$$

where $\mu_{fi,j}$ are the components of the structure tensor (5).

The procedures for computing the features are developed. For obtaining the histogram of orientation angles of grayscale image ridges the following operations are performed: (a) image rotation and scaling; (b) extension of image dynamic range; (c) creating a mask of informative fragment; (d) creating a craquelure and damage mask; (e) combining masks; (f) image masking; (g) Gaussian blurring; (h) localizing ridges of grayscale image relief; (i) defragmenting obtained image ridges; (j) filtering connected components of ridges by size; (k) computing orientation angle values and building a histogram. For extraction the second feature, firstly the image is downsampled with factor 2. Then the operations (a)-(g) listed above are performed. After this, for each pixel marked out by the mask the components of structure tensor (5) are

computed and neighborhood orientation angle (6) is determined. Finally, a histogram of simple neighborhood orientation is obtained for a particular image fragment. The procedure for creating a craquelure mask includes operations of "black top-hat"; adaptive thresholding, interactive selection of connected components, morphological opening, and dilation.

## V. Comparing Images of Paintings

For comparing fragments of artworks, the statistical tests [10], cluster analysis, and classification techniques [5, 6] are used. In this paper, the image samples are compared using information-theoretical measure of difference, because this measure fits the features represented by distributions. The measure is constructed on the basis of Kullback-Leibler divergence as follows [9]:

$$d\left(x_j^{is}, x_k^{is}\right) = \frac{1}{2}\left[\sum_{\varphi\in H} p_\Phi(\varphi)\log\frac{p_\Phi(\varphi)}{q_\Phi(\varphi)} - \sum_{\varphi\in H} q_\Phi(\varphi)\log\frac{q_\Phi(\varphi)}{p_\Phi(\varphi)}\right], \quad (7)$$

where $p_\Phi(\varphi)$ and $q_\Phi(\varphi)$ are the probabilities of the event when orientation angle values in samples $u_j^i$ and $u_k^i$ are equal to $\varphi$; $H$ is the alphabet of random variable $\Phi$ representing the orientation angle. The measure $d\left(x_j^{is}, x_k^{is}\right)$ is non-negative and symmetric. To compare the paintings we aggregate measures as described by the expressions (1) – (2).

## VI. Computing Experiment

Computing experiment was carried out for tuning feature extraction algorithms and testing the proposed feature description for applicability to attribution tasks.

In the experiment we use images of three portraits by F. Rokotov and eight portraits by other artists dated to seventeenth-nineteenth centuries.

For comparing images of paintings the feature description (4), (6) is used. Firstly, according to the procedure proposed above, we create craquelure masks for specified portrait regions (forehead, nose, and cheek, see Fig. 1). We apply created masks to image patches at the step of computing feature description. According to the developed feature extraction procedure, we obtain histograms of orientation angles defined by expressions (4) and (6). Secondly, using measure of difference (7) the distances $D_{jk}^i$ between fragments of type $i$ in portraits $j$ and $k$ are computed. At the next step as defined by expression (2), we aggregate computed distances $D_{jk}^i$ between homotypic portrait regions into the values of distances $D_{jk}$ between portraits $j$ and $k$.

For tuning feature extraction algorithms we computed distances

$$D_{jk}^s = \sqrt{\sum_{i=1}^{I}\left(d(x_j^{is}, x_k^{is})\right)^2} \quad (8)$$

between images represented by feature $s$ for different values of parameters.

The first parameter under consideration is the lower bound $b$ of the size of ridge connected components. Distances $D_{jk}^s$ between portraits for values of $b$ equal to 8 and 12 pixels were computed. Histograms of distances between images of paintings represented by orientation angle of ridge connected components at $b=8$ and $b=12$ are shown in Fig. 2. Here, distances between three portraits by F. Rokotov are denoted as "own" and distances between portraits by different artists are denoted as "alien".

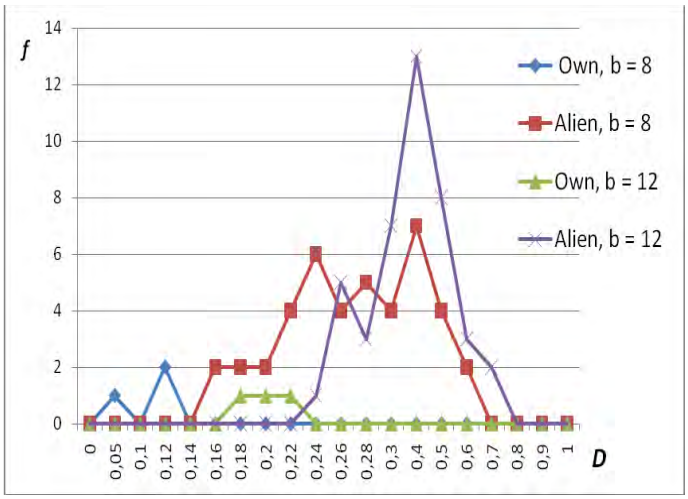



Fig. 2. Histograms of distances between images of paintings represented by orientation angle of ridge connected components at $b=8$ and $b=12$; $f$ is a frequency of a distance value $D$.

From Fig. 2 it can seen that "own" and "alien" portraits separated better at $b=8$.

The second parameter is the size of the window function $w$ in (5). Distances $D_{jk}^s$ between portraits for values of window function size $a$ equal to 3, 5, 7, and 11 pixels were computed. Averages and standard deviations found from histograms of distances between "own" and "alien" images of paintings represented by angle of simple neighborhood orientation $\varphi$ at different values of $a$ are presented in Fig. 3. The point at $a=3$ corresponds to noise components of painting's texture. Other values of window function size produce distances revealing different components of brushwork texture characterized by different spatial frequencies.

The results of comparing images using features extracted from forehead, nose, and cheek regions of eleven portraits at

$b = 8$ and $a = 5$ are obtained and presented in Fig. 4. Histogram of distances $D_{jk}$ between paintings by F. Rokotov is shown by a solid line (denoted as "own"). Dashed line designates histogram of distances between paintings created by different artists (denoted as "alien"). It follows from Fig. 4 that "own" and "alien" paintings represented by feature description used in this work, can be separated. Using the results shown in Fig. 4, a threshold value for attribution decision can be selected.

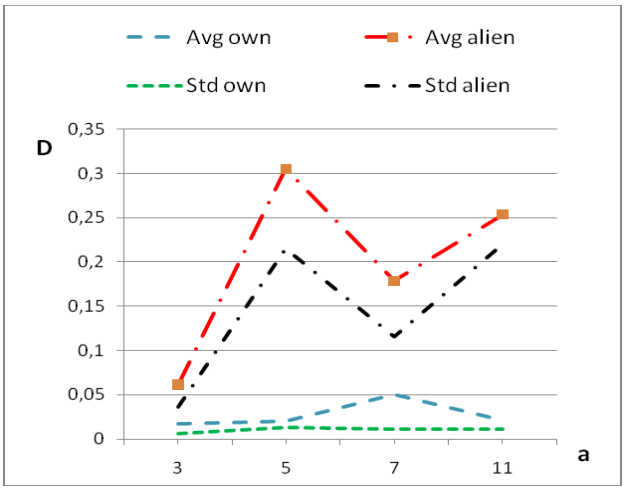



Fig. 3. Averages and standard deviations found from histograms of distances between "own" and "alien" paintings at different values of $a$.

## VII. Conclusion

A feature description of images of paintings based on textural characteristics is proposed. Selected image features in the form of orientation angle distributions give quantitative description of a painter artistic style and provide suitable accuracy of features computation. Feature evaluation does not require segmentation of a single brushstroke and is not sensitive to image acquisition conditions as opposed to conventional techniques. The parameters of feature extraction procedures are chosen. Results of the computing experiments showed the efficiency of the proposed features for comparing artistic styles. The proposed feature set may be used as a part of technological description of fine art paintings for restoration and attribution. The future research will be aimed at the extension of feature space and creating a procedure for making decisions on similarity of brushwork techniques of the researched paintings based on the extended feature space.

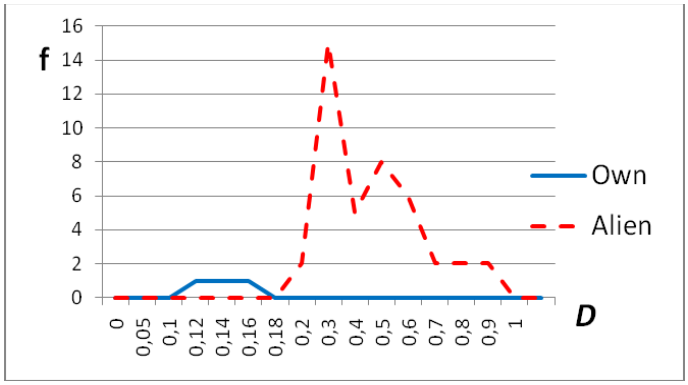



Fig. 4. Histograms of distances between paintings created by the same artist (solid line) and different artists (dashed line); $f$ is a frequency of a distance value $D$.

## References

[1] G. Morelli, Italian Painters. Critical Studies of Their Works. John Murray, London, 1900.

[2] N.S. Ignatova, "Analysis of oil painting textures,". In: Fundamentals of Oil painting Examination. The guidelines. Moscow, Grabar restoration Centre, vol. 1, pp. 12 – 26, 1994 (In Russian).

[3] R. Sablatnig, P. Kammerer, E. Zolda, "Structural Analysis of Paintings Based on Brush Strokes," Proc. of SPIE Scientic Detection of Fakery in Art, SPIE,. vol.. 3315, pp. 87–98, 1998.

[4] C. R.Johnson, E. Hendriks, I.J. Berezhnoy, E. Brevdo, S.M. Hughes, I. Daubechies, J. Li, E. Postma, J.Z. Wang, "Image processing for artist identification (Computerized analysis of Vincent van Gogh's painting brushstrokes)," Signal Processing Magazine, IEEE, vol. 25, No 4, pp. 37-48, 2008.

[5] G. Polatkan, S. Jafarpour, A. Brasoveanu, S. Hughes, I. Daubechies, "Detection of forgery in paintings using supervised learning," ICIP2009, IEEE, pp. 2921-2924, 2009.

[6] J. Li, L. Yao, E. Hendriks, J. Z.Wang, "Rhythmic brushstrokes distinguish van Gogh from his contemporaries: findings via automated brushstroke extraction," IEEE TPAMI, vol. 34, No 6, pp. 1159-1176, 2012.

[7] D. M. Murashov. "Localization of differences between multimodal images on the basis of an information-theoretical measure," Pattern Recognition and Image Analysis. Springer, vol. 24, Issue 1, pp. 133-143, 2014.

[8] D. Murashov, "Composing image feature space for painting attribution tasks," Proc. of the 11th Int. Conf. PRIA-11, Samara, IPSI RAS, vol. 2, pp. 674-677, 2013.

[9] F. Escolano, P. Suau, B. Bonev, Information Theory in Computer Vision and Pattern Recognition, London. Springer Verlag, 2009.

[10] D. Eberly, Ridges in Image and Data Analysis. Klewer Academic Publishers, Dordrecht/Boston, London, 1996.

[11] T. Lindeberg, Scale-space Theory in Computer Vision. The Kluwer International Series in Engineering and Computer Science. Kluwer Academic Publishers, 1994.

[12] B. Jähne. Digital Image Processing. 6th ed., Springer, 2005.

# Adaptivity of Conditional Random Field based Outdoor Point Cloud Classification

Dagmar Lang, Susanne Friedmann and Dietrich Paulus
Active Vision Group, University of Koblenz-Landau
Universitätsstr. 1, 56070 Koblenz, Germany
{dagmarlang, scfriedmann, paulus}@uni-koblenz.de

***Abstract*—In this paper we present how adaptable learned models of graphical models are and how they can be used for classification tasks of 3D laser point clouds with different distributions and density. In order to model the contextual information we use a pairwise conditional random field and an adaptive graph downsampling method based on voxel grids. As feature we apply the rotation invariant histogram-of-oriented-residuals operator to describe the local point cloud distribution. We validate the approach with data collected from different laser range finders with varying point cloud distribution and density. Our experiments imply, that conditional random field models learned from one dataset can be applied to another dataset without a significant loss of precision.**

## I. Introduction

In recent years, object and place classification of 3D point clouds, images, and fused sensor data have gained a lot of attention. Mainly object classification approaches for all three domains have been established to create semantic labels as presented in Section II. Probabilistical graphical models are commonly used to segment and classify indoor and outdoor point clouds or camera data into objects and places with semantic labels. Most approaches classify each 3D point or pixel in the scene with graphical models, some apply data reduction methods and use downsampled point clouds for feature extraction and classification.

With this work, we use a pairwise conditional random field (CRF) approach presented in [5] to classify 3D point clouds into semantic labels without loss of context information by downsampling the CRF graph adaptively and using the rotation invariant local "histogram-of-oriented-residuals" (HOR) operator to characterize the 3D points. In our experimental section, we show that learned CRF models for one dataset, one type of laser scanner, and one type of data acquisition can be applied to other datasets with other point cloud distribution without a significant loss of precision.

This paper is organized as follows. First, we discuss the related work for point cloud classification in Section II. Afterwards, we present in Section III the pairwise multi-label CRF definition and the adaptive graph downsampling method combined with the rotation invariant HOR operator. Experiments and results are depicted in Section IV. We draw our conclusion and point out future work in Section V.

## II. Related Work

We first present different graphical models and the development for classification tasks and explain the differences of the approaches and if used, the data reductions methods. The first approaches are only laser based classification methods. Afterwards we show approaches based on fused camera and laser range data.

Anguelov et al. [1] present one of the first approaches to classify 3D point clouds based on associative Markov networks (AMNs). Triebel et al. [14] propose an extension of this approach, where the authors show that adaptive data reduction not necessarily influences the classification results. In Triebel et al. [16] the authors argue, that the extension of AMNs by a nearest-neighbor classifier can improve the results for 2D and 3D point cloud classification tasks.

Munoz et al. [9] present an directional AMN for 3D point cloud classification and that the directional AMN performs better than the standard AMN formulation or support vector machines. A contextual classification of 3D point clouds or camera data using a linear associative max-margin Markov network approach was introduced by Munoz et al. [8]. Xiong et al. [17] depict a multi-scale inference procedure with a graphical model to capture the contextual relationship among 3D points by training point cloud statistics and to learn relational information over fine and coarse scales for different outdoor scenes.

Lim et al. [6] propose an adaptive data reduction method and use discriminative CRFs for 3D point cloud classification. The authors show that smaller sets of data samples containing relevant information within the support region of super-voxels produce similar results as using the whole point cloud for classification. A classification approach based on pairwise CRFs to segment terrestrial LIDAR point clouds was introduced by Niemeyer et al. [11]. The approach offers the opportunity to incorporate contextual information and learning of specific relations of label classes.

The problem of automatically labeling scenes without prior training and with a model representation that is refined and improved during the classification process for a sequence of 3D outdoor range scans by leveraging an online star cluster algorithm coupled with an incremental belief update in an evolving CRF is addressed by Triebel et al. [15].

Posner et al. [13] introduce a multi-level classification framework for high-order semantic classification of fused camera and 2D laser range data. The classification of spatial and temporal context was modeled by a Markov random field (MRF). A similar problem was addressed by Douillard et al. [2] with the focus on object classification based on 2D laser range and camera data. The authors propose an approach based on CRFs allowing modeling spatial and temporal correlations by extracting visual features from color imagery as well as

shape features from 2D laser scans. Munoz et al. [7] present a multi-modal scene analysis for camera and laser range data when there is no one-to-one correspondence across modalities available. The authors model the camera and 3D point cloud classification by graphical models and propagate information across domains during a co-inference procedure.

## III. Point Cloud Classification

In this section we present the pairwise CRF, an adaptive graph downsampling method and the HOR operator used to classify 3D point clouds into semantic labels presented by the authors in [5].

### A. Pairwise Conditional Random Field

We apply a conditional random field (CRF), which is a discriminative undirected graphical models used for image or point cloud classification modeling the conditional probability $P(\boldsymbol{y}|\boldsymbol{x})$. A set of conditional variables $\boldsymbol{x}$ representing a fully observed data sequence, e. g., the features for every 3D point in a point cloud, should be classified by a set of conditional variables $\boldsymbol{y}$ representing the label sequence, e. g., a finite set of learned classes to a given feature.

Our approach presented below uses a pairwise CRF model based on [10] that is defined as

$$P(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{Z(\boldsymbol{y})} \exp\left(\sum_{i\in\boldsymbol{x}} \boldsymbol{w}_u^T f_i(\boldsymbol{x}) + \sum_{i\in\boldsymbol{x}} \sum_{j\in \mathrm{MB(i)}} \boldsymbol{w}_p^T f_{ij}(\boldsymbol{x})\right), \quad (1)$$

with the partition function $Z(\boldsymbol{y})$. In contrast to all features $f$ of the data sequence $\boldsymbol{x}$, the first sum in the exponential function is called association potential. This potential determines the likeliest label $\boldsymbol{y}_i$ for each node in the graph. The feature vector $f_i(\boldsymbol{x})$ will be calculated for each random variable of $i \in \boldsymbol{x}$ respectively node in the graph. The resulting label sequence is depending on the weight $\boldsymbol{w}_u$ and the feature function $f_i(\boldsymbol{x})$. As smoothing factor based on the neighboring labels the second sum in the exponential function called interaction potential is applied. $f_{ij}(\boldsymbol{x})$ and the label sequence depending parameter vector $\boldsymbol{w}_u$ will be calculated for each random variable $\boldsymbol{x}$ in the graph and its elements $i$. The interaction feature function $f_{ij}(\boldsymbol{x})$ and the corresponding parameter vector $\boldsymbol{w}_p^T$ model the relationship between node $i$ and node $j$ in the graph. The neighborhood of node $\boldsymbol{x}_i$ is defined by the Markov blanket $\mathrm{MB(i)}$. For node labels $\boldsymbol{y}_i$ and $\boldsymbol{y}_j d$ of $\mathrm{MB_i}$, the edge feature vector $\boldsymbol{\mu}_{ij}$ is determined by the difference of the feature vectors $f_i$ and $f_j$ depending on $\boldsymbol{x}$. The interaction feature function is defined as $f_{ij}(\boldsymbol{x}) = \delta_{ij}\boldsymbol{\mu}_{ij}$, where similar labels are preferred by the Kronecker's delta $\delta_{ij}$. We train the CRF using pseudo log-likelihood training with an optimization by the L-BFGS algorithm [12]. For inference, we run loopy belief propagation with residual message update schedule as proposed in [3] until convergence.

### B. Adaptive Graph Downsampling

In literature, point based CRFs, where each 3D point in the cloud is associated with a node in the CRF graph and the Markov blanked is defined by the k-nearest neighbors of the node, yield to good and robust classification results. One drawback is the complexity of the graph structure which leads to expensive and slow computations for large point clouds. In this case, a reduction of the cost was achieved by downsampling the CRF graph respectively the point cloud, which however leads to a loss of information, especially in the presence of small objects.

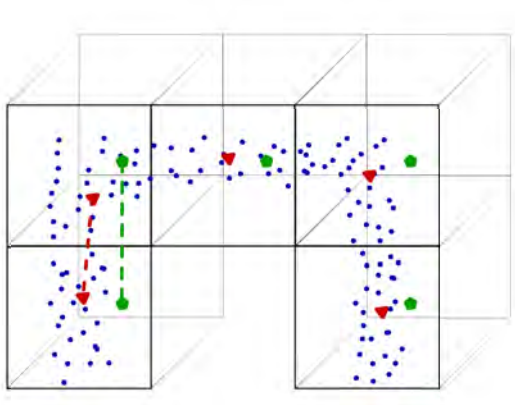


Fig. 1: Example for the adaptive graph downsampling. In blue a part of an example point cloud for a bollard is shown and in black the corresponding voxel grid. The geometric voxel centers of the voxel grid are highlighted in green and the center of mass in red. If the distance (highlighted as dotted line) between the center of mass between both voxels is lower than the distance between the voxel centers both voxels are merged into one. For this example the voxels will be merged.

In order to keep as much information as possible, we downsample the point based CRF graph by using a voxel grid with an adaptive cell size. The basic voxel grid consists of metrically equidistant voxels in each dimension and the nodes are integrated into the voxels. For each voxel, we compute the center of mass for all points in the voxel. The center of mass then becomes a new node in the CRF graph and the other nodes in the voxel are removed.

Since the structure of the voxel grid is fix, we loose a lot of information, if the boundary of the grid passes through small objects. Therefore, we perform a merge step, if the Euclidean distance between neighboring voxel nodes is smaller than the distance between their geometric voxel centers. We recompute a new node and remove the nodes corresponding to the center of mass for the merged voxels. In order to define the Markov blanket each voxel gets connected to its k-nearest neighbor voxel. After the adaptive downsampling, the graph is reduced by about 20 % of its original size.

### C. "Histogram-of-Oriented-Residuals" Operator

As feature we apply the HOR operator introduced by Krückhans [4] and extended by Lang et al. [5]. For each node $\boldsymbol{x}_i$ of the downsampled graph and the corresponding 3D point $\boldsymbol{p}_s$, the descriptor is a local descriptor based on all points $\boldsymbol{p}_{\mathcal{N}_j}$, $j = 1, \ldots, n$, in a neighborhood $\mathcal{N}$, by searching for all neighboring 3D points in a given radius in the point cloud.

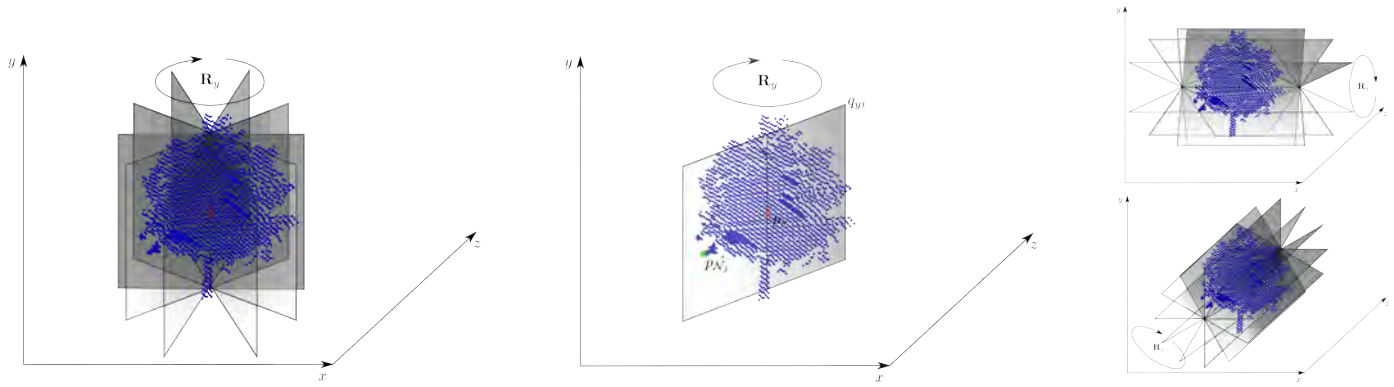


Fig. 2: In the following a sketch of the HOR operator is presented for an example tree point cloud. The schematic overview of the yHOR is shown in the left and middle image. All points of the neighborhood are highlighted blue and the origin of the descriptor $\boldsymbol{p}_s$ red. In the left image the planes defined for the descriptor and the corresponding rotation matrix are shown. Point $\boldsymbol{p}_{\mathcal{N}_j}$ is highlighted in green and one plane is shown in the middle image to demonstrate the calculation of the HOR operator. The upper image of the right row shows the xHOR and the lower image the zHOR.

The search radius is initially fixed, but will be adapted by the adaptive graph downsampling method such that the merged voxel is enclosed by the volume created by the radius.

The HOR operator exploits the difference between points and planes, so called residuals, to characterize local regions in a point cloud. Therefore, $m$ planes $\boldsymbol{q}_{ai} = \boldsymbol{R}_a \left(i\frac{2\pi}{m}\right) \boldsymbol{e}_{a_k}$ are defined by rotation $\boldsymbol{R}_a \left(i\frac{2\pi}{m}\right)$ around axis $a$ with incrementally increased step sizes $i = 0, \ldots, m$ and the unit vector $\boldsymbol{e}_{a_k}$ of a corresponding axis $a_k$. The residual $r_{ai\mathcal{N}_j}$ for axis $a$, step size $i$ and point $\boldsymbol{p}_{\mathcal{N}_j}$ of the neighborhood are calculated as

$$r_{ai\mathcal{N}_j} = \left\langle \begin{pmatrix} \boldsymbol{p}_{\mathcal{N}_j} \\ \|\boldsymbol{q}_{ai}\|^{-1} \end{pmatrix}, \begin{pmatrix} \boldsymbol{q}_{ai} \\ -\langle \boldsymbol{p}_s, \boldsymbol{e}_{a_k} \rangle \end{pmatrix} \right\rangle \quad . \tag{2}$$

The first vector of the scalar product is representing one point of the neighborhood, which is density invariant according to the fourth entry and the second vector is representing the Hesse normal form of the plane. The residuals $r_{ai\mathcal{N}_j}$ are calculated for all planes around axis $a$, number of steps $m$ and points $\boldsymbol{p}_{\mathcal{N}_j}$. The residuals are summarized into histogram $\boldsymbol{h}_a$.

The original descriptor is based on rotation $\boldsymbol{R}_z$ with $\boldsymbol{e}_x$, summarized in $\boldsymbol{h}_z$, and performs well in separating the label *ground* from *building*. Improving its discriminative strength, [5] added two additional rotation axes $\boldsymbol{R}_x$ with $\boldsymbol{e}_y$ and $\boldsymbol{R}_y$ with $\boldsymbol{e}_z$ to the original descriptor, such that $f(\boldsymbol{x}_i) = (\boldsymbol{h}_z, \boldsymbol{h}_y, \boldsymbol{h}_x)$.

## IV. Experiments and Results

In this section we evaluate the presented semantic point cloud segmentation algorithm with two real-world datasets with different point cloud structures and density. We show that with a learned CRF model datasets with different properties can be classified without significant influence on the results.

### A. Datasets

In the following, we present the properties of the datasets used for classification, e. g. the point cloud structure. We classify all datasets with the semantic labels *ground*, *building*, *vegetation*, *column*, *street lamp* and *bollards*. *Columns* can either be *tree trunk* or *pole* and *bollards* do not occur in the Koblenz dataset.

The Freiburg dataset[1] was captured using a wheeled robot equipped with a SICK LMS laser range scanner mounted on a pan-tilt unit and consists of 77 3D scans. Each 360° scan was acquired in a stop-and-go fashion and is composed by three tilted scans. Each 360° scan consists of 150,000-200,000 points.

Additionally, we evaluate our algorithm on the Koblenz dataset. It was also recorded by a wheeled robot in a stop-and-go fashion and consists of 35 3D scans. The robot was equipped with a pan-tilt unit where a Hokuyo UTM-30LX scanner was mounted pointing upwards. The 360° scans were acquired by a 190° rotation and each scan consists of 700,000-750,000 points.

### B. Model and Feature Parameters

The results for all examples were achieved using the presented adaptive graph downsampling method of [5] with an initial voxel size of 1 m × 1 m × 1 m and a Markov blanket with 6 neighbors. The initial search radius for the extended HOR operator calculation for all neighboring 3D points of the voxel center was set to 2.5 m. We calculate $f(\boldsymbol{x}_i)$ with 35 bins, the number of planes is set to $m = 8$ and each $\boldsymbol{h}_a$ consists of 15 bins.

### C. Evaluation Metric

We train the CRF model using small subsets of the point clouds, representing each class. Point clouds not involved in training will be used for evaluation of the performance of the classification. The ground truth was annotated by experts for both datasets. The classification performance for the presented algorithm is summarized in Figure 3 by confusion matrices, example images in Figure 4 and the runtime measurements in Table I.

[1] http://ais.informatik.uni-freiburg.de/projects/datasets/fr360/

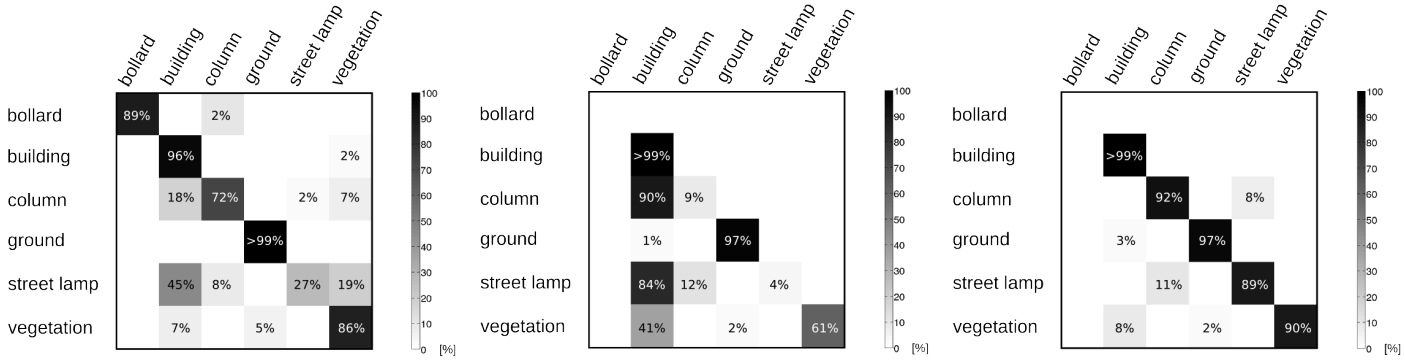



Fig. 3: This figure presents the graphical representation of the confusion matrices for experiment 1 (left), experiment 2 (middle) and experiment 3 (right). In the columns the classification results and the rows the ground truth is plotted. Note that entries on the diagonals represent the precision with which the particular class is classified.

For evaluation, we perform three experiments. In experiment 1 we train the CRF model with subsets of the Freiburg dataset and classify other parts of this dataset. The Freiburg CRF model trained in experiment 1 is used to classify parts of the Koblenz dataset and the results will be presented in experiment 2. Finally, in experiment 3, we train the model with subsets of the Koblenz dataset and classify other parts of this dataset.

### D. Experimental Results

**Experiment 1** In this experiment we compare the model trained with subsets of the Freiburg dataset and classified with other parts of the same dataset. In the upper row of Figure 4 an example for the segmentation result is shown. *Ground* (dark blue) was classified with a precision of >99 %, *building* (light blue) with 96 %, *bollards* (violet) with 89 %, *vegetation* (green) with 86 % and *columns* (red) with 72 %. As shown in the left confusion matrix of Figure 3 *street lamps* (yellow) reached only a precision of 27 % and are often confused with the semantic label *building*. Less frequently misclassifications of columns as *building* or *building* as *street lamp* are present. The approach reaches a overall precision for the Freiburg dataset of 96 %.

**Experiment 2** The adaptivity of CRF model to classify point clouds with different distributions is evaluated in this experiment. Therefore, the CRF model trained for experiment 1 was applied to classify the Koblenz dataset. *Bollards* do not occur in this dataset. The precision of the classification results is shown in the confusion matrix in the middle of Figure 3. *Building* achieved a precision of >99 %, *ground* of 97 % and *vegetation* of 61 % . *Street lamp* was classified only with a precision of 4 % due to *street lamp* was misclassified as *building*, as depicted in the right image of the second row of Figure 4. Furthermore, *column* achieved a precision of 9 % while points annotated in the ground truth as *building* were misclassified *column* as shown in Figure 4 and vice versa. The effect that the misclassification rate increases when the misclassification was high in experiment 1. This is an interesting side effect while the overall precision rate is still by 90 %.

**Experiment 3** We run this experiment to evaluate the precision which can be achieved with a CRF model trained and classified with the Koblenz dataset. As shown in the right confusion matrix of Figure 3 *building* achieved a precision of >99 %, *ground* of 97 % *column* of 92 %, *vegetation* of 90 % and *street lamp* of 89 %. With a frequency <10 % a misclassification of *vegetation* as *building* or *street lamp* as *column* and vise versa occurred. One example result is shown in the bottom row of Figure 4 and achieved a overall precision of 97 % for this dataset.

| Dataset | Process | Mean | Max |
|---|---|---|---|
| Freiburg | HOR calculation | 566 ms | 795 ms |
| | Inference | 804 ms | 841 ms |
| Koblenz | HOR calculation | 2259 ms | 2882 ms |
| | Inference | 883 ms | 1025 ms |

TABLE I: Runtime results for the CRF with adaptive graph downsampling.

**Runtime Experiment** To evaluate the runtime of the semantic classification process we run the classification process for all point clouds of the Freiburg and Koblenz datasets. The results are presented in Table I. In the worst case we needed approximately 4 s to classify a point cloud of ca. 725,000 3D points. This offers the opportunity to incorporate the semantic classification approach into a mobile system for semantic mapping. The HOR classification timing in Table I includes the adaptive downsampling.

## V. Conclusion and Future Work

In this paper we have shown, that learned CRF models can be applied to different datasets with different point cloud distribution for dense 3D laser range data without too much loss of precision. In summary, all classification results presented achieve high precision. The proposed semantic segmentation approach applying the extended HOR operator shows either for large or small detailed objects very good results. New local features based on local point cloud structures and fused camera data should be investigated. For example, a height feature could be ensure, that street lamps are above columns and not vice versa. Fused camera and 3D point clouds could offer the opportunity to divide labels such as ground into new semantic labels such as grass or man-made paths. Especially

the integration of color information and/or texture features should be investigated. Generally, it is very time consuming to annotate data by hand. Thus, it could be examined if incorporating unlabeled data into the training approach can produce considerable improvement in learning accuracy, which could lead to a semi-supervised learning approach.

## VI. ACKNOWLEDGMENTS

This work was partially funded by the Deutsche Forschungsgemeinschaft (DFG) under research contract PA 599/11-1.

## REFERENCES

[1] D. Anguelov, B. Taskar, V. Chatalbashev, D. Koller, D. Gupta, G. Heitz, and A. Y. Ng. Discriminative Learning of Markov Random Fields for Segmentation of 3D Scan Data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 169–176, 2005.

[2] B. Douillard, D. Fox, F. Ramos, and H. Durrant-Whyte. Classification and Semantic Mapping of Urban Environments. *International Journal of Robotics Research*, 30(1):5–32, 2011.

[3] G. Elidan, I. McGraw, and D. Koller. Residual Belief Propagation: Informed Scheduling for Asynchronous Message Passing. In *Proceedings of the Conference on Uncertainty in AI*, pages 165–173, 2006.

[4] M. Krückhans. Ein Detektor für Ornamente auf Gebäudefassaden auf Basis des "histogram-of-oriented-gradients"-Operators. Master's thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, 2010.

[5] D. Lang, S. Friedmann, and D. Paulus. Semantic 3D Octree Maps based on Conditional Random Fields. In *Proceedings the IAPR Conference on Machine Vision Applications*, pages 185–188, 2013.

[6] E. H. Lim and D. Suter. 3D Terrestrial LIDAR Classifications with Super-Voxels an Multi-Scale Conditional Random Fields. *Computer-Aided Design*, 41(10):701–710, 2009.

[7] D. Munoz, J. A. Bagnell, and M. Hebert. Co-Inference for Multi-modal Scene Analysis. In *Proceedings of the European Conference on Computer Vision*, pages 668–681, 2012.

[8] D. Munoz, J. A. Bagnell, N. Vandapel, and M. Hebert. Contextual Classification with Functional Max-Margin Markov Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 975–982, 2009.

[9] D. Munoz, N. Vandapel, and M. Hebert. Directional Associative Markov Network for 3-D Point Cloud Classification. In *Proceedings of the International Symposium on 3D Data Processing, Visualization and Transmission*, 2008.

[10] J. Niemeyer, C. Mallet, F. Rottensteiner, and U. Soergel. Conditional Random Fields for the Classification of LiDAR Point Clouds. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume 38, 2011.

[11] J. Niemeyer, F. Rottensteiner, and U. Soergel. Conditional Random Fields for LIDAR Point Cloud Classification in Complex Urban Areas. In *Proceedings of Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume I-3, pages 263–268, 2012.

[12] N. Okazaki. libLBFGS: A Library of Limited-Memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS), 2010.

[13] I. Posner, M. Cummins, and P. M. Newman. A Generative Framework for Fast Urban Labeling using Spatial and Temporal Context. *Autonomous Robots*, 26(2–3):153–170, 2009.

[14] R. Triebel, K. Kersting, and W. Burgard. Robust 3D Scan Point Classification using Associative Markov Networks. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 2603–2608, 2006.

[15] R. Triebel, P. Rohan, D. Rus, and P. M. Newman. Parsing outdoor scenes from streamed 3d laser data using online clustering and incremental belief updates. In *Proceedings of the Conference on Artificial Intelligence*, 2012.

[16] R. Triebel, R. Schmidt, Ó. Martínez Mozos, and W. Burgard. Instace-based AMN Classification for Improved Object Recognition in 2D and 3D Laser Range Data. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 2225–2230, 2007.

[17] X. Xiong, D. Munoz, J. A. Bagnell, and M. Hebert. 3-D Scene Analysis via Sequenced Predictions over Points and Regions. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 2609–2616, 2011.

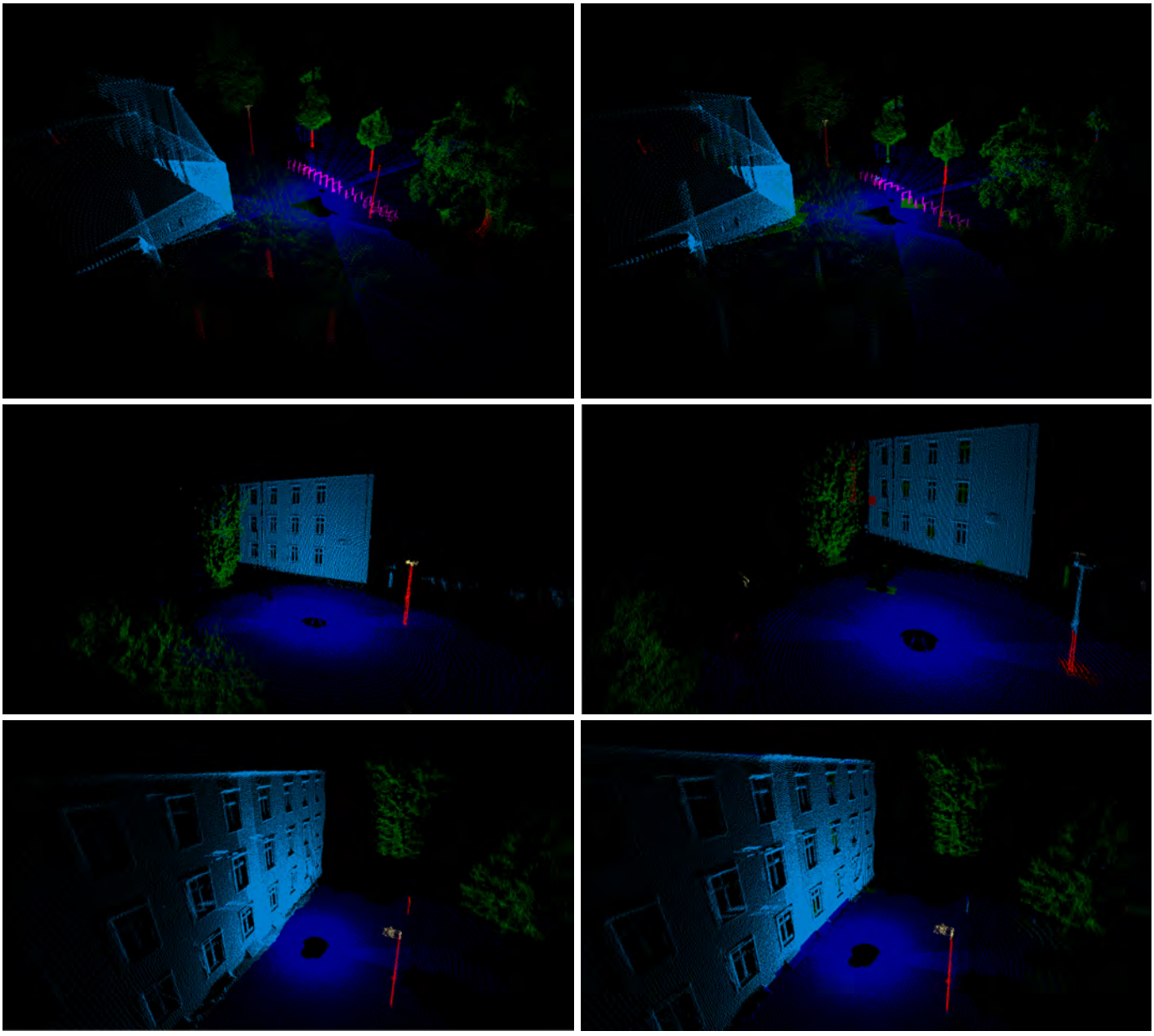


Fig. 4: Classification results for the Freiburg and Koblenz dataset. In the left column the labeled ground truth point cloud and in the right column the corresponding results are presented. One example for the results for experiment 1 is shown in the upper row, for experiment 2 in the middle row and for experiment 3 in the bottom row. The color-coding (best viewed in color) is wrt. to the ground truth (light blue = *building*, dark blue = *ground*, green = *vegetation*, red = *column*, violet = *bollard*, yellow = *street lamp*).

# Advanced 3-D Pose Estimation for Articulated Vehicles

Christian Fuchs, Frank Neuhaus and Dietrich Paulus[1]

*Abstract*— **The knowledge about relative poses within a tractor/trailer combination is a vital prerequisite for kinematic modelling and trajectory estimation. In case of autonomous vehicles or driver assistance systems, for example, the monitoring of an attached passive trailer is crucial for operational safety. We propose a camera-based 3-D pose estimation system based on a Kalman-filter. It is evaluated against previously published methods for the same problem.**

## I. Introduction and Motivation

Transportation scenarios all over the world rely on a wide range of different tractor and trailer combinations, also denoted as articulated vehicles. In complex steering scenarios, such as backwards driving, the control of these vehicles is challenging and is likely to be the cause of dangerous situations which are difficult to handle for both humans as well as autonomous systems. The majority of these problems is caused by passive trailers which do not have direct steering facilities. Subsequently connected passive trailers are being pulled or pushed by the tractor vehicle and change their direction according to the connection and fifth wheel configuration.

The goal of this work is to obtain an estimate of the pose (i. e. the combination of a *3-D* rotation and translation) of a trailer relative to the tractor, so that it could subsequently be used for autonomous driving or human driver assistance. Knowledge about the relative pose between tractor and trailer is the key prerequisite for predicting the correct kinematic behavior. Our work focuses on two-axle trailers (*general-3-trailer*), which exhibit the most difficult kinematic behavior in any two-vehicle setup. Yet our work can also be applied to one-axle trailers.

Examples for applications requiring accurate knowledge about the vehicle state include:

- **Autonomously Operating Vehicles**
  e. g. articulated vehicle control, automated transport or platooning
- **Advanced Driver Assistance Systems for Human Operated Vehicles**
  e. g. backing-up assistances and stability/slipping monitoring
- **Safety Systems**
  e. g. swaying detection and trailer monitoring (at high velocities)

We propose an advanced sensor system providing *3-D* pose information between tractor and trailer using a tracking mechanism. We use an optical sensor setup in order to reconstruct the state information. Active components only need to be installed on the tractor vehicle, making the system relatively inexpensive to use in practice. Our system represents an improvement on two previously published sensor systems for the same purpose. We evaluate our sensor system in a virtual test environment and compare the results against those obtained by previously published reconstruction methods.

[1]Active Vision Group, Institute for Computational Visualistics, Faculty of Computer Science, University of Koblenz-Landau, 56070 Koblenz, Germany {fuchsc, fneuhaus, paulus}@uni-koblenz.de

## II. State of the Art

A system for estimating the relative pose of a trailer comprises a number of components. We consider the state of the art for each of these components individually.

### A. Calibration

An optical approach implies the use of cameras as sensor devices. In order to deduce measurements from camera images, appropriate camera calibration techniques are necessary. The method developed by Tsai and Lenz [Tsa87, LT88] is commonly used to extract intrinsic parameters and distortion coefficients.

Cameras with wide opening angles (e. g. fish-eyes, omnidirectional cameras) cause heavy distortions, especially at the edges of the image frames, and cause problems with the original approach by Tsai and Lenz. The method/toolbox introduced by Scaramuzza focuses on this additional challenge and may be used instead in this scenario [Sca07].

Artificial markers are frequently used in the augmented reality (AR) or the computer vision communities. Therefore a large number of different libraries are freely available [KB99, SFH+02, Fia05, Ols11]. The library by Olson [Ols11] is currently used in the proposed system. A setup without artificial markers is generally possible, but not addressed in this publication.

### B. Tracking

Since the goal of the work is to provide a running estimate of the trailer's pose, it is generally a good idea to consider sequences of images instead of individual images. The Kalman-filter [Kal60] defines a probabilistically sound way to fuse measurements, obtained in the individual frames, with an appropriate motion model for the trailer. It is frequently used in tracking applications in computer vision.

### C. Trailer Pose Estimation

Balcerak, Zöbel and Weidenfeller propose a method to estimate the trailer pose, making the assumption that the tractor/trailer combination to move in a *2-D* plane [BZW06]. The method was later evaluated by Fuchs et al. in [FEKZ14].

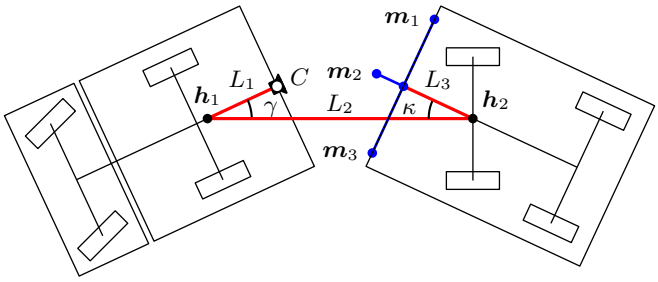



Fig. 1. Sensor setup on a tractor/trailer combination: $\boldsymbol{m}_1$, $\boldsymbol{m}_2$ and $\boldsymbol{m}_3$ define the markers in a passive marker system, $C$ denotes a camera facing backwards

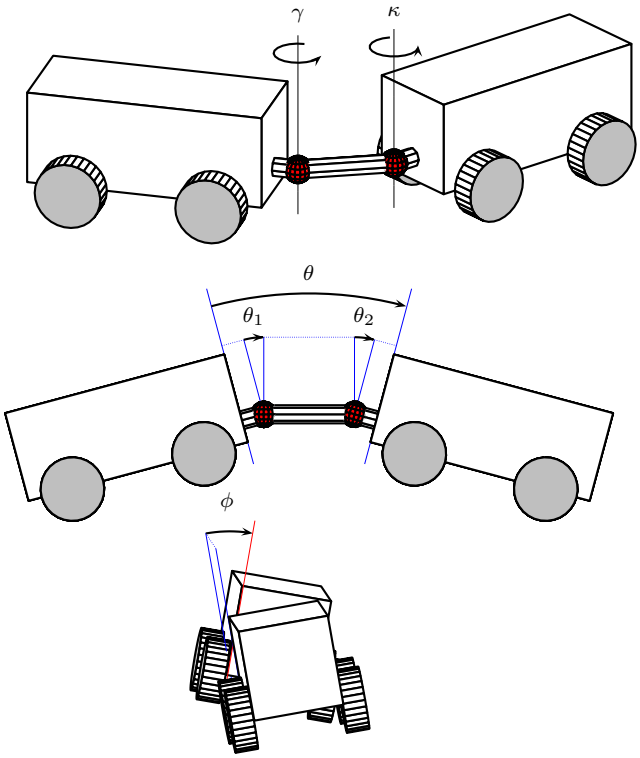



Fig. 2. (Simplified) *3-D* transformation model with angles $\gamma$ and $\kappa$ (yaw), $\theta$ (pitch), $\phi$ (roll)

The trailer's pose (*2-D*) is described using two yaw angles $\gamma$ and $\kappa$, which are extracted using a backwards-facing camera mounted on the truck which observes a set of markers, which was previously installed on the trailer [fig. 1].

The fundamental issue is that torsion in either roll or pitch direction, which is frequent especially on rough terrain, leads to errors as the underlying assumptions are violated.

To overcome these issues, Fuchs, Zöbel and Paulus extended the approach to *3-D* in a very similar setup. Each joint ($\boldsymbol{h}_1$ and $\boldsymbol{h}_2$, fig. 1) has three degrees of freedom [FZP14]. The reasonable assumption that $\theta_1 = \theta_2$ (pitch) and $\theta = \theta_1 + \theta_2$ is made [fig. 2]. The same can be done for $\phi$ (roll). This effectively reduces the degrees of freedom from six to four and leads to a considerable simplification of the computation without loss of accuracy. Figure 2 illustrates the degrees of freedom and variables used.

The authors use projective geometry to match detected markers [Ols11] in the camera images with an underlying *3-D* model of the marker pattern and the geometric tractor/trailer setup. After an initial camera calibration [Tsa87, LT88], knowledge about intrinsic and lens distortion is used to map features found in the camera image (markers) to the *3-D* model. The mapping is done by formulating and solving a joint optimization problem using all detected marker positions, thus minimizing the error of reprojection [Mar63, Lev44].

The perspective projection in case of a marker observation is depicted in fig. 3. As a result of the reprojection, the joint transformation (rotation and translation) is decomposed in order to obtain the desired angles [FZP14].

*Full "State of the Art" will be in the full paper.*

## III. System Setup

The setup in this work is identical to the setup used by Fuchs, Zöbel and Paulus in [FZP14]. The hardware of the sensor consists of two components: A *sensor component (active)* mounted at the tractor and a *passive marker component* mounted at the trailer. This has obvious advantages: Trailers do not have to be equipped with cost intense hardware in order to enable pose detection. Only the tractors, usually equipped with motors and/or electric energy supplies, need to carry digital hardware. The geometric setup of the hardware installation is shown in fig. 1. The markers are based on a system developed by Olson [Ols11].

The software extracts the artificial markers from each frame of the camera image and subsequently attempts to estimate the trailer's pose from the resulting corner-observations of each marker. The final step of the algorithm is to decompose the obtained pose into the relevant angles of the tractor/trailer combination.

## IV. Kalman-tracking of Markers

The naive way to reconstruct the trailer pose relative to the truck would be to perform a direct estimation of the pose in each frame using an approach described by Hartley and Zisserman [HZ03]. Since an initial guess for the pose is known from the previous frame however, a tracking mechanism can be used to enhance the estimation accuracy of the system.

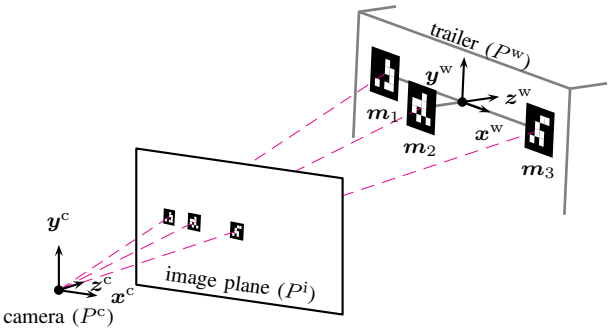



Fig. 3. Projection from the marker-pattern mounted on a trailer to the image plane


Fig. 4. Miniature truck model, scale 1 : 16

We use a Kalman-filter in order to fuse a constant-velocity motion model of the trailer with the measurements extracted from the current image frame. The trailer's state is represented as the pose of the trailer along with it's linear and angular velocity.

The Kalman-filter's measurement model predicts the current marker positions using an appropriate camera model. In the subsequent Kalman update step, the predictions are compared to the actual measurements, and the innovation is fused into the current estimate of the trailer's state.

Tracking the markers with the Kalman-filter allows to observe and incorporate an arbitrary amount of markers whose precise pose at the rear of the trailer is known. The approach naturally handles partial or full occlusions and detection failures and can run at high framerates.

## V. Virtual Test Environment

The evaluation of a sensor system is a crucial issue when no reference system is available. Therefore, a virtual test environment has been developed in by Fuchs et al. [FEKZ14] and extended by Fuchs, Zöbel and Paulus [FZP14]. A simulation unit virtualizes the geometry and the camera and provides synthetically rendered image data for further software-in-the-loop processing.

The approach enables effective regression testing and optimization of the algorithms and allows to derive precise accuracy results relative to a known ground-truth. The *3-D* renderer is capable of simulating different parameter settings and mounting positions for both markers and camera.

## VI. Evaluation

We presented an advanced *3-D* sensor system that is able to accurately reconstruct a trailer's pose relative to a tractor vehicle from video data. The approach is evaluated in different geometric setups and camera configurations. We compare the obtained accuracy with the results of Fuchs et al. and the results of a different approach from Fuchs, Zöbel and Paulus [FEKZ14, FZP14].

The configuration set used for evaluation shown in this publication is deduced from a model scaled truck which is available in our laboratory (values in cm, value of $\alpha$ in Deg):

| $L_1$ | $L_2$ | $L_3$ | $p_w$ | $p_h$ | $w_m$ | $\alpha$ |
|---|---|---|---|---|---|---|
| $-18.4$ | 16.4 | 3.5 | 12.5 | 4.3 | 2 | 67° |

The virtual test environment renders an artificial video stream, that is comparable to a real video image. For each image frame, the ground-truth angles $\gamma$, $\kappa$, $\theta$ and $\phi$ (simulation) are logged together with the corresponding reconstructed angles $\gamma'$, $\kappa'$, $\theta'$ and $\phi'$ obtained with our methods as well as with the methods we compare against.

*The final detailed evaluation will be available in the full paper. Evaluation will include the following steps:*

- *Configuration of the virtual test environment*
- *Evaluation using the plane constraint with comparison of the mentioned methods*
- *Evaluation without the plane constraint with different roll/pitch settings with comparison of all methods*

## References

[BZW06] Elisabeth Balcerak, Dieter Zöbel, and Thorsten Weidenfeller. *Patent DE 10 2006 056 408 A1*. Deutsches Patent- und Markenamt, 11 2006.

[FEKZ14] Christian Fuchs, Simon Eggert, Benjamin Knopp, and Dieter Zöbel. Pose detection in truck and trailer combinations for advanced driver assistance systems. In *13th International Conference on Intelligent Autonomous Systems*, 2014.

[Fia05] Mark Fiala. Artag, a fiducial marker system using digital techniques. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 590–596. IEEE, 2005.

[FZP14] Christian Fuchs, Dieter Zöbel, and Dietrich Paulus. 3-d pose detection for articulated vehicles. In *2014 IEEE Intelligent Vehicles Symposium*. IEEE, 2014.

[HZ03] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[Kal60] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82(1):35–45, 1960.

[KB99] Hirokazu Kato and Mark Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Augmented Reality, 1999.(IWAR'99) Proceedings. 2nd IEEE and ACM International Workshop on*, pages 85–94. IEEE, 1999.

[Lev44] Kenneth Levenberg. A method for the solution of certain problems in least squares. *Quarterly of applied mathematics*, 2:164–168, 1944.

[LT88] Reimar K Lenz and Roger Y Tsai. Techniques for calibration of the scale factor and image center for high accuracy 3-d machine vision metrology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(5), 1988.

[Mar63] Donald W Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics*, 11(2):431–441, 1963.

[Ols11] Edwin Olson. AprilTag: A robust and flexible visual fiducial system. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 3400–3407. IEEE, May 2011.

[Sca07] Davide Scaramuzza. *Omnidirectional vision: from calibration to robot motion estimation*. PhD thesis, Citeseer, 2007.

[SFH+02] Dieter Schmalstieg, Anton Fuhrmann, Gerd Hesina, Zsolt Szalavári, L Miguel Encarnaçao, Michael Gervautz, and Werner Purgathofer. The studierstube augmented reality project. *Presence: Teleoperators and Virtual Environments*, 11(1):33–54, 2002.

[Tsa87] Roger Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, 1987.

# An Automatic Initialization of Interactive Segmentation Methods Using Shortest Path Basins

Tomas Ryba
The University of West Bohemia
Pilsen, Czech Republic
Email: tryba@kky.zcu.cz

Milos Zelezny
The University of West Bohemia
Pilsen, Czech Republic
Email: zelezny@kky.zcu.cz

***Abstract*—Image segmentation is one of many fundamental problems in computer vision. The need to divide an image to a number of classes is often a part of a system that uses image processing methods. Therefore, lots of methods were developed that are based on different approaches. The image segmentation could be classified with respect to many criteria. One such a criterion is based on the degree of allowed interactivity. The interactivity could be of several types - interactive initialization, interaction while the computation is running or manual refinement of achieved result, for example. Especially the precise initialization plays an important role in many methods. Therefore the possibility to initialize the method manually is often invaluable advantage and information obtained this way could be the difference between good and poor results. Unfortunately, in many cases it is not possible to initialize a method manually and the process needs to be automated. In this paper, an approach for such an automation is presented. It is based on shortest paths in a graph and deriving an area of influence for each obtained seed point. This method is called shortest path basins.**

## I. Introduction

The partioning of the image to several relevant classes, i.e. image segmentation, is one of the common image processing task. Thanks to all the research that has been done in this area lot of techniques for solving this task were developed. Based on the amount of interactivity these methods could be classified as interactive or autonomous. The information provided by the user can make all the difference between good and poor segmentation results. On the other hand, inexpert interactivity can confuse the algorithm, which can further lead to its collapse.

The critical part of many algorithms is their initialization. Inaccurate initialization can lead the algorithm to stuck in a local extrema of solution space that can be far away from the optimal solution. Therefore, the possibility to initialize the algorithm manually is very valuable. Due to the user intervention the manual initialization can be viewed as a type of interactivity. It is often done by marking several image points for each image class. These poins are called seed points and can be used for example as starting points for further propagation of the algorithm in the image domain or for creating intensity models. The user represents here an expert and in most cases has a critical influence on the results.

Unfortunately, the manual initialization can not be done in many cases and therefore needs to be automated. In this paper an approach for such an autonomous initialization is presented. This method iteratively looks for relevant seed points satisfying an energy function.

This paper is organized as follows. Section II describes several image segmentation algorithms with different amount of user interactivity. Two methods that are initialized using presented algorithm are detailed in Section III and the algorithm itself is described in Section IV. Section V starts with introduction of the evaluation technique and continues with performed experiments. The whole paper concludes in the last Section VI, where possible improvement of the algorithm is outlined.

## II. Related Work

In present time lot of interactive segmentation methods exists that differ in amount and type of used interactivity. The Magical rope algorithm is popular especially in software for photo editing. The user has to trace the boundary of an object and the algorithm refines the segmentation by snapping to edges. A representative of this class of methods is The Lazzy snaping algorithm [1], for example. The drawback of this approach is relatively higher demands on amount of the interactivity that is needed for obtaining good results.

A popular fully autonomous segmentation method is Fuzzy C-means [2], for example. The core of this method is deriving the so called partition matrix $W = w_{i,j} \in [0,1],\ i = 1,\ldots,n, j = 1,\ldots,c,$ , where $n$ determines the number of image pixels and $c$ the total number of classes. The value of matrix element $w_{i,j}$ tells the degree to which element $x_i$ belongs to cluster $c_j$. The goal of the method is to segment the image in a way that minimizes an objective function.

Another well known autonomous methods are the Felzenszwalb algorithm [3] and the Normalized cuts [4]. These methods are often used as standards for comparing to another segmentation methods and this paper is no exception.

In [5] an approach is presented, where the initialization is based on histogram analysis. The main idea there is an assumption that individual classes are homogeneous and therefore they should correspond to peaks in image intensity histogram. Hence the peaks are localized and used for generating seeds of individual classes. A region growing method called Grow cut [6] is then initialized by these seeds.

From our point of view the manual initialization plays important role. The user simply marks few points that belongs to individual classes. These points can then be used for estimating intensity models for each class or as starting points for further propagation of the algorithm. Typical representatives of this kind of methods are Graph cut on MRF [7] and

(a) (b) (c) (d) (e)

Fig. 1. Image segmentation of input image (a) to four classes using the Random Walker algorithm. Figure (a) shows marked seed points and outlined resulting segments. Figure (b) to (d) shows the probability of assigning a pixel to the first, second, third and fourth class, respectively.

Random walker [8], for example. These methods are used to demonstrate presented algorithm and therefore they are covered in following section.

In [9] a method is presented that reduces the amount of interactivity used in the initialization step. The method uses pseudolikeliohood algorithm which jointly learns the color mixture and coherence parameters for foreground and background respectively. Unfortunately, using the pseudolikelihood learnig a limitation of usable models arises. Moreover, we are not trying to reduce the amount of interactivity but to eliminate it completely.

Another promising approach is described in [10] that is based on adaptive evolutionary algorithm and immune system dynamics.

## III. Used segmentation methods

There are many interactive segmentation methods but maybe the most popular are the Graph cuts on markov random fields and the Random walker algorithm. The initialization of this method is often done by specifying few seed points for each image class. This initialization approach is replaced by presented algorithm and both methods are then compared with another two fully autonomous methods in Section V.

### A. Graph Cuts on Markov Random Fields

A theory that tries to understand the laws of human ability of understanding visual scenes and recognizing individual objects is called the gestalt theory and is described in [11]. This theory shows that both boundary and region information should be used in computer vision system to acquire precise results. One way how to be consistent with the Gestalt theory is by using the contextual information.

The use of contextual information is ultimately necessary for proper image understanding as shown in [12]. The first use of contextual information for solving an image analysis problem is published in [13]. Markov random fields (MRF) form a branch of probability theory and provide tools suitable for the characterization of the contextual constraints.

To define a MRF it is common to use the graph theory. The MRF can be defined as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, 2, \ldots, N\}$ denotes the set of nodes. Each node is associated with a random variable $F_p$ for $p = 1, 2, \ldots, N$, each random variable can take a value $f_p \in \mathcal{L}$, where $\mathcal{L}$ is the set of all possible labels. To define the connectivity of the graph either the set of edges $\mathcal{E}$ or the neighborhood system $\mathcal{N}$ must be known. The set of nodes to which the node $p$ is adjacent defines its neighborhood $\mathcal{N}_p$, i.e. $q \in \mathcal{N}_p \iff (p, q) \in \mathcal{E}$. The set $\mathcal{N}_p$ is called the Markov blanket of node $p$.

A realization of the MRF $f$, i.e. an event where each node takes a value $f = \{f_1, f_2, \ldots, f_m\}$, is called a configuration of that field. The probability that a random variable $F_p$ takes the value $f_p$ is denoted $P(F_p = f_p) = P(f_p)$ and the probability of a configuration $f$ is denoted $P(F = f) = P(f)$.

An important property of every MRF is the Markovianity:

$$P(f_p|\{f_q\}_{j\in\mathcal{V}\setminus i}) = P(f_p|q \in \mathcal{N}_p) \tag{1}$$

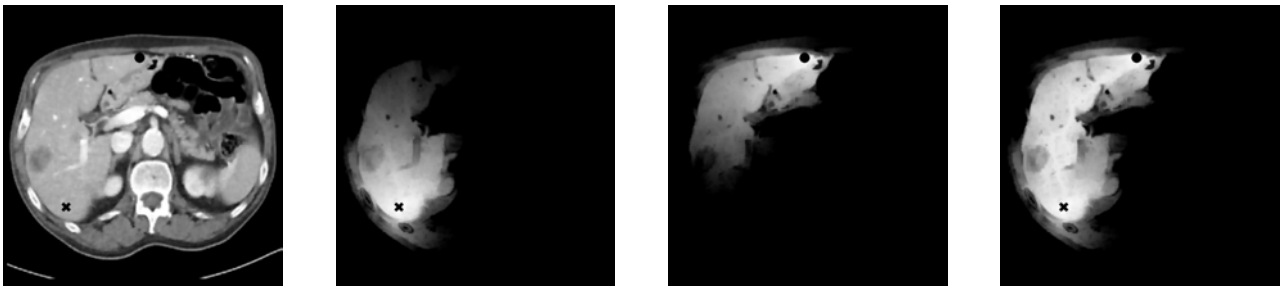
This feature describes a so called knock-on effect where explicit short-range linkages give rise to implied long-range correlations ([14]). It is the great attraction for using MRFs and incorporates the contextual information that is very valuable for proper image understanding.

To specify a MRF the joint probability $P(f)$ should be determined. Thanks to the Hammersley-Clifford theorem described in [15] the Gibbs distribution can be used for calculating this probability. The advantage of using this distribution is that the joint probability can then be specified on clique potentials. The maximization of the joint probability equals the minimization of energy $E(f)$ that can be rewritten in a form that couples cliques based on their order. In most applications only the first $V_1(f_p)$ and the second $V_2(f_p, f_q)$ order of clique potentials are taken into account. The energy then takes the well known form:

$$\begin{aligned} E(f) &= \sum_{\{p\}\in\mathcal{C}_1} V_1(f_p) + \sum_{\{p,q\}\in\mathcal{C}_2} V_2(f_p, f_q) = \\ &= E_{\text{data}}(f) + E_{\text{smoothness}}(f) \end{aligned} \tag{2}$$

where $\mathcal{C}_i$ is the set of all cliques of the order $i$ in a graph. The first term $E_{\text{data}}$ in equation 2 is often called simply the data term and evaluates the fit of the labeling to the observation. The second term $E_{\text{smoothness}}$ is often called simply the smoothness term and encourages homogeneous regions.

For solving the MRF, i.e. finding the most probable configuration, an optimization technique must be used. While it is possible to use any kind of technique, e.g. the simulated annealing, a genetic algorithm etc., in our work the graph cut method with large moves was used. This method was



(a) (b) (c) (d)

Fig. 2. In the input image (a) two seeds (cross and circle) are chosen. In figures (b) and (c) local energy of given seeds are shown, fig. (d) shows accumulative penalizing energy.

introduced in [7] and become one of the most used optimization technique for solving a MRF. The authors also discussed conditions for achieving a solution within a known factor from the global optimum.

### B. Random Walker

A method based on random walks is another popular representative of methods with the possibility of interactive initialization. This method was introduced in [8].

The input image is represented as a graph. In each node of that graph a random walker is placed that randomly goes through the graph. When this walker steps on one of the seeds it propagates no further. The starting point is then assigned to the class represented by that seed point. In more mathematical way, for each unlabeled node $n$ and seed $s$ the algorithm derives a probability $p(n, s)$ that a random walk starting at $n$ ends in $s$.

However, starting a random walk in each node is very inefficient. Fortunately, in [16] and [17] is shown that the above mentioned probability $p(n, s)$ is equal to the solution of Dirichlet problem with boundary conditions placed in seed points. More information about this calculus can be found for example in [18] and [19].

The Dirichlet problem is a problem of finding a harmonic function subject to its boundary values. Such a harmonic function minimizes the Dirichlet integral for a field $u$ and a region $\Omega$:

$$D[u] = \frac{1}{2} \int_{\Omega} |\nabla u|^2 d\Omega. \tag{3}$$

In case of $K$ different labels, the solution of a combinatorial formulation of the Dirichlet problem is given by solving only $K - 1$ sparse linear systems. More information about the calculus is given for example in [8].

Moreover, if the image is represented as a graph then this algorithm has an analogy with propagation of the electric potential. The probability $p(n, s)$ is given as the solution of a problem from circuit theory that is the combinatorial version of Dirichlet problem [17]. Firstly, the potential of every seed that did not belong to the class $s$ is setted to zero and the potential of the seeds representing class $s$ is setted to one. The steady potential of each unlabeled node then corresponds to the probability $p(n, s)$. Example of image segmentation using the Random walker algorithm is shown in fig. 1.

When all probabilities are derived, a node is assigned to the class to which it has the highest probability. Deriving all probabilities allows us to identify another classes to which the node is 'near'. This can be used for example in subsequent processing steps.

## IV. Shortest path basins

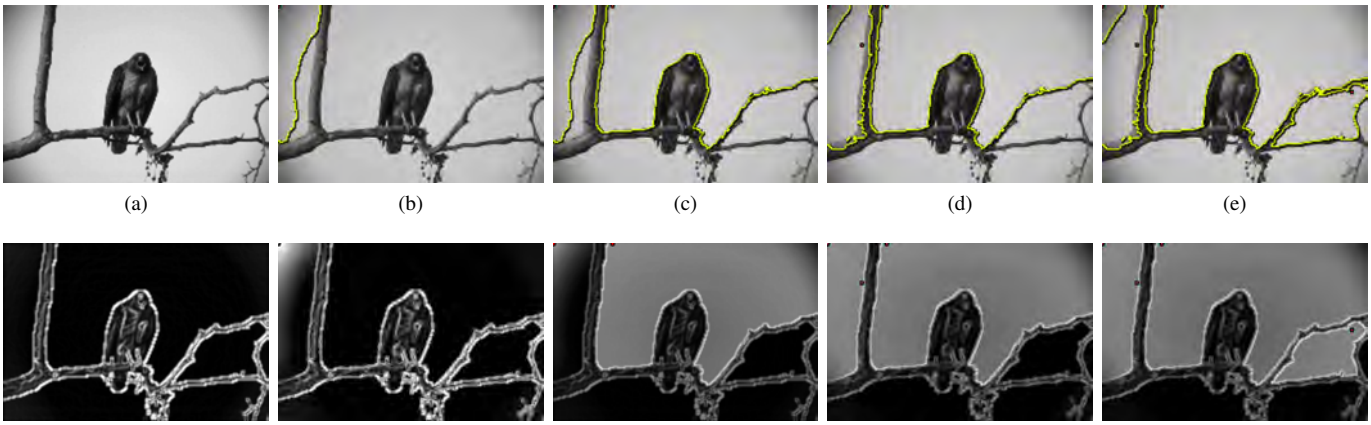
Motivation for creation of this algorithm is to develop an alternative approach for deriving seed points, which is fully autonomous. The method was introduced in [20] as an autonomous algorithm for image segmentation. In this paper a possibility of using it for initialization is presented. The method is described in more details in following subsection.

### A. Algorithm overview

The principle of the algorithm is iterative seeking for seed points that represents individual image classes the best way. To be such an algorithm effective it is necessary to define a heuristic function that will control the selection of the seed points. The heuristic used in this paper is based on finding seed points that are maximally different. In other words, each new seed should be as different from the previously chosen seeds as possible. That way the corresponding classes should be different as well.

When new seed point is derived it is appropriate to define image points that are close to the seed. This area is then penalized so that it is less possible that new seed point will be chosen from this area and thus making him similar to an already created seed point. The area that contains points that are similar to a seed point is called the basin. Certainly, there are many approaches for defining such a basin. We use the shortest path in a graph therefore the basin is called shortest path basin (SPB).

The core for creating a SPB is the inverse Dijkstra's algorithm. It propagates from a starting point (source) to its neighborhood and for each point derives a value that reflects the closeness of that point to the source. This value is determined as the cost of the shortest path from the source to this point. The algorithm terminates when there are no points that are closer than a predefined maximal distance threshold $T$. In other words, all points that are closer from the source then the threshold $T$ create the SPB of the source. By using the algorithm for computing the shortest path in a graph it is possible to define the proximity of points that depends not



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Fig. 3. Example of deriving seed points in input image (a). Figures (b) - (e) show derived seed points and their basins. Figure (f) shows initial overall energy $E_o(t)$ and figure (g)-(j) shows first four iterations.

only on intensity difference but on geometric distance as well. This is achieved by defining graph edges in a common way:

$$w_{i,j} = 1/\exp\left(-\frac{(I_i - I_j)^2}{2\sigma^2}\right), \tag{4}$$

where $I_j$ is the intensity of a pixel $i$ in an image $I$ and $\sigma$ could be interpreted as expected intensity difference inside the segmented object [21].

If there are two points with similar intensity that are geometrically far away, there is noticeable chance that these points will be classified to different classes - the geometrical distance outweighs the similarity of intensities. To overcome this problem a system of siblings is used. Each point becomes a sibling of the source if it lies in a basin and has the same (or very similar) intensity as the source. Each sibling is then taken as another source of the same class and the algorithm can propagate to farther image areas.

Possible alternative to creating basins using shortest path is to use Euclidean distance, which will be much more efficient. Using this approach yields to suppression of the intensity values and the basins would be spherical. The main drawback is that points in such a basin could easily belong to different classes, i.e. they could be of very different intensity.

### B. Penalizing energy

The above mentioned heuristic is based on iterative updating of a so called overall energy $E_o(t)$, where $t$ denotes the iteration index. This energy reflects penalization of each seed point that was derived so far. In each iteration a new seed $s$ is derived and the corresponding SPB needs to be penalized. The energy of the basin of the seed $s$ is denoted $E(s)$ and is defined by the costs of shortest path of each point in the basin:

$$E(s) = T - dist(p, s),\ \forall p \in \text{Im} \tag{5}$$

where $T$ represents distance threshold for algorithm propagation and $dist(p, s)$ is the distance of an image point $p$ to a seed point $s$ in the sense of shortest path.

A problem that can easily arise is that a new seed point will be derived near an edge in the image. To overcome this problem the energy can be calculated as a weighted mean of the energy and image gradient at the same point:

$$E_f(s) = \alpha \cdot E(s) + (1 - \alpha) \cdot \nabla \text{Im} \tag{6}$$

Once a seed point is chosen its energy $E_f(p)$ is added to the overall (accumulative) energy:

$$E_o(t) = E_o(t-1) + E_f(s) \tag{7}$$

where $t$ defines current iteration.

The next seed point will be chosen as a point with the smallest value of this accumulative energy:

$$s(t+1) = \arg\min_{\forall p \in \text{Im}} (E_o(t)) \tag{8}$$

Process of updating the energy $E_O(t)$ is shown in figure 2. The algorithm terminates after a certain number of iterations or if the minimum of the accumulative energy exceeds the predefined threshold value. The process of iteratively deriving seed points is shown in 3.

## V. Experiments

It is generally known that the image segmentation is ill-defined process. There is not only one ground truth in most cases, which makes the objective comparison of segmentation methods a challenging task. Due to this fact it is reasonable to compare tested method with several manual segmentations from different users. Fortunately, the Probabilistic rand index (PRI) is a measure for this kind of evaluation and was introduced in [22].

The main principle of this approach lies in analysing the labels of each pixel pair. Consider an input image of $N$ pixels $X = x_1, x_2, \ldots x_i, \ldots, x_N$ and a set of manually segmented (ground truth) images $S_1, S_2, \ldots, S_K$. Let $S^t$ be the segmentation that is to be compared and let a the label of a point $x_i$ be denoted as $l_i$ in the segmentation $S^t$ and $l_i^{(k)}$ in the manual segmentation $S_k$. To be able to objectively compare a segmentation $S$ with several manual segmentations, it is necessary to determine expected interrelation $P(l_i = l_j)$

TABLE I. COMPARISON OF METHODS THAT IS BASED ON PROBABILISTIC RAND INDEX (PRI). THE MEAN AND STD VALUES ARE CALCULATED OVER 50 DIFFERENT IMAGES.

| | GC-SPB | RW-SPB | FHC | NC |
|---|---|---|---|---|
| 'man' | 0.832 | 0.849 | 0.831 | 0.867 |
| 'bird' | 0.834 | 0.835 | 0.882 | 0.592 |
| 'woman' | 0.697 | 0.745 | 0.749 | 0.737 |
| 'boy' | 0.667 | 0.619 | 0.652 | 0.661 |
| mean | 0.733 | 0.709 | 0.662 | 0.718 |
| std | 0.127 | 0.141 | 0.249 | 0.144 |

and $P(l_i = l_j)$ for each pixel pair $x_i$ and $x_j$ following the next equation:

$$P(l_i = l_j) = \frac{1}{K} \sum_{k}^{K} \mathbb{I}\left(l_i^{(k)} = l_j^{(k)}\right) \tag{9}$$

$$P(l_i \neq l_j) = \frac{1}{K} \sum_{k}^{K} \mathbb{I}\left(l_i^{(k)} \neq l_j^{(k)}\right) \tag{10}$$

The equation 9 represents an expectation that both pixels will have the same label. The equation 10 represents an expectation that the pixels will have different labels. These terms reflects the interrelation of pixel pairs through the whole set of manual segmentations.

The PRI is then defined as follows:

$$PRI(S, \{S_k\}) = \frac{1}{\binom{N}{2}} \sum_{\substack{i,j \\ i \neq j}} [\mathbb{I}(l_i = l_j)\, P(l_i = l_j) + \mathbb{I}(l_i \neq l_j)\, P(l_i \neq l_j)] \tag{11}$$

The PRI reflects how many pair pixels have the same interrelation (same or different labels). This value is also weighted by the expected values given by equations 9 and 10. The PRI measure takes values in the interval $[0, 1]$, where 0 means that tested and manual segmentations have no similarities and 1 means that all segmentations are identical.

The accuracy of presented method was tested on 50 images that were taken from the well known Berkeley Segmentation Dataset and Benchmark described in [23]. As the tested segmentation methods we use the GC and RW that were initialized with the SPB algorithm. These algorithms were named GC-SPB and RW-SPB. Furthermore, another two methods were included in the final comparison, the Felzenszwalb's algorithm [3] (FHC) and the Normalized cuts (NC) [4]. The set of ground truth segmentations $\{S_k\}$ was formed by five manual segmentations from above mentioned benchmark. The resulting PRI on several concrete images as well as the mean and std values are listed in table I. These results shows that the use of SPB initialization provides meaningful segmentations that is more then comparable with popular autonomous methods. The resulting image segmentation of compared algorithms are shown in 4, 5, 6.

## VI. CONCLUSION

The possibility of interactive initialization by making seed points in an image provides very valuable advantage in solving the image segmentation task. Unfortunately, not always is such an initialization possible and thus needs to be automated. In this work an approach for such an initialization is presented. This algorithm automatically places seed points in the image and determines their impact area - so called shortest path basin (SPB). To place the seed points in a meaningful way an accumulative energy is calculated. When a new seed is derived this energy is updated to penalize the SPB of this seed. The seed point in next iteration is chosen as a point with smallest accumulative energy.

Because the image segmentation is ill-defined problem one can hardly find a single ground truth. Therefore it is reasonable to compare tested segmentation with several manual segmentations. In this work we use the evaluating measure called Probabilistic rand index, which provides a meaningful way for such a comparison.

To compare our algorithm we use it for initialization of the Graph cut (GC-SPB) and the Random walker (RW-SPB). These methods were then compared with two fully autonomous algorithms, the Felzenszwalb's algorithm (FHC) and the Normalized cuts (NC). The experiments show that the GC-SPB method gives the best results among compared algorithms followed by NC, RW-SPB and FHC.

A big disadvantage of our algorithm is the time complexity. While the NC and the FHC take less than a second our algorithm needed tens of seconds. To overcome this problem it is possible to reduce the spatial resolution using superpixels. On the other hand, using such an approach could yield to degenerating the image gradient, which is of great importance in our approach.

## ACKNOWLEDGMENT

The work has been supported by the grant of The University of West Bohemia, project No. SGS-2013-032 and by the European Regional Development Fund (ERDF), project New Technologies for Information Society (NTIS), European Centre of Excellence, ED1.1.00/02.0090.

## REFERENCES

[1] Y. Li, J. Sun, C.-K. Tang, and H.-Y. Shum, "Lazy snapping," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 303–308, 2004.

[2] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. Norwell, MA, USA: Kluwer Academic Publishers, 1981.

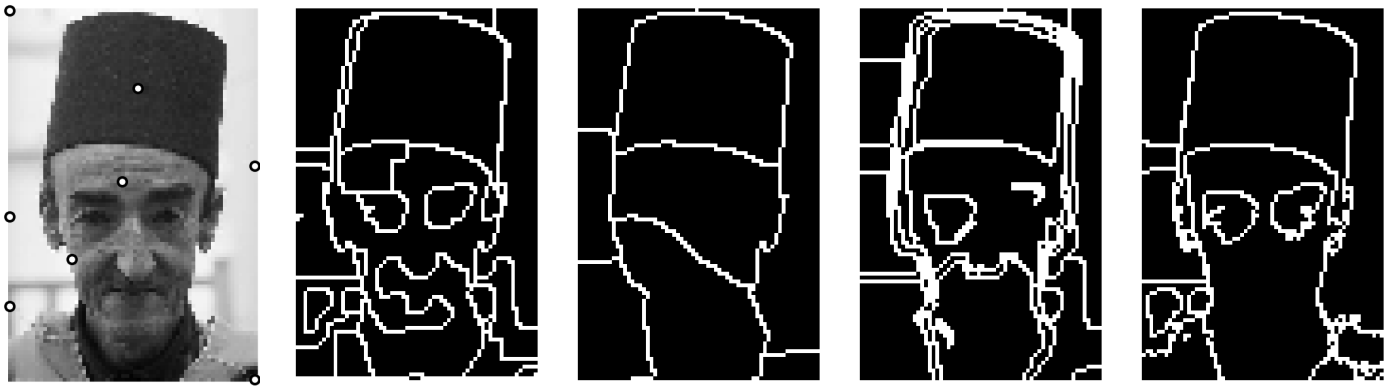
[3] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient Graph-Based Image Segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167–181, Sep. 2004. [Online]. Available: http://link.springer.com/10.1023/B:VISI.0000022288.19776.77

[4] J. Shi and J. Malik, "Normalized cuts and image segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 8, pp. 888 –905, aug 2000.

[5] T. Ryba, M. Jirik, and M. Zelezny, "An automatic liver segmentation algorithm based on grow cut and level sets," *Pattern Recognition and Image Analysis*, vol. 23, no. 4, pp. 502–507, 2013. [Online]. Available: http://dx.doi.org/10.1134/S1054661813040147

[6] V. Vezhnevets and V. Konouchine, "Growcut - interactive multi-label n-d image segmentation by cellular automata," *Cybernetics*, p. 150156, 2004.

[7] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001. [Online]. Available: http://dx.doi.org/10.1109/34.969114



(a) Input image with seeds (b) GC, pri = 0.832 (c) RW, pri = 0.849 (d) FHC, pri = 0.831 (e) NC, pri = 0.867

Fig. 4. Segmentation of the image called 'man'.

(a) Input image with seeds (b) GC, pri = 0.834 (c) RW, pri = 0.835 (d) FHC, pri = 0.882 (e) NC, pri = 0.592

Fig. 5. Segmentation of the image called 'bird'.

(a) Input image with seeds (b) GC, pri = 0.667 (c) RW, pri = 0.619 (d) FHC, pri = 0.652 (e) NC, pri = 0.661

Fig. 6. Segmentation of the image called 'boy'.

[8] L. Grady, "Random walks for image segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 11, pp. 1768 –1783, nov. 2006.

[9] A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr, "Interactive image segmentation using an adaptive gmmrf model," in *in ECCV*, 2004, pp. 428–441.

[10] V. Pathak, P. Dhyani, and P. Mahanti, "Autonomous image segmentation using density-adaptive dendritic cell algorithm," in *in International Journal of Image, Graphics and Signal Processing*, vol. 5, no. 10, 2013, pp. 26 – 35.

[11] M. Wertheimer, *Laws of organization in perceptual forms*. London: Harcourt, Brace & Jovanovitch, 1938.

[12] T. Pavlidis, "A critical survey of image analysis methods," *ICPR*, pp. 502–511, 1986.

[13] C. K. Chow, "A recognition method using neighbor dependence," *Electronic Computers, IRE Transactions on*, vol. EC-11, no. 5, pp. 683–690, Oct 1962.

[14] A. Blake, P. Kohli, and C. Rother, *Markov Random Fields for Vision and Image Processing*. MIT Press, 2011.

[15] J. M. Hammersley and P. E. Clifford, "Markov random fields on finite graphs and lattices," 1971.

[16] S. Kakutani, "Markov processes and the dirichlet problem," *Proc. Japanese Academy*, vol. 21, pp. 227–233, 1945.

[17] P. G. Doyle and J. L. Snell, "Random walks and electric networks," *American Mathematical Monthly*, vol. 94, no. January, p. 202, 2000. [Online]. Available: http://arxiv.org/abs/math/0001057

[18] R. Courant and D. Hilbert, *Methods of Mathematical Physics, Volume II*. John Wiley & Sons, 1966, vol. 6, no. 4.

[19] R. Hersh and R. J. Griego, *Brownian Motion and Potential Theory*, 1969, vol. 220.

[20] T. Ryba, M. Jirik, and M. Zelezny, "An automatic image segmentation algorithm involving shortest path basins," in *International Conference on Pattern Recognition and Image Analysis - Proceedings of Extended Abstracts*, 2013.

[21] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient n-d image segmentation," *International Journal of Computer Vision*, vol. 70, no. 2, pp. 109–131, 2006. [Online]. Available: http://www.springerlink.com/index/10.1007/s11263-006-7934-5

[22] R. Unnikrishnan and M. Hebert, "Measures of similarity," in *Application of Computer Vision, 2005. WACV/MOTIONS'05 Volume 1. Seventh IEEE Workshops on*, vol. 1. IEEE, 2005, pp. 394–394.

[23] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th Int'l Conf. Computer Vision*, vol. 2, July 2001, pp. 416–423.

# *Analysis of ionospheric parameters and geomagnetic field variations by the wavelet-transform and multicomponent models*

O.V. Mandrikova, Yu.A. Polozov, I.S. Solovev, N.V. Fetisova(Glushkova)
Institute of Cosmophysical Research and Radio Wave Propagation (IKIR FEB RAS)
Paratunka village, Russia
e-mails: oksanam1@mail.ru, up_agent@mail.ru, kamigsol@yandex.ru, nv.glushkova@yandex.ru

M.S. Kupriyanov
Saint-Petersburg Electrotechnical University "LETI"
Saint-Petersburg, Russia
e-mail: mikhail.kupriyanov@gmail.com

Dmitriev A.
Institute of Space Science National Central University
Jhongli, Taiwan
alexei_dmitriev@yahoo.com

***Abstract*—The work is aimed at the creation of techniques for geophysical data analysis and detection of abnormal changes during ionospheric-magnetospheric disturbances. This paper describes methods for modeling of ionospheric parameters based on combination of wavelets with a class of autoregressive-integrated moving average models. These methods allow us to approximate complex dependencies of data, to obtain forecasts on parameter variations and to detect anomalies during ionospheric disturbances. In order to perform a detailed study of anomalous variations in the ionosphere and magnetosphere, the authors suggest using the continuous wavelet transform. Computing solutions for automatic detection of anomalous variations in ionospheric parameters, geomagnetic field variations and estimation of their parameters (moments of occurrence, duration and scales of anomalies and intensity) are described. The application of the developed techniques is shown for Kamchatka region. The analysis of dynamic mode of ionosphere and magnetosphere in the analyzed region is performed during increased solar activity.**

***Keywords—wavelet transform, autoregressive-integrated moving average model, the ionospheric critical frequency, geomagnetic field variations, ionospheric-magnetospheric disturbances.***

## I. INTRODUCTION

The state of near-Earth space as one of the important factors of space weather affects different spheres of our life, including the operation of satellite systems, radio propagation and etc.; its investigation is based on the analysis of registered parameter variations of the environment. The recorded parameters have nonstationary structure, contain uneven local peculiarities, appearing at random times and having different form and duration, and, as a rule, carrying the main information about the processes under the investigation [1, 2]. Traditional methods and approaches applied to study the processes in the magnetospheric-ionospheric system, are not effective enough especially during disturbances, they mask the process dynamics and result in significant losses of important data [3, 4]. In addition to the complex nature of the process, in the most cases the problem is also associated with sparse network of measuring instruments as well as with the fragmentation of used technology and methods.

At present, adaptive filtration methods, neural nets, and wavelet-transform are intensively developing in the tasks of monitoring and forecast of magnetosphere-ionosphere system state [1-3, 5, 6]. Application of wavelet-transform method ionograms in recognition algorithms [5] allowed one to significantly increase the speed and the quality of their recognition in automatic regime for the region of Sourh-Eastern Asia, thus, optimizing the process of monitoring of ionosphere parameters and early detection of ionospheric inhomogeneities. Application of neural nets allowed one to solve the task of display of complicated non-linear dependences in of environment parameter modeling [6], to develop a method of modeling of geomagnetic field variations for low-latitudinal regions [6]. On the basis of neural net apparatus, models of forecast for ionospheric variations and storms in interactive regime near Tokyo were developed [1, 2].

Extending the area of traditional techniques for modeling and analysis of time series, the authors suggest a method for ionosphere parameter modeling, based on combination of wavelets and autoregressive-integrated moving average (ARIMA) models. On the basis of this method, approximations of ionosphere parameter time variation were constructed, data analysis was carried out, and anomalies in the ionosphere were detected. It was suggested to use continuous wavelet-transform

The present paper and the research are supported by the grant of the Russian Science Foundation, project no. 14-11-00194, the grant of President Fellowship of the Russian Federation, project no. 2976.2013.5 and the Foundation for Advancement of Small Businesses in Science and Technology, project U.M.N.I.K. (Participant of Youth Science and Innovation Contest) no. 11754r/17262 of April 5, 2013.

for the detailed investigation of anomalous changes in the ionosphere and magnetosphere. Computational solutions for automatic detection of anomalous changes in ionosphere parameters, geomagnetic field variations and estimation of their parameters (the times of appearance, duration, anomaly scales and intensity are estimated) are described

## II. Description of the Methods

### A. Multicomponent model identification and assessment

Consider a random time series, containing stationary components and noise, as $f_0$. On the basis of multiresolution wavelet decomposition to $m$ level [7] the $f_0$ time series is presented as a linear combination of multiscale components [8]: $f\left[2^{-m}t\right]$ is a smoothed component of $m$ scale and $g\left[2^{j}t\right]$ are detailing components of $j=\overline{-1,-m}$ scales:

$$f_0(t)=\sum_{j=-1}^{-m} g\left[2^{j}t\right]+ f\left[2^{-m}t\right] \qquad (1)$$

On the basis of changing the decomposition level $m$ we can obtain various representations of a time series. In order to determine the best representation, which extracts stationary components from noise and allows us to obtain an adequate ARIMA model for them, the following steps are to be made:

1. We obtain a representation of a time series in the form (1) for the levels $m=\overline{1,M}$ (the maximum acceptable level $M$ of decomposition is determined by the length of $N$ time series: $M\le \log_2 N$) and a set of smoothed components as follows $f\left[2^{-m}t\right]=\sum_k c_{-m,k}\phi_{-m,k}(t)$, $m=\overline{1,M}$.

2. We determine stationary components from a set of the components $f\left[2^{-m}t\right], m=\overline{1,M}$. Applying the traditional approaches [9] we determine models from ARIMA model class for approximation of $f\left[2^{-m}t\right]$ stationary components. We obtain presentations for each components as follows: $f_{-m}(t)=\sum_k s_{-m,k}\phi_{-m,k}(t)$,

where $s_{-m,k}=\sum_{l=1}^{p}\gamma_{-m,l}\omega_{-m,k-l}-\sum_{n=1}^{h}\theta_{-m,n}a_{-m,k-n}$ is assessed value of a smoothed component; $\omega_{-m,k}=\nabla^{\nu}c_{-m,k}$, $\nabla^{\nu}$ is the difference operator of order $\nu$; $p,\gamma_{-m,l}$ are autoregression model order and parameters of smoothed component; $h$, $\theta_{-m,n}$ are model order and parameters of a moving average of smoothed component; $a_{-m,k}$ are residual errors of model.

3. We carry out evaluation of the component model errors:

$$E_m=\sum_{k=1}^{K}\sum_{q=1}^{Q} e_{k+q}^{m},$$

where $e_{k+q}^{m}=\left(s_{-m,k+q}^{actual}-s_{-m,k+q}^{model}\right)^2$ is component model error at the point $k$ with time advance $q$; $s_{-m,k+q}^{actual}$ are actual values of time series component; $s_{-m,k+q}^{model}$ are model values of time series component; $Q$ is the length of data time advance; $K$ is length of time series component.

4. We consider that the best representation of a time series is the representation corresponding to a multiresolution wavelet decomposition to level $m^*$, where $m^*: E_{m^*}=\min_m E_m$

5. We determine stationary components from a set of detailing components $g\left[2^{j}t\right], j=\overline{-1,-m^*}$. Applying the traditional approach [9] we determine models from ARIMA model class for approximation of stationary components $g\left[2^{j}t\right]$.

6. Components $g\left[2^{j}t\right]$, which are not stationary, contain local features and noise.

7. Using the expression (1) we combine the obtained component models into a joint multi-component construction representing data changes in the time domain

$$f_0(t)=\sum_{\mu=\overline{1,\mathrm{T}}}\sum_{k=\overline{1,N_j^{\mu}}} s_{j,k}^{\mu}b_{j,k}^{\mu}(t) \qquad (2)$$

where $s_{j,k}^{\mu}=\sum_{l=1}^{p_j^{\mu}}\gamma_{j,l}^{\mu}\omega_{j,k-l}^{\mu}-\sum_{n=1}^{h_j^{\mu}}\theta_{j,n}^{\mu}a_{j,k-n}^{\mu}$ is the assessed value of $\mu$-th component; $p_j^{\mu},\gamma_{j,l}^{\mu}$ are order and parameters of the $\mu$-th component autoregression; $h_j^{\mu}$, $\theta_{j,k}^{\mu}$ are model order and parameter of a moving average of $\mu$-th component; $\omega_{j,k}^{\mu}=\nabla^{\nu^{\mu}}\beta_{j,k}^{\mu}$, $\nu^{\mu}$ is the order of difference of $\mu$-th component; $\beta_{j,k}^{1}=c_{j,k}$, $\beta_{j,k}^{\mu}=d_{j,k}, \mu=\overline{2,\mathrm{T}}$, T is the number of modeled components; $a_{j,k}^{\mu}$ are residual errors of $\mu$-th component model; $N_j^{\mu}$ is the length of $\mu$-th component; $b_{j,k}^{1}=\phi_{j,k}$ is a scaling function; $b_{j,k}^{\mu}=\Psi_{j,k}, \mu=\overline{2,\mathrm{T}}$ is a wavelet basis of $\mu$-th component.

Prediction of $s_{j,k+q}^{\mu}$ value, $q\ge 1$ determines the prediction of $s_{j,k}^{\mu}$ value at the point $k$ with time advance $q$. $s_{j,k+q}^{\mu}$ value can be determined by the $\mu$-th component model:

$s^{\mu}_{j,k+q} = \sum_{l=1}^{p^{\mu}_j} \gamma^{\mu}_{j,l} \omega^{\mu}_{j,k+q-l} - \sum_{n=1}^{h^{\mu}_j} \theta^{\mu}_{j,n} a^{\mu}_{j,k+q-n}$. Residual errors of $\mu$-th component model are determined as a difference between actual and predicted values at the point $k+q: a^{\mu}_{j,k+q} = s^{\mu,actual}_{j,k+q} - s^{\mu,predict}_{j,k+q}$. The obtained multicomponent model (MCM-model) (2) represents typical changes of approximated data. During the periods of abnormal changes of data, residual error absolute values of component models will rise. For this reason detection of anomalies can be carried out by the following conditional test:

$$\varepsilon_{\mu} = \sum_{q=1}^{Q_{\mu}} \left| a^{\mu}_{j,k+q} \right| > T_{\mu},$$

where $Q_{\mu}$ is length of data time advance on the basis of $\mu$-th component model; $T_{\mu}$ is a threshold value of $\mu$-th component, which defines the presence of anomalies.

### B. *Detection of ionospheric anomalies and assessment of their parameters on the basis of continuous wavelet-transform*

Continuous wavelet-transform is determined relatively every basic wavelet $\Psi$ by the following formula [7].

$$W_{\Psi} f_{b,a} := |a|^{-1/2} \int_{-\infty}^{\infty} f(t) \Psi\left( \frac{t-b}{a} \right) dt,$$
$$f \in L^2(R), a, b \in R, a \neq 0.$$

When scale $a$ decreases, amplitudes of the coefficients $\left| W_{\Psi} f_{b,a} \right|$ have fast decreases up to zero in the regions where $f$ function does not have local peculiarities [7]. Basing on this property we apply the following threshold function to detect anomalies in the ionosphere:

$$P_{T_a}(W_{\Psi} f_{b,a}) = \begin{cases} W_{\Psi} f_{b,a}, если \left| W_{\Psi} f_{b,a} - W_{\Psi} f^{med}_{b,a} \right| \geq T_a \\ 0, если \left| W_{\Psi} f_{b,a} - W_{\Psi} f^{med}_{b,a} \right| < T_a \end{cases},$$

were the threshold $T_a = U * St_a$ determines anomalies on scale $a$ in the vicinity of point $\xi$, which is contained in the carrier $\Psi_{b,a}$, $U$ is a threshold coefficient, $St_a = \sqrt{\frac{1}{\Phi - 1} \sum_{k=1}^{\Phi} \left( W_{\Psi} f_{b,a} - \overline{W_{\Psi} f_{b,a}} \right)^2}$, $\overline{W_{\Psi} f_{b,a}}$ and $W_{\Psi} f^{med}_{b,a}$ is an average and a mediana, determined in a moving time window of length $\Phi$. Considering the diurnal variation of $f_0F2$ data, the average $\overline{W_{\Psi} f_{b,a}}$ and the median $W_{\Psi} f^{med}_{b,a}$ were calculated for every hour separately. As long as the carrier $\Psi_{b,a}$ on the scale $a$ is equal to $[b - \Omega a, b + \Omega a]$, the cone of influence of point $\xi$ on scale $a$ is determined by the inequality $|b - \xi| \leq \Omega a$.

Anomaly time duration on scale $a$ is determined by the cone of influence of point $\xi$ and is equal to $H_j = 2\Omega a$. Anomaly intensity at time $t = b$ is defined as $Y_b = \sum_a \left| W_{\Psi} f_{b,a} \right|$.

### C. *Determination of anomalous periods in variations of geomagnetic field and assessment of intensity of geomagnetic disturbances*

Considering the equivalence of discrete and continuous wavelet transforms [10], it is possible to introduce the way of calculation of intensity of geomagnetic perturbations at the time point $t = b$ on the scale $a$. This intensity can be found as

$$e_{b,a} = \left| (W_{\Psi} f)(b,a) \right| \quad (3)$$

Then, by applying a threshold function to the value $e_{b,a}$, we can estimate state of the coefficients and allocate time-frequency intervals containing weak and strong geomagnetic disturbances on the analyzed scale $a$:

$$P_{T_{a,1}}(e_{b,a}) = \begin{cases} 0, if \ e_{b,a} < T_{a,1} \\ e_{b,a}, if \ e_{b,a} \geq T_{a,1} \end{cases},$$

$$P_{T_{a,2}}(e_{b,a}) = \begin{cases} 0, if \ e_{b,a} < T_{a,2} \\ e_{b,a}, if \ e_{b,a} \geq T_{a,2} \end{cases},$$

where the threshold $T_{a,1}$ allows us to allocate weak and strong perturbations, and the threshold $T_{a,2}$ allocates strong perturbations. On the basis of (3), the intensity of disturbances of the field at time moment $t = b$ is: $E_b = \sum_a e_{b,a}$.

## III. Analysis of the Data

The Fig. 1 and 2 show the results of the ionospheric and geomagnetic data processing during magnetic storms on 14-22 March 2013 and 12-16 December 2013. On the eve of the storm on the 14-22 March 2013 (Fig. 1) local increase of geomagnetic activity was observed in the geomagnetic field, simultaneously a large-scale positive anomaly occurred in the ionosphere (it is shown in red, it characterizes the increase of electron density in the ionosphere). A large-scale negative anomaly occurred in the ionosphere during the storm (it is shown in blue, it characterizes reduction of electron density in the ionosphere) with duration of more than a day and small scale anomalies were also observed.

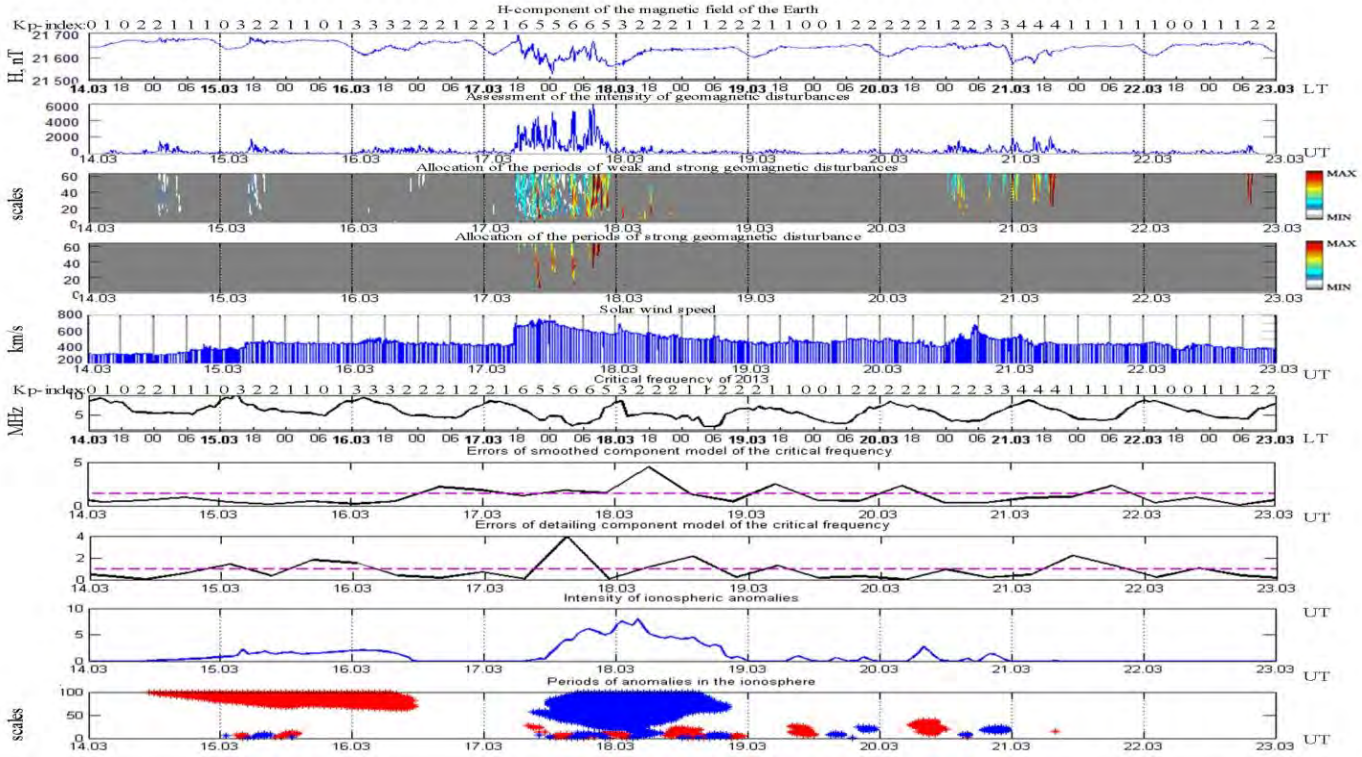



Fig. 1. Results of processing of geomagnetic and ionospheric data on March 14-23, 2013.

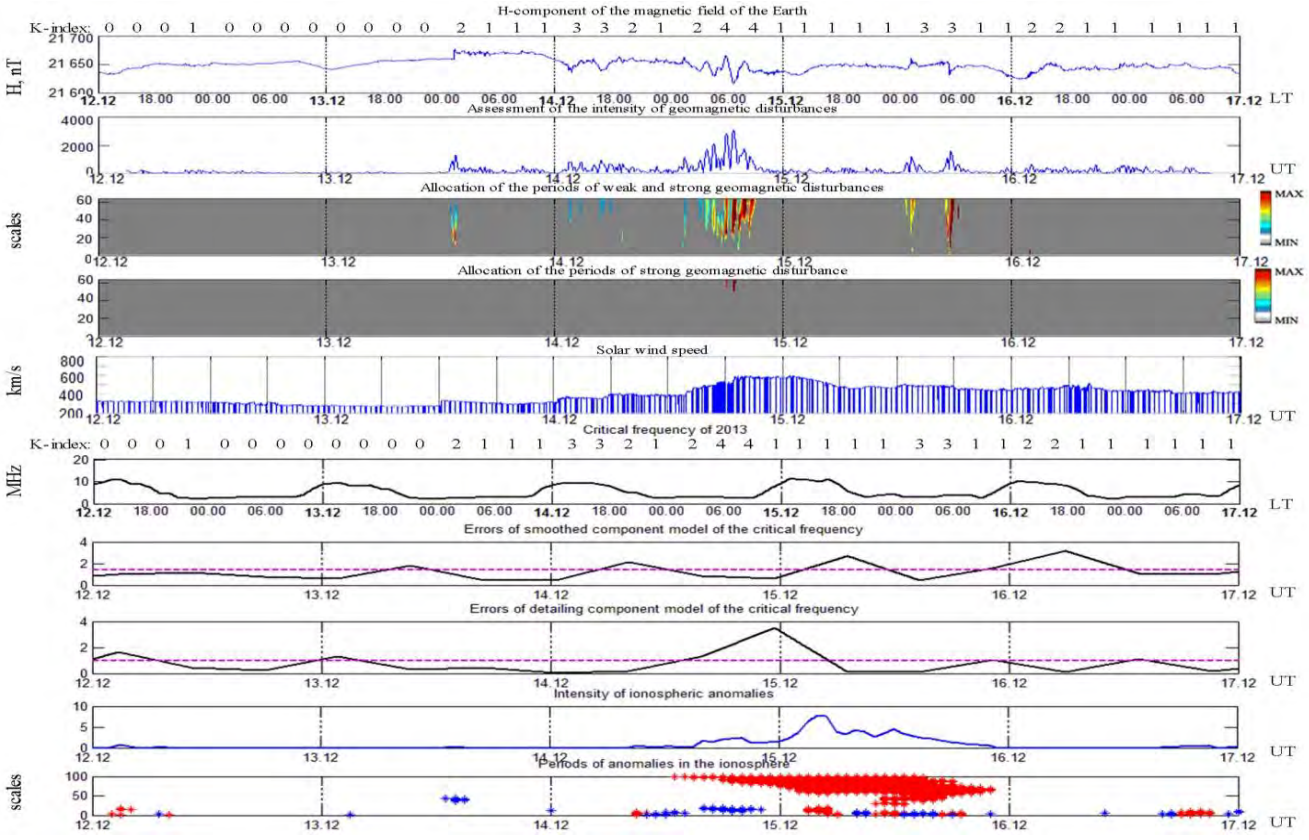



Fig. 2. Results of processing of geomagnetic and ionospheric data on December 12-16, 2013.

On the eve of the storm on the 14-15 December 2013 (Fig. 2) a local increase of geomagnetic activity was observed in the geomagnetic field (December 13, from 12:00 UT to 13:00 UT, the value of K-index reaches the value of 2), while a short-term negative anomaly appeared in the ionosphere. During the magnetic storm a two-day large-scale positive anomaly occurred in the ionosphere and some small-scale anomalies. The significant increases of MCM model errors also indicate the presence of anomalies in the ionosphere during the storms.

The analysis results show that the proposed methods can obtain detailed and reliable information of the state of the ionosphere-magnetosphere system in the analyzed geographic region.

## ACKNOWLEDGMENT

The authors are grateful to the institutes supporting the registration stations of ionospheric parameters and magnetic observatories whose data were used in the research.

## REFERENCES

[1] M. Nakamura, T. Maruyama and Y. Shidama, "Using a neural network to make operational forecasts of ionospheric variations and storms at Kokubunji, Japan", Journal of the National Institute of Information and Communications Technology, 2009, vol. 56, pp.391–406.

[2] K. Watthanasangmechai, P. Supnithi, S. Lerkvaranyu, T. Tsugawa, T. Nagatsuma and T. Maruyama, "TEC prediction with neural network for equatorial latitude station in Thailand", Earth, Planets and Space, 2012, vol. 64, pp. 473–483.

[3] S. Marple, Digital Spectral Analysis with Applications. New-Jersey: Prentice-Hall, 1987.

[4] S. Pervak, V. Choliy and V. Taradiy, "Spectral analysis of the ionospheric irregularities from GPS observations", Proc. of the 17th Annual Conf. of Doctoral Students – WDS 2008, Prague, Part II – Physics of Plasmas and Ionized Media, 2008, pp. 189–191.

[5] H. Kato, Y. Takiguchi, D. Fukayama, Y. Shimizu, T. Maruyama and M. Ishii, "Development of automatic scaling software of ionospheric parameters", Journal of the National Institute of Information and Communications Technology, 2009, vol. 56, pp. 465–474.

[6] J. Uwamahoro, L.A. McKinnell and J.B. Habarulema, "Estimating the geoeffectiveness of halo CMEs from associated solar and IP parameters using neural networks", Annales Geophysicae, 2012, vol. 30, pp. 963.

[7] S. Mallat, A Wavelet Tour of Signal Processing. London: Academic Press, 1999.

[8] O.V. Mandrikova, N.V. Glushkova and I.V. Zhivet'ev, "Modeling and analysis of ionospheric parameters by a combination of wavelet transform and autoregressive models", Geomagnetism and Aeronomy, 2014, vol. 54, № 5, pp. 593-600. DOI: 10.1134/S0016793214050107

[9] G. Box and G. Jenkins, Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day, 1970.

[10] O.V. Mandrikova, I. Solovjev, V. Geppener, R. Taha Al-Kasasbehd and D. Klionskiy, "Analysis of the Earth's magnetic field variations on the basis of a wavelet-based approach", Digital Signal Processing, 2013, vol. 23, pp. 329-339.

# Analysis of near-field diffraction patterns of gaussian beams for surface defects detection

Dmitrey A. Savelyev[1,2] and Svetlana N. Khonina[1]
[1] Image Processing Systems Institute of the RAS, Molodogvardeyskay 151, Samara, Russia
[2] Samara State Aerospace University named after academician S.P. Korolyov, Moskovskoye Shosse 34, Samara, Russia
dmitrey.savelyev@yandex.ru, khonina@smr.ru

***Abstract* – Numerical simulation of diffraction of Gaussian laser beams with circular input polarization is calculated using the finite-difference time-domain method. The opportunity of surface defects (asperities and pits) recognition and defect dimensions change is shown.**

***Keywords – vortex phase singularity; defects defection; FDTD-method; Meep software***

## I. Introduction

Bright-field and dark-field microscopy are often used for surface inspection [1-3]. The majority of surface defects can be divided into two optical classes: specified for bright-field and dark-field microscopy. Intensity, reflected and refracted beams direction, reflected and refracted radiation coherence are changed because of defects presence. In this case, the importance of dark-field illumination is growing with the defect dimension decrease. With the help of dark-field microscopy it became possible to detect defects smaller than 10 μm on a moving target [2]. In the paper [3] cylindrical beams were used to analyze boundary singularities of surface dark-field imaging.

Vortex (radial vortex phase change from 0 to $2\pi$) phase singularity introduction into an incident beam [4] lets us strengthen the longitudinal component of laser beams with the homogeneous polarization on the optical axicon in the focal plane. Due to their unusual characteristics, singular beams can be used in different applications such as capturing and manipulation of micro-objects [5] and objects microstructure research [6].

In this paper, we consider such defects as asperities and pits. Besides that, we research the diffraction pattern change after defect's height and width modification. The numerical simulation of diffraction is calculated by means of the finite-difference time-domain method (FDTD), implemented in the Meep software package.

## II. Surface Defects Detection

Numerical simulation was made using the computational cluster with the power of 775 GFlops. The cluster's characteristics are the following: 116 cores, computing nodes – 7 dual servers HP ProLiant 2xBL220c, RAM volume 112 Gbit.

Simulation parameters are the following: wavelength $\lambda = 0.532$ μm, computational domain size $x, y, z \in [-7\lambda; 7\lambda]$. Absorbing layer PML thickness is $1.5\lambda$, space discretization is $\lambda/21$, time discretization is $\lambda/(42c)$ where c is the light speed. Light source is disposed within the pad at the distance $0.1\lambda$ from the asperity (pit). Refractive index $n$ is 1.68. Asperity (pit) height is chosen in order to correspond with the phase advance for given wavelength $\lambda$ and refractive index $n$:

$$h = \pi / k(n-1) = \lambda / 2(n-1) \approx 0.74\lambda \quad (1)$$

Two types of laser beams, which can be generated in laser resonators and preserve their structure propagating in free space with circular polarization of laser irradiance in direction opposite to that of vortex phase singularity, are researched: Gaussian beam and Laguerre-Gauss mode (0,1). The beams being researched are shown in Figure 1.

a) b) c)

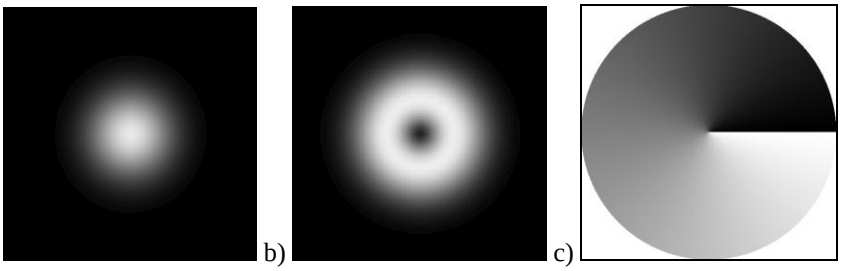
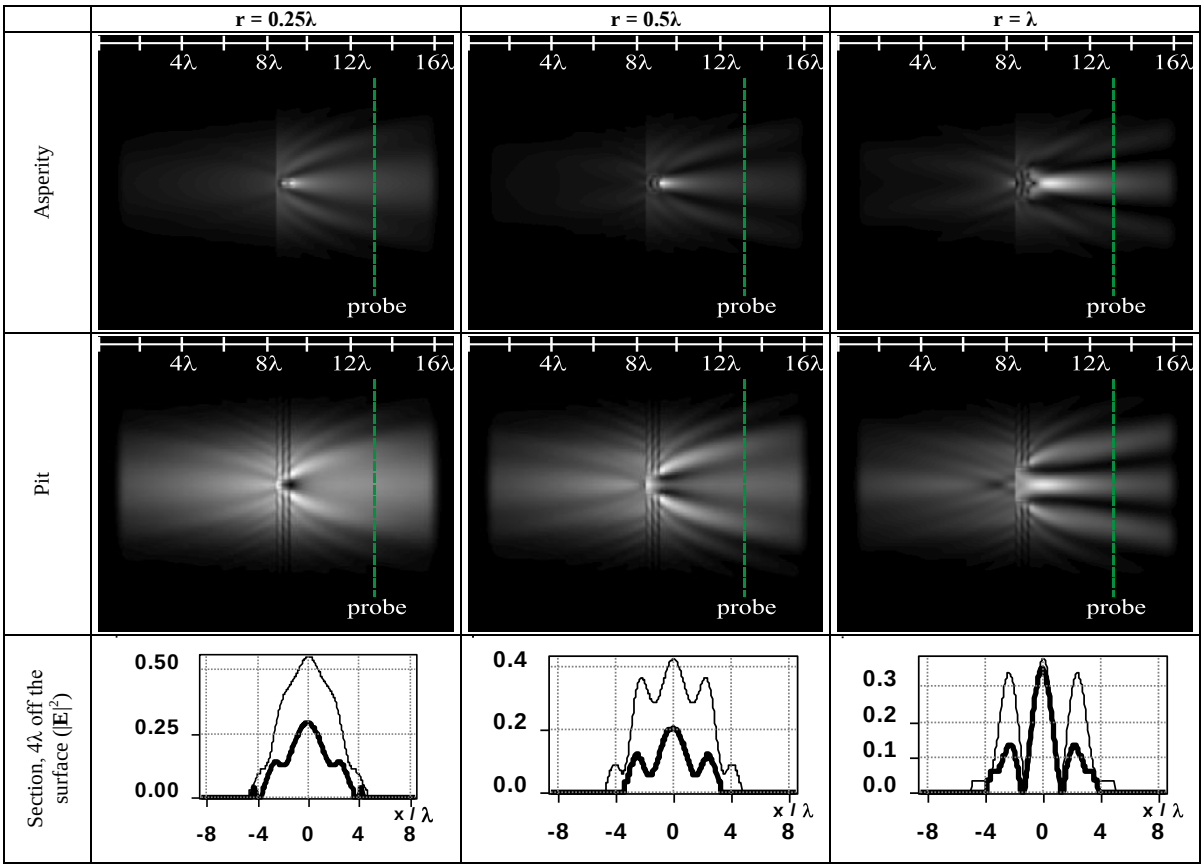


Fig. 1. Incident beams: a) Gaussian beam intensity, b) Laguerre-Gauss mode (0,1) intensity, c) Laguerre-Gauss mode (0,1) phase

In [7] was experimentally show that square micro-steps (asperities) can use for tight focusing of laser beams. Now we consider round surface defects, such as asperities and pits, varying their size. Table 1 show longitudinal sections of found Gaussian beam diffraction pattern and plots of transverse sections intensity ($|\mathbf{E}|^2$) at the distance $4\lambda$ from the surface. Asperity is shown as a heavy line, and pit is shown as a thin one.

TABLE I. LONGITUDINAL SECTION, GAUSSIAN BEAM, GENERAL INTENSITY




It should be noted that using the fundamental Gaussian mode, either for the asperity or for the pit, lets us detect the size change of the asperity or of the pit: with the radius growing, side petals appear and the gap between these petals and the central peak increases. When radius r = 0.5λ, it is possible to distinguish between the asperity and the pit due to the difference between central peak and side petals heights: peaks maximums are comparable in height for a pit case.

Now we do the same research, introducing the phase singularity into the beam, i.e. we consider the Laguerre-Gauss mode (0,1). Table 2 shows longitudinal diffraction patterns and plots of transverse sections intensity at the distance 4λ and λ from the surface.

As we can see from the plots, the defect smaller than some value (smaller than λ) can't be detected. In this case, laser beam size needs to be decreased. Nevertheless, if radius r = λ, it is possible to distinguish the asperity from the pit due to the difference between side petals height and central intensity value. Moving the probe closer to the element's surface lets us define the type of a defect due to the appearance of a bright light spot in the center for the case of the asperity.

Now we fix the asperity size and then change the height, increasing it two times: $h$ = 1.48λ. Figure 2 shows two transverse diffraction patterns for an asperity.

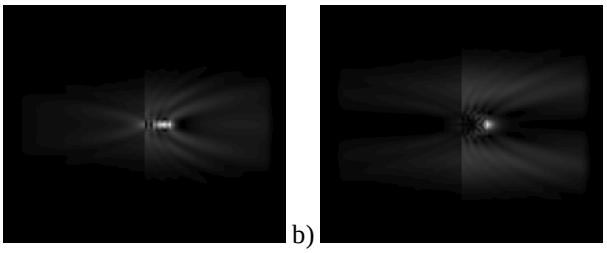



Fig. 2. Transverse diffraction patterns for an asperity (intensity) when h = 1,48λ: a) Gaussian beam, r = 0.5λ b) Laguerre-Gaussian mode (0,1), r = λ

Table 3 shows plots of transverse sections intensity at a fixed distance from the surface. Since heights are compared, the probe is located at the fixed distance 4λ and λ from the surface when $h$ = 0.74λ, i.e. when $h$ = 1.48λ, this distance is 3.26λ and 0.26λ, respectively. The initial height $h$ = 0.74λ is shown as a heavy line, $h$ = 1.48λ is shown as a thin one.

As we can see from the Table 3, for a Gaussian beam asperity height increase is characterized by increase of the light spot width, the central peak is smaller than side petals, when h = 1.58λ, in contrast to the pit for h = 0.74λ (see Table 1, line 3).

Introducing phase singularity to a beam, if the probe is located close enough (~λ) to the surface for the focusing effect to take place, height increase is characterized by light spot decrease. If the probe is located far from the surface (> 3.26λ), asperity height is increasing, central intensity is decreasing and the plot is smoothing.

TABLE II. LONGITUDINAL SECTION, LAGUERRE-GAUSS MODE (0,1), GENERAL INTENSITY

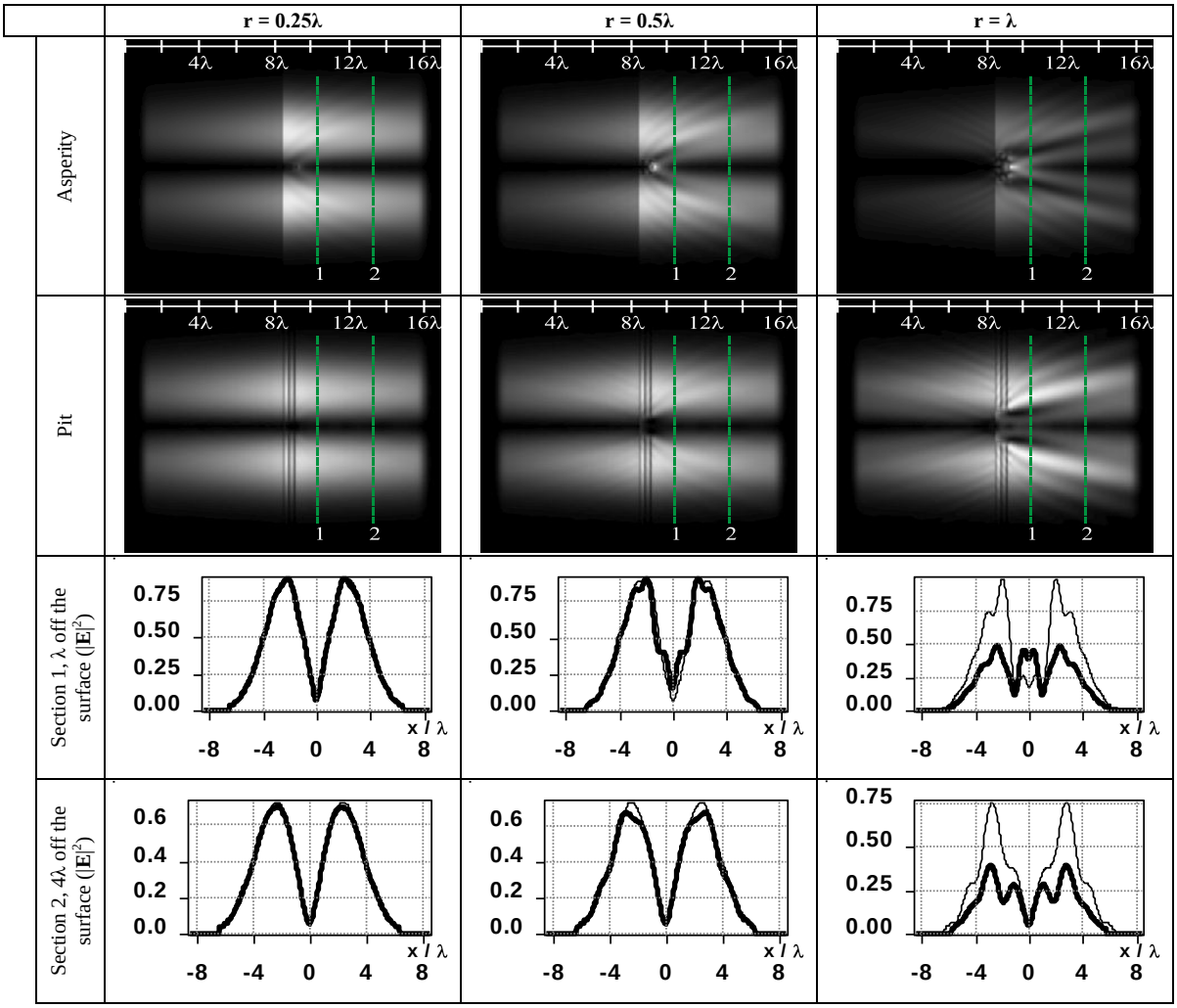
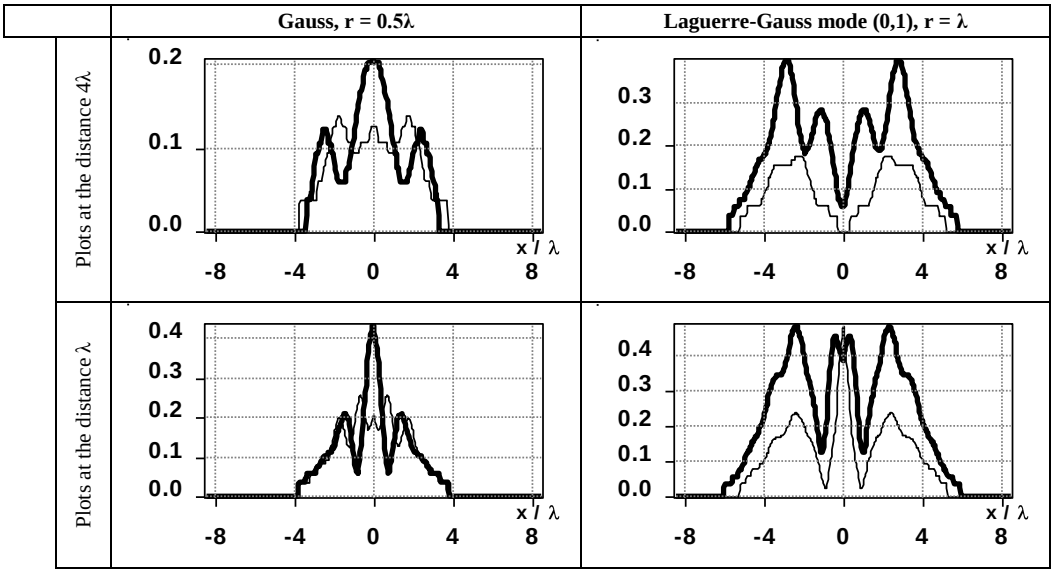



TABLE III. DYNAMICS OF ASPERITY HEIGHT CHANGE, INTENSITY ($|\mathbf{E}|^2$)



## III. Conclusion

In this paper, we suggest using Gaussian beams, in particular those with the vortex phase singularity, to detect surface defects, such as an asperity or a pit, and the possibility of these defects recognition is shown, in particular when their width and height change.

When using the fundamental Gaussian mode:

- when radius r = 0.5λ, it is possible to distinguish between the asperity and the pit due to the difference between central peak and side petals heights because peaks maximums are comparable in height for a pit;
- size change of the asperity or of the pit is detected the following way: with the radius growing, side petals appear and the gap between these petals and the central peak increases;
- asperity height increase is characterized by increase of the light spot width.

When using the Laguerre-Gaussian mode (0,1):

- if radius r = λ it is possible to distinguish the asperity from the pit due to the difference between side petals height and central intensity value, moving the probe closer to the element's surface lets us define the type of a defect due to the appearance of a bright light spot in the center for the case of the asperity;
- asperity height increase is characterized by light spot decrease, if the probe is located close enough (~λ) to the surface, and if the probe is located far from the surface (> 3.26λ), asperity height is increasing, central intensity is decreasing and the plot is smoothing.

It is shown that it is possible to detect defects smaller than 1 μm, however, in this case, near-field microscopes are necessary to use; but contrast to scanning electron microscopes metal sputtering is not required, therefore considered object deformation doesn't take place.

## Acknowledgment

We acknowledge a partial financial support from the Russian Foundation for Basic Research, grant No 14-07-31079 mol_a.

## References

[1] J. Zhang, M.C. Pitter, S. Liu, C. See and M.G. Somekh, "Surface-plasmon microscopy with a two-piece solid immersion lens: bright and dark fields," *Applied Optics*, vol. 45, no. 31, pp. 7977-7986, 2006.

[2] A. Yazaki, C. Kim, J. Chan, A. Mahjoubfar, K. Goda, M. Watanabe and B. Jalali, "Ultrafast dark-field surface inspection with hybrid-dispersion laser scanning," *Appl. Phys. Lett.*, vol. 104, p. 251106, 2014.

[3] D.P. Biss, K.S. Youngworth and T.G. Brown, "Dark-field imaging with cylindrical-vector beams," *Applied Optics*, vol. 45, no. 3, pp. 470-479, 2006.

[4] S.N. Khonina and I. Golub, "Optimization of focusing of linearly polarized light," *Opt. Lett.*, vol. 36, no.3, pp. 352-354, 2011.

[5] W. M. Lee, X.-C. Yuan and W.C. Cheong, "Optical vortex beam shaping by use of highly efficient irregular spiral phase plates for optical micromanipulation," *Optics Letters*, vol. 29, no. 15, pp. 1796-1798, 2004.

[6] J. Masajada, M. Leniec, E. Jankowska, H. Thienpont, H. Ottevaere and V. Gomez, "Deep microstructure topography characterization with optical vortex interferometer," *Optics Express*, vol. 16, no. 23, pp. 19179-19191, 2008.

[7] V.V. Kotlyar, S.S. Stafeev and A.Y. Feldman, "Photonic nanojets formed by square microsteps," *Computer Optics*, vol. 38, no 1, pp. 72-80, 2014.

# APPLICATION OF MIXED MODELS TO SOLVE THE PROBLEMS OF RESTORATION AND ESTIMATION OF IMAGE PARAMETERS

**K. Vasiliev, V. Dementiev, N. Andriyanov[1]**

**[1] Ulyanovsk state technical university, 32, Severnyi Venec, Ulyanovsk, Russia, 432027, vkk@ulstu.ru, 8422778082**

Methods of estimation of the parameters of images mixed autoregressive models are considered. The possibility of using the estimated parameters for image restoration was investigated. We propose an algorithm for image restoration based on the combination of pseudogradient and Kalman estimates. Comparative analysis of the effectiveness of various procedures is carried out.

## Introduction

For solving a number of problems of representation and processing of images it is advisable to use stochastic methods [1-5]. These problems include anomaly detection, image enhancement in the presence of interference, estimating deformations of multidimensional images and others. In many cases, data transmission errors, shading image the problem of recovering of missing patches arises [6-8]. To solve it, we can use singular value decomposition of matrices with gaps [6, 7], but this approach leads to large errors and significant computational cost.

The report proposes to restore the portion of the image based on the data available to carry out estimation of parameters in a sufficiently general doubly stochastic model [3-5].

However, this approach does not guarantee precise estimates, since it contains at least twice a random additive value. Improvement of the accuracy of the restoration can be achieved by the consecutive use of pseudogradient algorithm [1] to estimate the model parameters and vector Kalman filter [5] to determine the parameters in the vicinity of the damaged area. After that there appears a possibility of using the fitted model to implement forecasting (interpolation) of the existing image into the absent area.

## Pseudogradient parameter estimation of doubly stochastic model

We consider the case when the square patch of the resulting image is subjected to severe damage, so that in a certain area of it can be considered as noise:

$$x_{ij} = N,\ i = \{i_0,..., i_0 + q - 1\},\ j = \{j_0,..., j_0 + q - 1\}, \quad (1)$$

where q – the size of the damaged area; N – noise.

The problem of estimating the parameters can be relatively easy solved in the case of the formation of a random field (RF) $X = \{x_{\bar{i}}, \bar{i} \in \Omega\}$ by a simple autoregressive model:

$$x_{i,j} = \rho_x x_{i-1,j} + \rho_y x_{i,j-1} - \rho_x \rho_y x_{i-1,j-1} + \xi_{i,j},\ i = \overline{1...M_1};\ j = \overline{1...M_2}, \quad (2)$$

where $\rho_x$ и $\rho_y$ – correlation coefficients of adjacent elements in columns and rows, respectively; $\{\xi_{i,j}\}$ – independent Gaussian random variables (RV) with zero mean $M\{\xi_{i,j}\} = 0$ and variance $\sigma_\xi^2 = M\{\xi_{i,j}^2\} = (1-\rho_x^2)(1-\rho_y^2)\sigma_x^2;\ \sigma_x^2 = M\{x_{i,j}^2\};\ M_1 \times M_2$ – size of the simulated image.

Indeed, the correlation parameter estimates for further recovery section (1) in the model (2) can be easily obtained on the basis of the undamaged picture element:

$$\rho_x = \frac{\sum_{i=1}^{M_1-1}\sum_{j=1}^{M_2} x_{i,j}x_{i+1,j}}{\sigma_x^2 \times (M_1 - 1) \times M_2}, \qquad \rho_y = \frac{\sum_{i=1}^{M_1}\sum_{j=1}^{M_2-1} x_{i,j}x_{i,j+1}}{\sigma_x^2 \times M_1 \times (M_2 - 1)}.$$

However, a more interesting case is doubly stochastic model [2-5]. For it and the main field and the fields of the correlation coefficients may be obtained according to the expression:

$$x_{ij} = \rho_{xij}x_{i-1\,j} + \rho_{yij}x_{ij-1} - \rho_{xij}\rho_{yij}x_{i-1\,j-1} + \sigma_x^2\sqrt{1-\rho_{xij}^2}\sqrt{1-\rho_{yij}^2}\,\xi_{ij}, \quad (3)$$

where parameters $\rho_{xij} = \tilde{\rho}_{xij} + m_{\rho x}$, $\rho_{yij} = \tilde{\rho}_{yij} + m_{\rho y}$

$$\tilde{\rho}_{xij} = r_{11}\tilde{\rho}_{x(i-1)j} + r_{12}\tilde{\rho}_{xi(j-1)} - r_{11}r_{12}\tilde{\rho}_{x(i-1)(j-1)} + \sigma^2_{\rho x}\sqrt{1-r_{11}^2}\sqrt{1-r_{12}^2}\varsigma_{xij}$$

$$\tilde{\rho}_{yij} = r_{21}\tilde{\rho}_{y(i-1)j} + r_{22}\tilde{\rho}_{yi(j-1)} - r_{21}r_{22}\tilde{\rho}_{y(i-1)(j-1)} + \sigma^2_{\rho y}\sqrt{1-r_{21}^2}\sqrt{1-r_{22}^2}\varsigma_{yij}$$

$m_{\rho x}$ and $m_{\rho y}$ - average values of the correlation coefficients of the ground field, $\sigma_x^2$ - the variance of the main RF, $\xi_{ij}$, $\varsigma_{xij}$, $\varsigma_{yij}$ - Gaussian RV with $M\{\xi_{ij}\} = 0$ and $M\{\xi_{ij}^2\} = \sigma_\xi^2 = 1$, $i = 1,..., M_1$ and $j = 1,..., M_2$.

In this approach, one implementation of model parameter estimation (3) is the use of pseudogradient algorithms for search of estimates for the coefficients $r_{11}$, $r_{12}$, $r_{21}$ and $r_{22}$ [1,8].

Pseudogradient procedure may be described by the following expression [1]:

$$\hat{\alpha}_{t+1} = \hat{\alpha}_t - \Lambda_{t+1}\beta_{t+1}(J(Z_{t+1}, \hat{\alpha}_t),$$

where $\alpha$ - estimated parameter vector; $t$ - number of iterations; $\Lambda$ – approximation matrix; $\beta$ – pseudogradient of objective function $J$, which characterizes the quality of the evaluation; $Zt$ - local sampling observations used on t-th iteration.

Pseudogradient algorithms provide accuracy of estimating sufficient for most practical applications without requiring a priori knowledge of the gradient of the objective function and the features of its behavior.

The algorithm operation reduces to a comparison of errors and finding the optimal direction of motion at each step. So for the s-th step model correlation parameters can be rewritten as:

$$\tilde{\rho}_{xij} = (r_{11s} \pm \Delta r_{11s})\tilde{\rho}_{x(i-1)j} + (r_{12s} \pm \Delta r_{12s})\tilde{\rho}_{xi(j-1)} - (r_{11s} \pm \Delta r_{11s})(r_{12s} \pm \Delta r_{12s})\tilde{\rho}_{x(i-1)(j-1)} + \sigma^2_{\rho x}\sqrt{1-(r_{11s} \pm \Delta r_{11s})^2}\sqrt{1-(r_{12s} \pm \Delta r_{12s})^2}\varsigma_{xij},$$

$$\tilde{\rho}_{yij} = (r_{21s} \pm \Delta r_{21s})\tilde{\rho}_{y(i-1)j} + (r_{22s} \pm \Delta r_{22s})\tilde{\rho}_{yi(j-1)} - (r_{21s} \pm \Delta r_{21s})(r_{22s} \pm \Delta r_{22s})\tilde{\rho}_{y(i-1)(j-1)} + \sigma^2_{\rho y}\sqrt{1-(r_{21s} \pm \Delta r_{21s})^2}\sqrt{1-(r_{22s} \pm \Delta r_{22s})^2}\varsigma_{yij}$$

The second method consists in estimating the correlation parameters in the sliding window:

$$\rho_{xij} = \frac{1}{\sigma^2_{xij}(W-1)W}\sum_{l=1}^{W}\sum_{k=1}^{W} x_{(i+l)(j+k)}x_{(i+l-1)(j+k)}, \quad \rho_{yij} = \frac{1}{\sigma^2_{xij}(W-1)W}\sum_{k=1}^{W}\sum_{l=1}^{W} x_{(i+l)(j+k)}x_{(i+l)(j+k-1)}, \qquad (4)$$

where $W$ – length and width of the sliding window.

The table shows the results of the internal parameters using pseudogradient method and the assumption that the estimates (4) are the realization of subsidiary RF of the model (3). We compared the cases where $r_{ij}$ are close to 0, and when $r_{ij} \sim 0.5$ for $\sigma_x^2 = 1$, $m_{\rho x} = m_{\rho y} = 0.8$, $\sigma^2_{\rho x} = \sigma^2_{\rho y} = 0.001$, $q = 3$, $M_1 = M_2 = 100$

Table 1. Estimates of the correlation parameters

| | $r_{11}$ | $r_{12}$ | $r_{21}$ | $r_{22}$ | $r_{11}$ | $r_{12}$ | $r_{21}$ | $r_{22}$ |
|---|---|---|---|---|---|---|---|---|
| Real value | 0.6 | 0.6 | 0.3 | 0.3 | 0.1 | 0.1 | 0.1 | 0.1 |
| W=5 | 0.35 | 0.35 | 0.35 | 0.35 | 0.41 | 0.41 | 0.41 | 0.41 |
| W=9 | 0.33 | 0.33 | 0.33 | 0.33 | 0.37 | 0.37 | 0.37 | 0.37 |
| W=15 | 0.27 | 0.27 | 0.27 | 0.27 | 0.27 | 0.27 | 0.27 | 0.27 |
| Pseudogradient | 0.56 | 0.56 | 0.31 | 0.34 | 0.12 | 0.14 | 0.19 | 0.19 |

The analysis presented in the table shows that the estimates obtained by using the pseudogradient algorithm are 3-5 times more accurate than the estimates in the sliding window.

## Image restoration

We now consider the problem of reconstructing the damaged portion of the image. For simplicity, we rewrite the model (3) as follows:

$$x_{ij} = \rho_{ij}x_{(i-1)j} + \rho_{ij}x_{i(j-1)} - \rho_{ij}\rho_{ij}x_{(i-1)(j-1)} + \xi_{ij}, \qquad (5)$$

where $\rho_{ij} = \tilde{\rho}_{ij} + m_\rho$ - base field correlation coefficients for rows and columns; $\{\xi_{ij}\}$ - Gaussian RV with zero mean and unit variance;

$$\tilde{\rho}_{ij} = r\tilde{\rho}_{(i-1)j} + r\tilde{\rho}_{i(j-1)} - r^2\tilde{\rho}_{(i-1)(j-1)} + \varsigma_{ij},$$

where $r$ - correlation coefficient for the row and the column for base RF; $m_\rho$ - the average value of the correlation of the base RF; $\{\varsigma_{ij}\}$ - RV with a Gaussian distribution, zero mean and unit variance.

We will carry out restoration by estimating correlation coefficients for rows and columns based on the model (2). After that, the damaged area will use this model for prediction. In a similar algorithm based on a doubly stochastic model, first estimation of the field correlation parameters is carried out, and then the estimates are used at restoration of the damaged area. Thus the error obtained at the first realization increases in the second one.

To reduce the error, it is needed to estimate the parameter field in the vicinity of the damage. We use a special case of the vector Kalman filter for this task:

$$\begin{pmatrix}\hat{x}\\ \hat{\rho}\end{pmatrix} = \begin{pmatrix}x_э\\ \rho_э\end{pmatrix} + PV_n^{-1}\begin{pmatrix}1\\ 0\end{pmatrix}(z - x_э), \qquad (6)$$

where $z$ - the corrupted signal; $P = P_э(E + V_n^{-1}P_э)^{-1}$; $P_э = P_{э1} + P_{э2} + P_{э3} + V_e$.

It should be noted that in the formula (6) the estimation is carried out at each point *(i, j)* of the RF. At the same time $P_{э1}, P_{э2}, P_{э3}$ - the contributions of the respective elements, such that:

$$P_{э1} = \begin{pmatrix}\rho_0 + \hat{\rho}_{i,j-1} & \hat{x}_{i,j-1}\\ 0 & r\end{pmatrix} P_{пред} \begin{pmatrix}\rho_0 + \hat{\rho}_{i,j-1} & \hat{x}_{i,j-1}\\ 0 & r\end{pmatrix}^T, \qquad P_{э2} = \begin{pmatrix}\rho_0 + \hat{\rho}_{i-1,j} & \hat{x}_{i-1,j}\\ 0 & r\end{pmatrix} P_{пред} \begin{pmatrix}\rho_0 + \hat{\rho}_{i-1,j} & \hat{x}_{i-1,j}\\ 0 & r\end{pmatrix}^T,$$

$$P_{э3} = \begin{pmatrix}-(\rho_0 + \hat{\rho}_{i-1,j-1})^2 & \hat{x}_{i-1,j-1}\\ 0 & -rr\end{pmatrix} P_{пред} \begin{pmatrix}-(\rho_0 + \hat{\rho}_{i-1,j-1})^2 & \hat{x}_{i-1,j-1}\\ 0 & -rr\end{pmatrix}^T,$$

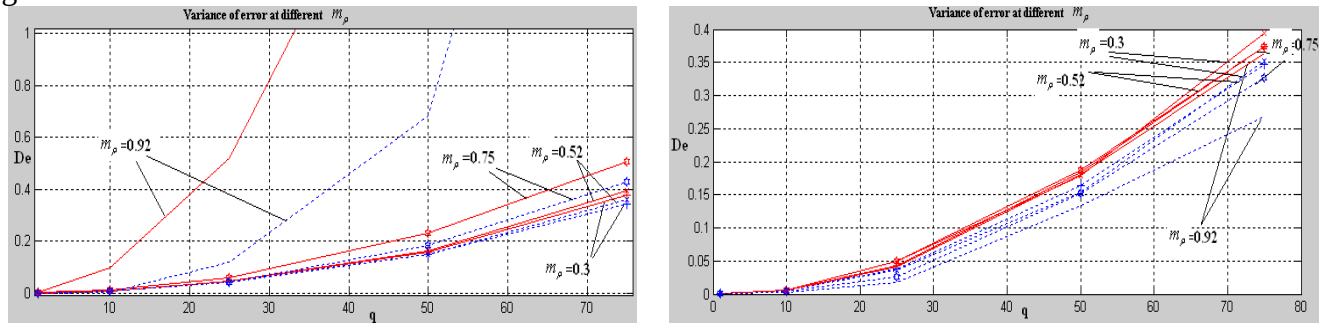
where the estimate of $r$ obtained by pseudogradient algorithm:

$$r = \min_{r_0 \pm \Delta r}\{(\hat{x}(r_{0i} \pm \Delta r_i) - z)^2\},$$

where $\hat{x}(\ldots)$ - implementation of the model (5) at different $r$.

Thus, the filter (6) allows to measure $\rho_{ij}$ in point $(i_0 - 1, j_0 - 1)$. Then you can restore the field correlation parameters $\rho_{ij}$ in the damaged area, and then the corrupted elements $x_{ij}$.

Figure 1 shows the dependence of the error variance reduction with the simplest model (2) and the mixed model (5). By increasing the size of the damage area the error increases, the smallest error is possible to achieve when restoring the mixed model, at that the higher r the smaller the error. In the case of a rapid change in the field of correlation parameters, i.e. $r \approx 0.5$ the variance of the reconstruction error accumulates rapidly with increasing damage size $q$ for the case of the main field with a high correlation. At r = 1 the best estimates are obtained at large $m_\rho$. However, regardless of the value of $r$, pseudogradient algorithm in conjunction with the Kalman filter at the same parameters gives the best estimates.




a) r = 0.5 b) r = 0.9

Fig. 1 The dependence of the error variance versus the size of the damages at different parameters of the model: the dark - recovery with a simple AR model, light - pseudogradient and filter

Fig. 2 shows the results of image reconstruction.

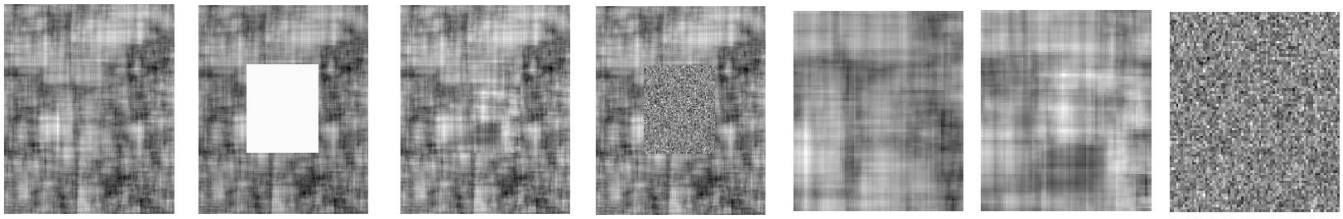


Fig. 2: Restoration of the image. From left to right: original image, the damaged image, restoration by the mixed method, recovery with a simple AR model, the damaged area: in a mixed method, in the simplest AR model.

We see from Fig. 2 that the recovery using a combination of pseudogradient and filter for model (3) are much closer to real image than the implementation of the simplest model (2). This is also confirmed by the graphs in Fig. 1.

## Conslusion

Thus, in this paper we propose a new method for image restoration based on a complex estimation of the parameters of a doubly stochastic model. Moreover, for such recovery, a combination of pseudogradient algorithm for estimating the rate of the correlation parameters change, and nonlinear Kalman filtering, which gives an accurate estimation of image parameters in the vicinity of the damaged fragment. Studies have shown high accuracy of the approaches to restoration of various fragments of two-dimensional images.

## References

1. Vasiliev KK, Tashlinsky AG, VR Krasheninnikov, Statistical analysis of sequences of multidimensional images // High Tech, 2013, №5, p. 5-11
2. Vasiliev KK, Dement'ev VE Autoregressive model of multidimensional images // High Tech, 2013, vol. 14, №15, p. 12-15
3. K.. Vasiliev, V. Dementiev, N. Andriyanov Twice stochastic models of images The 11th International Conference «PATTERN RECOGNITION and IMAGE ANALYSIS: NEW INFORMATION TECHNOLOGIES» (PRIA-11-2013) - Samara, The Russian Federation, 2013. vol. 1, p. 334-337.
4. Andriyanov NA Vasiliev KK, Dement'ev VE Identification of the parameters of doubly stochastic autoregressive models of stochastic processes // 16-th International Conference "Digital Signal Processing and its application - DSPA-2014", Moscow, Russia, reports, vol. 1, p. 109 - 113.
5. Vasiliev K.K., Dement'ev V.E., Andriyanov N.A. Parameter Estimation of doubly stochastic random fields // Radiotekhnika. - 2014. - №7, p. 103-106
6. Gorban AN, SV Makarov, Russia AA An iterative method of principal components for tables with spaces // Third Siberian Congress on Industrial and Applied Mathematics (INPRIM-98), June 22-27, 1998, Abstracts. Part 5. Novosibirsk: Publishing House of the Institute of Mathematics, 1998.
7. Larionov I.B. Clustering matrices with gaps as a method of recovery of graphical information // Mathematical Structures and Modeling, 2009, vol. 20, p. 97-106.
8. Vasiliev KK, VR Krasheninnikov Statistical analysis of multidimensional images / KK Vasiliev, VR Krasheninnikov. - Ulyanovsk Ulyanovsk State Technical University, 2007 - 170 p.

# BIOGRAPHY OF SPEAKERS

***Vasiliev Konstantin Konstantinovich**, was born in 1948. He graduated from the Leningrad Electrotechnical Institute, "Radio-electronic devices" in 1972 His focus research are statistical analysis of random processes and fields. In 1975 he defended his thesis, and in 1985 - PhD, Honored Worker of Science and Technology of the Russian Federation, Corresponding Member of the Academy of Sciences of the Republic of Tatarstan. He is the chair of "Telecommunications" department in Ulyanovsk State Technical University. [e-mail: vkk@ulstu.ru]*

**Dementiev Vitaly Eugenievich**., *was born in 1982 He graduated from the Ulyanovsk State Technical University, majoring in "Applied Mathematics" in 2004. In 2007 he defended his thesis on "Detection of signals on multispectral images." Line of research I statistical analysis of random processes and fields. Associate Professor of "Telecommunications" department in Ulyanovsk State Technical University. [e-mail: dve@ulntc.ru]*

***Andryianov Nikita Andreevich**, was born in 1990. In 2013 he graduated from the Faculty of Radio Engineering, Ulyanovsk State Technical University. Graduate student of "Telecommunications" department in Ulyanovsk State Technical University. He has articles in the field of statistical signal processing methods. [e-mail: nikita-and-nov@mail.ru].*

# Automatic Image Analysis Algorithm for Quantitative Assessment of Breast Cancer Estrogen Receptor Status in Immunocytochemistry

Dobrolyubova D.A., Kravtsova T.A., Artyukhova O.A., Samorodov A.V.
Chair for Biomedical Technical Systems
Bauman Moscow State Technical University
Moscow, Russian Federation
daria.dobrolyubova@mail.ru, vigilax@rambler.ru, artyukhova@bmstu.ru, avs@bmstu.ru

***Abstract*—The paper presents an algorithm for quantification the degree of receptor expression to steroid hormones by automatic analysis of microscope images of immunocytochemical specimens. During experiments a high correlation between the results of the automatic analysis and visual expert assessment was shown and the possibility to apply the proposed algorithm to automate immunocytochemical analysis was confirmed.**

***Keywords—automatic image analysis; immunocytochemistry; estrogen receptor status quantification; color deconvolution***

## I. Introduction

The importance of semi-quantitative and quantitative evaluation of specific tumor markers (such as estrogen and progesterone receptors, Ki-67, and others) is emphasized by many authors [1-4]. The most popular methods for this task are immunocytochemistry (ICC) and immunohistochemistry (IHC). Specimen preparation for ICC and IHC often includes double or triple staining, which greatly complicates the task of visual assessment, especially in the cases when the dyes are co-localized and cannot be fully spectrally separated and, therefore, can hardly be distinguished visually [5-7].

Cytological identification of breast cancer nuclear markers which characterize the degree of receptor expression to steroid hormones – estrogen receptor (ER) status of breast cancer, – refers to the case of double staining. Nuclei of tumor cells containing receptors are stained with chromogen DAB, which gives them a brown color, and then all the nuclei are stained with hematoxylin, which gives them a blue color. Interpretation of ICC and IHC specimens include determination of the degree of estrogen receptor expression based on the intensity of 3,3'-Diaminobenzidine (DAB) staining, which require the gradation of brown hues. Subjectivity of color perception and inability to visually separate the two components of the nuclei color intensity may lead to errors. In general, this could not provide the standardization of this analysis and reproducibility of visually obtained results in different laboratories. Automation of image analysis could improve such situation [3].

Nowadays there are numerous examples of image analysis automation in IHC. But corresponding software could not be adapted for ICC specimens. In this paper, we propose an algorithm for quantification of the ER status of breast cancer by automatic analysis of digital microscope images of ICC breast specimens obtained with fine-needle aspiration biopsy.

## II. Automatic Image Analysis Algorithm Description

The algorithm designed for quantitative assessment of breast cancer estrogen receptor status includes four steps: image pre-processing, color deconvolution, segmentation, and evaluation of diagnostic Allred score and H-Score (a schematic diagram of the first three image processing steps is presented in Fig. 1).

### *A. Image pre-processing*

Image preprocessing is an important step of the algorithm, since its purpose is to eliminate the influence of image capture conditions (camera settings, light level, etc.) on image color characteristics. Standardization of color rendering is necessary to ensure reproducible assessment of the degree of expression, based on the nuclei staining intensity estimation.

The first step of the pre-processing is color correction. ICC specimen has transparent background due to the procedure of preparation. So the true color of image background corresponds to the color temperature of light source, which could be distorted by the camera setting. The aim of color correction in the pre-processing stage is to eliminate the influence of the color temperature of the light source by the color transformation resulting in the grayscale background.

To select the proper color correction algorithm comparative study of the three such algorithms was carried out. These algorithms were our modification of Gray World [8], White Patch [8], and iterative color correction algorithm [9]. For this study 150 images of ICC specimens were used: 50 images with low color temperature, 50 images with high color temperature, and 50 images with color temperature near 6500 K. All images were captured using camera PixeLINK PLB873 (image resolution of 1600×1200 pixels, sampling 0.16 μm/pixel). Color correction algorithms were compared using the criterion *C*, reflecting the closeness

of average intensity values $\bar{I}_R$, $\bar{I}_G$, and $\bar{I}_B$ in RGB color model to each other. The smallest value of $C$ corresponds to the best balance of RGB channels:

$$C = \sqrt{(\bar{I}_R - \bar{I}_G)^2 + (\bar{I}_B - \bar{I}_G)^2} .$$

The experimental results are presented in Table I. The smallest value of the criterion corresponds to the Gray World algorithm. This algorithm makes the following pixel-wise color transformation:

$$\begin{cases} \tilde{I}_i^R = (\bar{I}_{BACK}^R / \bar{I}_{BACK}^G) \cdot R_i, \\ \tilde{I}_i^G = G_i, \\ \tilde{I}_i^B = (\bar{I}_{BACK}^B / \bar{I}_{BACK}^G) \cdot B_i, \end{cases}$$

where $\tilde{I}_i^R$, $\tilde{I}_i^G$, $\tilde{I}_i^B$ are channel intensities of corrected image and $\bar{I}_{BACK}^R$, $\bar{I}_{BACK}^G$, $\bar{I}_{BACK}^B$ – the average values of background intensity in the channels R, G, B respectively.

The second pre-processing step is the correction of image brightness histogram, to make the image background white. The brightness correction is also performed in the RGB color space. We use the same transformation for each channel to avoid color distortion. An example of the image pre-processing results is shown in Fig. 1 (Step 1).

### B. Color deconvolution

The necessity of stains separation is provided by the fact that due to the double staining procedure hematoxylin dye could influence the result of DAB brightness evaluation. The proposed algorithm uses the color deconvolution method in the case of two dyes [10]. The color deconvolution method is based on the Beer–Lambert law

$$I = I_0 \cdot \exp(-k_1 c_1 \Delta l - k_2 c_2 \Delta l),$$

where $I$ and $I_0$ are the intensities of incident and transmitted light respectively; $c_1$ and $c_2$ are relative concentrations of H and DAB, $k_1$ and $k_2$ are absorption coefficients of hematoxylin and DAB, and $\Delta l$ denotes the thickness of the absorbing layer (here taken as unit value).

To find the relative dyes concentrations $c_1$ and $c_2$ the optical densities $OD_i^R$, $OD_i^G$, $OD_i^B$ in RGB channels are calculated using intensity values $I_i^R$, $I_i^G$, $I_i^B$ of each of $i$-th image pixel:

$$\begin{cases} OD_i^R = \lg(255/I_i^R), \\ OD_i^G = \lg(255/I_i^G), \\ OD_i^B = \lg(255/I_i^B). \end{cases}$$

TABLE I. COMPARISON OF COLOR CORRECTION ALGORITHMS

| Color Correction Algorithm | Mean Value of Criterion *C* |
|---|---|
| Modified Gray World | 0.0116 |
| White Patch | 0.0172 |
| Algorithm with iterative procedure | 0.0122 |

The dependence of optical densities values on dyes concentrations is presented by the system of linear equations, which is solved by the least squares method:

$$\begin{cases} OD_i^R = k_1^R \cdot c_1^i + k_2^R \cdot c_2^i, \\ OD_i^G = k_1^G \cdot c_1^i + k_2^G \cdot c_2^i, \\ OD_i^B = k_1^B \cdot c_1^i + k_2^B \cdot c_2^i. \end{cases}$$

The estimation of absorption coefficients was carried out on a training set of images, for which the areas of nuclei stained only by hematoxylin and both by hematoxylin and DAB were segmented manually after image pre-processing. An example of relative concentration profiles for DAB and hematoxylin are shown in Fig. 1 (Step 2).

### C. Segmentation

Segmentation is carried out to allocate areas of the image related to the nuclei with steroid hormone receptors and the nuclei without receptors. For all cells nuclei segmentation Niblack adaptive binarization algorithm was applied to a grayscale image obtained from the original one using the standard RGB to gray transform. The sliding window size for the algorithm equals three times the mean cell size. Allocation of DAB-stained nuclei is performed by global thresholding of the DAB relative concentration profile. Examples of masks for all nuclei and for DAB-stained nuclei after additional morphological processing, as well as the segmented image, are shown in Fig. 1 (Step 3).

### D. Evaluation of diagnostic scores

On the fourth step of the algorithm two types of scores characterizing ER-tumor status are calculated: Allred score [11] and H-Score [12].

According to Allred quantification method, the Total Score (TS) is calculated as the sum of Proportion Score (PS) and Intensity Score (IS). PS corresponds to the proportion of cells with DAB-stained nuclei, and IS reflects the intensity of their staining. PS takes its value within the range from 0 to 5, and IS – from 0 to 3. We proposed the approximation of Allred PS and IS semi-quantitative scales by power functions:

$$PS = 5 \cdot \sqrt[3]{N_{DAB} / (N_H + N_{DAB})},$$
$$IS = 3 \cdot (C_{DAB} - 1)^5 + 3,$$

where $N_H$, $N_{DAB}$ is number of image pixels in the area of hematoxylin -stained and DAB-stained nuclei respectively. $C_{DAB}$ is the weighed mean of relative DAB concentration estimated over complete specimen. Corresponding plots are shown in Fig. 2.

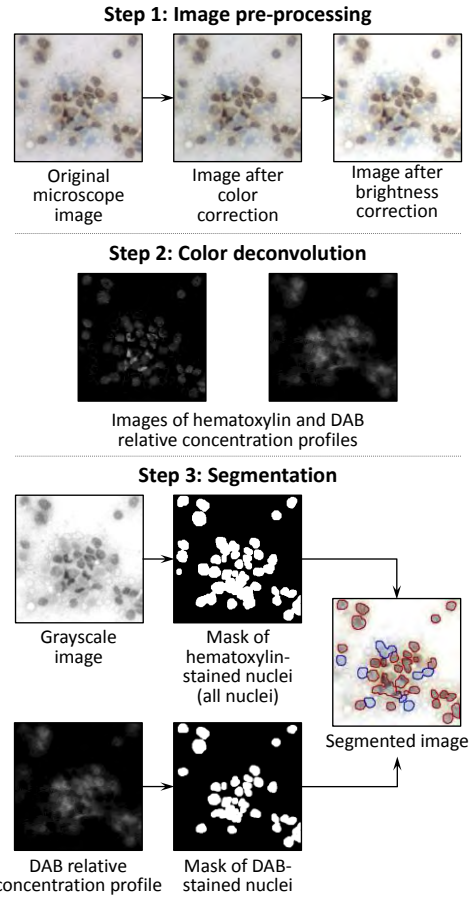



Fig. 1. Schematic diagram of the first three microscopic image processing steps. Step 1 – pre-processing (color and brightness correction); Step 2 – color deconvolution; Step 3 – image segmentation.

During the ICC specimen image analysis the proportion of cells with DAB-stained nuclei is calculated on the base of the segmentation results. $C_{DAB}$ for complete specimen is estimated by the histogram of DAB relative concentration profile considering DAB-stained nuclei mask as

$$C_{DAB} = \sum_k N_{DAB}^k \cdot C_{DAB}^k \Big/ \sum_k N_{DAB}^k ,$$

where $N_{DAB}^k$ – the number of image pixels with concentration $C_{DAB}^k$.

The intensity of staining is calculated using $C_{DAB}$. The range of DAB relative concentration values was obtained from the training set of images.

H-Score has a range of values from 0 to 300 and is calculated as follows:

$$HS = \frac{100}{N_{DAB} + N_H} \cdot \sum_k N_{DAB}^k \cdot IS_k ,$$

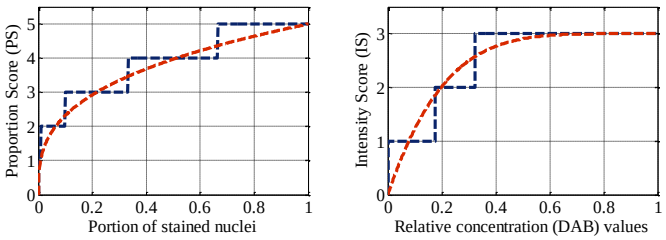



Fig. 2. Approximation of Allred's PS (left) and IS (right) semi-quantitative scales

where $N_{DAB}^k$ – the number of image pixels corresponding to the relative DAB concentration $k$, $IS_k$ – staining intensity determined by the approximating curve shown in Fig. 2 for the relative DAB concentration $k$.

## III. Experimental Results

A comparative study of the proposed method and visual ICC specimens examination was conducted. Images of 24 ICC specimens were captured with the camera PixeLINK PLB873 (image size 1600×1200 pixels, sampling 0.16 μm/pixel). Images cover the entire informative regions on ICC specimen. Selected specimens represent the entire range of possible Allred score and H-Score values. These images were analyzed by medical expert in accordance to the methods of Allred score [11] and H-Score [12] calculation, and using the proposed algorithm. Decision on positive/negative ER-status was made based of the recommendations given in [13]. The results are shown in Table II.

The Spearman rank correlation coefficients of the results of visual and automatic evaluation of Allred score (TS) and H-Score (HS) for these 24 ICC specimens were 0.96 ($P < 0.001$) and 0.99 ($P < 0.001$) respectively. Decision on ER-status coincided for all ICC specimens. The decisions on ER-status for specimen № 10 defined by two evaluation systems (Allred score and H-Score) are not the same. This fact represents the complication of ER-status determination for cases with weak staining of small number of cells and does not indicate incorrect performance of the proposed algorithm.

## IV. Conclusion

An algorithm for evaluation of the diagnostic Allred score and H-Score for immunocytochemical estimation of breast cancer ER status is proposed. It is based on the steps of pre-processing making the proposed automatic image analysis algorithm invariant to the conditions of image capture, of color deconvolution separating hematoxylin and DAB stains, of segmentation utilizing fast local and global thresholding, and quantitative assessment of Allred score and H-Score by means of the proposed scores approximation. Comparative experimental study has shown high correlation between the results of visual and automatic assessment of these diagnostic scores and the possibility to apply the proposed algorithm to automate immunocytochemical analysis.

TABLE II. COMPARISON OF VISUAL EXPERT ASSESSMENT AND AUTOMATIC EVALUATION OF SCORES

| Number of ICC specimen | Visual Expert Assessment | | | Automatic Analysis | | |
|---|---|---|---|---|---|---|
| | ***TS*** | ***HS*** | ***ER-Status*** | ***TS*** | ***HS*** | ***ER-Status*** |
| 1 | 0 | 0 | - | 0.0 | 0 | - |
| 2 | 0 | 0 | - | 0.0 | 0 | - |
| 3 | 0 | 0 | - | 0.0 | 0 | - |
| 4 | 0 | 0 | - | 0.0 | 0 | - |
| 5 | 0 | 0 | - | 0.0 | 0 | - |
| 6 | 0 | 0 | - | 0.0 | 0 | - |
| 7 | 0 | 0 | - | 0.0 | 0 | - |
| 8 | 0 | 0 | - | 0.0 | 0 | - |
| 9 | 0 | 0 | - | 0.0 | 0 | - |
| 10 | 2 | 2 | -/+ | 2.5 | 2 | -/+ |
| 11 | 3 | 7 | + | 3.7 | 13 | + |
| 12 | 3 | 7 | + | 3.8 | 12 | + |
| 13 | 4 | 20 | + | 5.3 | 66 | + |
| 14 | 5 | 43 | + | 5.7 | 83 | + |
| 15 | 6 | 99 | + | 6.5 | 128 | + |
| 16 | 6 | 192 | + | 7.2 | 205 | + |
| 17 | 6 | 104 | + | 6.1 | 105 | + |
| 18 | 7 | 125 | + | 6.5 | 145 | + |
| 19 | 7 | 134 | + | 6.8 | 170 | + |
| 20 | 7 | 164 | + | 7.2 | 210 | + |
| 21 | 7 | 171 | + | 7.0 | 188 | + |
| 22 | 8 | 228 | + | 7.0 | 192 | + |
| 23 | 8 | 192 | + | 7.4 | 222 | + |
| 24 | 8 | 156 | + | 7.1 | 188 | + |

## REFERENCES

[1] O. Brouckaert, R. Paridaens, G. Floris, E. Rakha, K. Osbome, and P. Neven, “A critical review why assessment of steroid hormone receptors in breast cancer should be quantitative,” Ann. Oncol., vol. 24, no. 1, pp. 47-53, July 2012.

[2] R.A. Walker, “Quantification of immunohistochemistry-issues concerning methods, utility and semiquantitative assessment I,” Histopathology, vol. 49, no. 3, pp. 406-410, October 2006.

[3] C.R. Taylor, R.M. Levenson, “Quantification of immunohistochemistry-issues concerning methods, utility and semiquantitative assessment II,” Histopathology, vol. 49, no. 4, pp. 411-424, October 2006.

[4] S. Di Cataldo, E. Ficarra., E. Macci, “Computer-aided techniques for chromogenic immunohistochemistry: status and directions,” Comput. Biol. Med., vol. 42, no. 10, pp. 1012-1025, October 2012.

[5] R. Levenson, P.J. Cronin, K.K. Pankratov, “Spectral imaging for brightfield microscopy,” in Spectral imaging: instrumentation, applications, and analysis II, R.M. Levenson, G.H. Bearman, and A. Mahadevan-Jansen, Eds. Proceedings of the SPIE, vol. 4959, pp. 27-33, 2003.

[6] R.M. Levenson and R. Mansfield, “Multispectral imaging in biology and medicine: slices of life,” Cytometry A, vol. 69, no. 8, pp. 748-758, August 2006.

[7] C.M. van der Loos, “Multiple immunoenzyme staining: methods and visualizations for the observation with spectral imaging,” J. Histochem. Cytochem, vol. 56, no. 4, pp. 313-328, April 2008.

[8] E.Y. Lam and G.S.K. Fung, “Automatic white balancing in digital photography,” in Single-sensor imaging: methods and applications for digital cameras. Boca Raton: CRC Press, 2008, pp. 267-294.

[9] J. Huo, Y. Chang, J. Wang, and X. Wei, “Robust automatic white balance algorithm using gray color points in images,” IEEE Transactions on Consumer Electronics, vol. 52, no. 2, pp. 541-546, May 2006.

[10] A.C. Ruifrok and D.A. Johnston, “Quantification of histochemical staining by color deconvolution,” Anal Quant Cytol Histol, vol. 23, no. 4, pp 291-299, August 2001.

[11] D.C. Allred, “Assessment of prognostic and predictive factors in breast cancer by immunohistochemistry,” Connection, vol. 9, pp. 4-5, June 2006.

[12] K.S. Jr. McCarty, E. Szabo, J.L. Flowers et al., “Use of a monoclonal anti-estrogen receptor antibody in the immunohistochemical evaluation of human tumors,” Cancer Research, vol. 46, no. 8, pp. 4244-4248, August 1986.

[13] M.E. Hammond, D.F. Hayes, M. Dowsett et al., “American society of clinical oncology/college of american pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer,” Arch Pathol Lab Med, vol. 134, pp. 907-922, July 2010.

# Blood Pressure Rhythm Estimation based on Shape Patterns for Analytic Spectra*

Vjacheslav Antsiperov
and Gennady Mansurov
Kotel'nikov Institute of Radio
Engineering and Electronics of RAS
Moscow, Russia
Antciperov@cplire.ru

Basil Bonch-Bruevich
Joint Stock Company NEUROCOM
Moscow, Russia
v.bonch-bruevich@neurocom.ru

*Abstract*— **The report presents the latest results of developing new methods, based on Multiscale Correlation Analysis for bio-medical signals processing. It is shown that in case a signal has the form of repeated wave pulses, MCA naturally leads to the technique previously introduced by the authors and called the analytical spectra. It is also demonstrated, that detecting typical MCA fragments by appropriate patterns, gives the possibility to significantly improve the repetition estimation. The report discusses the application of these methods to the problems of the blood pressure rhythm monitoring. In relation to monitoring BP some characteristics of the method are under discussion.**

***Keywords— bio-medical signals processing; analytical spectra; rhythm estimation; pattern recognition; blood pressure monitoring;***

## I. INTRODUCTION

The great progress of the modern technological innovation in the field of computer, communication and multimedia gives hope for similar impressive advances in new medical devices and instruments generation. Therefore we observe now an unprecedented growth of interest in such instrumentation, especially in compact mobile medical devices (gadgetry). Such modern tools will provide undoubtedly principally new and potentially unlimited health services to the wide range of people.

Medicine becomes now a considerably data-rich science. The power of above mentioned technologies enables us to measure and process a lot of signals, parameters and other data that characterize the patient functional status. It will significantly improve the quality of treatment, help in diagnosis and patient monitoring, and allow some patients to stay home instead of hospital bed thanks to remote monitoring possibility. It makes true the opinion of Norbert Wiener – the father of cybernetics – who wrote in [1]: "As far back as four years ago, the group of scientists and myself had already become aware of the essential unity of the set of problems centering about communication, control, and statistical mechanics, whether in the machine or in living tissue." So Wiener's vision today is reality.

The authors acknowledge financial support from RFBR grant № 14-07-00496 A.

After the new hardware comes the need for appropriate software, which is equally important as it helps in linking the functionality of the parts of hardware, grows over and over. It is important to outline, that it concerns not only the system software such as drivers, interfaces and utilities, but also the special purpose applications, designed for data analysis, automatic diagnostics, forecasting, interoperability scenarios, etc. It should be stressed that the most specific of special purpose applications are those, which are destined for primary processing of acquired data (digital signal processing, DSP). Specificities, special features of such processing determine the most adequate information extraction and, in *ultima analysi*, the medical device specificity.

It should be also noted that, despite impressive progress in the development of microprocessor technology, and the expansion of programming mobile devices opportunities for the programmability of personal computers, the task of developing a robust, adequate and effective DSP algorithms and methods is still in the field of attention.

In this report we present the latest results of developing new methods of bio-medical signals processing, based on Multiscale Correlation Analysis (MCA) [2, 3]. We show that in case of signals having a form of repeated wave pulses (motives), MCA leads naturally to the technique previously introduced by the authors and called the analytical spectra (AS) [4,5]. We also demonstrate, that detecting typical MCA fragments by appropriate patterns, gives us the possibility to significantly improve the periodicity estimates. In the center of discussion is the application of these methods to the problems of the arterial blood pressure (ABP) rhythm monitoring automation.

## II. ABP AND ECG: SIMILARITIES AND DIFFERENCES

The problem of ABP rhythm monitoring is very similar to the well-known problem of instantaneous heart rate estimation from electrocardiogram (ECG) waveforms [5], since both signals are due to the cardiac cyclic contractions. However, since ABP and ECG reflect different aspects of the cardiovascular system operating – ECG reflects the heart electrical activity, and the ABP – mechanical response to this excitation, their repeated waveforms are markedly different

(see Fig. 1), and therefore the estimating procedures can differ essentially.

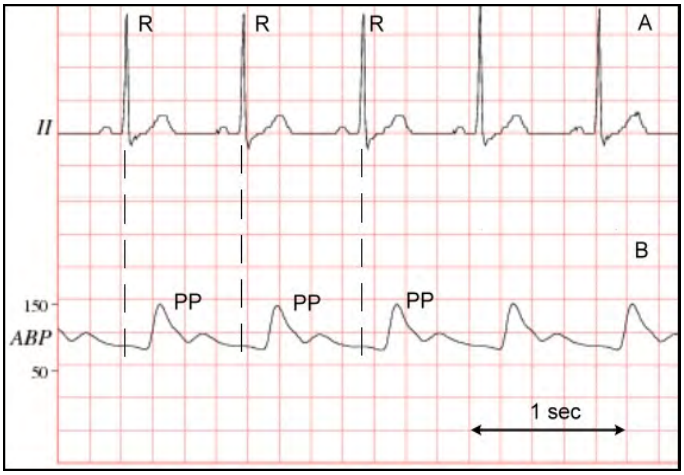



Fig. 1. Normal arterial blood pressure waveform (B) and its relation to the ECG wave (A). Record mimic2db/a40008 from MIMIC II Waveform DB, v2 [6]; PhysioNet ATM screenshot

The complete ABP waveform results from ejection of blood from the left ventricle into the aorta during systole, followed by peripheral arterial runoff of this stroke volume during diastole (Fig.1). The systolic components follow the ECG wave (with characteristic R–peak) and consist of a steep pressure upstroke, peak (PP), and decline and correspond to the period of left ventricular systolic ejection. Note that the systolic upstroke of the artery pressure trace does not appear for 120 to 180 milliseconds after inscription of the ECG R–peak. This interval reflects the sum of times required for spread of left ventricular contraction, aortic valve opening, left ventricular ejection, etc.

Thanks to such a feature of the ECG signal as sharp R–peaks (or rather whole QRS complexes) already in the early years of heart rate detection, an algorithmic structure was developed that is now shared by many algorithms [7]. As a rule it is divided into a preprocessing and rate estimation stage, including peak detection and decision logic. Due to the fact that ABP waveform does not contain as well expressed peak as ECG does, moreover, its pressure peak (PP) can often be greatly deformed (depending on measuring conditions), using algorithms similar in structure with QRS detecting generally does not bring success [8,9].

## III. MCA and Analytical Spectra for Repetitive Wave Pulses, in particular ABP

Fortunately, there are other effective approaches to solve the aforesaid problem. We have shown previously [5] that our Multiscale Correlation Analysis (MCA) method [2,3] is a very effective approach to bio-medical signal processing if it contains a set of repeating events (motives). The MCA basic tool is a symmetric estimation $R_x(\tau|t)$ of a signal two-dimensional autocorrelation function (ACF), which in the simplest case has the form:

$$R_x(\tau \mid t) = \frac{1}{\tau}\int_{t-\tau/2}^{t+\tau/2} x(t'+\tau/2)\,x(t'-\tau/2)\,dt' \qquad (1)$$

where $x$ is one-dimensional source signal, $t$ – current time (the time moment under analysis), $\tau$ – variable (multi) scale. An example of multi-scale ACF (1) for a real ABP signal is shown in Fig. 2.

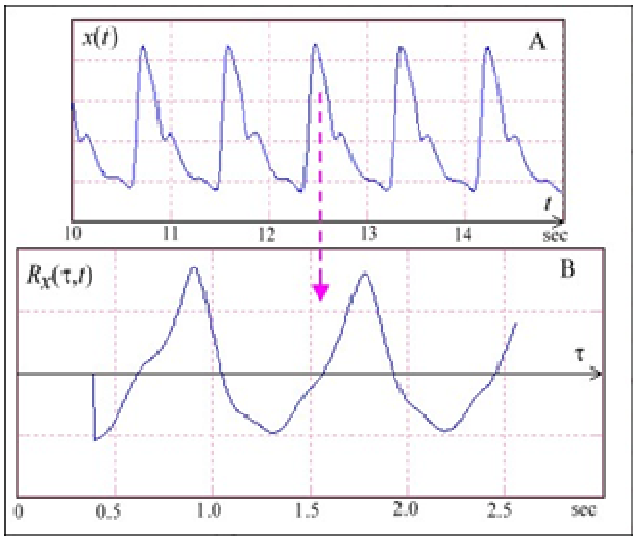



Fig. 2. Real ABP signal fragment (A) and its ACF (B) for some time moment highlighted by arrow

Figure clearly demonstrates that ACF (1) has side peaks corresponding to the local (for time moment $t$ ) rhythm characteristic times [3] (first peak location corresponds to rhythm period, second – to double period, etc.). So, positions of these peaks can be utilized as rhythm period estimates instead of maxima of source signal. The advantage of ACF (1) is that it can be formed on the small signal basis equal to doubled maximal scale $\tau$, whose value can be assigned, for example, in 1.5 sec. The traditional ACF estimates usually need a much bigger signal base to avoid peaks smoothing due to the rhythm variability.

So in order to have automatically generated local period estimation, we need to determine reliably the position of ACF (1) side peaks. Our experience has shown that direct peaks detection by means of ACF maximum search is not a robust procedure because autocorrelation (1) is a multimodal function. Much more stable and reliable is the maximum detection by means of the generalized spectrum of ACF (Fig.3). Generalized spectrum $G(\tau|t,\sigma)$ represents the ACF decomposition in window functions (patterns) $W(\tau'|\tau,\sigma)$ having a form of a Morlaix wavelet. In contrast to the classical wavelet analysis we consider as spectral variable window position $\tau$ instead of scale $\sigma$:

$$G(\tau \mid t,\sigma) = \int_0^{\infty} W(\tau' \mid \tau,\sigma) R_x(\tau' \mid t)\,d\tau'$$
$$W(\tau' \mid \tau,\sigma) = \frac{1}{\sqrt{\pi\sigma^2}} \exp(-(\tau'-\tau)^2/\sigma^2)\cos\left(\pi(\tau'-\tau)/\sigma\right) \qquad (2)$$

An example of ACF (1) shown on the background of spectral components { $W(t'|\tau,\sigma)$ } and resulting generalized spectrum $G(\tau|t,\sigma)$ are presented in Fig.3, (A) and (B) consequently.

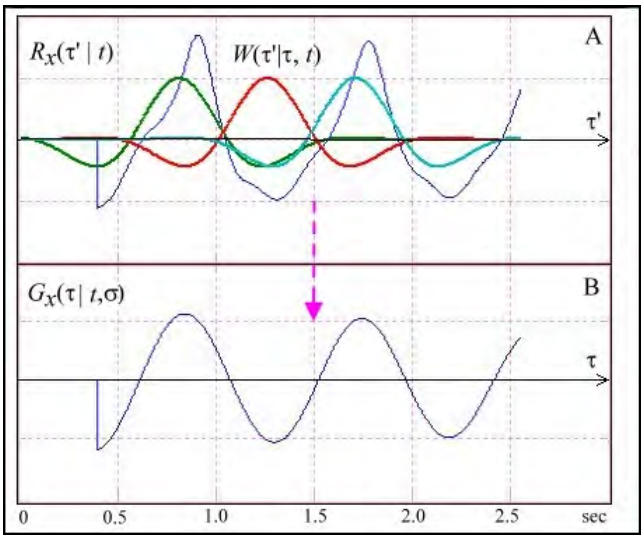



Fig. 3. ACF (Fig.2) with a set of spectral windows (A) and resulting generalized spectrum (B).

Applying twice convolution theorem for the Fourier transform, it is easy in approximation $\tau > \sigma$ to get a different representation of the spectrum (2):

$$G(\tau,\sigma \mid t) = \frac{1}{\tau}\mathrm{Re}\int_{-\infty}^{\infty}\gamma_{Pt}(\nu)\gamma_{Ft}(\nu)\times \quad , \qquad (3)$$
$$\times \exp\left(-\pi^2(\sigma\nu - 1/2)^2\right)\exp\left(2\pi i\tau\nu\right)d\nu$$

where

$$\begin{cases}\gamma_{Ft}(\nu) = \int_0^{\infty} x(t+t')\exp(-2\pi i\nu\, t')dt' \\ \gamma_{Pt}(\nu) = \int_0^{\infty} x(t-t')\exp(-2\pi i\nu\, t')dt'\end{cases} . \qquad (4)$$

are analytic spectra (AS), introduced and discussed in [4].

## IV. Notes about Realization

Numerical realization of the spectrum computing algorithm (3) does not cause any problems. Obviously it is the inverse Fourier transform from the weighted by Gaussian window product of analytical spectra (4), which, in turn, are the result of a Fourier transform of local signal future $x(t+t')$ and local signal past $x(t-t')$ consequently. Using for Fourier transformations the fast algorithms (FFT), makes the whole procedure (3) of the spectrum computation fast as well. In addition, as mention above, the rhythm estimation on the base of generalized spectrum (3) is also robust, with good accuracy, even in a considerable noise conditions.

## References

[1] N. Wiener. “Cybernetics: or Control and Communication in the Animal and the Machine, Camb. Mass. (MIT Press) 1948.

[2] V.E. Antsiperov, Y.V. Obukhov. “Multiscale correlation analysis of real medical and biological signals and their graphical-based representation,” in Proceedings of VIII International scientific conference "Physics and radioelectronics in medicine and ecology FREME’2008", Vladimir-Suzdal, Vol. 1, 2008, pp. 180-184.

[3] V.E. Antsiperov. “Multiscale correlation analysis of nonstationary signals containing quasi-periodic fragments,” in J. of Communications Technology and Electronics Vol. 53, № 1, 2008, pp. 65-77.

[4] V.E. Antsiperov. “The concepts of analytic local past / future signal spectra and their use for constructing and analysis of the bilinear time-frequency representations,” in Proceedings of the 16-th international conference “Digital signal processing and its applications, DSPA-2014”, Moscow, Vol.1, 2014, pp. 113-117.

[5] V.E. Antsiperov, S.A. Nikitov. “Heart rate monitoring based on analytical spectra technique,” in Proceeding of Russian-German (until 2012 – Russian-Bavarian) conference on biomedical engineering, S.Petersburg, 2014, pp. 169-171.

[6] M. Saeed, M. Villarroel, A.T. Reisner, et al. “Multiparameter intelligent monitoring in intensive care II (MIMIC-II): A public-access ICU database,” in Critical Care Medicine, Vol. 39, № 5, 2011, pp. 952-960.

[7] B-U. Kohler, C. Hennig, and R. Orglmeister. “The Principles of Software QRS Detection,” in IEEE Engineering in Med. and Biol., Vol. 21, 2002, pp. 42-57.

[8] S. Carrasco, R. Gonzalez, et al. “Comparison of the heart rate variability parameters obtained from the electrocardiogram and the blood pressure wave,” in Journal of Medical Engineering & Technology, Vol. 22, № 5, 1998, pp. 195-205.

[9] Bing Nan Li, Ming Chui Dong, Mang I. Vai. “On an automatic delineator for arterial blood pressure waveforms,” in Biomedical Signal Processing and Control Vol. 21, № 1, 2010, pp. 76–81.

# Centroid-Based Ensemble Clustering: Algorithms for Hyperspectral Images Segmentation*

V.Berikov
Sobolev Institute of Mathematics SB RAS,
Novosibirsk State University
Novosibirsk, Russia
berikov@math.nsc.ru

I.Pestunov, P.Melnikov
Institute of Computational Technologies SB RAS
Novosibirsk, Russia
pestunov@ict.nsc.ru

G.Gonzalez
Ecole des Ponts ParisTech
Paris, France
guillaumegonz@yahoo.fr

***Abstracts*—A novel method of ensemble clustering for hyperspectral image segmentation is proposed. The basic idea of the method is to use a series of *k*-means algorithms as a preliminary step to reduce the amount of data under analysis. Clustering results on real hyperspectral image demonstrate the efficiency of the proposed algorithms.**

***Keywords—clustering ensemble; hyperspectral image analysis; k-means; centroids***

## I. INTRODUCTION

The purpose of data clustering (unsupervised classification, taxonomy) is to partition a set of objects into homogeneous groups. Clustering algorithms are widely used in various applications, such as finding of similar subsets of genes, automatic classification of internet documents or segmentation of satellite images. There is a number of different approaches in the theory and methods of data clustering [1]: decomposition of distribution mixture models, hierarchical methods, finding of "centers of gravity" among data, graph-based methods etc.

One of the rapidly developing fields of research in data analysis is ensemble clustering [2]. It aims to get a stable solution to the clustering problem by combining several clustering partitions. There is no universal method of grouping. Each clustering algorithm has its "niche" i.e. specific data structures or distributions for which this algorithm, in a sense, gives the best solution. The reasonable way is to produce the collective result using a number of different algorithms. Each algorithm in the ensemble should supplement others i.e. compensate its disadvantages. On the other hand, every single clustering algorithm depends on various parameters or installation-specific settings (e.g. initial centroids coordinates, types of similarity measure or desired number of clusters) which are usually hard to pick. From this point of view, an ensemble approach allows one to get a stable consensus decision based on a number of partial results of clustering obtained with a collection of different parameters/settings [3].

In this paper, we consider a problem of hyperspectral images segmentation which is one of the research areas appealing for enhanced clustering techniques. The peculiarities of hyperspectral images consist in the following [4]: huge amount (up to millions) of data objects (pixels); large number (up to several hundred) of features (image channels); the presence of "noise" (in data objects and/or features); features correlation.

Existing image segmentation algorithms do not entirely take into account all these conditions [5]. In particular, available ensemble clustering methods are too time-consuming due to the label correspondence problem (i.e., class labels are assigned to objects in an arbitrary way, thus all permutations of labels should be considered) or the need for pairwise comparison of data objects.

In this paper we propose an efficient method of ensemble clustering of hyperspectral images. Its idea is to use preliminary data compression instead of exhaustive analysis of label permutations or huge amount of pairwise comparisons. Data compression step consists of a series of *k*-means clustering algorithms partitioning initial data into a large number of groups. At the same time, the partitioning results are considered as base elements of clustering ensemble: the obtained group centroids (prototypes) are used for the ensemble clustering. We have developed two variants of the method. The first is based on simple averaging of appropriate centroid coordinates over all clustering variants. The second one performs ensemble clustering by the analysis of co-association matrix constructed over pairs of obtained prototypes.

The rest of the paper is organized as follows. Necessary notations and mathematical background is introduced in Section II. The proposed algorithms of ensemble segmentation of hyperspectral images are described in Section III. Experimental results are provided in Section IV. Finally, the work is summarized in the conclusion.

This work was partially supported by the Russian Foundation for Basic Research, projects 14-07-00249a, 13-07-12202-ofi_m and by V. Potanin Foundation.

## II. BASIC NOTIONS

Let us consider a set of data objects $s=\{o_1,...,o_N\}$ characterized by a collection of variables $X_1,...,X_d$. We denote the vector of these variables for an object o by $x=(x_1,...,x_d)\in R^d$ where $x_j=X_j(o)$, $j=1,...,d$. In image analysis problems, an object corresponds to a pixel, and a variable represents an image channel (e.g. any RGB image has three channels: Red, Green, and Blue). Each variable contains the brightness intensity; its values are bounded by 0 and 255.

The purpose of the analysis is to group objects into $K\geq 2$ clusters. Homogeneous segments of an image are represented by the clusters obtained. The number of clusters may be either given beforehand or not (in the latter case an optimal number of groups should be determined automatically). We assume desired number of clusters to be a predefined parameter. Cluster centroids (i.e. vectors of mean coordinates over clusters) are denoted by $c_1,...,c_K$. The set of prototypes (i.e. data points closest to the cluster centroids) are denote by $p_1,...,p_K$.

Let $\mu$ be an algorithm used to partition $s$ into clusters $C_1,...,C_K$. Cluster labels obtained by different algorithms do not matter, so it's convenient to consider the equivalence relation instead (i.e. to indicate whether the algorithm assigns each pair of objects to the same class or not). We denote

$$h_{\mu(i,j)} = \mathbf{I}[\mu(o_i) = \mu(o_j)],$$

where $\mathbf{I}[\cdot]$ is indicator function: $\mathbf{I}$[true] = 1, $\mathbf{I}$[false] = 0; $\mu(o_i)$ is cluster number assigned to object $o_i$ by $\mu$; $i, j\in\{1,...,N\}$.

Following ensemble clustering approach, we pick a collection of different algorithms (or single algorithm with different parameter settings) and use it to assign cluster labels $\mu_1(o_i)$, ... ,$\mu_L(o_i)$ to each $o_i$ from $s$. Co-association matrix-based ensemble clustering is considered as a two-stage process. Firstly, collective co-association matrix $\mathbf{H} = h(i,j)$ with elements

$$h(i,j) = 1/L\ \Sigma_l\, h_{\mu_l}(i,j),\ (i, j = 1,..., N;\ i \neq j)$$

is produced by the set of labels. The final consensus partition is performed at the second stage. The elements of **H** can be considered as pairwise distances between objects so any distance-based algorithm is applied to partition a sample into the desired number of clusters. Hierarchical agglomerative clustering algorithm based on averaged linkage distance ("dendrogram" algorithm) was used in our experiments.

## III. SUGGESTED ALGORITHMS OF ENSEMBLE SEGMENTATION FOR HYPERSPECTRAL IMAGES

*K*-means is a popular clustering algorithm due to its simplicity and effectiveness. The centroids produced by *k*-means could be used for the design of an ensemble solution by averaging of its coordinates. The proposed algorithm of averaged centroid-based ensemble clustering consists of the following steps.

**The "Averaged centroids" algorithm:**

**Input**: dataset $\mathbf{x} = (x_i)$, $i=1,...,N$, where $x_i \in R^d$.
**Output**: partition of **x** on a given number of clusters ($K$).

1. Produce $L$ variants of $k$-means clustering by random initialization of algorithm's settings (such as initial centroids coordinates, number of centroids etc.). The number of clusters for each variant may be different.
2. For each $i^{\text{th}}$ object, find closest centroid $c_{i,l}$ in each $l^{\text{th}}$ variant of partition and set $\mathbf{y}_i := 1/L\ \Sigma_l\, c_{i,l}$ (the object is mapped to the average of its group centroids).
3. Apply $k$-means with the given number $K$ of clusters to the amended dataset **y**.
4. Form a final partition of **x** by matching cluster labels of **y**.

This algorithm is really efficient because the number of computations is quite low. However, it works well only if all the ensemble partitions are quite similar. An alternative method uses co-association matrix on prototypes and combines pairwise ensemble approach and the proposed $k$-means based compression.

**The "Prototype co-association matrix" algorithm:**
**Input**: dataset $\mathbf{x} = (x_i)$, $i=1,...,N$, where $x_i \in R^d$.
**Output**: partition of **x** on a given number of clusters ($K$).

1. Using $k$-means method produce $L$ variants of dataset partitions on $K_l$ clusters ($l=1,...,L$) similarly to the previous algorithm.
2. For each $l^{\text{th}}$ partition, form a co-association matrix using pairs of prototypes $p_{k,l}$ obtained for all partitions ($l=1,...,L$, $k=1,...,K_l$).
3. Calculate collective (averaged) coassociation matrix **H** over all clustering variants.
4. Perform dendrogram clustering of the set of prototypes on $K$ groups using matrix **H**.
5. Form a final partition of **x** by assigning each object to the closest prototype's cluster.

## IV. EXPERIMENTAL EVALUATION

The suggested algorithms were tested on *Indian Pines* hyperspectral image [6]. The image size is 145×145 pixels; each pixel is characterized by a vector of 224 spectral intensities in 400-2500 nm range.

Fig. 1 illustrates the image: a grayscale representation is obtained from the $20^{\text{th}}$ channel. The result of clustering with conventional $k$-means algorithm (with 10 clusters) is shown on

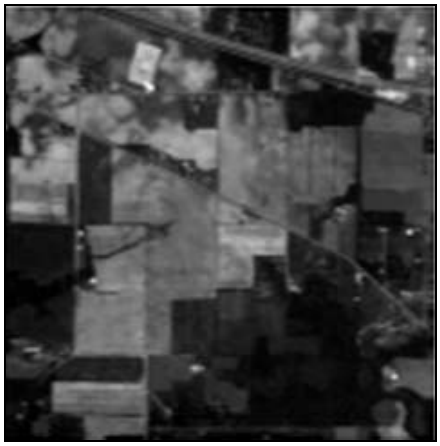


Fig. 1. *Indian Pines* hyperspectral image: $20^{\text{th}}$ channel.

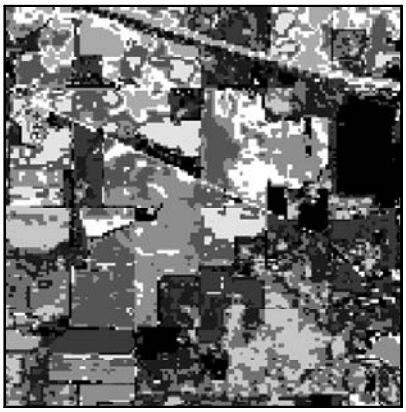


Fig. 2. Results of *k*-means clustering (*K*=10).

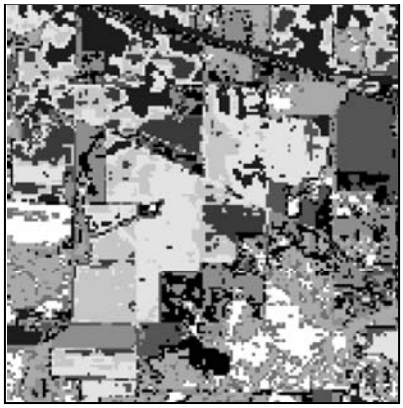


Fig. 3. Results of the "averaged centroids" algorithm (*K*=10).

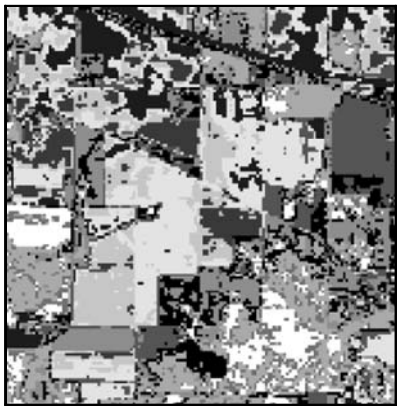


Fig. 4. Results of "prototype co-association matrix" algorithm (*K*=10).

Fig. 2. The next two figures present the results of suggested algorithms: "averaged centroids" (Fig. 3) and "prototype co-association matrix" (Fig. 4).

It is rather difficult to compare the quality of images visually so we used a modification of Dunn's cluster validity index [7] to numerically evaluate the degree of compactness and separability of the classes. The index has the following form:

$$DI=\Sigma_{i,j}\delta(C_i,C_j)/\Sigma_l\Delta(C_i),$$

where $\delta(C_i,C_j)$ is distance between clusters $C_i$ and $C_j$ (the Euclidian distance between their closest points), $\Delta(C_i)$ is the diameter of cluster $C_i$ (the distance between its farthest points). The greater the index, the more compact and separated are the clusters found.

Table 1 shows the indexes averaged over 40 repetitions of algorithms with randomly chosen initialization settings. One can see that the "averaged centroids" algorithm works better than simple *k*-means, and the "prototype co-association matrix" algorithm is better than the "averaged centroids".

Note that there is a calibrated representation for the Indian Pines hyperspectral image. It contains 16 areas (maize, soya, wheat plantations etc.) of the earth surface. However, the presence of "ground truth" class labels is rather uncommon in image segmentation, so we prefer to use internal validity index as a measure of "natural" clustering of a dataset.

TABLE I. AVERAGED DI INDEX

| ***k*-means clustering** | **Averaged centroids** | **Prototype co-association matrix** |
|---|---|---|
| 4.5477 | 4.5809 | 4.6203 |

The next experiment aims to test the stability of suggested algorithms in the processing of "noisy" dataset. A clustering algorithm considered stable if adding "noisy" variables or clustering only a random part of pixels doesn't change the solution too much. Fig. 5 and 6 illustrate the clustering results for "noisy" variables (data points for these variables are randomly distributed over its ranges). Under these conditions one can compare the stability of simple *k*-means algorithm and the "averaged centroids" method. Even if the visual result is not so obvious, the quality index for the ensemble solution based on 100 repetitions of the algorithm is better than the one with the simple clustering: 0.2653 for *k*-means and 0.2845 for "averaged centroids" algorithm. The result of ensemble clustering is obviously better than the single one. It is also fair for datasets without "noisy" channels. We calculated the relative difference of indexes without "noisy" variables to compare these two treatments properly. It's 2.066% compared

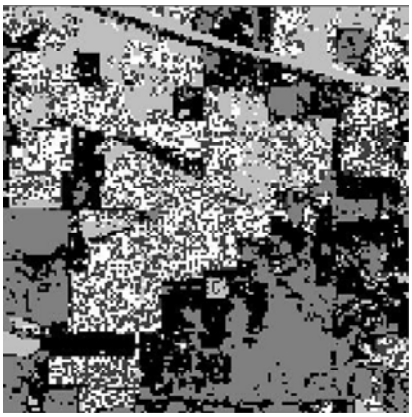


Fig. 5. Clustering with *k*-means for 180 "noisy" channels (*K*=5).

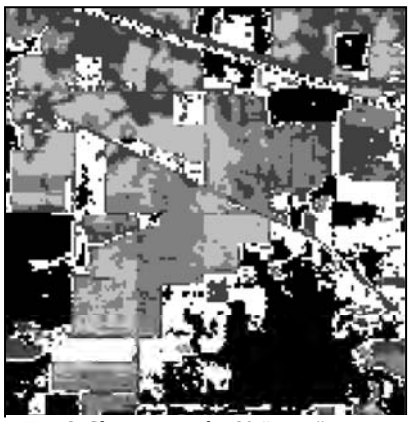


Fig. 6. Clustering with 180 "noisy" channels: "averaged centroids" algorithm (*K*=5).

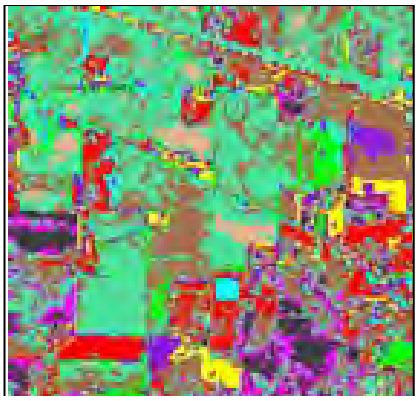


Fig. 7. Clustering results with different number of ensemble elements $L$ ($K$=10, $B$=40, $L$=20).

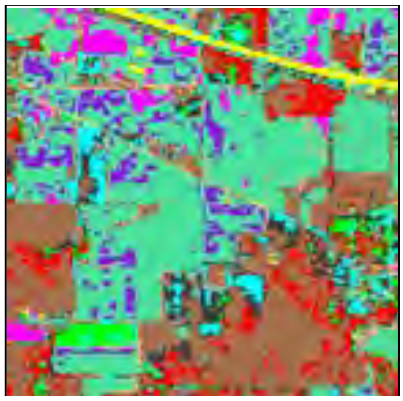


Fig. 8. Clustering results with different number of ensemble elements $L$ ($K$=10, $B$=40, $L$=50).

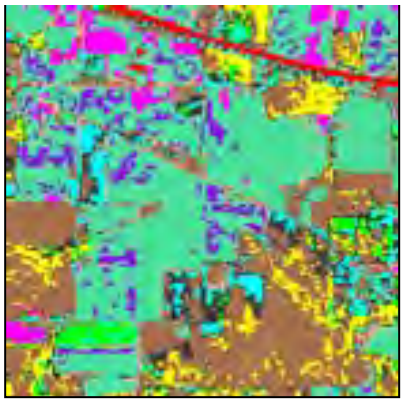


Fig. 9. Clustering results with different number of ensemble elements $L$ ($K$=10, $B$=40, $L$=170).

with 6.749% for datasets with noisy channels. This result confirms the stability improvement due to ensemble clustering.

Another modification of the ensemble algorithm exploits random subspace method. The overview of the algorithm is as follows.

**The "Random subspace averaged centroids" algorithm:**
**Input**: dataset $\mathbf{x} = (x_i)$, $i$=1,...,$N$, where $x_i \in R^d$.
**Output**: partition of $\mathbf{x}$ on a given number of clusters ($K$).
1. For each ensemble iteration 1,…, $L$:
1.1. Select fixed number $B$ of a hyperspectral image channels.
1.2. Perform $k$-means clustering for the selected channels.
1.3. Calculate centroids for the resulting clusters in feature space (using all available channels of the image).
1.4. Create a hyperspectral image with every pixel having a value of the centroid of the corresponding cluster.
2. Take all produced images and store average spectral value for each pixel in a new image.
3. Perform $k$-means clustering of the resulting image.

Clustering results for the "random subspace averaged centroids" algorithms with different number of ensemble elements are shown on Fig. 7-9. One can see that the results tend to stabilize after first 50 ensemble iterations.

## V. Conclusion

A novel method of ensemble clustering is proposed. It is realized as a number of hyperspectral image segmentation algorithms.

The basic idea of the method is to use an ensemble of $k$-means algorithms as a preliminary step to reduce the amount of data under analysis. Clustering results on real hyperspectral image demonstrate the efficiency of the proposed algorithms.

One of potential extensions of the method is finding the consensus partition with use of algorithm's weights proportional to obtained cluster validity indexes.

## References

[1] A.K. Jain, "Data clustering: 50 years beyond K-means", Pattern Recognition Letters, Vol. 31(8), 2010. pp. 651-666.

[2] J. Ghosh, A. Acharya, "Cluster ensembles", Wiley Interdisciplinary Reviews, Data Mining and Knowledge Discovery, Vol. 1, 2011, pp. 305-315.

[3] V. Berikov, "Weighted ensemble of algorithms for complex data clustering", Pattern Recognition Letters, Vol. 38, 2014, pp. 99-106.

[4] A. Plaza, J.A. Benediktsson, J.W., Boardman, J. L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, G. Trianni, "Recent advances in techniques for hyperspectral image processing", Remote sensing of environment, Vol. 113, 2009, pp. S110-S122.

[5] R. Ablin, C.H. Sulochana, "A Survey of Hyperspectral Image Classification in Remote Sensing", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Is. 8, 2013, pp. 2986-3003.

[6] Hyperspectral Remote Sensing Scenes [URL: http://www.ehu.es/ccwintco/index.php/HyperspectralRemote SensingScenes].

[7] J.C. Dunn, "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", Journal of Cybernetics, Vol. 3(3), 1973, pp. 32-57.

Proceedings of the 9th Open German-Russian Workshop on Pattern Recognition and Image Understanding
Dec 1-5, 2014, Koblenz, Germany

# Combinatorial Optimization Problems Related to Machine Learning Techniques

Michael Khachay
Krasovsky Institute of
Mathematics and Mechanics
16 S. Kovalevskoy str., 620990 Ekaterinburg
Omsk State Technical University
11 Mira ave., 644050 Omsk, Russia
Email: mkhachay@imm.uran.ru

Maria Poberii
Krasovsky Institute of
Mathematics and Mechanics
16 S. Kovalevskoy str., 620990 Ekaterinburg
Omsk State Technical University
11 Mira ave., 644050 Omsk, Russia
Email: maschas_briefen@mail.ru

***Abstract*—A brief survey of computational complexity and approximability results concerning efficient cluster analysis techniques and learning procedures in the class of piece-wise linear majority classifiers is provided. Also, new results confirming the connection between the structural minimization risk principle and theoretic combinatorics (along with rigorous proofs) are presented.**

## I. Introduction

Optimization and machine learning appear to be extremely close fields of the modern computer science. Various areas in machine learning: SVM-learning and kernel machines (see, e.g. [1]), PAC-learning [2] and boosting [3], [4], cluster analysis, etc. are continuously presenting new challenges for designers of optimization methods due to the steadily increasing demands on accuracy, efficiency, space and time complexity and so on.

Sometimes a learning problem can be successfully reduced to some kind of combinatorial optimization problems: max-cut, p-median, TSP, etc. To this end, all the known results for the latter problem (approximation algorithms, polynomial-time approximation schemas, approximation thresholds) can find their application in design precise and efficient learning algorithms for the former.

In this paper we try to observe some new results proving such a mutual cooperation between CO and ML.

This is chiefly a survey paper. In Section II we overview several recent results obtained for two special cases of well-known Weighted Clique Problem and their application in cluster analysis. Further, in Section III, some aspects of ensemble learning are covered. In particular, in Subsection III-A we give a brief survey on combinatorial results concerning the Minimum Affine Separating Committee (MASC) problem. Finally, in Subsection III-B we present new results bridging well-known Structural risk minimization principle and the Integer Partition Problem (IPP), which is an object of interest in combinatorics and number theory.

## II. Cluster Analysis and the Weighted Clique Problem

We start with giving an informal description of the special class of cluster analysis problems. Suppose, we deals with a collection of uniform objects. Any object can be in two states, we call them *on-line* and *off-line*. During the experiment, for any object in question, its current state can be revealed up to some additive noise. The input dataset contains such results. The problem is to distinguish the given number of active objects.

It is convenient, to formulate this problem mathematically in terms of graph theory. Let a complete, weighted, undirected graph $G = (V, E, w)$ be given. Here, $w$ is a weighting (cost) non-negative function $w : E \to \mathbb{R}_+$ defined on the edge-set of the graph $G$. For any subgraph $G' = (V', E') \subset G$, the weight $w(G')$ is defined by the equation $w(G') = \sum_{e \in E'} w(e)$. As usual, we call any complete subgraph $G'$ of the graph $G$ *a clique*.

*Problem 1 (min-EWCP):* Input: a graph $G$ of order $n$ and a natural number $m < n$. It is required to find a clique of order $m$ having the minimum possible weight.

As it is proved in [5], the min-EWCP is intractable and inapproximable in the general case.

*Theorem 1:* (i) The Min-EWCP is strongly NP-hard. (ii) For any ratio $r = O(2^n)$, the Min-EWCP has no polynomial time r-approximation algorithms unless $P = NP$.

Following [5], we describe two special cases of the Min-EWCP problem: we call them *metric* (Min-EWCP-M) and *square Euclidean* (Min-EWCP-SE):

(i) in the metric subclass, the weighting function $w$ is defined by symmetric zero-main-diagonal $n \times n$-matrix $W = \|w_{ij}\|$ satisfying the triangle inequality $w_{ij} + w_{jk} \geq w_{ik}$;

(ii) in the square Euclidean case, vertices of the graph $G$ are points in some finite-dimensional Euclidean space and entries of the weight matrix $W$ is obtained as squared Euclidean among them.

As the general min-EWCP, the defined above its special cases are intractable as well.

*Theorem 2:* The min-EWCP-M and min-EWCP-SE problems are strongly NP-hard.

Fortunately, the both problems belong to Apx class. In [5], for each of these problems, a polynomial-time 2-approximation

algorithm is developed. For brevity, we skip their formal description. Theorem 3 summarizes their main properties.

*Theorem 3:* (i) for min-EWCP-M problem there exist a polynomial time 2-approximation algorithm with running time of $O(n^2)$ and asymptotically tight approximation ratio; (ii) for min-EWCP-SE problem there exist a polynomial time 2-approximation algorithm with running time of $O(n^2)$ and tight approximation ratio;

## III. Committee Separability Related Problems

### A. Minimum Affine Separating Committee (MASC) Problem

We start with the simplest two-class setting of the classic pattern recognition problem. For a finite sample

$$\xi = ((x_1, y_1), \ldots, (x_m, y_m)), \tag{1}$$

where $x_i \in \mathbb{R}^d$ and $y_i \in \{-1, 1\}$, for a given class

$$\mathcal{H} \subset [\mathbb{R}^d \to \{-1, 1\}]$$

of two-valued functions, which are called *classifiers*, and a given utility functional $I : H \to \mathbb{R}_+$, it is required to find an optimal (or suboptimal) classifier $h = h_\xi$ w.r.t. the functional $I$. Usually, the functional $I$ has a form of expected misclassification loss

$$I[h] = \int\limits_{X \times Y} L(y - h(x)) dP(x, y)$$

depending on unknown probability measure $P$. Due to this uncertainty, the initial problem

$$\min_{h \in H} I[h] \tag{2}$$

could not be solved as is and should be refined. According to the famous Empirical Risk Minimization (ERM) principle, the unformalized functional $I$ is replaced with the empirical mean functional (of the misclassification loss)

$$I_{emp}[h] = \frac{1}{m} \sum_{j=1}^{m} L(y_j - h(x_j)), \tag{3}$$

which is completely defined by training sample (1). A classifier $h_0$ making no classification errors on sample (1) is called *perfect* w.r.t. this sample.

The Minimum Affine Separating Committee (MASC) combinatorial optimization problem is equivalent to the minimization problem of the functional $I_{emp}$ over the special class $\mathcal{H}_{asc}$ of piece-wise linear classifiers

$$h(x) = \operatorname{sign} \sum_{i=1}^{k} a_i \operatorname{sign}(w_i^T x - b_i) \tag{4}$$

for some positive integers $a_i$, where decisions of *weak* affine classifiers $\operatorname{sign}(w_i^T x - b_i)$ are aggregated by the simple majority voting rule. The value $n = \sum_{i=1}^{k} a_i$ is called *a length* of the classifier $h$. Any perfect committee classifier w.r.t. sample (1) of minimum length is called *minimum affine separating committee*.

Actually, the MASC problem appears to be one of mathematical formalizations of the well-known Vapnik-Chervonenkis structural risk minimization principle [6], which is of searching of the most precise classifier in the most narrow subfamily of feasible ones.

We use the traditional mathematical notation $\mathbb{N}$ and $\mathbb{R}$ for the sets of natural and real numbers, and $\mathbb{N}_k$ for the set $\{1, \ldots, k\}$. For convenience, we introduce subsets $A, B \subset \mathbb{R}^d$ consisting of $x_i$ from sample (1) and defined by the equations

$$A = \{x_i \colon y_i = 1\}, \quad B = \{x_i \colon y_i = -1\}. \tag{5}$$

In [7], it is proved the following criterion of perfect learnability in the class of affine committee classifiers.

*Theorem 4:* Class $\mathcal{H}_{asc}$ contains a perfect classifier if and only if

$$A \cap B = \varnothing. \tag{6}$$

Hereinafter, we assume training sample (1) to be *regular*, i.e.

(i) satisfying condition (6);

(ii) the set $A \cup B$ is in *general position* (see Definition 1)

*Definition 1:* A set $D \subset \mathbb{R}^d$, $|D| > d$, is said to be in general position, if, for any subset $D' \subseteq D$, $|D'| = n + 1$, the equality $\dim \operatorname{aff} D' = d$ is valid.

*Problem 2 (MASC-GP):* For a given sample (1) it is required to construct a minimum affine separating committee.

To emphasize the important special case of the MASC-GP problem, in which the dimensionality $d$ is fixed in advance, we call this problem MASC-GP(d). Complexity of the both problems is described in Theorem 5 [8] and 6 [9].

*Theorem 5:* The MASC-GP problem is strongly NP-hard.

*Theorem 6:* The MASC-GP($d$) problem is polynomially solvable for $d = 1$ and NP-hard for any fixed $d > 1$.

The state-of-the-art *Boosting Greedy Committee (BGC)* approximation algorithm for the MASC-GP(d) problem is proposed in [10]. This algorithm has the best knonwn approximation ration and a rather huge but polynomial complexity bound.For brevity, we skip its formal description but recall main properties in Theorem .

In sequel, we call an instance of the MASC-GP($d$) problem *nice* if there exists a minimum affine separating committee (4) of odd length $n$ such that, for any $t = 1, \ldots, (n-1)/2$, the following conditions

$$(w_t^T x - b_t > 0) \quad \vee \quad (w_{t+1}^T x - b_{t+1} > 0), \quad (x \in A),$$
$$(w_t^T x - b_t < 0) \quad \vee \quad (w_{t+1}^T x - b_{t+1} < 0), \quad (x \in B),$$

are valid.

*Theorem 7:* The BGC algorithm has the approximation ratio of $O(\ln(m))$ for the nice instances of the MASC-GP($d$) problem and

$$O\left(\left(\frac{m \ln m}{d}\right)^{1/2}\right),$$

otherwize. Its time-complexity is of $O(m^{3d})$ for any $d > 2$.

### B. Committee Minimal Partitions

The Integer Partition Problem (IPP) is one of the fundamental problems, which is studied in combinatorics and number theory. In this problem, for a given natural number $n$ it is required to enumerate all of ways to represent of this number us a sum of other non-negative integer numbers $n = a_1 + \ldots + a_k$. The finite sequence $(a_1, \ldots, a_k)$ is called *integer partition* of the number $n$. For brevity, the fact 'the sequence $A$ is a partition of the number $n$' is denoted by $A \vdash n$.

We consider the restricted version of the IPP; we call this problem Minimum Committee Integer Partition Problem (MC-IPP) which is closely related to the minimum committee notion considered above. In MC-IPP, for a given training sample $\xi$, a given class $\mathcal{F}$ of weak classifiers, and for a given number $n$, it is required to enumerate only such partitions $(a_1, \ldots, a_k)$, for which there exists a perfect (w.r.t. the sample $\xi$) minimum committee classifier $h$ such that

$$h(x) = \operatorname{sign} \sum_{i=1}^{k} a_i f_i(x), \quad (f_i \in \mathcal{F}). \tag{7}$$

Hereinafter, to emphasize that the classifier $h$ is defined by the partition $(a_1, \ldots, a_k)$ and the weak classifiers $f_1, \ldots, f_k$, we use the notation $h(\cdot|a_i, f_i)$.

We introduce the following partial order over the set of all partitions of the length $k$.

*Definition 2:* Suppose $A = (a_1, \ldots, a_k)$ and $B = (b_1, \ldots, b_k)$ be finite sequences of non-negative integers. The relation $A \succcurlyeq B$ is defined by the following equations

$$\sum_{i=1}^{k} a_i = n \quad \geq \quad m = \sum_{i=1}^{k} b_i, \tag{8}$$

$$(J \subseteq \mathbb{N}_k) \quad \left(\sum_{i \in J} a_i > \frac{n}{2}\right) \quad \Longleftrightarrow \quad \left(\sum_{i \in J} b_i > \frac{m}{2}\right) \tag{9}$$

Notice that, if $A \succcurlyeq B$ then, for any sample $\xi$ and for any weak classifiers $f_1, \ldots, f_k$, the committee classifiers $h(\cdot|a_i, f_i)$ and $h(\cdot|b_i, f_i)$ are perfect w.r.t. the sample $\xi$ simultaneously and the length of the latter committee does not exceeds the length of the former.

We call a sequence $A = (a_1, \ldots, a_k)$ *decreasing* if $a_1 \geq \ldots \geq a_k$. It is convenient to restrict our consideration to decreasing partitions only.

*Assertion 1:* Let sequences $A = (a_1, \ldots, a_k)$ and $B = (b_1, \ldots, b_k)$ be partitions of numbers $n$ and $m$, respectively, satisfying equation (9). If the partition $A$ is decreasing, then, for the number $p$ there exists a decreasing partition $C = (c_1, \ldots, c_k)$ such that, for $A$ and $C$ the equation (9) is valid, as well.

*Proof:* W.l.o.g., it is sufficient to verify that, if $b_1 < b_2$ then the sequence $B' = (b'_1, b'_2, \ldots, b'_k)$, for which $b'_1 = b_2$, $b'_2 = b_1$, and $b'_i = b_i$ for $i > 2$, is also satisfies (together with $A$) condition (9).

Choose any $J \subset \mathbb{N}_k$ such that

$$\sum_{i \in J} b'_i > \frac{m}{2}.$$

If $\{1, 2\} \cap J = \varnothing$ or $\{1, 2\} \subset J$ then

$$\sum_{i \in J} b_i = \sum_{i \in J} b'_i > \frac{m}{2};$$

therefore,

$$\sum_{i \in J} a_i > \frac{n}{2},$$

by condition. If $1 \in J$ and $2 \notin J$, then

$$\sum_{i \in J \setminus \{1\} \cup \{2\}} b_i = \sum_{i \in J} b'_i > \frac{m}{2}.$$

Therefore,

$$\sum_{i \in J \setminus \{1\} \cup \{2\}} a_i > \frac{n}{2},$$

by condition, and consequently

$$\sum_{i \in J} a_i \geq \sum_{i \in J \setminus \{1\} \cup \{2\}} a_i > \frac{n}{2},$$

as $a_1 \geq a_2$. Finally, if $1 \notin J$ and $2 \in J$, then $\sum_{i \in J} b_i > \sum_{i \in J} b_i$, since $b_2 > b'_2 = b_1$. Hence,

$$\sum_{i \in J} a_i > \frac{n}{2}.$$

Assertion 1 is proved. ■

Hereinafter, for any partition $A$, we assume that $A$ is decreasing. Assertion 1 suggests us to extend (using the zero-padding technique) the defined above partial order $A \succcurlyeq B$ onto sequences of unequal lengths.

*Definition 3:* Suppose, for a sequence $A$ and for any other sequence $B$, either $B \succcurlyeq A$ or the sequences $A$ and $B$ are incomparable. Then, the sequence $A$ is called *a committee minimal (c-minimal) sequence*.

In the sequel, we apply Definition 3 to partitions of natural numbers. Evidently, if committee (7) is perfect for some sample $\xi$, then the partition $(a_1, \ldots, a_k)$ is c-minimal. The inverse claim is also valid.

*Theorem 8:* Let $A = (a_1, \ldots, a_k)$ be a c-minimal partition of some number $n$. There exist a training sample $\xi$ and a class $\mathcal{F}$ of weak classifiers such that committee (8) (for some classifiers $f_1, \ldots, f_k$) is perfect w.r.t. the sample $\xi$.

Theorem 8 is announced in [11] and can be proved using the technique proposed in [12].

To define the MC-IPP, we introduce some additional notation. For an arbitrary odd number $n$ and natural number $s \leq n$, we denote by $\Lambda(n)$ and $\Lambda(n, s)$ the set of all partitions and the set of $s$-fold partitions of $n$, respectively.

*Problem 3 (MC-IPP):* For an odd number $n$ it is required to enumerate all the c-minimal elements of $\Lambda(n)$.

Further we propose the following efficiently verifiable c-minimality condition.

*Condition 1:* Let a partition $A = (a_1, \ldots, a_k$ be given. For any $1 \leq i_1 < i_2 \leq k$ there exists a subset $J \subset \mathbb{N}_k$ such that $\{i_1, i_2\} \subset J$ and $\sum_{i \in J} a_i = \lceil n/2 \rceil$.

*Lemma 1 (Necessary condition):* For any odd number $n$ and any c-minimal partition $A = (a_1, \dots, a_k) \vdash n$, Condition 1 is valid.

*Proof:* Assume by contradiction that there exist numbers $i_1 < i_2$ such that, for any $J \subset \mathbb{N}_k$ the condition

$$\left(\{i_1, i_2\} \subset J,\ \sum_{i\in J} a_i > \frac{n}{2}\right) \Rightarrow \left(\sum_{i\in J} a_i > \left\lceil \frac{n}{2} \right\rceil\right) \quad (10)$$

is valid. We consider the partition $B = (b_1, \dots, b_k)$ of the number $n-2$ defined by equations

$$b_{i_1} = a_{i_1} - 1, b_{i_2} = a_{i_2} - 1,\ b_i = a_i\ (i \in \mathbb{N}_k \setminus \{i_1, i_2\}) \quad (11)$$

and verify that $A \succcurlyeq B$. Indeed, let, for some subset $J \subset \mathbb{N}_k$, the equation $\sum_{i\in J} b_i > n/2 - 1$ be valid. If $\{i_1, i_2\} \cap J \neq \varnothing$, then $\sum_{i\in J} a_i > n/2$ by construction of the partition $B$.

Otherwise, we obtain

$$\sum_{i\in J} a_i > \frac{n}{2} - 1.$$

To this end, if

$$\sum_{i\in J} a_i < \frac{n}{2},$$

then

$$\sum_{i\in J} a_i = \left\lfloor \frac{n}{2} \right\rfloor;$$

therefore,

$$\sum_{i\notin J} a_i = \left\lceil \frac{n}{2} \right\rceil.$$

This equation contradicts to our assumption (10), since $\{i_1, i_2\} \subset \mathbb{N}_k \setminus J$. Lemma is proved. ■

Further, we prove that, to verify whether a given partition $A = (a_1, \dots, a_k)$ is c-minimal, it is sufficient to examine partitions $B = (b_1 \dots, b_l)$, for which $l \le k$.

*Lemma 2:* (i) Suppose, $A = (a_1, \dots, a_k)$, $a_k > 0$, is not a c-minimal partition of an odd number $n$. For some number $m < n$, there exists a partition $B = (b_1, \dots, b_l) \vdash m$ such that $A \succcurlyeq B$ and $l \le k$. (ii) If, in addition, for the partition $A$, Condition 1 is valid, then $l = k$.

*Proof:* Indeed, since the partition $A$ is not c-minimal, there exists a partition $C = (c_1, \dots, c_l) \vdash m$ such that $c_l > 0$ and $A \not\succcurlyeq C$. Assume, $l > k$; and introduce the following notation

$$d = \sum_{i=k+1}^{l} c_i. \quad (12)$$

Let $d$ be an even number. We consider the partition $B = (c_1, \dots, c_k) \vdash (m-d)$ and show that $A \succcurlyeq B$. For an arbitrary subset $J \subset \mathbb{N}_k$ such that

$$\sum_{i\in J} c_i > \frac{m-d}{2} > \frac{m}{2} - d,$$

we obtain

$$\sum_{i\in J\cup\{k+1,l\}} c_i > \frac{m}{2};$$

therefore,

$$\sum_{i\in J} a_i = \sum_{i\in J} a_i + \sum_{i=k+1}^{l} 0 > \frac{n}{2},$$

by choice of the partition $C$.

On the other hand, let $d$ be an odd number. To this end, we consider the partition $B = (b_1, b_2, \dots, b_k) \vdash (n-d-1)$ defined by the equations

$$b_k = c_k - 1,\ b_i = c_i\ (i \in \mathbb{N}_{k-1}).$$

As above, we define $d$ by (12). For any $J \subset \mathbb{N}_k$ such that

$$\sum_{i\in J} b_i > \frac{m-d-1}{2} > \frac{m}{2} - d,$$

we obtain

$$\sum_{i\in J\cup\{k+1,l\}} c_i > \frac{m}{2}.$$

Consequently,

$$\sum_{i\in J} a_i > \frac{n}{2}.$$

Claim (i) is proved.

To prove claim (ii), assume that, for the partition $A$, the necessary condition of c-minimality (proven in Lemma 1) is valid. In this case, we prove that, for any partition $B = (b_1, \dots, b_l)$,

$$(A \succcurlyeq B,\ b_l > 0) \Rightarrow (l \ge k).$$

Assume, by contradiction that there is a partition $B = (b_1, \dots, b_l) \vdash m$, for which $b_l > 0$ and $l < k$. Since the partition $A$ satisfies the necessary condition, there exists a subset $J \subset \mathbb{N}_k$ such that $\{1, k\} \subset J$ and

$$\sum_{i\in J} a_i = \left\lceil \frac{n}{2} \right\rceil.$$

As $A \succcurlyeq B$,

$$\sum_{i\in \mathbb{N}_l \cap J} b_i > \frac{m}{2}.$$

On the other hand,

$$\sum_{i\in \mathbb{N}_k\setminus J} a_i + a_k > \sum_{i\in \mathbb{N}_s\setminus J} a_i = \left\lfloor \frac{n}{2} \right\rfloor.$$

Therefore, $\sum_{i\in \mathbb{N}_l\setminus J} b_i > m/2$, i.e. $\sum_{i=1}^{l} b_i > m$ that contradicts to the choice of the partition $B \vdash m$. Claim (ii) and lemma are proved. ■

Further, we prove several assertions providing an approach to recurrent construction of c-minimal partitions.

*Assertion 2:* Let $A = (a_1, \dots, a_k)$ be a c-minimal partition of some odd number $n$. The partition

$$B = (a_1, \dots, a_k, 1, 1) \vdash (n+2) \quad (13)$$

is c-minimal as well.

*Proof:* As $A$ is a c-minimal partition, for this partition Lemma 1 is valid. It is easy to verify, that the same necessary condition is valid for the partition $B$ as well. Assume, by contradiction that there exists a partition $C = (c_1, \dots, c_l) \vdash m$

such that $B \succcurlyeq C$ and $m \le n$. Due to claim (ii) of Lemma 2, $l = k+2$ and $c_{k+2} > 0$. By choice of the partition $C$, for any $J \subset \mathbb{N}_k$, the equation

$$\sum_{i \in J} c_i + c_{k+1} > \frac{m}{2}$$

implies

$$\sum_{i \in J} a_i + 1 > \frac{n+2}{2}, \text{ i.e. } \sum_{i \in J} a_i > \frac{n}{2}.$$

Let $c_{k+1} + c_{k+2} \equiv 0 \pmod 2$. Consider the partition

$$D = (c_1, \dots, c_k) \vdash m - (c_{k+1} + c_{k+2}).$$

Assume that, for some $J \subset \mathbb{N}_k$

$$\sum_{i \in J} c_i > \frac{m - c_{k+1} - c_{k+2}}{2}.$$

Then,

$$\sum_{i \in J} c_i > \frac{m}{2} - c_{k+1},$$

since the partition $C$ is decreasing. Therefore,

$$\sum_{i \in J} c_i + c_{k+1} > \frac{m}{2};$$

hence,

$$\sum_{i \in J} a_i + 1 > \frac{n+2}{2}, \text{ and } \sum_{i \in J} a_i > \frac{n}{2}.$$

Thus, $A \succcurlyeq D$ and the partition $A$ is not c-minimal, since $m - (c_{k+1} + c_{k+2}) < n$.

In the case $c_{k+1} + c_{k+2} \equiv 1 \pmod 2$, we obtain $c_k > 1$ by construction of the partition $C$. Consider the partition

$$D = (c_1, \dots, c_k - 1) \vdash m - (c_{k+1} + c_{k+2} + 1).$$

As $c_{k+1} > c_{k+2}$, then

$$\frac{c_{k+1} + c_{k+2} + 1}{2} \le c_{k+1}.$$

It is easy to show that, for any $J \subset \mathbb{N}_k$, the condition

$$\sum_{i \in J} c_i > \frac{m - (c_{k+1} + c_{k+2} + 1)}{2}$$

implies

$$\sum_{i \in J} a_i > \frac{n}{2}.$$

Therefore, again, $A \succcurlyeq D$ and the partition $A$ is not c-minimal, since $m - (c_{k+1} + c_{k+2} + 1) < n$. The obtained contradiction completes the proof. ■

*Assertion 3 (Sufficient condition):* Let $A = (a_1, \dots, a_k) \vdash n$, $a_k > 0$, be a c-minimal partition for a given odd number $n$. For some $i \in \mathbb{N}_k$, let the partition $B = (b_1, \dots, b_k, b_{k+1})$ be defined by the equation

$$b_i = a_i + 1,\ b_j = a_j\ (j \in \mathbb{N}_k \setminus \{i\}), b_{k+1} = 1. \qquad (14)$$

If Condition 1 is valid for the partition $B$, then this partition is c-minimal.

*Proof:* Assume, by contradiction that there exists a partition $C = (c_1 \dots, c_{k+1} \vdash m)$ such that $B \succcurlyeq C$ and $m \le n$. Define the partition $D = (d_1, \dots, d_k) \vdash m - 2c_{k+1}$ by the formulas

$$d_i = c_i - c_{k+1},\ d_j = c_j\ (j \in \mathbb{N}_k \setminus \{i\})$$

Suppose, for some subset $J \subset \mathbb{N}_k$,

$$\sum_{j \in J} d_i > \frac{m}{2} - c_{k+1}.$$

If $i \in J$, then

$$\sum_{j \in J} c_j = \sum_{j \in J} d_j + c_{k+1} > \frac{m}{2}.$$

Therefore,

$$\sum_{j \in J} b_j + 1 = \sum_{j \in J} a_j + 2 > \frac{n+2}{2}.$$

Hence, as $a_j$ are integers and $n$ is an odd number,

$$\sum_{j \in J} a_j > \frac{n}{2}. \qquad (15)$$

Else, if $i \notin J$, then

$$\sum_{j \in J \cup \{k+1\}} c_j = \sum_{j \in J} d_j + c_{k+1} > \frac{m}{2};$$

therefore

$$\sum_{j \in J \cup \{k+1\}} b_j = \sum_{j \in J} a_j + 1 > \frac{n}{2} + 1.$$

Thus, equation (15) is valid again.

Since $m - 2c_{k+1} < n$ and $A \succcurlyeq D$, then $A$ is not c-minimal partition. The obtained contradiction completes the proof. ■

Assertions 2 and 3 provides an efficient algorithm for recurrent construction of c-minimal partitions. It can be easily verified that, if a c-minimal partition $B = (b_1, \dots, b_l)$ is constructed (from another c-minimal partition $A$) by equations (13) or (14), then the partitions $C = (b_1, \dots, b_l, 1, 1)$ and

$$D = (b_1, b_2, \dots, b_i + 1, \dots, b_l, 1)$$

are c-minimal as well for any $i \in \mathbb{N}_l$.

## IV. Conclusion

In the paper connections between several combinatorial optimization problems and machine learning techniques are shown. In particular,

(i) approximability of metric and square Euclidean Minimum Weighted Clique Problem closely related to Minimum Sum of Squares clustering method are observed;

(ii) computational complexity and approximability issues of Minimum Affine Separating Committee Problem related to optimal learning procedures in the class of piece-wise linear majority classifiers are listed;

(iii) new results establishing the connection between classic number theory and combinatorics and structural risk minimization machine learning principle are presented.

## ACKNOWLEDGMENT

This research was supported by Russian Science Foundation grant no. 14-11-00085.

## REFERENCES

[1] B. Schölkopf and A. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[2] L. G. Valiant, "A theory of the learnable," *Communications of the ACM*, vol. 27, no. 11, pp. 1134–1142, 1984.

[3] R. Schapire and Y. Freund, *Boosting: Foundations and algorithms*. MIT Press, 2012.

[4] Y. Freund, "Boosting a weak learning algorithm by majority," *Information and Computation*, vol. 121, pp. 256–285, 1995. [Online]. Available: http://libgen.org/scimag/index.php?doi=10.1006/inco.1995.1136

[5] I. Eremin, E. Gimadi, A. Kelmanov, A. Pyatkin, and M. Y. Khachai, "2-approximation algorithm for finding a clique with minimum weight of vertices and edges," *Proceedings of the Steklov Institute of Mathematics*, vol. 284, no. 1, pp. 87–95, 2014.

[6] V. Vapnik, *Statistical Learning Theory*. Wiley, 1998.

[7] V. Mazurov, "Committees of inequalities systems and the pattern recognition problem," *Kibernetika*, vol. 3, pp. 140–146, 1971.

[8] M. Khachai, "Computational and approximational complexity of combinatorial problems related to the committee polyhedral separability of finite sets," *Pattern Recognition and Image Analysis*, vol. 18, no. 2, pp. 236–242, 2008.

[9] M. Khachay and M. Poberii, "Complexity and approximability of committee polyhedral separability of sets in general position," *Informatica*, vol. 20, no. 2, pp. 217–234, 2009.

[10] M. Khachai and M. Poberii, "Scheme of boosting in the problems of combinatorial optimization induced by the collective training algorithms," *Automation and Remote Control*, vol. 75, no. 4, pp. 657–667, 2014.

[11] M. Khachai, "On one combinatorial problem concerned with the notion of minimal committee," *Pattern Recognition and Image Analysis*, vol. 11, no. 1, pp. 45–46, 2001.

[12] M. Khachay, "Estimate of the number of members in the minimal committee of a system of linear inequalities," *Computational Mathematics and Mathematical Physics*, vol. 37, no. 11, pp. 1356–1361, 1997.

# Compression algorithm for indexed images with the use of context-based modeling

A.V. Borusyak
Research Institute of Applied Mathematics and Cybernetics
Lobachevsky Nizhni Novgorod State University
National Research University
603005, Russia, Nizhni Novgorod, Ulyanov St.10, UNN
RIAMC
Sw-bor@yandex.ru

Yu.G.Vasin
Research Institute of Applied Mathematics and Cybernetics
Lobachevsky Nizhni Novgorod State University
National Research University
603005, Russia, Nizhni Novgorod, Ulyanov St.10, UNN
RIAMC
ya.vasinyuri@yandex.ru

***In this paper, we propose some techniques for an optimal compression algorithm of indexed raster images (IRI) based on entropy coding with the use of context-based modeling. A model and methods for context-based modeling for an efficient compression algorithm of indexed graphic images are developed. An efficient algorithm for the coding of indexed images is implemented, including optimization and parallelization.***

***Keywords— compression, indexed images, context modeling***

## I. INTRODUCTION

Currently, due to the development of information technologies and Internet technologies, the transfer of large volumes of graphic information, and the use of remote databases, great importance is attributed to the development of effective models and algorithms for data compression. In many areas, the need to develop specialized algorithms focused on a specific data type is relevant today. [1-4] With the knowledge of the internal structure and specificity of the compressed data, such algorithms in many cases offer significant advantages in compression. One such area is the compression of indexed raster images (IRI).

Among the different approaches to IRI compression, the methods and algorithms for entropy coding with context-based modeling play an essential role. In the number of these algorithms, the Prediction by Partial Matching (PPM) algorithm occupies a leading position [1].

## II. MODELS AND METHODS OF CONTEXT MODELING FOR EFFICIENT COMPRESSION OF INDEXED IMAGES

We used a modification of the PPM algorithm described in [3-4] as the basic algorithm. This modification of the algorithm has been adapted for working with 8-bit pixels. In the course of experiments, the form and size of the two-dimensional context were found that provided high compression efficiency.

To estimate the escape probability, the following method is proposed. Initially, the frequency of the escape symbol is equal to 0. With each new pixel encountered, the frequency of the escape symbol is increased by a unity. When the value of the escape symbol frequency reaches the amount equal to the number of different symbols in the image, the escape symbol frequency is zeroed to increase the degree of compression. It was found during the experiments that if a pixel is encountered in the given context for the first time, it is not rational to escape to the nearest smaller context. Instead, it is better to escape to the context that is 2-3 times smaller than the current one. In this case, when calculating the lower order context, all the pixel values of the calculated context are divided by two in order to spare RAM, to increase the speed of compression and to increase the compression ratio by combining the statistics of the nearest contexts. Experimental testing has confirmed the high efficiency of this approach. To increase the compression ratio (Rc), we used a technique for scaling the counter of the last encountered symbol. This technique is to multiply the frequency of the last symbol encountered by some factor. In this paper, we used the factor 4.

This work was supported by RFBR grants (Projects No.13-07-00521, No. 13-07-12211).

## III. Optimization of the time complexity of the algorithm

To increase the speed of compression and to reduce RAM consumption, a special storage structure has been developed for the context key. This structure makes it possible to convert the current context key into a lower order context key by changing only one value of the context length.

We added to each context model a reference to its ancestor context model in order to accelerate the operation of search for the ancestor context model, which is used for inheriting information and escaping to lower-order contexts. Trees of context models were bounded from the top in terms of the number of context models they contained. If the predetermined threshold was exceeded, the tree was completely cleared.

Initially, the algorithm used an AVL-tree structure for storing context models. We also implemented B-tree and Splay-tree structures and then compared the speed of all these structures. As a result of experimental testing, B-tree has proved more efficient in terms of speed and was adopted as the basic storage structure for the tree of context models.

Program performance was improved by transferring it to the Qt 4.8 programming environment; the program structure also became more flexible.

To further improve the performance, the refresh rate of the percentage indicator of the encoding and decoding process was reduced.

The above approaches have contributed to a significant (1.5-2 times) reduction in the encoding time; RAM consumption was also reduced.

## IV. Parallelization of the encoding algorithm

In order to further increase the speed of compression, we implemented image file encoding that involved processing parallelization into several threads at the same time. For this purpose, the possibility to partition the image into n parts was realized. Let us denote the vertical size (height) of the image by H, the horizontal size (width) by W. At this moment, the image is partitioned by two methods, depending on the number of parts into which the image should be partitioned. If the number of parts is equal to 4, the image is partitioned along the lines joining the midpoints of the image's opposite sides.

As a result, we obtain four rectangular areas P1, P2, P3, P4 with the coordinates ((x1, y1), (x2, y2)) of diametrically opposite vertices P1=((0,0), (W/2, H/2)), P2=((W/2,0), (W,H/2)), P3=((0,H/2), (W/2,H)), P4=((W/2,H/2),(W,H)). If $n \neq 4$, vertical sides of the image are divided into n-1 equal parts with a height equal to H/n, and one with a height (HH/n*(n-1)), the image being divided by opposite divisions of these edges. This results in n-1 rectangular areas P[i], (i=0..n-1) with the coordinates Pi = ((0,(H/n)*i),(W,(H/n)* (i+1))) and one rectangular area P[n] = ((0,(H/n)*(n-1)),(W,H)). After partitioning the image into several parts, each part of the image is compressed as a separate image in a separate thread. This approach allows the potential of modern computers to be utilized more fully by evenly distributing the load on the processor cores. In this case, the compression efficiency is insignificantly reduced. When making calculations on a quad-core PC1 (CPU - Core i5-3230M 2.6 GHz) and a dual-core PC2 (CPU - Core 2 Duo E7400 2.8 GHz), the following results in the Table.I were obtained for compression of the file artificial.bmp (3072x2048, color depth of 8 bits, 256 colors) from the set [5].

The first column of Table 1 shows the number of threads used to encode the image. In the second and the third columns, the corresponding encoding time is shown in seconds for PC1 and PC2. The last column shows the Rc of the file, depending on the number of threads used. It can be seen from the experiments that 4 threads provide better encoding efficiency when using dual-core and quad-core computers.

TABLE I. COMPARISON OF ENCODING TIME FOR THE FILE ARTIFICIAL.BMP DEPENDING ON THE NUMBER OF THREADS

| Number of threads | Time PC1 (sec.) | Time PC2 (sec.) | Rc |
|---|---|---|---|
| 1 | 7,94 | 93,657 | 13,301 |
| 2 | 6,563 | 11,141 | 13,173 |
| 4 | 4,671 | 16,187 | 13,041 |
| 8 | 4,494 | 29,656 | 12,766 |
| 16 | 4,323 | 27,515 | 12,344 |

## V. MULTI-PLATFORM FEATURE

Currently, a number of operating systems other than Windows find wide application in the world. A large number of devices use operating systems based on Linux, Unix, and we also witness widespread use of operating systems developed by Apple. Therefore, it is very important to have the possibility to write a program once and to have several versions of it for different operating systems. For this purpose, the program was transferred to the multi-platform program development environment Qt 4.8. This transfer has added a multi-platform feature to this application, improved its performance and made the program structure more flexible.

## VI. PECULIARITIES OF THE ENCODING ALGORITHM IMPLEMENTATION

We use the context models of the 8th, 5th, 2nd, 1st and zero order. Pixels in the context of the 8th (maximum) order are arranged in a certain sequence that allows us to calculate contexts of a lower order n from a context of a higher order N(N>n). This approach allowed us to achieve a significant gain in the algorithm performance. In the context model of the zero order, pixel counters are equally probable at initialization and are equal to unity.

In order to further increase Rc and improve the algorithm performance, the method of exclusion was applied [1,3,4]. To improve the accuracy of estimates of symbols in context-based models of high order, we used the method of information inheritance [1,3,4].

An arithmetic coder is used as a statistical coder [1]. For efficient storage of context models in the memory, a B-tree is used. A context-base model for a specific context is created and stored in the tree only after the given context was encountered in image processing for the first time.

The decoding process is symmetrical to the encoding process.

## VII. RESULTS

We have developed a program that implements the proposed algorithm and enables the compression of indexed images.

To test the effectiveness of the algorithm, experiments were conducted on two sets of test images [5] and comparison was made with the most common universal and specialized algorithms for image compression: JpegLS, PNG, 7z, rar. These experiments have demonstrated that the algorithm developed is significantly superior to other special-purpose and universal lossless compression algorithms (its compression ratio being on average 1.15 to 2.13 times higher).

This program requires an average of 50 - 500 MB of RAM to compress an image up to 6 Mpix, and the speed of encoding/decoding is approximately equal to 0.5 megapixels per second on a computer with a Core 2 Duo processor 2.8MHz using 2 threads.

## REFERENCES

[1] Vatolin D., Ratushnyak A., Smirnov M., Yurkin V. Methods of data compression. Construction of archivers, image and video compression. - Moscow: DIALOG - MEPhI 2003

[2] Vasin Yu.G. and Zherzdev S.V. Information Techniques for Hierarchical Image Coding // Pattern Recognition and Image Analysis, Vol. 13, №. 3, 2003, pp. 539–548.

[3] Borusyak A.V., Vasin Yu.G., Zherzdev S.V. “Compression of Binary Graphics Using Context Simulation” // Pattern Recognition and Image Analysis, Vol.23, № 2, 2013, pp207-210

[4] Borusyak A.V., Vasin Yu.G., Zherzdev S.V. “Optimizing the computational complexity of the algorithm for adaptive compression of binary raster images” The 11-th International Conference “Pattern Recognition and Image Analysis: new information technologies” Samara, September 23-28, 2014, pp.170-172

[5] A set of test images from the website http://www.imagecompression.info/test_images

# Conjugacy indicator for hyperspectral image thematic classification*

V. Fursov
Image Processing Institute
of the Russian Academy of Science
Samara, Russia
fursov@smr.ru

S. Bibikov and O. Bajda
Samara State Aerospace University
(National Research University)
Samara, Russia
bibikov.sergei@gmail.com

***Abstract*—We consider an algorithm of hyperspectral images thematic classification using conjugacy indicator as a proximity measure. We use the cosine of an angle between considered vector and subspace, which is spanned by class vectors, instead of spectral angle mapper. Paper describes modification of a method based on partitioning of the class into subclasses and based on reduction of vectors to zero mean value.**

***Keywords— hyperspecter imagery, classification, specter angle mapper, conjugacy indicator***

## I. PROBLEM STATEMENT

The problem of thematic classification of hyperspectral image has attracted a lot of attention recently. The number of publications on this topic is rapidly growing [1], [2], [3]. The most popular software, which allows to use a great number of functions for analyzing hyperspectral images, is ENVI [4]. One of the algorithms suggested in the program – the method of spectral angle – helps to solve the problem of thematic classification of spectral images [5], [6].

In general, the problem of thematic processing is posed as follows. Vector $N\times 1$ characterizing sample $j$ is introduced in the following way:

$$\mathbf{x}_j = \left[ x_1(j), x_2(j), ...., x_i(j), ..., x_N(j) \right]^T, \quad (1)$$

where $x_i(j)$ − value of the registered object's intensity of reflection in spectral band $i$ at point $j$ of hyperspectral image.

The task is to construct decision function $f: R^n \mapsto \{0,1,2,...,k\}$ to decide which class each vector $\mathbf{x}_j$ belongs to. It is assumed that $M$ of training vectors for each class are prescribed, i.e. $N\times M$ matrix is known:

$$\mathbf{X}_k = \left[ \mathbf{x}_1(k), \mathbf{x}_2(k), ..., \mathbf{x}_M(k) \right], \quad k = \overline{1,K} \quad (2)$$

The ENVI algorithm of classification, based on the method of spectral angle, is described in the following way. For each class, vector $\bar{\mathbf{x}}(k)$ is calculated:

$$\bar{\mathbf{x}}(k) = \frac{1}{M}\sum_{j=1}^{M} \mathbf{x}_j(k), \quad k = \overline{1,K}. \quad (3)$$

This vector characterizes class $k$. If maximum of spectral angle

$$\Theta = \cos^{-1}\left( \sum_{i=1}^{N} x_i(j)\bar{x}_i(k) \left( \sqrt{\sum_{i=1}^{N} x_i^2(j) \sum_{i=1}^{N} \bar{x}_i^2(k)} \right)^{-1} \right). \quad (4)$$

is reached, then the feature vector $\mathbf{x}_j$ and corresponding point $j$ are referred to class $k$

In this paper, we study the algorithm, which can be considered as generalization of SAM. That means that instead of measuring the angle between a tested vector and a prototype vector, we suggest calculating the angle between the tested vector and the subspace spanned by multiple vectors of the same class.

This method was first described in [6] and was developed further in [7], [8]. A number of studies [9], [10], [11], [12] consider the use of the method for face recognition. In this paper, we provide the results of experiments with test samples of hyperspectral images. They show much higher quality of classification in comparison with those where the method of spectral angle was used.

## II. DESCRIPTION OF CLASSIFICATION ALGORITHM

The suggested classifier is based on the use of a so-called indicator of conjugacy with the subspace spanned by the feature vectors of images from a given class. Let $\mathbf{x}_j$ be a feature vector (a prototype of sample $j$), which is tested for proximity to class $k$, and $\mathbf{X}_k$ is $N\times M$ matrix (2), made of training vectors (samples) of the given class.

*Work is supported by Russian Science Foundation grant # 14-31-00014.*

The algorithm is based on the following equations. For each class $k$ $N \times M$ - matrix $\mathbf{Q}_k$ is formed:

$$\mathbf{Q}_k = \mathbf{X}_k \left[ \mathbf{X}_k^T \mathbf{X}_k \right]^{-1} \mathbf{X}_k^T, \quad k = \overline{1,K}. \tag{5}$$

Decision function $f(\mathbf{x})$ is described as follows: Vector $\mathbf{x}_j$ is a tested sample, and it belongs to class $m$, i.e. $f(\mathbf{x}) = m, \quad m = 1,2,...k$,

if

$$R_m(j) = \max_{\forall k} R_k(j), \tag{6}$$

where

$$R_k(j) = \frac{\mathbf{x}_j^T \mathbf{Q}_k \mathbf{x}_j}{\mathbf{x}_j^T \mathbf{x}_j}. \tag{7}$$

So in order to construct the classifier, it is necessary to form $N \times M$ -matrix $\mathbf{Q}_k$ (2), using (5) for each class at the training stage. It is important to emphasize that the maximum conjugacy indicator $R_k(j)$ of $\mathbf{x}_j$ is calculated for all classes of hyperspectral image. It is clear that if we use this algorithm, the classification quality will largely depend on the way matrices of classes are formed (2). The simplest way of forming matrices $\mathbf{X}_k$ is a random choice of given numbers of vectors on some image area, which belongs to a certain class. In order to fully characterize a class, the number of class vectors should be large enough. In this case, the computational complexity of the algorithm is increasing, which may lead to classification errors due to a decrease in the condition of matrix $\mathbf{X}_k^T \mathbf{X}_k$.

In order to decrease computational complexity and improve the classification quality however, the classes may be divided into their subclasses. The division procedure has several steps: the first is to choose two vectors $x_1, x_M$ from initial class $M$. These vectors should be most different from each other, where $R_{1,M} = \langle x_1^T x_M \rangle / \|x_1\| \|x_M\|$ is minimal (8). As step 2, from the rest of $M$, another pair of vectors $x_2, x_{M-1}$ are chosen to add to the first pair vectors, where

$$R_{1,2} = \langle x_1^T x_2 \rangle / \|x_1\| \|x_2\|, \tag{9}$$

$$R_{M-1,M} = \langle x_{M-1}^T x_M \rangle / \|x_{M-1}\| \|x_M\| \tag{10}$$

have maximum values. The combinations of vectors $x_1, x_2$ and $x_{M-1}, x_M$ for two subspaces described by matrices $\mathbf{X}_{1,2}$ and $\mathbf{X}_{M-1,M}$ respectively. The next step (which is an iterative one) is to choose two vectors $x_3, x_{M-2}$ which are closely conjugated with the formed subspaces. The following equations illustrate the criteria of proximity:

$$R_{1,2,3} = \frac{\mathbf{x}_3^T \mathbf{X}_{1,2} \left[ \mathbf{X}_{1,2}^T \mathbf{X}_{1,2} \right]^{-1} \mathbf{X}_{1,2}^T \mathbf{x}_3}{\mathbf{x}_3^T \mathbf{x}_3}, \tag{11}$$

$$R_{M-2,M-1,M} = \frac{\mathbf{x}_{M-2}^T \mathbf{X}_{M-1,M} \left[ \mathbf{X}_{M-1,M}^T \mathbf{X}_{M-1,M} \right]^{-1} \mathbf{X}_{M-1,M}^T \mathbf{x}_{M-2}}{\mathbf{x}_{M-2}^T \mathbf{x}_{M-2}}. \tag{12}$$

As a result of Step $M/2$ (in case where $M$ is even) or Step $(M-1)/2$ (when $M$ is odd), two subclasses are formed. It is possible to divide each of these subclasses into two other subclasses (having four as total), using the same procedure. When the classified vector is referred (6) to one of the subclasses, it is considered that it belongs to the initial class. The given procedure improves the classification quality.

Further improvement of the classification quality can be reached with the help of pre-processing. i.e. subtraction of the mean vector from hyperspectral image. The mean vector is obtained by calculating the mean vectors in each band of the image. Such pre-processing increases angles between the vectors belonging to different classes. This increase leads to better discernibility of the classes. It should be noted that the feature vectors should have the same sizes, and their components should characterize the corresponding spectral band.

## III. Materials and Results of the Experiment

### A. Materials

To test the efficiency of the suggested method, an attempt was made to solve the problem of thematic classification of images. For hyperspectral analysis, an available in freeware MultiSpec image was used. This image had been obtained within the project AVIRIS (Airborne Visible/ Infrared Imaging Spectrometer). The image shows the test field Indian Pines located in the north-west of Indiana, the USA. The image pictures the main highway, railways, dwellings, different constructions, country roads, a forest, and agricultural fields. The size of the image taken by AVIRIS is 145×145 pixels, where each pixel contains 224 spectral samples in the band 0.4-2.5 mkm.

The MultiSpec developers also offered the same image, but containing 200 bands with the division into 16 classes. There is an unmarked area on the image that does not refer to any of the 16 classes. We did not use this area in the experiment. The classified image is provided in Figure 1, where the unused area is of white color.

### *B. Procedure*

In the experiment, two methods were compared to test the classification quality: the method of conjugacy indicator and the method of spectral angle. The criterion of quality was the ratio of accurately classified pixels of the image to the total number of pixels of the image.

To evaluate the quality of classification, we used the procedure of stratified cross-validation as there was only one image used in the experiment and no prior information about training data was available. This procedure allows to divide the original data by $N$ ways into two non- overlapping subsets of data, one of which is a training data subset, the other is a test data subset. As a result, we get $1/N$ of the test vectors, and $(N-1)/N$ of the training vectors.

However, the randomly obtained subsets have a serious disadvantage: the quantity of vectors belonging to different classes is not balanced. In this case, when there are not enough training vectors of some classes (or no such vectors at all), the quality of classification will significantly decrease. To solve this problem, we firstly divided the original data to get equal fractions of different classes in both training and test sets, and then used the procedure of stratified cross-validation. Thus, it was possible to evaluate the generalization ability of the algorithm.

The results of the comparative study of stratified cross-validation with the use of the methods of conjugacy indicator and spectral angle are shown in Table 1, as well as in Figures 1*a* and *b*, where classification errors are indicated with white pixels.

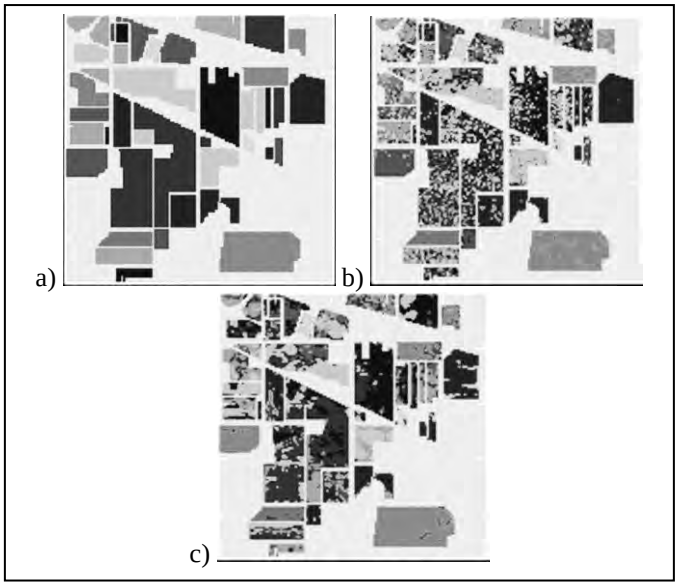



Fig. 1. Image divided into classes: a) original image; b) classified image using SAM; c) classified image using conjugacy indicator

TABLE I. TABLE 1 – RESULTS OF CLASSIFICATION

| № of experiment | % accurate classifications | |
|---|---|---|
| | *Conjugacy indictor* | *Spectral angle* |
| 1 | 62,7 | 50,9 |
| 2 | 64,1 | 48,7 |
| 3 | 61,0 | 48,7 |
| 4 | 61,6 | 49,4 |
| 5 | 65,2 | 50,1 |
| Mean result | 62,9 | 49,6 |

In addition, the procedure of class division into subclasses (see (8) – (12)) was also considered in the study. Table 2 and Figure 2a illustrate the results of the classification with the use of subclasses. White pixels again indicate errors of classification.

TABLE II. RESULTS OF CLASSIFICATION WITH PRIOR DIVISION OF CLASSES INTO TWO SUBCLASSES

| № of experiment | % of correct classifications |
|---|---|
| 1 | 67,1 |
| 2 | 68,7 |
| 3 | 66,4 |
| 4 | 67,7 |
| 5 | 66,4 |
| Mean result | 67,3 |

We also tested the dependency of subtraction of the mean vector on the classification quality. Table 3 and Figure 2b show the results of classification based on the method of conjugacy indicator with subtraction of the mean vector and division into subclasses.

TABLE III. RESULT WITH SUBTRACTION OF MEAN VECTOR AND DIVISION INTO SUBCLASSES

| № of experiment | % correct classifications |
|---|---|
| 1 | 71,2 |
| 2 | 71,8 |
| 3 | 71,2 |
| 4 | 74,0 |
| 5 | 69,6 |
| Mean | 71,6 |

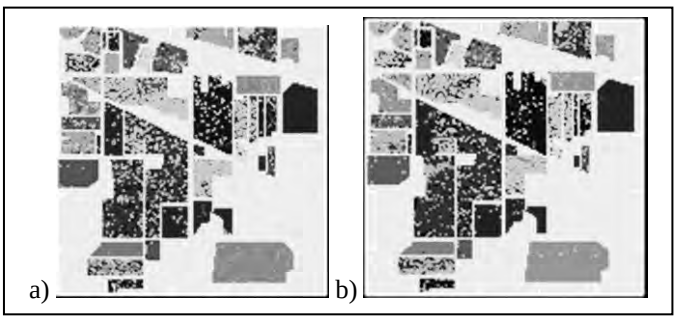



Fig. 2. Fig. 2. Result of classification: a) with division into two subclasses; b) with subtraction of mean vector and division into subclasses

## REFERENCES

[1] A.V. Kuznetsov and V.V. Myasnikov, "A comparison of algorithms for supervised classification using hyperspectral data," Computer optics, vol. 38, no. 3, pp. 494-502, 2014 (in Russian).

[2] R.A. Schowengerdt, "Remote sensing: models and methods for image processing," Academic Press, 2006, 560 p.

[3] L.N. Chaban, G.V. Vecheruk, T.V. Kondranin, S.V. Kudriavtsev, and A.A. Nikolenko, "Modeling and thematic processing of images identical to the imagery from workable and preparing for the space launch hiperspectral remote sensors," Current problems in remote sensing of the Earth from space, vol. 9, no. 2, pp. 111-121, 2012 (in Russian).

[4] ENVI 4.1 User's Guide, Research System Inc., 2004, 1150 p.

[5] O. A. de Carvalho, and P. R. Meneses, "Spectral correlation mapper: an improvement on the spectral angle mapper (SAM)," Summaries of the 9th JPL Airborne Earth Science Workshop, JPL Publication 00-18, Pasadena, CA: JPL Publication, 2000, vol. 9, 9 p.

[6] H. Z. M. Shafri, S. Affendi, and M. Shattri, "The performance of maximum likelihood, spectral angle mapper, neural network and decision tree classifiers in hyperspectral image analysis," Journal of Computer Science, no. 3(6), pp. 419-423, 2007.

[7] V. A. Fursov, "Training in pattern recognition from a small number of observations using projections onto null-space," Proc. 15th International Conference on Pattern recognition (ISPR) 2000, Barcelona, Spain, vol. 2, pp. 119-121, 2000.

[8] V. A. Fursov, I. A. Kulagina, and N. E. Kozin "Building of classifiers based on conjugation indices," Optical Memory and Neural Networks (Information Optics), vol. 16, no. 3, pp. 136-143, 2007.

[9] V. A. Fursov, I. A. Kulagina, and N. E. Kozin, "Building of classifier based on conjugation indexes," Proceedings of The 5-th International Conference on Machine Learning and Data Mining, Leipzig, Germany, 18-20 July, 2007, pp. 231-235, 2007.

[10] N.E. Kozin, and V.A. Fursov, "Constructing of classifier for face recognition using conjugation indexes," Computer optics, no 28, pp. 160-163, 2005 (in Russian).

[11] N.E. Kozin, and V.A. Fursov, "Constructing of Classifier for Face Recognition on the Basis of the Conjugation Indexes," Transactions on Engineering Computing and Technology, vol. 13, pp. 72-74, 2006.

[12] V.A. Fursov, and N.E. Kozin, "Recognition through constructing the eigenface classifiers using conjugation indices," 2007 IEEE International Conference on Advanced Video and Signal based Surveillance London (United Kingdom), 5-7 September 2007, pp. 465-469, 2007.

# Construction of the hybrid intelligent system of express-diagnostics of information security attackers based on synergy of several sciences and scientific directions

A.E. Yankovskaya
Tomsk State University of Architecture and Building, National Research Tomsk State University, Tomsk State University of Control Systems and Radioelectronics
TSUAB, TSU, TUSUR
Tomsk, Russia
ayyankov@gmail.com

A.A. Shelupanov, V.G. Mironova
Tomsk State University of Control Systems and Radioelectronics
TUSUR
Tomsk, Russia
saa@tusur.ru, mvg@security.tomsk.ru

***Abstract*—The paper is devoted to construction of the hybrid intelligent system (IS) of express-diagnostics of information security attackers (HIS DIVNAR) based on synergy of several sciences and scientific directions: test pattern recognition; discrete mathematics; threshold and fuzzy logic; FSM theory; theory of separating systems; theory of probability and mathematical statistics; artificial intelligence; logic-combinatorial (LC) and LC-probabilistic algorithms; fault-tolerance; soft computing; reliability; graphic cognitive means based on two approaches: naturalistic and without using display in ordinary reality. A proposed approach and basis of the mathematical apparatus are fragmentarily given. HIS DIVNAR consists of four subsystems. Further development of this approach is proposed.**

***Keywords— hybrid intelligent system, express-diagnostics, information security, attackers, organizational stress, depression, deviant behavior, synergy, test pattern recognition, fault-tolerance, logic-combinatorial (LC) and LC-probabilistic algorithms, threshold and fuzzy logic, cognitive means.***

## I. INTRODUCTION

A comprehensive use in the modern society of information technologies dictates the need in development of reliable information security systems. Information on attackers and threats to information security (InS) is necessary in creation of a reliable information security system (InSS). In this regard, the relevance of the problem of locating the InS attacker based on newest information technologies is uncontroversial. Currently, such technologies are unknown to us. Naturally, these technologies should be based on effective methods of data and knowledge analysis (DKA) [1] and decision-making, implemented in intelligent systems (IS) designed for detection of persons (subjects) – attackers of information security that could negatively affect and inflict harm to information and resources of information systems.

A model of the attacker [2] is needed in systematization of the information on types and capabilities of subjects, purposes of unauthorized attacks, and development of diagnostic decisions (quantitative assessment of the probability of security threats) and appropriate organizational and technical counteractions against attackers.

Development of hybrid IS (HIS) that combine several methods of representation and processing of data and knowledge on the basis of synergy of several sciences and scientific directions, for the first time proposed for data and knowledge analysis in [1], is absolutely necessary due to the fact that diagnostic decisions depend on a very large number of features, divided into 4 groups, which determine the presence in the investigated subject (subject under investigation) of organizational stress (OS), depression, deviant behavior, and a group of symptoms that defines characteristics of the subject from the viewpoint of a possibility of InS violations. Since different organizations, institutions, and government agencies use a great number of information systems, the task of creating HIS of express – diagnostics of IS violators (HIS DIVNAR) [3] allowing to quickly identify individuals capable of inflicting harm to information and to reduce the efficiency of the information infrastructure of these organizations, government institutions, and agencies, is extremely essential.

## II. MATHEMATICAL BASIS FOR CONSTRUCTION OF HIS OF EXPRESS – DIAGNOSTICS OF THE IS ATTACKER

Account of different requirements imposed at decision-making (minimization of the feature space; cost; damage (risk) inflicted in measurement (entry) of characteristic features values of the investigated subject; errors at entry deliberately made by the investigated subject; assurance of the predesigned reliability, etc.) leads to their account in the process of regularities revealing at DKA and decision-making [4]. Consequently, substantiation of implementation of the synergy of several sciences and scientific directions at construction of the HIS is uncontroversial.

The term "synergy" refers to an interpenetration of several sciences and scientific directions. The more the interpenetration, the higher the degree of convergence and the possibility of obtaining a higher synergistic effect [1]. Account of multiple requirements leads to the need for implementation of a coordinate alignment of a totality of individual items (several sciences): test pattern recognition [4, 5]; discrete mathematics; theory of FSM, theory of probability and mathematical statistics; artificial intelligence; logic-combinatorial (LC), LC-probabilistic algorithms [4]; fuzzy and threshold logic; theory of separating systems [4, 6]; fault-tolerance [7, 9]; soft computing; reliability[1, 4, 6, 10]; probability theory; graphic cognitive means based on two approaches: naturalistic and without using display in ordinary reality [4, 9].

The most important means of DKA are diagnostic tests (DT) [1, 4, 9] constructed at revealing of various kinds of regularities and used in decision-making in IS based on test pattern recognition methods [4, 7-10].

We shall outline mathematical basis of constructing HIS DIVNAR.

It is proposed to describe the IS attacker in the space of the four abovementioned groups of characteristic features (CF).

The idea of a three-stage diagnostics of OS [11] and the mathematical apparatus based on threshold and fuzzy logic, as well as graphic, including cognitive, means of visualization of information structures and results of substantiation of diagnostic and interventional decisions, have been used in identification of OS.

The same mathematical apparatus is used in identification of depression [12] as in identification of the OS. The same mathematical apparatus is proposed to be used in identification of persons with a deviant behavior, determined by such CFs as internal negative attitude towards social requirements; proneness to conflicts or weak development of communication skills; cognitive distortions of reality; abuse of substances causing a state of altered mental activity (alcoholism, drug addiction, smoking, etc.) aggression; immoral and amoral behavior.

For construction of the fourth component –IS of express – diagnostics of the IS attacker (IS DINARLOG2 ) in the reduced features space we shall use the CF included in the optimal subset of fault-tolerant unconditional irredundant diagnostic tests (FT UIDT) [8, 10] and FT of mixed DTs [4] representing the compromise between unconditioned and conditioned components [4], and constructed on its basis decision-making rules in a broad feature space using the applied IS DINARLOG1 constructed on the basis of the intelligent instrumental tool (IIT) IMSLOG [13] based on the matrix method of data and knowledge representation (description matrix Q in the space CF and distinction matrices R in the space of classification features (CF)) [4, 9].

It shall be noted that rows of matrices Q, R are associated with objects (subjects) of the training sample.

At construction of the description matrix for IS DINARLOG1 the CFs will include: territorial location of the IS attacker relative to the boundaries of the controlled area; the presence of access to the system by the attacker; the presence of rights to perform legal activities in IS; the use of vulnerable to attack the system and acquisition information; availability by the attacker of the information on the system and/or ISS; availability of documents regulating the operational procedure of a user in the system in terms of information security; professional qualities of IS attacker; moral qualities; the presence of an ally and a number of other CFs; as well as features resulting from diagnostics of OS, depression and deviant behavior.

For the distinction matrix R we shall list classes and values of their elements: k1 is the possibility of committing destructive actions: 1 - unable, 2 - borderline (the possibility of committing actions under conditions), 3 - able; k2 is the presence of conditions for committing destructive actions: 1 - present, 2 - absent; k3 is the probability of committing destructive actions: 1 - low, 2 - little, 3 - average, 4 - high, 5 - very high.

A number of previously revealed regularities based on IS DINARLOG1 will be taken into account at decision-making.

The result of the decision-making by IS DINARLOG2, as well as by IS DIVNAR, is the reference of the investigated subject to one of three classes: an unlikely IS attacker, an attacker of the average probability, a highly probable attacker.

If one of the requirements for decision-making is the account of reliability, then at revealing of regularities it is transformed to the necessity for constructing of fault-tolerant DTs, i.e. tests that are tolerant to a given number t of measurement errors (entry) of CFs, or DTs resulting in a minimum probability of error at decision-making.

Since the scope of the report does not allow to represent entirely the mathematical apparatus, we shall present only necessary and sufficient, as well as sufficient conditions for the consistency of data and knowledge and decision-making, tolerant to measurement errors and entry of characteristic values.

**Theorem.** To ensure consistency of data and knowledge and decision-making, tolerant to a number not exceeding t of measurement errors (entry) of CFs, in the matrix Q for each pair of objects from different classes at a fixed classification mechanism it is necessary and sufficient to ensure conditions $h=2t+1$ for any row of the implication matrix U, defining the distinction of objects from different classes at each classification mechanism.

To formulate a sufficient condition to ensure the consistency of data and knowledge and decision-making, tolerant to a number not exceeding t of measurement errors, we shall use the implication matrix constructed on the basis of consideration of object pairs from different patterns and denoted hereinafter by U. It is believed that objects with the same combination of CF values belong to the same pattern [1, 10].

**Confirmation.** In confirmation of the theorem, replacement of the analysis of each pair of objects from different classes in the considered classification mechanism on

the analysis of each pair of objects from different patterns (pattern-pattern, object-pattern, object-object) does not break its sufficiency in construction of the matrix U based on patterns.

## III. Brief description of HIS of express - diagnostics of the IS attacker

The developed by us HIS of express – diagnostics of the IS attacker (HIS DIVNAR) consists of four subsystems, a block diagram of which given on Fig. 1.

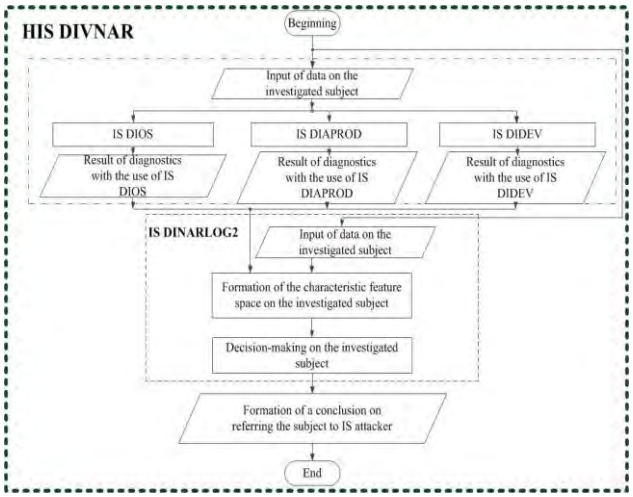



Fig. 1. Block diagram of HIS DIVNAR

Let's describe these subsystems: the 1st is IS DIOS designed for diagnostics of OS of the investigated person (subject), the 2nd – IS DIAPROD designed for diagnostics and prevention of depression, the 3rd – IS DIDEV designed for diagnostics of deviant behavior, and the 4th IS of a probable IS attacker (IS DINARLOG2) based on the results obtained from 1st 3rd subsystems and additionally formed CFs by means of additional inspection of the subject, provides support at decision-making on referring the subject to the attacker. IS DINARLOG2 is designed only for making and substantiation of decisions, unlike IS DINARLOG1 which is designed for revealing of different kinds of regularities, decisions-making and it's substantiation using graphic, including cognitive, means.

## IV. Conclusion

The urgency of construction of the HIS of express - diagnostics of the IS attacker, allowing to quickly identify persons capable to inflict harm to information and reduce the efficiency of the information infrastructure of organizations, government departments, and agencies, is formulated.

The mathematical apparatus of its creation, based on synergy of several sciences and scientific directions, is briefly outlined. Necessary and sufficient, as well as sufficient conditions for the consistency of data and knowledge and decision-making that are tolerant to measurement errors of characteristic values, are formulated.

Account of the presence of measurement errors (entry) of characteristic values that define the description of objects increases the reliability and the quality of decision making.

Implementation on the basis of IIT IMSLOG [13], IS DINARLOG1, IS DINARLOG2 will allow to construct HIS DIVNAR, which will broaden the scope of practical application of the created IS and HIS.

Further envestigation will be focused on construction of FT UIDT and decision-making based on synergy of several sciences and scientific directions [1], as well as on increase of the effectiveness of algorithms. First steps in this direction are presented in [3].

## Acknowledgment

Russian Foundation for Basic Research, projects no. 13-07-00373, 13-07-98037 and by Russian Humanitarian Foundation project no. 13-06-00709.

[1] A.E. Yankovskaya, "Analysis of data and knowledge based on synergy of several sciences and scientific directions", Intellectualization of information processing. Book of abstracts of the 8th Intern. Conf., M.: MAKS Press, pp. 196-199, 2010. (In Russian)

[2] V.G. Mironova, A.A. Shelupanov, "The model of an information security attacker", Informatics and Control Systems. № 31, vol. 1, 2012, pp. 28-35. (In Russian)

[3] A.E. Yankovskaya, A.A. Shelupanov, and V.G. Mironova, "Bases of intelligent system creation of probable attacker express – diagnostics of information security", Proc. of Intelligent Systems and Information Technologies Congr. (AIS-2014), Moscow, vol. 2, pp. 277-284, 2014. (In Russian)

[4] A.E. Yankovskaya, Logic tests and means of cognitive graphics, Saarbrucken, Germany: LAP Lambert Academic Publishing GmbH & Co. KG, 2011, p. 92 (In Russian)

[5] Yu.I. Zhuravlev, Pattern recognition and image analysis, Artificial intelligence. vol. 2. Models and methods, M.: Radio and communication. 1990, pp. 149-191. (In Russian)

[6] Yu.L. Sagalovich, Separating systems, M.: IITP RAS, 2012, p. 130. (In Russian)

[7] A.E. Yankovskaya, "Logic-combinational probabilistic recognition algorithms", Pattern Recogn. and Image Analysis, № 1, vol. 11, 2001, pp. 123–126.

[8] A.E. Yankovskaya, S.V. Kitler, "Intelligent system for parallel fault-tolerant diagnostic tests construction", Journal of Software engineering and Applications, № 4A, vol. 6, 2013, pp. 54-61.

[9] A. E. Yankovskaya, "An automaton model, fuzzy logic, and means of cognitive graphics in the solution of forecast problems", Pattern Recognition and Image Analysis, № 2, vol. 8, 1998, pp. 154–156.

[10] A.E. Yankovskaya, "Decision-making resistant to measurement errors of characteristic values in intelligent systems", Artificial intelligence. Intelligent systems (II-2009). Materials of the Xth International Scientific Conference, Taganrog: Publishing TTI SFEDU, 2009, pp. 127-130. (In Russian)

[11] A.E. Yankovskaya, S.V. Kitler, and R.V. Ametov, "Development and investigation of the intelligent system for diagnostics and intervention of organization stress", Pattern Recognition and Image Analysis, vol. 23, No 4, 2013, pp. 459-467.

[12] A. Yankovskaya, A. Kornetov, N. Ilyinskikh, A. Silaeva, and V. Obuhovskaya, "Psychodiagnostic data and knowledge structuring for an intelligent system for depression diagnosis and prevention", V International interdisciplinary academic conference: Innovations and humans. Turkey, Antalya, April 26 – May 7, 2014, pp. 114-120. (In Russian)

[13] A.E. Yankovskaya, A.I. Gedike, R.V. Ametov, and Bleikher A.M., "IMSLOG-2002 software tool for supporting information technologies of test pattern recognition", Pattern Recognition and Image Analysis., vol. 13, №. 4, 2003, pp. 650-657.

# Deconvolution and correction based approach to restore images captured using simple Fresnel lenses

A. Nikonorov, N. Kazansky
Image Processing Systems Institute of RAS
Samara, Russia
artniko@gmail.com, kazansky@smr.ru

M. Petrov, S. Bibikov
Samara State Aerospace University
Samara, Russia
max.vit.petrov@gmail.com, bibikov.sergei@gmail.com

***Abstract*—A good practical challenge is to create computational correction procedure for images captured by Fresnel (diffractive) lenses, which allows these lenses to become imaging lenses. This paper present a technique for restoring images captured with single Fresnel lens. Image restoration is made using several steps: deconvolution, edge analysis and color correction. The first two steps are based on combination of restoration techniques used for restoring images, which were obtained through simple refraction lenses. Color correction step is necessary to remove strong color shift caused by chromatic aberrations of simple Fresnel lens. This complex technique was tested on the real images, captured by a simple lens, which was made as three-step approximation of the Fresnel lens.**

***Keywords — simple Fresnel lens imaging; chromayic aberration; deconvolution; deblur;color correction***

## I. INTRODUCTION

Modern camera lenses became very complex. They consist of more than dozen elements, which are necessary for removing optical aberrations. Recently, simple lenses with one or two optical elements were proposed [1]. Such lenses are similar to lenses used hundreds years ago, and chromatic aberration is still a big problem for images captured using simple lenses [1]. The optical correction to such images is usually made using methods of digital processing.

A chromatic aberration is a correlation between optical system characteristics and wavelength of registered light. Chromatic aberrations result in appearing a chroma in achromatic objects and/or in coloring the contours.

The algorithmic correction of such kind of aberrations is the postprocessing of distorted images using one of two different approaches. The first approach is based on blind or semi-blind deconvolution using PSF estimation. Another uses an analysis of contours in different color channels of the image [5] to remove aberrations. In paper [7], a combination of these approaches is used.

This paper considers using these approaches to improve images obtained with Fresnel lenses. This type of lenses is described in [6]. Such lenses can be defined as a stepped approximation of the Fresnel lens (Fig. 1). $\varphi(u,v)$ in Fig. 1 is the phase shift. Fresnel lenses can be created by consecutive etching with several different binary masks.

*Work is partially supported by RFBR #13-07-13166 OFI-M RZhD, and Ministry of Education and Science of Russian Federation by SSAU Roadmap Program (task 1.4) and project 14.575.21.0083.*

Fresnel lenses have advantages over refractive lenses in weight and linear sizes. Obviously, these advantages are greater in case of long-focal-length cameras, where a complex set of refractive lenses can be replaced with only one Fresnel lens. However, there is a disadvantage of using Fresnel lenses: the dependence of image blurring on light wavelength and some other distortions like a moire. Therefore typical usage of Fresnel lenses is optical collimator or concentrator but not an imaging lens [8].

From a perspective of aberration correction, simple Fresnel lenses have the following features. Such lenses have much stronger chromatic aberration than simple refractive lenses do. One of the color channels (in this paper we use the green channel) has less blurring and can be used as a reference channel for the correction of the two other channels.

A good practical challenge is to create correction procedure for images captured by Fresnel lenses, which allows these lenses to become imaging lenses. In this paper we propose the model for correction of chromatic aberrations in the images obtained with Fresnel lenses. Then we implement a technique for such images. This technique consists of deconvolution, edge analysis and color correction procedures. Finally we present experimental results for images captured using lens obtained as a three-step approximation of Fresnel surface.

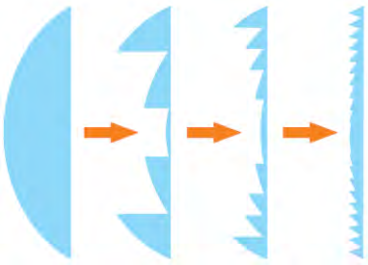


Fig. 1. Conceptual illustration of collapsing of aspheric refraction lens into Fresnel (diffractive) lens

## II. IMAGE CORRECTION SCHEME FOR FRESNEL LENSES

The chromatic aberration, which occurs in refraction lenses, is described by the general defocus model [1]. In this model, the point spread function (PSF) is supposed to be linear, at least in local spatial area, as shown in:

$$j_k(\mathbf{x}) = \mathbf{B} \otimes i_k(\mathbf{x}) + \mathrm{n}, \quad k = 1..3, \tag{1}$$

where $j_k(\mathbf{x})$ is the *k*-th color channel of the blurred image, and $i_k(\mathbf{x})$ is the channel of the underlying sharp image, $\mathbf{B}$ is a blur kernel, *n* is additive image noise.

Paper [2] shows that the lens PSF varies substantially being a function of aperture, focal length, focusing distance, and illuminant spectrum. So, blur kernel B in (1) being a constant is not accurate enough, especially for Fresnel lenses with strong chromatic aberration.

For such strong aberration kernel B is space-varying. There are two different types of distortions in the image: space-varying blur along the edges and color shift in the regions with plain colors. Therefore, to handle such distortions, we use the following modification of (1):

$$j_k(\mathbf{x}) = \mathbf{B}_k \otimes \left(i_k(\mathbf{x}) + D_k(i_k(\mathbf{x}))\right) + \mathrm{n}, \quad k = 1..3. \tag{2}$$

Here $D_k(i_k(\mathbf{x}))$ is a component characterizing the color shift. Color correction [4] of this shift is shown in Section V. Blurring kernels $\mathbf{B}_k$ in (2) are different for different color channels; let us call these kernels the chromatic blur.

According to (2), the correction consists of two stages – chromatic deblurring and the correction of the color shift:

$$i_k^1(\mathbf{x}) = \mathbf{B}_k^{-1} \otimes \left(j_k(\mathbf{x})\right), \quad k = 1..3, \tag{3}$$

$$i_k(\mathbf{x}) = F(i_k^1(\mathbf{x})), \quad k = 1..3. \tag{4}$$

Here operation $\mathbf{B}_k^{-1} \otimes$ is a deconvolution for the chromatic deblurring, with an intermediate image $\mathbf{i}^1(\mathbf{x})$ as a result. $F()$ is a color correction transformation.

In the present paper, we propose the following technique based on model (3) - (4):

1) the chromatic deblurring (3) of the green channel based on the deconvolution, described in Section III);
2) the chromatic deblurring (3) of the blue and red channels using the contours analysis (this approach is described in Section IV);
3) the color correction (4), which is described in Section V.

## III. PRIMAL-DUAL DECONVOLUTION ALGORITHM FOR CHROMATIC DEBLURRING

To solve the image deconvolution problem (4), we derive optimization methods based on the optimal first-order primal-dual framework by Chambolle and Pock [3]. In this section, we present a short overview of this optimization framework. We refer the reader to the original work by Chambolle and Pock [3] for an in-depth discussion.

Let X and Y be finite-dimensional real vector spaces for the primal and dual space, respectively. Consider the following operators and functions:

$\mathbf{K}: \mathrm{X} \to \mathrm{Y}$ is a linear operator from X to Y;

$\mathbf{G}: \mathrm{X} \to [0, +\infty)$ is a proper, convex, (l.s.c.) function;

$\mathbf{F}: \mathrm{Y} \to [0, +\infty)$ is a proper, convex, (l.s.c.) function, where l.s.c. stands for lower-semicontinuous.

The optimization framework of (5) considers general problems of the form

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \mathbf{F}(\mathbf{K}(\mathbf{x})) + \mathbf{G}(\mathbf{x}). \tag{5}$$

To solve problem in form (5), the following algorithm is proposed in paper [3].

Initialization step: choose - $\tau, \sigma \in R_+$, $\theta \in [0,1]$, $(\mathbf{x}_0, \mathbf{y}_0) \in \mathrm{X} \times \mathrm{Y}$ – some initial approximation, $\bar{\mathbf{x}}_0 = \mathbf{x}_0$.

Iteration step: $n \geq 0$, iteratively update $\mathbf{x}_n, \mathbf{y}_n, \bar{\mathbf{x}}_n$ as follows:

$$\mathbf{y}_{n+1} = prox_{\sigma F^*}(\mathbf{y}_n + \sigma \mathbf{K} \bar{\mathbf{x}}_n), \tag{6}$$

$$\mathbf{x}_{n+1} = prox_{\tau \mathbf{G}}(\mathbf{x}_n + \tau \mathbf{K}^* \mathbf{y}_{n+1}), \tag{7}$$

$$\bar{\mathbf{x}}_{n+1} = \mathbf{x}_{n+1} + \theta(\mathbf{x}_{n+1} - \mathbf{x}_n). \tag{8}$$

Following paper [5], a proximal operator with respect to **G** in (7), is defined as:

$$\begin{aligned} prox_{\tau \mathbf{G}}(\tilde{\mathbf{x}}) &= (\mathbf{E} + \tau \partial \mathbf{G})^{-1}(\tilde{\mathbf{x}}) = \\ &= \arg\min_{\mathbf{x}} \frac{1}{2\tau} \left\| \mathbf{x} - \tilde{\mathbf{x}} \right\|_2^2 + \mathbf{G}(\mathbf{x}) \end{aligned}, \tag{9}$$

where **E** is identity matrix. The proximal operator in (6) $prox_{\sigma F^*}$ is the same.

In order to apply the described algorithm to the deconvolution model, we follow [3]:

$$\mathrm{F}(\nabla i) = \left\| \nabla i \right\|_1, \tag{10}$$

$$G(i) = \left\| \mathbf{B} \otimes i - j \right\|_2^2. \tag{11}$$

Using (10) and (11), it is possible to obtain the proximal operators for steps (6) and (7) of the algorithm. Further details are available in [3]. The deconvolution algorithm based on the total variance has an ability to preserve sharp edges.

This deconvolution step is applied to the sharpest channel of the distorted image. Restoration of the other two channels is made using edge processing procedure, described in the next section.

## IV. Chroma Blur Correction Using Color Contours Processing

We propose a modification of the known algorithm [5]. The original algorithm uses an approach based on the contours analysis. One of the color channels is used as a reference channel. Let $Z(p)$ be a transition zone from one color to another. Let $l(p)$ be the left border, and $r(p)$ be the right border of such area. Obviously, in the transition area, an abrupt change of values in color channels (R, G, or both) occurs. The goal of the algorithm is to transform R and B signals into G signal in the transition area as closely as possible.

To do this, let us define differences between signals: $D_R$ is the difference between red and green channels, and $D_B$ is the difference between blue and green signals. For each pixel in area $Z(p)$, these differences must be fewer than the differences on the borders of the transition area. If this requirement is not fulfilled, R and G components of such pixels need to be corrected in one of the following ways.

If the color difference between red and green channels $D_R(k)$ at pixel $k \in Z(p)$ is more than maximum of color difference on the left and right borders

$$D_R(k) > \max(D_R(l(p)), D_R(r(p))),$$

then

$$R(k) = \max(D_R(l(p)), D_R(r(p))) + G(k).$$

If $D_R(k)$ is fewer than minimum of color difference on the left and right borders

$$D_R(\mathrm{k}) < \min(D_R(l(p), D_R(r(p))), \quad \text{then}$$

$$R(k) = \min(D_R(l(p)), D_R(r(p))) + G(k).$$

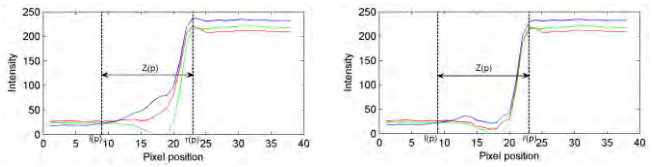
The correction of the blue channel is made in the same manner.

Therefore, color differences $D_R$ and $D_B$ are decreased, and R and B signals in transition area look more like G. An example of the result of the algorithm at work is shown in Fig. 2.

There are several pixels (#16-19) with near to zero values in green channel (Fig. 2(a)). This means that there is no significant information in green channel for correcting pixels in red and blue channels in this part of transition area. We propose some modifications of the algorithm to solve this problem:

1) Preprocessing the green channel in order to handle near to zero values. We use median filter, so we replace pixel with near to zero value with the middle value in the some window. If the new value is also near to zero, size of window is increased.
2) Using dilation after the correction.
3) Handling too bright pixels.

After correction of two color channels using color contours processing we use color correction, described in the following section, to remove strong color shift, caused by chromatic aberration.




(a) (b)

Fig. 2. Result of the algorithm work: (a) original image (b) image after color contours processing

## V. Color Correction for Chromatic Aberration Correction

The proposed chromatic aberration correction approach includes color correction in the final stage. A detailed description of the color correction approach is provided in [4]. As shown in [4], the problem consists of correcting non-isoplanatic deviation in illumination $I(\lambda, \mathbf{x})$ and restoring an image with the given illumination $I_0(\lambda)$:

$$\begin{aligned} &\mathbf{p}(\mathbf{x}) = \int R(\lambda, \mathbf{x}) I(\lambda, \mathbf{x}) \mathbf{T}(\lambda, \mathbf{x}) d\lambda \rightarrow \\ &\rightarrow \mathbf{p}_0(\mathbf{x}) = \int R(\lambda, \mathbf{x}) I_0(\lambda) \mathbf{T}(\lambda) d\lambda \end{aligned}, \qquad (12)$$

where *R* and *I* are $R \times Z_2 \rightarrow [0,1]$ functions of wavelength $\lambda$. *R* is the spectral reflectance reflectance of the scene surfaces. *I* is the spectral irradiance that is incident at each scene point. $\mathbf{T}(\lambda) = [T^{\{1\}}(\lambda), ..., T^{\{K\}}(\lambda)]^T$ is the spectral transmittance distribution of color sensors.

We propose using prior knowledge of the colors of small isolated patches in the image in the same way as a color correction specialist does. These small patches, limited in color and space, are defined in [4] as *spectral shape elements*, SSE.

Using SSE, the task of the correction function identification takes the following form:

$$\mathbf{a}^* = \arg\min_{\mathbf{a}} \left\| F(\mathbf{u}_i, \mathbf{a}), \mathbf{u}_i^0 \right\|, \qquad (13)$$

where $\{\mathbf{u}_i\}$ is a set of distorted SSE, and $\{\mathbf{u}_i^0\}$ is a set of distortion-free SSE.

Using SSE matching condition theorem from [4], the identification problem of the correction function looks like that:

$$\mathbf{a}^* = \arg\min_{\mathbf{a}}((F(p(\mathbf{u}_i),\mathbf{a}) - p(\mathbf{u}_i^0))^2), \quad F'(p(\mathbf{u}_i),\mathbf{a}) \geq 0. \tag{14}$$

In the problem of color correction for Fresnel lenses, we use a color checker scale for identification (Fig. 3(e) and Fig. 3(f)). The original colors of the scale are used as distortion-free SSE, $\mathbf{u}_i^0$. The same scale captured using a Fresnel lens is used for getting distorted SSEs, $\mathbf{u}_i$.

After identification of the color correction transform parameters we apply this transform to the image as a final step of the technique based on model (3) - (4).

## VI. Results

The results of the correction are shown in Fig 3. The original picture was captured using a digital camera with a single Fresnel lens. The lens was made as three-step approximation of the Fresnel lens. Firstly we removed the blur from green channel using deconvolution, after that we used edge analysis for red and blue channels. Color correction transform was identified using color checker table (Fig 3(e)), and finally color correction was applied to the image.

The proposed technique implements three steps for image restoration. First of all, we use deconvolution (2) to increase the quality of the best color channel. Then using edge analysis, we improve another two color channels. These two steps allow restoration of the edge information. Finally, we use the color correction technique to restore the plain-colored regions. As it is shown in Fig. 3, the proposed correction technique provides restoration of both colors and edges information from distorted images, captured by simple difractional lenses.

## VII. Conclusion

The presented paper shows that simple difractional lenses could be used for imaging. Strong aberrations that are inherent to this kind of optics could be restored by image processing techniques.

There are two approaches used for correction of images captured through simple lenses: deconvolution and contour analysis. We have obtained good results for these approaches combination for the simple Fresnel lenses case. The color correction is also useful in this case.

The obtained image restoration is significant but not complete. Two main directions for further research are 1) increasing the quality of deconvolution, taking into account the estimation of space-varying PSF and 2) combining an edge analysis and color correction in one filter. It allows increase quality of chromatic blur removing and also suppresses processing noise.

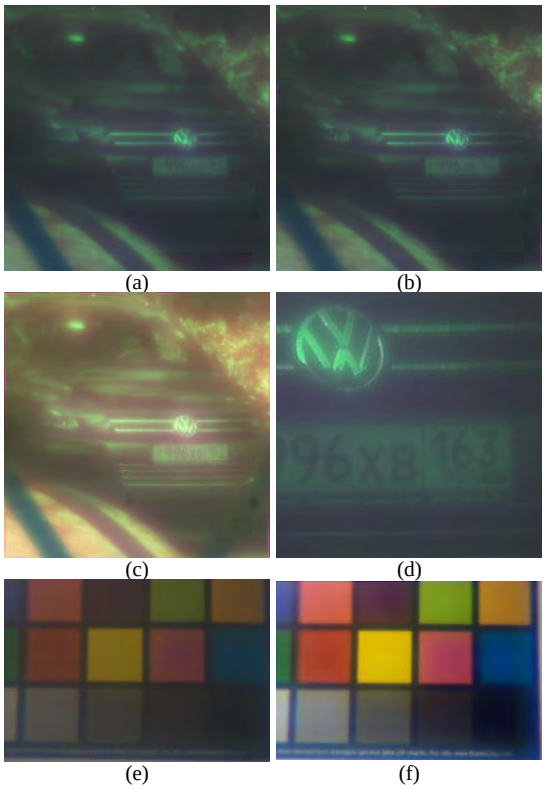



Fig. 3. Example of chromatic aberration correction: (a) original image (b) image after deconvolution (c) image after color correction (d) zoomed image after deconvolution (e) color checker image for correction identification (f) color checker image after correction

## References

[1] F. Heide, M. Rouf, M.B. Hullin, B. Labitzke, W. Heidrich, A. Kolb, High-Quality Computational Imaging Through Simple Lenses, ACM Transactions on Graphics, V. 32 I. 5, No. 149, 2013.

[2] Y. Shih, B. Guenter, N. Joshi, Image Enhancement Using Calibrated Lens Simulations, ECCV 2012, Lecture Notes in Computer Science, V. 7575, 2012, pp. 42-56.

[3] A. Chambolle, T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. 40, 2011, 120–145.

[4] A. Nikonorov, S. Bibikov, P. Yakimov, V. Fursov, Spectrum shape elements model to correct color and hyperspectral images, 8th IEEE IAPR Workshop on Pattern Recognition in Remote Sensing, 2014.

[5] S.-W. Chung, B.-K. Kim, W.-J. Song, Removing chromatic aberration by digital image processing, Optical Engineering 49(6), 067002, 2010.

[6] V.A. Soifer, Computer Design of Diffractive Optics, Woodhead Publishing, 2012, 896 p., ISBN: 9781845696351.

[7] S. B. Kang, Automatic Removal of Chromatic Aberration from a Single Image, CVPR 2007, pp.1-8.

[8] A. Davis, F. Kuhnlenz, Optical Design using Fresnel Lenses - Basic Principles and some Practical Examples, Optik & Photonik, Volume 2, Issue 4, pages 52–55, 2007.

# Design and Implementation of the Alida Framework to Ease the Development of Image Analysis Algorithms

Stefan Posch and Birgit Möller
Institute of Computer Science
Martin Luther University Halle-Wittenberg
{stefan.posch, birgit.moeller}@informatik.uni-halle.de

***Abstract*—**Solving image analysis problems is not restricted to the pure delineation of algorithms suitable to tackle the task at hand. Rather these also need to be made available to the users promptly and equipped with handy user interfaces to foster progress in the intended field of application. `Alida` is a software framework to advance the *integrated* development of algorithms and appropriate user interfaces. It automatically generates user interfaces for implemented algorithms, offers an automatic documentation of analysis procedures, and ships with a graphical editor for designing complex workflows. `Alida`'s Java implementation is licensed under GPL 3.0 and publicly available at http://www.informatik.uni-halle.de/alida.

## I. Introduction

The ubiquitous availability of image data of various types demands more than ever for automatic analysis and interpretation. Such image analysis processes can be understood as a set of individual processing steps applied to different data. Each step is typically realized as a functional unit like a method or function and will be called *operator* in the following. Operators may use other operators and can be combined programmatically into pipelines which need not necessarily to be sequential.

Besides this reusability another important aspect is to enable users to invoke operators directly. This may be accomplished via a command line interface, using scripts, or interactively with a graphical user interface (GUI). The later may be used to invoke single operators or be extended to interactively compose pipelines in an easy to use and intuitve way.

The results of analysis processes are often stored for later use, e.g., to be included in publications or to compare different parameter settings. When archiving it is of great interest to augment these results with information about the pipeline executed to achieve the results. Such a documentation subsumes not only which processing steps were applied in which sequence to which data, but also parameter settings and software versions of all steps [1].

Both aspects, the generation of suitable user interfaces (UIs) and the support of detailed documentation is typically associated with a substantial overhead to the programmer when developing image analysis operators. Here we propose `Alida`, an *Advanced Library for Integrated Development of Data Analysis Applications*, as a flexible framework facilitating both tasks. Implementing operators in `Alida` allows to focus on the functional issues where a graphical (cf. Fig. 1) and a command line UI are automatically available with no additional effort. Likewise a detailed documentation is automatically generated and stored with the results upon execution. In addition, `Alida` comprises a graphical editor supporting the easy development of pipelines based on operators implemented in the framework (Fig. 2). `Alida` is publicly available under the GPL license as a Java library. It can easily be integrated into existing image analysis systems like ImageJ [2], and also serves as basis for our Microscope Image Analysis Toolbox `MiToBo`[1]. Sample applications are [3], [4], [5], [6], [7].

## II. Related Work

The steadily increasing number of image processing toolboxes and algorithmic collections is a decisive indicator of the emerging demand for efficient and flexible image analysis solutions. By gathering implementations of state-of-the-art algorithms in toolboxes and libraries they become readily available to the public and support straightforward solutions to manifold problems. Prominent examples for such toolboxes are OpenCV [8], ITK [9], JAI [10] or ImageJ [2] on the open source, and Matlab [11] or Halcon [12] on the commercial side, respectively. While these tools mostly focus on implementations of stand-alone algorithms, the increasing importance of designing workflows to tackle more complex problems is mirrored by projects like KNIME [13] or Kepler [14]. They seek to offer not only independent functional units, but also provide additional support for easy pipeline design and integrated data analysis in terms of graphical front ends for workflow development.

Under the hood most toolboxes and libraries rely on a fundamental set of data types on which the various algorithms operate and which alleviate data exchange and I/O. The implementation of functionality, however, varies significantly not only between different tools, but also *within* certain libraries. Only partially attempts have been undertaken to unify the usage of functions and operators on the programming level. E.g., JAI and KNIME define unified interfaces for their operators, and also the new ImageJ 2.0 release [15] aims at unifying command implementation and usage as well.

And the situation is even worse on the user side. In many fields of application, like in the life sciences, short release times are inevitable for software not to loose its relevance because of changing requirements in the mean-time. This

[1] MiToBo, http://www.informatik.uni-halle.de/mitobo

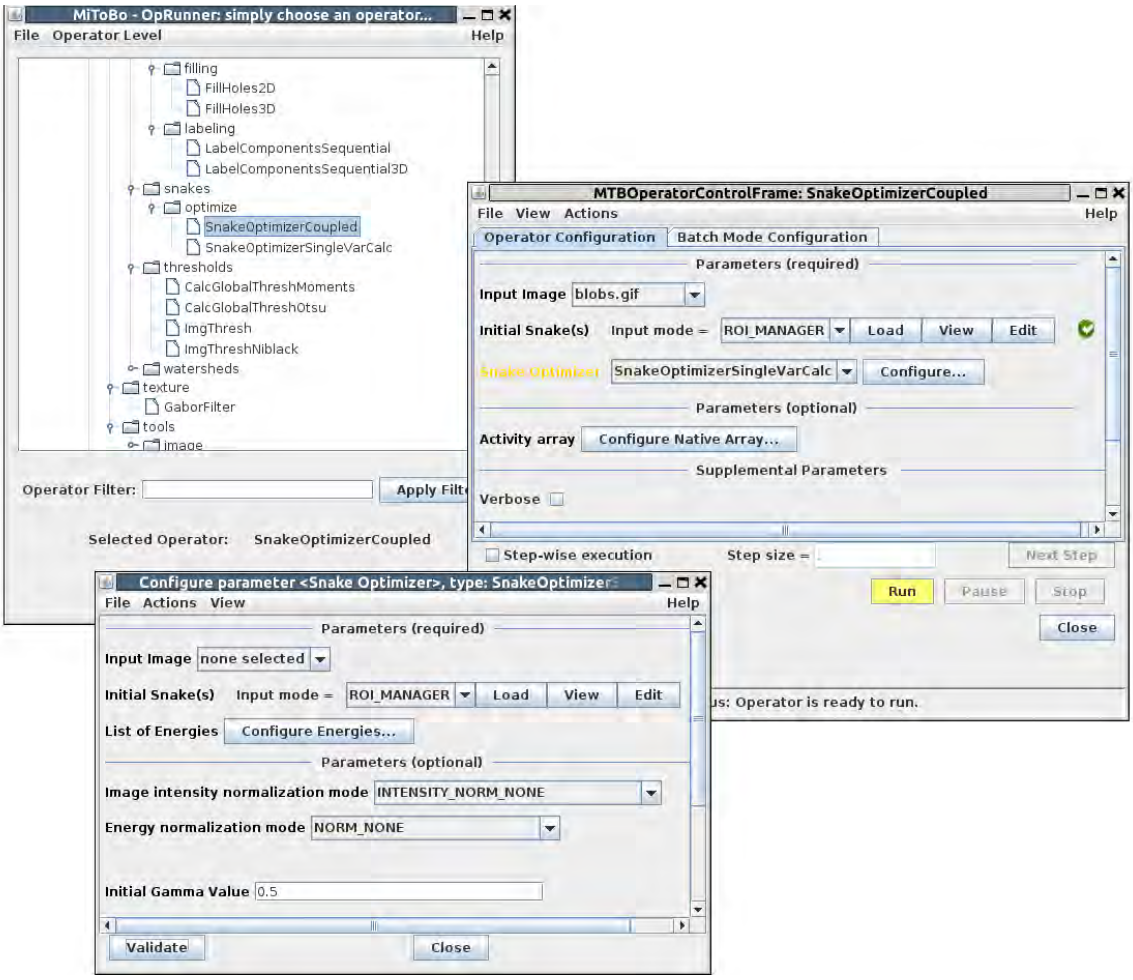



Fig. 1. The main window of Alida's operator runner (top left) and two automatically generated control and configuration windows for operators.

naturally conflicts with the additional amount of time necessary to design handy user interfaces. As a consequence user-friendly interfaces are too often neglected throughout the development process.

Some recent projects seek to change this situation aiming at the inherent integration of user interface design into the algorithmic development process. One of the pioneers in this direction is the ImageJ 2.0 project, which offers built-in functionality for automatic generation of GUIs. As pointed out , however, GUIs are just one building block on the way towards integrated image analysis frameworks. Equally important are command line interfaces and built-in documentation capabilities. While some tools support at least a rudimentary documentation of analysis procedures (e.g., CellProfiler [16], Icy [17]), software subsuming all of these features is still rare.

## III. Framework Requirements

As described in the introduction, the overall goal of Alida aims to provide a framework for the development of image analysis algorithms which facilitates straightforward algorithm implementation featuring automatic UI generation and the inherent documentation of analysis processes at run-time.

The documentation of an analysis procedure is supposed to contain all information necessary to recover the results from the same input data at a later point in time. Given operators as the fundamental building blocks for data analysis procedures this information is two-fold. One the one hand it comprises the analysis pipeline with all operators involved and the data flow between these functional units which defines the *processing graph*. On the other hand all information required to re-run each single operator needs to be stored, i.e. the *operator configuration*. Besides the input data provided by the data flow between operators, this includes all control settings, e.g., thresholds, and also metadata like software versions.

For invoking single operators, UIs are required providing functionality to specify input data and control settings, subsumed as the operator's *input parameters* in the following. Subsequently, the operator is to be executed by the user and the resulting data of the analysis are to be presented in a user-friendly manner. Depending on the situation, the execution of single operators may be performed either interactively via a GUI or console-based, e.g., using scripts for parameter tuning. In contrast the design of a suitable workflow for a specific problem is most conveniently carried out graphically. The configuration of involved operators beyond the data flow shares the same requirements as for execution using a GUI.

From these objectives design requirements for the software architecture can be derived to be fulfilled by a framework like

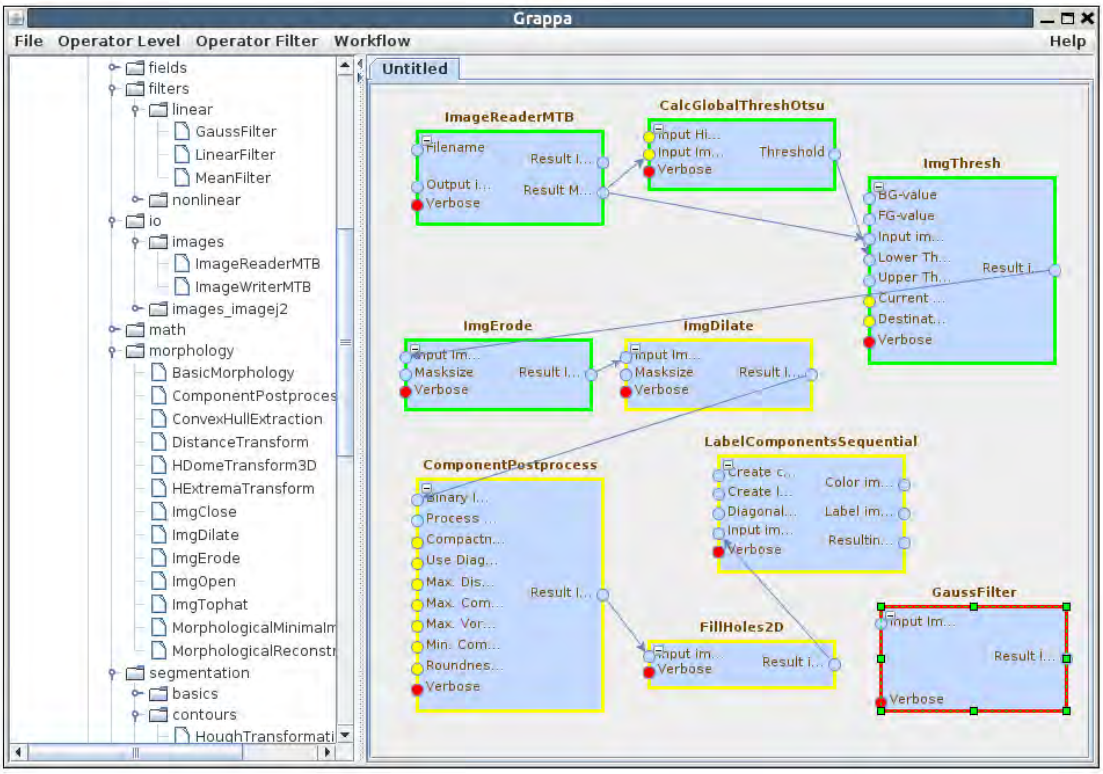



Fig. 2. Screen shot of Alida's graphical programming editor Grappa.

Alida. As all operator calls performed during an analysis are to be documented, it is inevitable that each operator invocation is registered by the framework. Most easily this can be guaranteed by a unified invocation procedure for operators. Furthermore, also the current parameter settings need to be documented. To this end operators have to provide unified means to access all input parameters at run-time which includes queries for their names and types as well as for their values.

This unified invocation procedure also fosters automatic generation of UIs. Execution via an automatically generated UI requires generic means to set input parameters of an operator prior to execution. Values are requested from the user either graphically, i.e. asking the user to input data via automatically generated graphical components, or by parsing command line inputs. Finally, UIs need to present result data to the user in an appropriate manner. Regarding the ability to set and retrieve parameter values, it is of particular importance that the set of supported parameter data types is not statically defined by the framework, but may be dynamically extended.

## IV. Object-Oriented Design

In this section Alida's approach to satisfy the requirements identified in the last section within an object-oriented paradigm is described and details of the Java implementation are provided in the subsequent section.

### A. Operators and Parameters

In an object-oriented approach a natural choice is to implement each operator in its own class. Alida adopts this strategy and requires each operator to extend the abstract class ALDOperator. This base class implements the method runOp() which is the only admissible way to invoke an operator. This unified invocation procedure allows to track all operator invocations and to retrieve the processing graph of the analysis pipeline. In addition, the class defines an abstract method operate() which must be implemented by each operator and contains its data processing functionality. This method is implicitly invoked by the generic runOp() method subsequent to initial validation and registration procedures.

All data to be processed by an operator, controlling its manipulation, or to be returned as result are consistently denoted as *parameters*. Alida propagates mechanisms for the programmer of a new operator to easily define all parameters including their properties in a way which allows Alida to supply methods to query an operator object at run-time for all its parameters and their properties. Likewise generic getter and setter methods are supplied automatically for all parameters of an operator. As stated in Sec. III this is a prerequisite for automatic documentation and provides all information to generate user interfaces. For interface generation a further requirement is the possibility to instantiate operator objects via a default constructor, and to set all parameters subsequently by the user.

Each parameter of an operator is characterized by its name and data type as well as further properties. The role of a parameter is specified via its *direction*, which may be IN, OUT, or INOUT. A typical example for an operator in image processing is a filter applied to an input image (direction IN) where the filtered image is returned in a newly allocated data

structure as a parameter with direction OUT. If the filter acts destructively in place, this is described by a single parameter of direction INOUT. A parameter controlling the filter operation, e.g., a bandwidth, is provided as an IN parameter. Parameters of direction IN and INOUT may be either *required* or *optional*. In addition to parameters affecting or representing the result of data manipulation, `Alida` supports *supplemental* parameters which must not influence the processing results. Examples include flags to control the output of debug information or intermediate results to be returned from the operator.

To make use of an operator's functionality on the programmatic level an instance of the class needs to be created and its parameters have to be set. Processing is invoked using the generic `runOp()` method. Upon termination the results may be retrieved using the generic getter methods for output parameters.

### B. Automatic User Interface Generation

To automatically generate UIs `Alida` needs to know how to query values for parameters from the user, how to instantiate parameter objects from these values, and how to present parameter values to the user, e.g., graphically or via console. This knowledge is specific for each data type, and the set of potential parameter data types is unknown in advance as new operators implemented may require additional data types. Thus, a mechanism is required that allows for extensibility and links the I/O knowledge to specific data types rather than incorporating it into the core of `Alida`. Hence we propose the idea of data I/O *providers* which are functional units performing input and output operations for specific data types. Each of these providers is naturally implemented in its own class, and all providers have to implement a common interface defining methods for reading and writing data types. The meaning of 'reading' and 'writing' depends on the specific type of the UI. Therefore different sets of such provider classes are required, each dedicated to a specific type of UI. For the command line a provider typically needs to parse textual user input into appropriate objects. For output a transformation of parameters into a textual representation displayed in the console or printed to file is required. For GUIs, input of parameters requires to display a graphical element to the user suitable for specifying parameter values, and upon request to instantiate a data object from the specified values. For output a data object has to be visualized graphically.

The set of I/O providers is not static over time in the sense that new providers can easily be added. So-called I/O *managers* keep track of the set of providers currently available. Typically for each type of UI a dedicated manager is responsible. This implies a registration mechanism for providers upon initialization of the framework. Once this information has been gathered, `Alida` can query the manager responsible for the given type of UI to supply a provider able to read or write a specific data object.

### C. Workflows

Image analysis problems usually require a combination of several operators to be applied to data rather than only a single operator. Thus, `Alida` supports to represent and manipulate complete pipelines for data processing comprising several operators and the data flow in between by the class `ALDWorkflow`. It needs to provide entry and exit points into and out of the whole pipeline. As these points have essentially the same role as parameters for operators, the class `ALDWorkflow` naturally extends the base class `ALDOperator`.

`ALDWorkflow` basically models a processing pipeline as a graph, i.e. operators are represented by nodes, and the `OUT` and `IN` parameters of different operators are connected by edges to describe the flow of data.

When connecting parameters of different nodes the validity needs to be verified. For example, an input parameter may have at most one incoming edge, and the data types of parameters connected by an edge need to be compatible. In general the `operate()` method of a workflow object invokes all selected operators in topological order and forwards output data between operators according to the data flow.

### D. Automatic documentation

Since each operator execution is realized by a method invocation in `Alida`, the processing pipeline can be understood as a subgraph of the dynamic call graph of the analysis process. This processing graph may also be interpreted as a hierarchical graph where each invocation of an operator is represented by a node (see Fig. 3). If an operator is invoked by another operator, its *parent operator*, the corresponding node is modelled as a nested child of the parent operator's node. To represent the data flow, each node features input and output ports for each of the operator's input and output parameters. The port of each input parameter is linked to the origin of its value upon invocation. If this data object was already manipulated by a previous operator invocation its origin may be an input port of the parent operator or an output port of a sibling operator. Otherwise the origin is a node of type *data port* which represents the instantiation of a data object resulting, e.g., from reading data from file. At any point in time the processing graph for a specific data object may be extracted which in general forms a subgraph of the overall graph.

## V. Alida's Implementation and Tools

### A. Operators and Parameter Handling

As described in the last section each operator in `Alida` is implemented as a class extending the abstract base class `ALDOperator`. For each parameter a member variable is defined. Java's annotation mechanism is used to declare these members as parameters using `Alida`'s @Parameter annotation which is currently an extended version of the corresponding annotation of ImageJ 2.0. The elements of this annotation are used to define the properties of the parameter, e.g., the direction and if the parameter is required or supplemental. Java's reflection mechanism is used to implement methods to query an operator for its parameters including their data types and properties, as well as generic getter and setter methods for all parameters.

Java's annotations are also employed to annotate the operators themselves. This allows to easily collect all operators available during framework initialization and facilitates, e.g., to abbreviate operator names in the command line user interface, or to list all operators in the GUI to choose from.

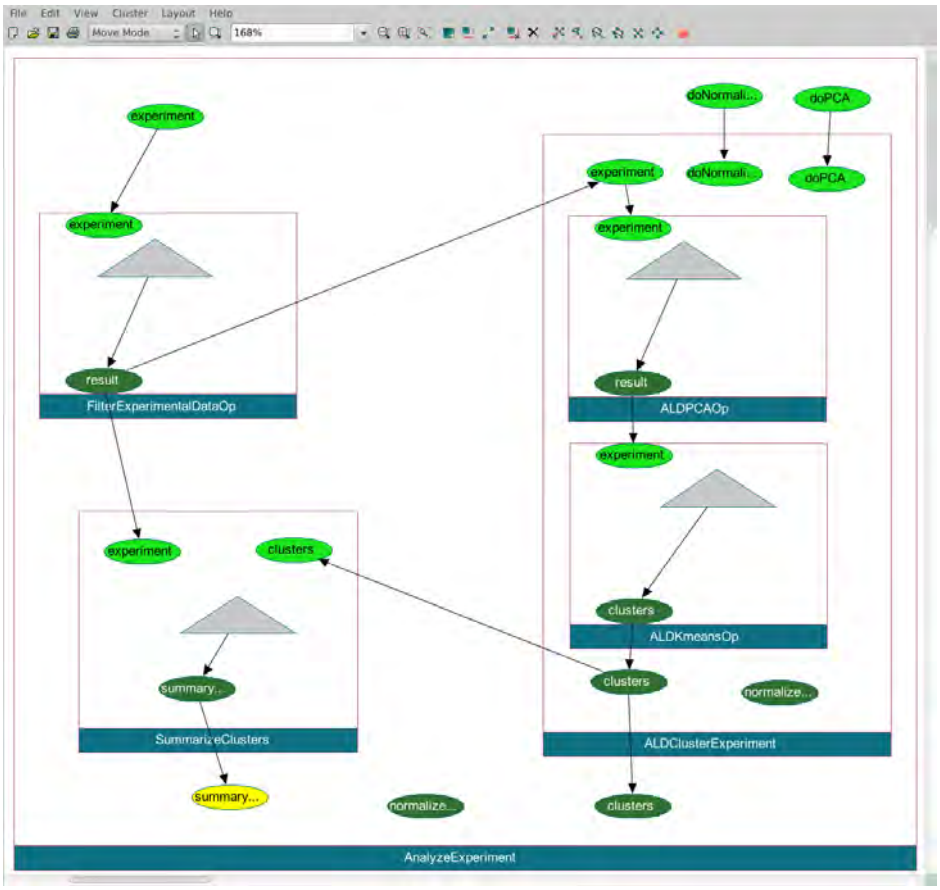



Fig. 3. Example processing graph representing the history of operations for producing the data object shown as yellow ellipse. Each operator invocation is represented by a blue rectangle. Light and dark green ellipses are input and output ports of operators, gray triangles depict data ports representing newly generated data

The base class `ALDOperator` implements the method `runOp()` which is the only admissible way to invoke an operator. The purpose of the method is threefold: (i) the parameters are validated, (ii) all information for automatic documentation are generated and stored (see Sec. V-D), and (iii) the operator functionality is finally applied calling the `operate()` method overwritten in the extending classes. The validation includes a generic part enforcing all required input parameters to have non-null values, and a custom validation as defined by the operator overwriting the method `validateCustom()`. The later allows to enforce, e.g., admissible intervals for numeric values.

### B. User Interfaces

The key idea for automatically generating user interfaces within `Alida` is the flexible I/O manager and provider concept introduced above. In this section its implementation is described.

*1) Data I/O managers and providers:* Alida's data I/O concept is currently implemented for GUIs based on the Java Swing framework, and for the command line. Consequently, `Alida` ships with two I/O managers which provide convenience methods for data type input and output, encapsulating the actual I/O provider objects. The command line I/O manager basically implements two methods:

```
Object readData(Field, Class, String);
String writeData(Object, String);
```

The first allows for reading parameter objects of the given class from a source string, while the second transforms a given object into a string. The field argument in the first method allows to hand over additional information about the parameter, and the string argument in the second method allows to pass additional formatting information to the method, e.g., to achieve varying behavior when writing to a file compared to console. Both methods internally instantiate a provider object for the requested class, potentially falling back to default providers, e.g., for enumerations. Subsequently the corresponding methods of the provider are called for reading and writing data, passing through all arguments, and returning values to the caller.

The functionality of the Swing manager is slightly different as data input requires user interaction. It implements three methods,

```
ALDSwingComponent createGUIElement(
        Field,Class,ALDParameterDescriptor);
Object readData(Field,Class,
     ALDSwingComponent);
JComponent writeData(Object,
     ALDParameterDescriptor);
```

The first method is supposed to generate graphical components for the requested object class by which values can be specified. `ALDSwingComponent` is an `Alida` base class basically wrapping `JComponent`. The parameter descriptor argument allows for passing additional information to the manager, e.g., for enhancing the GUI components with information about the parameter not necessary for reading values, but helpful for the user. Once graphical components have been generated and are presented to the user (see below), values can be queried from them via `readData(...)` which returns an object of the requested class from the given component. Finally, the `writeData(...)` method transforms an object into a graphical component, again optionally using information from the parameter descriptor.

Upon initialization both managers gather the set of available providers and supported data types exploiting the Java annotation mechanism. Each provider is to be annotated with the `@ALDDataIOProvider` annotation and needs to implement either the command line interface `ALDDataIOProvider-Cmdline` or the Swing interface `ALDDataIOProvider-Swing`.

Currently, `Alida` features general purpose providers for all primitive data types, enumeration types, arrays, collections, and so-called *parameterized classes*. An arbitrary class may be declared as parameterized class, and any subset of its member variables declared as *class parameters*, both via annotations. This is sufficient for `Alida`'s general purpose provider to handle this class as an operator parameter. Likewise operators may act as parameters of other operators.

*2) Graphical Operator Runner:* Provided that suitable data I/O providers are available, `Alida` can automatically generate a GUI for each operator. This functionality is offered to the user via its graphical operator runner (cf. Fig. 1) where all available operators are listed and can be selected for execution. Upon selection the runner instantiates an operator object of the corresponding class and queries the Swing I/O manager for graphical components for all of its parameters and arranges them in a control window. The window contains buttons for executing the operator, and upon termination the runner collects the values of all output parameters from the operator object via its generic getters, requests graphical components for all non-null values from the Swing I/O manager and arranges the components in a result window (Fig. 4).

*3) Command Line Operator Runner:* As a second generic user interface, `Alida` features a command line operator runner (CLR) to invoke operators via the console or scripts. All input parameters are supplied as arguments by 'name=value' pairs. To ease the handling of class inheritance and operators as parameters in a generic fashion, the CLR features a flexible parser for argument preprocessing. It allows for parsing CLR calls like the one shown in Fig. 5. The `MiToBo` operator `SnakeOptimizerCoupled` for multiple snake segmentation not only takes initial snakes and an image as input, but also an operator instance of `SnakeOptimizerSingleVarCalc` dealing with a single snake (cf. Fig. 1, bottom). The syntax of the individual parameter value strings is defined by the specific I/O providers they are finally passed to and which, by convention, also allow values to be read from file. Analogously output parameters

```
java de.unihalle.informatik.Alida.\
 tools.ALDOpRunner SnakeOptimizerCoupled \
 initialSnakes=RoiSet.xml inImg=cell.tif \
 snakeOptimizer=\
  '$SnakeOptimizerSingleVarCalc:\
  {energySet={energies=\
   [$MTBSnakeEnergyCD_CVRegionFit:\
    {lambda_in=1.0,lambda_out=5.0}],\
   weights=[1.0]}}' outSnakes=snakesOut.xml
```

Fig. 5. Example call of an operator from command line. The operator `SnakeOptimizerCoupled` called here among others takes an operator of type `SnakeOptimizerSingleVarCalc` as input parameter.

are specified as 'name=value' pairs allowing to, e.g., redirect output into files, as an alternative to formatting the values onto standard out.

### C. Workflows

As outlined in Sec. IV-C, workflows are represented in `Alida` by the class `ALDWorkflow` extending the operator base class. The implementation is designed to represent the model of a workflow in a model-view-controller software architecture. For all relevant modifications of the workflow, e.g., adding or removing nodes and edges, or for the execution of operators, appropriate methods are provided. Besides the execution of the complete workflow via its `operate()` method also partial execution is supported. Execution may be threaded in order to, e.g., allow external control of the operator execution by the user. Invocation of all of these methods triggers corresponding events passed to all registered listeners, e.g., the graphical editor (cf. Sec. V-E).

The nodes of the workflow have associated states reflecting the state of the operator instance represented by the node. States include RUNNABLE which indicates that all required input parameters are set or may be retrieved from results of other runnable or ready operators. Likewise, an operator may be in the state RUNNING, i.e. is currently executed, or READY. These states are consistently updated by `ALDWorkflow` considering the dependencies between nodes, their configurations and execution. Finally, workflows may be saved to and loaded from file including the operators' configurations defined by the values of their parameters.

### D. Automatic Process Documentation

To prepare for later documentation of processing results, the processing graph is continuously constructed and extended as described in Sec. IV-D. To this end, `runOp()` creates an instance of the class `ALDOpNode` representing the call to the operator in the processing graph. If the operator is invoked in a nested fashion from a parent operator the `ALDOpNode` is registered to the corresponding `ALDOpNode` of the enclosing operator. For each data object handled as input or output parameter of an operator invocation the current origin is stored in a global weak hash map. Upon invocation via the `runOp()` method the origin of all input parameters is retrieved from this map and the corresponding link set in the `ALDOpNode` object. Subsequently the origin is updated to the input port of the current operator invocation and restored upon return from `operate()`.

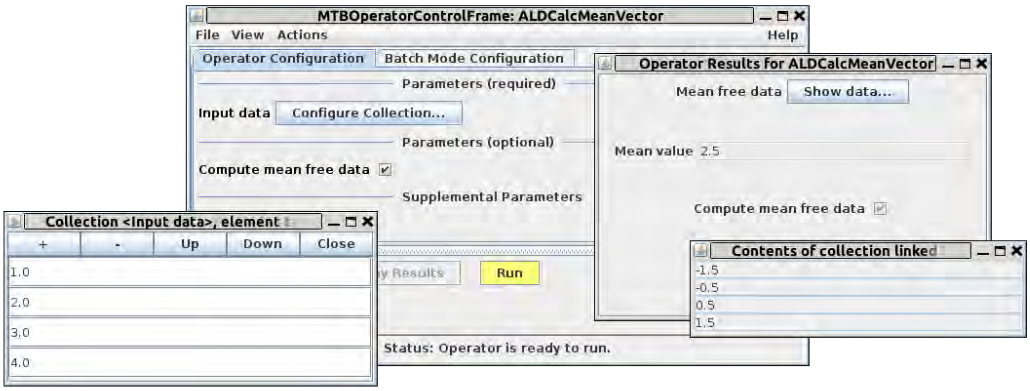



Fig. 4. After termination `Alida` summarizes all operator results in a Here the operator `ALDCalcMeanVector` calculated the mean of an input array containing doubles (left) and generated an array with zero mean.

In this way the global processing graph is incrementally constructed during data processing. Supplementing the construction of the graph, upon invocation the method `runOp()` retrieves the current values of all input parameters and stores these in the `ALDOpNode` object created. As the final necessary information the current software version is acquired. This is accomplished using the abstract class `ALDVersionProvider` where a concrete implementation may be passed to the operator mechanism at run-time.

### *E. Grappa*

Grappa is a tool for designing and editing workflows based on graph edit operations. As introduced in Sec. IV-C, a workflow is basically given by a set of nodes with associated `Alida` operators which are linked by edges encoding the flow of data and control. In Grappa all `Alida` operators are available as workflow nodes, and each node owns a set of ports linked to the operator's parameters, which can be connected by edges. Via intuitive mouse actions the nodes can be placed and moved within a workbench, and ports can be linked by edges. Grappa performs workflow consistency checks, i.e. prohibiting cyclic workflow graphs. Likewise automatic type checking is performed which also incorporates an extensible mechanism for type conversion. As an example, a numerical data type may be converted to a data type of less precision on user request. The complete workflow, subgraphs, or single nodes can be executed adopting `Alida`'s generic execution mechanisms. For node configuration a graphical configuration window for each node is available utilizing `Alida`'s functionality for automatic GUI generation.

## VI. Conclusion

The framework `Alida` presented in this article seeks to narrow the gap between user and developer requirements during algorithm development. By providing built-in functionality like the automatic generation of user interfaces or the automatic documentation of analysis procedures `Alida` enables developers to focus on algorithms rather than infrastructure, and users to efficiently solve their problems without requiring too much knowledge about the internal matter of the software. `Alida`'s concepts can easily be generalized to different areas of data analysis, and they have already proven their suitability for every-day's scientific algorithm design as basis for the image analysis toolbox `MiToBo`. Both, `Alida` and `MiToBo`, have been released under GPL and are publicly available from the internet.

## References

[1] Linkert, M., Rueden, C., et al.: Metadata matters: access to image data in the real world. The Journal of Cell Biology **189**(5777-782) (2010)

[2] Schneider, C., Rasband, W., Eliceiri, K.: NIH Image to ImageJ: 25 years of image analysis. Nature Methods **9** (2012) 671–675

[3] Möller, B., Posch, S.: Comparing active contours for the segmentation of biomedical images. In: Proc. of IEEE International Symposium on Biomedical Imaging (ISBI), Barcelona, Spain (2012) 736–739

[4] Glaß, M., Möller, B., Zirkel, A., Wächter, K., Hüttelmaier, S., Posch, S.: Cell migration analysis: Segmenting scratch assay images with level sets and support vector machines. Pattern Recognition **45**(9) (2012) 3154–3165

[5] Glaß, M., Möller, B., Posch, S.: Scratch assay analysis in ImageJ. In: Proc. of ImageJ User & Developer Conference, Mondorf-les-Bains, Luxembourg (2012) 211–214

[6] Greß, O., Möller, B., Stöhr, N., Hüttelmaier, S., Posch, S.: Scale-adaptive wavelet-based particle detection in microscopy images. In Meinzer, H.P., Deserno, T.M., Handels, H., Tolxdorff, T., eds.: Bildverarbeitung für die Medizin, Berlin, Springer (2010) 266–270

[7] Möller, B., Stöhr, N., Hüttelmaier, S., Posch, S.: Cascaded segmentation of grained cell tissue with active contour models. In: Proceedings International Conference on Pattern Recognition (ICPR). (2010) 1481–1484

[8] Bradski, G.: The OpenCV Library (2000) Dr. Dobb's Journal of Software Tools, Library website: http://opencv.org/.

[9] Ibanez, L., Schroeder, W., Ng, L., Cates, J.: The ITK Software Guide, http://www.itk.org/ItkSoftwareGuide.pdf. Second edn. (2005)

[10] Sun Microsystems Palo Alto, CA 94303, USA: Programming in Java Advanced Imaging. (1999) Rel. 1.0.1.

[11] MATLAB: The MathWorks, Inc., Natick, Massachusetts, United States (accessed September 2014) http://www.mathworks.com.

[12] MVTec Software GmbH: Halcon - HDevelop User's Guide (2012) Version 11.0.1, http://www.mvtec.com/halcon/.

[13] Berthold, M.R., et al.: KNIME - the Konstanz Information Miner: version 2.0 and beyond. SIGKDD Explor. Newsl. **11**(1) (2009) 26–31

[14] Ludäscher, B., Altintas, I., Berkley, C., Higgins, D., Jaeger, E., Jones, M., Lee, E.A., Tao, J., Zhao, Y.: Scientific workflow management and the kepler system. Concurrency and Computation: Practice and Experience **18**(10) (2006) 1039–1065

[15] ImageJDev project: ImageJDev project (accessed September 2014) http://developer.imagej.net/.

[16] CellProfiler: cell image analysis software (accessed September 2014) http://www.cellprofiler.org/.

[17] Icy: An open community platform for bioimage informatics (accessed October 2014) http://icy.bioimageanalysis.org/.

# Development of the Logic Programming Approach to the Intelligent Monitoring of Anomalous Human Behaviour (Extended Abstract)

Alexei A. Morozov and Alexander F. Polupanov
Kotel'nikov Institute of Radio Engineering and Electronics of RAS, Mokhovaya 11, Moscow, Russia
Moscow State University of Psychology & Education, Sretenka 29, Moscow, Russia
Emails: morozov@cplire.ru, sashap55@mail.ru

***Abstract*—A research software platform is developed that is based on the Actor Prolog concurrent object-oriented logic language and a state-of-the-art Prolog-to-Java translator for experimenting with the intelligent visual surveillance. We demonstrate an example of the application of the method to the monitoring of anomalous human behaviour that is based on the logical description of complex human behaviour patterns and special kinds of blob motion statistics. The logic language is used for the analysis of graphs of tracks of moving blobs; the graphs are supplied by low-level analysis algorithms implemented in a special built-in class of Actor Prolog. The blob motion statistics is collected by the low-level analysis procedures that are of the need for the discrimination of running people, people riding bicycles, and cars in a video scene. The first-order logic language is used for implementing the fuzzy logical inference based on the blob motion statistics.**

Human activity recognition is a rapidly growing research area with important application domains including security and anti-terrorist issues [1]. Recently logic programming was recognized as a promising approach for dynamic visual scenes analysis (see surveys of logic-based recognition systems in [2], [3]). The idea of the logic programming approach is in usage of logical rules for description and analysis of people activities. Knowledge about object co-ordinates and properties, scene geometry, and human body constraints is encoded in the form of certain rules in a logic programming language and is applied to the output of low-level object / feature detectors.

The distinctive feature of our approach to the visual surveillance logic programming is in application of general-purpose concurrent object-oriented logic programming features, but not in the development of a new logical formalism. We use the Actor Prolog object-oriented logic language [4], [5], [6], [7], [8], [3] for implementation of concurrent stages of video processing. A state-of-the-art Prolog-to-Java translator is used for efficient implementation of logical inference on video scenes. Special built-in classes of the Actor Prolog language were developed and implemented for the low-level video storage and processing.

Let us consider an example of logical inference on video. The input of a logic program written in Actor Prolog is a standard sample provided by the BEHAVE team [9]. The program will use no additional information about the content of the video scene, but only co-ordinates of reference points in the ground plane that are necessary for estimation of physical distances in the scene.

Fig. 1. An example of BEHAVE video with a case of a street offence: one group attacks another.

The video (see Fig. 1) demonstrates a case of a street offence—a probable conflict between two groups of persons. The input data for the logic program will be supplied by automatic low-level algorithms that trace objects in a video scene and estimate average speed in different segments of the trajectories [10], [3]. This low-level processing is implemented in Java (but not in Prolog) and includes extraction of foreground blobs, tracking of the blobs over time, detection of interactions between the blobs, creation of connected graphs of linked tracks of the blobs, and estimation of average speed of the blobs in separate segments of the tracks (see Fig. 6). The input data include a special set of blob motion statistics to discriminate running pedestrians, bicycles, and cars during the logical inference.

We have created a set of blob motion metrics based on the windowed coefficient of determination of the temporal changes of the length of the contour of the blob. They are supposed to be reliable when images of moving objects are noised and / or fuzzy, that is necessary for real applications.

The coefficient of determination $R^2$ indicates the proportionate amount of variation in the given response variable $Y$ explained by the independent variables $X$ in the linear regression model. Thus, the $R^2$ metrics is supposed to be useful for discrimination of vehicles and running pedestrians, when the $X$ variable is the time and the $Y$ variable is the

area or the length of the contour of the moving blob. In the general case, vehicles will be characterised by bigger values of the $R^2$ metrics than running persons, because the contour of the running person changes permanently in the course of his motion when he waves his arms and moves his legs.

The standard definition of the (unadjusted) coefficient of determination is the following one:

$$R^2 = 1 - SSE/SST \tag{1}$$

where $SSE$ is the sum of squared error and $SST$ is the sum of squared total.

We use a windowed modification of the $R^2$ metrics, that is, the trajectory of a moving blob is characterised by a set of instantaneous values of the $R^2$ metrics computed in each point of the trajectory. Suppose $t_B$ is the beginning time point of the trajectory and $t_E$ is the end time point. Thus, the windowed $R^2$ metrics is a set:

$$wR^2 = \left\{R^2_{t,w}\right\} \tag{2}$$

where $t$ is the time ($t \in \{t_B + w/2 \dots t_E - w/2\}$) and $w$ is the width of the widow (the neighbourhood of $t$) to be used for computation of the $R^2$ value.

In the scope of this paper, we will use two statistical metrics that characterise the motion of the blob, namely, the mean of the $wR^2$ distribution:

$$mean(wR^2) = \frac{\sum_{i=1}^{n} wR_i^2}{n} \tag{3}$$

and the bias-corrected skewness of the $wR^2$ distribution:

$$skewness(wR^2) = \frac{\sqrt{n\,(n-1)}}{n-2} s(wR^2) \tag{4}$$

$$s(wR^2) = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(wR_i^2 - \overline{wR^2}\right)^3}{\left(\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(wR_i^2 - \overline{wR^2}\right)^2}\right)^3} \tag{5}$$

where $n$ is the number of elements in the $wR^2$ set.

Skewness is a measure of the asymmetry of the data around the sample mean. If skewness is negative, the data are spread out more to the left of the mean than to the right. If skewness is positive, the data are spread out more to the right. Thus, in the framework of the moving blobs discrimination problem, one can expect that the vehicles will be characterised by bigger values of the $mean(wR^2)$ metrics and smaller values of the $skewness(wR^2)$ metrics than running persons.

The properties of the $mean(wR^2)$ and the $skewness(wR^2)$ metrics can be illustrated by the example of the BEHAVE data set [9]. For that, we have created the tracks of moving blobs in this data set by the blob extraction methods implemented in Actor Prolog [10], [3]. Then we have selected 193 samples of tracks including 85 alone walkers, 58 groups of walkers, 20 alone running persons, 15 groups of running persons, 6 bicycles, and 9 cars. Each blob has the following attributes: the trajectory of the central point of the rectangle blob; the area and the length of the contour of the blob in each point of time; the lower boundary estimation of the speed of the blob [10] in each point of time; and others. The values of the $mean(wR^2)$ and the $skewness(wR^2)$ metrics of these blobs are computed on the base of blob contour length values with sampling rate 25 Hz and full window width $w$=0.4 sec (10 points).

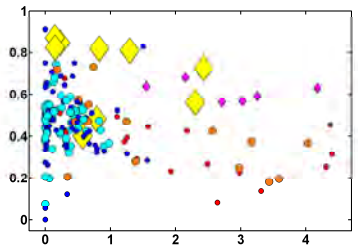



Fig. 2. The values of the $mean(wR^2)$ metrics of the blobs. The $x$ co-ordinate is the speed of the object; the $y$ co-ordinate is the $mean(wR^2)$ metrics value. Pedestrians are depicted by circles: small blue circles denote single walking persons; big cyan circles denote groups of walking persons; small red circles denote single running persons; big orange circles denote groups of running persons. Vehicles are depicted by diamonds: small magenta diamonds denote bicycles and big yellow diamonds denote cars.

The "speed-$mean(wR^2)$" and the "speed-$skewness(wR^2)$" diagrams (Fig. 2, 3) show clearly that these metrics allow the discrimination of fast moving persons and vehicles, but are rather useless for the discrimination of slow objects.

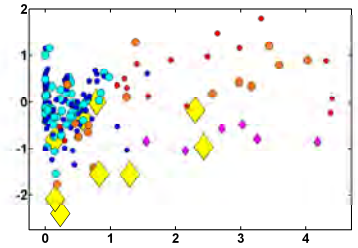



Fig. 3. The values of the $skewness(wR^2)$ metrics of the blobs. The $x$ co-ordinate is the speed of the object; the $y$ co-ordinate is the $skewness(wR^2)$ metrics value. Pedestrians and vehicles are depicted in the same manner as in Fig. 2.

These metrics are highly correlated, but are not equal still. Thus, we will use both of them as input arguments for the fuzzy logical inference. Note, that the sample mean (3) and the sample skewness (4) metrics have sense only if the sample is big enough; therefore we should use the cardinality (the number of values in the $wR^2$ set) as an additional argument in the fuzzy logical inference.

The size of blobs could be used as an additional statistical metrics for the discrimination of running people and vehicles. It is difficult to estimate the real physical size of objects in the video scene, but one can use a standardised size of the blobs that is a ratio:

$$StandArea(t) = \frac{BlobArea(t)}{CharactLength(x(t), y(t))^2} \tag{6}$$

where $BlobArea(t)$ is the area of the blob (in pixels) and $CharactLength(x, y)$ is the characteristic distance. The characteristic distance is a function of $x$ and $y$, where $(x, y)$ are

co-ordinates of the blob anchor point that is a function of time. It is convenient to define the anchor point $(x, y)$ as the centre of the rectangle blob. The following method of characteristic distance computation is used in Actor Prolog. Let $C$ be a one meter diameter circle created in the ground plane and the centre of the circle is $(x, y)$. Then the characteristic distance in the $(x, y)$ point is the maximal diameter (in pixels) of $C$ in the pixel space (see Fig. 4). The transfer between the real space and the pixel space is implemented using the projective transform matrix. In our example, the average value of the standardised area will be used for the discrimination of cars and other objects.

Fig. 4. An example of the computation of the characteristic distances in given positions of the ground plane. The green ellipse corresponds to the one meter diameter circle created around the position in the physical space. The red line corresponds to the characteristic distance, i.e., maximal diameter of the ellipse in the pixel space. The blue stick is an estimation of the vertical direction in the physical space.

The diagram in the Fig. 5 demonstrates that the $mean(StandArea)$ metrics really does work.

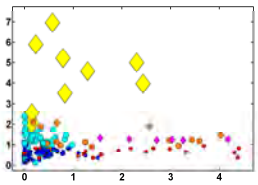


Fig. 5. The values of the $mean(StandArea)$ metrics of the blobs. The $x$ co-ordinate is the speed of the object; the $y$ co-ordinate is the $mean(StandArea)$ metrics value. Pedestrians and vehicles are depicted in the same manner as in Fig. 2.

We have implemented the $mean(wR^2)$, the $skewness(wR^2)$, and the $mean(StandArea)$ metrics in Java in the Vision standard package of the Actor Prolog language and use them for experimenting with the intelligent visual surveillance.

All results of the low-level analysis are received by the logic program in a form of Prolog terms describing the list of connected graphs. The connected graph of tracks is a list of underdetermined sets denoting separate edges of the graph. The nodes of the graph correspond to the points where tracks cross, and the edges are pieces of tracks between such points.

Here is an Actor Prolog program code with brief explanations. The logic program checks the graph of tracks and looks for the following pattern of interaction among several persons: "If two or more persons meet somewhere in the scene and one of them runs after the end of the meeting, the program should consider this scenario as a kind of a running away and a probable case of a sudden attack or a theft." So, the program has to alarm if this kind of sub-graph is detected in the total connected graph of tracks. In this case, the program will mark all persons in the inspected graph by yellow rectangles and outputs the "Attention!" warning in the middle of the screen (see Fig. 6).


Fig. 6. The logical inference has found a possible case of a street offence in the graph of blob trajectories. All probable participants of the conflict are marked by yellow rectangles. Multicoloured lines denote tracks of the blobs. The program estimates the velocity of the blobs and depicts it by different colours. Direct blue lines depict possible links between the blobs.

```
CLAUSES:
is_a_kind_of_a_running_away([E2|_],G,E1,E2,E3):-
     E2 == {inputs:O,outputs:B|_},
     B == [_,_|_],
     contains_a_running_person(B,G,E3),
     is_a_meeting(O,G,E2,E1),!.
is_a_kind_of_a_running_away([_|R],G,E1,E2,E3):-
     is_a_kind_of_a_running_away(R,G,E1,E2,E3).
contains_a_running_person([N|_],G,P):-
     get_edge(N,G,E),
     is_a_running_person(E,G,P),!.
contains_a_running_person([_|R],G,P):-
     contains_a_running_person(R,G,P).
is_a_meeting(O,_,E,E):-
     O == [_,_|_],!.
is_a_meeting([N1|_],G,_,E2):-
     get_edge(N1,G,E1),
     E1 == {inputs:O|_},
     is_a_meeting(O,G,E1,E2).
get_edge(1,[Edge|_],Edge):-!.
get_edge(N,[_|Rest],Edge):-
     N > 0,
     get_edge(N-1,Rest,Edge).
```

A fuzzy definition of the running person concept is as follows:

```
is_a_running_person(E,_,E):-
     E == {    frame1:T1,
               frame2:T2,
               mean_velocity:V,
               mean_standardized_area:A,
               wr2_mean:M,
               wr2_skewness:S,
               wr2_cardinality:C|_},
     is_a_fast_object(T1,T2,V),
     fast_object_is_a_runner(A,M,S,C),!.
```

```
is_a_running_person(E,G,P):-
     E == {outputs:B|_},
     contains_a_running_person(B,G,P).
```

The graph edge corresponds to the running person if and only if two conditions hold:

1) This edge is recognised as a fast object, i.e., the speed and the length of the graph edge satisfy the fuzzy definition of the fast object (see the logic program code below).
2) The values of the $mean(wR^2)$, the $skewness(wR^2)$, the $mean(StandArea)$ metrics, and the cardinality of the $wR^2$ metrics satisfy the fuzzy definition of the running pedestrian.

```
is_a_fast_object(T1,T2,V):-
     M1== ?fuzzy_metrics(V,1.7,0.7),
     D== (T2 - T1) / 25,
     M2== ?fuzzy_metrics(D,0.5,0.25),
     M1 * M2 >= 0.5.
fast_object_is_a_runner(A,M,S,C):-
     MC== ?fuzzy_metrics(C,7,2),
     MA== 1 - ?fuzzy_metrics(A,2.75,0.75),
     MM== 1 - ?fuzzy_metrics(M,0.49,0.10),
     MS== ?fuzzy_metrics(S,0.25,1.00),
     MA * MM * MS * MC >= 0.1.
```

The values of fuzzy thresholds used in the rules were computed on the basis of the BEHAVE samples. The fuzzy rules discriminate correctly all humans and vehicles that are fast objects in accordance to the fuzzy definition. More precisely, only 22 blobs from 193 are recognised as fast objects; the rules properly recognise 15 running pedestrians and 7 fast moving vehicles.

An auxiliary function that calculates the value of the fuzzy metrics is represented below:

```
fuzzy_metrics(X,T,H) = 1.0 :-
     X >= T + H,!.
fuzzy_metrics(X,T,H) = 0.0 :-
     X <= T - H,!.
fuzzy_metrics(X,T,H) = V :-
     V== (X-T+H) * (1 / (2*H)).
```

Even a simple video surveillance logic program has to contain a lot of elements, including video information gathering, low-level image analysis, high-level logical inference control, and reporting the results of intelligent visual surveillance; we have to emphasise that the logic programming approach allows one to implement all stages of the video data processing using the single logic language that is a prominent step in the approach to the intellectual visual surveillance.

We have created a research software platform based on the Actor Prolog concurrent object-oriented logic language and a state-of-the-art Prolog-to-Java translator for studying the intelligent visual surveillance. The platform includes the Actor Prolog logic programming system and an open source Java library of Actor Prolog built-in classes [11]. It is intended to facilitate the study of the intelligent monitoring of anomalous people activities, the logical description and analysis of people behaviour (see Web Site [12]).

Authors are grateful to Abhishek Vaish, Vyacheslav E. Antciperov, Vladimir V. Deviatkov, Aleksandr N. Alfimtsev, Vladislav S. Popov, and Igor I. Lychkov for co-operation.

We acknowledge a partial financial support from the Russian Foundation for Basic Research, grant No 13-07-92694.

## REFERENCES

[1] P. V. K. Borges, N. Conci, and A. Cavallaro, "Video-based human behavior understanding: A survey," *IEEE Trans. Circuits Syst. Video Techn.*, pp. 1993–2008, 2013.

[2] A. Skarlatidis, A. Artikis, J. Filippou, and G. Paliouras, "A probabilistic logic programming event calculus," *Theory and Practice of Logic Programming*, vol. FirstView, pp. 1–33, 9 2014. [Online]. Available: http://journals.cambridge.org/article_S1471068413000690

[3] A. A. Morozov and A. F. Polupanov, "Intelligent visual surveillance logic programming: Implementation issues," in *CICLOPS-WLPE 2014*, ser. Aachener Informatik Berichte, T. Ströder and T. Swift, Eds., no. AIB-2014-09. RWTH Aachen University, Jun. 2014, pp. 31–45. [Online]. Available: http://aib.informatik.rwth-aachen.de/2014/2014-09.pdf

[4] A. A. Morozov, "Actor Prolog: an object-oriented language with the classical declarative semantics," in *IDL 1999*, K. Sagonas and P. Tarau, Eds., Paris, France, Sep. 1999, pp. 39–53. [Online]. Available: http://www.cplire.ru/Lab144/paris.pdf

[5] A. Morozov and Y. Obukhov, "An approach to logic programming of intelligent agents for searching and recognizing information on the Internet," *Pattern Recognition and Image Analysis*, vol. 11, no. 3, pp. 570–582, 2001. [Online]. Available: http://www.cplire.ru/Lab144/pria570m.pdf

[6] A. A. Morozov, "On semantic link between logic, object-oriented, functional, and constraint programming," in *MultiCPL 2002*, Ithaca, NY, USA, Sep. 2002, pp. 43–57. [Online]. Available: http://www.cplire.ru/Lab144/multicpl.pdf

[7] A. A. Morozov, "Logic object-oriented model of asynchronous concurrent computations," *Pattern Recognition and Image Analysis*, vol. 13, no. 4, pp. 640–649, 2003. [Online]. Available: http://www.cplire.ru/Lab144/pria640.pdf

[8] A. A. Morozov, "Operational approach to the modified reasoning, based on the concept of repeated proving and logical actors," in *CICLOPS 2007*, V. S. C. Salvador Abreu, Ed., Porto, Portugal, Sep. 2007, pp. 1–15. [Online]. Available: http://www.cplire.ru/Lab144/ciclops07.pdf

[9] R. Fisher, "BEHAVE: Computer-assisted prescreening of video streams for unusual activities. the EPSRC project GR/S98146." 2013. [Online]. Available: http://groups.inf.ed.ac.uk/vision/BEHAVEDATA/INTERACTIONS/

[10] A. A. Morozov, A. Vaish, A. F. Polupanov, V. E. Antciperov, I. I. Lychkov, A. N. Alfimtsev, and V. V. Deviatkov, "Development of concurrent object-oriented logic programming system to intelligent monitoring of anomalous human activities," in *BIODEVICES 2014*, A. C. Jr., G. Plantier, T. Schultz, A. Fred, and H. Gamboa, Eds. SCITEPRESS, Mar. 2014, pp. 53–62. [Online]. Available: http://www.cplire.ru/Lab144/biodevices2014.pdf

[11] A. A. Morozov, "A GitHub repository containing source codes of Actor Prolog built-in classes," 2014. [Online]. Available: https://github.com/Morozov2012/actor-prolog-java-library

[12] A. A. Morozov and O. S. Sushkova, "Intelligent visual surveillance logic programming Web Site," Kotel'nikov Institute of Radio Engineering and Electronics of RAS, Moscow, Russia, 2014. [Online]. Available: http://www.fullvision.ru/actor_prolog_2014

# Efficient Multi-temporal hyperspectral signatures classification using a Gaussian-Bernoulli RBM based approach

Selim Hemissi
Telecom Bretagne, Brest, France
Ecole nationale des Sciences de l'Informatique
Email: selim.hemissi@ensi.rnu.tn

Imed Riadh Farah
Telecom Bretagne, Brest, France
Ecole nationale des Sciences de l'Informatique
Email: riadh.farah@ensi.rnu.tn

***Abstract*—This paper presents an efficient Gaussian-Bernoulli Restricted Boltzmann Machines (GR-RBM) framework in order to better address the classification challenge of remotely sensed images. The proposed approach relies on well-designed features for a new 3D modality of spectral signature. First, mesh smoothing is introduced to reduce noise while conserving the main geometric features of the multi-temporal spectral signature. Then, we propose the use of an RBM framework as stand-alone non-linear classifier. The adapted framework focuses on a cooperative integrated generative-discriminative objective allowing us to model features of input layer and their classification in one-pass algorithm. The main benefit of the proposed approach is the ability to learn more discriminative features. We evaluated our approach within different scenarios and we demonstrated its usefulness for noisy high dimensional hyperspectral images.**

## I. Introduction

The great spectral resolution of hyperspectral images offers an exceptional characterization of land-cover types with unrivaled accuracy. Multi-temporal Hyperspectral acquisition is the fastest growing technology in the fields of remote sensing, change detection, environmental monitoring etc. It produces spectra of various hundred wavelengths with fine spatial resolution. Therefore, each pixel in the hyperspectral cube is depicted by a rich spectral information allowing a better characterization of land-cover types.

Nevertheless, the processing of multi-temporal images turns out to be more challenging than common aerial or multispectral images. This is mainly due to many reasons. First, the high dimensionality or curse of dimensionality of data known as Hughes Phenomenon [1]. In fact, the rich number of spectral bands coupled with the unavailability of training samples exacerbates this problem. Consequently, overfitting result must be alleviated by putting more attention to the choice and the configuration of the classifier.
The second problem is related to the probable non-linear spread of the data classes and the singular noise and uncertainty level. Such uncertainties and non-linearities may result from several factors, including temporal variability, heterogeneities of mixed pixels, as well as acquisition distortions [2] . Such facts lead to a distinct non-linear aspect of the classifier. This paper presents an efficient approach to better address these challenges.

To overcome these issues, proper classifier should be privileged. The classifier has to deal with voluminous data and non-linear multi-class data. Moreover, thoughtful features must be used and afforded to the classifiers to reach a high accuracy rate.

## II. Proposed approach

The few years have seen significant increase in the amount and the availability of multi-temporal images, and it brings a great challenge to the traditional pattern recognition approaches. To overcome these drawbacks, we proposed in [3] a novel model of the multi-temporal spectral signature. This model is defined by putting the classical spectral signatures of the same pixel at different times in the same space. We obtain a bounded domain $\Omega$ and denote $E = \{p_1, \ldots, p_n\}$ the set of points in $R^3$ called sites. The Delaunay triangulation of $E$, noted $Del(E)$, is the geometric dual of $Vor(E)$. We denote by $Vor(E)$ the Voronoi diagram of $E$ which is the subdivision of space investigated by the Voronoi cells $V(p_1), \cdots, V(p_n)$. Therefore, we obtain for each pixel a 3D multi-temporal signatures illustrated by figure 1. This new model incorporates the time ($T$) and the spectral ($\lambda$) dimensions. So, we can express the reflectance ($Ref$) at each pixel using the equation 1.

$$Ref_{Pixel_{i,j}} = f(T, \lambda) \quad (1)$$

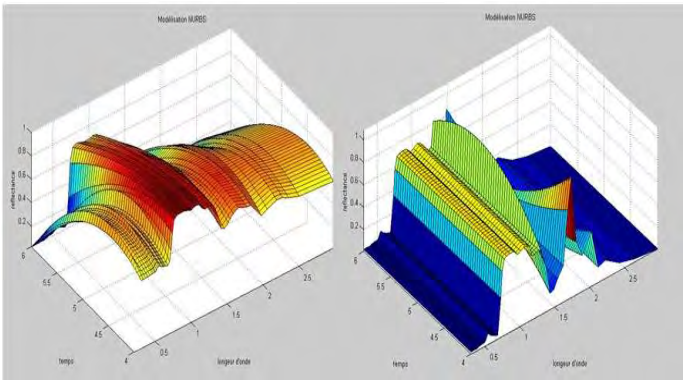


Fig. 1. Multi-temporal hyperspectral signature

### A. 3D hyperspectral signature smoothing for noise reduction

In this stage, the atmospheric distortions affecting the reflectance values of pixels alter the position of points in the 3D reconstruction space. Subsequently, this noise turns into a geometric noise which requiring a specific treatment in order to reduce its effects on the classification accuracy. The first contribution of this paper is the use of Laplacian smoothing in order to limit the impact of noise. Our choice was guided by the aim to establish a simple and robust optimization of 3D signatures, while maintaining both sampling rate and connectivity. The choice of Laplacian smoothing technique is motivated by the fact that it changes the position of nodes without modifying the topology of the multi-temporal signatures.
Laplacian smoothing is a well studied method for polygonal mesh [4]. We denoted by $S$ a triangular mesh surface, $X$ is the vertices of the mesh, $L$ is the Laplacian, and $\lambda$ is a scalar that controls the diffusion speed. Our goal is to produce a new mesh surface $S'$ with an enhanced quality ratio. This quality reflects a critical factor in the accuracy and stability of pattern recognition tasks. The key idea is to, incrementally, move the vertices of the mesh $S$ in the direction of the Laplacian. Since that the Laplacian operator is linear, the smoothing equation is :

$$X(n+1) = (I + \lambda dtL)X(n) \tag{2}$$

This investigation allows us to generate a smoothed 3D signature which is close as possible to $S$ and preserves the features of $S$. Then we generate for each pixel a feature vector denoted $x_i$ [3].

### B. 3D hyperspectral signatures classification using GB-RBM algorithm

The data produced by the previous stage are characterised maily by their high dimensionality and their probable non-linear nature. To improve their classification, we will need more sophisticated and adapted framework. The Gaussian-Bernoulli Restricted Boltzmann Machines (GR-RBM) is a stochastic neural network including a non-linear generative model. The theoretical proofs of this model correspond to the requirements of our data. Recently, RBMs have attracted much attention to address a wide range of pattern recognition problems. But, they are mainly employed to initialize deep neural networks [5]. Briefly, the RBM model mainly includes $m$ visible units $v = (v_1, \ldots, v_m)$ and $n$ hidden units $h = (h_1, \ldots, h_n)$, with fully connecting between them (Figure 2).

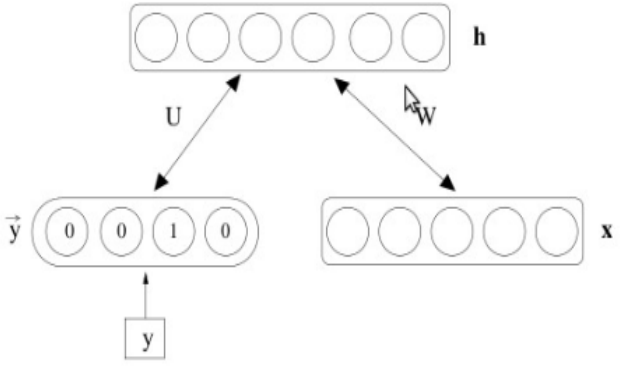



Fig. 2. RBM Architecture

The visible units are combined with the feature vectors extracted in the previous section.
In our case, we assume given a training set $Train = \{(x_i, y_i)\}$ involving for the $i^{th}$ pixel an input feature vector $x_i$ and a target class $y_i \in \{1, \cdots, C\}$. As known, the generative learning objective of standard RBMs aims to minimize the energy of the inputs and to induce the most truthful representation of the input. Nevertheless, this representation, which aims to model the intra-class variation of the learning samples, is not automatically useful for a classifier. Therefore, the goal of this part is to adapt the well-known learning objective to the classification of multi-temporal hyperspectral images.
Recent researches such as Larochelle et al. [6] and Louradour et al. [7] prove that RBMs offers a self-contained framework to implement an efficient non-linear classifier. Consequently, we propose to use a learning objective that promotes inter-class separability. To achieve this goal, the Gaussian-Bernoulli RBM (GB-RBM) framework was adopted including $v_m$ visible units with real-value and $h_n$ binary hidden units. Following the same theory, the energy function of the GB-RBM is expressed as :

$$E(v, h|\theta) = -\sum_{i=1}^{m}\sum_{k=1}^{n}\omega_{ij}h_j\frac{v_i}{\sigma_i^2} - \sum_{i=1}^{m}\frac{(v_i - b_i)^2}{2\sigma_i^2} - \sum_{j=1}^{n}c_i h_j, \tag{3}$$

Under this configuration, the computation of conditionnal probabilities is done by using the following experssions :

$$p(v_i = v|\mathbf{h}) = \boldsymbol{N}(v|b_i + \sum_j h_j\omega_{ij}\sigma_i^2) \tag{4}$$

$$p(h_i = 1|v) = sigmoid(c_j + \sum_i \omega_{ij}\frac{v_i}{\sigma_i^2}) \tag{5}$$

where $\boldsymbol{N}$ denotes the Gaussian probability density function with mean $\mu$ and variance $\sigma^2$. As shown, the proposed framework is a derived RBM model that models the conditional distribution $p(y\,|x)$ instead of $p(y, x)$. To train the GR-RBM model, we use the contrastive divergence algotihm.

Training RBMs is often proportional to the choice of learning rate and its scheduling. In accordance with the experiments results shown later, SVM classification outperforms GB-RBM in view of its sensitivity. Moreover, must studies reveal that if the learning rate is not annealed over time approaching zero, GB-RBMs diverges easily after initial convergence. To overcome this issue, we limit the maximum learning rate.

### C. Experimental results

This section illustrates the performance of the proposed method in a challenging multi-temporal classification case. The proposed approach was tested on two different data sets. The datasets involve several types of data, and with dimensions ranging from 176 to 183 bands. The first dataset, *Hyperion*, contains vegetation type data, is divided into five classes, has 183 spectral bands and has a pixel size of $30m$. The second set is from an airborne sensor (*AVIRIS*), divided into 7 classes, has 176 spectral bands and a pixel size of $18m$. First, we present experiments that assess the classification accuracy of the proposed approach (PA). We also included an SVM classifier and a Multilayer Perceptron (MP) classifier in our comparison as a baseline. Table (I) and table (II) summarizes the obtained results.

TABLE I. SEQUENCE OF REAL IMAGES (RED/GREEN/BLUE (RGB) COMPOSITION, BANDS [6,19,33]) AND THEIR CORRESPONDING TRUE CLASSIFICATION MAPS (FIRST TIME SERIE : 2009/2010).

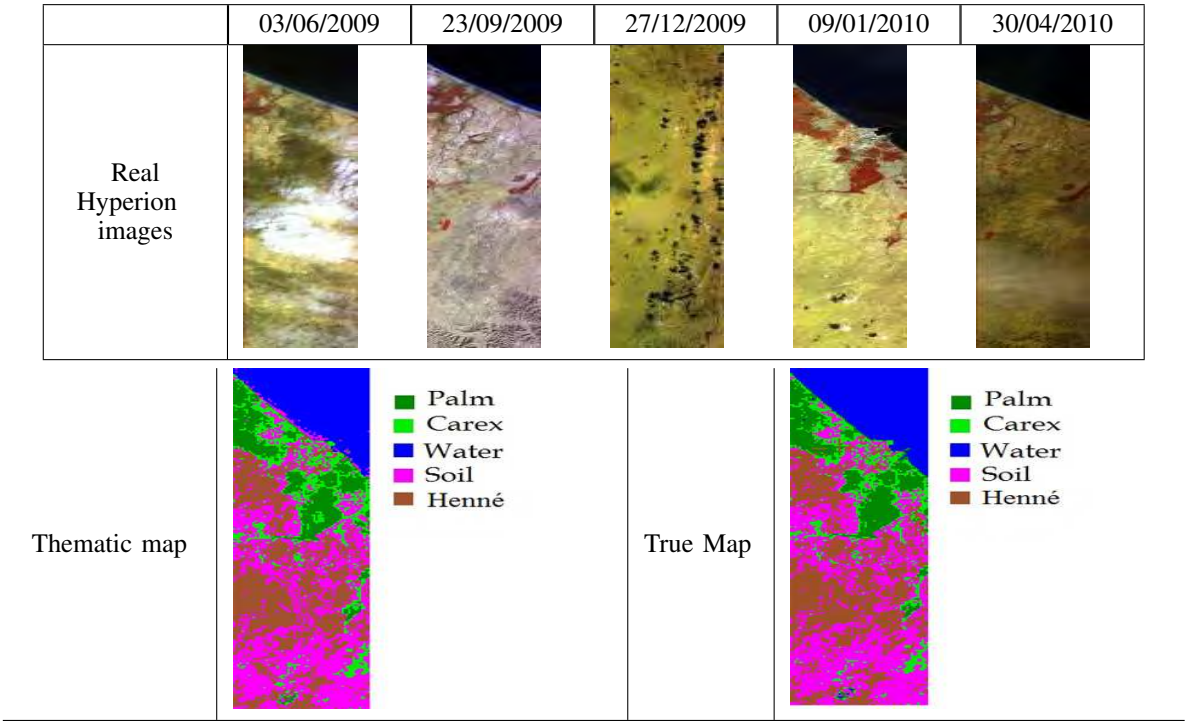



TABLE II. EVALUATION OF THE PROPOSED APPROACH COMPARED TO SEVERAL CONVENTIONAL APPROACHES

| Classifier | Overall Accuracy (%) | Kappa(%) |
|---|---|---|
| Multilayer Perceptron (MP) | 84,65 | 0.68 |
| Proposed Approach | 89.63 | 0.76 |
| Support Vector Machines (SVMs) | 98.11 | 0.75 |

### *D. Discussion and conclusions*

Recent advances in multi-temporal hyperspectral imaging offer the opportunity to easily integrate temporal and spatial facets by using a $3D$ model of the spectral signature. While improving objects recognition, some challenges such as the curse of dimensionality and noisy data are still difficult to address and need further improvements. To overcome these problems, this paper introduces two main innovations : I) the noise affecting the multi-temporal signatures is modeled and reduced by a post-processing Laplacian smoothing method, and II) the concept of Restricted Boltzmann Machines (RBM) is introduced for hyperspectral data classification and the Gaussian-Bernoulli RBM framework is adapted to handle this challenge. The experiments prove the competitive results of the proposed approach.

## REFERENCES

[1] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *Information Theory, IEEE Transactions on*, vol. 14, no. 1, pp. 55–63, Jan 1968.

[2] J. Chi and M. M. Crawford, "Selection of landmark points on nonlinear manifolds for spectral unmixing using local homogeneity," *IEEE Geosci. Remote Sensing Lett.*, vol. 10, no. 4, pp. 711–715, 2013. [Online]. Available: http://dx.doi.org/10.1109/LGRS.2012.2219613

[3] S. Hemissi, I. Farah, K. Saheb Ettabaa, and B. Solaiman, "Multi-spectro-temporal analysis of hyperspectral imagery based on 3-d spectral modeling and multilinear algebra," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 51, no. 1, pp. 199–216, Jan 2013.

[4] A. Nealen, T. Igarashi, O. Sorkine, and M. Alexa, "Laplacian mesh optimization," in *Proceedings of the 4th International Conference on Computer Graphics and Interactive Techniques in Australasia and South-east Asia*, ser. GRAPHITE '06. ACM, 2006, pp. 381–389.

[5] T. Kuremoto, S. Kimura, K. Kobayashi, and M. Obayashi, "Time series forecasting using a deep belief network with restricted boltzmann machines," *Neurocomputing*, vol. 137, pp. 47–56, 2014.

[6] H. Larochelle, M. Mandel, R. Pascanu, and Y. Bengio, "Learning algorithms for the classification restricted boltzmann machine," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 643–669, Mar. 2012. [Online]. Available: http://dl.acm.org/citation.cfm?id=2503308.2188407

[7] J. Louradour and H. Larochelle, "Classification of sets using restricted boltzmann machines," *CoRR*, vol. abs/1103.4896, 2011. [Online]. Available: http://arxiv.org/abs/1103.4896

Proceedings of the 9th Open German-Russian Workshop on Pattern Recognition and Image Understanding
Dec 1-5, 2014, Koblenz, Germany

# Evaluation of established line segment distance functions

Stefan Wirtz and Dietrich Paulus
University of Koblenz-Landau
Universittsstrae 1
56070 Koblenz
Email: wirtzstefan@uni-koblenz.de

***Abstract*—In this paper we present an evaluation of six well established line segment distance functions within the scope of line segment matching. We show analytically, using synthetic data, the properties of the distance functions with respect to rotation, translation, and scaling. The evaluation points out the main characteristics of the distance functions. In addition, we demonstrate the practical relevance of line segment matching and introduce a new distance function.**

## I. Introduction

Evaluating the similarity of two geometric shapes is an important issue in different fields of computer science including computer vision and pattern recognition.

We will introduce a new distance function and present an evaluation study of this method and six established ones within the scope of line matching. The evaluation lines out the characteristics of these distance functions. The requirements of line matching differ depending on the application domain. In pairwise image matching applications it is useful to be invariant against blurring, scaling, rotation and translation. In other cases, it is import to be variant against all these attributes. For example, if the aim is to search the best fitting rendering of a taken image using a set of 2-D renderings of a 3-D object, it is is important to notice differences in scaling, rotation and translation.

The knowledge of the requirements of an application is indispensable to choose eligible distance functions. For that reason, we describe and analyze in this approach established distance functions for lines or line segments: Hausdorff-distance (HD), Trucco-distance (TD), Modified line segment Hausdorff-distance (MHD), Modified perpendicular line segment Hausdorff-distance (MPHD), Midpoint-distance (MD), Closest point-distance (CD) and our new Straight line-distance function (SD) – definitions and references will be given below. These functions are applied on and analyzed with simulated data which allows an exact evaluation. This approach focus on the evaluation of the distance function independently of the matching algorithm – knowing that there exist a lot of sophisticated approaches for line matching.

We introduce the related work in Section II. Section III describes the distance function and analysis in detail. We show and discuss our experiments and results in Section IV. A summary can be found in Section V.

## II. Related Work

We describe approaches dealing with chamfer, line, and segment matching. For these topics an extensive literature exists which we summarize to a limited overview in the following.

Bay et al. [6] present an approach for matching line segments between two uncalibrated wide-baseline images. This approach is able to estimate the fundamental matrix robustly even from line segments only. In that approach no prior knowledge about the scene or camera positions is needed. The authors generate an initial set of line segment correspondences, which is iteratively increased by adding matches consistent with the topological structure of the current ones. Finally, a coplanar grouping stage allows to estimate the fundamental matrix.

Schmid and Zisserman [4] describe two algorithmic approaches. These methods require apriori the fundamental matrix of an image pair, or the trifocal tensor of an image triple, and cover the cases of both short and long range motion. Both methods use cross-correlation as matching scores for similarity measurements of line segments, whereby the method for long range motion has to adapt additionally the correlation measure.

Werner and Zisserman [5] present a fully automatic approach for reconstructing of buildings from multiple images. They fit geometric models by sweeping and introduce for that purpose sweeping scene planes about a line at infinity, correlation based search for building edges using translating rectangles, and inter image homographies; they search for local gradient maxima along translating line segments.

Witt and Weltin [11] present the approach *Iterative Closest Lines* which extends the algorithm *Iterative Closest Points* (ICP)[7] on line segment matching. The approach works similar to ICP: Line segments are detected, line correspondences to this line segments with minimal Euclidean distances are searched and a 3d rigid transformation is applied on the correspondences.

Liu et al. [2] introduce a study dealing with the object localization problem in images given a single hand-drawn example or a gallery of shapes as the object model. They present an approach for improving the accuracy of chamfer matching while reducing its computational cost. For these purposes, they proposed a method for incorporating the edge orientation in the cost function and solving the matching problem in the orientation augmented space. The novel cost

function is smooth and can be computed in sublinear time in the size of the shape template.

Sudarshan [8] addresses the problem of determining the path of a vehicle on a given vector map of roads, based on tracking data such as that obtained from onboard GPS receivers. He presents techniques for modifying existing methods for map-matching based on the idea of segment-wise matching, where a segment is contiguous sequence of track points. Track points and segments are assigned scores that quantify the expected accuracy of matching them to map features. Therefore, the Frechet distance is used, which takes the position along paths into account.

## III. Approach

We consider in our application that we have two sets of line segments and do not care where they come from (e.g. Canny edge detector or a rendered 3-D CAD model). The source of the line segments is unimportant for the examination of the distance function, because we want to evaluate these functions independently of line segment extraction mechanism.

In this approach, we work with points in pixel coordinates and define a line segment l by $\mathrm{l}_i = p_{1,\mathrm{l}_i} \times p_{2,\mathrm{l}_i}$. The geometric object G is the generalization of the set of line segments L.

A distance function for line segments is described as $\mathrm{d}^{(1)} : \mathrm{l} \times \mathrm{l} \mapsto \mathbb{R}^+$ and measures the dissimilarity of line segments, where a score of zeros means that that they are identical.

For the matter of line matching, we match the sets $\mathrm{G}_i = \{g_{1,i}, g_{2,i}, \ldots, g_{N_i,i}\}$ with $i = \{1, 2\}$. Resulting we get sets of matched geometric objects $\mathrm{G}_i^{(m)}$ and sets of unmatched geometric objects $\overline{\mathrm{G}_i^{(m)}}$ with $\mathrm{G}_i^{(m)} \cup \overline{\mathrm{G}_i^{(m)}} = \mathrm{G}_i$. The assignment $\gamma$ will be optimized to yield $\gamma^*$ by using a distance function and a penalty function $\delta_p$ which punishes unmatched geometric objects.

$$\gamma^* = \underset{\gamma}{\operatorname{argmin}} \sum_{g \in \mathrm{G}_i^{(m)}, g' \in \overline{\mathrm{G}_1^{(m)}}, g'' \in \overline{\mathrm{G}_2^{(m)}}} \mathrm{d}(\gamma(g), g) + \delta_{p_1}(g') + \delta_{p_2}(g'') \tag{1}$$

The penalty functions are described as $\delta_{p_j} : \overline{\mathrm{G}_j^{(m)}} \mapsto \mathbb{R}^+$, $j = \{1, 2\}$ and can be identical with $\delta_{p_1} = \delta_{p_2}$.

### A. Distance functions

For the evaluation we use the (i) Hausdorff-distance, (ii) Trucco-distance, (iii) Modified line segment Hausdorff-distance, (iv) Perpendicular-distance, (v) Midpoint-distance, (vi) Closest point-distance, and (vii) our own Straight line-distance function. which all measure the distance of two given line segments $l_1$ and $l_2$.

The *(i) Hausdorff line segment-distance* [9] is illustrated in figure 1 and described as following:

$$\mathrm{d}_{\mathrm{l}_1,\mathrm{l}_2}(\mathrm{l}_1, \mathrm{l}_2) = \max_{\boldsymbol{p} \in \mathrm{l}_1} \min_{\boldsymbol{q} \in \mathrm{l}_2} \|\boldsymbol{p} - \boldsymbol{q}\| \tag{2}$$

$$\mathrm{d}_{\mathrm{l}_2,\mathrm{l}_1}(\mathrm{l}_1, \mathrm{l}_2) = \max_{\boldsymbol{p} \in \mathrm{l}_2} \min_{\boldsymbol{q} \in \mathrm{l}_1} \|\boldsymbol{p} - \boldsymbol{q}\| \tag{3}$$

$$\mathrm{d}_{\mathrm{Hausdorff}}(\mathrm{l}_1, \mathrm{l}_2) = \max(\mathrm{d}_{\mathrm{l}_1,\mathrm{l}_2}, \mathrm{d}_{\mathrm{l}_2,\mathrm{l}_1}). \tag{4}$$

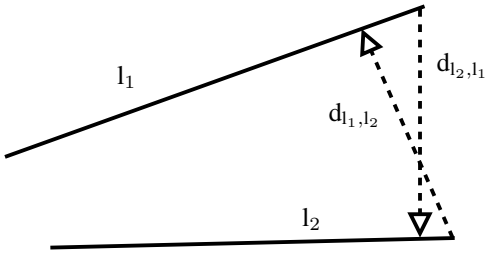



Fig. 1. Visualization of the used quantities of the Haussdorf-distance.

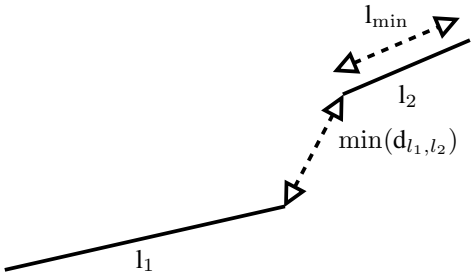



Fig. 2. Visualization of the used quantities of the Trucco-distance.

The *(ii) Trucco-distance* [10] is illustrated in figure 2 and described as following:

$$\mathrm{l}_{\min} = \min(\|l_1\|, \|l_2\|) \tag{5}$$

$$\min(\mathrm{d}_{l_1,l_2}) = \min(\|p_{1,l_1} - p_{1,l_2}\| \|p_{1,l_1} - p_{2,l_2}\|, \|p_{2,l_1} - p_{1,l_2}\|, \|p_{2,l_1} - p_{2,l_2}\|) \tag{6}$$

$$\mathrm{d}_{\mathrm{Trucco}}(l_1, l_2) = \left( \frac{\mathrm{l}_{\min}}{\min(\mathrm{d}_{l_1,l_2})} \right)^2. \tag{7}$$

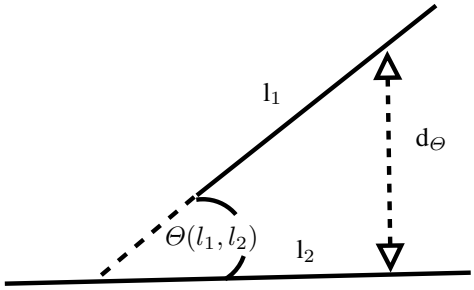



Fig. 3. Visualization of the used quantities of the Modified line segment Hausdorff-distance.

The *(iii) Modified line segment Hausdorff-distance* [3] is illustrated in figure 3 and uses the angle function $\Theta(l_1, l_2) : \mathbb{R}^2 \times \mathbb{R}^2 \mapsto \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ and includes them into

$$\mathrm{d}_{\Theta}(l_1, l_2) = \min(\|l_1\|, \|l_2\|) \ \sin(\Theta(l_1, l_2)). \tag{8}$$

The *(iv) Modified perpendicular line segment Hausdorff-distance* [1] is illustrated in figure 4 and uses the perpendicular-

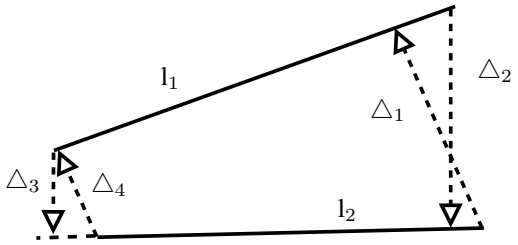



Fig. 4. Visualization of the used quantities of the Modified line segment Modified perpendicular line segment Hausdorff-distance with $\triangle_1 = \max(\mathrm{d}_{\perp,\mathrm{l}_1,1}, \mathrm{d}_{\perp,\mathrm{l}_1,2})$, $\triangle_2 = \max(\mathrm{d}_{\perp,\mathrm{l}_2,1}, \mathrm{d}_{\perp,\mathrm{l}_2,2})$ $\triangle_3 = \min(\mathrm{d}_{\perp,\mathrm{l}_1,1}, \mathrm{d}_{\perp,\mathrm{l}_1,2})$ and $\triangle_4 = \min(\mathrm{d}_{\perp,\mathrm{l}_2,1}, \mathrm{d}_{\perp,\mathrm{l}_2,2})$.

distance $\mathrm{d}_\perp(l_1, l_2) = s_\perp$ and including them into

$$s_{\perp,1} = \min(\max(\mathrm{d}_{\perp,\mathrm{l}_1,1}, \mathrm{d}_{\perp,\mathrm{l}_1,2}), \max(\mathrm{d}_{\perp,\mathrm{l}_2,1}, \mathrm{d}_{\perp,\mathrm{l}_2,2})) \tag{9}$$

$$s_{\perp,2} = \min(\min(\mathrm{d}_{\perp,\mathrm{l}_1,1}, \mathrm{d}_{\perp,\mathrm{l}_1,2}), \min(\mathrm{d}_{\perp,\mathrm{l}_2,1}, \mathrm{d}_{\perp,\mathrm{l}_2,2})) \tag{10}$$

$$w_i = \frac{s_{\perp,i}}{(s_{\perp,1} + s_{\perp,2})} \text{ with } i = \{1, 2\} \tag{11}$$

$$\mathrm{d}_{\perp\mathrm{mod}}(l_1, l_2) = \frac{1}{2}(w_1\ s_{\perp,1} + w_2\ s_{\perp,2}). \tag{12}$$

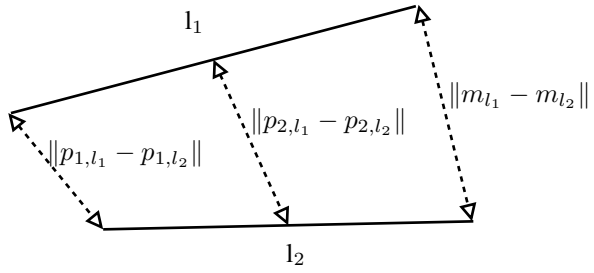



Fig. 5. Visualization of the used quantities of the Midpoint-distance.

The *(v) Midpoint-distance* is illustrated in figure 5 and uses the midpoints of the line segments $m_{l_i} = (p_{2,l_i} - p_{1,l_i})/2$ and includes this into

$$\mathrm{d}_{\mathrm{midpoint}}(l_1, l_2) = \|p_{1,l_1} - p_{1,l_2}\| + \|p_{2,l_1} - p_{2,l_2}\| + 3\|m_{l_1} - m_{l_2}\|. \tag{13}$$

The *(vi) Closest point-distance* [5] is illustrated in figure 6 and described as

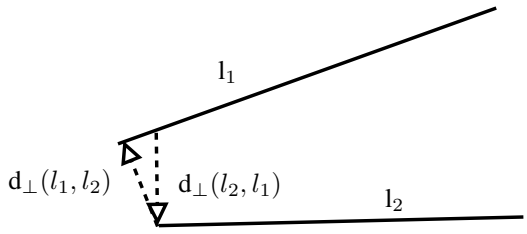



Fig. 6. Visualization of the used quantities of the Closest Point-distance.

$$\mathrm{d}_{\mathrm{closestpoint}}(l_1, l_2) = min(\mathrm{d}_\perp(l_1, l_2), \mathrm{d}_\perp(l_2, l_1)). \tag{14}$$

The *(vii) Straight line-distance* uses $\mathrm{d}_1 = \|p_{1,l_1} - p_{1,l_2}\|$, $\mathrm{d}_2 = \|p_{1,l_1} - p_{2,l_2}\|$, $\mathrm{d}_3 = \|p_{2,l_1} - p_{1,l_2}\|$ and $\mathrm{d}_4 = \|p_{2,l_1} - p_{2,l_2}\|$ and includes them into:

$$\mathrm{d}_t(l_1, l_2) = \frac{\mathrm{d}_1 + \mathrm{d}_2 + \mathrm{d}_3 + \mathrm{d}_4}{4} - \frac{\|l_1\| + \|l_2\|}{4} \tag{15}$$

$$\mathrm{d}_{\mathrm{ST}}(l_1, l_2) = \mathrm{d}_{\mathrm{closestpoint}} + 0.25\ \mathrm{d}_\Theta(l_1, l_2) + \mathrm{d}_t \tag{16}$$

$$\mathrm{d}_{straightLine}(l_1, l_2) = min(\mathrm{d}_{\mathrm{ST}}(l_1, l_2), \mathrm{d}_{\mathrm{ST}}(l_2, l_1)). \tag{17}$$

### B. Analytical Evaluation

Distance functions have to distinguish line segments using their parameters which are the position, length and orientation.

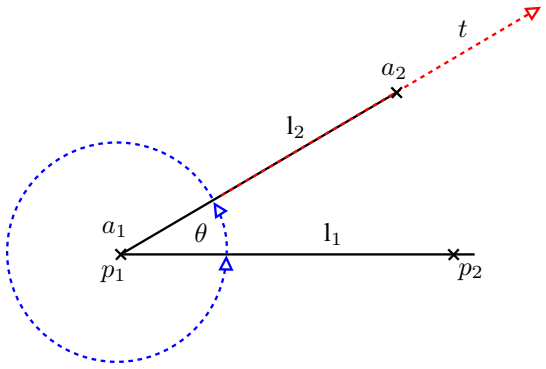



Fig. 7. Process of the evaluation.

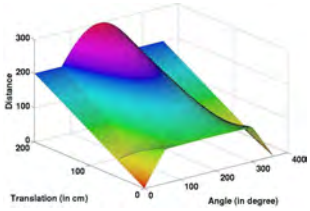
In general we assume that the metric of the line segments does not influence the measurement of the distance function. Therefore, we choose as measurement unit of centimeters and not pixels, because the number of pixel (resolution) in an image can vary from less than VGA*(=640x480)* up to multiple of Full HD*(=1920x1080)*.

Therefore, we generate two line segments $l_1$ and $l_2$ of length 100cm. The line segment $l_2$ will be rotated around $l_1$, translated in direction of $l_2$ and scaled where the point $p_1$ stays equal. We rotate with $\theta = [0,\ 2\pi]$, translate with $t = [0\mathrm{cm},\ 200\mathrm{cm}]$ and scale line segment $l_2$ from $[1\mathrm{cm},\ 200\mathrm{cm}]$ (see Fig. 7). In section IV the results are presented and discussed.

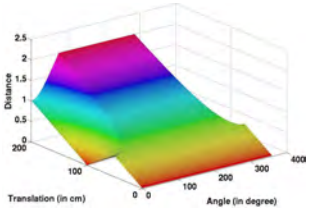
## IV. Experiments and Results

The results of the analytical evaluation are shown in the Fig. 8(a)-8(g). Fig. 8 clarifies the meaning of translation and rotation for the respective distance function. The Trucco- and the Closest point-distance (Fig. 8(b) and 8(f)) are inappropriate to distinguish line segments with different angles. Also, the Trucco-distance is not monotone increasing. Monotonicity is important if the aim is to interpret the score of the distance function as dissimilarity-degree.

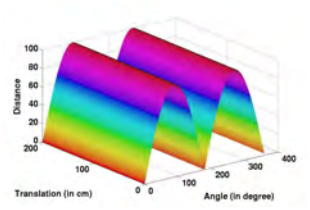
The Modified line segment Hausdorff- and the Modified perpendicular line segment Hausdorff-distance (Fig. 8(c) and 8(e)) distinguish very well differences of the angles, but are not able to distinguish line segments which are adjoining, what becomes apparent at the angle of $180°$ where the line segments lie next to each other. The Hausdorff-distance (Fig. 8(a)) like the Trucco-distance does not increase monotone for the test scenario, but distinguish very well angle- and translation changes. The missing monotony property yields that the distance outputs are not clearly interpretable. The remaining



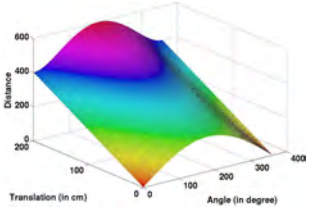

(a) Hausdorff line segment-distance



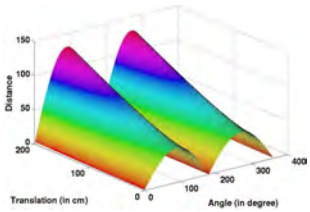

(b) Trucco-distance



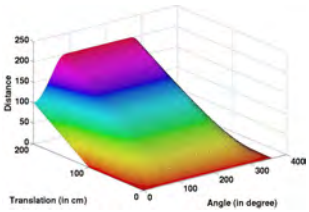

(c) Modified line segment Hausdorff-distance



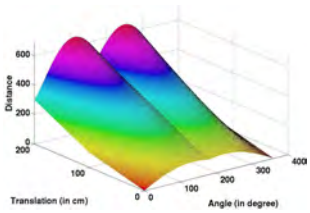

(d) Midpoint-distance




(e) Modified perpendicular line segment Hausdorff-distance

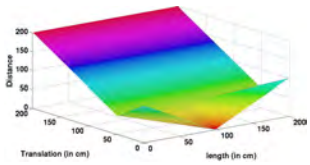



(f) Closest point-distance

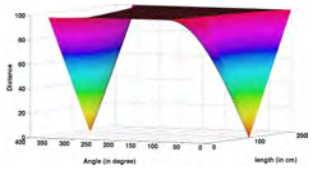



(g) Straight line-distance

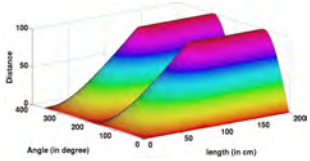
Fig. 8. Distance functions with the angle in degree on the $x$-axis, the translation in centimeter on the $y$-axis and the distance on the z-axis.



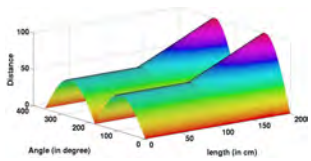

(a) Hausdorff line segment-distance - translation to length



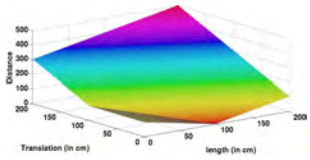

(b) Hausdorff line segment-distance - angle to length



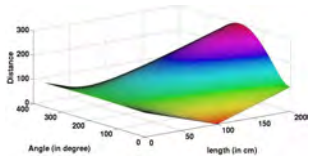

(c) Modified line segment Hausdorff-distance - angle to length



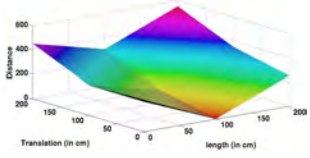

(d) Modified perpendicular line segment Hausdorff-distance - angle to length



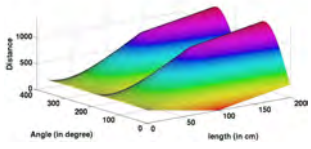

(e) Midpoint-distance - translation to length




(f) Midpoint-distance - angle to length




(g) Straight line-distance - translation to length




(h) Straight line-distance - angle to length

Fig. 9. Distance functions which are affected by the length of the line segments with the length on the $x$-axis and the angle in degree or the translation in centimeter on the $y$-axis and the distance on the z-axis. The not plotted value (translation or angle) is set to zero.

Midpoint- and Straight line-distance (Fig. 8(d) and 8(g)) seem quite appropriate to estimate rotation and translation. They are also monotone increasing.

Next to translation and rotation errors, which could be caused by imprecision of the used data or due to segmentation errors in the image, we have to handle incomplete data, which results in differences between the length of the corresponding line segments. Fig. 9(a) - 9(h) show the distance functions which are affected by the length of the line segments.

One distance function which is independent of the line segment length is the closest point-distance. The other one is the Trucco-distance which uses a normalization of the line segment length and is thus also independent of the line segment length. However, the Hausdorff-distance depends on the line length, but from a certain error in translation or angle the length difference is of no consequence (see Fig. 9(a) and 9(b)). The Modified line segment Hausdorff-distance uses the shortest line segment length, so that shorter lines segments reduce the error (see figure 9(c)). This behavior could be inappropriate in cases where the the length of the line segments covers a broad range.

In extreme case, this could lead to a matching of a very short line segment with a long line segment which is perpendicular to the short line segment. The opposite behavior occurs in the modified perpendicular line segment Hausdorff-distance, there the longest line impacts the score of the distance function (figure 9(d)). Even so, the differences of the line segments lengths are ignored.

We show for the modified Hausdorff distances only the angle plot, because there the length is independent of the translation. The Midpoint-distance and the Straight line-distance are both depending on the difference between the line segment lengths, which is shown in the Fig. 9(e) - 9(h). Table I gives an overview of the properties of the distance functions.

| | HD | TD | MHD | MPHD | MD | CD | SD |
|---|---|---|---|---|---|---|---|
| angle | $\oplus$ | $\ominus$ | $\oplus$ | $\oplus$ | $\oplus$ | $\ominus$ | $\oplus$ |
| translation | $\oplus$ | $\oplus$ | $\ominus$ | $\oplus$ | $\oplus$ | $\oplus$ | $\oplus$ |
| scale | $\odot$ | $\ominus$ | $\odot$ | $\odot$ | $\oplus$ | $\ominus$ | $\oplus$ |

TABLE I. WE PRESENT THE DEPENDENCIES OF THE DISTANCE FUNCTIONS TO ANGLE-, TRANSLATION- AND SCALE-/LENGTH-DIFFERENCES. THE SIGN $\oplus$ STANDS FOR DEPENDENT OF THIS QUANTITY, $\odot$ STANDS FOR DEPENDENT WITH LIMITATIONS AND $\ominus$ STANDS FOR INDEPENDENT OF THIS QUANTITY.

## V. CONCLUSION

We presented six well known distance functions for comparison of line segments and a new one which is variant to scale, translation and rotation. These functions were analytically evaluated by simulated data, so that we were able to stress out the properties of each distance functions concerning angle, translation, and scaling changes. The extracted properties make it now easy choosing an adequate distance function to a specific application.

## REFERENCES

[1] Y. X. Supervisor and M. Leung, "Hausdorff distance for shape matching," *The 4th LASTED International Conference on visualization image and image processing*, pp. 1–25, 2004.

[2] M.-Y. Liu, O. Tuzel, A. Veeraraghavan, and R. Chellappa, "Fast directional chamfer matching," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 2010, pp. 1696 –1703.

[3] J. Chen, M. K. Leung, and Y. Gao, "Noisy logo recognition using line segment hausdorff distance," *Pattern Recognition*, vol. 36, no. 4, pp. 943–955, 2003.

[4] C. Schmid and A. Zisserman, "Automatic line matching across views," in *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*. Washington, DC, USA: IEEE Computer Society, 1997, p. 666.

[5] T. Werner and A. Zisserman, "New techniques for automated architecture reconstruction from photographs," in *European Conference on Computer Vision*, vol. 2. Springer-Verlag, 2002.

[6] H. Bay, V. Ferrari, and L. J. V. Gool, "Wide-baseline stereo matching with line segments," in *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1*. Washington, DC, USA: IEEE Computer Society, 2005, pp. 329–336.

[7] S. Rusinkiewicz and M. Levoy, "Efficient variants of the icp algorithm," in *Third International Conference on 3D Digital Imaging and Modeling (3DIM)*, 2001.

[8] Sudarshan, "Segment-based map matching," in *Intelligent Vehicle, IEEE Symposium*, 2007, pp. 1190–1197.

[9] H. Alt, B. Behrends, and J. Blömer, "Approximate matching of polygonal shapes," *Annals of Mathematics and Artificial Intelligence*, vol. 13, no. 3–4, pp. 251–265, 1995, this is the full version of SoCG'91.

[10] E. Trucco and A. Verri, *Introductory Techniques for 3-D Computer Vision*. New York: Prentice Hall, 1998.

[11] J. Witt and U. Weltin, "Robust stereo visual odometry using iterative closest multiple lines," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE/RSJ, 2013.

Proceedings of the 9th Open German-Russian Workshop on Pattern Recognition and Image Understanding
Dec 1-5, 2014, Koblenz, Germany



# Experiments with automatic segmentation of liver parenchym using texture description

Miroslav Jiřík, Petr Neduchal

***Abstract*—This paper provides summary of our experiments with automatic segmentation of liver parenchym. It presents methods and classifiers that we used on computer tomography medicine data. In introduction there are a description of our motivation to do this research. Second part contains information about our approach, list of methods and classifiers. In part called results, we presents figure with subset of our experiments and described evaluation. Summary at the end of this paper presents future research of this topic.**

***Keywords*—liver parenchym, computer tomography, texture description, texture analysis**

## I. Introduction

COLORECTAL Carcynoma (CRC), also known as cancer of the gastrointestinal tract, is one of the most frequently diagnosted malignant tumour in the Western world. There are many factors causing CRC. Particularly life style and alimentation habits (alcohol, fried and grilled food etc.). Other factors are heredity and enviroment conditions. In many cases, CRC is often acompanied by metastasis in liver parenchym. If metastasis are resectable, the surgeon will do resection of liver tissue.

Before operation, its necessary to use modern imaging methods to judge, if it is possible to remove damaged liver tissue. The most common imaging technique is computer tomography (CT). In special cases it is followed by magnetic resonance (MRI). During operation, surgeon cuts out relatively big part of the liver parenchym. It is important to preserve function of the remaining part of patient liver. Statistics show that there are 25-40% of people who outlive longer than 5 years after resection and 20% of people who survive longer than 10 years after operation. Thanks to modern operating techniques, the postoperative mortality decreased from 15% to only 5% of patients.

As we mentioned in previous paragraph, it is necessary to do preoperative CT examination of liver parenchym. The next important step is marking liver area in CT data in order to compute volume of the liver and both parts after resection. There are some ratio of volume of healthy part of liver tissue to patient weight. The most common technique of marking liver in CT data is highlighting contours of liver slide by slide. This marking process is really time-consuming. It takes approximately 30 minutes. The question is, it is possible to do it automaticaly and significantly faster using computer?

Indications of dealing with this task are described in various papers. As in other scientific branches, there is a problem with comparing of different approaches to automatic marking of liver tissue. Within 3D Segmentation in the Clinic : A Grand Challenge there arised competition SLIVER07 and comparsion metodology to our task. Results of this competition is in [2]. Summary description of different approaches is in papers [4] and [1]. One of the most sucessful method is described in [3]. Paper [5] describes method which is based on segmentation of portal vein. Nowadays (2014) that method has the best score in SLIVER07. The dataset contains CT data and ground truth data (manually segmented) that can be used for evaluation.

M. Jiřík and P. Neduchal is with the Department of Cybernetics, University of West Bohemia, Pilsen, Czech Republic, e-mail: mjirik@kky.zcu.cz and neduchal@kky.zcu.cz

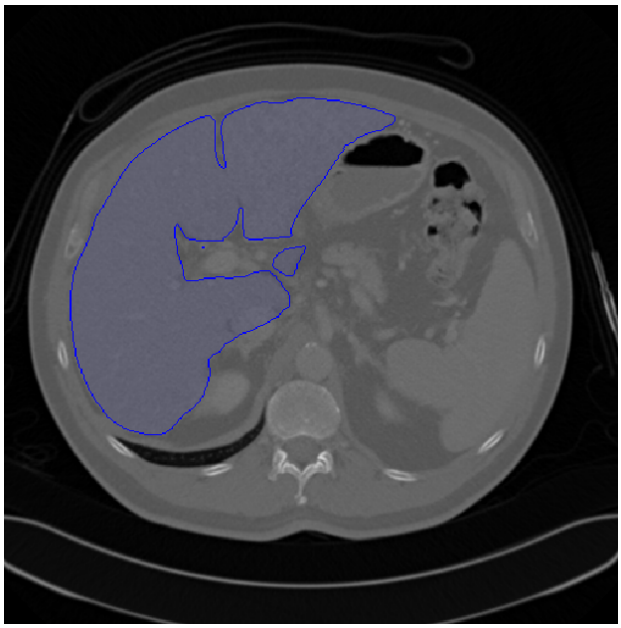


Fig. 1. Example of SLIVER07 computer tomography data. Blue region is manually segmented liver parenchym - i.e. ground truth for our experiments

## II. Methods and Classifiers

We tried couple well known texture description methods. Our approach contains of data decomposition on smaller tiles. Tile is a $X \times Y \times S$ block, where $X$ is width of tile, $Y$ is height and $S$ is number of data slices. We applied methods on them in order to get local description of the tile. This approach had some advantages and few disadvantages. The advantage is that it is easy to implementation. Disadvantage on the other hand is unknown ideal size of the tile. Too small tile is bad because of small amount of information and large tile is bad because the information hasnt local character. Unfortunately,

estimation of tile size is not simple task. Below, there are list of texture description methods which we used.

Each CT data in our dataset has different size of voxels. To suppress this issue a normalization algorithm was used. It is based on resampling input data to same voxel size. We used 1 mm for each axis.

### A. Texture descriptors

*1) Histogram method:* Simplest method of our approach. It creates histogram directly from data tile. Generated histogram is whole description of texture in local area which is defined by size of the tile.

*2) Gabor filters:* Principle of this method is creation of bank of 2D filters with different orientation. The core of Gabor filter is assembled of Gaussian function and modulated by sin function. Result is created as convolution of source data and filter itself. Each response contains information about occurence of wave with specific frequency omega and orientation sigma.

*3) Gray Level Co-occurence Matrix (GLCM):* Co-occurence matrix is square matrix of size NxN where N is number of gray levels. On the position [i,j] is information about number of co-occurence between gray level i and gray level j. There are also defined maximum distance of co-occurence and angle in which the method searches co-occurences.

$$\mathbf{Img} = \begin{bmatrix} 0 & 2 & 1 \\ 1 & 1 & 0 \\ 0 & 2 & 0 \end{bmatrix} \Rightarrow \mathbf{P}(0,1) = \begin{bmatrix} 0 & 1 & 3 \\ 1 & 2 & 1 \\ 3 & 1 & 0 \end{bmatrix}$$

where $\mathbf{Img}$ is source data and $\mathbf{P}(0,1)$ is co-occurence matrix for angle $0$ and distance $1$.

*4) Local Binary Patterns (LBP):* Local Binary pattern method transforms all values in neighbourhood of center pixel to one binary number. That number defines texture of whole local area. Approach is shown on the example:

$$\mathbf{G} = \begin{bmatrix} g_1 & g_2 & g_3 \\ g_8 & g_0 & g_4 \\ g_7 & g_6 & g_5 \end{bmatrix} \Rightarrow \begin{bmatrix} 0 & 3 & 1 \\ 7 & 2 & 1 \\ 9 & 1 & 4 \end{bmatrix} \Rightarrow$$

$$\sum_{i=1}^{n} sg(g_i - g_0) \cdot 2^{n-1} \Rightarrow b = [11010010] \Rightarrow 210,$$

where sg is 1 if $g_i < g_0$ and 0 otherwise, $n$ is number of pixels $g_i$ in neighbourhood of center pixel $g_0$.

At the end we get LBP numbers for all local neighbourhoods in tile. We are able to create histogram that contains all information about texture of the tile. There are a lot of improvements of LBP method, but we used basic algorithm we dont have to deal with rotation in CT data.

Histogram method and LBP are neighbourhood independent methods. i.e. It is possible to use them on N-dimensional data. On the other hand, the GLCM and gabor filters can be used only on 2-dimensional data - i.e it is necessarz to do decomposition of tile to slices.

We implemented histogram features and LBP (based on [7]). Scikits-image implementation of GLCM and Gabor Filters was used [6].

### B. Classifiers

In our experiments we worked with seven different classifiers. Naive Bayes, Support Vector Machine (SVM), Gaussian Mixture Model (GMM), Decision Trees (DT), Random Forest, Quadratic discriminant analysis (QDA) and Linear discriminant analysis (LDA). Python module Scikits-learn implementation of these algorithms was used. Results

## III. Experiments

Our experiments are composed of three parts. In first step, the appliaction loads first half of SLIVER07 dataset and makes smaller tiles. In second part, script applied one (or more) of texture description methods on each tile and trains one of classifiers listed above. Ground truth from dataset is used as supervisor information that decides whether pixel belongs to liver or to non-liver region. Third part consist of applying method on test data and classifying acquired result.

After classification we have marked each pixel by number 1 or 0. Number 1 is pixel classified as liver and 0 is non-liver pixel. The accuracy of our experiments is limited by siye of the tile. The best scenario would be tile of size $1 \times 1 \times 1$. Unfortunately, it is impossible to applied texture description methods on the single pixel. This is the reason, why we never get absolutely accurate results.


Fig. 2. LIver Surgery Analyser (Lisa). Software package developed by M. Jiřík et al. We used Lisa to our experiments.

## IV. Results

As we mentioned in previous section, we did set of segmentation experiments of liver tissue. We combined texture description methods with classifiers. For our experiments we used Sliver07 training dataset. One half of twenty CT images was used for training classifier with texture descriptor. After that, remaining data was used for evaluating experiment.

Evaluation is based on five diferent metrics: Volumetric overlap error (in percent), Relative absolute volume difference (in percent), Average symmetric surface distance (in milimeters), Root Mean Square symmetruc surface distance

(in milimeters) and Maximum symetric surface distance (in milimeters). These metric are rescaled to score from 0 (negative values are aligned to zero) to 100. Total score is the average of the individual scores. Complete description of methodology can be found in [2].

In the figure 3 you can see scores of individual combinations of descriptor and classifier. From total of 44 experiments a subset with significant score is selected.

We used acronyms to describe each combination of method and classifier. First part of acronym is the label of method: fh - feature histogram, glcm - gray level co-ocurence matrix, gb - gabor filter and lbp - local binary patterns. In some cases there are multiple methods used. The second part is the label of classifier: dt - decision tree, g - naive Bayes classifier and lda - LDA. Last part of acronym is the information about voxel size normalization. Column marked as h+glcm_dt_n has the best performance. Feature vector is a combination of histogram method and gray level co-ocurence matrix. As classifier we used decision tree classifier in this case.

Generally the decision tree classifier was better than other classifiers used in our experiments. The LDA classifier had good results too. On the other side of score table were SVM and naive Bayes classifier, which had very poor results.

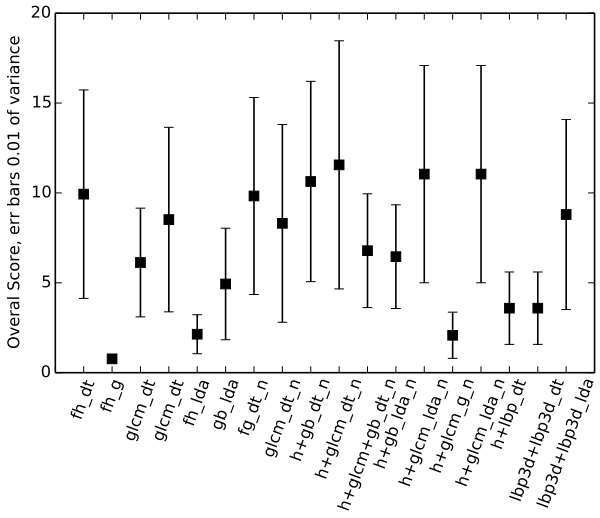



Fig. 3. Subset of experimental results. Acronyms are described below.

## V. Conclusion

As you can see in previous section, results are dependent on texture description method, classifier and amount of noise in processed data. Because of that, there are big difference between various combinations. Texture description itself is not the best way to classify tiles in noisy CT data.

There are a lot of opportunities to improve results by trying new combinations of descriptor and classifier, trying different type of tiles - i.e. overlapping tiles - or adding some kind of support information about processed data.

Other approach in this task might be description vector created by some feature detection method as SIFT, SURF, etc. instead of texture descriptor. Neural network classification could be interesting to. We want to try several ways in our future research of this topic.

## Acknowledgment

The work has been supported by the grant of The University of West Bohemia, project number SGS-2013-032 and GRANT IGA MZ R 13326 2012-2015 and postdoctoral project LFP04 CZ.1.07/2.3.00/30.0061

## References

[1] S. Luo, X. Li andJ. Li, *Review on the Methods of Automatic Liver Segmentation from Abdominal Images* Journal of Computer and Communications, 2014(January), 17

[2] T. Heimann, B. van Ginneken, M. Styner, Y. Arzhaeva, V. Aurich, C. Bauer, I. Wolf *Comparison and evaluation of methods for liver segmentation from CT datasets* IEEE Transactions on Medical Imaging, 28(8),2009, 125165

[3] L. Rusko and G. Bekes, *Fully automatic liver segmentation for contrast-enhanced CT images* Segmentation in the Clinic, 18.

[4] A. M. Mharib, A. R. Ramli, S. Mashohor, R. B. Mahmood, *Survey on liver CT image segmentation methods* Artificial Intelligence Review, 2011, 37(2), 8395.

[5] A. S. Maklad, M. Masuhiro, H. Suzuki, Y. Kawata, H. Niki, N. Moriyama, M. Shimada, *Blood vessel-based liver segmentation through the portal phase of a CT dataset* In SPIE Medical Imaging (p. 86700X86700X), 2013.

[6] S. van der Walt, J. L. Schnberger, J. Nunez-Inglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu and scikit-image contributors, *Scikit-image: Image processing in Python* PeerJ 2:e453 (2014)

[7] T. Mäenpää, M. Turtinen, M.Pietikäinen, *Real-Time Surface Inspection by Texture* Real-Time Imaging Volume, Issue 5 (2003), 289-296

**Miroslav Jiřík** was born in Klatovy, Czech Republic in 1984. He received his Bc. and Ing. (similar to M.S.) degrees in cybernetics from the University of West Bohemia, Pilsen, Czech Republic (UWB) , in 2006 and 2008 respectively. As a Ph.D. candidate at the Department of Cybernetics, UWB his main research interests include computer vision, machine learning, medical imaging, image segmentation, texture analysis. He is a teaching assistant at the Department of Cybernetics, UWB.

**Petr Neduchal** was born in Rokycany, Czech Republic in 1989. He received his Bc. and Ing. (similar to M.S.) degrees in cybernetics at University of West Bohemia, Pilsen, Czech Republic (UWB) , in 2011 and 2013 respectively. As a Ph.D. candidate at the Department of Cybernetics, UWB his main research interests including computer vision, estimation theory, simultaneous localization and mapping, medical imaging and thermography.

# Fast implementation of the Niblack binarization algorithm for microscope image segmentation of cell cultures infected with Chlamydia

Artyukhova O.A., Samorodov A.V.
Chair for Biomedical Technical Systems
Bauman Moscow State Technical University
Moscow, Russian Federation
artyukhova@bmstu.ru, avs@bmstu.ru

***Abstract*—An algorithm for the segmentation of cells and Chlamydial inclusions on microscope images, containing the steps for color deconvolution and adaptive local binarization is considered. A fast way to implement the Niblack binarization algorithm is described. It uses not only the integral image for the local mean values calculation, but also the second order integral image for the local variance calculation. Following the proposed approach the time of segmentation has been significantly reduced providing the possibility of its use in practice. The generalization of integral image representation, called '*k*-order integral image' could be used for fast calculation of higher order local statistics.**

***Keywords—k-order integral image; fast algorithm; Niblack binarization; segmentation; Chlamydial inclusions analysis***

## I. INTRODUCTION

Chlamydial infection is an urgent medical and social problem of our time. Nowadays more than 100 million new cases of the disease are registered around the world every year [1]. Cultural method is the "gold standard" of Chlamydial infection diagnostics, the reference method for assessing the effectiveness of antibiotic treatment, and the only method allowing analysis of Chlamydia resistance to antibiotics in the task of therapy selection and the development of new antibacterial agents [2]. This method consists of preparing and microscopic study of eukaryotic cell cultures infected with Chlamydia. Currently such investigation is carried out visually using a fluorescence microscope. It comprises the following parts: (1) determination of Chlamydial inclusions presence or absence in the cells, (2) counting the number of inclusion-forming units per unit volume of the sample by means of calculating the proportion of infected cells in the preparation, (3) qualitative evaluation of the Chlamydial inclusions size.

Visual examination has several significant disadvantages such as adverse effects on the researcher's eyes, high labor intensity, long analysis time of one preparation, a small amount of cells analyzed visually and, ultimately, low reliability of analysis results. Development of automated methods for microscope image segmentation and analysis allows to increase the reliability of the obtained results, reduce the time of research, expand the scope of its application.

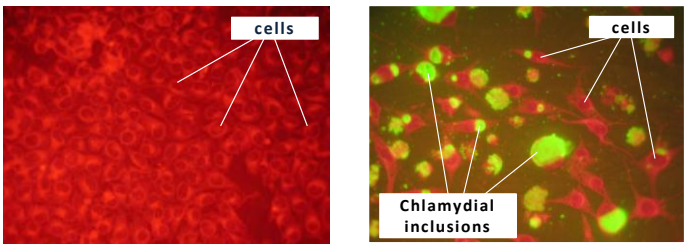
## II. ALGORITHM DESCRIPTION

The procedure of the specimen preparation lies in infecting the cell culture (fibroblasts) with infective suspension and its incubation, then fixation and staining of the obtained specimen with a fluorescent dye FITC and Evans blue. The obtained specimens are analyzed using a fluorescence microscope. Fig. 1 shows the fluorescence microscope image of a pure cell culture and of cell culture infected with Chlamydia. In the fluorescence images Chlamydial inclusions appear bright green on the reddish-brown background of stained cells.

The segmentation of the images under consideration is the most important stage of their analysis, and its aim is to allocate the areas of cells, Chlamydial inclusions and background. The absence of quantitative biomedical studies so far makes it impossible to set substantiated requirements for segmentation accuracy beforehand.

For the segmentation of color images the methods of thresholding, clustering, region growing, construction of physical models of imaging and some others are commonly used. Each of these methods can use different color models.

To select a color representation model for the considered images providing their easiest and highest quality segmentation, a quantitative analysis of color differences between Chlamydial inclusions, cells and background was carried out in channels of the following color spaces: RGB, CMY, HSV, Lab, YCbCr, YIQ.




(a) (b)

Fig. 1. Fluorescence microscopy images of (a) – pure cell culture, (b) – cell culture infected with Chlamydia.

Also, FITC and Evans blue concentration profiles obtained as the result of color deconvolution method proposed in [3] and adapted for fluorescence microscopy in [4] were used as the color channels.

To quantify the comparison of color channels we used equal error rate (EER) of the pixel-wise threshold classification. On 20 microscopic images three classes of regions: Chlamydial inclusions, cells and background – were labeled manually. EER was calculated for two classification problems: "cells and Chlamydial inclusions against the background" ($P_1$) and "Chlamydial inclusions against the cells and the background" ($P_2$). The results are shown in Table I.

The FITC concentration profile allows to significantly reduce the EER of pixels belonging to Chlamydial inclusions threshold classification (to an average of 3%), but Evans blue concentration profile does not provide a similar result in the classification of pixels belonging to cells. This is consistent with the peculiarities of the specimens preparation, as Evans blue is often associated not only with cellular structures, but also with the extracellular matrix. Thus, for cells segmentation it is expedient to use channel V of the color space HSV, since it has a minimum EER value.

Despite the small EER value for pixels of Chlamydial inclusions classification, obtained for images used, often during the preparation of cell culture specimens the background fluorescence caused by the presence of a fluorescent dye not only in the microbiological organisms being studied, but in general on the entire surface of the preparation, is not completely suppressed. For this reason, for the segmentation of cells, as well as of Chlamydial inclusions in the selected color channels it is expedient to use local adaptive binarization methods, the most famous of which is the Niblack method.

TABLE I. RESULTS OF DIFFERENT COLOR CHANNELS COMPARISON

| Color space | Color channel | EER $P_1$ | EER $P_2$ |
|---|---|---|---|
| RGB | R | 0.27 | 0.43 |
| | G | 0.48 | 0.11 |
| | B | 0.31 | 0.18 |
| CMY | C | 0.28 | 0.45 |
| | M | 0.37 | 0.17 |
| | Y | 0.44 | 0.16 |
| HSV | H | 0.36 | 0.17 |
| | S | 0.48 | 0.13 |
| | V | 0.25 | 0.34 |
| Lab | L | 0.28 | 0.17 |
| | a | 0.31 | 0.23 |
| | b | 0.32 | 0.16 |
| YCbCr | Y | 0.30 | 0.15 |
| | Cb | 0.39 | 0.11 |
| | Cr | 0.30 | 0.29 |
| YIQ | Y | 0.30 | 0.15 |
| | I | 0.29 | 0.40 |
| | Q | 0.30 | 0.14 |
| The results of color deconvolution | FITC concentration profile | 0.48 | 0.03 |
| | Evans blue concentration profile | 0.34 | 0.18 |

The idea of Niblack method is to calculate the binarization threshold $T(x, y)$ by local mean $m(x, y)$ and standard deviation $s(x, y)$ of the neighbouring pixels intensity values:

$$T(x, y) = m(x, y) + k \cdot s(x, y), \quad (1)$$

where $k$ is the coefficient, determined experimentally. For $k$=0 only local average is used for calculation of the threshold value.

The parameters of binarization algorithm that provide the best quality of segmentation were determined experimentally. The size of a square window used for the calculation of threshold value is equal to 700 pixels for the segmentation of cells and 500 pixels for the segmentation of Chlamydial inclusions for images captured with pixel size 0.14 μm (image size 3072×2304 pixels). This corresponds to three times the average diameter of these objects. Coefficient $k$ in (1) is equal to –1 and 2 for segmentation of cells and Chlamydial inclusions respectively.

Image segmentation stages are shown in Fig. 2.

## III. SECOND ORDER INTEGRAL IMAGE REPRESENTATION

The significant size of the window used for the calculation of local statistics leads to significant time loss for the direct implementation of the Niblack algorithm and the impossibility of its use in practice.

The problem of speeding up the local statistics calculation is not new. In [5] a box-filtering technique was proposed for fast calculation of image local mean value. In [6] another idea of fast local mean value calculation was published, and later it was popularized under the name of integral image representation in [7].

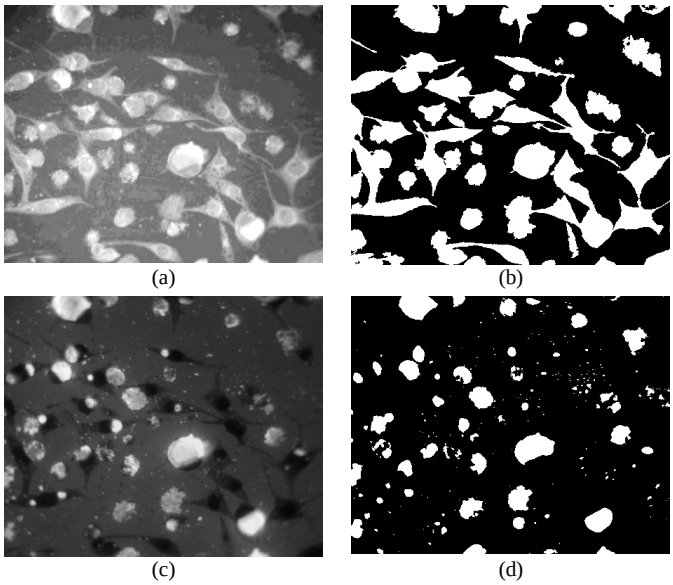



Fig. 2. The stages of image segmentation: (a) – grayscale image in the color channel V of HSV color space, (b) – the result of cells segmentation, (c) – FITC concentration profile obtained by color deconvolution, (d) – the result of Chlamydial inclusions segmentation.

In [8] the box-filtering technique was extended for fast calculation of local statistics of different orders for *N*-dimensional images. Box-filtering technique compared to integral image representation is less memory consuming, but is dependent on the size of local window which is intended to be used for local statistics calculation. If local statistics with different local window sizes have to be calculated for one image, the box-filtering technique seems not to be the best choice, regarding the calculations time.

Absence of simple and fast algorithm for local standard deviation calculation resulted, e.g., in attempts to replace local standard deviation in Niblack algorithm by local intensity mean deviation [9]. Here we continue to develop a simple and fast approach for local standard deviation calculation based on the expansion of integral image representation, initially presented in [4] under the name of 'integral squared image'.

Integral image $I(x_0, y_0)$ is calculated from the initial image $f(x, y)$ as follows [7]:

$$I(x_0, y_0) = \sum_{x=1}^{x_0} \sum_{y=1}^{y_0} f(x, y). \quad (2)$$

With integral image the local mean value inside rectangle area, multiplied by the number of area pixels, can be calculated using only tree addition operations:

$$\begin{aligned} m(x_0, y_0) \cdot N = I(x_0 + \Delta x, y_0 + \Delta y) + \\ + I(x_0 - \Delta x - 1, y_0 - \Delta y - 1) - \\ - I(x_0 - \Delta x - 1, y_0 + \Delta y) - I(x_0 + \Delta x, y_0 - \Delta y - 1), \end{aligned} \quad (3)$$

where $m(x_0, y_0)$ – local mean value inside rectangle area with center coordinates $(x_0, y_0)$, upper left and the lower right corners coordinates $(x_0-\Delta x, y_0-\Delta y)$ and $(x_0+\Delta x, y_0+\Delta y)$ respectively,

$N = (2 \cdot \Delta x+1) \cdot (2 \cdot \Delta y+1)$ – the number of rectangle area pixels.

The generalization of integral image representation is *k*-order integral image:

$$I_k(x_0, y_0) = \sum_{x=1}^{x_0} \sum_{y=1}^{y_0} \left[ f(x, y) \right]^k. \quad (4)$$

The representation (4) can be effectively used for the calculation of *k*-order local statistics in the same manner as initial integral image.

The main time loss during Niblack binarization is associated with the calculation of the local standard deviation of image pixel intensities. To accelerate these computations a second order integral image representation $I_2(x_0, y_0)$ was used:

$$I_2(x_0, y_0) = \sum_{x=1}^{x_0} \sum_{y=1}^{y_0} \left[ f(x, y) \right]^2. \quad (5)$$

Considering the well-known equation in statistics where the variance is calculated as the difference between the mean square value and squared mean value, local variance $s^2(x_0, y_0)$, multiplied by $(N-1)$, could be calculated using second order integral image representation (5) as follows:

$$\begin{aligned} s^2(x_0, y_0) \cdot (N-1) = I_2(x_0 + \Delta x, y_0 + \Delta y) + \\ + I_2(x_0 - \Delta x - 1, y_0 - \Delta y - 1) - \\ - I_2(x_0 - \Delta x - 1, y_0 + \Delta y) - \\ - I_2(x_0 + \Delta x, y_0 - \Delta y - 1) - \frac{1}{N} \cdot \left[ m(x_0, y_0) \cdot N \right]^2. \end{aligned} \quad (6)$$

In Table II computation complexity of direct calculation of local standard deviation $s(x_0, y_0)$ for an image of size $N_1 \times N_2$, and of its calculation using the proposed integral representation is given. It could be clearly seen that the second order integral image representation provides significant reduction in calculation complexity and independence on the local window area *N*.

To calculate the Niblack local binarization threshold for one image pixel by direct calculation of the local mean and standard deviation values inside local window with area *N*, approximately $(3 \cdot N-1)$ addition operations, $(N+3)$ multiplication and squaring operations, and one square rooting are required. The use of integral representations (2) and (5) provides the independence from local window area and reduction in the number of mathematical operations to about twelve additions, four multiplications, two squaring and one square rooting per each local threshold value calculation.

## IV. Experimental Results

The use of second order integral image representation in Niblack algorithm provided the reduction of the average time of one image automatic analysis from few days (when direct calculation of the local image statistics was used) to 9 seconds (on computer with processor Intel Core i5, 2.27 GHz, RAM 4.00 GB and 64-bit OS). This provides the possibility to use the proposed segmentation algorithm in practice.

An example of the segmented image of cell cultures infected with Chlamydia is presented in Fig. 3. The assessment of the proposed segmentation algorithm was performed using fluorescence microscope images provided by the Laboratory of Chlamydial infections of Gamaleya R&D Institute of Epidemiology and Microbiology. 50 images of Chlamydial inclusions with substantially different sizes, and 30 images of cell cultures each containing about 200 cells, were segmented both manually and automatically.

TABLE II. Computational Complexity of Calculation of the Standard Deviation Directly and Using the Supposed Integral Representation

| Operation | Number of operations | |
|---|---|---|
| | ***By direct calculation*** | ***Using integral representation $I_2$*** |
| addition / subtraction | $1+(2 \cdot N-1) \cdot N_1 \cdot N_2$ | $1+3 \cdot N_1 \cdot N_2-N_1-N_2$ |
| multiplication / squaring | $1+(N+1) \cdot N_1 \cdot N_2$ | $1+4 \cdot N_1 \cdot N_2$ |
| division | 1 | 1 |
| square rooting | $N_1 \cdot N_2$ | $N_1 \cdot N_2$ |

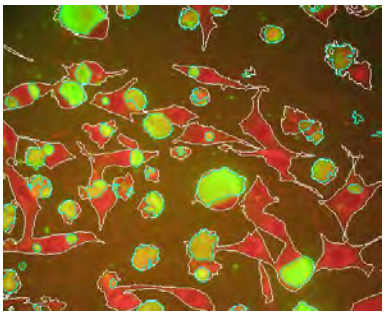


Fig. 3. An example of segmentation results, obtained by the suggested algorithm (white contours define the boundaries of cells and the cyan ones – the boundaries of Chlamydial inclusions).

The quality of the segmentation of cells and Chlamydial inclusions was assessed by the relative error of their areas determination. These errors for cells and Chlamydial inclusions did not exceed 5.2 % and 3.0 % respectively. The proposed segmentation algorithm enables for the first time to carry out quantitative microscope studies of cells infected with Chlamydia (see, e.g., [10]), in which substantiated requirements for the segmentation error of the considered images will also be established, allowing to confirm adequacy of the proposed segmentation algorithm or to make a reasonable decision on the need of its improvement.

## V. Conclusion

This study presents the fast Niblack binarization algorithm and its implementation for microscope image segmentation of cell cultures infected with Chlamydia. The expanded integral image approach provides the significant reduction in segmentation time which opens the possibility of practical use of the developed automatic segmentation algorithm. Its areas of possible application are the quantifying of morphological parameters of Chlamydial inclusions as well as the degree of cell culture infection, which is an essential step in the analysis of new antibacterial agents.

The proposed *k*-order integral image representation is the generalization of well-known integral image representation. The use of the second order integral image provides fast variance (and standard deviation) calculation. Higher order local statistics can be further calculated by means of higher order integral image representations.

## References

[1] World Health Organization, “Global strategy for the prevention and control of sexually transmitted infections: 2006 – 2015,” 2007.

[2] E.V. Shipitsina, A.M. Savicheva, Савичева А.М., T.A. Khusnutdinova et al., “Resistance of Chlamydia trachomatis to antibiotics in vitro: methodological aspects and clinical significance,” Clinical Microbiology and Antimicrobial Therapy, vol. 6, no. 1, pp. 54-64, January 2004, in Russian.

[3] A.C. Ruifrok and D.A. Johnston, “Quantification of histochemical staining by color deconvolution,” Anal Quant Cytol Histol, vol. 23, no. 4, pp 291-299, August 2001.

[4] O.A. Artyukhova and A.V. Samorodov, “Elaboration of an automatic segmentation algorithm for fluorescence microscopic images of cell cultures preparations for the problems of microbiology,” Science and education: the electronic scientific and technical publishing, no. 6, June 2013, URL: http://technomag.edu.ru/doc/574140.html, in Russian.

[5] M.J. McDonnell, “Box-filtering techniques,” Computer graphics and image processing, vol. 17, pp. 65-70, September 1981.

[6] F.C. Crow, “Summed-area tables for texture mapping,” Computer graphics, vol. 18, no. 3, pp. 207-212, July 1984.

[7] P. Viola and M.J. Jones, “Robust real-time face detection,” International Journal of Computer Vision, vol. 57, no. 2, pp. 137-154, May 2004.

[8] C. Sun, “Fast algorithm for local statistics calculation for N-dimensional images,” Real-time imaging, vol. 7, pp. 519-527, December 2001.

[9] T.Romen Singh, Sudipta Roy, O.Imocha Singh, Tejmani Sinam, Kh.Manglem Singh, “A new local adaptive thresholding technique in binarization,” International journal of computer science issues, vol. 8, issue 6, no. 2, pp. 271-277, November 2011.

[10] E.A. Kost, “The effect of an inhibitor of type III secretion system on the development of experimental infection caused by Chlamydia trachomatis,” PhD thesis. Moscow, 122 p., 2014, in Russian.

# Fast Model Based Object Recognition and Pose Estimation Using Local Deviation Grids

**ABSTRACT**

Benjamin Hohnhäuser Stephan Brodkorb André Moltmann Frank Püschel

Gesellschaft zur Förderung angewandter Informatik
Volmerstraße 3
12489 Berlin, Germany

The fast automatic object recognition and pose estimation of work pieces from 3D point clouds is production processes a tough challenge, especially under typical industrial production conditions with a high type variety. Even more if vast variety of nearly equal objects of different types has to be distinguished and identified in very short time. A reliable discrimination, classification, and pose estimation with the application of classical 3D matching algorithms (e.g. ICP) is not a promising approach at all. Furthermore it should easily be possible to add new work pieces to the data base in absence of a proper CAD model.

A typical application field for such algorithms are recognition systems for automated loading of robot driven glazing lines in sanitary ceramic production. In this context misrecognition has to be avoided.

1Figure 2: Workpiece with laser lines in the recognition stage

The data acquisition for the detection process of the ceramic work pieces is achieved using a monocular camera system in combination with laser plane (see figure 1), which yields 3D data by triangulation. By rotating the work piece with an appropriate actuator the data points can be transformed into an object describing point cloud (see figure 2).After the preprocessing stage and the automatic removal of artefacts the recognition and pose estimation is performed. The process is divided in several stages. First by means of moments and the projection of the measurement points into a plane the major axis are determined. The approximated surface area, the height, and the area of projections a pre-classification is performed by using the volume point cloud. In a further stage the object is classified by means of local distances of certain surface points to sparse 3D-grids that are calculated from prototype models. The distance in feature space is calculated by da distance function which yields the likeliness, respectively. The best matching model with highest likeliness level is chosen but only if a parameterized threshold is exceeded. Otherwise it is rejected. The

matching process directly provides the pose of the object if it is recognized. The discriminatory power and the rejection rate can be adapted by parametrizable thresholds.

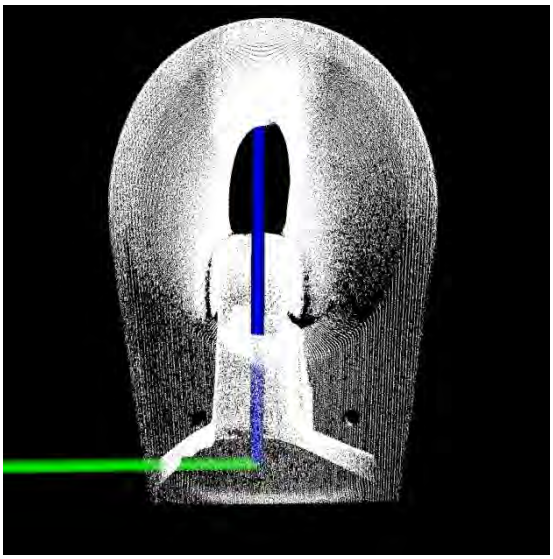


Figure 2: Point cloud of a scanned work piece

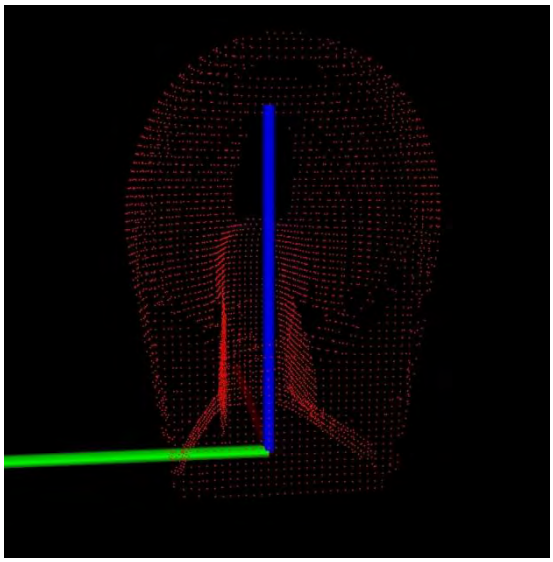


Figure 3: Distance grid points of the work piece

The developed system is already used in the productive manufacturing. At the moment 150 objects with differences that are barely visible with naked eyes can be reliably distinguished. The time for pose estimation and object recognition is less than 1 second. The processed point clouds have between 5 and 10 Million measurement points.

# Filters Based On Aggregation Operators

Labunets V.
[1]*Ural Federal University*
[2]*Samara State Aerospace University*,
[1]*Yekaterinburg*, [2]*Samara, Russia*,
*vlabunets05@yahoo.com*

Osthaimer E.
*Capricat LLC*
*Pompano Beach, Florida, USA*
katya@capricat.com

*Abstract*— **In this work we introduce and analyze a new class of nonlinear filter which have their roots in aggregation operators theory. We show that a large body of non-linear filters proposed to date constitute a proper subset of aggregation filters.**

***Keywords— nonlinear filtering, multicolor images, fggregation operators***

## I. INTRODUCTION

The basic idea behind this paper is the estimation of the uncorrupted image from the distorted or noisy image, and is also referred to as image "denoising". To denoise images is to filter out the noise. The challenge is to preserve and enhance important features during the denoising process. For images, for example, an edge is one of the most universal and crucial features There are various methods to help restore an image from noisy distortions. Each technique has its advantages and disadvantages Selecting the appropriate method plays a major role in getting the desired image. Noise removal or noise reduction can be done on an image by linear or nonlinear filtering. The more popular linear technique is based on average (on mean) linear operators. Denoising via linear filters normally does not perform satisfactorily since both noise and edges contain high frequencies. Therefore, any practical denoising model has to be nonlinear. In this paper, we propose a new type of nonlinear data-dependent denoising filter called the *aggregation digital filter (ADF)*.

Let us introduce the observation model and notion used throughout the paper. We consider noise images of the $\vec{\mathbf{f}}(\mathbf{x})=\vec{\mathbf{s}}(\mathbf{x})+\vec{\boldsymbol{\eta}}(\mathbf{x})$, where $\vec{\mathbf{s}}(\mathbf{x})=\left(s_1(\mathbf{x}),s_2(\mathbf{x}),...,s_K(\mathbf{x})\right)$, is the original multichannel signal, $\vec{\boldsymbol{\eta}}(\mathbf{x})=\left(\eta_1(\mathbf{x}),\eta_2(\mathbf{x}),...,\eta_K(\mathbf{x})\right)$ denotes the multichannel noise introduced into the signal $\vec{\mathbf{s}}(\mathbf{x})$ to produce the corrupted image $\mathbf{f}(\mathbf{x})=\left(f_1(\mathbf{x}),f_2(\mathbf{x}),\ldots,f_K(\mathbf{x})\right)$ and $\mathbf{x}=(i,j)\in\mathbf{Z}^2$ (or $\mathbf{x}=(i,j,k)\in\mathbf{Z}^3$) are a 2D (or 3D) coordinates that belong to the image domain and represent the pixel location. If $\mathbf{x}\in\mathbf{Z}^2,\mathbf{Z}^3$ then $\vec{\mathbf{f}}(\mathbf{x})$, $\vec{\mathbf{s}}(\mathbf{x})$, $\vec{\boldsymbol{\eta}}(\mathbf{x})$ are 2D and 3D multichannel images, respectively.

The aim of image enhancement is to reduce the noise as much as possible or to find a method which, given $\vec{\mathbf{s}}(\mathbf{x})$, derives an image $\hat{\vec{\mathbf{s}}}(\mathbf{x})$ as close as possible to the original $\vec{\mathbf{s}}(\mathbf{x})$, subject to a suitable optimality criterion [1].

In a 2D standard linear and median filters with a square window $\left[\mathrm{M}_{(i,j)}(m,n)\right]_{m=-r,n=-r}^{m=+r,n=+r}$ of size $(2r+1)\times(2r+1)$ is located at $(i,j)$ the arithmetic mean and median replace the central pixel

$$\hat{\vec{\mathbf{s}}}(i,j)=\underset{(m,n)\in M_{(i,j)}}{\mathbf{Arithm}}\left\{\vec{\mathbf{f}}(m,n)\right\},\ \hat{\vec{\mathbf{s}}}(i,j)=\underset{(m,n)\in M_{(i,j)}}{\mathbf{Med}}\left\{\vec{\mathbf{f}}(m,n)\right\},$$

where $\hat{\vec{\mathbf{s}}}(i,j)$ is the filtered image, $\left\{\vec{\mathbf{f}}(m,n)\right\}_{(m,n)\in M_{(i,j)}}$ is image block of the fixed size $(2r+1)\times(2r+1)$ extracted from $\vec{\mathbf{f}}$ by moving window $\mathrm{M}_{(i,j)}$ at the position $(i,j)$. Symbols **Arithm** and **Med** are the arithmetic mean (average) and median operators, respectively. When those filters are modified as follows

$$\hat{\vec{\mathbf{s}}}(i,j)=\underset{(k,l)\in M(i,j)}{\mathbf{Agg}}\left\{\vec{\mathbf{f}}(k,l)\right\},\qquad(1)$$

it becomes an aggregation digital filter, where **Agg** is a *generalized average* or an *aggregation operator* [2,3].

## II. AGGREGATION OPERATORS

The aggregation problem consist in aggregating *n*-tuples of objects $(x_1,x_2,...,x_N)$ all belonging to a given set $D$, into a single object of the same set $D$, i.e., **Agg**: $D^N\to D$. In fuzzy logic theory the set $D$ is an interval of the real $D=[0,1]\subset\mathbf{R}$. In image processing theory $D=[0,255]\subset\mathbf{Z}$. In this setting, an aggregation оператор is simply a function, which assigns a number $y$ to any $N$-tuple $(x_1,x_2,...,x_N)$ of numbers that satisfies:

1) $\mathbf{Agg}_N(x_1,x_2,...,x_N)$ is continuous and monotone in each variable; to be definite, we assume that $\mathbf{Agg}_N$ is increasing in each variable.
2) The aggregation of identical numbers is equal to their common value: $\mathbf{Agg}_N(x,x,...,x)=x$.
3) $\mathbf{min}(x_1,...,x_N)\le\mathbf{Agg}(x_1,...,x_N)\le\mathbf{max}(x_1,...,x_N)$. Here $\mathbf{min}(x_1,x_2,...,x_N)$ and $\mathbf{max}(x_1,x_2,...,x_N)$ are the *minimum*

and the *maximum* values among the elements of $(x_1, x_2, ..., x_N)$.

4) $\mathbf{Agg}(x_1, x_2, ..., x_N)$ is a symmetric function:

$$\mathbf{Agg}\left(x_{\sigma(1)}, x_{\sigma(2)}, ..., x_{\sigma(N)}\right) = \mathbf{Agg}\left(x_1, x_2, ..., x_N\right),$$

$\forall \sigma \in \mathbf{S}_N$ of $\{1, 2, ..., N\}$, where $\mathbf{S}_N$ is the set of all permutations of $1, 2, ..., N$. In this case $\mathbf{Agg}(x_1, ..., x_N)$ is invariant (symmetric) with respect to the permutations of the elements of $(x_1, x_2, ..., x_N)$. In other words, as far as means are concerned, the *order* of the elements of $(x_1, x_2, ..., x_N)$ is - and must be - completely irrelevant.

**Proposition 1.** (Kolmogorov). If conditions 1)–4) are satisfied, the aggregation $\mathbf{Agg}_N(x_1, x_2, ..., x_N)$ of the average type is of the form:

$$\mathbf{Kolm}\left(K \mid x_1, x_2, ..., x_N\right) = K^{-1}\left[\frac{1}{N}\sum_{i=1}^{N} K\left(x_i\right)\right],$$

where $K$ is a strictly monotone continuous function in the extended real line [4].

We list below a few particular cases of means:

1) Arithmetic mean ( $K(x) = x$ ):

$$\mathbf{Arithm}\left(x_1, x_2, ..., x_N\right) = \frac{1}{N}\sum_{i=1}^{N} x_i,$$

2) Geometric mean ( $K(x) = \ln(x)$ ):

$$\mathbf{Geo}\left(x_1, x_2, ..., x_N\right) = \sqrt[N]{\prod_{i=1}^{N} x_i} = \exp\left(\frac{1}{n}\sum_{i=1}^{N} \ln x_i\right).$$

3) Harmonic mean ( $K(x) = x^{-1}$ ):

$$\mathbf{Harm}(x_1, x_2, ..., x_N) = \left(\frac{1}{N}\sum_{i=1}^{N} \frac{1}{x_i}\right)^{-1}.$$

4) A very notable particular case corresponds to the function $K(x) = x^p$. We obtain then power mean:

$$\mathbf{Power}_p\left(x_1, x_2, ..., x_N\right) = \left(\frac{1}{N}\sum_{i=1}^{N} x_i^p\right)^{\frac{1}{p}}.$$

In mathematics, the power mean, also known as Hölder mean (named after Otto Holder), is an abstraction of the Pythagorean means including arithmetic, geometric, and harmonic means.

## III. Generalized Average Aggregation Filters

### A. Kolmogorov filters

The simplest aggregation filters associated with arithmetic means are ordinary linear filters:

$$\hat{s}(i, j) = \underset{(m,n)\in M_{(i,j)}}{\mathbf{Arithm}}\left\{f(m,n)\right\},$$

$$\hat{s}(i, j) = \underset{(m,n)\in M_{(i,j)}}{\mathbf{Arithm}}\left\{\overline{w}(m,n)\, f(m,n)\right\}.$$

Many extensions of the simple ordinary linear filters are defined as Kolmogorov filters:

$$\hat{s}(i, j) = \underset{(m,n)\in M_{(i,j)}}{\mathbf{Kolm}}\left\{K \mid f(m,n)\right\} =$$

$$= K^{-1}\left[\frac{1}{N}\sum_{(m,n)\in M_{(i,j)}}^{N} K\left(f(m,n)\right)\right].$$

Using Hölder means we obtain the Hölder (or power) filters of the form:

$$\hat{s}(i, j) = \underset{(m,n)\in M_{(i,j)}}{\mathbf{Hold}^p}\left\{f(m,n)\right\} = \sqrt[p]{\frac{1}{N}\sum_{(m,n)\in M_{(i,j)}}^{N} f^p(m,n)}.$$

This family is particularly interesting, because it generalizes a group of common filters, only by changing the value of $p$.

In particular, using Hölder filters we can construct new Kolmogorov-Lehmer filters as:

$$\underset{(m,n)\in M_{(i,j)}}{\mathbf{Lehm}^p}\left\{f(m,n)\right\} = \frac{\underset{(m,n)\in M_{(i,j)}}{\mathbf{Hold}^p}\left\{f(m,n)\right\}}{\underset{(m,n)\in M_{(i,j)}}{\mathbf{Hold}^{p-1}}\left\{f(m,n)\right\}}.$$

### B. Heronian filters

The classical Heronian mean and median definitions of two positive real numbers $a$ and $b$ are

$$\mathbf{MeanHeron}(a,b) = \left(a + \sqrt{ab} + b\right)/3 =$$

$$\left(\sqrt{aa} + \sqrt{ab} + \sqrt{bb}\right)/3, \qquad (2)$$

$$\mathbf{MedHeron}(a,b) = \mathbf{Med}\left\{\sqrt{aa}, \sqrt{ab}, \sqrt{bb}\right\}.$$

*Hero of Alexandria* is the Greek mathematician[5].

Let $(x_1, x_2, ..., x_N)$ be an $N$-tuple of positive real numbers. An obvious way to generalize Eq.(2) is by including inside the parentheses the square roots of all possible products of two elements.

**Definition 1.** The 2-generalized Heronian mean and median of an $N$-tuple of positive real numbers $(x_1, x_2, ..., x_N)$ are defined as

$$\mathbf{MeanHeron}_2(x_1, x_2, ..., x_N) = \frac{2}{N(N+1)}\sum_{i \le j}\sum \sqrt{x_i x_j},$$

$$\mathbf{MedHeron}_2(x_1, x_2, ..., x_N) = \mathbf{Med}\left[\left\{\sqrt{x_i x_j}\right\}_{i \le j}\right].$$

Now we cam generalize this definition using $k$-*th* roots of all possible distinct products of $k$ elements of $(x_1, x_2, ..., x_N)$, again with repetition. The number of all such products corresponds to extracting $k$ elements from a bag of $N$, with replacement, where $C_{N+k-1}^{k}$ is the binomial coefficient. This determines the normalization factor.

**Definition 2.** The generalized Heronian $k$-mean and $k$-median of an $N$-tuple of positive real numbers $(x_1, x_2, ..., x_N)$ are defined as

$$\mathbf{MeanHeron}_k(x_1, x_2, ..., x_N) =$$

$$= \frac{2}{C_{N+k-1}^{k}}\sum_{i_1 \le}\sum_{i_2 \le}...\sum_{\le i_2} \sqrt[k]{x_{i_1} x_{i_2} \cdots x_{i_k}}. \qquad (3)$$

$$\mathbf{MedHeron}_k(x_1,x_2,...,x_N) = \\ = \mathbf{Med}\left[\left\{\sqrt[k]{x_{i_1}x_{i_2}\cdots x_{i_k}}\right\}_{i_1\le i_2\le...\le i_k}\right]. \quad (4)$$

Let us introduce $(k+1)$-valued function on mask $\sigma: M \to \{0,1,...,k\}$. Obviously, $w(\sigma) = \sum_{(n,m)\in M}\sigma(n,m)$ is the weight of a function $\sigma$ and $w(\sigma)\in\{0,1,...,Nk\}$, where $N = |M|$. If $B^N_{k+1} = \{\sigma \mid \sigma: M \to \{0,1,2,...,k\}\}$ is the set of all $(k+1)$-valued functions, then

$$B^N_{k+1} = {}^0B^N_{k+1}\cup...\cup {}^rB^N_{k+1}\cup...\cup {}^{Nk}B^N_{k+1} = \bigcup_{r=0}^{Nk} {}^rB^N_{k+1},$$

where

$${}^iB^N_{k+1} = \left\{{}^i\sigma \mid \left({}^i\sigma: M\to\{0,1,...,k\}\right)\&\left(w({}^i\sigma)=i\right)\right\}$$

is the set of all $(k+1)$-valued function with weight $r$ and $\left|{}^rB^N_{k+1}\right|$ - cardinality of the set ${}^rB^N_{k+1}$. Now we define a product of pixels $f^{r_\sigma} := \prod_{(n,m)\in M_{(i,j)}}\left(f(n,m)\right)^{r\sigma(n,m)}$ associated with $(k+1)$-valued function ${}^r\sigma$. Using this product we define generalized aggregation Heronian filter as

$$\hat{s}(i,j) = \underset{(m,n)\in M_{(i,j)}}{\mathbf{GenHeron}^r}\left\{f(m,n)\right\} = \\ = \underset{\forall {}^r\sigma\in {}^rB^N_2}{\mathbf{Aggreg}}\left(\sqrt[r]{\prod_{(n,m)\in M_{(i,j)}}\left(f(n,m)\right)^{r\sigma(n,m)}}\right).$$

In particular cases we have

1) the arithmetic $r$-Heronian filter

$$\hat{s}(i,j) = \underset{(m,n)\in M_{(i,j)}}{\mathbf{MeanHeron}^r}\left\{f(m,n)\right\} = \\ \underset{\forall {}^r\sigma\in {}^rB^N_2}{\mathbf{Mean}}\left(\sqrt[r]{\prod_{(n,m)\in M_{(i,j)}}\left(f(n,m)\right)^{r\sigma(n,m)}}\right) = \\ = \frac{1}{\left|{}^rB^N_{k+1}\right|}\sum_{{}^r\sigma\in {}^rB^N_2}\sqrt[r]{\prod_{(n,m)\in M_{(i,j)}}\left(f(n,m)\right)^{r\sigma(n,m)}},$$

2) the median $r$-Heronian filter

$$\hat{s}(i,j) = \underset{(m,n)\in M_{(i,j)}}{\mathbf{MedHeron}^r}\left\{f(m,n)\right\} = \\ = \underset{(n,m)\in M_{(i,j)}}{\mathbf{Med}}\left[\left\{\sqrt[r]{\left(f(n,m)\right)^{r\sigma(n,m)}}\right\}_{\forall {}^r\sigma\in {}^rB^N_{k+1}}\right].$$

3) the Kolmogorov- Heronian filter

$$\hat{s}(i,j) = \underset{(m,n)\in M_{(i,j)}}{\mathbf{KolmHeron}^r}\left\{K \mid f(m,n)\right\} = \\ = K^{-1}\left[\underset{\forall {}^r\sigma\in {}^rB^N_{k+1}}{\mathbf{Mean}}\left(K\left(\sqrt[r]{\prod_{(n,m)\in M_{(i,j)}}\left(f(n,m)\right)^{r\sigma(n,m)}}\right)\right)\right],$$

where $r\le k$. In Fig.1 we present examples of $\underset{(m,n)\in M_{(i,j)}}{\mathbf{MeanHeron}^2}$- and $\underset{(m,n)\in M_{(i,j)}}{\mathbf{MedHeron}^2}$-filtering for $N = |M| = 5\times 5$.

a) b) PSNR =25dB

c) $\underset{(m,n)\in M_{(i,j)}}{\mathbf{MedHeron}^2}$, d) $\underset{(m,n)\in M_{(i,j)}}{\mathbf{MeanHeron}^2}$,

Fig. 1. Original (a) and noise (b) images.
Denoised images (c) PSNR=31dB ,and (d) PSNR=28dB

It is easy to see that $\sum_{i_1\le}\sum_{i_2\le}...\sum_{\le i_k}\sqrt[k]{x_{i_1}x_{i_2}\cdots x_{i_k}}$ is a symmetric polynomial in the variables $x_1,...,x_N$. There are a few types of symmetric polynomials in variables $x_1,...,x_N$ which that are associated new symmetric means.

*C. Symmetric filters*

Any monomial in $x_1,x_2,...,x_N$ can be written as $x_1^{p_1}x_2^{p_2}...x_N^{p_N}$, where the exponents $p_i$ are natural numbers (possibly zero); writing $\mathbf{p} = (p_1,p_2,...,p_N)$ this can be abbreviated to $\mathbf{X}^{\mathbf{p}} = x_1^{p_1}x_2^{p_2}...x_N^{p_N}$. If $p = p_1+p_2+...+p_N$ then we write $\sqrt[p]{\mathbf{X}^{\mathbf{p}}} = \sqrt[p]{x_1^{p_1}x_2^{p_2}...x_N^{p_N}}$.

**Definition 3.** The monomial symmetric polynomials of the first and second kinds are defined as the sums of all monomials $\mathbf{X}^{\mathbf{q}}$ or $\sqrt[q]{\mathbf{X}^{\mathbf{q}}}$, where $\mathbf{q}$ ranges over all distinct permutations of $\mathbf{p}$:

$$\mathbf{Mon}^I_{\mathbf{p}}(x_1,x_2,...,x_N) = \sum_{\sigma\in S_N}x_1^{\sigma(p_1)}x_2^{\sigma(p_2)}\cdots x_N^{\sigma(p_N)},$$

$$\mathbf{Mon}^{II}_{\mathbf{p}}(x_1,x_2,...,x_N) = \sum_{\sigma\in S_N}\sqrt[p]{x_{\sigma(1)}^{p_1}x_{\sigma(2)}^{p_2}...x_{\sigma(N)}^{p_N}}.$$

**Definition 4.** Let $x_1,x_2,...,x_N$ be positive real numbers and $\mathbf{p} = (p_1,p_2,...,p_N)\in R^N$. The $\mathbf{p}$-Muirhead symmetric polynomials of the first and second kinds are defined by

$$\mathbf{Mui}_{\mathbf{p}}^{I}(x_1,x_2,...,x_N)=\sum_{\sigma\in S_N}x_{\sigma(1)}^{p_1}x_{\sigma(2)}^{p_2}...x_{\sigma(N)}^{p_N},$$

$$\mathbf{Mui}_{\mathbf{p}}^{II}(x_1,x_2,...,x_N)=\sum_{\sigma\in S_N}\sqrt[p]{x_{\sigma(1)}^{p_1}x_{\sigma(2)}^{p_2}...x_{\sigma(N)}^{p_N}},$$

For example,

$$\mathbf{Mui}_{(1,0,...,0)}^{I}(x_1,x_2,...,x_N)=\sum_{i=1}^{N}x_i=\mathbf{Arithm}(x_1,x_2,...,x_N),$$

$$\mathbf{Mui}_{(1,1,...,1)}^{I}(x_1,x_2,...,x_N)=x_1x_2...x_N,$$

$$\mathbf{Mui}_{(\underbrace{1,1,...,1}_{k},0,...,0)}^{I}(x_1,x_2,...,x_N)=\sum_{i_1}\sum_{i_2}...\sum_{i_2}x_{i_1}x_{i_2}\cdots x_{i_k}.$$

$$\mathbf{Mui}_{(1,0,...,0)}^{II}(x_1,x_2,...,x_N)=\sum_{i=1}^{N}x_i=\mathbf{Mean}(x_1,x_2,...,x_N),$$

$$\mathbf{Mui}_{(1,1,...,1)}^{II}(x_1,x_2,...,x_N)=\sqrt[N]{x_1x_2\cdots x_N}=\mathbf{Geo}(x_1,x_2,...,x_N),$$

$$\mathbf{Mui}_{(\underbrace{1,1,...,1}_{k},...,1)}^{II}(x_1,x_2,...,x_N)=\mathbf{Heron}_k(x_1,x_2,...,x_N)=.$$

For each nonnegative integer $0\le k\le N$ the elementary $\mathbf{El}_k^I(x_1,x_2,...,x_N)$ and homogeneous $\mathbf{Hom}_k^I(x_1,x_2,...,x_N)$ symmetric polynomials are the sums of all distinct products of $k$ distinct variables:

$$\mathbf{El}_k^I(x_1,x_2,...,x_N)=\sum_{x_{i_1}<x_{i_2}<....<x_{i_k}}x_{i_1}x_{i_2}\cdots x_{i_k},$$

$$\mathbf{El}_k^{II}(x_1,x_2,...,x_N)=\sum_{x_{i_1}<x_{i_2}<....<x_{i_k}}\sqrt[k]{x_{i_1}x_{i_2}\cdots x_{i_k}}$$

and

$$\mathbf{Hom}_k^I(x_1,x_2,...,x_N)=\sum_{x_{i_1}\le x_{i_2}\le....\le x_{i_k}}x_{i_1}x_{i_2}\cdots x_{i_k},$$

$$\mathbf{Hom}_k^I(x_1,x_2,...,x_N)=\sum_{x_{i_1}\le x_{i_2}\le....\le x_{i_k}}\sqrt[k]{x_{i_1}x_{i_2}\cdots x_{i_k}}.$$

We then define

$$\mathbf{El}_{k_1k_2...k_r}^{I}(x_1,x_2,...,x_N)=\prod_{t=1}^{r}\mathbf{El}_{k_t}^{I}(x_1,x_2,...,x_N),$$

$$\mathbf{El}_{k_1k_2...k_r}^{II}(x_1,x_2,...,x_N)=\prod_{t=1}^{r}\mathbf{El}_{k_t}^{II}(x_1,x_2,...,x_N),$$

$$\mathbf{Hom}_{k_1k_2...k_r}^{I}(x_1,x_2,...,x_N)=\prod_{t=1}^{r}\mathbf{Hom}_{k_t}^{I}(x_1,x_2,...,x_N),$$

$$\mathbf{Hom}_{k_1k_2...k_r}^{II}(x_1,x_2,...,x_N)=\prod_{t=1}^{r}\mathbf{Hom}_{k_t}^{II}(x_1,x_2,...,x_N).$$

To each of polynomial $\mathbf{Mon}_{p_1,p_2,...,p_N}^{I,II}$, $\mathbf{Mui}_{p_1,p_2,...,p_N}^{I,II}$, $\mathbf{El}_{k_1k_2...k_r}^{I,II}$, $\mathbf{Hom}_{k_1k_2...k_r}^{I,II}$ we will associate normalized symmetric function:

$$\overline{\mathbf{Mon}}_{p_1,...,p_N}^{I,II}(x_1,...,x_N)=\mathbf{Mon}_{p_1,...,p_N}^{I,II}(x_1,...,x_N)/\mathbf{Mon}_{p_1,...,p_N}^{I,II}(1,...,1),$$

$$\overline{\mathbf{Mui}}_{p_1,...,p_N}^{I,II}(x_1,...,x_N)=\mathbf{Mui}_{p_1,...,p_N}^{I,II}(x_1,...,x_N)/\mathbf{Mui}_{p_1,...,p_N}^{I,II}(1,...,1),$$

$$\overline{\mathbf{El}}_{k_1...k_r}^{I,II}(x_1,...,x_N)=\mathbf{El}_{k_1...k_r}^{I,II}(x_1,...,x_N)/\mathbf{El}_{k_1...k_r}^{I,II}(1,...,1),$$

$$\overline{\mathbf{Hom}}_{k_1...k_r}^{I,II}(x_1,...,x_N)=\mathbf{Hom}_{k_1...k_r}^{I,II}(x_1,...,x_N)/\mathbf{Hom}_{k_1...k_r}^{I,II}(1,...,1).$$

We obtain eight families of a generalized symmetric means:

$$\mathbf{MonMean}_{p_1,p_2,...,p_N}^{I}(x_1,...,x_N)=\sqrt[p]{\overline{\mathbf{Mon}}_{p_1,p_2,...,p_N}^{I}(x_1,...,x_N)},$$

$$\mathbf{MuiMean}_{p_1,p_2,...,p_N}^{I}(x_1,...,x_N)=\sqrt[p]{\overline{\mathbf{Mui}}_{p_1,p_2,...,p_N}^{I}(x_1,...,x_N)},$$

$$\mathbf{ElMean}_{k_1k_2...k_r}^{I}(x_1,...,x_N)=\sqrt[k]{\overline{\mathbf{El}}_{k_1k_2...k_r}^{I}(x_1,...,x_N)},$$

$$\mathbf{HomMean}_{k_1k_2...k_r}^{I}(x_1,...,x_N)=\sqrt[k]{\overline{\mathbf{Hom}}_{k_1k_2...k_r}^{I}(x_1,...,x_N)}.$$

and

$$\mathbf{MonMean}_{p_1,p_2,...,p_N}^{II}(x_1,...,x_N)=\overline{\mathbf{Mon}}_{p_1,p_2,...,p_N}^{II}(x_1,...,x_N),$$

$$\mathbf{MuiMean}_{p_1,p_2,...,p_N}^{II}(x_1,...,x_N)=\overline{\mathbf{Mui}}_{p_1,p_2,...,p_N}^{II}(x_1,...,x_N),$$

$$\mathbf{ElMean}_{k_1k_2...k_r}^{II}(x_1,...,x_N)=\overline{\mathbf{El}}_{k_1k_2...k_r}^{II}(x_1,...,x_N),$$

$$\mathbf{HomMean}_{k_1k_2...k_r}^{II}(x_1,...,x_N)=\overline{\mathbf{Hom}}_{k_1k_2...k_r}^{II}(x_1,...,x_N),$$

where $k=k_1+k_2+...+k_r$, and $p=p_1+p_2+...+p_N$.

Using generalized symmetric means of the first kind we can construct the following families of symmetric filters of the first kind:

$$\hat{s}(i,j)=\underset{(m,n)\in M_{(i,j)}}{\mathbf{MonMean}_{p_{m,n}}^{I}}\{f(m,n)\}=$$

$$=\sqrt[p]{\frac{1}{N!}\sum_{\sigma\in S_N}\prod_{(m,n)\in M_{(i,j)}}\left[f(m,n)\right]^{\sigma(p_{m,n})}},$$

$$\hat{s}(i,j)=\underset{(m,n)\in M_{(i,j)}}{\mathbf{MuiMean}_{p_{m,n}}^{I}}\{f(m,n)\}=$$

$$=\sqrt[p]{\frac{1}{N!}\sum_{\sigma\in S_N}\prod_{(m,n)\in M_{(i,j)}}\left[f\left(\sigma(m,n)\right)\right]^{p_{m,n}}}.$$

Let $|M(n,m)|=(2r+1)\times(2r+1)=N$ and

$$\{f_{(n,m)}\}_{(n.m)\in M(i,j)}=$$

$$=\{f_{(-r,-r)}(i,j),...,f_{(0,0)}(i,j),...,f_{(r,r)}(i,j)\}=$$

$$=\{f_1(i,j),f_2(i,j),...,f_N(i,j)\}.$$

Then using this we define generalized aggregation **ElMean**- and **HomMean**-filters as

$$\hat{s}(i,j)=\underset{(m,n)\in M_{(i,j)}}{\mathbf{ElMean}_{k_1k_2...k_r}^{I}}\{f_{(m,n)}\}=$$

$$=\underset{(m,n)\in M_{(i,j)}}{\mathbf{ElMean}_{k_1k_2...k_r}^{I}}\{f_1(i,j),f_2(i,j),...,f_N(i,j)\}=$$

$$=\sqrt[k]{\mathbf{El}_{k_1k_2...k_r}^{I}(f_1,f_2,...,f_N)/\mathbf{El}_{k_1k_2...k_r}^{I}(1,1,...,1)}$$

$$\hat{s}(i,j)=\underset{(m,n)\in M_{(i,j)}}{\mathbf{HomMean}_{k_1k_2...k_r}^{I}}\{f_{(m,n)}\}=$$

$$=\underset{(m,n)\in M_{(i,j)}}{\mathbf{HomMean}_{k_1k_2...k_r}^{I}}\{f_1(i,j),f_2(i,j),...,f_N(i,j)\}=$$

$$=\sqrt[k]{\mathbf{Hom}_{k_1k_2...k_r}^{I}(f_1,f_2,...,f_N)/\mathbf{Hom}_{k_1k_2...k_r}^{I}(1,1,...,1)}.$$

In Fig.2 we present examples of **MonMean-. MuiMen-, ElMean**- and **HomMean**-filtering.

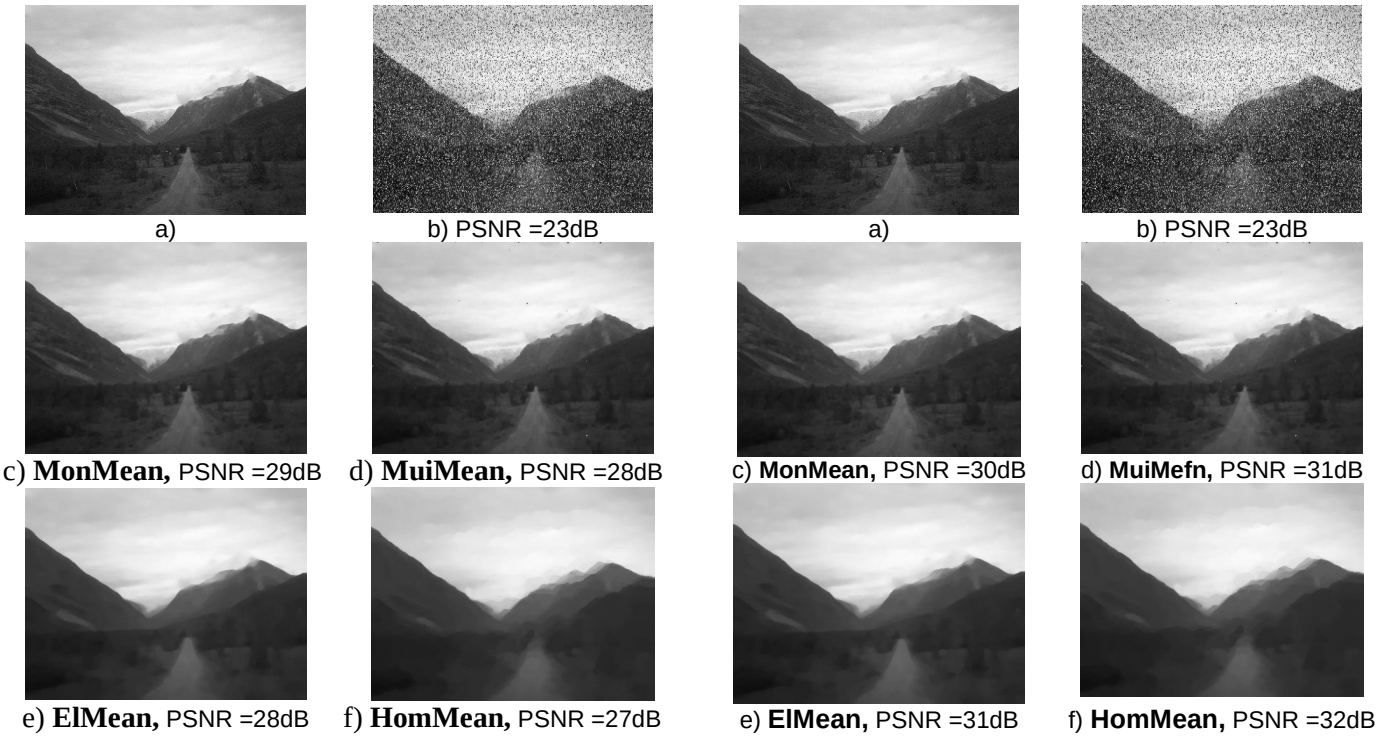



*Fig. 2. Original (a) and noise (b) images. Denoised images c)* PSNR =23dB*, d), e), f).*




*Fig. 3. Original (a) and noise (b) images. Denoised images c), d), e), f).*

Using generalized symmetric means of the second kind we can construct the following families of symmetric filters:

$$\hat{s}(i,j) = \mathbf{MonMean}^{II}_{\substack{p_{m,n} \\ (m,n)\in M_{(i,j)}}}\left\{ f(m,n)\right\} =$$

$$= \frac{1}{N!}\sum_{\sigma\in S_N} \sqrt[p]{\prod_{(m,n)\in M_{(i,j)}} \left[ f(m,n)\right]^{\sigma(p_{m,n})}},$$

$$\hat{s}(i,j) = \mathbf{MuiMean}^{II}_{\substack{p_{m,n} \\ (m,n)\in M_{(i,j)}}}\left\{ f(m,n)\right\} =$$

$$= \frac{1}{N!}\sum_{\sigma\in S_N} \sqrt[p]{\prod_{(m,n)\in M_{(i,j)}} \left[ f\left(\sigma(m,n)\right)\right]^{p_{m,n}}}$$

$$\hat{s}(i,j) = \mathbf{ElMean}^{II}_{\substack{k_1k_2\ldots k_r \\ (m,n)\in M_{(i,j)}}}\left\{ f_{(m,n)}\right\} =$$

$$= \mathbf{ElMean}^{II}_{\substack{k_1k_2\ldots k_r \\ (m,n)\in M_{(i,j)}}}\left\{ f_1(i,j), f_2(i,j),..., f_N(i,j)\right\} =$$

$$= \mathbf{El}^{II}_{k_1k_2\ldots k_r}\left( f_1, f_2,..., f_N\right) / \mathbf{El}^{II}_{k_1k_2\ldots k_r}\left(1,1,...,1\right),$$

$$\hat{s}(i,j) = \mathbf{HomMean}^{II}_{\substack{k_1k_2\ldots k_r \\ (m,n)\in M_{(i,j)}}}\left\{ f_{(m,n)}\right\} =$$

$$= \mathbf{HomMean}^{II}_{\substack{k_1k_2\ldots k_r \\ (m,n)\in M_{(i,j)}}}\left\{ f_1(i,j), f_2(i,j),..., f_N(i,j)\right\} =$$

$$= \mathbf{Hom}^{II}_{k_1k_2\ldots k_r}\left( f_1, f_2,..., f_N\right) / \mathbf{Hom}^{II}_{k_1k_2\ldots k_r}\left(1,1,...,1\right).$$

In Fig.3 we present examples of **MonMean-. MuiMen-, ElMean**- and **HomMean**-filtering.

## IV. Vector median Frechet filters

Median filtering has been widely used in image processing as an "edge preserving" filter. The basic idea is that the pixel value is replaced by the median of the pixels contained in a window around it. In this subsection this idea is extended to vector-valued images, based on the fact that the median is also the value that minimizes the $L_1$ norm between all the pixels in the window. In vector-valued case, we need to define a distance between pairs of vectors on the domain of application. Let $\langle \mathbf{D}, d\rangle$ be a metric space (color or multicolor space), where $d$ is a distance function for pairs of objects in $\langle \mathbf{D}, d\rangle$ (that is, $d : \mathbf{D}\times\mathbf{D} \to \mathbf{R}^+$). Let $w_1, w_2, ..., w_N$ be $N$ weights summing to 1 and let $\mathbf{x}^1, \mathbf{x}^2, ..., \mathbf{x}^N \in \mathbf{D} \subset \mathbf{R}^n$ be $N$ observations from $\langle \mathbf{D}, d\rangle$.

**Definition 5** [6]**.** The weighted Fréchet median or generalized vector-valued median is the point, $\mathbf{c}_{opt} \in \mathbf{D}$, that minimizes the Fréchet function $\sum_{i=1}^{N} w_i d\left(\mathbf{c}, \mathbf{x}^i\right)$ (the weighted sum distances from an arbitrary point $\mathbf{c} \in \mathbf{D}$ to each point $\mathbf{x}^1, \mathbf{x}^2, ..., \mathbf{x}^N \in \mathbf{D} \subset \mathbf{R}^K$) and is formally defined as

$$\mathbf{c}_{opt} \equiv \mathbf{Frech}\left(\mathbf{x}^1, \mathbf{x}^2, ..., \mathbf{x}^N\right) =$$

$$= \mathbf{VectMed}\left(\mathbf{x}^1, \mathbf{x}^2, ..., \mathbf{x}^N\right) = \arg\min_{\mathbf{c}\in\mathbf{D}} \sum_{i=1}^{N} w_i \rho\left(\mathbf{c}, \mathbf{x}^i\right).$$

Note that **argmin** means the argument for which the sum is minimized. In this case, it is the point $\mathbf{c}_{opt}$ from $\mathbf{D} \subset \mathbf{R}^n$ for one the sum of all distances to the $\mathbf{x}^i$'s is minimum. This generalizes the ordinary median, which has the property of minimizing the sum of distances for scalar-valued pixels, and provides a central tendency higher dimensions [1,2].

Let $\mathbf{D} \subset \mathbf{R}^n$ be a multicolor space. We introduce the notion of aggregation space $\langle \mathbf{D}, d\text{-}\mathbf{Agg} \rangle$. Here $d$-**Agg** is an aggregation pseudo-distance function $d\text{-}\mathbf{Agg} : \mathbf{D} \times \mathbf{D} \to \mathbf{R}^+$ for pairs of pixels in $\mathbf{D}$:

$$d\text{-}\mathbf{Agg}_1\left(\mathbf{c}, \mathbf{x}\right) = \mathbf{Agg}_1\left(c_1 - x_1, c_2 - x_2, ..., c_K - x_K\right),$$

where $\mathbf{Agg}_1$ is an arbitrary aggregation operator acting like pseudo-distance function.

The new generalized aggregation operators **FrechAgg** applied to $(\mathbf{x}^1, \mathbf{x}^2, ..., \mathbf{x}^N)$ are formally defined as

$$\mathbf{FrechAgg}\left(\mathbf{x}^1, \mathbf{x}^2, ..., \mathbf{x}^N\right) :=$$

$$= \arg\min_{\mathbf{c}\in\mathbf{D}} \left[ \mathbf{Agg}_2\left(d\text{-}\mathbf{Agg}_1\left(\mathbf{c}, \mathbf{x}^1\right), ..., d\text{-}\mathbf{Agg}_1\left(\mathbf{c}, \mathbf{x}^N\right)\right)\right],$$

**FrechAgg** depends on two aggregation operators $\mathbf{Agg}_1$, $\mathbf{Agg}_2$. For example, if, $\mathbf{Agg}_2 = \sum$ then

$$\mathbf{FrechAgg}\left(\mathbf{x}^1, \mathbf{x}^2, ..., \mathbf{x}^N\right) = \arg\min_{\mathbf{c}} \left[ \sum_{i=1}^{N} d\text{-}\mathbf{Agg}\left(\mathbf{c}, \mathbf{x}^i\right) \right]$$

and if $d\text{-}\mathbf{Agg}\left(\mathbf{c}, \mathbf{x}^i\right) = d\left(\mathbf{c}, \mathbf{x}^i\right)$ is a metric distance, then we have the classical Fréchet median.

Now we can use generalized aggregation operators **FrechAgg** in (1)

$$\hat{\tilde{\mathbf{s}}}(i, j) = \mathbf{FrechAgg}_{(k,l)\in M(i,j)}\left\{\tilde{\mathbf{f}}(k, l)\right\},$$

it becomes an aggregation digital Fréchet filter.

In Fig.4 we present examples of **FrechAgg**-filtering.

We developed a new theoretical framework for image filtering using aggregation operators. The main goal of the work is to show that aggregation operators can be used to solve problems of image filtering in a natural and effective manner. Some properties of a nonlinear aggregation filters are exploited in this paper. Unlike the linear masking filter, they avoid amplification thanks to the nonlinearity of the response to luminance variations; unlike the classical linear and median filters, they are able to sharpen even small details as its impulse response demonstrates.

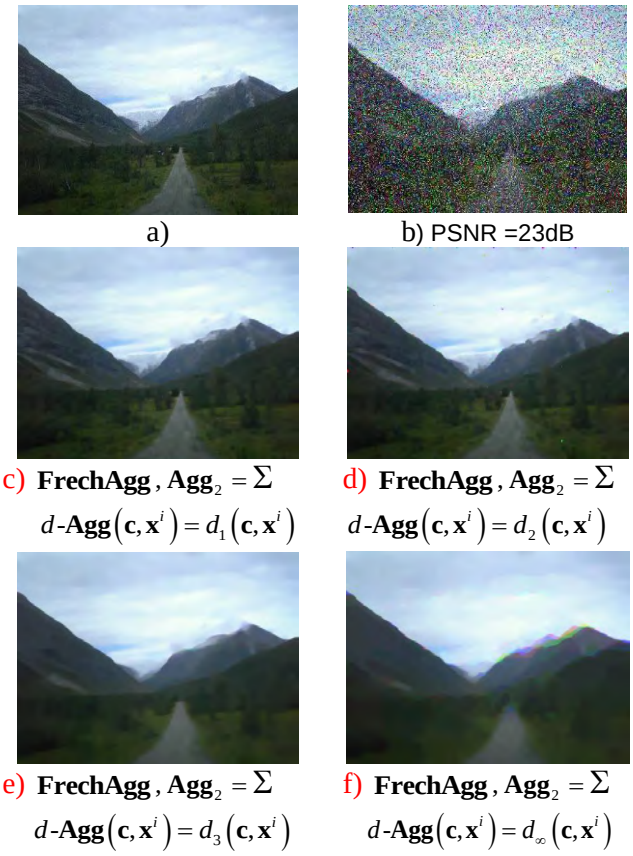



*Fig. 4. Original (a) and noise (b) images. Denoised images c)* PSNR =32dB, *d)* PSNR =31dB, *e)* PSNR =30dB, *f)* PSNR =29dB.

## Acknowledgment

This work was supported by grants the RFBR № 13-07-12168, RFBR № 13-07-00785.

## References

[1] R.C. Gonzalez and R.E. Woods, Digital Image Processing. New York: Addison-Wesley, 1992.

[2] G. Mayor and E. Trillas, "On the representation of some Aggregation functions," Proceeding of ISMVL, 1986, pp. 111-114.

[3] S. Ovchinnikov, On Robust Aggregation Procedures, Aggregation Operators for Fusion under Fuzziness. Bouchon-Meunier B. (eds.), 1998, pp. 3-10.

[4] A. Kolmogorov, "Sur la notion de la moyenne," Atti Accad. Naz. Lincei, 1930, vol. 12, pp. 388–391.

[5] S. Sykora, "Mathematical Means and Averages: Generalized Heronian means". Stan's Library, Ed.S.Sykora, Vol.III, 2009.

[6] C.Bajaj, "Proving geometric algorithms nonsolvability: An application of factoring polynomials," Jornal of Symbolic Computaton, 1986, No. 2, pp. 99–102.

# Formation and recognition of metabolic profile of cancer on the basis of chromatography-mass spectrogram of urine volatile metabolite image analysis

Rozhentsov A.A.
Volga state technical university organization
Yoshkar-Ola, Russian Federation acceptable
rozhencovaa@volgatech.net

Lychagin K.A.
Volga state technical university organization
Yoshkar-Ola, Russian Federation acceptable
1349@lenta.ru

Ryzhkov V.L.
Republican Clinical Hospital
Yoshkar-Ola, Russian Federation acceptable
viktorryzhkov79@rambler.ru

Mitrakova N.N.
Republican Clinical Hospital
Yoshkar-Ola, Russian Federation acceptable
endomitrakova@mail.ru

Furina R. R.
Republican Clinical Hospital
Yoshkar-Ola, Russian Federation acceptable
furina_raisa@mail.ru

***Abstract*— The key idea of the paper is to introduce the basic principles gas chromatography-mass-spectrometry as a new diagnostic method. The engineering approach to the problem based on the use of a gas chromatograph with mass spectrometry to suggest innovative solutions for method of screening early cancer diagnostics. Besides, the problem of the identification of specific biomarkers thoroughly considered. The result of laboratory studies on metabolic profile handling are analyzed in detail.. The material presented can open new prospects for further research studies.**

***Keywords*— *early diagnosis of cancer; volatile organic metabolites; chromatography-mass spectrometry; solid microextraction; image processing; metabolic profiles***

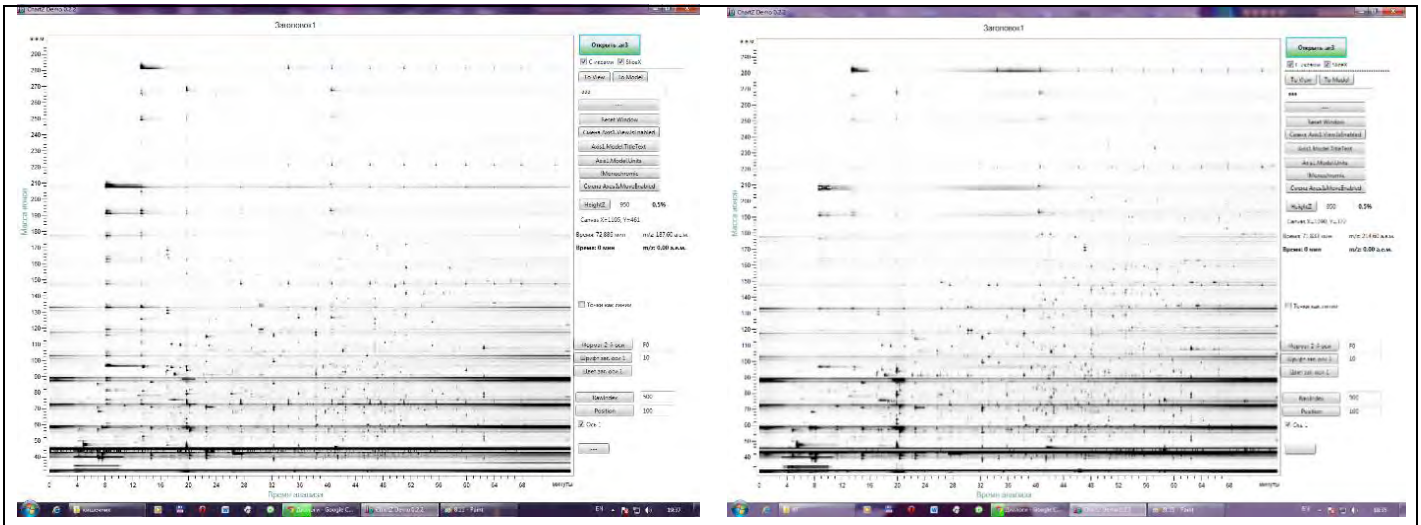
## I. INTRODUCTION

Early diagnosis is essential for successful treatment of cancer. Currently, there are many approaches to solving this problem [1-3], however, the majority of examinations is not always possible due to the lack of availability of quality health care, financial and organizational problems, low efficiency of the necessary equipment. One of the possible approaches is the developing a method for screening on the basis of the analysis of the composition of the volatile fractions of urine. The aim of this study is to develop methods of the analysis of the composition of volatile organic metabolites in urine (VOM), preparation of metabolic profiles of cancer on the basis of the image analysis by gas chromatography-mass spectrograms VOM, evaluation of the effectiveness of the methods of cancer detection based on the analysis of the VOM composition.

## II. FORMATION AND IMAGE PROCESSING OF THE GAS CHROMATOGRAPHY-MASS SPECTROGRAMS OF VOLATILE METABOLITES IN URINE

The analysis of the composition of volatile metabolites in urine was made by the gas chromatography-mass spectrometry. For sample preparation the method of the solid microextraction was applied. Fig. 1 shows the example of the images of the chromatography-mass spectrogram. The signal level at the

output of the mass the spectrometer detector is encoded by the brightness of the image.



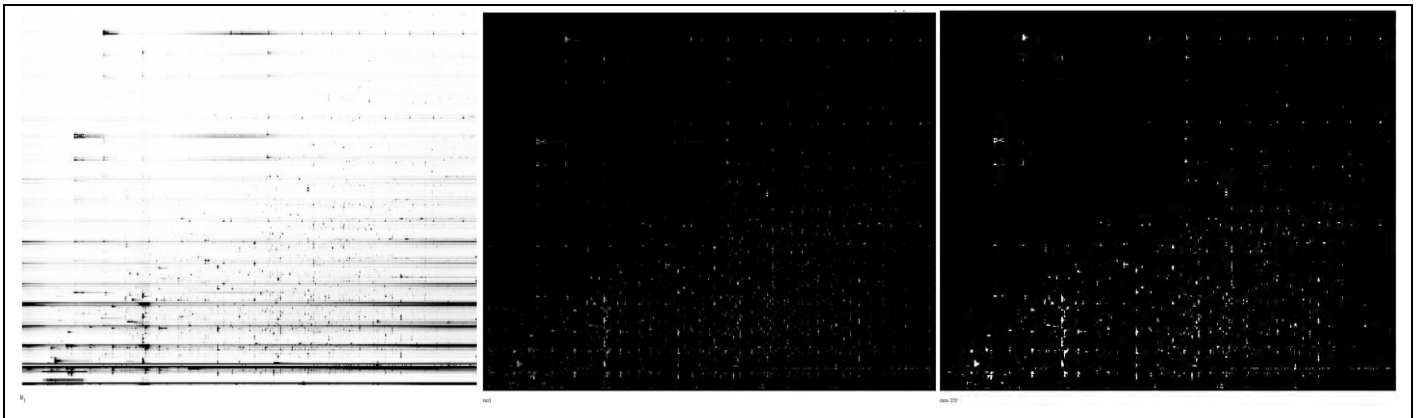
*a)* *b)*

Fig. 1. Image 3D spectrogram: control group, (b) colorectal cancer

The standard procedure of the chromatography-mass spectrogram processing is the identification of the substances that meet the selected in the chromatogram peaks by comparing the received and reference mass spectra. The reference mass spectra are taken from the libraries of the mass spectral data, for example, NIST 02, NIST 05, WILEY, etc. The depended chromatography-mass spectrometry data may be accompanied by the errors caused by the inaccurate selection of chromatographic peaks, errors in identification of substances by mass spectra, the incompleteness of the used libraries, etc. In this regard, the possibility of the formation of metabolic profiles of cancer was considered directly on the images of the chromatography-mass spectrograms of volatile metabolites in urine.

The procedure of the formation of the chromatography-mass spectrograms image corresponding to the metabolic profiles of different cancers, consists of several stages. In addition to useful information the original image of the chromatography-mass spectrogram contains a significant amount of interferences caused by the contact of detector materials internal coating column, impurities in gas-carrier, etc. They are shown as horizontal lines observed throughout chromatography-mass-spectrogram (Fig. 1). The drift of the base line of the chromatogram, which can be manifested as changes in the average brightness of the image over time is also shown. It is possible to eliminate the interference at the first stage of preliminary filtering of the image. Figure 2 shows an example of the image processing of the gas chromatography-mass spectrogram. The next stage is to identify the typical for patient metabolites. For this purpose the threshold processing of the received image is performed as the result are selected the

| | Control group | Bowel cancer | Lung cancer | Esophageal cancer | Gastr. cancer | Correct diagnosis, % |
|---|---|---|---|---|---|---|
| Control group | 29 | 0 | 3 | 0 | 0 | 90,625 |
| Bowel cancer | 0 | 9 | 2 | 1 | 0 | 75 |
| Lung cancer | 0 | 0 | 28 | 1 | 0 | 96,55172414 |
| Esophageal cancer | 0 | 0 | 0 | 8 | 0 | 100 |
| Gastric cancer | 0 | 0 | 0 | 1 | 26 | 96,2962963 |



*a)* *b)* *c)*

Fig. 2. Example of removing noise from the image chromatography-mass spectrograms: (a) original image, b) image after pre-filtering; c) image after thresholding processing

significant in metabolites (figure 2,b).

For the formation of the metabolic profile of a particular form of cancer according to the above mentioned method both all images of the chromatography-mass spectrograms of patients and the data for healthy people from the control group are processed,. After that the next group of the images containing information on all the metabolites of patients of this group is formed. For this purpose the operation of logical OR is performed a many all the images of this group.

At the next stage of processing in each image the detected unique elements and the reference image metabolic profiles were detected in accordance with rule:

$$J^{эт_m}_{x,y} = \begin{cases} 1, if\ J^{обр_n}_{x,y} = 0, m \neq n, n = 1,..,5 \\ 0, \text{otherwise}, \end{cases} \quad m = 1,..,5,$$

Finally, that is, the resulting image of this group of patients contained only those metabolites that were not met in the images of other groups. Figure 3 shows the results of imaging of the chromatography-mass spectrograms reference metabolic profiles.

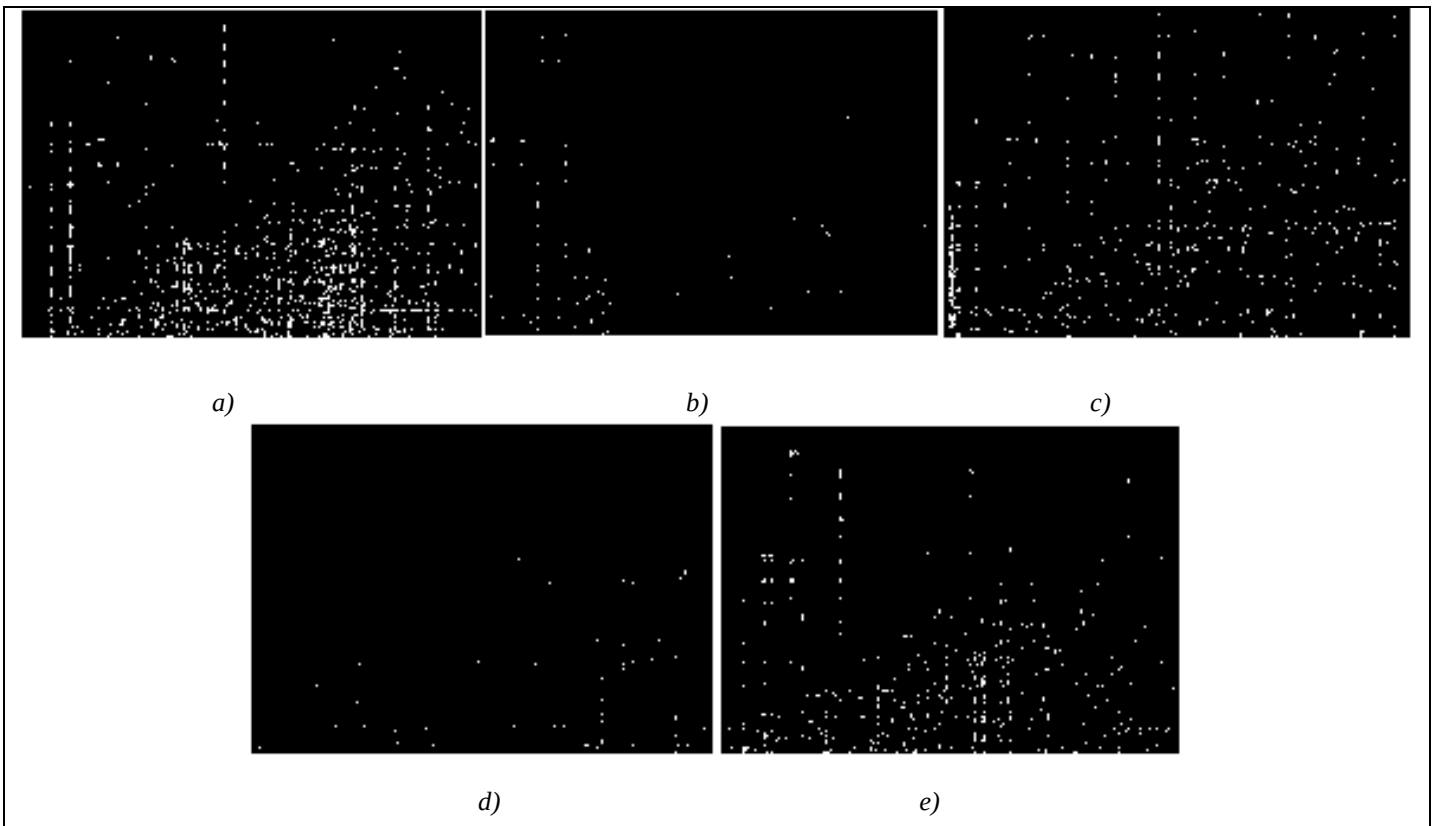



Fig.3. The reference image metabolic profiles: a) control group, b)bowel cancer, c) lung cancer, d) esophageal cancer, e) gastric cancer

## III. Evaluation of the effectiveness of the techniques for the detecting cancer images of the chromatography-mass spectrograms

To assess the possibility of recognition of malignant images chromatography-mass spectrograms the group of patients with clinically confirmed cancers including gastric cancer (27 people), lung cancer (29), bowel cancer (12), esophageal cancer (8 people) was examined. The control group of the healthy people containing 32 people was also.

The table shows the results of the evaluation of the efficiency of cancer detection based on the image analysis by the gas chromatography-mass spectrograms of volatile metabolites in urine.

According to the obtained results, the assessment of the sensitivity of the method to the considered sample is 100%, specificity is 90,62%, the probability of errors of the first type is 0, the probability of the second type errors is 0,0938.

## IV. Conclusion

In the paper a method for early diagnosis of cancer based on the analysis of the composition of volatile metabolites in urine is proposed. The image chromatography-mass spectrograms of metabolic profiles for various forms of cancer, including colon cancer, lung cancer, esophageal cancer, stomach cancer, and for healthy people were formed. The effectiveness of the techniques for detection of cancer based on the analysis of the composition of the metabolites is evaluated.. The obtained results can be considered as the basis for the research in the field of diagnosis of other forms of cancer on the basis of the analysis of the composition of volatile

metabolites in urine and can be also used in clinical practice for the primary diagnosis of cancer.

## ACKNOWLEDGMENT

The authors are grateful to the company «JSC Chromatec» for the provided equipment and assistance when conducting research.

## REFERENCES

[1] Phillips M., Cataneo R. N., Cummin A.R.C. и др. Detection of Lung Cancer With Volatile Markers in the Breath. Preliminary report// CHEST / 123 / 6 / JUNE, 2003.

[2] Phillips M, Gleeson K, Hughes JM, et al. Volatile organic compounds in breath as markers of lung cancer: a crosssectional study. Lancet 1999; 353:1930–1933.

[3] J Woo HM, Kim KM, Choi MH. et al. Mass spectrometry based metabolomic approaches in urinary biomarker study of women's cancers. Clin Chim Acta. 2009;400:63-9.

[4] Statistics of cancer URL: http://www.knigamedika.ru/novoobrazovaniya-kologiya/statistika-zabolevaemosti-rakom.html ( date of access 10.03.2014).

[5] Oliveira PA, Colaco A, Chaves HR. et al. Chemical carcinogenesis. An Acad Bras Cienc. 2007;79:593-616.

[6] Furina R.R., Mitrakova N.N., Ryzhkov V.L., Safiullin I.K. Metabolomicheskie issledovanija v medicine// Kazanskij medicinskij zhurnal. – 2014 – T.XCV,№1 – p.1- 6.

[7] Carev N.I., Carev V.I., Katrakov I.B. Prakticheskaja gazovaja hromatografija. – Barnaul: Izd-vo Alt. Un-ta – 2000 – 156p.

[8] Kouremenos K.A., Pitt J., Marriott P.J. Metabolic profiling of infant urine using comprehensive two-dimensional gas chromatography: Application to the diagnosis of organic acidurias and biomarker discovery// Journal of Chromatography A – 2010 – Vol. 1217 – P.104–111.

[9] Patti G.J., Yanes O., Siuzdak G. Innovation: Metabolomics: the apogee of the omics trilogy // Nat. Rev. Mol. Cell. Biol. – 2012 – Vol. 13 – P. 263-269.

[10] Pauling L, Robinson A B, Teranishi R., Cary P. Quantitative analysis of urine vapor and breath by gas–liquid partition chromatography // Proc. Natl Acad. Sci. USA – 1971 – Vol.68 – P. 2374-2376.

[11] Silva C.L., Passos M., Camara J.S. Investigation of urinary volatile organic metabolites as potential cancer biomarkers by solid-phase microextraction in combination with gas chromatography-mass spectrometry// British Journal of Cancer – 2011- Vol.105 – P. 1894 – 1904.

[12] Zimmermann D., Hartmann M., Moyer M. P. et al. Determination of volatile products of human colon cell line metabolism by GC/MS analysis.// Metabolomics – 2007 – Vol. 31 – P. 13-17.

# Genetic Algorithm Application in Image Segmentation

Pavel Jedlička
The University of West Bohemia
Plzeň, Czech Republic
Email: skely@kky.zcu.cz

Tomáš Ryba
The University of West Bohemia
Plzeň, Czech Republic
Email: tryba@kky.zcu.cz

***Abstract*—In consideration of living organisms' ability to endure for years, and their ability to adapt to surrounding environment, the mechanism of evolution is the inspiration for creating a new genetic algorithm. The goal of this paper is to examine possibilities of genetic algorithm application for segmentation of digital image data, implementation of this algorithm, and to create tools for its testing. The next goal is to examine possible choices of algorithm's parameters, and to compare quality of the results with other segmentation methods within various image data.**

## I. Introduction

The ability of living organisms to endure for many years and their ability to adapt to surrounding environment is a good inspiration for creating mechanisms in technical applications. There were attempts to simulate function of brain which is responsible for making decisions of living organisms. The solution of this is artificial neural network. Another important task is simulation of the mechanism of evolution and natural selection, first mentioned in [1]. This simulation is called genetic algorithm (GA).

GA is used for solving tasks as mathematical optimization, functional analysis and approximative solving NP-complete problem [2].

Image segmentation is one of many fundamental problems in computer vision. There are many more or less successful methods for solving this problem. The goal of this process is to partition digital image into multiple segments based on visual or logical similarity [3]. The goal of this paper is to examine application of GA in image segmentation.

## II. Genetic algorithm methodology

Principle of GA is an iterative creating of populations of individuals. Each individual represents one solution to a process. Evolution of individuals through generations and evaluation of each individual in generation provides convergence of the process to the optima, though global optimum is not ensured [2].

Initialization of GA is usually made by creating of individuals with random set of attributes (genes). Set of $n$ attributes of individual $i$ is represented as vector:

$$i = [i_1, i_2, ..., i_n], \tag{1}$$

In each iteration the cost function is calculated. Cost function is used for evaluate of quality of an individual as a solution.

There are two main methods that generates a new individual in our version of GA, crossover and mutation. Crossover is based on combination of parents' genes. In basic version selection of parents is made randomly. In our version selection of parents is based on value of cost function to increase speed of convergence. This method is called roulette wheel selection [5]. Value of cost function is transformed to probability:

$$p_j = \frac{f_j}{\sum_{j=1}^{J} f_j}, \tag{2}$$

where $p_j$ is probability of choosing an individual $j$ as a parent, $f_j$ stands for value of cost function for individual $j$ and $J$ is number of individuals in generation.

There are some different versions of crossover method. Number of parents can be more than two. Although it does not have equivalent in nature moreover crossover method become more complex and there is no prove that it provides any improvement. Two parent individuals are usually chosen and also there are two new individuals created. The reason is simply to keep number of individuals same among generations. Single point crossover is the simplest variant, which uses a randomly generated number that determines a gene which is last gene inherited from first patent for first descendant. Second descendant is created same way except the order of parents is swapped. Crossover by random vector is used for purpose of this paper. There is randomly generated binary vector with same length as gene of an individual. The first descendant's gene on positions where randomly generated vector is equal 1 inherit first parent's gene otherwise second parent gene is inherited. The second descendant is created the same way except the order of parents. This can be defined: Let $a$ and $b$ are parent individuals defined by gene vectors $a = [a_1, a_2, \ldots, a_n]^T$ and $b = [b_1, b_2, \ldots, b_n]^T$, where $n$ is number of genes and let $r$ is randomly generated binary vector $r = [r_1, r_2, \ldots, r_n]^T$. Descendants $c$ and $d$ are defined:

$$\begin{aligned} c = [&a_1 \cdot r_1 + b_1 \cdot \overline{r_1}, a_2 \cdot r_2 + b_2 \cdot \overline{r_2}, \ldots \\ &\ldots, a_n \cdot r_n + b_n \cdot \overline{r_n}]^T, \\ d = [&a_1 \cdot \overline{r_1} + b_1 \cdot r_1, a_2 \cdot \overline{r_2} + b_2 \cdot r_2, \ldots \\ &\ldots, a_n \cdot \overline{r_n} + b_n \cdot r_n]^T, \end{aligned} \tag{3}$$

where $\overline{x}$ is binary negation of $x$.

Mutation is method which provides possibility of appearance a gene that is not present in previous generation. It also provides ability of algorithm to disrupt convergence which can be useful to reach better local optima. Mutation is operation which changes value of one or more genes of an individual. There are some different versions of mutation method as single or multiple gene mutation. Maximum size of change is required to avoid breaking convergence.

One of algorithm's input parameters are probabilities of crossover and mutation. It means that this methods are not applied in each iteration. Probability of crossover is chosen usually high (about $90\%$) because this method is meant to be main source of new individuals and it provides convergence of algorithm. Probability of mutation is recommended to be lower (about $5 - 10\%$) to avoid breaking convergence.

After defined number of iterations (generations) algorithm ends. Individual with the lowest value of cost function is declared as a solution of GA's task. There is another method which does not have it's equivalent in nature but is very useful to avoid loss of the best individual which can appear among generations, elitism selection is made. This selection means that chosen percentage of the best individuals measured by cost function are moved to new generation unchanged. Same amount of worse-valued individuals in the new generation is discarded to keep size of population constant.

## III. Image representation

There are more different strategies how to represent image as a individual or it's genome. Due to high time demands of GA it is useful to reduce amount of input data. There are methods for creating superpixels (SP) (eg. Felzenszwalb's method [4], quickshift method [6], SLIC method [7], etc.) that partially solve this task.

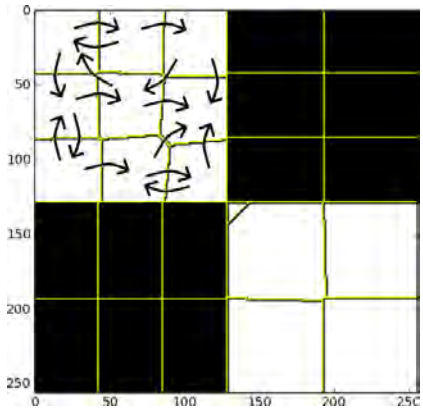



Fig. 1: Ilustration of locust swarm segmentation of SP.

For each individual in generation array of attributes is made. Dimension of array is $n \times m$, where $n$ is number of SP in picture and $m$ is number of pointers. For each individual each SP is connected through pointers with defined number of it's spatial neighbours. Areas made by these connections responds to one segment in segmentation task. This representation is inspired in method locust swarm [8]. Example of connecting SPs is illustrated in Fig. 1. Targets of pointers are independent on each other. This means that one SP can be connected with only one neighbour with all it's pointers. SP can also point on itself. It is because some pictures can contain small objects, which can be preprocessed to only one superpixel.

## IV. Cost function

Cost function which is used for evaluating results of iterations in this algorithm is graphcut cost function [9]. Value of the cost function determines which individual is more preferable to be used for crossovering in next iteration and selection of individuals for elitism method. For each segmentation $f$ the cost function $C$ is computed:

$$C(f) = C_{data}(f) + C_{smooth}(f), \tag{4}$$

where $C_{data}$ is cost caused by difference of two SP in same segment and $C_{smooth}$ is cost caused by dispersion of segments. Let the whole set of SP is $I$ and let $N$ is set of all pairs of adjacent SP $(p, q)$, $p, q \in I$. Each area $i$ recognized as segment is marked as $L_i$. Segmentation of whole image is $L = (L_1, L_2, ..., L_{|I|}$. Finding minima cost function $C$ responds to finding optimal segmentation. It is possible to weight both parts (data similarity $R(L)$ and dispersion $B(L)$) according to required result using parameter $\lambda$:

$$C(L) = \lambda \cdot R(L) + B(L), \tag{5}$$

where

$$R(L) = \sum_{p \in I} R_p(L_p), \tag{6}$$

$$B(L) = \sum_{p,q \in I} B_{p,q} \cdot \delta(L_p, L_q), \tag{7}$$

$$\delta(L_p, L_q) = \begin{cases} 1 & : \quad L_p \neq L_q \\ 0 & : \quad otherwise. \end{cases} \tag{8}$$

$$R_p = -lnP(I|L_p), \tag{9}$$

$$B_{p,q} = exp(-\frac{(I_p - I_q)^2}{2\sigma^2}), \tag{10}$$

where $P(I|L_p)$ is probability of belonging SP $I$ to segment $L_p$. Value of SP $p$ is $I_p$ and $I_q$ is value of neighbouring SP.

## V. Testing dataset

Efficiency of GA was tested on image datasets [10] which includes at least 5 different manual segmentations for each image. Example of one testing dataset is illustrated in Fig. 2. It is evident that manual segmentation is very subjective and depends on purpose of segmentation.

## VI. Results

Results of GA were compared to Felzenszwalb's method (FH) [4] which is automatic segmentation method and random walker (RW) [11] which is semi-automatic method. FH method has one input parameter. Parameter of FH method was chosen to be optimal for SE and SP evaluating method. In real task optimal value of this parameter is not probable and therefore results in real task should be slightly worse. RW

Fig. 2: Testing dataset #227092. Original picture and manual segmentations.

method is semi-automatic method. User have to select some pixels and determine which segment these pixels belonging to. To reduce human factor five different users were chosen to set up this method.

Sensitivity (SE) and specificity (SP) measures have been chosen for quantification of results. Each manual segmentation was chosen as the gold standard. SE is percentage of pixels signed as edges in manual segmentation found by algorithm. SP is percentage of not-edge signed pixels in manual segmentation correctly found by algorithm.

Results for datasets #135069, #227092 and #42049 are summarized in tables 1, 2 and 3 respectively. Source image and its segmentation made by GA is illustrated in Fig. 3 and Fig. 4.

On average GA has better or similar results as FH. Both methods are worse then RW which has advantage in additional information from user, because RW is semi-automatic method. There are detailed result in Table 1. Average and median columns are statistics from different manual segmentations. GA statistics in rows are maximums and medians from 20 runs of algorithm, because of stochasticity of GA. RW statistics in rows are maximums and medians from 5 different users settings, because RW is semi-automatic.

GA appears to be valuable automatic image segmentation method. Main disadvantage is high time demand. Stochasticity of GA may cause some problems, but this effect can be reduced by multiple running of GA and eg. median result should be used as a final result.

In further work it worth focus on different representation of image. Another field of improvement should be time efficiency which is main weakness of this algorithm.

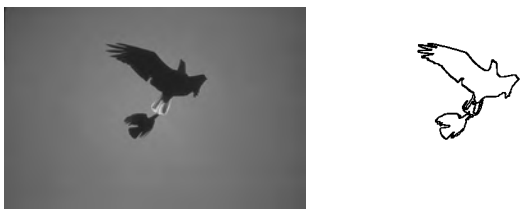


Fig. 3: Testing input image #227092 and GA segmentation.

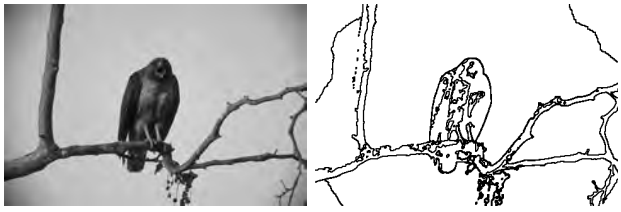


Fig. 4: Testing input image #42049 and GA segmentation.

| | average | median |
|---|---|---|
| GA max SE | 23.07 | 23.75 |
| GA max SP | 96.92 | 96.92 |
| GA median SE | 22.54 | 23.21 |
| GA median SP | 96.85 | 96.85 |
| FH SE | 14.88 | 14.75 |
| FH SP | 96.96 | 96.94 |
| RW max SE | 24.66 | 25.36 |
| RW max SP | 96.71 | 96.69 |
| RW median SE | 24.06 | 24.77 |
| RW median SP | 96.67 | 96.65 |

Table 1: Results resume for testing dataset #135069.

| | average | median |
|---|---|---|
| GA max SE | 18.87 | 18.29 |
| GA max SP | 89.87 | 87.21 |
| GA median SE | 16.84 | 17.36 |
| GA median SP | 86.95 | 83.38 |
| FH SE | 10.86 | 13.15 |
| FH SP | 89.87 | 87.21 |
| RW max SE | 18.53 | 19.31 |
| RW max SP | 87.99 | 84.33 |
| RW median SE | 16.72 | 17.85 |
| RW median SP | 87.70 | 83.93 |

Table 2: Results resume for testing dataset #227092.

| | average | median |
|---|---|---|
| GA max SE | 11.10 | 10.94 |
| GA max SP | 93.13 | 93.17 |
| GA median SE | 9.80 | 9.71 |
| GA median SP | 92.81 | 92.75 |
| FH SE | 11.64 | 11.77 |
| FH SP | 93.85 | 93.79 |
| RW max SE | 14.03 | 13.49 |
| RW max SP | 92.29 | 92.17 |
| RW median SE | 13.21 | 12.75 |
| RW median SP | 92.15 | 91.98 |

Table 3: Results resume for testing dataset $\#42049$.

## ACKNOWLEDGMENT

The work has been supported by the grant of The University of West Bohemia, project No. SGS-2013-032 and by the European Regional Development Fund (ERDF), project New Technologies for Information Society (NTIS), European Centre of Excellence, ED1.1.00/02.0090.

## REFERENCES

[1] C. Darwin, *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*, 1st ed. London: John Murray, 1859.

[2] S. Baluja, *Population-Based Incremental Learning: A Method forIntegrating Genetic Search Based Function Optimization and Competetive Learning* Carnegie Mellon University, 1994.

[3] V. Hlavac et al., *Image Processing, Analysis, and Machine Vision*, 3rd ed. Thomson Learning, 2008

[4] P.F. Felzenszwalb and D.P. Huttenlocher, *Efficient graph-based image segmentation*, International Jornal of Computer Vision, 2004

[5] T. Bäck, *Evolutionary Algorithms in Theory and Practice*, Oxford University Press, 1996

[6] A. Vedaldi and S. Soatto, "Quick shift and kernel methods for mode seeking", S. European Conference on Computer Vision, 2008

[7] R. Achanta et al.,"SLIC Superpixels Compared to State-of-the-art Superpixel Methods", TPAMI, 2012

[8] J. Kenedy and R. Eberhart, "Particle Swarm Optimization", Purdue School of Engineering and Technology, Indianapolis, 1995

[9] Y. Boykov and M-P. Jolly, "Interactive Graph Cuts for Optimal Boundary and Region Segmentation of Objects in N-D Images", Proceedings of Internation Conference on Computer Vision, 2011

[10] D. Martin at al., "A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics", Proc. 8th Int'l Conf. Computer Vision, vol. 2 July 2001.

[11] L. Grady, "Random Walks for Image Segmentation", IEEE Transactions on pattern analysis and machine inteligence, 2006

Proceedings of the 9th Open German-Russian Workshop on Pattern Recognition and Image Understanding
Dec 1-5, 2014, Koblenz, Germany

# Hand-Eye Calibration of SCARA Robots

Markus Ulrich
MVTec Software GmbH
Neherstr. 1, 81675 München
www.mvtec.com
Telephone: +49 89 4576950
Email: ulrich@mvtec.com

Andreas Heider
MVTec Software GmbH[1]
Neherstr. 1, 81675 München
www.mvtec.com
Telephone: +49 89 4576950
Email: andreas@heider.io

Carsten Steger
MVTec Software GmbH
Neherstr. 1, 81675 München
www.mvtec.com
Telephone: +49 89 4576950
Email: steger@mvtec.com

***Abstract*—In SCARA robots, which are often used in industrial applications, all joint axes are parallel, covering three degrees of freedom in translation and one degree of freedom in rotation. Therefore, conventional approaches for the hand-eye calibration of articulated robots cannot be used for SCARA robots. In this paper, we present a new linear method that is based on dual quaternions and extends the work of [1] for SCARA robots. To improve the accuracy, a subsequent nonlinear optimization is proposed. We address several practical implementation issues and show the effectiveness of the method by evaluating it on synthetic and real data.**

## I. Introduction

SCARA (Selectively Compliant Arm for Robotic Assembly) robots ( [2], [3]) are used in many industrial applications. They differ from articulated (antropomorphic) robots in that their movements are more restricted. The arm of an articulated robot typically has 6 rotary joints that cover 6 degrees of freedom (3 translations and 3 rotations). In contrast, SCARA robots have at least 2 parallel rotary joints and 1 parallel prismatic joint that cover only 4 degrees of freedom (3 translations and 1 rotation). Fig. 1 shows typical SCARA setups with 3 rotary joints: The camera can either be mounted on the robot's end effector (tool) and is moved to different positions by it or it can be mounted stationary outside the robot. Compared to other robot types, SCARA robots offer faster and more precise performance. They are best suited for pick and place, packaging, and assembly applications, and are often preferred if only limited space is available.

We assume that the (moving or stationary) camera observes the workspace of the robot. Objects are detected and localized in the camera coordinate system. To be able to grasp the object, the object pose must be transformed into the base coordinate system of the robot. The process of determining the required transformation between the camera and the robot base coordinate systems for a stationary camera or between the camera and the robot tool coordinate system for a moving camera is called hand-eye calibration.

Fig. 1 shows the transformations that are involved in the hand-eye calibration process for the case of a stationary and a moving camera. ${}^{c2}\boldsymbol{H}_{c1}$ is a rigid 3D transformation, represented by a $4\times4$ homogeneous transformation matrix, that transforms 3D points from the coordinate system $c1$ into $c2$. For stationary cameras, the fixed and unknown transformations are ${}^{cam}\boldsymbol{H}_{base}$ and ${}^{tool}\boldsymbol{H}_{cal}$. Note that although the calibration object is rigidly attached to the robot tool, in general the exact relative pose of the calibration object with respect to the robot tool cannot be gauged accurately by hand, and hence is unknown. The remaining two transformations ${}^{base}\boldsymbol{H}_{tool}$ and ${}^{cam}\boldsymbol{H}_{cal}$ of the closed chain of transformations depend on the robot pose and are known from the robot kinematics and from the algorithm that determines the pose of the calibration object in the camera coordinate system based on a calibration image. For moving cameras, the fixed and unknown transformations are ${}^{cam}\boldsymbol{H}_{tool}$ and ${}^{base}\boldsymbol{H}_{cal}$, while the known transformations are the same as for stationary cameras.

In the following, we will concentrate on the case of a stationary camera. The case of a moving camera can be treated in an equivalent way. We assume that the robot is calibrated, i.e., ${}^{base}\boldsymbol{H}_{tool}$ is known accurately. Furthermore, we assume a calibrated camera, i.e., the interior camera parameters are known, and hence ${}^{cam}\boldsymbol{H}_{cal}$ can be computed [4, chapter 3.9]. For simplicity reasons, we also assume w.l.o.g. that all joint axes of the SCARA robot are parallel to the $z$ axis of the robot base and the tool coordinate systems.[2] This is in accordance with [3], where the $z$ axis of base and tool coordinate system are parallel to the robot joint axes. The input for the calibration is obtained by moving the robot's tool to $n$ different poses and for each pose acquiring an image of the calibration object that is attached to the tool. Thus, the input consists of one pair of calibration poses ${}^{base}\boldsymbol{H}_{tool}$ and ${}^{cam}\boldsymbol{H}_{cal}$ for each of the $n$ robot states.

It should be noted that the hand-eye calibration can also be performed without a calibration object by using approaches that are able to determine the 3D pose of arbitrary objects in a monocular image, e.g., [6] or [7]. Furthermore, instead of a camera, a 3D sensor can be used to observe the workspace of the robot. In this case, ${}^{cam}\boldsymbol{H}_{cal}$ can be obtained by approaches that are able to determine the pose of objects in 3D sensor data, like [8], [9], or [10].

### A. Hand-Eye Calibration of Articulated Robots

The hand-eye calibration of articulated robots is well understood. One of the most common problem formulations is based on closing the chain of transformations [11]:

$$ {}^{tool}\boldsymbol{H}_{cal} = {}^{tool}\boldsymbol{H}_{base}\,{}^{base}\boldsymbol{H}_{cam}\,{}^{cam}\boldsymbol{H}_{cal}. \qquad (1) $$

[1] The work emerged during an internship of Andreas Heider at MVTec Software GmbH

[2] In [5], is it shown that if the joint axes are not aligned with the $z$ axes, the coordinate system can simply be rotated to achieve alignment. Aligning the joint axes with the $z$ axis has the advantage that the direction of the indeterminate translation (see below) is well known.

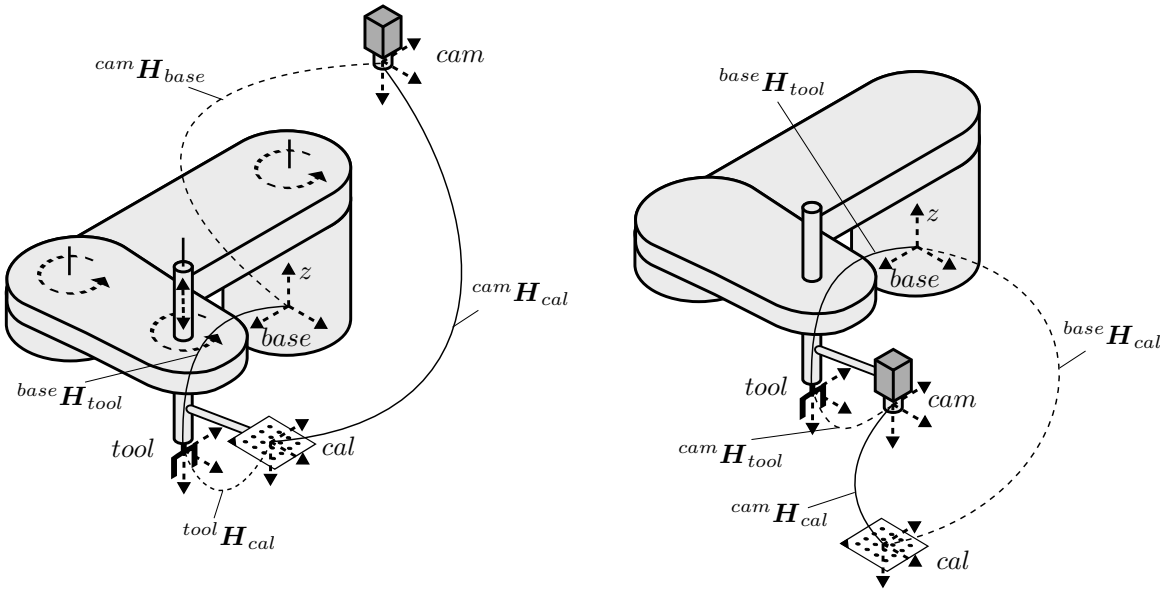



Fig. 1. Coordinate systems and transformations that are involved in the hand-eye calibration for the case of a stationary camera (left) and a moving camera (right). The coordinate systems are: camera ($cam$), robot base ($base$), calibration object ($cal$), robot tool ($tool$). Unknown transformations that are determined by the hand-eye calibration are visualized by dashed lines.

For each pair of calibration poses we set $\boldsymbol{A}_i = {}^{tool}\boldsymbol{H}_{base}$ and $\boldsymbol{B}_i = {}^{cam}\boldsymbol{H}_{cal}$, $i = 1 \dots n$. By setting the unknown poses to $\boldsymbol{X} = {}^{base}\boldsymbol{H}_{cam}$ and $\boldsymbol{Y} = {}^{tool}\boldsymbol{H}_{cal}$, we obtain $\boldsymbol{Y} = \boldsymbol{A}_i\boldsymbol{X}\boldsymbol{B}_i$. The pose $\boldsymbol{Y}$ can be eliminated by combining two pairs of poses from two different states $i$ and $j$, resulting in $\boldsymbol{A}_i\boldsymbol{X}\boldsymbol{B}_i = \boldsymbol{A}_j\boldsymbol{X}\boldsymbol{B}_j$. Rearranging yields $\boldsymbol{A}_j^{-1}\boldsymbol{A}_i\boldsymbol{X} = \boldsymbol{X}\boldsymbol{B}_j\boldsymbol{B}_i^{-1}$. The relative poses $\boldsymbol{A} = \boldsymbol{A}_j^{-1}\boldsymbol{A}_i$ and $\boldsymbol{B} = \boldsymbol{B}_j\boldsymbol{B}_i^{-1}$ represent the motions between the poses $i$ and $j$ of the tool and of the calibration object, respectively. Finally, the hand-eye calibration must solve

$$\boldsymbol{A}\boldsymbol{X} = \boldsymbol{X}\boldsymbol{B} \tag{2}$$

with respect to the unknown $\boldsymbol{X}$. Simple linear algorithms typically handle rotation and translation separately [12], [13]. This results in rotational errors propagating and increasing translational errors. More advanced linear methods, typically based on screw theory (see section II-A for a short introduction to screw theory), avoid this by solving for rotation and translation simultaneously [1], [14], [15]. Based on the results of linear methods, nonlinear optimizations can be used to further improve accuracy [1], [15], [16].

Another class of approaches computes $\boldsymbol{X}$ and $\boldsymbol{Y}$ simultaneously [17]–[19].

### B. Hand-Eye Calibration of SCARA Robots

Unfortunately, it is not possible to simply apply the methods for articulated robots to SCARA robots. In [13], the authors state that the error of the hand-eye calibration of articulated robots is inversely proportional to the sine of the angle between the screw axes (robot motions can be represented by screws, see below). Because the rotation axes of SCARA robots are parallel, all screw axes are parallel, too. Consequently, the error would be infinite. In particular, it is shown in [14] that if all screw axes are parallel, one parameter cannot be determined by the hand-eye calibration.

The hand-eye calibration of SCARA robots has drawn considerably less attention compared to articulated robots. In [20] and [21], linear methods are presented for solving the rotational part of the hand-eye calibration that work for robots with only one degree of freedom in rotation. Both approaches assume restricted specifically designed robot motions during calibration. In [5], a linear solution is presented that does not impose such restrictions on the robot motion while the calibration data is gathered. The linear solution is then used as an initial condition within an iterative optimization framework to further improve the accuracy. The author emphasizes that only up to five of the six unknown pose parameters can be determined and claims that the missing sixth parameter is irrelevant for the pose measurement of the sensor. Like [12] and [13], the linear method proposed in [5] solves for rotation and translation separately, which also results in the disadvantage that rotational errors propagate and increase translational errors.

In this paper, we extend the approach of [1], which is based on dual quaternions and handles rotation and translation simultaneously, for the calibration of SCARA robots. Furthermore, we show that for practical applications it is indeed necessary to know all of the six pose parameters. Therefore, we propose a pragmatic solution to determine the unknown sixth parameter. Finally, experiments on synthetic as well as on real data are presented.

## II. Hand-Eye Calibration of SCARA Robots Using Dual Quaternions

### A. Screw Theory

To better understand the relation between the motion of the tool and calibration object, screw theory can be used. It was first proposed for hand-eye calibration in [14]. A screw $(d, \theta, \vec{L})$ describes a 3D rigid transformation (rotation and translation) through a translation $d$ along the screw axis $\vec{L}$ and

a rotation by the rotation angle $\theta$ around the same axis. It is known from Chasles' theorem that any rigid 3D transformation can be described by a screw [1].

In [14], the motions $\boldsymbol{A}$ and $\boldsymbol{B}$ in (2) are expressed as screws. This allows to simultaneously solve for rotation and translation during hand-eye calibration. Because the calibration object is rigidly connected to the robot tool, the transformation between both coordinate systems is constant over time. The screws $\boldsymbol{A}$ and $\boldsymbol{B}$ represent the same rigid 3D transformation seen from different coordinate systems. The hand-eye calibration problem can be interpreted as finding the rigid transformation that aligns both screws [1]. The screw congruence theorem further tells us that the rotation angle $\theta$ and the translation $d$ of two screws are identical if one screw is the result of a rigid transformation applied to the other screw [14]. Consequently, the task of aligning screws reduces to aligning their screw axes, i.e., lines, in 3D space.

In [14], it is shown that at least two non-parallel robot motions are necessary to fix all degrees of freedom of the alignment of screw axes. For SCARA robots, all rotation axes, and hence all screw axes, are parallel to the $z$ axis of the robot base coordinate system. Two parallel motions only fix five of the six parameters of the unknown transformation between the base and camera coordinate system. Therefore, for SCARA robots, the translation $t_z$ along the $z$ axis of the robot base coordinate system is undetermined [5], [14].

Screw theory is the foundation of the calibration method for articulated robots presented in [1], which simultaneously solves for rotation and translation. It will be described and extended for SCARA robots in the next section.

### B. Dual Quaternions

Quaternions $\boldsymbol{q} = (q, \vec{q})$ with scalar part $q$ and vector part $\vec{q}$ are an extension of complex numbers to $\mathbb{R}^4$. For every rotation about an axis $\vec{n}$ ($\|\vec{n}\| = 1$) with an angle $\theta$, there exist two corresponding unit quaternions $\boldsymbol{q} = (\cos(\theta/2), \vec{n}\sin(\theta/2))$ and $-\boldsymbol{q}$ that each map a vector $\vec{p} \in \mathbb{R}^3$ to the rotated vector $\boldsymbol{q}(0, \vec{p})\bar{\boldsymbol{q}}$, where $\bar{\boldsymbol{q}}$ is the conjugate quaternion $(q, -\vec{q})$.

Dual numbers are defined as $\hat{z} = a + \epsilon b$ with the real part $a$, the dual part $b$, and the dual unit $\epsilon^2 = 0$. Dual 3-vectors $\hat{\vec{z}}$ with orthogonal real and dual parts are a representation of lines in $\mathbb{R}^3$, known as Plücker coordinates, where the real part is the direction of the line and the dual part is its moment. The line with direction $\vec{l}$ through the point $\vec{p}$ is $\hat{\vec{z}} = \vec{l} + \epsilon(\vec{p} \times \vec{l})$.

Dual quaternions are an extension of quaternions to dual numbers. A general introduction to dual quaternions can be found in [22], a brief outline is included in [1]. Similar to how unit quaternions are a tool to easily manipulate rotations in 3D space, dual unit quaternions can be used as a representation of rigid 3D transformations. Dual quaternions directly encode screws and have advantages over homogeneous transformation matrices when transforming lines in 3D [23].

A dual quaternion $\hat{\boldsymbol{q}} = \boldsymbol{q} + \epsilon\boldsymbol{q}'$ consists of the quaternions $\boldsymbol{q}$ (real part) and $\boldsymbol{q}'$ (dual part) with 4 elements each and is also often written as a single 8-vector. For unit dual quaternions, the following two conditions hold:

$$\boldsymbol{q}\bar{\boldsymbol{q}} = 1 \Rightarrow \boldsymbol{q}^\top \bar{\boldsymbol{q}} = 1 \tag{3}$$

and

$$\bar{\boldsymbol{q}}\boldsymbol{q}' + \boldsymbol{q}\bar{\boldsymbol{q}}' = 0 \Rightarrow \boldsymbol{q}^\top \boldsymbol{q}' = 0. \tag{4}$$

### C. Hand-Eye Calibration

Dual vectors (lines) can be written as pure dual quaternions (scalar part 0). The rigid transformations of the screw axes can be written concisely by using dual quaternions [23]:

$$\hat{\boldsymbol{a}} = \hat{\boldsymbol{x}}\hat{\boldsymbol{b}}\bar{\hat{\boldsymbol{x}}}, \tag{5}$$

where $\hat{\boldsymbol{a}} = \boldsymbol{a} + \epsilon\boldsymbol{a}'$ and $\hat{\boldsymbol{b}} = \boldsymbol{b} + \epsilon\boldsymbol{b}'$ are unit dual quaternions that represent the screws for the tool and the calibration object, respectively, for a single robot motion. $\hat{\boldsymbol{x}} = \boldsymbol{x} + \epsilon\boldsymbol{x}'$ is a unit dual quaternion that represents the unknown transformation $\boldsymbol{X}$. In [1], the definition of dual quaternion multiplication is used to rewrite (5) as

$$\boldsymbol{S}\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{x}'\end{pmatrix} = 0 \tag{6}$$

with the $6 \times 8$ matrix

$$\boldsymbol{S} = \begin{pmatrix}\vec{a} - \vec{b} & [\vec{a} + \vec{b}]_\times & \vec{0} & \vec{0}\\ \vec{a}' - \vec{b}' & [\vec{a}' + \vec{b}']_\times & \vec{a} - \vec{b} & [\vec{a} + \vec{b}]_\times\end{pmatrix} \tag{7}$$

encoding the result of the dual quaternion multiplication, where $\boldsymbol{a} = (0, \vec{a})$, $\boldsymbol{a}' = (0, \vec{a}')$, $\boldsymbol{b} = (0, \vec{b})$, and $\boldsymbol{b}' = (0, \vec{b}')$ represent the screw axes. Stacking the matrices $\boldsymbol{S}_i$ from all $m$ motions results in the $6m \times 8$ matrix $\boldsymbol{T}^\top = \left(\boldsymbol{S}_1^\top \boldsymbol{S}_2^\top \dots \boldsymbol{S}_n^\top\right)^\top$, which can be decomposed by the SVD as:

$$\boldsymbol{T} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top. \tag{8}$$

For articulated robots, $\boldsymbol{T}$ has rank 6, and hence results in two vanishing singular values with the two corresponding right singular vectors $\vec{v}_7$ and $\vec{v}_8$. The set of solutions is the null space of $\boldsymbol{T}$:

$$\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{x}'\end{pmatrix} = \lambda_1\vec{v}_7 + \lambda_2\vec{v}_8. \tag{9}$$

The remaining two degrees of freedom can be fixed by substituting the two unity constraints (3) and (4) into (9). This leads to a system of two quadratic polynomials in $\lambda_1$ and $\lambda_2$, which can be directly solved for $\hat{\boldsymbol{x}}$ [1].

For SCARA robots, the rank of the matrix $\boldsymbol{T}$ is 5 in the noise-free case. This results in three vanishing singular values and three matching right singular vectors $\vec{v}_6$, $\vec{v}_7$, and $\vec{v}_8$, yielding the set of solutions:

$$\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{x}'\end{pmatrix} = \lambda_1\vec{v}_6 + \lambda_2\vec{v}_7 + \lambda_3\vec{v}_8. \tag{10}$$

We know that there exists a 1D space of equivalent solutions, i.e., solutions with identical algebraic error. Therefore, it is sufficient to obtain one arbitrary solution out of this set (i.e., a solution for an arbitrary $t_z$). For this, we set $\lambda_3 = 0$, and hence reduce (10) to (9), which again can be solved as describe above. A geometric interpretation shall justify this approach: From (10), we know that we have a 3D solution space spanned by the three 8D vectors. The constraints (3) and (4) represent a 1D manifold within this solution space. Setting $\lambda_3$ to 0 restricts the original 3D solution space to a 2D hyperplane. The intersection of the 1D manifold with the 2D hyperplane results in a single solution. This assumes that the 1D manifold always intersects the 2D hyperplane. In [1], it is shown that

(9) always has a solution. Since we reduced the SCARA case from (10) to (9), it must have a solution, too. It follows that the 1D manifold always intersects the 2D hyperplane.

In section III-E, we will present methods to determine the real value of the undetermined $t_z$, which is essential for most practical applications.

## III. Implementation

### A. Selection of Robot Motions

In [24], it is pointed out that combining poses into motions in their chronological order is not the best choice in terms of numerical stability. Instead, it is proposed to select combinations of poses such that the orientation of the rotation axes of both poses are maximally different. While this criterion is not applicable for the calibration of SCARA robots, the basic idea is still useful.

The first criterion that we apply is the rotation angle of the screws. If the rotation angle is small, the screw axis is not well-defined and might be unstable in noisy conditions. Therefore, robot motions with larger rotation angles should be preferred.

The second criterion is again based on the screw congruence theorem, which for dual quaternions requires that the scalar parts of $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$ are equal. Because of measurement errors, inaccuracies of the robot, and noise, this condition is not perfectly fulfilled. Therefore, motion pairs $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$ with smaller differences in their scalar parts should be preferred.

In [24], a score is computed for all possible pose pairs. Then, the pairs with highest scores are selected for calibration. This may result in an unfavorable weighting, where some robot poses are represented multiple times while others are completely ignored. We propose the following approach instead: To apply the first criterion, for each pose $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ of the $n$ robot poses, find the robot poses $\boldsymbol{A}_j$ and $\boldsymbol{B}_j$ with maximum rotation angle $|r_a| + |r_b|$, where $r_a$ and $r_b$ are the screw angles of the corresponding robot motions $\boldsymbol{A} = \boldsymbol{A}_j^{-1}\boldsymbol{A}_i$ and $\boldsymbol{B} = \boldsymbol{B}_j\boldsymbol{B}_i^{-1}$. This results in $n$ robot motions. To apply the second criterion, for each pose pair $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$, find the pose pair $\boldsymbol{A}_j$ and $\boldsymbol{B}_j$ for which the rotation and translation components of the corresondings screws $\boldsymbol{A}$ and $\boldsymbol{B}$ differ the least, i.e., for which the scalar parts of $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$ differ the least. Hence, after eliminating duplicates, up to $2n$ robot motions are selected for calibration.

### B. Antiparallel Screw Axes

If the camera is mounted such that its $z$ axis is parallel to the $z$ axis of the base coordinate system, all vectors $\vec{a}$ are either $(0, 0, s)^\top$ or $(0, 0, -s)^\top$ (with $s = \sin(\theta/2)$) depending on whether the $z$ axes point in the same or in opposite directions. Furthermore, all vectors $\vec{b}$ are always either $(0, 0, s)^\top$ or $(0, 0, -s)^\top$. Equation (7) involves a vector product with the vector $\vec{a} + \vec{b}$. Therefore, it might happen that all $\vec{a}$ are antiparallel to the corresponding $\vec{b}$, and hence their sum vanishes. In this case, $\boldsymbol{T}$ degrades to rank 4. For articulated robots, in general the screw axes of most robot motions are not parallel, even if the $z$ axes of the coordinate systems are. For SCARA robots, if all screw axes $\vec{a}$ and $\vec{b}$ point in opposite directions ($\vec{a}^\top\vec{b} < 0$), we propose to applying a rotation $\hat{\boldsymbol{r}}$ of $180\,^\circ$ around the $x$ axis to both sides of (5), yielding $\hat{\boldsymbol{r}}\hat{\boldsymbol{a}}\bar{\hat{\boldsymbol{r}}} = \hat{\boldsymbol{r}}\hat{\boldsymbol{x}}\hat{\boldsymbol{b}}\bar{\hat{\boldsymbol{x}}}\bar{\hat{\boldsymbol{r}}}$. By substituting $\hat{\boldsymbol{a}}^\star := \hat{\boldsymbol{r}}\hat{\boldsymbol{a}}\bar{\hat{\boldsymbol{r}}}$ and $\hat{\boldsymbol{x}}^\star := \hat{\boldsymbol{r}}\hat{\boldsymbol{x}}$, we obtain a modified problem formulation $\hat{\boldsymbol{a}}^\star = \hat{\boldsymbol{x}}^\star\hat{\boldsymbol{b}}\bar{\hat{\boldsymbol{x}}}^\star$ with $\hat{\boldsymbol{a}}^\star$ and $\hat{\boldsymbol{b}}$ now pointing to the same half space. After performing the hand-eye calibration, the solution of the original formulation can be obtained by $\hat{\boldsymbol{x}} = \bar{\hat{\boldsymbol{r}}}\hat{\boldsymbol{x}}^\star$.

### C. Sign Ambiguity of Screws

In the same way that two unit quaternions $\boldsymbol{q}$ and $-\boldsymbol{q}$ represent the same rotation, every rigid transformation can be described by two screws $\hat{\boldsymbol{q}}$ and $-\hat{\boldsymbol{q}}$. Extracting the translation and rotation components from both screws results in $d^+ = -d^-$ and $\theta^+ = 2\pi - \theta^-$. The screw congruence theorem that is assumed in (5) requires $d_a = d_b$ and $\theta_a = \theta_b$. Consequently, it is important to choose the signs of the screws $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$ consistently.

In [14], two approaches for solving the sign ambiguity are proposed. The first orients the screw axis such that $d > 0$. For $d$ close to 0, this method becomes unstable. The second approach constrains the rotation angle $\theta$ to $0 \leq \theta < \pi$, which becomes unstable for $\theta$ close to 0 or $\pi$. In [1], the scalar parts, which represent both $\theta$ and $d$, are checked for equality for all robot motions. Because for SCARA robots all screw axes are parallel, it is sufficient to determine the sign only for a single robot motion and fix the signs of all other motions accordingly. For this, we select the most unambiguous motion pair as the reference motion, i.e., the motion pair for which $\theta$ differs most from 0 and $\pi$. Then, the screw axes of all other motions are compared to the axis of the reference motion and flipped if necessary.

### D. Refinement by Nonlinear Optimization

The resulting poses of the described linear approach can be used as initial values in a nonlinear optimization framework to further increase the accuracy. For this, we compute the error matrix

$$\boldsymbol{E} = {}^{tool}\boldsymbol{H}_{cal} - {}^{tool}\boldsymbol{H}_{base}\,{}^{base}\boldsymbol{H}_{cam}\,{}^{cam}\boldsymbol{H}_{cal} \tag{11}$$

and minimize

$$e = \sum_{i=0}^{n} \operatorname{tr}(\boldsymbol{E}_i\boldsymbol{W}\boldsymbol{E}_i^\top), \quad \text{where } \boldsymbol{W} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 9 \end{pmatrix}, \tag{12}$$

over all $n$ robot poses by using the Levenberg-Marquardt algorithm. The matrix $\boldsymbol{W}$ balances the different number of entries in $\boldsymbol{E}$ for rotation and translation (9 and 3, respectively). Furthermore, the translation part of all input matrices is scaled by $1/D$, where $D$ is the maximum extent of the workspace defined by the robot poses. This ensures that the results are scale-invariant and errors in rotation and translation are weighted appropriately. Because $t_z$ cannot be determined, the minimization is performed by varying only 11 of the 12 pose parameters and leaving $t_z$ fixed.

### E. Determination of the Real $t_z$

For most practical applications, it is essential to know $t_z$. Otherwise, objects for which the pose was determined in the camera coordinate system cannot be grasped by the robot. The following approach works well in practice:

Let us assume that ${}^{tool}\boldsymbol{H}_{cal}$ and ${}^{base}\boldsymbol{H}_{cam}$ (see (1)) are known from the calibration up to the $z$ component of the translation part of ${}^{tool}\boldsymbol{H}_{cal}$. To determine $z$, the calibration object is detached from the robot and placed at an arbitrary position such that it can be observed by the camera. The pose of the calibration object is then automatically determined in the camera coordinate system to obtain ${}^{cam}\tilde{\boldsymbol{H}}_{cal}$. From this, we can compute the $z$ component of the translation of ${}^{base}\tilde{\boldsymbol{H}}_{cal} = {}^{base}\boldsymbol{H}_{cam}\,{}^{cam}\tilde{\boldsymbol{H}}_{cal}$, which we will denote $z_{calib}$. Then, the tool of the robot is manually moved to the origin of the calibration object, which gives us ${}^{tool}\tilde{\boldsymbol{H}}_{base}$. The $z$ component of the translation of ${}^{base}\tilde{\boldsymbol{H}}_{tool}$ represents the true translation and is denoted as $z_{true}$. $z_{true}$ and $z_{calib}$ represent the same physical distance, and hence must be identical. This can be achieved by modifying the $z$ component of ${}^{tool}\boldsymbol{H}_{cal}$ by $z_{true} - z_{calib}$. Finally, ${}^{base}\boldsymbol{H}_{cam}$ can be computed by closing the chain of transformations (1).

The case of a moving camera can be treated similarly. However, for some setups it is not possible for the camera to observe the calibration object if the tool is moved to the origin of the calibration object. Let us assume that ${}^{base}\boldsymbol{H}_{cal}$ and ${}^{cam}\boldsymbol{H}_{tool}$ are known from the calibration up to the $z$ component of the translation part of ${}^{base}\boldsymbol{H}_{cal}$. In this case, the robot is manually moved to two poses. First, the tool is moved such that the camera can observe the calibration object. Now, an image of the calibration object is acquired and the tool pose is queried, which gives us ${}^{cam}\tilde{\boldsymbol{H}}_{cal}$ and ${}^{base}\tilde{\boldsymbol{H}}_{tool}$. Second, the tool of the robot is moved to the origin of the calibration object, yielding ${}^{base}\tilde{\boldsymbol{H}}_{tool}$. $z_{true}$ is the $z$ component of the translation of ${}^{tool}\tilde{\boldsymbol{H}}_{base}\,{}^{base}\tilde{\boldsymbol{H}}_{tool}$. $z_{calib}$ is the $z$ component of the translation of ${}^{tool}\boldsymbol{H}_{cam}\,{}^{cam}\tilde{\boldsymbol{H}}_{cal}$. Again, $z_{true} - z_{calib}$ can be used to correct the $z$ component of ${}^{base}\boldsymbol{H}_{cal}$.

Actually, it is sufficient in the above approaches to move the tool to a point with the same $z$ coordinate in the base coordinate system as the origin of the calibration object. Sometimes, however, the origin or even a point with the same $z$ coordinate cannot be reached by the tool. In this case, the tool should be moved to a point with known height (i.e., vertical distance in $z$ direction of the base coordinate system) above or below the origin. The $z$ component of the transformation must additionally be corrected by this height.

## IV. Experimental Results

We tested our algorithms on synthetic and real data. For the experiments with synthetic data, we simulated different setups with a stationary camera, which was mounted about 1.5 m above the robot base and a calibration object that was attached to the robot tool at a distance of about 0.2 m. The robot poses were randomly created within a workspace of size $0.6 \times 0.6 \times 0.6\,\mathrm{m}^3$.

In the first experiment, we investigated the influence of noise on the accuracy of the calibration. For this, we added noise of different amplitudes to the position and rotation components of ${}^{cam}\boldsymbol{H}_{cal}$, while assuming error-free robot poses ${}^{tool}\boldsymbol{H}_{base}$. For each noise amplitude, we created 15 random robot poses, performed the calibration, and repeated the experiment 500 times. Fig. 2(a) and (b) show the mean position and rotation errors of the resulting pose ${}^{base}\boldsymbol{H}_{cam}$ for the linear approach using dual quaternions and for the nonlinear optimization. All errors increase linearly with the noise amplitude. Furthermore, the advantage of the subsequent nonlinear optimization is clearly visible.

In the second experiment, we investigated the influence of the number of robot poses on the accuracy of the calibration. For this, we kept the noise magnitude fixed at 0.4 mm and 0.1 ° and repeated the experiment 500 times again. Fig. 2(c) and (d) show the corresponding mean errors. It should be noted that the errors rapidly decrease for the first 10 images. Further increasing the number of poses only slightly improves the accuracy.

For the experiments with real data, we attached a HALCON calibration plate to the tool of a DENSO robot HS-45452E/GM. A calibrated camera (IDS uEye UI-2240-M, 1/2", $1280 \times 1024$, $f$ =16 mm) was mounted stationary 0.9 m above the robot. At each of 12 calibration poses, we acquired an image of the calibration plate and determined ${}^{cam}\boldsymbol{H}_{cal}$ by using HALCON [4, chapter 3.9]. The hand-eye calibration using the linear approach results in RMS / maximum errors of 0.30 mm / 0.60 mm in translation and 0.06 ° / 0.11 ° in rotation. The nonlinear optimization further decreases the errors to 0.20 mm / 0.31 mm in translation and 0.05 ° / 0.10 ° in rotation.

## V. Conclusions and Outlook

In [1] a method for hand-eye calibration of articulated robots based on dual-quaternions is proposed, which shows advantages over methods that handle rotation and translation separately. In this paper, we extended the work of [1] to SCARA robots. For this, we argued why it is feasible to reduce the three degrees of freedom in (10) by setting $\lambda_3 = 0$ to the original problem with two degrees of freedom, which can be solved easily. In future work we will further justify this aproach by giving a sound mathematical proof.

To further improve the accuracy, we proposed a subsequent nonlinear optimization. We addressed several practical implementation issues and showed the effectiveness of the method by evaluating it on synthetic and real data. It was already shown by [1] that the dual-quaternions-based method is supperior to the separate estimation of rotation and translation. Nevertheless, in the future we will extend our evaluation by comparing our dual-quaternion-based approach to the linear approach proposed in [5].

## References

[1] K. Daniilidis, "Hand-eye calibration using dual quaternions," *International Journal of Robotics Research*, vol. 18, no. 3, pp. 286–298, 1999.

[2] ISO 8373:1994, "Manipulating industrial robots — Vocabulary," Geneva, Switzerland, 1994.

[3] ISO 9787:1999, "Manipulating industrial robots — Coordinate systems and motion nomenclatures," Geneva, Switzerland, 1999.

[4] C. Steger, M. Ulrich, and C. Wiedemann, *Machine Vision Algorithms and Applications*. Weinheim: Wiley-VCH, 2007.

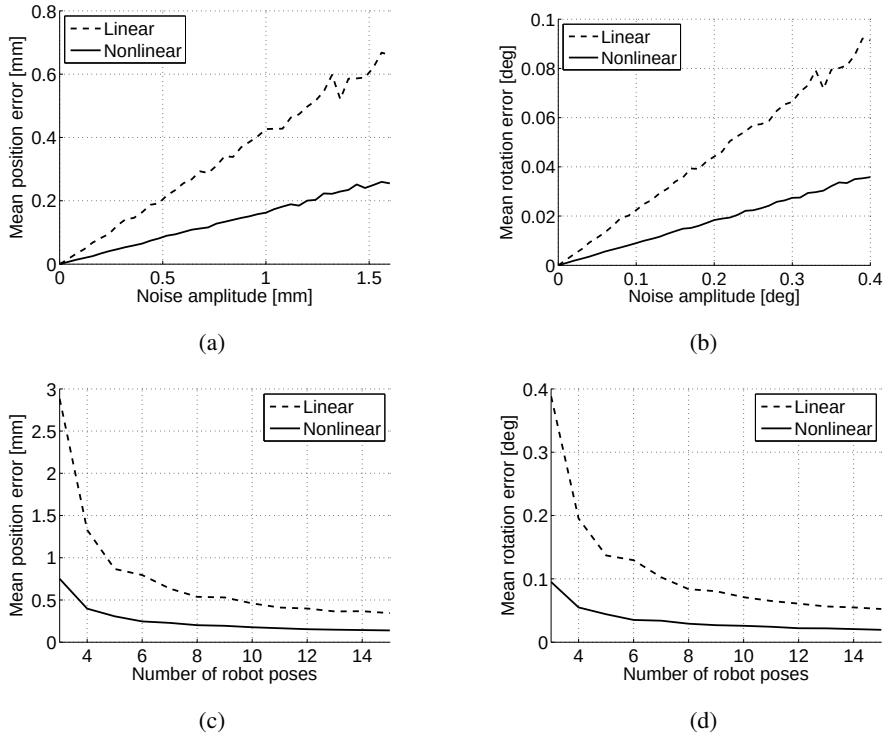



Fig. 2. Mean error in position (a) and rotation (b) of the hand-eye calibration for different noise levels and 15 random robot poses when using the linear approach using dual quaternions (dashed line) and when performing a subsequent nonlinear optimization (solid line); Mean error in position (c) and rotation (d) for different number of robot poses and a constant noise level.

[5] H. Zhuang, "Hand/eye calibration for electronic assembly robots," *IEEE Transactions on Robotics and Automation*, vol. 14, no. 4, pp. 612–616, 1998.

[6] S. Hinterstoisser, S. Benhimane, and N. Navab, "N3M: Natural 3d markers for real-time object detection and pose estimation," in *IEEE International Conference on Computer Vision*, 2007, pp. 1–7.

[7] M. Ulrich, C. Wiedemann, and C. Steger, "Combining scale-space and similarity-based aspect graphs for fast 3d object recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 10, pp. 1902–1914, 2012.

[8] B. Drost and S. Ilic, "3d object detection and localization using multimodal point pair features," in *International Conference on 3D Imaging, Modelling, Processing, Visualization and Transmission (3DIMPTV)*, 2012, pp. 9–16.

[9] B. Drost, M. Ulrich, N. Navab, and S. Ilic, "Model globally, match locally: Efficient and robust 3d object recognition," in *Computer Vision and Pattern Recognition*, 2010, pp. 998–1005.

[10] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3d scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 5, pp. 433–449, 1999.

[11] K. H. Strobl and G. Hirzinger, "Optimal hand-eye calibration," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006, pp. 4647–4653.

[12] J. C. K. Chou and M. Kamel, "Finding the position and orientation of a sensor on a robot manipulator using quaternions," *International Journal of Robotics Research*, vol. 10, no. 3, pp. 240–254, 1991.

[13] R. Y. Tsai and R. K. Lenz, "A new technique for fully autonomous and efficient 3d robotics hand/eye calibration," *IEEE Transactions on Robotics and Automation*, vol. 5, no. 3, pp. 345–358, 1989.

[14] H. H. Chen, "A screw motion approach to uniqueness analysis of head-eye geometry," in *Computer Vision and Pattern Recognition*, 1991, pp. 145–151.

[15] R. Horaud and F. Dornaika, "Hand-eye calibration," *International Journal of Robotics Research*, vol. 14, no. 3, pp. 195–210, 1995.

[16] B. Kaiser, R. A. Tauro, and H. Wörn, "Extrinsic calibration of a robot mounted 3d imaging sensor," *International Journal of Intelligent Systems Technologies and Applications*, vol. 5, no. 3/4, pp. 374–379, 2008.

[17] F. Dornaika and R. Horaud, "Simultaneous robot-world and hand-eye calibration," *IEEE Transactions on Robotics and Automation*, vol. 14, no. 4, pp. 617–622, 1998.

[18] S. Rémy, M. Dhome, J. M. Lavest, and N. Daucher, "Hand-eye calibration," in *International Conference on Intelligent Robots and Systems*, 1997, pp. 1057–1065.

[19] H. Zhuang, Z. S. Roth, and R. Sudhakar, "Simultaneous robot/world and tool/flange calibration by solving homogeneous transformation equations of the form AX = YB," *IEEE Transactions on Robotics and Automation*, vol. 10, no. 4, pp. 549–554, 1994.

[20] L. J. Everett and M. S. Burnett, "Experimentally registering position sensors," in *IEEE International Conference on Robotics and Automation*, 1996, pp. 641–646.

[21] S. D. Ma, "A self-calibration technique for active vision systems," *IEEE Transactions on Robotics and Automation*, vol. 12, no. 1, pp. 114–120, 1996.

[22] B. Kenwright, "Dual-quaternions from classical mechanics to computer graphics and beyond," 2012, online; accessed 2014-05-08, 11 pages. [Online]. Available: www.xbdev.net/misc_demos/demos/dual_quaternions_beyond/paper.pdf

[23] J. Rooney, "A comparison of representations of general spatial screw displacement," *Environment and Planning B*, vol. 5, no. 1, pp. 45–88, 1978.

[24] J. Schmidt, F. Vogt, and H. Niemann, "Robust hand-eye calibration of an endoscopic surgery robot using dual quaternions," in *Pattern Recognition*, ser. Lecture Notes in Computer Science. Berlin: Springer-Verlag, 2003, pp. 548–556.

# Hierarchical ensemble clustering algorithm for multispectral image segmentation*

*S.Rylov, I.Pestunov*
Institute of Computational Technologies SB RAS
Novosibirsk, Russia
RylovS@mail.ru, pestunov@ict.nsc.ru

*V. Berikov*
Sobolev Institute of Mathematics SB RAS
Novosibirsk, Russia
berikov@math.nsc.ru

***Abstract*—A novel hierarchical clustering algorithm based on nonparametric estimation of the global probability density function of the data points is proposed. Special similarity metric is introduced to deal with overlapping classes. Ensemble approach allows combining multiple hierarchical partitionings and improving the quality of results in exploring complicated hierarchical structures. High computing efficiency allowing interactive multispectral satellite image processing is achieved by the use of grid-based approach. Experimental results on both synthetic and real datasets demonstrate the effectiveness of the proposed algorithm.**

***Keywords—hierarchical ensemble clustering; grid-based approach; density-based approach; fast image segmentation; multispectral satellite images***

## I. Introduction

One of the most important stages of the digital image analysis is segmentation. It consists in image partitioning into non-overlapping areas on the basis of similarity of spectral or spatial (texture, size, shape, etc.) characteristics. The most common approach to the satellite images segmentation is based on the data clustering algorithms [1].

Clustering methods could be divided into two large groups: partitional and hierarchical. The partitional approach produces a single (flat) partition of the data points, while the hierarchical approach gives a nested clustering results in the form of a dendrogram (cluster tree), which allows to obtain different levels of partitions. The hierarchical clustering scheme is a popular choice when different levels of cluster structure are desired or when the exact number of the clusters could not be determined. However, it is generally recognized that the traditional hierarchical methods have some limitations. For example, Single Linkage method produces the "chaining" effect, while Complete Linkage and Average Linkage usually work well only for spherical-shaped clusters. Besides, these methods often fail to work with overlapping clusters [2]. In addition, they are usually too slow to be applied to multispectral images, which present large-scale datasets.

It's shown [3,4] that the ensemble approach improves the quality of the hierarchical clustering methods. However, these methods are also extremely computationally expensive.

In this paper, a new computationally efficient hierarchical clustering algorithm based on grid-based approach and the ensemble method of combining obtained hierarchical partitionings is proposed. This work extends the previous research [5].

* Partially supported by the Russian Foundation for Basic Research (grants no. 14-07-00249a, 14-07-31320-мол-а) and Russian Science Foundation (grant no. 14-14-00453).

## II. Basic Notions

Let the set of objects $X$ be classified consists of $d$-dimensional vectors lying in the attribute space $R^d$: $X = \{x_i = (x_i^1, \ldots, x_i^d) \in R^d, i = 1,\ldots,N\}$. Vectors $x_i$ are bounded by a hyper-rectangle $\Omega = [l^1, r^1]\times\ldots\times[l^d, r^d]$, where $l^j = min\ x_i^j$, $r^j = max\ x_i^j$, $x_i \in X$. *Grid structure* is formed by dividing $\Omega$ with hyperplanes $x^j = (r^j - l^j) \cdot i / m + l^j$, $i = 0,\ldots,m$ where $m$ is the number of partitions in each dimension. Here the minimum structure element is a cell (a closed hyper-rectangle bounded by hyperplanes). A common numbering of the cells is introduced (sequentially, one layer of cells after another).

Cells $B_i$ and $B_j$ ($i \neq j$) are *adjacent* if their intersection is not empty. The set of cells adjacent to $B$ will be denoted as $A_B$. The *density* $D_B$ of cell $B$ is defined by $D_B = N_B / V_B$ where $N_B$ is the number of elements from $X$ belonging to the cell $B$; $V_B$ is the volume of $B$. Cell $B$ is assumed to be *nonempty* if $D_B > 0$.

The nonempty cell $B_i$ is *directly connected* to the nonempty cell $B_j$ ($B_i \rightarrow B_j$) if $B_j$ is a cell with the maximum number that satisfies the conditions $B_j = arg\ max_{B_k \in A_{B_i}} D_{B_k}$ and $D_{B_j} \geq D_{B_i}$. The nonempty cells $B_i$ and $B_j$ are *directly connected* ($B_i \leftrightarrow B_j$) if $B_i \rightarrow B_j$ or $B_j \rightarrow B_i$. The nonempty cells $B_i$ and $B_j$ are *connected* ($B_i \sim B_j$) if there exist $k_1,\ldots,k_l$ such that $k_1 = i$, $k_l = j$ and for all $p = 1,\ldots,l-1$ we have $B_{k_p} \leftrightarrow B_{k_{p+1}}$. The introduced connectedness relation provides a natural partition of nonempty cells into the connected components $\{G_1,\ldots,G_S\}$. The connected component is defined as the maximum set of cells connected to each other. *Representative cell Y(G)* of the connected component $G$ is defined as a cell with the maximum number that satisfies the condition $Y(G) = arg\ max_{B \in G} D_B$. The connected components $G_i$ and $G_j$ are *adjacent* if there exist adjacent cells $B_i$ and $B_j$ such that $B_i \in G_i$ and $B_j \in G_j$.

## III. Hierarchical Similarity Metric

We define the distance between adjacent connected components $G_i$ and $G_j$ as

$$h_{ij} = min_{P_{ij} \in \Re_{ij}} [\ 1 - min_{B_{k_t} \in P_{ij}} D_{B_{k_t}} / min(D_{Y_i}, D_{Y_j})\ ]\ .$$

Here $\Re_{ij} = \{P_{ij}\}$ is a set of all possible paths between $Y(G_i)$ and $Y(G_j)$, $P_{ij} = \langle Y(G_i) = B_{k_1}, \ldots, B_{k_t}, B_{k_{t+1}}, \ldots, B_{k_l} = Y(G_j) \rangle$ such that for all cells of the path $t = 1, \ldots, l-1$: 1) $B_{k_t} \in G_i \cup G_j$; 2) $B_{k_t}, B_{k_{t+1}}$ are adjacent cells.

Let's define $\Theta(G_i, G_j) = \Theta_{ij} = \{Q_{ij}\}$ as a set of all possible paths consisting of the connected components $Q_{ij} = \langle G_i = G_{k_1}, \ldots, G_{k_t}, G_{k_{t+1}}, \ldots, G_{k_l} = G_j \rangle$ such that for all $t = 1, \ldots, l-1$: $G_{k_t}, G_{k_{t+1}}$ are adjacent connected components. Now the distance between arbitrary connected components $G_i$ и $G_j$ is defined as:

$$\hat{h}_{ij} = min_{\,Q_{ij} \in \Theta_{ij}} [max_t\, h_{k_t,k_{t+1}}] .$$

For an empty set $\Theta_{ij}$ we assume $\hat{h}_{ij} = 1$.

Constructed distance measure $\{\hat{h}_{ij}\}$ is an ultrametric [6], which means it is a metric satisfying the inequality $\hat{h}_{ij} \le max(\hat{h}_{ik}, \hat{h}_{kj}), \forall\, i, j, k$. Each ultrametric dissimilarity matrix corresponds to a single dendrogram [4]. Therefore, hierarchical result in form of a dendrogram could be generated from ultrametric dissimilarity matrix.

Computation of the ultrametric $\{\hat{h}_{ij}\}$ from the distance $\{h_{ij}\}$ is in fact a min-transitive closure operation [3]. This computing step could be implemented via single-linkage clustering method (SLINK) inheriting the complexity of $O(n^2)$, where $n$ is a number of connected components.

Lemma. Dendrogram constructed from the dissimilarity matrix $\{h_{ij}\}$ by SLINK method coincides with the dendrogram constructed from the ultrametric matrix $\{\hat{h}_{ij}\}$.

The proposed clustering scheme highlights the hierarchical structure of the clusters. The base elements of hierarchical structure consist of the connected components instead of original data elements. The relatively small number of connected components allows constructing the hierarchical result with minimal computational cost.

Experimental studies have shown that the results of the proposed algorithm are considerably dependent on the parameter $m$, which determines the size of a grid structure element. Therefore, the ensemble approach is applied to improve the quality and the robustness of clustering.

## IV. Hierarchical Clustering Combination

The problem of instability of the results comes from the following fact. A fine grid causes a large number of connected components and the growing noise influence, while on the other hand coarse grid causes problems with the accurate partitioning of overlapping classes.

The ensemble approach is applied to improve the quality of clustering results by efficient combination of multi-scale information. However, most of the existing ensemble techniques are designed for partitional clustering methods, while only few research efforts have been reported for hierarchical ones [3,4].

In this paper a new ensemble hierarchical clustering algorithm ECCAH is proposed. It combines several results of the proposed above algorithm with different values of the grid-size parameter $m$. The collective solution is obtained by estimating the average distances. It allows considering information from different grids equally and to provide the most proper estimation of the data structure.

Thereby, the proposed algorithm runs $L$ times with different values of $m$. Consequently, we obtain $L$ ultrametric distance matrices $\{\hat{h}_{ij}^{(1)}\}, \ldots, \{\hat{h}_{ij}^{(L)}\}$. Consolidated connectivity matrix $\{H_{ij}\}$ (it's size matches the size of $\{\hat{h}_{ij}^{(L)}\}$ which is the finest grid used) is constructed as follows:

$$H_{ij} = \frac{1}{L} \sum_{k=1}^{L} \hat{h}^{(k)}(G_i^{(k)}, G_j^{(k)})$$

Here $G_i^{(k)}$ is the connected component containing the representative cell of the connected component $G_i^{(L)}$ from the $k$-th run. So the problem of matching the labels from different partitions is solved via representative cells of connectedness components.

Obtained consolidated connectivity matrix may be non-ultrametric. To construct the final hierarchical result average linkage clustering method (UPGMA) is applied to $\{H_{ij}\}$. Experimental studies have shown the superiority of the UPGMA over SLINK on this stage.

The proposed algorithm ECCAH improves the quality of clustering results and its stability to the grid-size parameter change, which is confirmed by experimental studies. Fig. 1 shows clustering accuracy improvement with the growing number of the used grids. Grid-size parameter $m$ was taken from the set $\{m, m+2, \ldots, m+2\cdot(L-1)\}$.

## V. Experiments

The proposed methods were investigated on numerous simulated datasets and real images. All experiments were performed on Intel Core i7 3.2 GHz CPU. Software framework ELKI [7] was used to compare the proposed method with other algorithms in terms of quality and speed. It includes well-known clustering algorithms (e.g. k-means, EM, DBSCAN, OPTICS, DeLiClu, SLINK).

Fig. 3 (a) shows a model dataset consisting of eight normally-distributed clusters. Clusters are grouped into three well-separated groups, one of them contains significant class overlapping. The proposed ECCAH algorithm is capable of separating all eight clusters (Fig. 3 (b)) as well as exposing the hierarchical structure of the data (Fig. 3 (b-d)).

Complex model dataset containing 8 clusters different in shape and size (including normally-distributed clusters, rings and spirals) is shown in Fig. 2 (a). Fig. 2 (c-d) demonstrates the results of nonparametric clustering algorithms OPTICS, DeLiClu and SLINK (the most optimal parameters were chosen). The proposed ensemble method ECCAH achieves clustering accuracy 99.3% on this model (Fig. 2 (b)).

Fig. 4 shows WorldView-2 satellite image and the clustering result of ECCAH algorithm. Four spectral channels (1, 4, 6, 7) were used. Image size is 2048×2048 pixels. Clustering time is only 0.4 s.

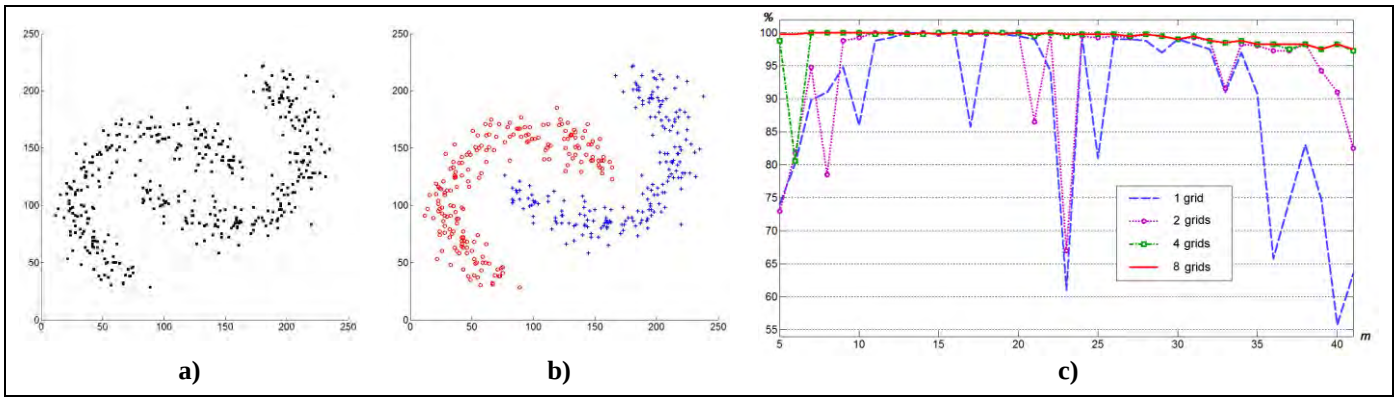



Fig. 1. Clustering results for model dataset "Bananas": original dataset (a), correct clustering (b) and ECCAH accuracy against the initial m value for different number of used grids (c).

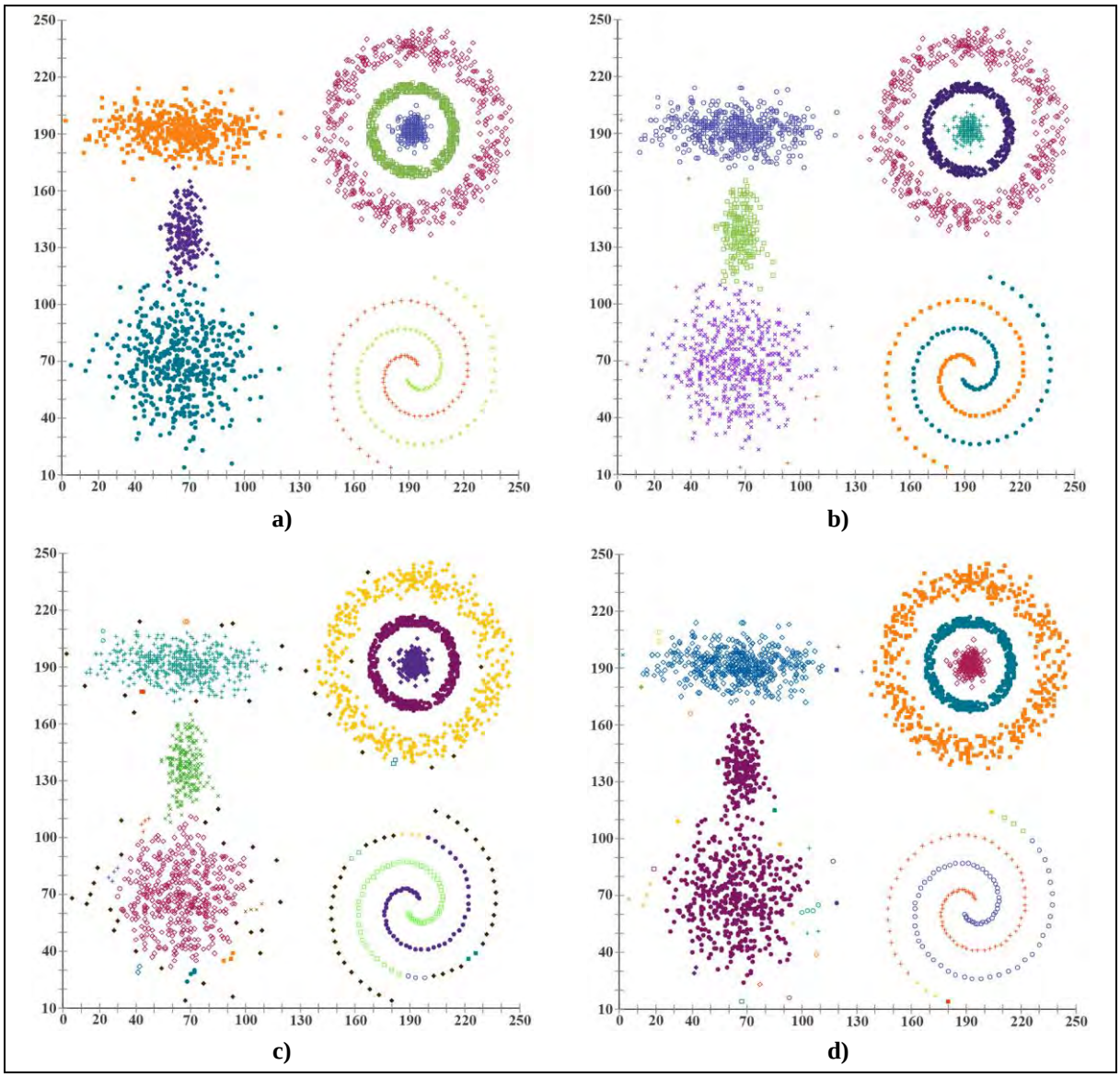



Fig. 2. Clustering results for model dataset: original labeling (a), proposed algorithm ECCAH (b), DeLiClu/OPTICS algorithms (c), Single Linkage (d).

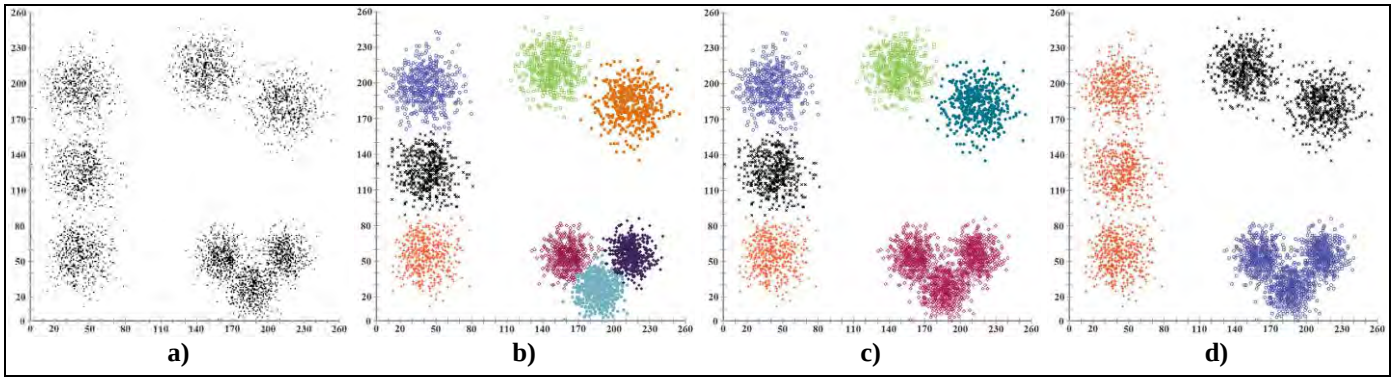



Fig. 3. Model dataset with 8 normally-distributed clusters (a) and ECCAH clustering result on different hierarchical levels (b, c, d).

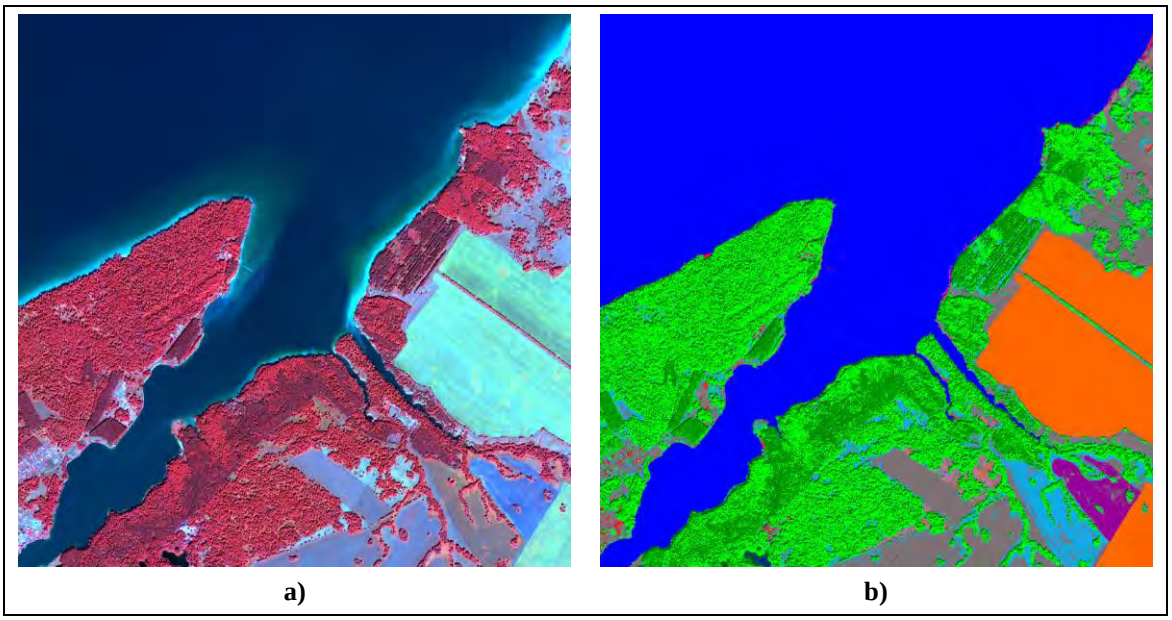



Fig. 4. Satellite image clustering: (a) WorldView-2 image (composite from channels 7, 4, 1); (b) clustering result of the proposed ECCAH algorithm.

Table below presents runtime of different clustering algorithms on the introduced models and RGB-color image of the size 452×588. It clearly demonstrates the advantage of the proposed method in speed.

TABLE I. RUNTIME COMPARISON OF CLUSTERING ALGORITHMS (THE MOST OPTIMAL PARAMETERS WERE USED).

| | **Model "Bananas"** | **Model 2 (fig. 2)** | **RGB-color image** |
|---|---|---|---|
| Data size | 400 | 4 000 | 265776 |
| ECCAH | 0.005 s | 0.003 s | 0.02 s |
| K-means Lloyd | 0.03 s | 0.08 s | 1 s |
| EM | 0.65 s | 18.4 s | 78 s |
| DBSCAN | 0.06 s | 0.35 s | 136 s |
| OPTICS | 0.03 s | 0.4 s | 390 s |
| DeLiClu | 0.19 s | 0.73 s | 430 s |
| SLINK | 0.01 s | 0.52 s | 2752 s |

## VI. CONCLUSION

In this paper, a new hierarchical ensemble clustering algorithm ECCAH for multispectral image segmentation is proposed. The advantage of hierarchical clustering is that number of clusters is not required by the algorithm and an expert may analyze clustering dendrogram to discover data structure in the best way. The algorithm is based on grid-approach and the ensemble method for combining obtained hierarchical partitionings.

Experimental results on both synthetic and real datasets confirming high quality of the obtained solutions and their stability to parameters change are presented. The possibility of obtaining different levels of cluster structure simplifies the process of interpreting the results. High performance of the proposed algorithm allows interactive satellite image processing. In addition, the developed algorithm allows parallelization.

## REVIEWER'S COMMENTS

Reviewer 1.

The original hierarchical clustering algorithm based on nonparametric estimation of the global probability density function of the data points is proposed. The authors also offer a method for constructing an ensemble of hierarchical grid algorithms clustering for segmentation multispectral satellite images. Ensemble approach allows combining multiple hierarchical partitioning and improving the quality of results in exploring complicated hierarchical structures. Experimental results on both synthetic and real datasets demonstrate the effectiveness of the proposed algorithm.

Reviewer 2 left no comments.

## REFERENCES

[1] I.A. Pestunov, Yu.N. Sinyavsky, "Clustering algorithms in problems of segmentation of satellite images," Bull. Kemerovo State Univ. 2012. No 4/2(52), pp. 110-125.

[2] R. Xu, D.I. Wunsch, "Survey of clustering algorithms," IEEE Trans. on Neural Networks. 2005. Vol. 16, N 3, pp. 645-678.

[3] L. Zheng, T. Li, C. Ding, "Hierarchical Ensemble Clustering," Proc. of 2011 IEEE Intern. Conf. on Data Mining. IEEE, 2010, pp. 1199-1204.

[4] A. Mirzaei, M. Rahmati, "A novel hierarchical-clustering-combination scheme based on fuzzy-similarity relations," IEEE Trans. Fuzzy Syst. 2010. Vol. 18, N 1, pp. 27-39.

[5] I.A. Pestunov, V.B. Berikov, E.A. Kulikova, S.A. Rylov, "Ensemble of clustering algorithms for large datasets," Optoelectronics, instrumentation and data processing. 2011. Vol. 47, N 3, pp. 245-252.

[6] B. Leclerc, "Description combinatoire des ultramétriques," Math. Sci. Humaines. 1981. Vol. 73, pp. 5-37.

[7] E. Achtert, H. Kriegel, E. Schubert, A. Zimek, "Interactive Data Mining with 3D-Parallel-Coordinate-Trees," Proc. ACM Intern. Conf. on Management of Data (SIGMOD). NY, 2013, pp. 1009-1012.

Proceedings of the 9th Open German-Russian Workshop on Pattern Recognition and Image Understanding
Dec 1-5, 2014, Koblenz, Germany

# Human Body Part Classification in Monocular Soccer Images

Andreas Bigontina, Michael Herrmann, Martin Hoernig and Bernd Radig
Image Understanding and Knowledge-Based Systems
Technische Universität München
Boltzmannstr. 3, 85748 Garching, Germany
Email: {bigontia,herrmmic,hoernig,radig}@in.tum.de

***Abstract*—This paper addresses the problem of finding body parts in images, which can be an essential first step for body pose estimation. The core component of the presented method is the pixel-based classification of body parts using Random Forests. This technique is applied to find the body part positions of soccer players in broadcast images. As this approach is usually used with depth data, we analyze how this method can be adapted to work with monocular images. In this context we identify the image representations with the best classification results. Although monocular images leave some ambiguities, our approach to body part classification achieves satisfying results: 90.32% of the pixels in the test set are correctly classified.**

## I. Introduction

The aim of articulated pose estimation is to determine the position of body parts and/or joints of a model of the human body in 2D or even 3D space. Hence, finding the position of body parts in an image, which is in the focus of this paper, can be seen as an essential part of a pose estimation algorithm. Combined with a skeleton fitting method it can be used to estimate 2D or 3D poses (e.g. by applying an inverse kinematic technique similar to Wei et al. [1]). In this paper we examine the particular application of detecting the body parts of soccer players in monocular images of soccer matches. As we work with cropped images of players from distance shots, we require a solution that can handle low resolution images.

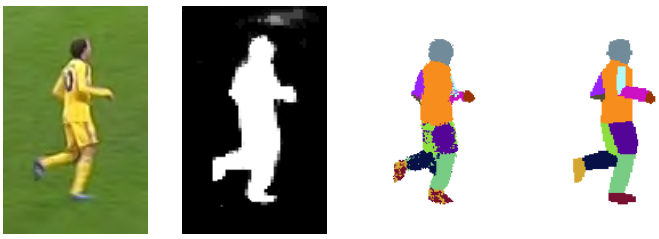
Despite its use as part of a pose estimation system, the pixel-based classification yields precious information on its own. During the analysis of a soccer video sequence, for instance, it might help to indicate whether the ball was hit by the player's foot, head, or hand. But the presented method can also be applied in other areas, given a monocular image and a person's silhouette. Some of the many areas of application are autonomous cars, human-robot interaction, surveillance, video indexing, and sports. Especially, when already employed cameras should be used or already recorded material should be analyzed, a solution that works with monocular images is required.

In summary, given the surrounding bounding box of a player in an image, our method performs two successive steps: In the first step we estimate the body orientation of the player. The result serves as input for the second step: For each pixel within the bounding box we determine the body part of the player to which the pixel most probably belongs.

Our proposed approach builds upon the work of Shotton et al. [2] who developed the pose estimation system for the Kinect gaming platform [3]. Similar to their approach, a Random Forest is used to classify each pixel of an image based on the state of the pixels in the neighborhood. The aim is to assign a body part or background class to every pixel in the image, as depicted in Fig. 1. This paper focuses on finding alternative image features to replace the missing depth information that was originally provided by the Kinect sensor. Therefore, several preprocessing steps are examined.

In addition, the body part classification of soccer players is more challenging than the classification of Kinect players since there is a greater variation in viewpoints. Therefore, an idea of Andriluka et al. [4] is followed who suggest to estimate the viewpoint beforehand. For this task, the orientation of a player is categorized into eight classes. A Random Forest is trained to estimate the orientation class of a player using the same techniques as for body part classification.

The main contribution of this paper is the adaption of the algorithm by Shotton et al. [2] to work with monocular images and the comparison of image features for this task. Furthermore, the problem of finding the orientation of a player is examined and corresponding results are presented.



(a) (b) (c) (d)

Fig. 1: Body part classification: (a) original player image, (b) silhouette retrieved using [5], (c) estimated body parts, (d) ground truth annotations

### A. Related Work

The problem of finding the pose of a person from sensor data has been investigated by many researchers before. For surveys on this topic see the work of Moeslund et al. [6] or Poppe [7]. With the upcoming of the Kinect gaming platform [3], developed by Microsoft, an accurate and fast pose estimation algorithm was presented [2] that is working on depth information. The main advantage of their algorithm

is a fast classification using a Random Forest with the idea of classifying every pixel separately. This allows for parallelization and implementation on GPUs, as shown by Sharp [8]. The presented approach differs from Shotton et al. [2] as it does not rely on depth information but works with monocular images.

Nevertheless, there are also several solutions for monocular images presented in literature. Many of these rely on the pictorial structures model, originally introduced by Fischler and Elschlager [9] and later applied to human pose estimation by Felzenszwalb and Huttenlocher [10]. The final results of a system using their techniques, for instance that of Andriluka et al. [11] who even estimate a 3D pose, are quite impressive. The complex series of actions that is required, however, also comes with a computational burden. For instance, Andriluka et al. [11] report an overall runtime of about 50 seconds for a $167 \times 251$ pixel image. In comparison, the method by Shotton et al. [2] can perform in real-time. Wei et al. [1] extended their approach and estimated a pose using the pixel-based classification, still keeping real-time performance. The presented approach clearly differs from classical pose estimation algorithms for monocular images due to the pixel-based classification concept and, thus, has computational advantages.

## II. Body Part Classification

As mentioned above, the classification of body parts is a core component of our method. Since images from broadcast videos might have a very low resolution, a rather coarse division into parts is necessary. Furthermore, the body model should be divided into more or less rigid parts to minimize the variation in appearance of each part. We use the following 14 classes: head, torso, left/right lower/upper arm, left/right lower/upper leg, hands, and feet (similar to [12]). The presented method largely follows the idea of Shotton et al. [2]. A Random Forest is used to classify every pixel of an image. Shotton et al. compare depth values of pixels in the neighborhood to decide about the class of a pixel. As such depth information is not available in monocular scenarios the following alternatives are examined: silhouette information, skin, color, gradient images, Haar-like features on gradients, HOG descriptors, pixel positions, and estimated player orientations. Since some of these features might give partial information about certain body parts only, feature combinations are examined as well. The silhouettes are retrieved using the method of Hoernig et al. [5] which is especially applicable in soccer and similar field sports. Silhouettes are very expressive but naturally leave some details unrecognized. Skin features are extracted from the image using the method described by Gomez and Morales [13] and are intended to complement the silhouette information. Additionally, raw color values are examined as well to see if the applied Random Forest is powerful enough to extract the most useful information by itself. Another way of finding hidden details inside silhouettes are gradients and their magnitudes. However, the most useful, strong gradients usually exist at thin edges and are therefore hard to utilize by the presented approach. The idea of summing up the gradients within an image area is implemented using Haar-like features (similar to [14] ). Finally, Histograms of Oriented Gradients (HOG) [15] are examined, as they have shown good results in human detection.

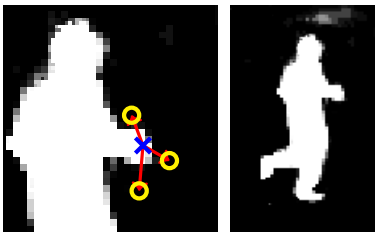


Fig. 2: Illustration of the classification method for silhouettes: the blue cross marks the pixel that has to be classified, the yellow circles mark positions in the neighborhood at which foreground probabilities are compared to a threshold

We trained Random Forests using these features and feature combinations to estimate the orientation of a player. Each node in the forest examines a feature value at a certain position in the image and compares it to a threshold. The most expressive position that contains the most distinctive feature values and the threshold were chosen during the training phase. The estimated orientation can be important for further pose recovery tasks, but it also gives hints about body part classes. Therefore, the result of the orientation classification serves as a feature for the pixel-based body part classification.

While orientation classification is concerned with the orientation of a whole player image, body part classification tries to find a class for a single pixel. Hence, an additional information that is available is the pixel position relative to the player's bounding box. Therefore, it is considered as a feature for the body part classification as well.

For all features, a simple axis-aligned weak learner is used. This means that a simple comparison between a feature attribute value and a threshold is performed in each node of a Random Tree. In the training phase an expressive feature attribute has to be chosen and an appropriate threshold has to be found. For a detailed description of Random Forests see, for example, the extensive work by Criminisi et al. [16].

We trained different Random Forests on features and combinations of these and performed an evaluation using a challenging data set containing 70 images of players in various poses, resulting in millions of test pixels. Surprisingly, a rather simple combination of features showed to perform best: silhouette, color, orientation, and pixel position. Therefore, these four features are described in detail.

### A. Silhouette

The silhouette information retrieved using [5] is not binary but provides foreground probabilities for every pixel. The applied weak learner of the Random Forest compares foreground probabilities in the neighborhood of a target pixel to a threshold. Which pixel in the neighborhood should be examined is specified by an offset that has to be found in the training phase. This is illustrated in Fig. 2.

### B. Color

Color contains full information, hence, it is more challenging to extract the most relevant information. Similar to the idea of Shotton et al. [2] two different strategies are encoded in

TABLE I: Orientation classification accuracy

| Features | Accuracy |
|---|---|
| Silhouette | 51.4% |
| Silhouette + Skin | 49.3% |
| Silhouette + HOG | 40.7% |
| Silhouette + Gradients | 38.5% |
| Silhouette + Color | 35.0% |
| Color | 29.3% |
| Silhouette + Haar-like | 28.6% |

TABLE II: Body part classification accuracy

| Features | Accuracy |
|---|---|
| Sil. + Orientation + Position + Color | 90.32% |
| Sil. + Orientation + Position + Gradients | 89.54% |
| Sil. + Orientation + Position + Skin | 89.46% |
| Sil. + Orientation + Position + HOG | 89.43% |
| Sil. + Orientation + Position + Haar-like | 89.37% |
| Sil. + Orientation + Position | 89.34% |
| Silhouette | 87.13% |
| Silhouette + Orientation | 87.03% |
| Color | 85.23% |

the applied weak learner of the Random Forests. One option is to compare the value of a color channel to a threshold. The other option is to compare the value difference between two pixels regarding one color channel to a threshold. This feature involves choosing one of these options and finding one or two offsets, a color channel, and a threshold during the training.

### C. Pixel Position

This paper is only concerned with finding the body parts of players, thus, it is assumed that players are already detected and that their bounding boxes are known. A further improvement of the classification accuracy can be achieved by adding the position of a pixel relative to a players bounding box as a feature (see Table II). To optimally utilize the position data, the Random Forest is provided with redundant information. In particular, a pixel's distance to the top, left, right, and bottom of the bounding box are provided, as well as the coordinates relative to the center. The applied weak learner selects one of these measurements and compares it to a threshold.

### D. Player Orientation

As mentioned above, the orientation of a person is estimated as well. The Random Forest for body part classification contains a weak learner that compares the estimated probability of one of the eight orientation classes to a threshold.

## III. Experiments

To evaluate the performance of the proposed system and to find the best feature or feature combination an extensive evaluation was performed. Concentrating on the special application of broadcast videos of soccer matches a corresponding data set was created that contains 200 images of soccer players from various viewpoints, in various qualities and containing motion blur and similar challenges. The average size of these images is about $78 \times 120$ pixels, but every image was scaled to a height of 200 pixels (with fixed aspect ratio) to cope with different image sizes to some extend. The data set was split into 130 training and 70 test images. To enhance the data set a flipped version of every image was added, while assuring that players from test and training set are not mixed up. In all images of the data set body parts were manually annotated on pixel-level.

Concerning the classification of a player's orientation, a Random Forest that uses the silhouette only, achieves the best results for the soccer data set (see Table I). Classifying the orientation is a difficult task, only $51.4\%$ of the player images are classified correctly, however, it is sufficient to support body part classification. Furthermore, it is a useful information for upcoming pose estimation tasks. It is likely that performance could be improved by increasing the number of training samples. This might also help to utilize more complex features better. Although, 130 images yield millions of training pixels for the body part classification, they only yield 130 samples to train the orientation classifier.

As already mentioned above, our most accurate feature combination for body part classification is silhouette, color, position, and orientation. An example result is visualized in Fig. 1. Using this feature combination $90.32\%$ of the pixels in the test set are correctly classified. Here, a pixel is considered correctly classified if the class with the highest probability equals the ground truth class. As can be seen from Table II, the results for other combinations are only slightly worse. Note, however, that a significant part of the pixels of a player image are background and are already indicated as such by the silhouette. In fact, if a simple threshold is applied to the silhouette probabilities and the resulting foreground pixels are classified as torso (the most frequent class), already $79.95\%$ of the pixels in the test set are correctly classified. Thus, a good classifier is expected to improve significantly over this baseline.

The best performing Random Forest within our test set contained 20 trees, which are trained to a depth of 20 but are also limited by a minimal amount of samples per leaf of 92. Every tree is trained on a random subset of the training data. During training only a random subset of the theoretically possible splits is examined as it is common practice for Random Forests. The neighborhood of a pixel that is examined is limited to a $50 \times 50$ pixels patch for images of height 200 pixels. An examination of this Random Forest shows that most information is retrieved from the silhouette, while the orientation, the color and the pixel position only improve results slightly.

The classification of orientation and body parts can be performed in less than a second on a recent CPU. As already mentioned, Random Forests can be implemented on GPUs which promises a significant performance improvement.

## IV. Conclusion

The performed experiments have shown that it is indeed possible to use pixel-based classification on image data from monocular cameras to estimate the position of body parts with high accuracy, even for low resolution and low quality images. Furthermore, it has been shown that simple features are usually sufficient to receive good results. However, some challenges remain such as distinguishing the left and right body parts, and scoping with erroneous silhouettes, which contain field lines, advertisement banners, or similar background objects.

## REFERENCES

[1] X. Wei, P. Zhang, and J. Chai, "Accurate realtime full-body motion capture using a single depth camera," *ACM Transactions on Graphics*, vol. 31, no. 6, p. 1, Nov. 2012.

[2] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *Computer Vision and Pattern Recognition*, vol. 2, no. 3. IEEE, 2011, pp. 1297–1304.

[3] Microsoft Corp. Redmond WA, "Kinect for Xbox 360," Redmond WA, 2010.

[4] M. Andriluka, S. Roth, and B. Schiele, "Monocular 3D pose estimation and tracking by detection," in *Computer Vision and Pattern Recognition*. IEEE, Jun. 2010, pp. 623–630.

[5] M. Hoernig, M. Herrmann, and B. Radig, "Real Time Soccer Field Analysis from Monocular TV Video Data," in *11th International Conference on Pattern Recognition and Image Analysis: New Information Technologies (PRIA-11-2013)*, vol. II, Samara, 2013, pp. 567–570.

[6] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Computer Vision and Image Understanding*, vol. 104, no. 2-3, pp. 90–126, Nov. 2006.

[7] R. Poppe, "A survey on vision-based human action recognition," *Image and Vision Computing*, vol. 28, no. 6, pp. 976–990, Jun. 2010.

[8] T. Sharp, "Implementing Decision Trees and Forests on a GPU," *Computer Vision - ECCV 2008*, vol. 5305, pp. 595–608, 2008.

[9] M. A. Fischler and R. A. Elschlager, "The Representation and Matching of Pictorial Structures," *IEEE Transactions on Computers*, vol. C-22, no. 1, pp. 67–92, Jan. 1973.

[10] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial Structures for Object Recognition," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 55–79, 2005.

[11] M. Andriluka, S. Roth, and B. Schiele, "Discriminative Appearance Models for Pictorial Structures," *International Journal of Computer Vision*, vol. 99, no. 3, pp. 259–280, Oct. 2011.

[12] A. Hernández-Vela, M. Reyes, V. Ponce, and S. Escalera, "GrabCut-based human segmentation in video sequences." *Sensors (Basel, Switzerland)*, vol. 12, no. 11, pp. 15 376–93, Jan. 2012.

[13] G. Gomez and E. F. Morales, "Automatic feature construction and a simple rule induction algorithm for skin detection," in *Proc. of the ICML workshop on Machine Learning in Computer Vision*, 2002, pp. 31–38.

[14] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition*, vol. 1. IEEE, 2001, pp. I–511–I–518.

[15] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition*, vol. 1, no. 3. IEEE, 2005, pp. 886–893.

[16] A. Criminisi, J. Shotton, and E. Konukoglu, "Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning," *Foundations and Trends in Computer Graphics and Vision*, vol. 7, no. 2-3, pp. 81–227, 2011.

# Image warping as an image enhancement post-processing tool

Andrey Krylov, Alexandra Nasonova, Andrey Nasonov*†

**Abstract**

A method to improve the results of image enhancement is proposed. The method is based on pixel grid warping, the main idea is moving pixels in the direction of the nearest image edges. Warping allows to make edges sharper while keeping textured areas almost intact. Experimental applications of the proposed method for image enhancement algorithms show the improvement of image quality.

## 1 Introduction

There are some fairly powerful techniques for image deblurring [1], [2], [3]. The typical problem of image deblurring methods is finding optimal parameters for a compromise between smooth result with blurry edges and sharp result with artifacts like ringing or noise amplification. In this work we present a new post-processing algorithm for image deblurring with enhancement of edge sharpness.

We localize the area of interest to the neighborhood of the edges. The idea is to transform the neighborhood of the blurred edge so that the neighboring pixels move closer to the edge, and then resample the image from the warped grid to the original uniform grid.

The warping approach for image enhancement was introduced in [4]. The warping of the grid in this work is performed according to the solution of a differential equation derived from the warping process constraints. The solution of the equation is used to move the edge neighborhood closer to the edge, and the areas between edges are stretched. The method has several parameters, and the choice of optimal values for the best result is not easy. Due to the global nature of the method the resulting shapes of the edges are often distorted. In [5] the warping map is computed directly using the values of left and right derivatives. In both methods [4] and [5] the pixel shifts are proportional to the gradient values. It results in oversharpening of already sharp and high contrast edges and insufficient sharpening of blurry and low contrast edges. Both methods also introduce small local changes in the direction of edges and produce aliasing effect due to calculation of horizontal and vertical warping components separately.

## 2 Warping technique

In this section we describe the idea of a single edge enhancement using a pixel grid transformation. The profile of a blurred edge is more gradual compared to a sharp edge profile. So in order to make the edge sharper its transient width should be decreased (see Fig. 1).

### 2.1 Warping of a one-dimensional signal

The idea of one-dimensional edge sharpening is based on the assumption that the edge can be approximated by a step-edge function $H(x)$ smoothed with a Gaussian filter $G_\sigma$ with a standard deviation $\sigma$ [6]:

$$E_\sigma(x) = [H * G_\sigma](x), \quad \text{where } H(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0, \end{cases} \tag{1}$$

One-dimensional grid warping (2) is performed according to the following equation:

$$\tilde{x} - x = AG'_\sigma(\tilde{x}), \tag{2}$$

*A. Krylov, A. Nasonova and A. Nasonov are with the Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, Moscow, Russia, e-mail: kryl@cs.msu.ru, nasonova@cs.msu.ru, nasonov@cs.msu.ru.

†The work was supported by Russian Science Foundation grant 14-11-00308.

a. Proposed approach: pixels are shifted

b. Typical image enhancement approach pixel values are modified

Figure 1: The idea of edge sharpening

where $x$ is the old position of pixel $E_\sigma(x)$, $\tilde{x}$ is the new position, $A > 0$ controls the strength of grid warping. This model ensures that the shape of the edge is not distorted and the grid transformation is smooth.

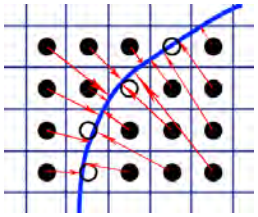
To avoid a discontinuity of the solution of the warp equation (2), the strength parameter $A$ should be such that $A < \frac{1}{\max\limits_{x \in \Re} G''_\sigma(x)}$. We use $A = 0.99 \cdot \frac{1}{\max\limits_{x \in \Re} G''_\sigma(x)}$ in order to get a strong sharpening effect.

## 2.2 Image warping

The idea of the proposed algorithm for one-dimensional signal warping needs to be adopted for application for two-dimensional images. Two variants of the image warping are suggested.

### 2.2.1 Two-dimensional extension

In the two-dimensional case, the pixels are surrounded by a number of edges. Choosing the right direction of pixel warping requires additional analysis. The simplest algorithm is finding the nearest edge for every pixel and apply the one-dimensional algorithm using that edge. It consists of the following steps:
1. Estimate the blur level (the average standard deviation of Gaussian filter) for the edges.
2. For all pixels in the neighborhood of the edge compute the distance $d$ to the nearest edge point.
3. Compute pixels' displacements (see Fig. 2) using equation (2) with $x \equiv d$.
4. Interpolate the image from the warped grid to the old uniform grid.



The thick line represents the exact edge location, white circles represent edge pixels, black circles represent pixels from the edge neighborhood

Figure 2: Displacements for two-dimensional grid warping

The edge map at the input of the algorithm has a great influence on the result of grid warping as only detected edges will be sharpened. In our work we use the result of Canny edge detection [7] as the input of the algorithm. The result of image warping is highly dependent on the input edges. The parameters of the Canny method ($\sigma$ and high threshold $T_{high}$) are chosen individually for each image.

### 2.2.2 Poisson warping

To take into account several edges simultaneously to avoid possible discontinuities between edges, we propose a 2D image warping algorithm which computes the displacement field directly as a 2D vector field $\vec{d}(x, y)$ with some obvious constraints (see Fig. 2).

1) The shapes of the edges cannot be warped, so $\vec{d}(x_e, y_e) = 0$ for each edge point $(x_e, y_e)$.

2) Also, there cannot be any turbulence: rot $\vec{d} = 0$. Since rot$\nabla u \equiv 0$, the displacement field is assumed to be gradient of some scalar function $u(x, y)$: $\vec{d}(x, y) = \nabla u(x, y)$.

3) The edge neighborhood points situated farther from the edge cannot move closer to the edge than the neighborhood points situated nearer to the edge: $\operatorname{div} \vec{d} \geq -1$.

Since $\operatorname{div}\nabla \equiv \Delta$, where $\Delta$ is a Laplacian, the warping problem can be posed as a Dirichlet problem for the Poisson equation in the area of the image:

$$\begin{cases} \Delta u & = p(x,y) - 1, \\ u(x,y) & = 0 \text{ at image borders} \end{cases}, \quad \text{where } p(x_0,y_0) = \frac{\sum\limits_{(x,y)\in E(x_0,y_0)} p(x_n) G_\sigma(x_t) |\vec{g}(x,y)|}{\sum\limits_{(x,y)\in E(x_0,y_0)} |\vec{g}(x,y)|} \tag{3}$$

The second constraint here is the boundary condition: the displacements at image borders should be equal to zero; $p(x,y)$ is the proximity function that describes the distance between adjacent pixels after image warping; $E(x_0,y_0)$ is the set of edge points in the neighborhood of $(x_0,y_0)$; the values $x_n$ and $x_t$ are projections of the vector $(x_0 - x, y_0 - y)$ on the edge gradient vector $\vec{g}(x,y)$ and on the tangent to the edge respectively; $p(x_n) = 1 - AG''_\sigma(x)$; $G_\sigma(x_t)$ is the weighting function with standard deviation equal to the edge's blur $\sigma$.

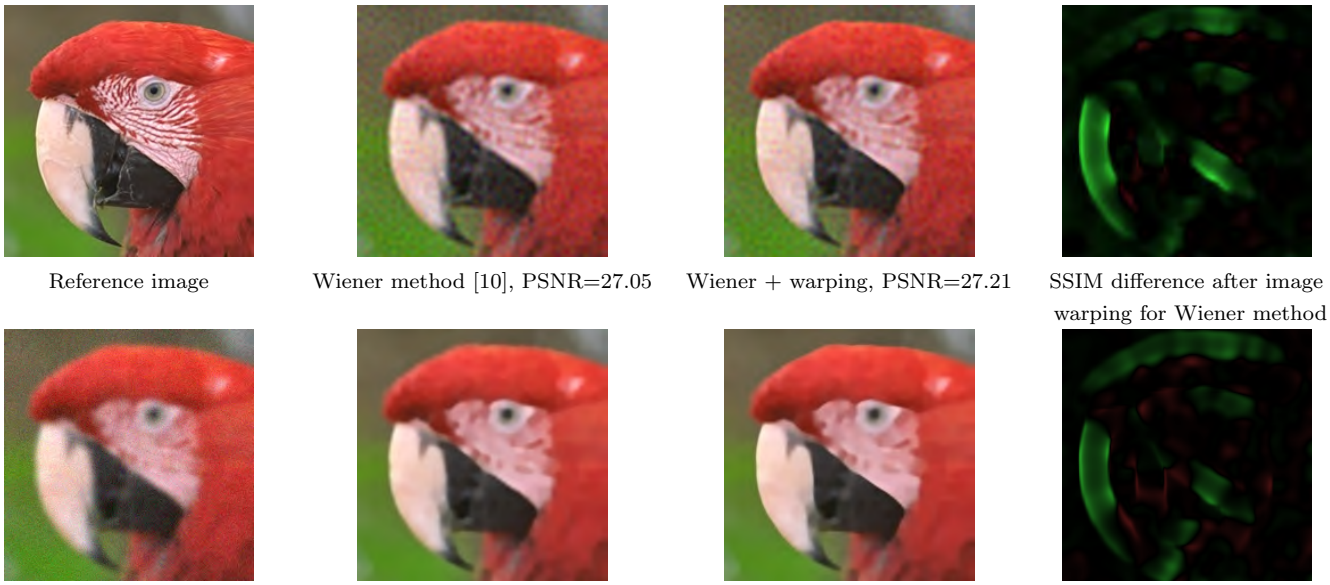
We solve the equation (3) by Gauss-Zeidel method.

## 3 Results and experiments

We applied the image warping as a post-processing algorithm for image deblurring and TV image enhancement algorithms. The proposed method was tested on 29 images from LIVE database [8]. The images were blurred with Gaussian kernel with random radius in the range $[1, 6]$, then Gaussian white noise with random standard deviation in the range $[0, 10]$ was added. After that we applied existing deblurring algorithms followed by image warping using known blur level. Table 1 represents the result. Preliminary experiments with automatic estimation of the unknown edge width [9] also show the enhancement of deblurring methods.

The Poisson warping shows a bit better results than 2D extension of 1D warping. It produces smoother edges but is about 10 times more computationally complex.

The example of the proposed Poisson warping is shown on Fig. 3. It can be seen that the edges become better and the overall quality is improved. Nevertheless, small SSIM degradation regions exist. They correspond to the initially blurred regions of the image. Of course, an unwanted sharpening effect for the blurred areas of the original image can appear but it is a rare case.



Reference image | Wiener method [10], PSNR=27.05 | Wiener + warping, PSNR=27.21 | SSIM difference after image warping for Wiener method

Degraded image, PSNR=25.99 | TVMM [2], PSNR=28.66 | TVMM + warping, PSNR=28.72 | SSIM difference after image

Figure 3: Poisson warping algorithm example. Green areas show improvement of SSIM, red — degradation.

In [12] was shown that the proposed warping approach is a good post-processing tool in image ringing suppression and resampling.

| Method | No warping | 2D ext warping | Poisson warping |
|---|---|---|---|
| Blurred and noisy images | 22.84 | 23.29 | 23.25 |
| Unsharp masking | 23.00 | 23.54 | 23.36 |
| TV regularization | 23.30 | 23.35 | 23.40 |
| Low-frequency TV reg. [11] | 23.08 | 23.15 | 23.18 |
| TVMM [2] | 23.31 | 23.33 | 23.48 |
| Lucy-Richardson [10] | 23.83 | 23.94 | 23.99 |
| Wiener [10] | 24.00 | 24.08 | 24.17 |
| MatLab blind deconvolution | 23.79 | 23.93 | 23.96 |
| Average | 23.39 | 23.58 | 23.60 |

Table 1: Average PSNR values for images from LIVE database with added blur and noise

# 4 Conclusion

The proposed image warping method has a great potential to improve the results of existing image enhancement algorithms. It is especially effective for total variation based image enhancement algorithms because image warping does not significantly change total variation value. It can also be used as a standalone image sharpening algorithm. It is a good choice in the presence of strong noise and complex and non-uniform blur kernel.

In comparison to existing sharpening approaches, the proposed method introduces no artifacts like ringing effect nor noise amplification, the resulting images look natural and do not inevitably become piecewise constant.

# References

[1] M. Almeida and M. Figueiredo, "Parameter estimation for blind and non-blind deblurring using residual whiteness measures," *IEEE Trans. Image Processing*, vol. 22, pp. 2751–2763, 2013.

[2] J. Oliveira, J. M. Bioucas-Dia, and M. Figueiredo, "Adaptive total variation image deblurring: A majorization-minimization approach," *Signal Process.*, vol. 89, pp. 1683–1693, 2009.

[3] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Variational bayesian blind deconvolution using a total variation prior," *IEEE Trans. Image Process.*, vol. 18, pp. 12–26, 2009.

[4] N. Arad and C. Gotsman, "Enhancement by image-dependent warping," *IEEE Trans. Image Proc.*, vol. 8, pp. 1063–1074, 1999.

[5] J. Prades-Nebot et al., "Image enhancement using warping technique," *IEEE Electronics Letters*, vol. 39, pp. 32–33, 2003.

[6] A. A. Chernomorets and A. V. Nasonov, "Deblurring in fundus images," in *22-th Int. Conf. GraphiCon'2012*, Moscow, Russia, 2012, pp. 76–79.

[7] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679–698, 1986.

[8] H. Sheikh, M. Sabir, and A. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

[9] A. A. Nasonova and A. S. Krylov, "Determination of image edge width by unsharp masking," *Computational Mathematics and Modelling*, vol. 25, pp. 72–78, 2014.

[10] J. G. Nagy, K. Palmer, and L. Perrone, "Iterative methods for image deblurring: A matlab object-oriented approach," *Numerical Algorithms*, vol. 36, no. 1, pp. 73–93, 2004.

[11] A. S. Krylov and A. V. Nasonov, "Adaptive image deblurring with ringing control," *Fifth International Conference on Image and Graphics (ICIG '09)*, pp. 72–75, 2009.

[12] A. V. Nasonov and A. S. Krylov, "Grid warping in total variation image enhancement methods," in *International Conference on Image Processing (ICIP2014) (in print)*, 2014.

Proceedings of the 9th Open German-Russian Workshop on Pattern Recognition and Image Understanding
Dec 1-5, 2014, Koblenz, Germany

# Image-based characterization of the pulp flows

Mikhail Sorokin*, Nataliya Strokina*†, Tuomas Eerola*, Lasse Lensu* and Heikki Kälviäinen*

*Machine Vision and Pattern Recognition Laboratory (MVPR), Department of Mathematics and Physics,
Lappeenranta University of Technology (LUT), Lappeenranta, Finland
Email: firstname.lastname@lut.fi
†Computer Vision Group, Department of Signal Processing, Tampere University of Technology, Tampere, Finland
Email: firstname.lastname@tut.fi

***Abstract*—Material flow characterization is important in the process industries and its further automation. In this study, close-to-laminar pulp suspension flows are analyzed based on double-exposure images captured in laboratory conditions. The correlation-based methods including autocorrelation and the particle image pattern technique were studied. During the experiments, synthetic and real test data with manual ground truth were used. The particle image pattern matching method showed better performance achieving the accuracy of 90.0% for the real data set with linear motion of the suspension and 79.2% for the data set with flow distortions.**

## I. Introduction

Material flow characterization has an important role in the process industries. The role is emphasized in the development of more resource-efficient process equipment and in the further automation of process control to meet the desired product quality. In the pulp industry, pulp flow characterization implies the estimation of the flow velocity as well as the detection of anomalies, e.g., turbulence and vortices, in the flow that can signal process malfunctioning and also less-than-optimal process equipment design.

Particle Image Velocimetry (PIV) is a common optical measurement technique for industrial fluid flow analysis [1]. The methods for the PIV analysis can be divided into two groups [2]: image patch-based techniques [3] and methods utilizing the principles of optical flow [4], [5]. In the first group of methods, a PIV image is divided into overlapping or non-overlapping patches. For each patch from the current image a corresponding patch from the consecutive images is sought, utilizing the maximum of the cross-correlation. Flow velocity vectors are the vectors going through the centers of the found patches. The main problem of the patch-based methods is that the estimated vector field is often incoherent and unsmooth which requires post-processing, e.g., median filtering [2].

The connection between the fluid and optical flows was formally described by Liu and Shen in [6]. The majority of the developed methods, based on the optical flow estimation, were built upon the original Horn and Schunk (HS) method [7], as it is stated in [8]. The flow is modelled as a global energy function over an image. Optimizing this model function, the velocity vectors are found. The methods developed later differ by the modelled objective function, by its approximation, and by the computational method for the objective function optimization. For example, in [9], the optical flow constraint equation is compensated with the fitted higher-order term by matching the corner points extracted by the Harris corner point detector [10]. In [8] Sun et al. studied the state-of-the-art methods and carried out a systematic experiments varying the model and the optimization method. This experiment allowed the authors to conclude that none of the modifications significantly improved over the baseline HS method.

In this work, double-exposed images by using two short laser pulses for each captured image frame (examples shown in Fig. 1) are used to estimate a dense set of 2D pulp flow velocity vectors. When compared to standard PIV the fibers in the pulp flow are the studied particles. To characterize the flow, the common autocorrelation technique [11] is compared to the modified Particle Image Pattern (PIP) matching [12]. A framework utilizing global and local techniques for the pulp flow velocity estimation is proposed where a synthetic image set and a real-world image set were used for testing in [13]. In this work, it is assumed that the motion of the pulp flow is planar and close-to-laminar which allows to use the correlation-based techniques.

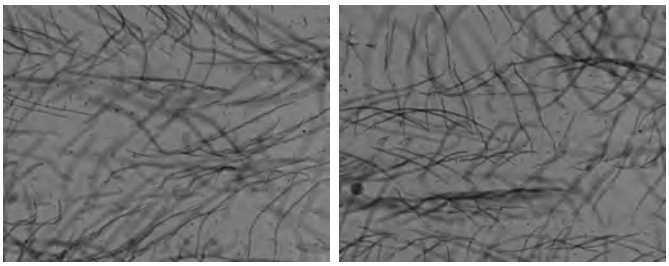


Fig. 1. Example of images captured with a double exposure.

## II. Pulp flow velocity estimation

### A. Estimation of the global displacement

The estimation of the optical flow global displacement allows the measurement of a large-scale motion velocity not taking into account local anomalies. The method of global displacement estimation for double-exposed images is based on normalized autocorrelation [14]. Before the autocorrelation matrix is computed, the background is subtracted which eliminates the background noise.

The second local maximum of the matrix is localized providing the global displacement $D_g$ since the result always contains a self-correlation peak located at the origin. Two peaks describing the global displacement of the pulp flow image locate symmetrically around the self-correlation peak [3].

Therefore, the direction of the flow, whether the flows moves right or left, cannot be determined.

### B. Estimation of the local displacement

The computed global displacement gives a good estimate for the coarse motion in the images, but by nature, the optical flow can contain irregularities which should be also detected. In this work, the local displacement is estimated by the autocorrelation technique [14] and the pattern matching technique [12].

*Autocorrelation technique:* As in the global displacement computation, the background noise is reduced by subtracting the mean intensity from the image. The image is subsequently split into parts, starting with four, and for each part the normalized autocorrelation matrix is computed, the second local maximum of the matrix is found, providing an estimate for the local displacements. The parts of the image are split into subsequent parts until the size of the part is less than the estimated global displacement $D_g$. The autocorrelation technique is summarized in Algorithm 1.

**Algorithm 1** Autocorrelation method for the local displacement estimation

**Input:** a double-exposed grayscale image **I**, an estimate of the global displacement $D_g$
**Output:** a set of computed local displacements $D = \{D_i\}$.

1: Compute the mean intensity of the image $\bar{I} = \frac{1}{n}\sum_{i=1}^{n} I(x_i, y_i)$ where $n$ is the number of pixels in the image.
2: Extract the mean intensity from the image pixels $I_m = I - \bar{I}$.
3: **repeat**
4: Split the image $I_m$ into $n = 4$ parts $J_1, ...J_4$ each of size $N$.
5: **for all** $J_i$ **do**
6: Compute the autocorrelation matrix.
7: Compute the local displacement $D_i$ as the second local maximum.
8: **end for**
9: **until** $N \leq D_g$

*PIP matching technique:* The original Particle Image Pattern (PIP) matching for individually exposed images was introduced in [15]. In [12] the PIP matching for the double-exposed images was presented and verified. Given a double-exposed image $I(x, y)$, the task is to estimate the flow displacement in the point $(x_0, y_0)$, restricted to the maximum displacement $D_g$. $\Delta X$ and $\Delta Y$ as the $x$- and $y$- components of the pulp flow shift between two exposures respectively. The interrogation PIP, or IPIP [12], of size $2N + 1 \times 2N + 1$ equals to

$$IPIP(m, n) = I(x_0 + m, y_0 + n), \quad m, n = -N, -N + 1, ..., N. \tag{1}$$

The search PIP, or SPIP [12], of size $2M + 1 \times 2M + 1$ equals to

$$SPIP(m, n) = I(x_0 + \Delta X + m, y_0 + \Delta Y + n), \quad m, n = -M, -M + 1, ..., M. \tag{2}$$

The background noise is compensated by the subtraction of the mean value. The normalized cross-correlation matrix is computed for each image part (IPIP) and whole image (SPIP). In this work, the SPIP is equal to the whole double-exposed image. Therefore, the correlation matrix is

$$\gamma(\Delta X, \Delta Y) = \frac{\sum_{x,y}\left[SPIP(x,y) - \overline{SPIP}\right]\left[IPIP(x-\Delta X, y-\Delta Y) - \overline{IPIP}\right]}{\sqrt{\sum_{x,y}\left[SPIP(x,y) - \overline{SPIP}\right]^2 \sum_{x,y}\left[IPIP(x-\Delta X, y-\Delta Y) - \overline{IPIP}\right]^2}}. \tag{3}$$

The second local maximum is sought for each cross-correlation matrix to estimate the displacement. The image is subsequently split into parts and the cross-correlation is computed for each part until the size of the image part is less than the global displacement $D_g$. The PIP matching method is summarized in Algorithm 2.

**Algorithm 2** PIP matching method for the local displacement estimation

**Input:** a double-exposed grayscale image $SPIP$, an estimate of the global displacement $D_g$
**Output:** a set of computed local displacements $D = \{D_i\}$.

1: Compute the mean intensity of the image $\overline{SPIP} = \frac{1}{N}\sum_{i=1}^{N} I(x_i, y_i)$ where $N$ is the number of pixels in the image.
2: Subtract the mean intensity from the image pixels $SPIP_m = SPIP - \overline{SPIP}$.
3: **repeat**
4: Split the image $SPIP_m$ into $n = 4$ parts $IPIP_1, ...IPIP_4$ each of size $N$.
5: **for all** $IPIP_i$ **do**
6: Compute cross-correlation matrices between $IPIP_i$ and the whole image $SPIP$ (Eq. 3).
7: Compute the local displacement $D_i$ as the second local maximum.
8: **end for**
9: **until** $N \leq D_g$

*Postprocessing:* In order to restrict the location of the maximum in the cross-correlation matrix and take into account the estimated global displacement, the low-pass Butterworth filter [16] is applied to the cross-correlation matrix. The size of the Butterworth filter is defined by estimated displacement from the previous level of splitting. The radius of the Butterworth filter depends on the split level and has to be more than the displacement computed in the previous level. In this case, the Butterworth filter does not remove the peak of the cross-correlation matrix which corresponds to the local displacement.

As the post-processing step, the local displacement is compared to the global displacement and if the difference exceeds a threshold, the value of the local displacement is replaced by the global displacement.

## III. Synthetic data generation

In order to evaluate the methods, the experiments were performed first on the synthetic images for which there exists the ground truth (GT). The Thin Plate Spline (TPS) [17] is a commonly used basis function for representing coordinate mappings from $\mathbb{R}^2$ to $\mathbb{R}^2$. The TPS models deformations by interpolating the displacement between the source and target points [17]. For the set of reference points $P_0 =$

$\{(p_i = \{x_i, y_i\} | i = 1, 2, \cdots, n\}$ the TPS transformation [18] is defined as

$$f(x,y) = \Phi_s(x,y) + R_s(x,y) = a_1 + a_x x + a_y y + \sum_{i=1}^{n} \omega_i U(r_i) \quad (4)$$

where $U(r_i) = r_i^2 \log r_i^2$ and $r_i = |p_i - (x,y)| = \sqrt{(x_i - x)^2 + (y_i - y)^2}$. The TPS transformation consist of two parts: the affine part and the elastic part. The affine part $\Phi_s(x,y)$ is a sum of polynomials with coefficients $\mathbf{a} = [a_1\ a_x\ a_y]$. A sum of Radial Basis Functions (RBF) with coefficients $\omega = [\omega_1\ \omega_2\ \cdots\ \omega_n]$ corresponds to the elastic parts $R_s(x,y)$. It is assumed that the locations $(x_i, y_i)$ are all different and are not collinear. The TPS interpolant $f(x,y)$ minimizes the bending energy

$$E_{TPS}(f) = \iint \left| \left| \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial x \partial y} + \frac{\partial^2 f}{\partial y^2} \right|^2 \right| dxdy \quad (5)$$

In order to find the coefficients $(\mathbf{a}, \omega)$, the following linear system needs to be solved:

$$\begin{cases} \mathbf{K}\omega + \mathbf{P}\mathbf{a} = \upsilon \\ \mathbf{P}^T \omega = \mathbf{0} \end{cases} \quad (6)$$

where $\mathbf{K}$ is the $n \times n$ matrix given by $K_{ij} = U(r_{ij})$, $f(x_i, y_i) = \upsilon_i$, $i = 1, 2, \cdots, n$, $\mathbf{P}$ is the $n \times 3$ matrix and the $i$th row of $\mathbf{P}$ is $[1\ x_i\ y_i]$, $\mathbf{0}$ is a $3 \times 1$ column vector of zeros, and $\upsilon = [\upsilon_1 \cdots \upsilon_n]$ [18]. After that, the transformation is applied to the reference points $P_0 = \{(p_i = \{x_i, y_i\} | i = 1, 2, \cdots, n\}$.

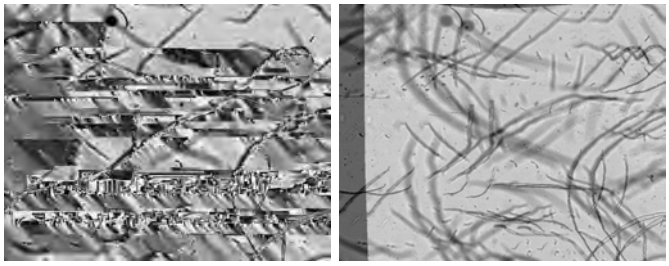
In this work, TPSs are applied to a set of single exposure images to synthesize the second exposure. The reference points are located on a uniform grid. After computing the target image by using the warping technique [19], a synthetic double-exposed image is produced by computing the mean value of the gray levels for each pixel in the target and reference images.

## IV. Experiments and discussion

### A. Synthetic data

Images of the birch pulp flow produced by the CEMIS-OULU Laboratory were taken as the source images. There were 100 images of size 896x704 pixels taken with 5x magnification and 80 ns exposure. Two data sets of the synthetic images were produced using the method described in Section III. In the first set, only linear motion along the $x$-axis was considered. In the second set, polar co-ordinates were used and the parameters $r$ and $\theta$ were sampled from two normal distributions with $\mu = 80$ and $\sigma^2 = 10$, and $\mu = 0$ and $\sigma^2 = \pi/32$. Examples of the original and the synthetic images are presented in Fig. 2.

The results of an experiment including 8500 displacement vectors are presented in Table I. The second column contains the percentage of the correctly computed vectors. A correctly computed vector is a vector with a relative length error less than 10%. $\delta\hat{L}$ is a relative error of the computed displacement vectors and $\Delta\hat{\alpha}$ is an average angle between vectors. The last column contains the execution time of the method per image. The methods were implemented in Matlab and executed on a PC with a 2.6 GHz CPU.



(a) (b)

Fig. 2. Synthetic data: (a) An original image with an individual exposure; (b) A synthetic double-exposed image.

TABLE I. Performance of the methods on the synthetic images.

| The method and the data set | Total correct, [%] | $\delta\hat{L}$, [%] | $\Delta\hat{\alpha}$ | $t$, [s] |
|---|---|---|---|---|
| Autocorrelation, the 1st synthetic set | 83.3 | 2.9 | -0.01 | 18 |
| Autocorrelation, the 2nd synthetic set | 72.5 | 9.3 | -0.04 | 17 |
| PIP matching, the 1st synthetic set | 99.8 | 0.0 | 0.00 | 163 |
| PIP matching, the 2nd synthetic set | 80.1 | 4.3 | 0.00 | 220 |

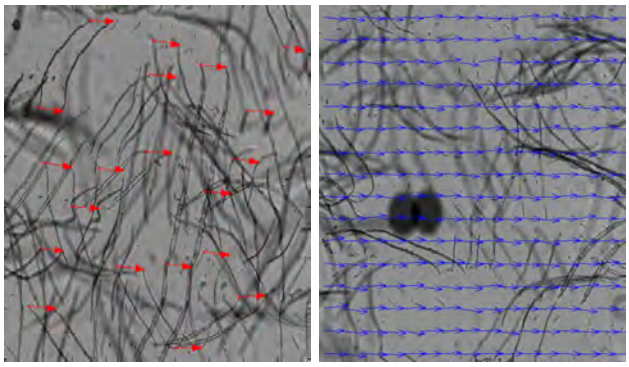
Since the displacement in the first set of images was linear the results of the both methods were better than with the second set. The relative error of the vector length computation for the autocorrelation method reached 2.9%, while the PIP matching produced the correct results in almost all the cases. The second data set is more similar to the real data. The relative errors of the vector length computation for the second data set were 4.3% for the PIP matching method and 9.3% for the autocorrelation method. On the synthetic data, the PIP matching outperformed the autocorrelation technique, but it required much more time for computation.

### B. Real-world test data

There were two data sets of the real data, provided by the CEMIS-OULU Laboratory. The images were captured with a CCD camera QImaging Retiga-2000R. In the first set of images, there were 100 images (896x704 pixels) of the birch pulp captured with 5x magnification and 100 us delay between the short laser pulses. The image in the first data sets had very little distortions in the pulp flow. Images in the second set were taken with different measurement setup. Most of the fibers in the images are blurred. This data set contained 80 images (400x300 pixels) of eucalyptus pulp, captured with 2.5x magnification and 2 ms delay between the light pulses. The GT was produced for each image manually by a non-expert. It contained a set of vectors, each corresponding to the displacement of fibers in that area. An example of the GT markings is presented in Fig. 3(a). The GT contains the length and location of the vectors.

The vectors resulting from the computation were compared to the GT by searching for the closest GT vector. If the length and the angle of the vector of the one for the GT, the vector was considered as correctly computed. The accuracy of the result is computed as the ratio between the vectors' lengths. The results for the first data sets with 8500 displacement vectors and for the second data set with 7480 displacement vectors are presented in Table II. The second column presents the

percentage of correctly computed vectors. $\delta\hat{L}$ is the relative error between the computed vector length and the nearest vector length in the GT. $\Delta\hat{\alpha}$ is the average angle between vectors. The last column in Table II contains the execution time of the method per image.



(a) (b)

Fig. 3. Real data: (a) An example of the ground truth image; (b) Velocity vectors obtained with the PIP matching.

TABLE II. PERFORMANCE OF THE METHODS ON THE REAL IMAGES.

| The method and the data set | Total correct, [%] | $\delta\hat{L}$, [%] | $\Delta\hat{\alpha}$ | $t$, [s] |
|---|---|---|---|---|
| Autocorrelation, the 1st set | 71.1 | 12.6 | 0.00 | 17 |
| Autocorrelation, the 2nd set | 66.2 | 30.1 | 0.00 | 3 |
| PIP matching, the 1st set | 84.2 | 8.3 | -0.01 | 193 |
| PIP matching, the 2nd set | 74.1 | 19.7 | 0.01 | 10 |

An example of the velocity vector estimation is presented in Fig. 3(b). From Table II, it can be seen that the percentage of the correctly computed vectors for both of the developed methods on the first data set is greater than on the second. In 5% of the cases, the global displacement was computed incorrectly and caused errors in the local displacement estimation. The PIP matching outperformed the autocorrelation similarly to the experiments on the synthetic data. The accuracy for the PIP matching technique is higher than for the autocorrelation method. However, the autocorrelation required less computational time, since the PIP matching was implemented without optimization.

## V. CONCLUSION

Two methods for estimating local pulp flow velocity were compared: the PIP matching and the autocorrelation technique. A set of experiments was performed on two synthetic data sets and two real data sets with the manually marked ground truth. On the synthetic data sets, the PIP matching demonstrated on the average an accuracy of 90.0% whereas the accuracy of the autocorrelation technique was 77.9 %. On the real-world data sets, the PIP matching and the autocorrelation methods achieved on the average an accuracy of 79.2% and 68.7%, correspondingly. The PIP matching method outperformed the autocorrelation method for the estimation of the local displacements for each data set. However, the autocorrelation method requires less computation time than the PIP matching method.

## ACKNOWLEDGMENTS

The research was carried out in the PulpVision project (TEKES Projects No. 70010/10 and 70040/11) funded by the European Union and the participating companies. The authors wish to acknowledge Kyösti Karttunen from the Measurement and Sensor Laboratory CEMIS-OULU of the University of Oulu for providing the image data.

## REFERENCES

[1] H. Fock and A. Rasmuson, "Computation of fluid and Particle Motion From a Time-Sequenced Image Pair: A Global Outlier Identification Approach," *Nordic Pulp & Paper Research Journal*, vol. 23, no. 1, pp. 120–125, 2008.

[2] N. Ray, "Computation of fluid and particle motion from a time-sequenced image pair: A global outlier identification approach," *IEEE Transactions on Image Processing*, vol. 20, no. 10, pp. 2925–2936, 2011.

[3] M. Raffel, S. Willert, S. Wereley, and J. Kompenhans, *Particle Image Velocimetry: A Practical Guide.* New York: Springer-Verlag, 2007.

[4] D. Heitz, P. Héas, E. Mémin, and J. Carlier, "Dynamic consistent correlation-variational approach for robust optical flow estimation," *Experimental Fluids*, vol. 45, no. 4, pp. 595–608, 2008.

[5] T. Corpetti, D. Heitz, G. Arroyo, E. Mémin, and A. Santa Cruz, "Fluid experimental flow estimation based on an optical-flow scheme," *Experimental Fluids*, vol. 40, no. 1, pp. 80–97, 2006.

[6] T. Liu and L. Shen, "Fluid flow and optical flow," *Journal of Fluid Mechanics*, vol. 614, pp. 253–291, 2008.

[7] B. Horn and B. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.

[8] D. Sun, S. Roth, and M. Black, "Secrets of optical flow estimation and their principles," pp. 2432–2439, 2010.

[9] L. Ziyun and L. Wei, "The Compensated HS Optical flow Estimation Based on Matching Harris Corner Points," in *2010 International Conference on Electrical and Control Engineering, ICECE*, Hefei, China, June 2010, pp. 2279–2282.

[10] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the Fourth Alvey Vision Conference*, 1988, pp. 147–151.

[11] R. D. Keane and R. J. Adrian, "Optimization of particle image velocimeters. Part I: Double pulsed systems," *Measurement Science and Technology*, vol. 1, no. 11, pp. 1202–1215, 1990.

[12] H. Huang, "An extension of digital PIV-processing to double-exposed images," *Experiments in Fluids*, vol. 24, pp. 364–372, 1998. [Online]. Available: http://dx.doi.org/10.1007/s003480050184

[13] M. Sorokin, "Image-based characterization of the process flows in pulping," Master's thesis, Lappeenranta University of Technology, 2012.

[14] R. D. Keane and R. J. Adrian, "Theory of cross-correlation analysis of PIV images," *Applied Scientific Research*, vol. 49, pp. 191–215, 1992. [Online]. Available: http://dx.doi.org/10.1007/BF00384623

[15] H. T. Huang, H. E. Fiedler, and J. J. Wang, "Limitation and improvement of PIV," *Experiments in Fluids*, vol. 15, pp. 168–174, 1993. [Online]. Available: http://dx.doi.org/10.1007/BF00189883

[16] S. Butterworth, "Theory of filter amplifiers," *Experimental wireless and the wireless engineer*, vol. 7, pp. 536–541, 1930.

[17] J. Li, X. Yang, and J. Yu, "Compact support Thin Plate Spline algorithm," *Journal of Electronics (China)*, vol. 24, no. 4, pp. 515–522, 2007. [Online]. Available: http://dx.doi.org/10.1007/s11767-005-0236-1

[18] G. Donato and S. Belongie, "Approximate thin plate spline mappings," in *Proceedings of the 7th European Conference on Computer Vision-Part III, ECCV*. London, United Kingdom: Springer-Verlag, 2002, pp. 21–31.

[19] G. Wolberg, *Digital Image Warping*. IEEE Computer Society Press, 1990.

# JOINT ANALYSIS OF ELECTROENCEPHALOGRAM, ELECTROMYOGRAM, AND TREMOR IN THE EARLY STAGE OF PARKINSON'S DISEASE

Sushkova O.S.[1], Gabova A.V.[2], Karabanov A.V.[3],
Kershner I.A.[4], Obukhov K.Y.[4], Obukhov Y.V.[1]

o.sushkova@mail.ru, Mokhovaya 11-7, Moscow, 125009, Russia,
[1] Kotel'nikov Institute of Radioengineering and Electronics of RAS, Moscow, Russia
[2] Institute of Higher Nervous Activity and Neurophysiology of RAS, Moscow, Russia
[3] Scientific Centre of Neurology of RAS, Moscow, Russia
[4] Moscow Institute of Physics and Technology, Moscow, Russia

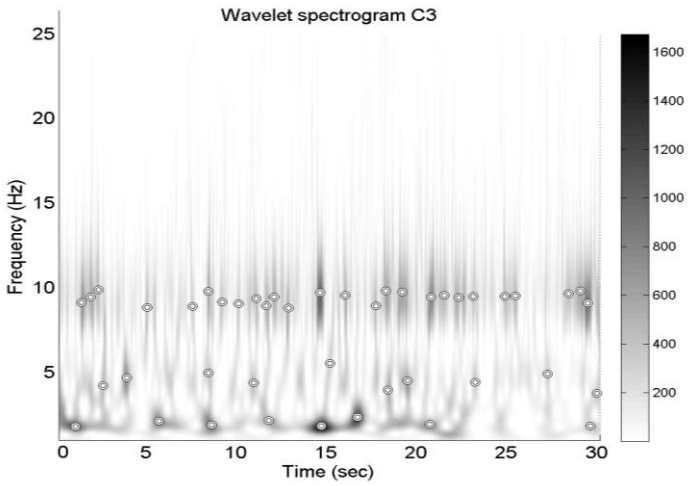
***Abstract* — Electroencephalogram, electromyogram and tremor of the non-treated early stage Parkinson's disease (PD) patients were investigated jointly. The control group and the 1st stage of Hoehn-Yahr scale patients group were consisted of 16 and 25 people respectively. The time-frequency distribution of extrema points of the electroencephalogram, envelope of electromyogram and tremor wavelet spectrograms were analyzed to extract quantitative PD features. Frequency synchronization of electroencephalogram, electromyogram and tremor were found out.**

***Keywords— Parkinson's disease, mechanical tremor, electroencephalogram, electromyogram, accelerometer, frequency synchronization, wavelet spectrogram, electromyogram envelope.***

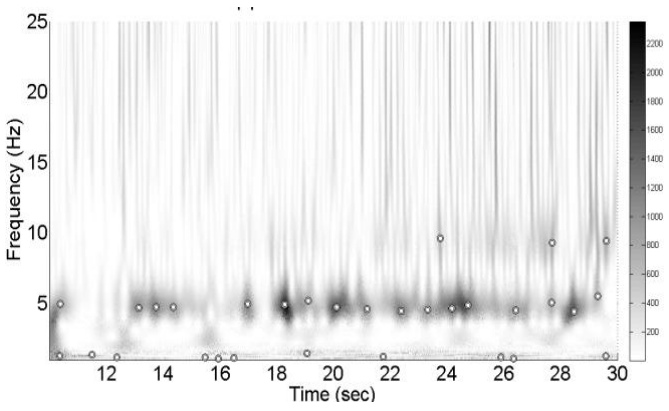
## I. INTRODUCTION

One of the ways of search for features of Parkinson's disease (PD) is a joint analysis of signals of different modalities. There are electroencephalogram (EEG) which shows the electrical brain activity, electromyogram (EMG), and mechanical tremor (MT) measured by accelerometers which are shown the mechanical movement disorders. Such analysis can lead to the understanding of the quantitative characteristics of the time-frequency structure of the EEG, EMG, and MT as well as to a more reliable recognition of PD in the early stages. The time-frequency synchronization of EEG, EMG and MT can be estimated by the time-frequency distribution of the extrema of wavelet spectrograms of the EEG [1], EMG, and MT and the envelope of amplitude-modulated high-frequency EMG.

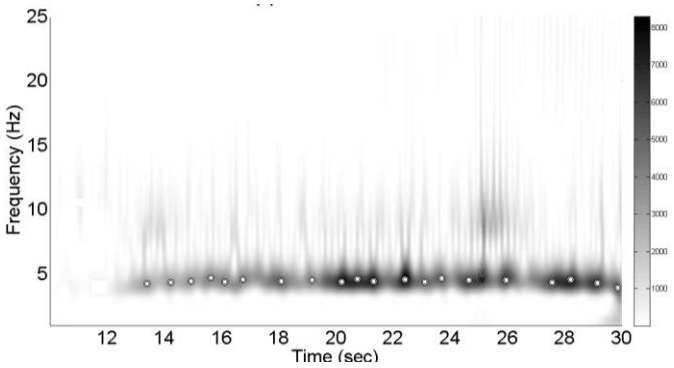
For obtaining the information about synchronization these signals, the Morlet wavelet spectrograms of EEG, EMG, and MT were computing. Wavelet spectrograms and the extrema points of the EEG, EMG envelope, and MT are presented in figure 1a, b, c, respectively.




a




b




c

Fig. 1. Wavelet spectrum of EEG (a), shape of EMG (b) and MT (c)

The method is based on the idea that the efforts of the muscles act create movement, which is similar in appearance to the flow of the curve, the EMG envelope. Since the information about hand tremor is not in the EMG signal, but in its shape, so one should compute a shape EMG. EMG shape is calculated using the Hilbert transform [2].

To isolate the amplitude and phase of an arbitrary signal u(t) (the modulated high-frequency signal) you need to create on its basis the analytical signal (1):

$$w(t) = u(t)+iv(t) \qquad (1).$$

The real part of the analytical signal is the original signal u(t). The imaginary part of w(t) is a Hilbert transform of u(t). Compute Hilbert transform as follows (2):

$$v(t) = \int_{-\infty}^{+\infty} \frac{u(\tau)}{\pi(t-\tau)} d(\tau) \qquad (2).$$

Representing (1) in a representative form (3), you can identify an shape (4)

$$w(t) = u(t)+iv\,(t) = a(t)e^{i\pi(\omega t)} \qquad (3),$$

where a(t) is the shape of the signal

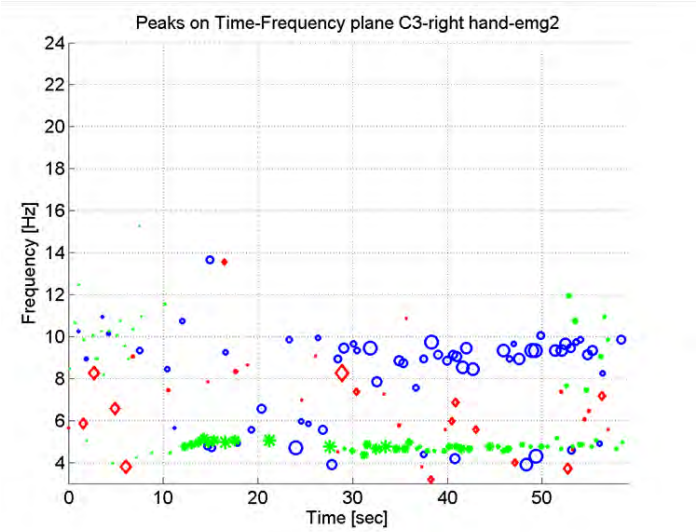
$$a(t) = \sqrt{(u(t))^2 + (v(t))^2} \qquad (4).$$

Digitized records were processed by EMG Butterworth filter of the fourth order to remove frequencies below 60 Hz.

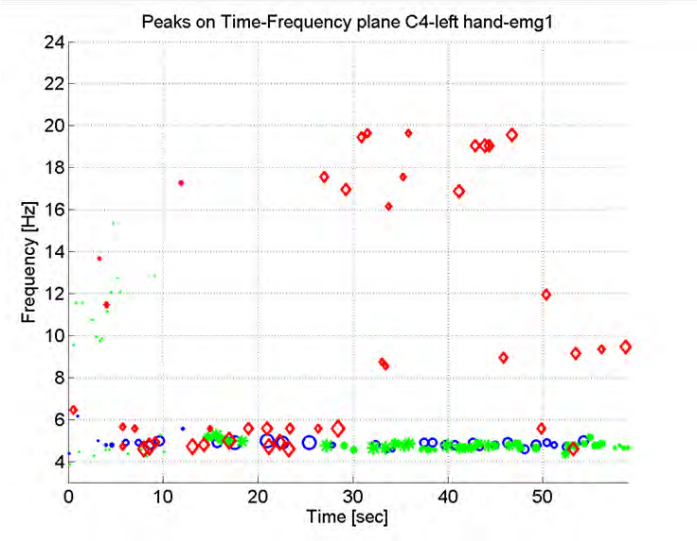
The calculation of a wavelet spectrogram of the electroencephalogram (EEG) of leads C3 and C4 of the patient with Parkinson's disease at the early stage was calculated as well. These areas are responsible for motor functions. On the spectrograms obtained, the time-frequency coordinates of the local maxima were defined.

Movement of limbs may be registered by the accelerometers set on the upper extremities (hands on). The wavelet spectrogram of the accelerometer of left and right hands to find the coordinates of the local maxima were calculated.

On the Figure 2b one can see, that the extrema in the motor area of the right hemisphere were connected with the extremes of MT and EMG. On the contrary, such connectivity was not in clinically healthy hemisphere (Fig. 2a). Magnitudes of MT extrema of the left hand was on 2 orders smaller than that of the right hand.



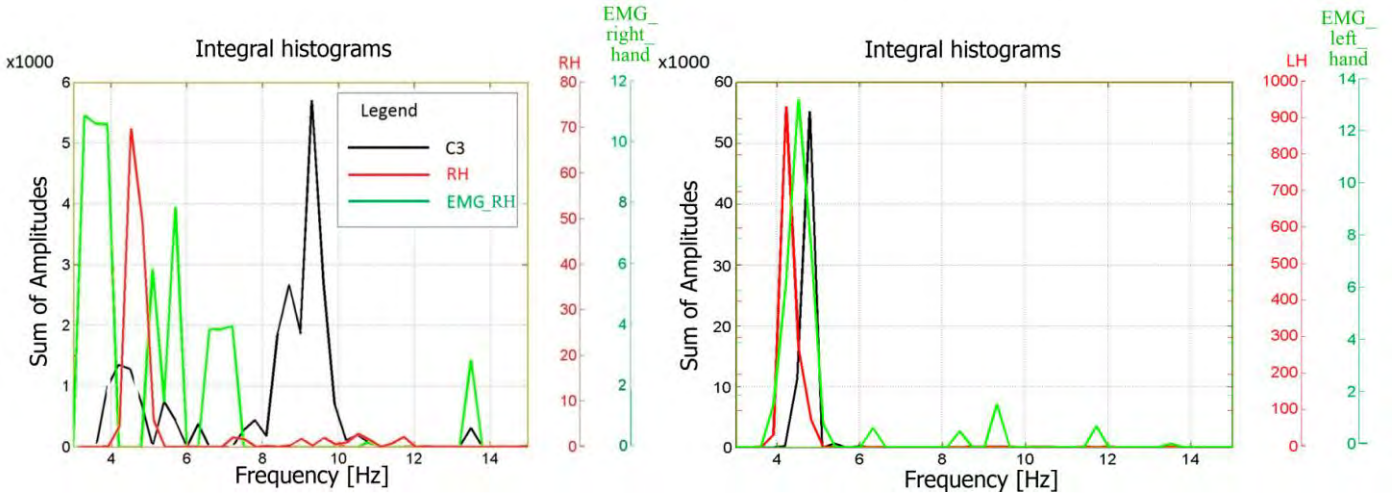

a




b

Fig. 2. Wavelet spectrograms extrema points in the time-frequency plane of EEG leads in the motor cortex of the brain C3 (circles) and contralateral MT (asterisks) and EMG (diamonds) of patient of the first stage of PD on a qualitative scale of Hoehn-Yahr (a) and maxima in the time-frequency range between hemispheric symmetrical C4 leads and contralateral MT and EMG of patient of the first stage of PD on a qualitative scale of Hoehn-Yahr (b)

Figure 3a and 3b illustrates the interhemispheric asymmetry of the motor cortex shown by time-frequency extrema points distribution, and the connectivity of 4-6 Hz electroencephalogram rhythm in one cortex area with the contralateral electromyogram and tremor.




a b

Fig 3. Frequency histograms of EEG, EMG, MT (a, b)

The coincidences with the clinical diagnosis for joint electroencephalogram, electromyogram and tremor investigations is equal to 96 percent's for the 1st stage PD patients and 87,5 percent's for the control group (see Table 1).

TABLE I. ANALYSIS OF EEG, MT AND THE AVERAGE CORRELATIONS FOR C3 AND C4 LEADS

AND THEIR MEAN-SQUARE DEVIATIONS

| № | Clinical diagnosis | EEG, MT analysis result | $A_\theta/A_\alpha$ | | LH/RH or RH/LH | r(C4)/r(C3) or r(C3)/r(C4) | σ(C3)/σ(C4) or σ(C4)/σ(C3) | R |
|---|---|---|---|---|---|---|---|---|
| | | | C3 | C4 | | | | |
| **Pathients** | | | | | | | | |
| 1. | 1 | 1 | 0,6 | 0 | 0,30 | 0,98 | 1,25 | 0,96 |
| 2. | 1 | 1 | 0,8 | 0,6 | 0,06 | 0,77 | 1,23 | 1,41 |
| 3. | 1 | 1 | 0,5 | 0,95 | 0,06 | 0,88 | 0,89 | 1,44 |
| 4. | 1 | 1 | 0 | 0 | 0,15 | 1,23 | 0,71 | 0,93 |
| 5. | 1 | 1 | 0 | 0,3 | 0,12 | 0,99 | 1,03 | 0,93 |
| 6. | 1 | 1 | 0,7 | 1,7 | 0,00 | 0,55 | 1,32 | 2,16 |
| 7. | 1 | 1 | 0,47 | 0 | 0,00 | 0,72 | 0,89 | 1,14 |
| 8. | 1 | ? | 0 | 0 | 0,88 | 1,04 | 1,05 | 0,13 |
| 9. | 1 | 1 | 0,4 | 0 | 0,29 | 1,40 | 0,70 | 0,96 |
| 10. | 1 | 1 | 0 | 33 | 0,03 | 0,86 | 1,24 | 33,02 |
| 11. | 1 | 1 | 125 | 1,3 | 3,57 | 1,45 | 1,09 | 125,03 |
| 12. | 1 | 1 | 3 | 0 | 0,90 | 1,23 | 0,69 | 3,03 |
| 13. | 1 | 1 | 2,17 | 70 | 1,05 | 0,42 | 0,57 | 70,04 |
| 14. | 1 | 1 | 0 | 0 | 0,02 | 0,95 | 1,29 | 1,02 |
| 15. | 1 | 1 | 0 | 0 | 0,04 | 0,84 | 1,50 | 1,09 |
| 16. | 1 | 1 | 0 | 0 | 0,00 | 1,12 | 0,89 | 1,01 |
| 17. | 1 | 1 | 1,3 | 0,8 | 0,03 | 0,96 | 0,96 | 1,81 |
| 18. | 1 | 1 | 1,15 | 0 | 0,01 | 0,92 | 1 | 1,52 |
| 19. | 1 | 1 | 0 | 0 | 0,06 | 0,81 | 2,10 | 1,46 |
| 20. | 1 | 1 | 1,09 | 2,34 | 0,10 | 1,36 | 0,81 | 2,76 |
| 21. | 1 | 1 | 0,31 | 0,25 | 0,03 | 1,01 | 0,94 | 1,05 |
| 22. | 1 | 1 | 0 | 0,48 | 0,02 | 1,16 | 0,65 | 1,16 |
| 23. | 1 | 1 | 0,17 | 0,3 | 0,04 | 0,84 | 1,48 | 1,14 |
| 24. | 1 | 1 | 0,84 | 0,18 | 0,02 | 1,17 | 0,81 | 1,33 |
| 25. | 1 | 1 | 0 | 0 | 0,07 | 1,07 | 0,5 | 1,06 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Control group** | | | | | | | | |
| 1. | 0 | ? | 0 | 0 | 0,33 | 0,85 | 1,76 | 1,0 |
| 2. | 0 | 0 | 0 | 0 | 1,7 | 0,86 | 1,32 | 0,8 |
| 3. | 0 | 0 | 0 | 0 | 1,43 | 0,97 | 0,93 | 0,4 |
| 4. | 0 | 0 | 0 | 0 | 0,91 | 1,28 | 0,83 | 0,3 |
| 5. | 0 | 0 | 0 | 0 | 1,25 | 1,04 | 1,2 | 0,32 |
| 6. | 0 | 0 | 0 | 0 | 1 | 0,95 | 0,94 | 0,08 |
| 7. | 0 | 0 | 0 | 0 | 0,67 | 1,01 | 1,3 | 0,45 |
| 8. | 0 | 0 | 0 | 0 | 1,2 | 1,03 | 0,88 | 0,23 |
| **Abstract ideal person** | | | | | | | | |
| | | | 0 | 0 | 1 | 1 | 1 | |

## II. Conclusion

At least two main features of Parkinson's disease (PD) in the early stage were detected: 1) time-frequency characteristics of the motor cortex zones hemisphere asymmetry; 2) the arising EEG rhythm in these zones in the frequency range of 4-6 Hz and its connectivity with the EMG and mechanical tremor of the contra lateral limbs in the tremor form of PD.

## Acknowledgment

We acknowledge a partial financial support from the Russian Foundation for Basic Research, grant No 12-02-00611-a.

## References

[1] Y.V. Obukhov, M.S. Korolev, A.V. Gabova, G.D. Kuznetsova, M.V. Ugrumov, “Method of early stage Parkinson's disease electroencephalography diagnostics” // RF patent. - 2484766, 20.06.2013, (in Russian)

[2] D.E. Vakman, L.A. Weinstein, “The amplitude, phase, frequency - the basic concepts of the theory of vibrations” // Advances in Physical Sciences. – 2000. - Vol. 123, No.4, - P. 657-682.

Proceedings of the 9th Open German-Russian Workshop on Pattern Recognition and Image Understanding
Dec 1-5, 2014, Koblenz, Germany

# Latent Space Gaussian Process Gaze-Tracking

Nicolai Wojke, Jens Hedrich and Detlev Droege[1]

*Abstract*— **Commercial gaze-tracking devices provide accurate measurements of the visual gaze and are applied to a broad range of problems in marketing, human-computer interaction, and health care technology. In some applications commercial systems are either unavailable or unaffordable. Therefore, developing low cost solutions using *off the shelf* components is worthwhile. In the paper at hand, we apply a hierarchy of Gaussian processes, a class of probabilistic function regressors, to the problem of visual gaze-tracking for consumer grade cameras. Gaussian process latent variable models lead to a lower dimensional manifold which represents the gaze space. Finally, a Gaussian process mapping from screen coordinates to gaze manifold enables us to seek for the users visual gaze point given a previously unseen eye-patch. In our experiments, we achieve mean errors of approximately 2 cm for a consumer grade webcam that is positioned 30–40 cm in front of the user.**

## I. INTRODUCTION

The quality and accuracy of common eye and gaze-tracking devices provides a very solid base for gaze based interaction systems. However, they are usually not precise enough to select tiny screen elements of common graphical user interfaces, leading to the development of specific user interfaces for gaze interaction. Such interfaces are designed to be used with a much lower need for accuracy, not only due to the technical limitations, but also to account for the users capabilities [8]. Gaze interaction systems are often used by disabled users, who, depending on the type and severeness of their handicaps, might also not be able to position their gaze as exact as would be required by conventional user interfaces.

Depending on the service capability of the health care system in different countries, affected persons might often not be funded to obtain a gaze-tracking system. Therefore, several research groups work on systems using inexpensive *off the shelf* devices to compile simple gaze-tracking systems like e.g. [16]. Such systems will perform at a significantly lower accuracy than the established systems, but still good enough for being used with gaze interaction systems. While the goal of using cheap (US$ 20-30) *web cams* could not yet be met, system costs of less than US$ 300 are currently realistic.

A number of algorithms has been published to determine the pupil center in eye tracking systems like [3], [7], [9], and others. These algorithms were shown to work good on medium to high resolution images, for low resolution input however they leave room for improvement. Given the position of the pupil center in an image, the visual gaze can be computed from the known geometry of camera setup. These computations are prone to inaccuracies in pupil center detection. We therefore circumvent pupil center detection and geometric computations by learning (1) a low dimensional latent manifold of eye-patches, and (2) the transformation from screen coordinates to latent space using Gaussian process regression. Given such a mapping, we seek for the unknown visual gaze by maximizing the normalized cross correlation between expected and observed eye patches.

The remainder of the paper is structured as follows: In section II we give an overview of related work in the field of gaze-tracking. In section III we then introduce Gaussian processes as the main framework for our gaze-tracking system. In section IV we describe the details of our methodology and section V describes the experiments we used to evaluate our approach. In section VI we summarize the paper and conclude from our results.

## II. RELATED WORK

Numerous approaches to gaze tracking from video images use the pupil center as primary feature. The visual gaze is then recovered using a geometric model. A thorough overview of methods for pupil center detection is given by Hansen and Ji [4]. Most of these algorithms were developed with specific setups and conditions in mind, as the usage scenarios for eye detection and tracking are quite different.

In the context of iris recognition, close-up high resolution images of the subjects eye are self-evident. Appropriate oculars, specific illumination and appropriate cameras deliver almost perfect images of the iris, however accept a distraction of the eye using extra illumination. Such images were the basis for Daugmans algorithm described in [2]. He defines the integrodifferential operator, based on a circular integral around the currently estimated pupil/iris center. Increasing the radius for this measure gives notable changes at the pupil and iris border, providing new estimates for the radii. These can then be used to find new center estimates.

In [3] two methods to determine the center of a pupil are presented. In both cases the first step is to find the transition from dark luminance values in the pupil to brighter values in the iris. This transition is considered to be monotonic when approximated by a polynomial to determine the pupil rim with sub-pixel accuracy. The *coordinates averaging* approach then forms horizontal and vertical scan lines between corresponding rim points. The middle of such lines should lie on the vertical resp. horizontal center axis of the pupil. The pupil center is estimated by averaging the horizontal mid points for the x-component and the vertical mid points for the

[1]Active Vision Group, Institute for Computational Visualistics, University of Koblenz-Landau, 56070 Koblenz, Germany {nwojke, jenshedrich, droege}@uni-koblenz.de

y-component. For images without artifacts from e.g. glints in the pupil this approach gives rather accurate results. For their *circle approximation* method the rim points are used to estimate a circle. Obvious outliers to the circle fitting are weighted down to find a good approximation after a few iterations.

An approach named *Starburst* is presented in [7]. After preprocessing the image, which includes the removal of glints, an initial rough guess for the pupil center is made. From here a number of radial rays is followed and observed for a significant jump in intensity, determined by a difference threshold. These pupil rim points are the origin for a number of secondary rays which are sent like a fan in the opposite direction in a range of $\pm 50°$. These rays provide additional points on the rim.

Peréz et al. [11] describe a similar technique, starting from an initial pupil center guess calculated from the center of gravity of those pixels which are considered pupil pixels by using a threshold. However, only primary rays are used for the pupil rim detection, which is done by employing a Laplace filter. If the detected points are not equidistant from the estimated center, a new center is chosen by using the mid points of the diagonals (where only diagonals of reasonable length are considered). This calculation is repeated until the points are equidistant or some iteration limit is reached.

The algorithm described in [9] is used in the FreeGaze system and called *double ellipse fitting*. A first guess for pupils is done by segmenting the image and looking for dark, round regions. From their center circular rays are followed and rim points are detected by a sudden raise of intensity, similar to Starburst. An ellipse is fitted to these points and its center is used as starting point for a second run. Here, the number of rays is doubled and points having significantly different distance from the center than can be expected are discarded. The remaining points are used for a second ellipse fitting.

This is not a comprehensive list of published algorithms on pupil center detection, several other approaches do exist, e.g. [12], [14], [10], but throughout follow similar principles.

Other methods involve appearance based approaches where machine learning techniques are applied to recover the visual gaze based on a set of image features. Zhang et al. [19] extract a set of color and gradient features from eye-patches taken from a head-mounted camera and learn the mapping to screen coordinates using a 2-layer regression neural network. Similarily, [18] learn a neural network based on the relative position of the pupil, cornea, and an equalized histogram of the image patch surrounding the eye. Williams et al. [17] learn a relevance vector machine on preprocessed eye-patches to obtain probabilstic estimates of visual gaze. The so obtained visual gaze is filtered over time using a linear Kalman filter with stationary motion model. Prisacariu and Reid [13] learn a shared space Gaussian process latent variable model, a variant of non-linear factor analysis, on segmented binary images of the eye. Preprocessing and feature extraction in all of these methods are determined by an expert in the field.

## III. GAUSSIAN PROCESSES

Since most of our work is based on Gaussian processes (GPs), we first introduce GPs following the function-space view described in [15]. Given data $\{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_N, y_N)\}$ consisting of inputs $\boldsymbol{x}_i$ and observations $y_i$, drawn from a noisy process

$$y_i = f_i + \epsilon_i = f(\boldsymbol{x}_i) + \epsilon_i$$

with noise $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, a Gaussian process estimates the posterior distribution of function $f$. Therefore, Gaussian processes are distributions over functions. Formally, in the GP framework the stochastic properties of function $f$ are characterized by a mean and a covariance function

$$m(\boldsymbol{x}) = \mathrm{E}[f(\boldsymbol{x})] \qquad k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \mathrm{Cov}[f(\boldsymbol{x}_i, \boldsymbol{x}_j)]$$

such that for any given subset of deterministic inputs $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$ random variables $(f_1, \ldots, f_N)$ have $N$-dimensional joint normal distribution

$$p(f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_N) \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) = \mathcal{N}(\boldsymbol{m}, \boldsymbol{K}). \qquad (1)$$

In the following we write $f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}_i, \boldsymbol{x}_j))$. Inference for GPs is possible due to the marginalization property of the Gaussian distribution. From Equation 1 follows that for training data $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]^{\mathrm{T}}$, $\boldsymbol{y} = (y_1, \ldots, y_N)^{\mathrm{T}}$ and previously unseen test point $\boldsymbol{x}_*$ with unknown, noisy function value $f_* = f(\boldsymbol{x}_*)$ the joint distribution is Gaussian. Assuming $m(\boldsymbol{x}) = 0$:

$$p(\boldsymbol{y}, f_*) = \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{K} + \sigma^2 \boldsymbol{I} & \boldsymbol{k}_* \\ \boldsymbol{k}_*{}^{\mathrm{T}} & k_* \end{bmatrix}\right),$$

where matrix $\boldsymbol{K}_{i,j} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ contains the covariance function evaluations and similarily $\boldsymbol{k}_{i,*} = k(\boldsymbol{x}_i, \boldsymbol{x}_*)$ and $k_* = k(\boldsymbol{x}_*, \boldsymbol{x}_*)$. Now, the posterior of unknown function value $f_*$ can be obtained by building the conditional Gaussian distribution:

$$f_* = \boldsymbol{k}^{*\mathrm{T}} (\boldsymbol{K} + \sigma^2 \boldsymbol{I})^{-1} \boldsymbol{y},$$
$$\sigma_* = k^* - \boldsymbol{k}_*{}^{\mathrm{T}} (\boldsymbol{K} + \sigma^2 \boldsymbol{I})^{-1} \boldsymbol{k}_*.$$

There exist many problem dependent choices for the covariance function. In all experiments we used the exponentiated quadratic covariance function

$$k_{\mathrm{EQ}}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \alpha^2 \exp\left(-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{x}_j)^{\mathrm{T}} \boldsymbol{\Lambda}^{-1} (\boldsymbol{x}_i - \boldsymbol{x}_j)\right) \qquad (2)$$

with parameters $\alpha$, $\boldsymbol{\Lambda}$. The exponentiated quadratic covariance function encodes *smoothness*, i.e. close points are expected to have similar function values. The parameters of the covariance function as well as noise $\sigma^2$ can be learned by maximizing the marginal likelihood

$$p(\boldsymbol{y} \,|\, \boldsymbol{X}, \sigma^2, \alpha, \boldsymbol{\Lambda}) = \int p(\boldsymbol{y} \,|\, \boldsymbol{f}, \boldsymbol{X}, \sigma^2) p(\boldsymbol{f} \,|\, \boldsymbol{X}, \alpha, \boldsymbol{\Lambda}) \,\mathrm{d}\boldsymbol{f}$$

where $p(\boldsymbol{f} \,|\, \boldsymbol{X}, \alpha, \boldsymbol{\Lambda}) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{K})$ is the Gaussian process prior on function values $\boldsymbol{f} = (f_1, \ldots, f_N)$ and

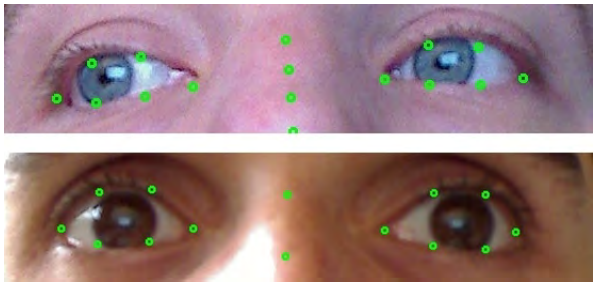


Fig. 1: Example of selected eye-patches. The green circles indicate landmarks given by the face tracking method proposed by Kazemi et al.[5].

$p(\boldsymbol{y} \mid \boldsymbol{f}, \sigma^2) = \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$ is the Gaussian observation likelihood. For completeness we write down the log marginal likelihood that is used for optimization:

$$\log p(\boldsymbol{y} \mid \boldsymbol{X}, \sigma^2, \alpha, \Lambda) = -\frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\left(\boldsymbol{K} + \sigma^2 \boldsymbol{I}\right)^{-1}\boldsymbol{y} - \frac{1}{2}\log|\boldsymbol{K}| - \frac{n}{2}\log 2\pi$$

where $n$ is the number of columns in $\boldsymbol{X}$. A more elaborated introduction to GPs as well as a practical implementation of Gaussian process regression can be found in [15].

## IV. LATENT-SPACE GAUSSIAN PROCESS GAZE-TRACKING

In this section we describe the methodology of the proposed gaze-tracker. We show the system architecture and illustrate how we apply Gaussian process latent variable models for learning a low dimensional manifold of eye-patches, as well as how Gaussian process regression is used for learning a mapping from screen coordinates to latent space for determining the visual gaze of the user.

The presented gaze-tracking system expects eye-patches of an interacting user who is sitting in front of a computer screen. In our setup we use a standard webcam[1] to observe the scene. In order to select the eye-patches we use the head tracking method proposed by Kazemi et al. [5]. This method uses a regression tree to detect 68 distinctive face landmarks. Six landmarks are used for each eye (c.f. Figure 1). In our test setup the landmarks are stable enough to extract stable eye-patches even for small head movements.

### A. System Overview

The proposed gaze tracker can be divided into four major steps. First, a small representative set of eye-patches and the corresponding gaze points in screen coordinates are captured during calibration phase. These patches are used to generate a lower dimensional gaze manifold using the Gaussian process latent variable model. In this specific application two latent dimensions were sufficient for capturing the relevant information. In order to identify the users gaze during tracking phase, we select nearest neighbor eye-patches in training data to initialize a non-linear optimization routine where we compare predicted eye-patches to the observed eye-patch. A simplified overview of the proposed method is given in Figure 2.

[1]Logitech Quickcam Pro 9000 Business


Fig. 2: System Overview of the gaze tracker. First, a small sample set of eye-patches and the corresponding gaze points in screen coordinates are captured. Second, the Gaussian process latent variable model [6] is used to compute a lower dimensional gaze manifold from eye-patches. Third, Gaussian process regression is used to establish a mapping from screen coordinates to eye-patches.

### B. Eye-Patch Latent Space Representation

From the image region around the eyes valuable information for determining the visual gaze may be obtained. While there exist algorithms to localize the pupil center as a primary feature [3], [7], [9], their accuracy is highly dependent on image resolution. Further, light reflections easily decrease stability of results. We therefore opt to circumvent error prone preprocessing and apply a non-linear dimensionality reduction method instead. The resulting low-dimensional manifold of eye-patches is then used as our sole feature for determining the visual gaze. In this section we describe the dimensionality reduction method that we have applied in our experiments. In section IV-C we show how to use this representation to determine the visual gaze.

Let $\boldsymbol{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N]^{\mathrm{T}}$ be a set of $N$ observations of dimensionality $D$ written as design matrix. We then may assume that the variations in the data is governed in a low dimensional manifold $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]^{\mathrm{T}}$ of dimension $Q$ with $Q << D$. In the gaze-tracking scenario, under stable lighting conditions, appearance variations in eye-patches are governed by only few parameters that describe mainly the position of the pupil, the eyelid, plus some noise. This motivates our application of (non-linear) dimensionality reduction for visual gaze discovery from image data.

Here, we apply the Gaussian process latent variable model (GP-LVM) of Lawrence et al. [6]. From calibration, we are given a set of eye-patches that make up our observations $\boldsymbol{Y}$. For dimensionality reduction, we define $D$ mappings

$$\boldsymbol{y}_j = g_j(\boldsymbol{x}_i) + \epsilon_j$$

from (unknown) latent sample $\boldsymbol{x}$ to the $j$-th element of observed eye-patch $\boldsymbol{y}$. If we assume independent Gaussian noise $\epsilon_j \sim \mathcal{N}(0, \sigma^2)$, the conditional of observed eye-patch $\boldsymbol{y}$ given latent sample $\boldsymbol{x}$ is

$$p(\boldsymbol{y} \mid \boldsymbol{x}) = \prod_{j=1}^{D} \mathcal{N}(\boldsymbol{y}_j \mid g(\boldsymbol{x}), \sigma^2)$$

and, assuming independence among observations $\boldsymbol{Y}$, the distribution of all eye-patches is

$$p(\boldsymbol{Y} \mid \boldsymbol{X}) = \prod_{i=1}^{N} \prod_{j=1}^{D} \mathcal{N}(\boldsymbol{Y}_{i,j} \mid g(\boldsymbol{X}_i), \sigma^2)$$

where $\boldsymbol{Y}_{i,j}$ is the $j$-th element of the $i$-ith observation and $\boldsymbol{X}_i$ is the $i$-th latent sample. In standard principal component analysis, mapping $g_j(\cdot)$ is linear and parametrized by the $j$-th basis vector that is found using eigen value decomposition. In GP-LVM the mapping is not modelled in parametric form, but a common Gaussian process prior is put on all mappings $g_1, \ldots, g_D$:

$$p(\boldsymbol{Y} \mid \boldsymbol{X}) = \prod_{i=1}^{N} \prod_{j=1}^{D} \mathcal{N}(\boldsymbol{Y}_{i,j} \mid \boldsymbol{0}, \boldsymbol{K} + \sigma^2 \boldsymbol{I})$$

where $\boldsymbol{K}$ is the matrix of covariance function evaluations. Parameter learning for the covariance function in GP-LVM is equivalent to the regression case in Section III. The positions of latent samples $\mathbf{X}$ is found along with covariance function parameters by maximizing the marginal likelihood.

In our experiments we have used a two dimensional latent space and the exponentiated quadratic covariance function (2). Fig. 3 shows a visualization of the learned model as well as eye-patches that are generated at different locations. Note that eye-patches corresponding to distant gaze points are placed apart in latent space. Further, the model successfully interpolates between locations where no data has been observed. Interestingly, not only the position of the pupil, but also the opening of the eyelids contains important information for estimating the visual gaze. By generating observations from the model, one can see that the eyelid is generally much more closed when the user looks at the bottom of the screen. This information is well captured by our model.

### C. Gaussian Process Gaze-Mapping

For determining the visual gaze we learn a mapping from screen coordinates to latent gaze space. For this, we use the set of gaze points $\{\boldsymbol{s}_i = (x, y)^{\mathrm{T}}\}_{i=1}^{N}$ that have been recorded during calibration phase and positions in latent gaze space $\{\boldsymbol{x}_i = (u_i, v_i)^{\mathrm{T}}\}_{i=1}^{N}$ of the corresponding eye-patches. We model the mapping with two independent GPs, one for each dimension of the latent space:

$$u = f_u(\boldsymbol{s}), \quad f_u \sim \mathcal{GP}(m(\boldsymbol{s}), k_{\mathrm{EQ}}(\boldsymbol{s}_i, \boldsymbol{s}_j)),$$
$$v = f_v(\boldsymbol{s}), \quad f_v \sim \mathcal{GP}(m(\boldsymbol{s}), k_{\mathrm{EQ}}(\boldsymbol{s}_i, \boldsymbol{s}_j)).$$

For both GPs we assume zero mean $m(\boldsymbol{s}) = 0$ and use the exponentiated quadratic covariance function (2). We then apply standard Gaussian process regression (GPR) to obtain covariance parameters by maximizing the marginal likelihood. Together with mapping

$$\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x})$$

from latent space to eye-patches described in section IV-B, we obtain a hierarchy of Gaussian processes

$$\boldsymbol{y} = \boldsymbol{h}(\boldsymbol{s}) = \boldsymbol{g}(f_u(\boldsymbol{s}), f_v(\boldsymbol{s}))$$

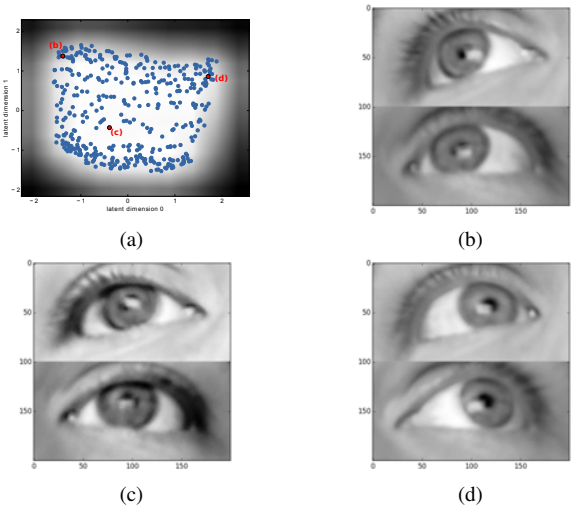



Fig. 3: (a) Visualization of the two dimensional latent gaze space. Training samples are shown as blue dots, uncertainty is gray-shaded (low uncertainty is shown in white, high uncertainty in black). Images (b)–(d) show eye-patches generated from the model at the annotated positions.

which outputs an *expected*[2] eye-patch for each location in screen coordinates. Further, we can compare the predicted eye-patch $\boldsymbol{y}_*$ of our model with any currently observed eye-patch $\boldsymbol{y}$ using normalized cross-correlation:

$$d(\boldsymbol{y}, \boldsymbol{y}_*) = \frac{\sum_i^N (\boldsymbol{y}_i \boldsymbol{y}_{*,i})}{\sqrt{\sum_i (\boldsymbol{y}_i)^2 \sum_i (\boldsymbol{y}_{*,i})^2}}. \tag{3}$$

So far we have explained how the mapping between screen coordinates and eye-patches is established. Further, we have defined normalized cross-correlation as a distance measure for computing similarity between prediceted eye-patches and the currently observed eye-patch. In order to find the visual gaze in screen coordinates during tracking, we solve a non-linear optimization problem in the following way:

1) Find $M$ nearest neighbors of newly acquired eye-patch $\boldsymbol{y}$ in training data that was collected during calibration. In our experiments we have used $M = 3$.
2) Initialize an estimate of the current gaze point in screen coordinates $\boldsymbol{s}_*$ as mean of the screen coordinates of the $M$ nearest neighbors in the training data.
3) Iteratively maximize the normalized cross-correlation (3) between predicted eye-patch $\boldsymbol{y}_* = \boldsymbol{h}(\boldsymbol{s}_*)$ and observed eye-patch $\boldsymbol{y}$ until convergence or a maximum number of iterations has been reached.

In our experiments we have used the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method for solving this optimization problem. Gradients have been computed numerically.

[2]Note that we do not propagate uncertainty from the GPR mapping through the GP-LVM. Therefore, the resulting eye-patch is not an expectation in the maximum likelihood sense.

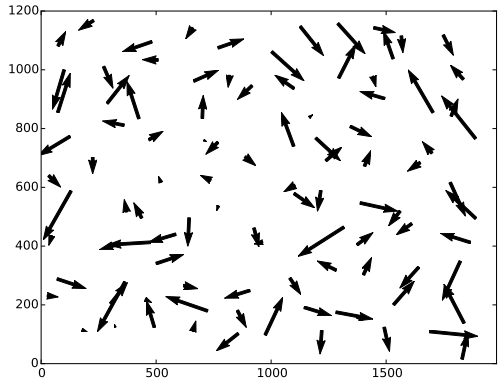



Fig. 4: Displacement field between ground truth and predicted position (in screen coordinates).

| | mean | std | max |
|---|---|---|---|
| $\epsilon_x$ | 12.865 mm | 8.881 mm | 38.903 mm |
| $\epsilon_y$ | 13.405 mm | 9.405 mm | 39.930 mm |
| $\epsilon_{\text{diag}}$ | 20.169 mm | 10.137 mm | 45.494 mm |

TABLE I: Mean, standard deviation and maximal error along the horizontal axis $\epsilon_x$, vertical axis $\epsilon_y$ and diagonal axis $\epsilon_{\text{diag}}$.

## V. EXPERIMENTS

The setup of our experiments was based on a consumer grade webcam which was positioned between the computer screen and the user. The distance from the camera to the user was approximate 30–40 cm. The user faced a 24 inch computer screen with a resolution of $1980 \times 1200$ pixels. The total length of the visible horizontal axis was $51.84$ cm and the total length of the visible vertical axis was $32.40$ cm. Further, the distance between the user and screen was approximate $60$ cm. During data acquisition, the user focused specific points on the computer screen that were highlighted by our evaluation software. The screen positions were saved together with the corresponding eye-patches taken from webcam images. We then partitioned the data into 400 samples used for training and 100 samples for evaluation.

The learned latent-space is shown in Fig. 3. Note that eye-patches corresponding to distant gaze points are placed apart in latent space. Further, the model successfully interpolates between locations where no data was observed. Interestingly, not only the position of the pupil, but also the opening of the eyelids contains important information for estimating the visual gaze. By generating observations from the model, one can see that the eyelid is generally much more closed when the user looks at the bottom of the screen. This information is well captured by our model.

Table 1 summarizes the results of our experiment. The mean error along the horizontal $\epsilon_x$ and the vertical axis $\epsilon_x$ are similar. In total we achieve errors between 12 and 45 mm. Fig. 4 depicts the displacement field between ground truth and predicted gaze points. While the error vectors in the lower right corner of the screen appear slightly longer than average, it can be seen that the model does not favor any specific region of the screen.


Fig. 5: (a), (c): Eye-patches of a previously unseen user. (b), (d): Predicted eye-patch of our model at the ground truth screen location.

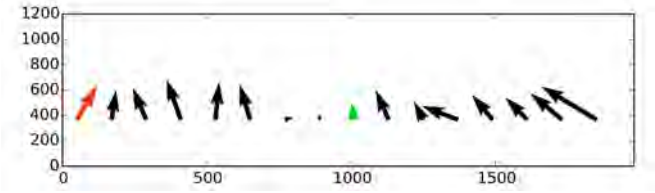



Fig. 6: Displacement field for a previously unseen user. The red arrow corresponds to the eye-patch shown in Fig. 5a, the green arrow corresponds to the eye-patch in Fig. 5c.

In an additional experiment we have tested how the trained model reacts to eye-patches of a previously unseen person with different appearance, i.e. different eye color, eye shape, and illumination when no recalibration is performed. Two eye-patches of this second experiment are shown in Fig. 5 together with the predicted eye-patches of our model at the corresponding ground truth screen locations. Note that the input images are notably darker and that eye sizes differ. This is because the second dataset was taken at a different time of day with slightly changed camera setup. As can be seen from the displacement field in Fig. 6, the model consistently predicts positions lower than ground truth. The average error was $36.590$ mm along the horizontal and $55.270$ mm along the vertical axis. It is not surprising that the model cannot be directly applied to estimate the visual gaze as the camera setup and therefore the mapping to screen coordinates has been altered. However, predictions are still within 6 cm of ground truth. Therefore, keeping the learned latent gaze space and only updating the mapping from screen coordinates may be possible when a new user is presented to the system. This requires further evaluation and is an open question for future work.

## VI. CONCLUSION

We have proposed a hierarchy of Gaussian processes for gaze-tracking. Using a head tracking method we extract stable eye-patches of the user from a consumer grade webcam. We then apply a GP-LVM to learn a two dimensional feature space that contains the relevant information for gaze-tracking. Using standard Gaussian process regression we establish a mapping between screen coordinates and the learned latent space of the GP-LVM to generate eye-

patches for each gaze point hypothesis in screen coordinates. During tracking, we use non-linear optimization to find the visual gaze point by comparing predicted eye-patches of our model with the observed eye-patches using normalized cross-correlation. The system was evaluated on a dataset of 500 samples out of which 100 were used solely for testing where we achieved errors between 10 and 45 mm. While this is not as accurate as professional gaze-tracking hardware, the method provides reasonable results for consumer grade hardware.

There is ample opportunity for future work. First, the proposed hierarchy of Gaussian processes could be jointly optimized in a variational framework that accounts for uncertainty propagation to obtain full posterior distributions of eye-patches [1]. Using this posterior, one can seek the maximum a posteriori estimate of the gaze point instead of using normalized cross correlation as distance measure. Second, our non-parametric machine-learning approach to gaze-tracking allows for easy intergration of further data that has been disregarded in our experiments. For example, locations of face landmarks could be used for head pose estimation which then could be integrated into the tracker to make the system more robust against head movements and perspective distortion.

## REFERENCES

[1] Andreas C. Damianou and Neil D. Lawrence. Deep gaussian processes. In *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2013, Scottsdale, AZ, USA, April 29 - May 1, 2013*, volume 31 of *JMLR Proceedings*, pages 207–215. JMLR.org, 2013.

[2] John Daugman. How iris recognition works. *Circuits and Systems for Video Technology, IEEE Transactions*, 14(1):21–30, 2004.

[3] Gintautas Daunys and Nerijus Ramanauskas. The accuracy of eye tracking using image processing. In *NordiCHI '04*, pages 377–380, New York, NY, USA, 2004. ACM.

[4] Dan Witzner Hansen and Qiang Ji. In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009. (in print).

[5] Vahid Kazemi and Josephine Sullivan. One millisecond face alignment with an ensemble of regression trees. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1867–1874, June 2014.

[6] Neil D. Lawrence, Matthias Seeger, and Ralf Herbrich. Fast sparse gaussian process methods: The informative vector machine. In Suzanna Becker, Sebastian Thrun, and Klaus Obermayer, editors, *Advances in Neural Information Processing Systems 15 [Neural Information Processing Systems, NIPS 2002, December 9-14, 2002, Vancouver, British Columbia, Canada]*, pages 609–616. MIT Press, 2002.

[7] Dongheng Li, David Winfield, and Derrick J. Parkhurst. Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. In *CVPR '05*, pages 79–86, Washington, 2005. IEEE.

[8] Päivi Majaranta and Kari-Jouko Räihä. Twenty years of eye typing. In *Proceedings of Eye Tracking Research and Applications*, pages 15–22, New York, 2002. ACM.

[9] Takehiko Ohno, Naoki Mukawa, and Atsushi Yoshikawa. Freegaze: A gaze tracking system for everyday gaze interaction. In *Proceedings of the symposium on ETRA 2002: eye tracking research and applications symposium*, pages 125–132, 2002.

[10] Kun Peng, Liming Chen, Su Ruan, and Georgy Kukharev. A robust algorithm for eye detection on gray intensity face without spectacles. *Journal of computer Science and Technology*, 5(3):127–132, 2005.

[11] A. Pérez, M.L. Córdoba, A. Garcia, R. Méndez, M.L. Munoz, J.L. Pedraza, and F. Sánchez. A precise eye-gaze detection and tracking system. In *WSCG POSTERS proceedings, February 3-7, 2003*, 2003.

[12] Ahmad Poursaberi and Babak Araabi. A novel iris recognition system using morphological edge detector and wavelet phase features. *ICGST International Journal on Graphics, Vision and Image Processing*, 05, 2005.

[13] Victor Adrian Prisacariu and Ian Reid. Shared shape spaces. In Dimitris N. Metaxas, Long Quan, Alberto Sanfeliu, and Luc J. Van Gool, editors, *ICCV*, pages 2587–2594. IEEE, 2011.

[14] H. Proenca and L.A. Alexandre. Iris segmentation methodology for non-cooperative recognition. *IEE Proceedings-Vision Image and Signal Processing*, 153(2):199–205, 2006.

[15] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005.

[16] Javier San Agustin, Henrik Skovsgaard, John Paulin Hansen, and Dan Witzner Hansen. Low-cost gaze interaction: ready to deliver the promises. In *CHI EA '09: 27th intl. conf. on Human factors in computing systems*, pages 4453–4458, New York, USA, 2009. ACM.

[17] Oliver Williams, Andrew Blake, and Roberto Cipolla. Sparse and semi-supervised visual mapping with the $s^3$p. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006), 17-22 June 2006, New York, NY, USA*, pages 230–237. IEEE Computer Society, 2006.

[18] Li-Qun Xu, Dave Machin, and Phil Sheppard. A novel approach to real-time non-intrusive gaze finding. In *Proceedings of the British Machine Vision Conference 1998, BMVC 1998, Southampton, UK, 1998*, pages 1–10, 1998.

[19] Yanxia Zhang, Andreas Bulling, and Hans Gellersen. Towards pervasive eye tracking using low-level image features. In *ETRA*, pages 261–264, 2012.

# Method of weak classifiers fuzzy boosting

Andrey V. Samorodov
Chair for Biomedical Technical Systems
Bauman Moscow State Technical University (BMSTU)
Moscow, Russia
avs@bmstu.ru

***Abstract*—Method of fuzzy boosting providing iterative weak classifiers selection and their quasi-linear composition construction is presented. The method is based on the combination of boosting and fuzzy integrating techniques, when at each step of boosting weak classifiers are combined by Choquet fuzzy integral. In the proposed FuzzyBoost algorithm 2-additive fuzzy measures were used, and method for their estimation was proposed. Although detailed theoretical verification of proposed algorithm is still absent, the experimental results, made on simulated data models, demonstrate that in the case of complex decision boundaries FuzzyBoost significantly outperforms AdaBoost.**

***Keywords—multiclassification, boosting, fuzzy integral, weak classifies, nonlinear class-separating surface, 2-additive fuzzy measures***

## I. INTRODUCTION

The classical pattern recognition task is to create and improve individual classification methods and algorithms. But beginning with the pioneering work of L.A. Rasstrigin, R.H. Ehrenstein and Y.L. Barabash [1, 2], and mainly in the last 15 years, the attention has been drawn from feature selection and construction of a better decision rule to multiclassification methods which find the best set of base classifiers and the method of combining their responses [3, 4, 5].

One of the challenges in this field is to create a multiclassification algorithm which can approximate nonlinear class-separating surfaces and, at the same time, which is relatively easy to learn and to use. In general, the nonlinearity of class-separating surfaces means that classes are incorrectly described and that their feature space is not optimal, resulting in noncompact classes with features that do not have the ergodic property. However, in some applications, such as object detection, decision boundaries, by definition, cannot be linear and even additive: compact class of objects of desired type is surrounded by a diffuse class of objects of all other types that could be met [6].

From the beginning of the 21st century boosting became one of the most popular techniques for multiclassification, which was used in numerous applications. J. Friedman et al. gave possible theoretical explanation and experimentally showed the potentials of AdaBoost. AdaBoost algorithms work well when the complexity of weak classifiers corresponds to the one of the decision boundary: for additive decision boundary stumps is the best choice, but for nonlinear decision boundary decision trees with more than two terminal nodes are considerably better than stumps. This comes from the nature of AdaBoost, which builds linear algorithmic composition of the responses of weak classifiers and could be regarded as an approximation method to the additive modelling on the logistic scale [7]. When the type of weak classifiers is not selected properly, the performance of boosting algorithms suffers greatly.

This paper presents the method and corresponding algorithm, which preserve the benefits of boosting technique and at the same time increase its performance in the case of nonlinear decision boundaries by the use of quasi-linear algorithmic composition of weak classifiers instead of linear one at each step of boosting.

## II. LINEAR AND QUASI-LINEAR MODELS FOR WEAK CLASSIFIERS AGGREGATION

For a two-class problem, an additive logistic model approximated by AdaBoost has the form

$$\log\frac{P(y=1\mid x)}{P(y=-1\mid x)}=F_M(x)=\sum_{m=1}^{M} f_m(x), \qquad (1)$$

where $f_m(x)$ is the response of $m$-th weak classifier, $M$ – the number of weak classifiers.

Different versions of AdaBoost algorithm can be interpreted as different stage-wise estimation procedures for fitting this model. The main difference between them is the way how $f_m$ are calculated.

Fuzzy integrating could be regarded as an extension of linear models [8]. The quasi-linear algorithmic composition of weak classifiers, based on Choquet fuzzy integral, has the form:

$$C_M^{\mu}(x)=\sum_{m=1}^{M} p_{\sigma(m)} f_{\sigma(m)}(x), \qquad (2)$$

where $\sigma$ is a permutation of weak classifiers' responses such that $f_{\sigma(1)}(x)\le\ldots\le f_{\sigma(M)}(x)$, $p_{\sigma(m)}$ are the coefficients determined as:

$$p_{\sigma(m)} = \mu\left(A_{\sigma(m)}\right) - \mu\left(A_{\sigma(m+1)}\right),$$

where $\mu$ is the fuzzy measure, $A_{\sigma(m)} = \{\sigma(m), \sigma(m+1), \ldots \sigma(M)\}$, $A_{\sigma(M+1)} = \varnothing$.

Sum of these coefficients $\sum_{m=1}^{M} p_{\sigma(m)} = 1$, so (2) could be considered as weighted sum of weak classifier responses.

Quasi-linearity of (2) lies in the dependence of coefficients $p_{\sigma(m)}$ on classifier responses $f_m(x)$, which is not the case for AdaBoost. The linear model (1) could be regarded as the special case of fuzzy integral (2) with equal values of $p_{\sigma(m)}$. If so,

$$F_M(x) = MC_M^{\mu}(x).$$

## III. FUZZYBOOST ALGORITHM DESCRIPTION

Choquet fuzzy integral represents much richer model than linear algorithmic composition. On the other hand, AdaBoost effectively selects weak classifiers, which could be hardly done using fuzzy integrating. Taking into account the complementary characteristics of AdaBoost and fuzzy integration algorithms, several approaches to their combining have been proposed in the literature. It was shown that weak classifiers, selected by AdaBoost, could be successfully aggregated by fuzzy integration [9]. Application of boosting method for learning fuzzy classifiers is considered in [10]. In [11] boosting sequential procedures were used for selection of approximate descriptive fuzzy rules.

In this paper, a new approach to the fusion of boosting and fuzzy integration methods initially proposed in [12] is developed. The method of fuzzy boosting implies the use of fuzzy integral for classification at each iteration step of boosting instead AdaBoost's own linear aggregation rule. General scheme of the two-class FuzzyBoost algorithm, implementing the method of fuzzy boosting, is shown in Fig. 1.

Each class could be characterized by different interactions between features. So for classification problem each class has to have its own fuzzy measure [13]. Moreover, the values $f_m(x)$ in (2) must be positive. That's why, in contrast to AdaBoost, the response of a weak classifier in the algorithm presented is calculated separately for each of the classes (Fig. 1, step 2.2). The resulting algorithmic composition is the difference of two fuzzy integrals, calculated separately for each class, using their own set of responses of weak classifiers and fuzzy measures (steps 2.5 and 2.6).

Given train data with $N$ examples $(x_i, y_i)$, $i = 1,\ldots,N$, with feature vector $x_i$ and class label $y_i \in \{-1;1\}$:

1. Start with weights $w_i = 1/N$, $i = 1,\ldots,N$.
2. Repeat for $m = 1,\ldots,M$
   - 2.1. Fit weak classifier $h_m(x) = \{h_m^{(1)},\ldots,h_m^{(j)},\ldots,h_m^{(L)}\}$ with $L$ terminal states $h_m^{(j)} \in \{-1;+1\}$, minimizing the weighted error probability.
   - 2.2. Set weak classifier responses $f_m^{(j)+}$, $f_m^{(j)-}$ for each terminal state in favor of each of the classes.
   - 2.3. For the train data examples estimate the initial data for the subsequent calculation of fuzzy measures in accordance with their type and the additive property.
   - 2.4. For the train data examples calculate the *m*-th weak classifier response for each of the classes. If an example activates *j*-th terminal state of weak classifier, then $f_m^+(x) = f_m^{(j)+}$, $f_m^-(x) = f_m^{(j)-}$.
   - 2.5. For the train data examples calculate the cumulative responses $C_m^{\mu+}(x)$ and $C_m^{\mu-}(x)$ of current weak classifiers ensemble for each of the classes with the use of fuzzy integrating.
   - 2.6. For the train data examples calculate the overall response of current weak classifiers ensemble $C_m^{\mu+,\mu-}(x) = C_m^{\mu+}(x) - C_m^{\mu-}(x)$.
   - 2.7. Update weights: $w_i = \exp\left(-y_i m C_m^{\mu+,\mu-}(x_i)\right)$.

Output the classifier $sign\left(C_M^{\mu+,\mu-}(x)\right)$.

Fig. 1. Two-class FuzzyBoost algorithm

Analogous to AdaBoost, weak classifier responses in FuzzyBoost could be calculated in different ways. In Gentle FuzzyBoost responses for *j*-th terminal state of a weak classifier are calculated as weighted posterior probabilities of classes:

$$f_m^{(j)\pm} = P_w(y = \pm 1 \mid j) = \frac{\sum_{i:y_i=\pm 1} w_i \cdot \left(h_m(x_i) = h_m^{(j)}\right)}{\sum_{i=1}^{N} w_i \cdot \left(h_m(x_i) = h_m^{(j)}\right)},$$

where the sign takes either '–' or '+' for the first or the second class support respectively.

In Real FuzzyBoost responses are calculated as logarithms of these probabilities:

$$f_m^{(j)\pm} = -\frac{1}{2}\lg P_w(y = \mp 1 \mid j).$$

Note, that the difference of the responses of a weak classifier for two classes in FuzzyBoost algorithm is the well-known response of weak classifier in AdaBoost.

The type and additive properties of the fuzzy measures determine the form of initial data, calculated on the step 2.3. In the previous work super-additive λ-measures were used in FuzzyBoost algorithm [12]. However, they do not really estimate interactions between weak classifiers. Moreover, with increasing number of weak classifiers in algorithmic composition the calculation of λ-measures becomes problematic. In [14] the concept of *k*-additive fuzzy measures was proposed and, as can be seen from a number of works in the field of fuzzy classification, 2-additive measures taking into account the pairwise weak classifiers interactions demonstrate the highest generalization ability in practice [8, 15, 16]. Therefore, the cumulative responses for each of the classes for the current set of weak classifiers could be effectively calculated using 2-additive fuzzy measures, resulting in the following form of Choquet fuzzy integral, proposed in [8]:

$$C_M^{\mu^{\pm}}(x) = \sum_{m=1}^{M} \phi_m^{\pm} f_m^{\pm}(x) - \frac{1}{2} \sum_{m \neq l} I_{ml}^{\pm} \left| f_m^{\pm}(x) - f_l^{\pm}(x) \right|, \qquad (3)$$

where $\phi_m^-$ and $\phi_m^+$ – Shapley indices for *m*-th weak classifier for the first and second class respectively, $I_{ml}^-$ and $I_{ml}^+$ – interaction indices for *m*-th and *l*-th weak classifiers for the same classes.

Shapley index characterizes the relative importance of weak classifier in the composition. Interaction index is the measure for joint behavior of a pair of weak classifiers; its positive or negative value reflects their positive or negative synergy.

In this study the Shapley indices were set to equal values, and interaction indices were determined as values proportional to the correlation coefficients of the responses of weak classifiers. At each iteration of FuzzyBoost algorithm the renormalization of Shapley and interaction indices is carried out to meet normalization (4) and monotonicity (5) conditions:

$$\sum_{m=1}^{M} \phi_m^{\pm} = 1, \qquad (4)$$

$$\sum_{m=1}^{M} \left| I_{ml}^{\pm} \right| \leq 2\phi_l, \forall l. \qquad (5)$$

Equation (5) is the compact interpretation of the formula, derived in [14].

Although the chosen way of interaction indices estimation is speculative, the fulfillment of the conditions (4) and (5) guarantee the existence of corresponding fuzzy measures, and thus the proper behavior of the Choquet integral (3) when the next weak classifier is added to the ensemble.

Note that FuzzyBoost algorithm take AdaBoost as the special case when 1-additive fuzzy measures (interaction indices equal to zero) with equal Shapley indices are used.

## IV. Experimental Results and Discussion

### A. Datasets description

The research of the FuzzyBoost algorithm performance was carried out on simulated two-class problems in 10-dimensional feature space. Simulated data represent two types of decision boundary: additive and nonlinear. The models used are analogues to ones proposed in [9] for the study of AdaBoost algorithms. In these models the ten input features are independent and are randomly drawn from ten-dimensional standard normal distribution. The decision boundary in the first test set is additive and is represented by ten-dimensional sphere. The rule for data labelling as first class ( $y_i = -1$ ) is

$$\sum_{j=1}^{10} x_j^2 < t,$$

where t is the threshold value.

Otherwise the observation is labelled as second class ( $y_i = +1$ ). The threshold value is chosen so that each class has approximately the same number of observations. In this study it equals to 9.2. For this data set it was 2000 training examples and 10000 test observations.

For the second data set, representing the nonlinear decision boundary model the observations are labelled as first class if

$$\sum_{j=1}^{6} x_j \left( 1 + \sum_{l=1}^{6} (-1)^l x_l \right) < 0. \qquad (6)$$

Otherwise the observation is labelled as second class. For this data set it was 5000 training examples and 10000 test observations.

Besides these two data models simulated ‘ring’ data model and well-known ‘banana’ dataset were also used. As previously, ‘ring’ dataset has 10-dimensional feature space with observations drawn in the same manner. The rule for data labelling as first class is

$$t_1 < \sum_{j=1}^{10} x_j^2 < t_2,$$

where threshold values were chosen so that each class has approximately the same number of observations and also that the number of observations inside the ‘ring’ is approximately the same as outside. The values for the first and the second thresholds were 6.75 and 12.56 respectively. For this data set it was 4000 training examples and 10000 test observations.

For each type of data training-test set combinations were three times independently drawn and corresponding average error was estimated.

'Banana' dataset contains 5300 examples in 2-dimentional feature space. For this dataset the error was estimated using 3-fold cross-validation technique.

### *B. Experimental results*

Comparative studies were carried out for the following classification algorithms: Gentle AdaBoost (GAB) and Real AdaBoost (their implementation was taken from GML AdaBoost Matlab Toolbox), Gentle FuzzyBoost (GFB) and Real FuzzyBoost, as well as *k*-nearest neighbors (*k*-NN) with *k*=3, naïve Bayes (NB) classifier, and decision trees (CART). The last three simple classification methods were used to feel the baseline in error rates.

Gentle and Real AdaBoost as well as Gentle and Real FuzzyBoost manifest comparable performance so only error rates for Gentle AdaBoost and Gentle FuzzyBoost are given here as functions of the number of iterations. Experimental results are presented in Fig. 2-8, where averaged error rates are plotted. Their 95 % confidence intervals in most cases did not exceed 8 % of corresponding mean values or were often less, with single exception for Fig. 3 where confidence intervals for FuzzyBoost error rate was near 13 % from 33$^{rd}$ to 61$^{st}$ iteration. Quantitative results are also presented in Table I.

The results indicate that in the case of additive decision boundary FuzzyBoost performance is equal to the one of AdaBoost with both stumps and decision trees with more than 2 terminal nodes used as weak classifiers (Fig. 2, Fig. 3). In the latter case (with 8 terminal nodes) FuzzyBoost seems to be slightly disturbed from 33$^{rd}$ to 61$^{st}$ iteration and near the 110$^{th}$ iteration, but all the time it returns to the error rates, demonstrated by AdaBoost.

The situation is dramatically different in the case of nonlinear decision boundary. Here FuzzyBoost provides a significant reduction in error rate as compared with AdaBoost when stumps are used as weak classifiers (Fig. 4). In the case of decision trees with 4 and 8 terminal nodes FuzzyBoost and AdaBoost demonstrate equal error rates, though in the latter case from 10$^{th}$ to 80$^{th}$ iteration FuzzzyBoost has slightly higher error rate than AdaBoost (Fig. 5, Fig. 6).

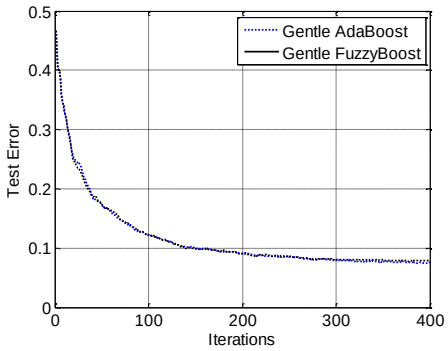



Fig. 2. The performance of classification algorithms (additive decision boundary, stumps)

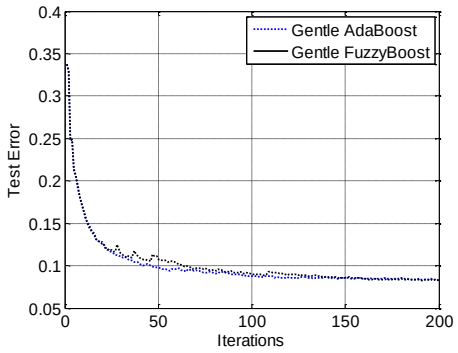



Fig. 3. The performance of classification algorithms (additive decision boundary, 8 terminal nodes decision trees)

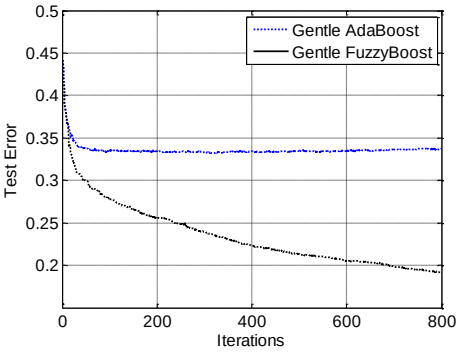



Fig. 4. The performance of classification algorithms (nonlinear decision boundary, stumps)

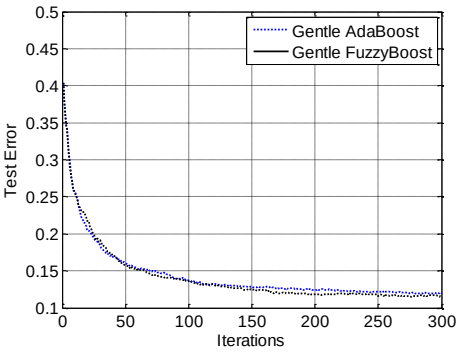



Fig. 5. The performance of classification algorithms (nonlinear decision boundary, 4 terminal nodes decision trees)

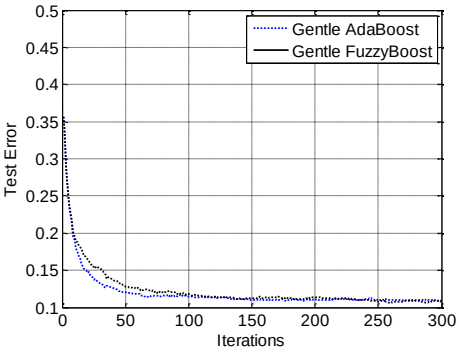



Fig. 6. The performance of classification algorithms (nonlinear decision boundary, 8 terminal nodes decision trees)

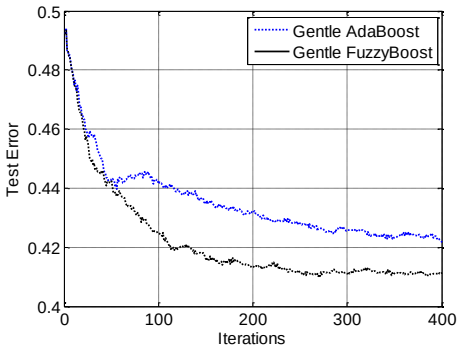



Fig. 7. The performance of classification algorithms ('ring' dataset, stumps)

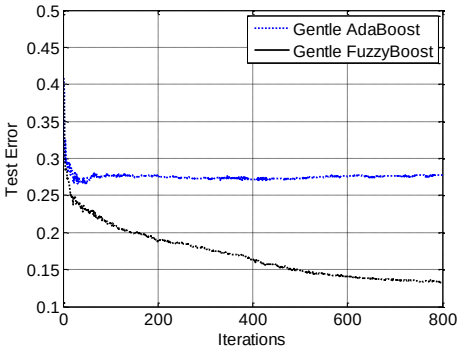



Fig. 8. The performance of classification algorithms ('banana' dataset, stumps)

For 'ring' dataset when using stumps as weak classifiers we see again the better performance of FuzzyBoost, though AdaBoost error rate seems to become closer to FuzzyBoost error rate the more iterations are made.

The results on 'banana' dataset demonstrate the example of AdaBoost overfitting (Fig. 8). Its error rate achieves the minimal value at approximately 40$^{th}$ iteration and then slightly grows. FuzzyBoost demonstrates another behavior; its error rate continues to decrease as in the case of simulated nonlinear decision boundary (see Fig. 4).

TABLE I. EXPERIMENTAL RESULTS

| Type of decision boundary | Number of terminal nodes in weak classifiers | Classification error rate | | | | |
|---|---|---|---|---|---|---|
| | | ***GAB*** | ***GFB*** | ***k-NN*** | ***NB*** | ***CART*** |
| Additive | 2 (stumps) (400$^{th}$ iteration) | 0.075 | 0.079 | 0.335 | 0.044 | 0.262 |
| | 8 (200$^{th}$ iteration) | 0.083 | 0.083 | | | |
| Nonlinear | 2 (stumps) (800$^{th}$ iteration) | 0.337 | 0.192 | 0.228 | 0.320 | 0.291 |
| | 4 (300$^{th}$ iteration) | 0.120 | 0.115 | | | |
| | 8 (300$^{th}$ iteration) | 0.110 | 0.109 | | | |
| 'Ring' | 2 (stumps) (400$^{th}$ iteration) | 0.422 | 0.411 | 0.436 | 0.494 | 0.435 |
| 'Banana' | 2 (stumps) (800$^{th}$ iteration) | 0.277 | 0.132 | 0.119 | 0.386 | 0.130 |

As it can be seen from the Table I, for the dataset with additive decision boundary and for 'banana' dataset FuzzyBoost has error rate which is close to the one of the best 'baseline' algorithms (NB and *k*-NN correspondingly). For the dataset with nonlinear decision boundary and for 'ring' dataset FuzzyBoost outperforms 'baseline' algorithms.

### C. Discussion

Experimental results show that the main advantage of FuzzyBoost over AdaBoost is that it could work with weak classifiers that do not fit the complexity of decision boundary. But if they do FuzzyBoost has no real advantages over AdaBoost and the latter is worth using because of less calculation complexity.

The tests performed indicate, that proposed FuzzyBoost algorithm is stable enough, and could less suffer from overfiting than AdaBoost. 'Banana' dataset is an example of noisy data when AdaBoost slightly manifest overfiting phenomenon. The observed behavior of FuzzyBoost algorithm (see Fig.8) is promising, but its resistance to overfiting should be further investigated.

The intuition for understanding why FuzzyBoost works better than AdaBoost with stumps as weak classifiers in the case of nonlinear decision boundary is that the data model (6) is the second order model with components representing pairwise products of the features. This could exactly fit the aggregation model, produced by the Choquet integral with respect to 2-additive fuzzy measures when pairwise interactions of weak classifiers are considered.

In real world applications nearly all decision boundaries, even if they are nonlinear should be smooth and thus could be locally represented as 2-order surfaces. So FuzzyBoost seems to fit the majority of the classification problems in real applications. But this of course should be proved by future researches.

## V. CONCLUSION

The study presents a new multiclassification method of fuzzy boosting, which is a combination of AdaBoost and fuzzy integration methods. The difference between AdaBoost and FuzzyBoost algorithms is in the way how weak classifier responses are aggregated at each steps of boosting. In FuzzyBoost the Choquet fuzzy integral with respect to 2-additive fuzzy measures is used, providing quasi-linear aggregation model. Results of experimental studies have shown that the proposed FuzzyBoost algorithm has better generalization ability than AdaBoost in the case of classification problems with nonlinear decision boundaries. First of all, that is because of taking into account the dependencies between the weak classifiers and of the construction of not the linear, as in AdaBoost algorithms, but a quasi-linear composition of weak classifiers at each iteration of boosting.

## REFERENCES

[1] L.A. Rasstrigin and R.H. Ehrenstein, Method of collective recognition. Moscow: Energoizdat, 1981, 80 p. (in Russian).

[2] Y.L. Barabash, Collective statistical decisions in recognition. Moscow: Radio i Svyaz, 1983, 224 p. (in Russian).

[3] T.K. Ho, "Multiple classifier combination: lessons and the next steps," Hybrid methods in pattern recognition, World Scientific Publishing, pp. 171-198, 2002.

[4] L.I. Kuncheva, Combining pattern classifiers: methods and algorithms. John Wiley & Sons, Inc., 2004, 350 p.

[5] J. Ghosh, "Multiclassifier systems: back to the future," LNCS 2364, pp. 1-15, June 2002.

[6] Handbook of face recognition, S.Z. Li and A.K. Jain, Eds. Springer Science+Business Media, Inc., 2005, 395 p.

[7] J. Friedman, T. Hastie, and R.Tibshirani, "Additive logistic regression: a statistical view of boosting," The Annals of Statistics, vol. 38, no. 2, pp. 337-374, 2000.

[8] M. Grabisch and Ch. Labreuche, "A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid," A Quarterly Journal of Operations Research, vol. 6, issue 1, pp. 1-44, March 2008.

[9] L.I. Kuncheva, ""Fuzzy" versus "nonfuzzy" in combining classifiers designed by boosting," IEEE Transactions on Fuzzy Systems, vol. 11, no. 6, pp. 729-741, December 2003.

[10] L. Junco and L. Sanchez, "Using the Adaboost algorithm to induce fuzzy rules in classification problems," Proc. ESTYLF, Sevilla, pp. 297-301, 2000.

[11] M.J. Del Jesus, F. Hoffmann, J.L. Navascues, and L. Sanchez, "Induction of fuzzy rule based classifiers with evolutionary boosting algorithms," IEEE Transactions on Fuzzy Sets and Systems, vol. 12, no. 3, pp. 296-308, June 2004.

[12] A.V. Samorodov, "Application of a fuzzy integral for weak classifiers boosting," Pattern Recognition and Image Analysis, vol. 21, no. 2, pp. 206-210, June 2011.

[13] M. Grabisch and J.-M. Nicolas, "Classification by fuzzy integral: performance and tests," Fuzzy Sets and Systems, vol. 65, pp. 255-271, August 1994.

[14] M. Grabish, "k-order additive discrete fuzzy measures and their representation," Fuzzy Sets and Systems, vol. 92, pp. 167-189, December 1997.

[15] L. Mikenina and H.-J. Zimmermann, "Improved feature selection and classification by the 2-additive fuzzy measure," Fuzzy Sets and Systems, vol. 107, pp. 197-218, October 1999.

[16] M. Grabish and C. Labreuche, "The Choquet integral for 2-additive bi-capacities," Proc. of 3-rd Int. Conf. of the European Soc. for Fuzzy Logic and Technology (EUSFLAT 2003), P. 300-303, 2003.

Proceedings of the 9th Open German-Russian Workshop on Pattern Recognition and Image Understanding
Dec 1-5, 2014, Koblenz, Germany

# Detection of Object Interactions in Video Sequences

Ali Al-Raziqi, Mahesh Venkata Krishna and Joachim Denzler
Computer Vision Group
Friedrich Schiller University of Jena
Ernst-Abbe-Platz 2 07743 Jena, Germany
Email: {ali.al-raziqi, mahesh.vk, joachim.denzler}@uni-jena.de

***Abstract*—In this paper, we propose a novel framework for unsupervised detection of object interactions in video sequences based on dynamic features. The goal of our system is to process videos in an unsupervised manner using Hierarchical Bayesian Topic Models, specifically the *Hierarchical Dirichlet Processes* (HDP). We investigate how low-level features such as optical flow combined with Hierarchical Dirichlet Process (HDP) can help to recognize meaningful interactions between objects in the scene, for example, in videos of animal interaction recordings, kicking ball, standing, moving around etc. The underlying hypothesis that we validate is that interactions in such scenarios are heavily characterized by their 2D spatio-temporal features. Various experiments have been performed on the challenging JAR-AIBO dataset and first promising results are reported.**

## I. Introduction

Application fields such as video-based surveillance systems, animal monitoring systems etc., often require us to distinguish the interactions between objects or the interactions between objects and their surroundings. Figure 1 shows an example scenario where various objects in a scene are interacting with each other. The meaningful interactions in a scene are characterized by the spatio-temporal dynamics of the objects within the scene.

Detecting interactions between objects in scenes is a challenging problem in computer vision. The challenge is compounded by various aspects such as occlusions, variations in objects sizes, illumination variations, noisy recordings etc. It is important that any system tackling the problem is robust with respect to such factors.

Further, in many of these application scenarios, the interactions are not well-known beforehand, and preparation of a well-labeled data-set covering all possible interactions for the purpose of training a machine learning algorithm may not be possible. For example, in the scenario where we observe interactions between animals, all the interactions the animals might be involved in can not be determined beforehand, and sometimes, even the exact number of possible interactions is impossible to predict. In such situations, use of unsupervised methods becomes imperative.

For unsupervised scenarios, as the kind of interactions are not known beforehand, interactions are defined as co-occurring actions from multiple actors or actors performing actions using some inanimate objects in the scene.

In the literature, *Hierarchical Dirichlet Processes* (HDP) and their derivatives have been used for unsupervised activity perception and analysis [1]–[3]. While they have been demonstrated for activity perception and detection for crowded scenes or individual actors, it is not clear whether HDP can be extended to analyze specific interaction between actors, or between actors and objects, in a scene. Further, determining the correct representation schemes for the current task remains a challenge.

According to our knowledge, most of the current object interactions modeling systems rely on supervised learning methods and some features such as histogram of oriented gradients (HOG), scale-invariant feature transform (SIFT), shape/appearance feature matching etc. [4]–[12]. These frameworks typically start with the localization of an object in the frames and then determining the relevant action. Some of these works done on learning the interactions applied on static images [8], [9], [13], [14]. However, object segmentation and localization are often error prone steps, leading to performance deterioration. They suffer from problems such as camouflage, noisy recording process, occlusions, or poor visibility.

Another interesting line of approaches are based on recognizing objects, actions and human poses [6], [13], and then detecting/recognizing interactions from static images of single object without using feature matching and motion analysis.

Also, in [11], the authors used network graphs framework to analyze the interaction between parts of an object. The body parts and objects are represented as nodes of social network graphs, the parts are tracked to extract the temporal features and the social network analysis features provide the spatial features. They then, used SVM and a Hidden Markov Model to classify the interactions of the object's parts. However, an approach free from object localization requirement and using features that better characterize the interactions in the scene is called for. As a solution, some methods focus on background subtraction [4], [8].

In contrast, to tackle the task of interaction detection in an unsupervised manner and without object localization / pose estimation, we combine the HDP model presented in [15], and low level features such as optical flow using [16]. Since, to the best of our knowledge, no such work has been done in the past, we evaluate the advantages and drawbacks of our HDP-based algorithm on the challenging JAR-AIBO dataset [17] and present the results.

## II. Optical Flow and HDP

Due to their wide applicability, clustering techniques are applied commonly in many areas of computer vision. Unlike supervised classification methods, in clustering, class labels are not supplied. There are two categories of clustering algorithms: partitioning and hierarchical. Most of the partitioning based

(a) Coming together

(b) Playing with the ball

Fig. 1: Examples of interactions between objects. In (a), in successive frames, the dogs are coming together from the corners of the marked area. (b) shows the four dogs playing with the ball in the middle. Images are from the JAR-AIBO dataset [17].

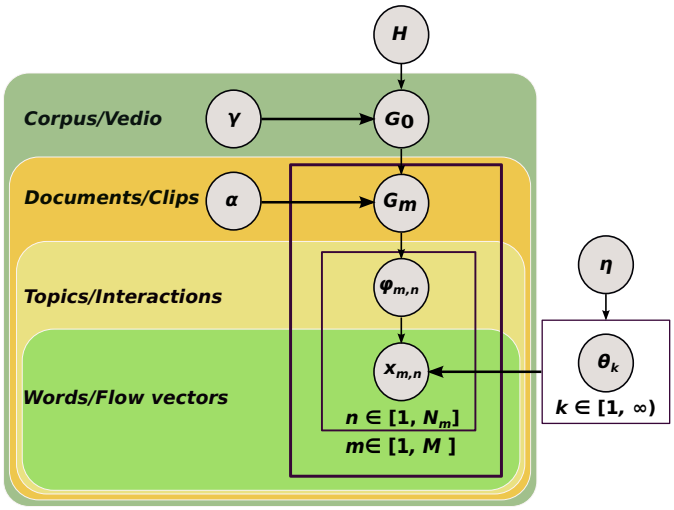



Fig. 2: HDP Model.

clustering techniques such as k-means and Latent Dirichlet Analysis (LDA), require a set of parameters, such as the number of clusters to be provided, which limits their applicability in many situations where such information is not available. In HDP, the number of clusters is deduced automatically from the data and hyper-parameters. As will be formally shown later (*cf.* 3), the number of resulting clusters in HDP can be controlled by the hyper-parameters $\alpha$, $\gamma$ and $\eta$. The hyper-parameters, especially $\eta$, determine the number of extracted clusters, in our case interactions.

HDP has been originally designed for clustering words in documents based on word co-occurrences. Figure 2 shows the basic HDP model. Suppose we are given an input data corpus, which is divided into $M$ documents and each document consists of a set of words $x_{m,n}$, where $n \in [1, N_m]$. The goal of the HDP model is to cluster these words into meaningful latent structures, or *topics*.

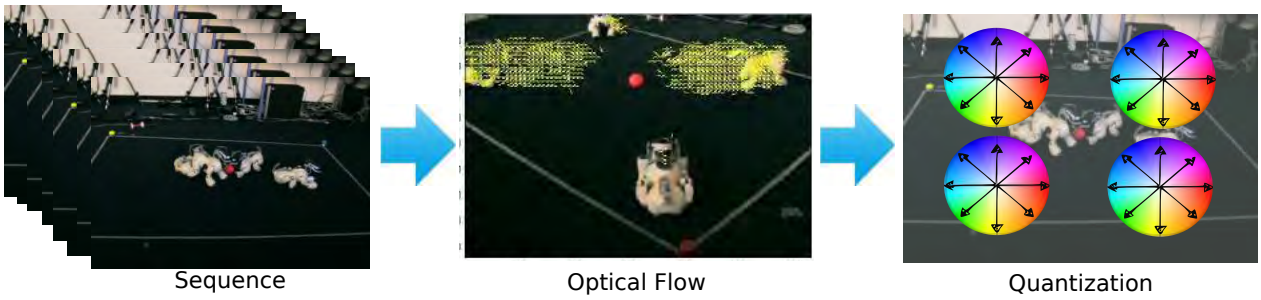
In our case, given an input video, optical flow features are extracted from each pair of successive frames using TV-L$^1$ algorithm [16]. The resulting optical flow is threshlolded to remove noise such as changing illumination or camera motion, and only significant motion is used for feature extraction. Subsequently, the optical flow vectors are quantized into eight directions. The optical flow features can be defined as X=($x$, $y$, $u$, $v$), where ($x$, $y$) is the location of a particular pixel in the image, and ($u$, $v$) are the flow values which represent the vector of optical flow. Based on the flow values, the magnitude and direction of the optical flow can be represented as $P = \sqrt{u^2 + v^2}$ and $\theta = \tan^{-1}\left(\frac{v}{u}\right)$ respectively. Figure 3 illustrates the complete procedure.

Then a dictionary or codebook is built with all possible flow words (flow words are four-tuples, x-y co-ordinates and associated flow values). The video is divided into small equally sized clips (e.g. 10 sec) without overlapping, and each clip is represented by a bag-of-words based on the dictionary. In our framework, clips and optical flow words correspond to documents and words, respectively.

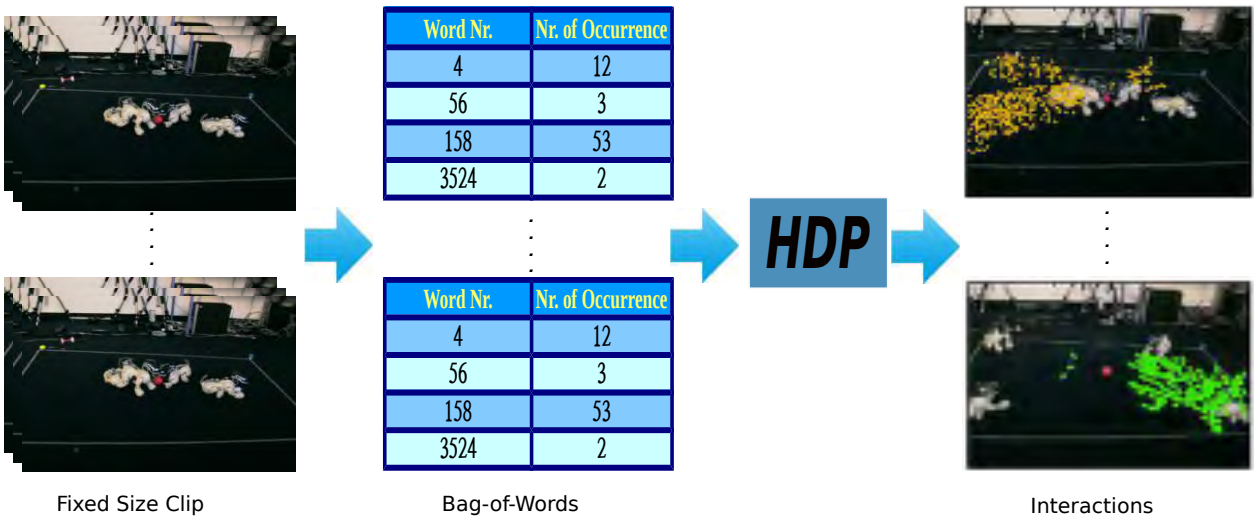
The HDP model generates the global list of interactions using a top level Dirichlet Process (DP) $G_0$. Then, the clip-specific interactions $G_m$ are drawn from the global list $G_0$ for each clip. Formally, we write the generative HDP formulation as shown in 1:

$$\begin{aligned} G_0 \mid \gamma, H &\sim DP(\gamma, H) \\ G_m \mid \alpha, G_0 &\sim DP(\alpha, G_0) \quad \text{for } m \in [0, M] \end{aligned} \tag{1}$$

where the hyper-parameters $\alpha$ and $\gamma$ are called the concentration parameters and the parameter $H$ is called the base distribution (Dirichlet distribution). Therefore, the observed words $x_{m,n}$ are seen as being sampled from the mixture priors $\phi_{m,n}$, which in turn are seen as being drawn from a Dirichlet Process $G_0$ The values of mixture components drawn from $\theta_k$




(a) Representation




(b) Clustering

Fig. 3: Illustration of the process of extracting optical flow features and arranging them according to a bags-of-words representation scheme.

Thus, the formulation of this construction can be written as,

$$\begin{aligned} \theta_k &\sim P(\eta) \quad \text{for } k \in [1, \infty) \\ \phi_{m,n} \mid \alpha, G_m &\sim G_m \quad \text{for } m \in [i, M], n \in [1, N_m] \\ x_{m,n} \mid \phi_{m,n}, \theta_k &\sim F(\theta_{\phi_{m,n}}) \end{aligned} \tag{2}$$

where $M$ is the number of clips in sequences, $N_m$ is the number of words in clip $m$, $P(\cdot)$ and $F(\cdot)$ are the prior distribution over topics and the prior word distribution given the topic respectively.

In our problem, we perform the Bayesian inference, where given the observed words, we *infer* the latent interactions. As a closed form solution for the inference process is not available for our case, we use the *Markov Chain Monte Carlo* (MCMC) approximation, specifically Gibbs sampling, using the Chinese Restaurant Franchise-based formulation. Following the formulation of [2], the conditional probability of the topic-word association for each iteration step evaluates to:

$$p(\phi_{m,n} = k, \alpha, \gamma, \eta, \theta, H) \propto (n_{m,k}^{\neg m,n} + \alpha\theta_k) . \frac{n_{k,t}^{\neg m,n} + \eta}{n_k^{\neg m,n} + V.\eta} \tag{3}$$

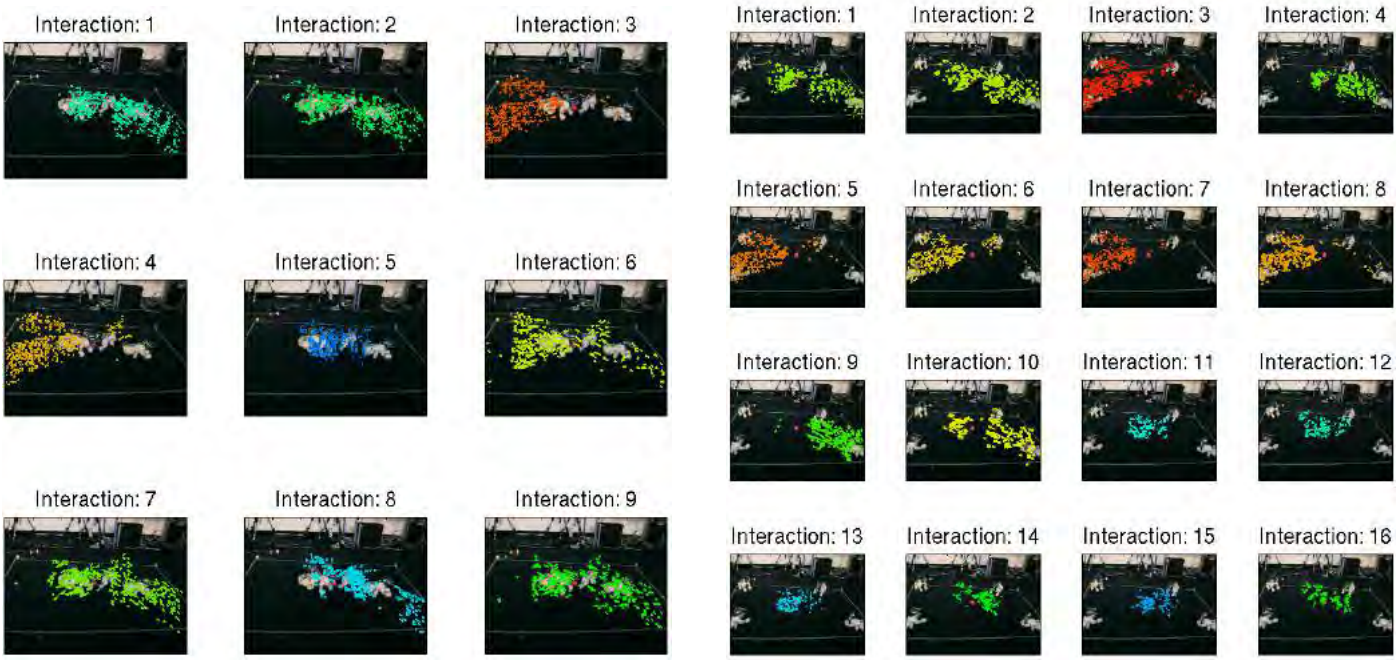
where $n_{m,k}$; $n_{k,t}$; and $n_k$ represent count statistics of the word-topic, topic-document and the topic-wise word counts, respectively.

The superscript $\neg m, n$ means that the current word $x_{m,n}$ must be eliminated from these statistics. $V$ is the size of the dictionary. The first part of the equation 3 reveals that the probability of assigning the current word to a topic is proportional to the number of words already assigned to that topic. This forms the basis of the clustering property of the HDP model. The second part (the probability of creating a new topic) shows that the hyper-parameters $\alpha, \gamma$ and especially $\eta$ can be used to determine the number of extracted topics We also perform hyper-parameter sampling to make our framework completely data-driven. For further details on the sampling procedure, we refer to [15].

## III. Experiments

### A. Data-set

We use the challenging JAR-AIBO dataset [17] to evaluate our system (*cf.* Fig. 1). JAR-AIBO dataset enables us to test our system in the face of many issues such as changing illumination, changing object view and occlusions. It contains 5 sequences taken of four SonyAIBO robot dogs performing




(a) Hyper-Parameters $\alpha = 0.5$, $\eta = 0.5$

(b) Hyper-Parameters $\alpha = 1.5$, $\eta = 1.5$

Fig. 4: Some qualitative results with various extracted interactions, coded by different colors. Note the impact of varying the value of the HDP's hyper-parameters $\alpha$ and $\eta$ on the number of extracted interactions.

actions autonomously, which are captured by six cameras at 17 fps with a resolution of 640 x 480 pixel. The camera feeds are synchronized to a frame level. In our experiments, we utilized the sequences of two cameras.

In all, we have approximately 15 interactions involving four dogs. Interactions include the dogs "converging" from all corners of the frame to the center, "playing with the ball", one or more dogs "leaving the group", one dog "walking around" the others, one dog "kicking the ball" as other dogs walk around, etc. Figure 1 shows some example frames of "coming together" and "playing with the ball" interactions. In all these, more than one dog is involved and the challenge is to detect these interactions without any prior knowledge about them.

### B. Experimental setup

The optical flow extraction is performed as follows. As we mentioned above, optical flow is computed using [16]. Each frame is divided into grid cells of size 8 x 8 pixels, and quantized into eight directions. Hence, the size of the dictionary is 80 x 60 x 8. In our experiments, in order to study the effects of clip lengths on performance, the video is divided into clips of various sizes ranging from 100 to 400 frames each – corresponding to approximately 5 to 23 seconds in the videos – and constructed bags-of-words representations for them.

Though the HDP model provides possibility of assigning multiple topics per word based on its context. In this paper, we also study the effect of changing the hyper-parameters $\alpha$, $\eta$ where their values ranging from 0.1 to 1.5. Further, as it gets re-sampled depending on the data and the initial value does not significantly affect performance, we initialize $\gamma$=1 in all experiments.

For quantitative performance evaluation, we use the true positive rate (TPR) and the false positive rate (FPR), defined as follows:

$$\text{TPR} = \frac{\text{TP}}{\text{TP}+\text{FN}}; \qquad \text{FPR} = \frac{\text{FP}}{\text{FP}+\text{TN}} \tag{4}$$

where TP, FP, FN, and TN stand for True Positives, False Positives, False Negatives, and True Negatives respectively.

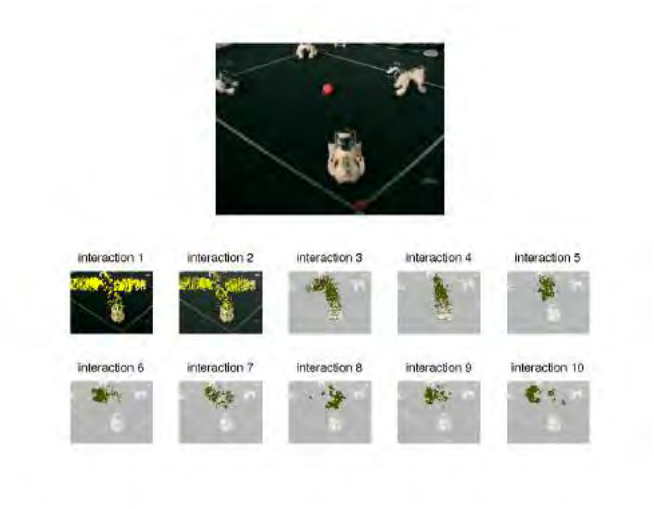
As the data-set does not contain ground truth in terms of object interactions, the video sequences were marked with clip-wise annotations regarding the interactions contained within them[1]. Then, following the procedure similar to [1], [3], the output of our system is manually mapped to the ground truth labels and the performance measures are calculated.

### C. Results and Discussion

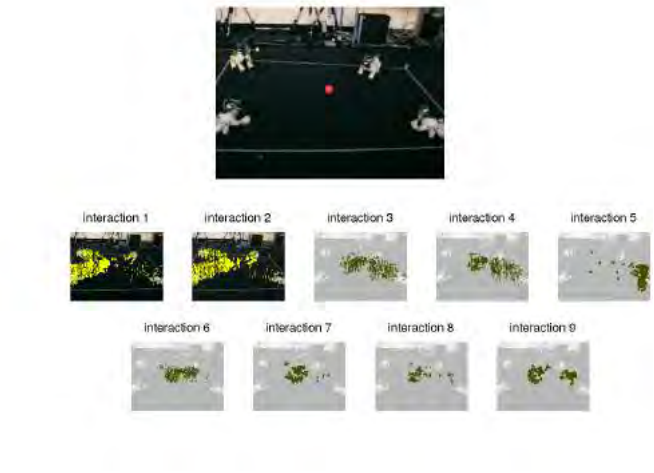
We can see some quantitative results in Fig. 4, for a video containing four dogs, where the dogs start from different corners of the frame, converge at the center, play with the ball, and finally one dog leaves the group to the bottom right corner of the frame. In Fig. 4a, interactions 1-4 and 7 represent the "converging" interaction, interactions 5 and 6 represent "playing with the ball" interaction, and interactions 8 and 9 represent the "dog leaving the group" interaction. Similar parallels can be seen in Fig. 4b.

The impact of varying the values of the HDP's hyper-parameters on the number of extracted interactions can be

[1]The ground truth will be made available as a part of the data-set




(a) View1




(b) View2

Fig. 5: Interactions extracted for multiple views. Note that despite the change of views, the interactions are still detected meaningfully.

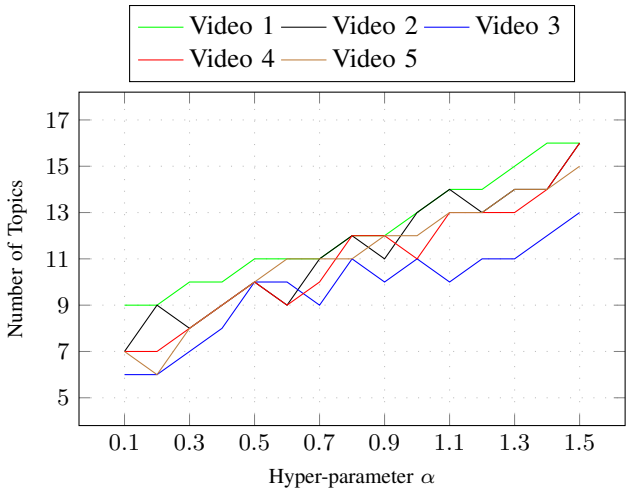



Fig. 6: Number of interactions extracted for each of the five videos as a function of the hyper-parameter $\alpha$. $\eta$ was held constant at 0.5.

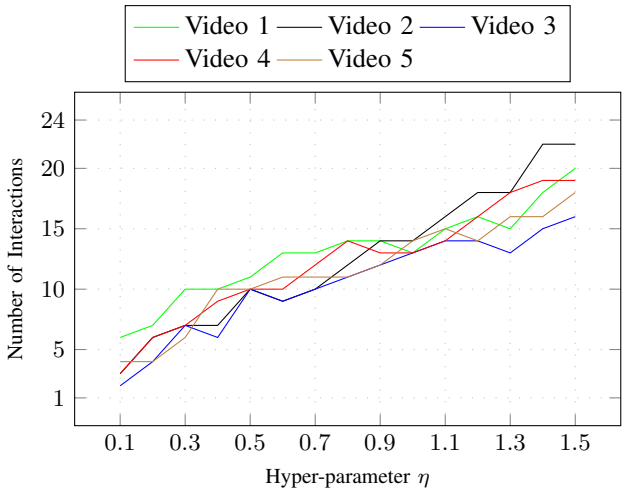



Fig. 7: Number of interactions extracted for each of the five videos as a function of the hyper-parameter $\eta$. $\alpha$ was held constant at 0.5.

clearly seen. The number of topics grows with increasing hyper-parameters values. Figures 4(a),(b) show that high values of hyper-parameters in situations with smaller number of interactions result in the creation of duplicate interactions. For example, in Fig. 4(a) interaction 4 is a duplicate of interaction 3, with only a few noisy flow vectors being the difference. In Fig. 4(b), this is more pronounced, where interaction 3, for example, is repeated four more times in interactions 5 to 8. Sometimes, due to high hyper-parameter values, a single interaction, such as interaction 5 in Fig. 4(a), is split into multiple smaller interactions, such as interactions 11 to 16 in Fig 4(b). This increase in the number of inferred interactions follows from the HDP inference process, where higher values of hyper-parameters imply a higher probability of drawing new interactions, and the presence of noisy features compounds the effect.

Quantitatively, Fig. 6 and 7 show the variations in number of interactions extracted as a function of the two hyper-parameters $\alpha$ and $\eta$ respectively. Clearly, the number of interactions extracted increases with the hyper-parameter values. However, it is interesting to note that the range of the number of interactions is larger in the case of hyper-parameter $\eta$. This is due to the fact that, being the parameter controlling the probability of generation of new interaction directly, it has larger effect on the resulting number of interactions. Therefore, a user can provide prior knowledge about the number of interactions through setting the hyper-parameters accordingly.

Figure 5 shows the extracted interactions for two different views. It can be clearly observed that despite a change in viewpoint, the extracted interactions are stable.

Table I shows the quantitative evaluation of our experiments. As can be observed, View 1 with frame size 400 has

TABLE I: Results of the HDP algorithms for two views. The effects of clip-sizes on the performance can be clearly observed.

| **View** | **View 1** | | | **View 2** | | |
|---|---|---|---|---|---|---|
| **Clip Size (Frames)** | 100 | 250 | 400 | 100 | 250 | 400 |
| **True Positive Rate%** | 77.14 | 78.60 | **82.35** | 77.14 | 78.57 | 76.47 |
| **False Positive Rate%** | 32.95 | 41.70 | 32.00 | 52.13 | 51.16 | **31.81** |

achieved the high value of TPR 82.35 % also lowest value of FPR 31.81 %, whereas the lower frames per clip values result in worse performance. This is likely due to the fact that, smaller clip sizes split the interactions into many sub-interactions, and consequently, performance suffers.

## IV. Conclusions and Future Work

The aim of this paper was to show how low-level optical flow features combined with a Hierarchical Dirichlet Process can be used to extract meaningful interactions in video sequences in an unsupervised manner. We compared the effect of several values of HDP's hyper-parameters, and the qualitative results obtained from the various experiments performed on the challenging JAR-AIBO dataset were promising.

Future research topic will be a comparison of different features combined with Hierarchical Dirichlet Processes and other similar topic models. Furthermore, in order to reduce testing time during deployment, we can use a step-wise combination of generative and discriminative methods, following the approach of [3]. Use of other clustering schemes, such as DP-means of [18] also seems interesting.

## Acknowledgments

The authors thank Golenur Khanam for her contribution in compiling the results presented in this work. The work was partially funded by Deutscher Akademischer Austauschdienst (DAAD) and ZEISS.

## References

[1] X. Wang, X. Ma, and W. Grimson, "Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 539–555, March 2009.

[2] M. Venkata Krishna, M. Körner, and J. Denzler, "Hierarchical dirichlet processes for unsupervised online multi-view action perception using temporal self-similarity features," in *Seventh International Conference on Distributed Smart Cameras (ICDSC)*, 2013, pp. 1–6.

[3] M. V. Krishna and J. Denzler, "A combination of generative and discriminative models for fast unsupervised activity recognition from traffic scene videos," in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2014, pp. 640–645.

[4] A. Patron-Perez, M. Marszalek, A. Zisserman, and I. Reid, "High five: Recognising human interactions in tv shows," 2010.

[5] B. Yao and L. Fei-Fei, "Grouplet: A structured image representation for recognizing human and object interactions," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 9–16.

[6] A. Gupta, A. Kembhavi, and L. S. Davis, "Observing human-object interactions: Using spatial and functional compatibility for recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 10, pp. 1775–1789, 2009.

[7] B. Yao, X. Jiang, A. Khosla, A. L. Lin, L. Guibas, and L. Fei-Fei, "Human action recognition by learning bases of action attributes and parts," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1331–1338.

[8] C. Desai, D. Ramanan, and C. Fowlkes, "Discriminative models for static human-object interactions," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010, pp. 9–16.

[9] A. Prest, C. Schmid, and V. Ferrari, "Weakly supervised learning of interactions between humans and objects," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 3, pp. 601–614, 2012.

[10] M. S. Ryoo and J. K. Aggarwal, "Spatio-temporal relationship match: Video structure comparison for recognition of complex human activities," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 1593–1600.

[11] G. Yang, Y. Yin, and H. Man, "Human object interactions recognition based on social network analysis," in *Applied Imagery Pattern Recognition Workshop: Sensing for Control and Augmentation, 2013 IEEE (AIPR*. IEEE, 2013, pp. 1–4.

[12] B. Yao and L. Fei-Fei, "Modeling mutual context of object and human pose in human-object interaction activities," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 17–24.

[13] B. Yao and L. FeiFei, "Recognizing human-object interactions in still images by modeling the mutual context of objects and human poses," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 9, pp. 1691–1703, 2012.

[14] S. Park and J. Aggarwal, "Semantic-level understanding of human actions and interactions using event hierarchy," in *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on*. IEEE, 2004, pp. 12–12.

[15] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical dirichlet processes," *Journal of the american statistical association*, vol. 101, no. 476, 2006.

[16] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime tv-l 1 optical flow," in *Pattern Recognition*. Springer, 2007, pp. 214–223.

[17] M. Körner and J. Denzler, "Jar-aibo: A multi-view dataset for evaluation of model-free action recognition systems," in *New Trends in Image Analysis and Processing–ICIAP 2013*. Springer, 2013, pp. 527–535.

[18] B. Kulis and M. I. Jordan, "Revisiting k-means: New algorithms via bayesian nonparametrics." in *International Conference on Machien Learning (ICML)*, 2012.

# New Approach to the Analysis of Geoacoustic Emission Signals

Alexander B. Tristanov, Yuriy V. Marapulets, Olga O. Lukovenkova, Alina A. Kim
Laboratory of acoustic research
Institute of Cosmophysical Research and Radio Wave Propagation of the Far Eastern Branch of Russian Academy of Science (IKIR FEB RAS)
Kamchatka, Russia
alextristanov@mail.ru

***Abstract* — The paper is devoted to the new approach of analysis of geoacoustic emission signals based on sparse approximation. A modified algorithm of matching pursuit is presented.**

***Keywords—sparse approximation; matching pursuit; geoacoustic emission; time-frequency analysis; geophysical signals***

This paper is devoted to the development of a new approach of analysis of geoacoustic emission signals. Acoustic emission in solid bodies is elastic oscillations, which are the result of dislocation changes in environment. The generated pulse radiation characteristics are directly associated with the features of elastic process. This fact determines the interest to research the geoacoustic emission for the development of methods for acoustic diagnostics of enviromnemt. The infra- and ultra-sonic frequency ranges are the most frequently studied.

The first one is aimed at the study of the seismic process; the second one is for the material strength.

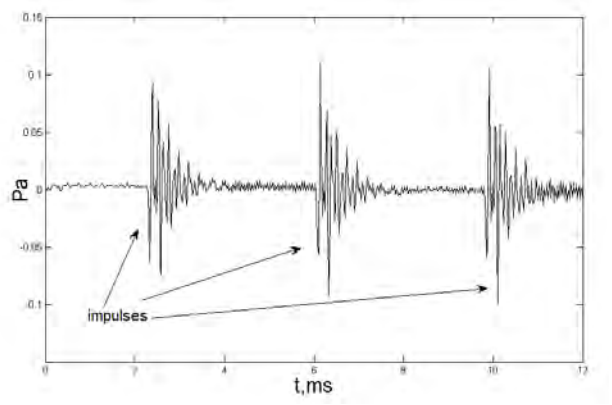



Fig. 1. An example of geoacoustic emission signal.

The researches carried out in Kamchatka showed that acoustic methods are effective to make diagnostics of natural environments on the scales corresponding to the sound oscillation wavelengths. The relation between the intensification of deformation processes and the behavior of acoustic emission, including earthquake preparation, was detected. It was determined, that emission anomalies in kilohertz frequency range are registered at the distance of the first hundreds of kilometers 1-3 days before strong earthquakes [1].

Applying sparse approximation, the authors suggested a method for attributive description of separate pulses of the emission process [2-5]. The main advantages of the suggested method are shown, which give more informative representation of pulse signal fine structure, compared to the classical spectral and time-frequency analysis. Time series sparse approximation is one of the most dynamically developing and perspective methods of shot time signal representation. Sparse approximation with redundant dictionary assumes the construction of a linear model with a few nonzero elements. These elements are chosen from the function family forming a redundant dictionary.

The idea of replacing of a complex signal by a few more simple components is widely used in many areas: compression and analysis of audio-, video- data, images, investigation of seismic data, medical signals, etc [6-8]. Sparse approximation is not a new concept and has been investigated for more than a hundred of years. Temlyakov V.N. and Tropp J.A. note that the so-called m-term approximation was first mentioned by Schmidt E. in the paper «Zur Theorie der linearen und nichtlinearen Integralgleichungen», published in 1908 [9-10]. During the recent ten years, the researches have been systematized and generalized, for example, in the papers by Mallat S., Sturm B.L., Zhang Z., Gribonval R., Donoho D.L., Elad M., and others [11-16].

The authors developed an algorithm which allows one to get compact signal decomposition with high precision when computational resources are limited [2-5]. The algorithm is based on the matching pursuit (MP) method with the procedure of refinement of the dictionary at each iteration and methods of parallel programming by CUDA technology.

The most time-consuming part of MP algorithm is the calculation of scalar product of dictionary atoms with a signal for each iteration. For the dictionary composed of M atoms with $L_{atom}$ counting length, and a signal with $L_{sig}$ counting

length, calculation of all scalar products will require $M \times (2L_{atom} - 1) \times L_{sig}$ addition and multiplication, the volume of calculation resources is proportional to M dictionary size. The authors applied two mechanisms to reduce the calculation efforts for MP algorithm. First, a modified algorithm of matching pursuit with refinement (Fig. 2) was developed. The essence of the modified MP algorithm with refinement is the search for a new, more significant decomposition element for each iteration of the algorithm in the neighborhood of the selected atom at the given iteration. The determined and refined atom and all its shifts are added into the dictionary, adapting it to signal specific peculiarities. The prevailing part of the time to perform one iteration of an adaptive algorithm with refinement on the dictionary with $N_p$-dimensional parameter space is summed up from the time of $(M + k \cdot 3^{N_p}) \times (2L_{atom} - 1) \times L_{sig}$ addition and multiplication with a fixed time of refining dictionary formation $3^{N_p}$ from atoms, multiplied by the number of iteration of k education, thus the calculated resource volume, required for the matching pursuit with refinement, proportionally depend on M dictionary size and the number of iterations of k education. For a dictionary of a definite size M, it is possible to find k value, which will result in high accuracy for the given calculation resources.

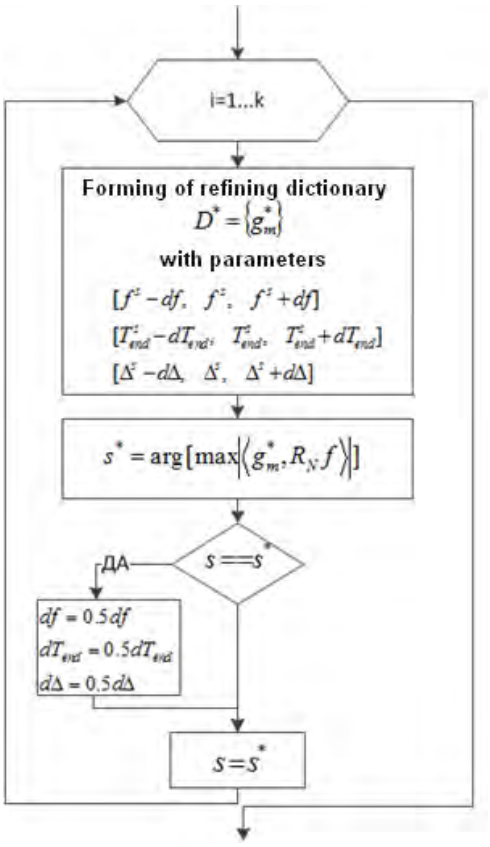



Fig. 2. The modificaded part of MP algorithm

One of the main problems of sparse approximation is the choice of the base function family (dictionary) and the criteria defining sufficient sparseness. The authors have researched dictionaries of different basic functions and suggested a combined dictionary which includes both functions having similarity to the morphological elements of the pulses under the study (Berlage's functions) and functions having the smallest square of the time-frequency window (Gauss' modulated functions).

It is shown that the application of the combined dictionary provides a more precise accuracy of signal approximation in comparison to mono-dictionary.

The cutoff of the approximation error was suggested as a criterion for the break-point decomposition. Multistep aposterior process of the decomposition analysis (composition and frequency-time atom distribution) allowed the authors to filter pulse structure, to separate the proper pulse of the acoustic emission from secondary pulse impact. The final analysis represents the inner structure of pulse in the framework of the chosen basic function dictionary.

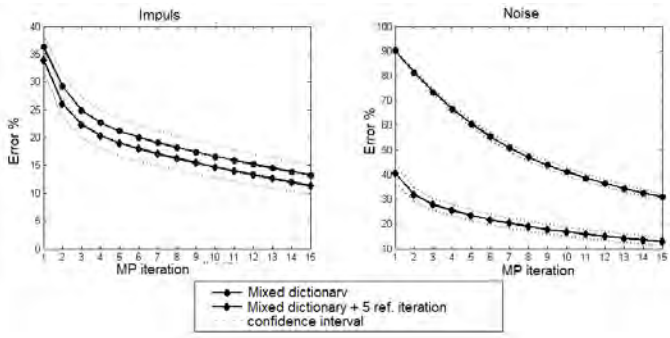



Fig. 3. Dicrease of the approximation error of the matching pursuit method with refinement and without it

Classical matching pursuit modification applying mixed dictionaries and involving the refinement in parameter space significantly increases the quality of acoustic emission signal approximation. In Fig. 3 the decrease of the approximation errors is shown applying the developed and classical algorithms. The suggested algorithm can be efficiently and effectively used in the processing and analyzing systems of geoacoustic emission signals.

In the framework of this paper the approach to the statistic analysis of decomposition and the pulse classification method are developed. It is shown that atom parameter distribution represents a mixture of distributions that indicates the presence of several classes in pulse selection under the investigation. The classification algorithm is based on symbol representation of feature space elements. Symbol approximation was investigated by E. Keogh [17]. Symbol approximation suggests the changing of the initial signal by the sequence of symbols. Each symbol has a corresponding local signal behavior. This behavior can be assigned differently and can correspond to different local models. In our case the local model is described by the atom filling frequency. The symbol sequence describes

the dynamics of time-frequency signal structure thus specifying a separate pulse class.

## REFERENCES

[1] Marapulets Yu.V., Tristanov A.B. Using The Sparse Approximation Method For The Problems Of Geoacoustic Emission Analysis //Digital Signal Processing. 2011, №2, P.13-17 (in Russian)

[2] Marapulets Yu.V., Shevtsov B.M. Mesoscale Acoustic Emission. – Vladivostok: Dalnauka, 2012 (in Russian)

[3] Afanaseva A.A. Lucovenkova O.O., Marapulets Yu.V., Tristanov A.B. Using sparse approximation and clastering for the acoustic emission time series description // Digital Signal Processing. 2013, №2, P.30-34 (in Russian)

[4] Marapulets Yu.V., Tristanov A.B. Shevtsov B.M. The structure of acoustic emission signals analysis with the methods of sparse approximation//Acoustic journal, 2014, Vol 60, #4 p. 398-406(in Russian)

[5] Tristanov A.B., Lucovenkova O.O. The Adaptive refining matching pursuit algorithm for combined dictionaries in analysis of the geoacoustic emission signals // Digital Signal Processing. 2014, №2, P.54-57 (in Russian)

[6] Sturm B.L. Similarity in Sparse Domains with Applications to Audio Signals. 2010. P. 1–36.

[7] Chakraborty A., Okaya D. Frequency-time decomposition of seismic data using wavelet-based methods. 1995. Vol. 60, № 6. P. 1906–1916.

[8] Li Y., Cichocki A., Amari S.-I. Blind estimation of channel parameters and source components for EEG signals: a sparse factorization approach. // IEEE Trans. Neural Netw. 2006. Vol. 17, № 2. P. 419–431.

[9] Temlyakov V.N. Nonlinear methods of approximation // Found. Comput. Math. 2003. Vol. 3. P. 33–107.

[10] Schmidt E. Zur Theorie der linearen und nichtlinearen Integralgleichungen. III. Teil // Math. Ann. 1908. Vol. 65. P. 370–399.

[11] Mallat S. A Wavelet Tour of Signal Processing: The Sparse Way. 2009 // Acad. Burlingt. 2009.

[12] Mallat S., Zhang Z. Matching Pursuit with time-frequency dictionaries // Semin. Oncol. Nurs. 1993. Vol. 28, № 3.

[13] Gribonval R. et al. Analysis of sound signals with high resolution matching pursuit TFTS ’ 96 : Analysis of sound signals with High Resolution Matching Pursuit. P. 1–4.

[14] Davis G., Mallat S., Zhang Z. Adaptive Time-Frequency Decompositions with Matching Pursuits 1 Introduction. P. 1–15.

[15] Donoho D.L., Elad M. Optimally sparse representation in general (nonorthogonal) dictionaries via l minimization. // Proc. Natl. Acad. Sci. U. S. A. National Academy of Sciences, 2003. Vol. 100, № 5. P. 2197–2202.

[16] Elad M. Sparse and redundant representations: from theory to applications in signal and image processing. 2010.

[17] Lin, J., Keogh E., Lonardi S., Chiu B. A symbolic representation of time series, with implications for streaming algorithms // DMKD '03 Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery, 2003 P.2-11

# On Coordination of Contour Descriptions for the Equivalence Class with a Group of Affine Transformations*

Leonid Ivanovich Lebedev
Research Institute of Applied Mathematics and Cybernetics
Lobachevsky Nizhni Novgorod State University
Nizhni Novgorod, Russia
lebedev@pmk.unn.ru

Yury Grigorievich Vasin
Research Institute of Applied Mathematics and Cybernetics
Lobachevsky Nizhni Novgorod State University
Nizhni Novgorod, Russia
pmk@unn.ac.ru

***Abstract*— In this paper, two theorems are presented, based on which we propose a method to coordinate contour descriptions for the equivalence class with a group of affine transformations. The method's high performance is ensured by eliminating the need to calculate similarity estimates. The essence of the method is explained by using an example of coordination of contour descriptions for trapezoids.**

***Keywords—affine transformation; equivalence class; similarity estimate; contour description; matrix; coordination of descriptions***

## I. INTRODUCTION

In [1-3], some methods for coordination of contour descriptions are described. It is shown that the method of relative displacements and the parabola method offer the best performance in coordination. A prerequisite for the use of these methods is the requirement for calculating similarity estimates of the object, with the reference. In this case, different initial points for the description of one of the contours must be specified. In this paper, we propose a method that does not require the calculation of similarity estimates.

## II. STATEMENT OF THE PROBLEM

Let us introduce the concept of an equivalence class $\mathbf{I(C^o, G)}$. We assign to this class a set of objects generated by the contour definition $\mathbf{C^o}$ and by the group of affine transformations $\mathbf{G = (A, \Delta P)}$, where $\mathbf{A}$ is the transformation matrix, and $\mathbf{\Delta P}$ is the displacement. Thus, if the description of the contour $\mathbf{C^o}$ is given by a sequence of vertices $\mathbf{P_i^o}$, $\mathbf{i = 1, 2, ..., n}$, namely, $\mathbf{C^o = \{P_1^o, P_2^o, ..., P_n^o\}}$, then any contour $\mathbf{C = \{P_1, P_2, ..., P_n\}}$ belongs to the class of equivalence $\mathbf{I(C^o, G)}$, if all the vertices of its description satisfy the condition $\mathbf{P_i = A \cdot P_k^o + \Delta P}$, where $\mathbf{k = (i + q)\%(n + 1) + (i + q)/(n + 1)}$, and $\mathbf{q}$ is the parameter that provides cyclicity in the setting of the starting point. It follows that the task of coordinating the descriptions of the two contours $\mathbf{C^o}$ and $\mathbf{C}$ belonging to the same equivalence class $\mathbf{I(C^o, G)}$ consists in finding the value of the parameter $\mathbf{q}$.

The work was supported by RFBR, project No.13-07-00521.

## III. METHODS OF SOLUTION

The solution of the problem of coordinating the descriptions of the two contours $\mathbf{C^o}$ and $\mathbf{C}$ will be based on finding the matrix of affine transformation $\mathbf{A}$. To do this, we take an arbitrary vertex $\mathbf{P_i^o}$ on the contour $\mathbf{C^o}$, find two vectors $\mathbf{b_1^i = \overrightarrow{P_i^o P_{i-1}^o}}$, $\mathbf{b_2^i = \overrightarrow{P_i^o P_{i+1}^o}}$ and form the matrix $\mathbf{B_i = (b_1^i, b_2^i)}$. Similarly, we determine two vectors $\mathbf{v_1^j = \overrightarrow{P_j P_{j-1}}}$, $\mathbf{v_2^j = \overrightarrow{P_j P_{j+1}}}$ for the vertex $\mathbf{P_j}$ of the contour $\mathbf{C}$ and define the matrix $\mathbf{V_j = (v_1^j, v_2^j)}$.

From the vectors $\mathbf{(b_1^i, b_2^i)}$ and $\mathbf{(v_1^j, v_2^j)}$, we find the matrix of the affine transformation $\mathbf{A_j^i}$, which establishes the relationship between the vertices $\mathbf{P_{i-1}^o, P_i^o, P_{i+1}^o}$ and $\mathbf{P_{j-1}, P_j, P_{j+1}}$ of the polygons $\mathbf{C^o}$ and $\mathbf{C}$ respectively. It is obvious that $\mathbf{V_j = A_j^i \cdot B_i}$. Since the contour $\mathbf{C^o}$ is a polygon without self-intersections, and $\mathbf{P_i^o}$ is its vertex, then the vectors $\mathbf{b_1^i}$, $\mathbf{b_2^i}$ are not collinear, and hence, $\det \mathbf{B^i} \neq 0$. From this we can find the components of the matrix of the affine transformation $\mathbf{A_j^i = V_j \cdot B_i^{-1}}$. Denote $\mathbf{A^i = \{A_1^i, A_2^i, ..., A_n^i\}}$.

**Theorem 1.** Among the elements of the set $\mathbf{A^i}$, there exists at least one matrix for which the equality $\mathbf{A = A_k^i}$ holds.

One can give the answer to the question which of the matrices $\mathbf{A_j^i}$, $\mathbf{j = 1, 2, ..., n}$ allows us to coordinate the descriptions, based of the calculation of similarity estimates, the number of which in the worst case will not exceed $\mathbf{n}$. However, it is possible to offer a more computationally efficient method for finding the components of the matrix $\mathbf{A}$.

**Theorem 2.** There is at least one value of $\mathbf{k}$, such that for a pair of matrices of the sets $\mathbf{A^i}$ and $\mathbf{A^{i+1}}$ the equality $\mathbf{A_k^i = A_{k+1}^{i+1}}$ holds for all $\mathbf{i}$, from which follows that $\mathbf{A = A_k^i}$.

On the basis of these theorems, for any pair of objects of the same equivalence class one can obtain not only the values of the components of the affine transformation matrix $\mathbf{A}$, but also the numbers of vertices to coordinate the descriptions using the values of $\mathbf{n}, \mathbf{k}$ and $\mathbf{i}$. For the final formulation of the algorithm for coordination of descriptions, we introduce the concept of $m-$ way symmetry of the contour with respect to the starting point of the description. The contour will be assumed $m-$ way symmetrical with respect to the starting point of the description, if in addition to the initial description there are more than $(m-1)$ descriptions, which do not require coordination. The example of a two-way symmetrical contour is a rectangle, and of a five-way symmetrical contour, a five-pointed star.

**Statement 1.** For contours that do not possess the property of symmetry with respect to the starting point of the description, there is only one value of $\mathbf{k}$ for which $\mathbf{A_k^i = A_{k+1}^{i+1}}$ for all values of $\mathbf{i}$.

**Statement 2.** For $m-$ way symmetrical contours, there are no more than $m$ different values of $\mathbf{k}$ for which $\mathbf{A_k^i = A_{k+1}^{i+1}}$ for all values of $\mathbf{i}$.

Thus, the proposed method for coordination of descriptions is reduced to the determination of the diagonal with the same elements in the matrix $\mathbf{S} = \left(\mathbf{A_j^i}\right)$ with the dimensions $\mathbf{n \times n}$. Below, one of the possible algorithmic implementations of this method is shown.

## IV. Algorithm for Coordination of Descriptions

The algorithm for coordination of descriptions can be represented as the following step by step procedure.

Step 1. We set $\mathbf{k = i = 1}$, $=0$. Matrices $\mathbf{V_j}$, $\mathbf{B_j^{-1}}$ are computed for all $\mathbf{j = 1, 2, ..., n}$.

Step 2. We find matrices $\mathbf{A_k^i}$, $\mathbf{A_{k+1}^{i+1}}$, where $\mathbf{A_k^i = V_k \cdot B_i^{-1}}$, $\mathbf{A_{k+1}^{i+1} = V_{k+1} \cdot B_{i+1}^{-1}}$.

Step 3. If $\mathbf{A_k^i = A_{k+1}^{i+1}}$, we proceed with Step 4, otherwise, if $=1$ we restore the parameters by setting $\mathbf{k = k_o}$, $\mathbf{i = 1}$, $=0$. If $=0$, $\mathbf{k}$ is increased by unity. If $\mathbf{k > n}$, we proceed with Step 5, otherwise, with Step 2.

Step 4. If $=0$, the parameter $\mathbf{k}$ is fixed $\mathbf{k_o = k}$. We set $=1$. If $=1$, the parameters $\mathbf{k}$ and $\mathbf{i}$ are increased by unity. If $\mathbf{i > n}$ we proceed with Step 5, otherwise, with Step 2.

Step 5. If $=0$, the algorithm terminates without coordinating the descriptions. When $=1$, from the value of $\mathbf{k_o}$ we find $\mathbf{q}$ for coordinating the descriptions as well as the affine transformation matrix $\mathbf{A}$.

## V. Results

Let us estimate the computational complexity of the proposed method. It is obvious that to obtain matrices $\mathbf{V_j}$, $\mathbf{B_j}$, $\mathbf{B_j^{-1}}$, $\mathbf{j = 1, 2, ..., n}$ it is necessary to perform $2f_{\pm} \cdot \mathbf{n}$, $2f_{\pm} \cdot \mathbf{n}$, $(6f_{\times} + f_{\pm}) \cdot \mathbf{n}$ operations, respectively, where symbols $f_{\pm}$ and $f_{\times}$ denote the operations of addition (or subtraction) and multiplication (division). To obtain any matrix $\mathbf{A_k^i}$, it is necessary to perform $(8f_{\times} + 4f_{\pm})$ operations and, therefore, to find the parameter $\mathbf{k_o}$ and to subsequently test the equality of the diagonal elements of the matrix $\mathbf{S}$, this number should be multiplied by $(\mathbf{2 \cdot k_o + n - 2})$. As a result, the computational complexity of the method for coordinating descriptions is determined from the expression:

$$C = (14f_{\times} + 9f_{\pm}) \cdot \mathbf{n} + (16f_{\times} + 8f_{\pm}) \cdot (\mathbf{k_o - 1}).$$

Let us demonstrate the algorithm by using the example of coordination of descriptions of trapezoids. The original description of the trapezoid in the plane XOY is given by the coordinates:

$$\mathbf{C^o} = \{(10,15), (30,60), (70,60), (80,15)\}.$$

The affine transformation $\mathbf{G = (A, \Delta P)}$ required to obtain the contour $\mathbf{C}$ is determined by the matrix $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix}$ and the displacement along the axes $\mathbf{\Delta P} = (-25, 100)^T$. The initial point of the description $\mathbf{C}$ is shifted along the contour by the value $\mathbf{q} = 2$. The obtained descriptions of trapezoids are shown in Fig.1.

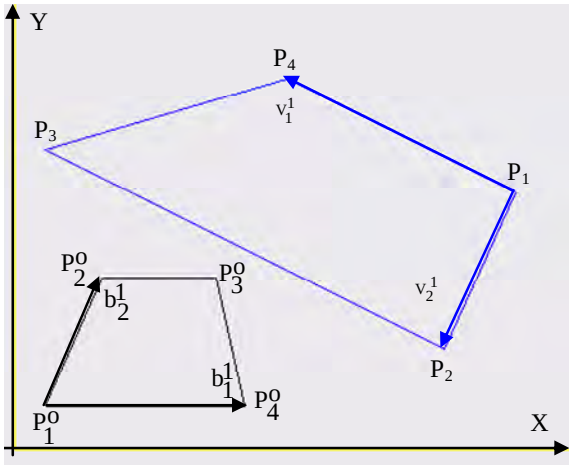



Fig. 1. Trapezoids of the equivalence class $\mathbf{I\ (C^o, G)}$

Table 1 illustrates the application of the proposed algorithm for coordinating the descriptions of trapezoids. It follows from the analysis of the table that the equality $\mathbf{A_k^1 = A_{k+1}^2}$ will initially be satisfied at $\mathbf{k = 3}$ which is taken as the value of $\mathbf{k_o}$. Subsequent checks revealed the identity of the diagonal elements of the matrix $\mathbf{S}$ for the obtained value of $\mathbf{k_o}$.

Consequently, $\mathbf{q} = \mathbf{k_o} - \mathbf{1} = \mathbf{2}$, and $\mathbf{A} = \mathbf{A_3^1}$. It follows from the logic of the algorithm that for solving the problem of coordination of contour descriptions, the calculation of a certain part of matrices $\mathbf{S} = \left(\mathbf{A_j^i}\right)$ is optional.

TABLE I. ELEMENTS OF MATRIX **S**

| $\mathbf{A_k^i}$ | k=1 | k=2 | k=3 | k=4 |
|---|---|---|---|---|
| i=1 | $\begin{pmatrix} -1.1 & -0.5 \\ 0.57 & -1.5 \end{pmatrix}$ | $\begin{pmatrix} 0.36 & -3.3 \\ 0.79 & 1.21 \end{pmatrix}$ | $\begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix}$ | $\begin{pmatrix} -1.2 & 2.32 \\ -0.4 & -0.7 \end{pmatrix}$ |
| i=2 | $\begin{pmatrix} -0.6 & 2.06 \\ -1.4 & -0.3 \end{pmatrix}$ | $\begin{pmatrix} -3.5 & 1 \\ 1.75 & -2 \end{pmatrix}$ | $\begin{pmatrix} 2.13 & -4.1 \\ 0.63 & 1.28 \end{pmatrix}$ | $\begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix}$ |
| i=3 | $\begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix}$ | $\begin{pmatrix} -0.6 & 2.97 \\ -1.4 & -1.9 \end{pmatrix}$ | $\begin{pmatrix} -3.5 & -2.7 \\ 1.75 & -0.2 \end{pmatrix}$ | $\begin{pmatrix} 2.13 & -1.3 \\ 0.63 & 1.03 \end{pmatrix}$ |
| i=4 | $\begin{pmatrix} 0.36 & -1.7 \\ 0.79 & 1.06 \end{pmatrix}$ | $\begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix}$ | $\begin{pmatrix} -1.2 & 2.84 \\ -0.4 & -1.6 \end{pmatrix}$ | $\begin{pmatrix} -1.1 & -2.1 \\ 0.57 & -0.4 \end{pmatrix}$ |

The cells in the table corresponding to these matrices have a darker background. For the initial conditions given in the example, their number will be half of the elements of the matrix **S** (the average over all **q** will be equal to 9).

## VI. CONCLUSION

The proposed algorithm for coordinating contour descriptions is undoubtedly the most efficient one in terms of speed among those listed in [1-3]. However, the possibility of its application in the presence of contour distortions requires further investigation.

## REFERENCES

[1] Vasin Yu.G.,Lebedev L.I. The problem of obtaining coherent contour descriptions in the calculation of similarity estimates. //8th Open German-Russian Workshop “Pattern recognition and image understanding” (OGRW-8-2011): Workshop Proceedings. 2011. Pp. 324-327.

[2] Lebedev L.I. Optimization of the computational complexity of correlation-extreme contour recognition methods. //Intellectualization of information processing: 9th International conference. Montenegro, Budva, 2012 / Proceedings. / M .: TORUS PRESS, 2012. Pp. 472-475.

[3] L.I. Lebedev The method of target location of coordinated descriptions in the recognition of image objects //Mathematical methods for pattern recognition (MMPR-16-2013): The 16th All-Russian Conference with international participation: Proceedings. / M .: "MAKS Press", 2013. P. 87.

# On the False Rejection Ratio of Face Recognition Based on Automatic Detected Feature Points

Kazuo OHZEKI † Masahiro TAKATSUKA †
Masaaki KAJIHARA † Yutaka HIRAKAWA †
Graduate School of Engineering and Science
Shibaura-Institute of Technology
3-7-5 Toyosu, Koutou-ku, Tokyo, 135-8548 Japan
{ohzeki@sic, ma14067@, ma14034@, hirakawa@}.shibaura-it.ac.jp

Kiyotsugu SATO

Dept. of Information Processing Engineering
College of Industrial Technology
1-27-1 Nishikoya, Amagasaki, Hyogo, 661-0047 Japan
kiyo@cit.sangitan.ac.jp

***Abstract*—The authors propose a new face recognition system with an evaluation function using feature points. The feature points are detected automatically by Milborrow's Stasm software. Before recognition, rotation compensation and size normalization are applied to the feature points. The main method is to calculate the squared error between the registered face and the input face as to length of a characteristic pair of feature points on face. The False Rejection Rate (FRR) for the registered and input face of the same person, and the False Acceptance Rate (FAR) for the registered face and a different person's input face are evaluated. The input is a video sequence. Stable recognition is obtained with small FRR and FAR for the video of a period of 0.5 second.**

***Keywords—Face Recognition; Feature Points; Normalization; Rotation Compensation; Individual Characteristics***

## I. INTRODUCTION

There are two major methods for face recognition [1] one is the feature-based method which uses feature points at endpoints of facial parts. The other is the holistic method which processes the whole face region without decomposing regions by feature points. The former method has become unpopular because it is difficult to detect accurate feature points automatically. The latter method is now popular. However, Kathryn Bonnen and Anil K. Jain newly propose the effectiveness of a component-based method which uses facial parts in detail, rather than globally recognizing face information [2]. The holistic method conventionally analyzed the whole face region at once for robust recognition. The component-based method utilizes separate regions of the eye-brow, eyes, nose and mouth, and performs dedicated recognition for each separated region, then integrates the results. The component-based method implies that it is possible now to detect facial parts before face recognition. The performance of the component-based method is better than that of a single holistic face recognition method.

Recently, detection technology of feature points of facial parts has improved greatly. Luxand detected 66 points automatically and at high speed [3]. Milborrow detected 77 points accurately and automatically [4][5]. Comparative experiments in face recognition were performed and reported in ref [2]. The recognition rate by the holistic method was 63.78% with FAR=1%, while the recognition rate by the component-based method was 76.88% with the same FAR=1%.

Based on this background, this paper aims to investigate the discriminating ability of face recognition using values of feature points incorporating facial feature points detected by software program of Milborrow. Once the feature points are detected accurately, the proposed method can accurately perform face recognition. Conventional methods such as ref [2] evaluated recognition rate usually with a non-zero value of FAR. This paper partially introduces experiments with FAR=0.0 and will lead to face recognition with FAR=0.

## II. FACE RECOGNITION SYSTEM

Face recognition in this paper will be done using feature points. If we can detect personal differences in facial feature points, face recognition can be performed because we can get much combination of feature point values of facial parts. One of the ideal situations for an input scene is a frontal face without expression. However, a usual situation involves variation from this frontal face and may include expressions. Thus, in our system, in the first input stage, rotation compensation and normalization as to the size are applied to reduce input variation.

Figure 1 depicts the processing system after input. For N-frame input images, the first image is registered for recognition. The other images after the first are tested by a recognition algorithm. After input, facial feature points are detected using Milborrow's software program called "Stasm" [6].

Stasm incorporates the Active Shape Model and stably detects 77 feature points on a face whose direction is nearly frontal as shown in Figure 2.

After detection of the feature points of the facial parts, rotation compensation and normalization as to the size are processed.

### *A. Rotation Compensation*

Let us take two points one is the right endpoint of the right eye and the other is the left endpoint of the left eye as shown in Fig. 2. Let the gradient between the right endpoint of the right eye and the left endpoint of the left eye as shown in Fig. 2. The gradient made of these two points is θ as shown in Fig.3. The rotation compensation is to transform the feature point coordinate values to make the θ become zero. The matrix for the rotation transform is shown in formula (1), applying all

feature points (x,y) to obtain new coordinate values (X,Y), placing both eyes on a horizontal line.

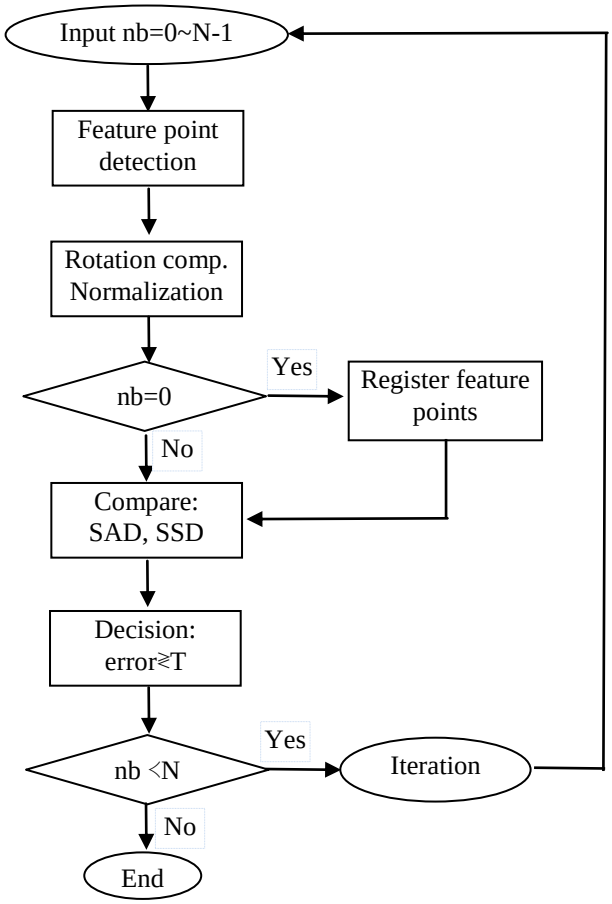



Fig. 1 Face Recognition Using Facial Feature Points.

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (1)$$

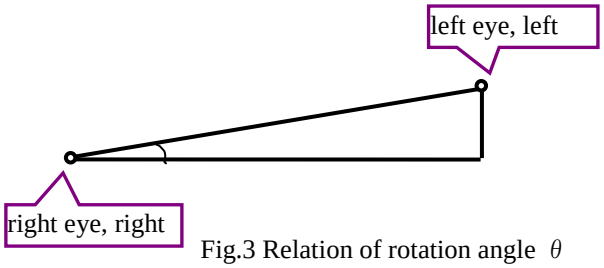



Fig.3 Relation of rotation angle $\theta$

### B. *Normalization as to the size*

Next, normalization as to the size is processed. The size of the face is different from person to person. The distance from camera to face is usually indefinite. To partially remove the indefinite factors of sizes, normalization as to one of the distances between two selected feature points is introduced. By this normalization, the difference in the face sizes between person to person diminishes. The distance between the two selected feature points is normalized to a fixed value "sc". Under this normalization condition, the relative difference from person to person is evaluated.

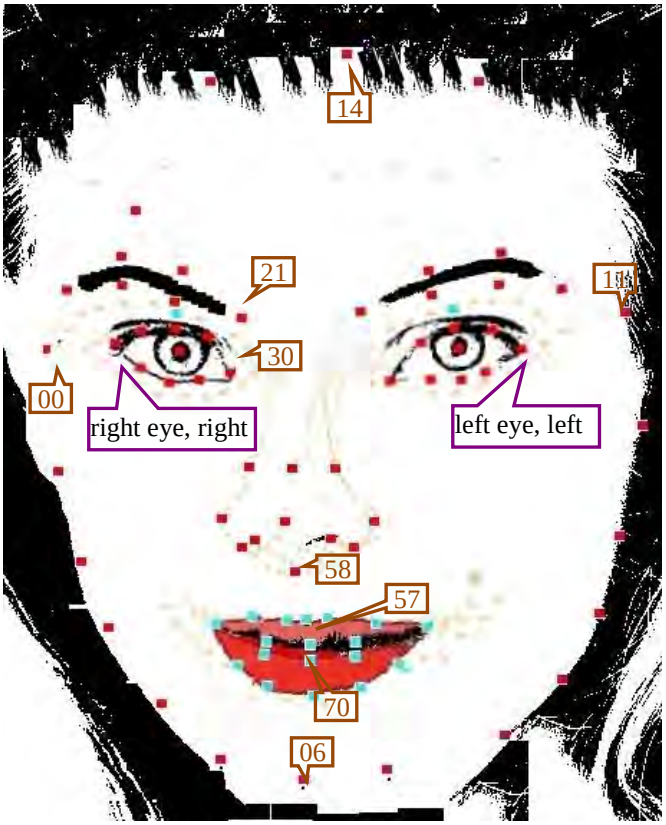



Fig.2 77 feature points on a face, indicated by black dots, detected by Stasm. (This image is licensed by Datacraft Corp.) The numbers are referred to in later sections.

To be more precise, let

NF=sc/(left_X of left eye – right_X of right eye),

and the normalization is to multiply this NF to all (X,Y) coordinate values.

Another normalization for position is finally applied to move all data in parallel. Actually selecting one point as an anchor point among facial feature points, all other data are moved in parallel. For example, if we select the first feature point (X,Y) coordinate values, whose registered data are (Reg(0),Reg(1)) and whose input data are (In(0),In (1)), then displacement amounts for moving in parallel are Reg(0)-In(0) for X component and Reg(1)-In(1) for Y component.

### C. *Recognition judgment*

Verification is done by comparing the registered facial feature points and the newly input data. As for evaluation, the sum of the absolute difference (SAD) between the two data values and another sum of square difference (SSD) are calculated. The sums are compared with respective thresholds. Let the registered (X,Y) coordinate values be (Reg(2i),Reg(2i+1)) and input (X,Y) be (In(2i),In(2i+1)), then,

$$SAD = \sum_{i=0}^{76} |Reg(2i) - In(2i)| + |Reg(2i+1) - In(2i+1)| \quad (2)$$

$$SSD = \sum_{i=0}^{76} (Reg(2i) - In(2i))^2 + (Reg(2i+1) - In(2i+1))^2 \quad (3)$$

## III. EXPERIMENTS

### A. *Basic experiments*

Computer simulation was carried out according to the construction above. Input sequences are recorded for scenes involving reading aloud manuscripts. Each sequence is composed of 2000 frames of images, which are about 67 seconds as an interval. The first frame is registered and other frames are processed for recognition to evaluate the False Rejection ratio (FRR). The False Acceptance Rate (FAR) is evaluated by replacing the first frame by another person's data.

After several preliminary experiments, six characteristic pairs of feature points as shown in Table 1 are selected for effective evaluation. The characteristics are from the subjective impression that humans usually have, such as a long face, round face, prominent forehead, turned-up chin.

Table 1 Criteria of six selected characteristics.

| | Characteristics | Pair of feature points |
|---|---|---|
| 1 | Face length | (14,06) |
| 2 | Round face | (11,01) |
| 3 | Prominent forehead | (14,30) |
| 4 | Eyes and nose | (58,21) |
| 5 | Under nose | (average of 57 and 70, 58) |
| 6 | Turned-up chin | (06, average of 57 and 70) |

The SAD and SSD of the results for Table 1 are shown in Figure 4. Figure 4(a) shows FRR with the evaluation for the same one person and SSD values densely packed below 600. This Fig. 4(a) suggests that the threshold value of SSD can be about 100 for recognition. Another experiment with a different registered person to investigate FAR is shown in Figure 4(b). There are spaces in this graph below 100 of SSD. One of two results obtained by registering faces of two other different persons is shown in Figure 4(c).

These experimental results for Table 1 showed generally lower FRR FAR values than those of the preliminary results. The recognition rates are not sufficient at this stage. Because the video sequences include variations in speaking motion and facial expressions, to get rid of these fluctuations and use fully frontal faces and remove faces with expressions will improve the recognition performance. The next section will try to remove isolated failures from video sequences.

### B. *Removal of isolated failures*

For 2000 frames of images, isolated failed frames among the images are removed after the decision of the recognition by the threshold applied to values of calculated SSD. The removal of isolated failures is done by logical function using the majority decision rule. The results for FRR are shown in Table 2. The results for FAR are shown in Table 3.

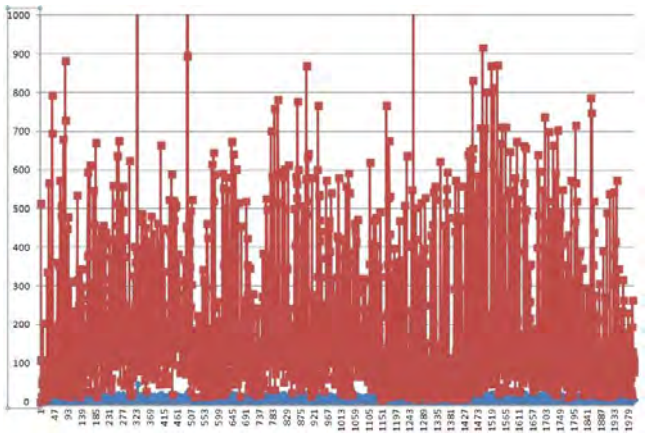


Fig. 6(a) SSD values for six characteristics. O-O

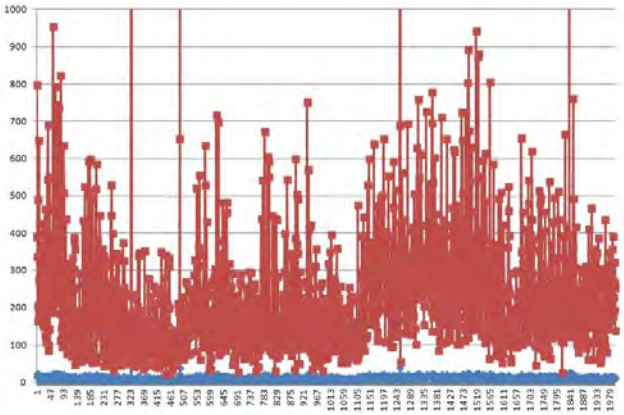


Fig. 6(b) SSD values for six characteristics. H-O）

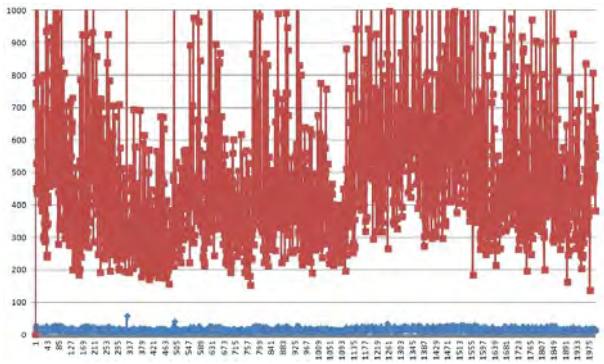


Fig. 6(c) SSD values for six characteristics. K-O

Table 2 FRR improvement.

| items | values | values | values |
|---|---|---|---|
| threshold | 100 | 100 | 100 |
| Number of isolated failures | 0 | 5 | 14 |
| Number of failures | 1212 | 296 | 15 |
| Total number | 1999 | 1999 | 1999 |
| FRR | 0.606303 | 0.613 | 0.00806 |

Table 3 FAR improvement.

| items | values | values | values |
|---|---|---|---|
| threshold | 100 | 100 | 100 |
| Number of isolated failures | 0 | 3 | 5 |
| Number of failures | 317 | 6 | 0 |
| Total number | 1999 | 1996 | 1994 |
| FAR | 0.1586 | 0.003006 | 0.0 |

Table 2 shows that FRR values are large without the removal of isolated failures. But after removal of isolated failures, FRR decreases greatly. Especially, for the case of the removal with the number of isolated failure frames of 14, FRR remarkably decreases even to 00806. Table 3 shows FAR improvements. With the removal of isolated failures, FAR can be zero if the number of isolated failure frames is 5.

## IV. Conclusions and Future Works

Face recognition using automatic detection of feature points and evaluating functions derived from feature point values was carried out. As for the evaluation function, using the overall average, which was tried in the preliminary experiments, did not show good performance. From the preliminary experiments, to use facial characteristics as an evaluation function rather than to use all featue points uniformly shows large discrimination ability in face recognition. Removal of isolated failures shows good improvements. Some results even show FAR=0 using the removal of isolated failures.

In the future, a more effective evaluation function composed of feature points with facial characteristics will be considered. More accurate rotation compensation and size normalization for input video sequences will be considered. Also, embedding to authentication login system in ref [7] will be considered.

## Acknowledgment

The authors are grateful for Project Research Grant of Joint University Collaboration offered by Shibaura Institute of Technology in Japan for supporting this research.

## References

[1] Tony Jebara, “Current Vision Systems for Face Recognition”, http://www.cs.columbia.edu/~jebara/htmlpapers/UTHESIS/node8.html

[2] Bonnen, K. Klare, B.F. Jain, A.K., "Component- Based Representation in Automated Face Recognition", IEEE Transactions on Information Forensics and Security, Vol.8, No.1 pp.239-253, Jan. 2013

[3] http://www.luxand.com/facesdk/#facialfeatures.

[4] Stephen Milborrow, Fred Nicolls, "Locating Facial Features with an Extended Active Shape Model", Proceedings of the 10th European Conference on Computer Vision:(ECCV '08 Proceedings) Part IV pp.504-513, Springer-Verlag Berlin, Heidelberg 2008

[5] S. Milborrow and F. Nicolls, "Active Shape Models with SIFT Descriptors and MARS", International Conference on Computer Vision Theory and Applications (VISAPP) pp.380-387. 2014

[6] http://www.milbo.users.sonic.net/stasm/

[7] Kazuo Ohzeki, YuanYu Wei, Yutaka Hirakawa, Toru Sugimoto, “Authentication System using Encrypted Discrete Biometrics Data”, Proceedings of TRUST 2014, Geece, Springer LNCS 8564 pp.210-211 June 30-July2 2014.

# Optical Signal Processing to Analyze Fluid Absorption inside the Skin Using Point by Point Photon Counting

Bushra Jalil*, Ovidio Salvetti*, Marco Righi*, Luca Poti†, and Antonio L'Abbate‡
*Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" CNR, Pisa, Italy
Email: bushra.jalil@isti.cnr.it
†Consorzio Nazionale Interuniversitario per le Telecomunicazioni, CNR, Pisa, Italy
‡Istituto di Fisiologia Clinica CNR, Pisa, Italy

***Abstract*—Time-correlated single photon counting (TCSPC) is popular in time resolved techniques due to its prominent performance such as ultra-high time resolution and ultra-high sensitivity. This paper presents advance signal processing techniques on the optical TCSPC signals obtained from the series of experiments on fabricated tissue like phantom. A pulsed laser sources at a wavelength of 830 nm transmits the light through the surface of phantom and finally at receiver side, photon counting device generates the histogram of the receiving signal. The noisy data obtained from the photon counter is processed with the splitting based denoising method. The method divide the signal into different subsets based on the transitions. Each subset is then processed individually and final merging of all subsets gives noise free signal. The main objective of this work is to analyze the signal obtained from photon counter in context of skin blood absorption. We had examined the signal obtained by varying the distance between transmitter and receiver to extract the features relevant to the diagnostic problem. Experimental results with our prototype shows more scattering with the increase in the distance at 3dB level and hence less absorption with increase in the distance.**

## I. Introduction

In optical instrumentation, time-resolved techniques measure time-dependent transmittance, reflectance, and fluorescence in response to illumination by an ultrashort light pulse. Two good examples are time-resolved diffuse optical tomography and fluorescence lifetime imaging. Both techniques require measurements in the time domain, in which short excitation pulses are used [1, 2, 3]. With periodic excitation, e.g., from a laser, it is possible to extend the data collection over multiple cycles of excitation [1, 2]. However, such techniques often suffer for high noise due to the measure time duration and the low optical signal. For that reason, averaging is required. In order to reduce the noise the number of acquisition gets higher and higher making the system time consuming and sensitive to time drift. In this work we have used advanced signal processing methods to reduce impact of noise for each individual acquisition in order to reduce both the number of acquisition and the time. We have used advanced signal processing methods to further investigate the signal acquired from such systems. The complete experimental setup is illustrated in section 2. The obtained signal is further denoised with the splitting based algorithm explained in section 3. Afterword, the extraction of parameters and retrieved information is finally explained in section 4.

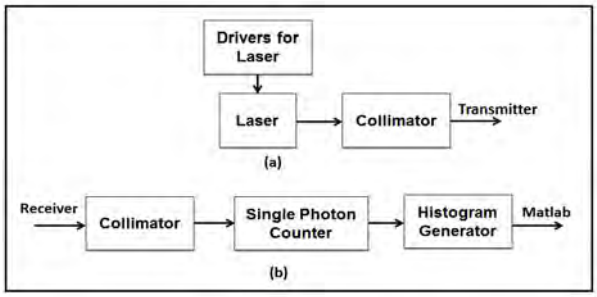



Figure 1. Block diagram of the experimental setup.

## II. Experimental setup

We propose to study the penetration of a pulsed laser light on human skin tissues in case of normal and pathological situation. The complete hardware setup is shown in Figure 1. 830nm laser source from picoquant is used to transmit the signal. We had conducted phantom, shown in Figure 2a, based experiments. The fabricated tissue like solid phantom mimics the real human skin. The flow chart of the processing is illustrated in Figure 2b. Each 1D signal belongs to the respective $(x, y)$ position. The scanning time of each measurement was 20sec, set at the constant rate.

## III. Denoising

In order to reduce number of acquisition, we had applied advanced signal processing to remove noise elements. In the past two decades, wavelet transform has been used as significant non parametric estimation tool to extract noise elements from the signal. Antoniadis et al provided an extensive review of the vast literature of wavelet shrinkage and wavelet thresholding estimator developed to denoise data [4]. Among these denoising techniques, the modulus maxima approach proposed by Mallat et al. has received the most attention in continuous and non-orthogonal domains [5,6]. Although

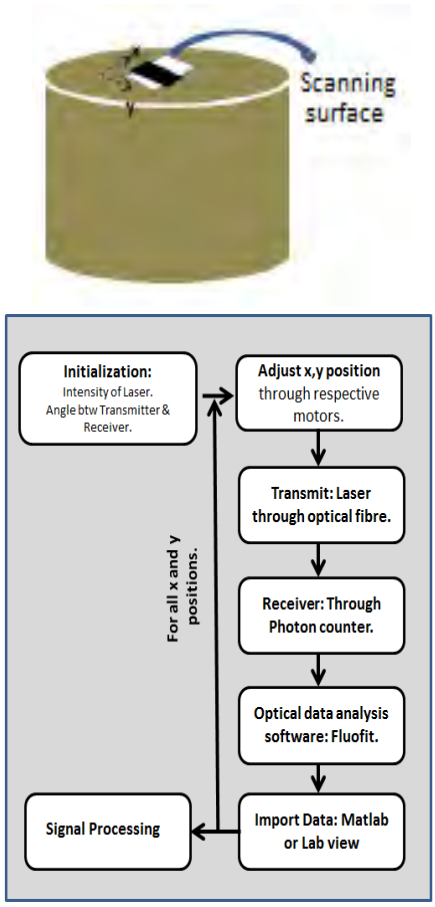



Figure 2. (a) Phantom based experiment. Black line on the surface blocks the photons to pass through the surface and result in reducing the peak of the histogram of photon counts, (b) Flow chart of the data acquisition method.

many researchers have proposed different methods to estimate signals from the evolution of the Wavelet Transform Modulus Maxima (WTMM) across different scales, still the proposed reconstruction process are either complicated or computationally expensive. Also, very few of them discussed about the singularities or interruptions present inside the signal. The main objective of these methods is to extract the noise free signal, but not on preserving the energy at the edges. In this work, we will examine the absorption scattering by changing the distance between Tx and Rx with respect to the peak of photon counting histogram. Therefore, it is of extreme importance to detect the peak signal more accurately [5]. In order to achieve our objectives, we had applied the splitting based denoising in order to preserve the peak amplitude. Figure 3 illustrate the block diagram of splitting based method.

### A. Principle of the Method

Suppose we have a noisy function $Y$ such that:

$$Y = F(t_i) + \epsilon_i, \qquad where \quad i = 1, .....N \tag{1}$$

$F(t_i)$ is the deterministic function with $t_i = (i-1)/N$, $N$ is the total number of samples and $\epsilon_i$ is white gaussian noise $N(0, \sigma^2)$. The aim of the current work is to estimate the function $F$ with the minimum error. In order to estimate the function $F$, the method propose to split the function $Y$ into its subsets in spatial domain on the basis of sharpness of transitions or edges such that :

$$Y \supseteq Y_{i=1,2...J} \tag{2}$$

$Y$ denotes the $set$ of all samples and each subset $Y_i$, with $i = 1, ..., J$ contains $K_i$ adjacent samples $y_{i,l}$ with $l \epsilon 1, ... K_i$. $J$ is the total number of subsets and $K_i$ is the length of respective subset. Piecewise analysis has been performed on each subset of the function $Y$ individually. The selection of subset is defined on the basis of Stein's Unbiased Estimate of the Risk (SURE) based non linear thresholding of Wavelet Transform Modulus Maxima (WTMM)[6].

### B. Splitting Method

In order to split the signal, continuous wavelet transform based multiscale analysis has been applied on signal $y(t)$ to compute the modulus maxima by using an integrable function [5]:

$$WT(u,s) = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} y(t) \Psi^*[\frac{(t-u)}{s}]dt \tag{3}$$

Where $WT(u,s)$ is the wavelet coefficient of the function $y(t)$, $\Psi(t)$ is the analyzing wavelet, $s(>0)$ is the scale parameter and $u$ is the position parameter. The Gaussian function $(\theta(t))$ has an important property of continuous differentiability, which makes this function suitable for the analysis of most types of signals. Therefore, the derivative of the Gaussian function has been used as a wavelet analyzing function $(\Psi(t))$ for the splitting method. These WTMM computed by using the derivative of the Gaussian wavelet is defined as any point $(u_0, s_0)$ such that $||Wf(u, s_0)||$ has a local maximum at $u = u_0$ [6].

In order to select the optimum threshold criterion for making the subsets, the level-dependent thresholds are derived from modulus maxima by regarding the different scale levels as independent multivariate normal estimation problems. SURE gives an estimate of the risk for a particular threshold value t; minimizing this in t gives a selection of the threshold level for that level j ( $j = 1, 2.....J$ ) [7]. SURE-based thresholding on modulus maxima results in splitting the signal into subsets based on the sharpness of transitions.

### C. Reconstruction

Piecewise analysis has been performed on each subset of the function $Y$ individually by following the algorithm as follows:

1) Lipschitz exponents computed from each subset result in identifying the regular or smooth points in respective subset [5, 6].
2) The reconstruction of each subset is based on data sample and smoothing. The restoration method between

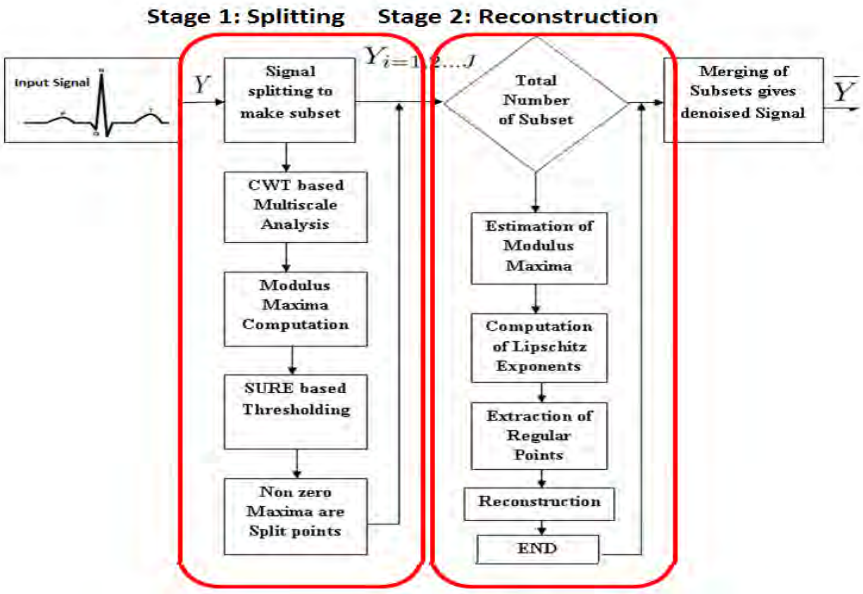



Figure 3. Block Diagram of the Splitting based denoising algorithm. Mainly two stages i) Splitting and ii) Reconstruction. Final merging of all reconstructed subsets gives complete denoised signal.

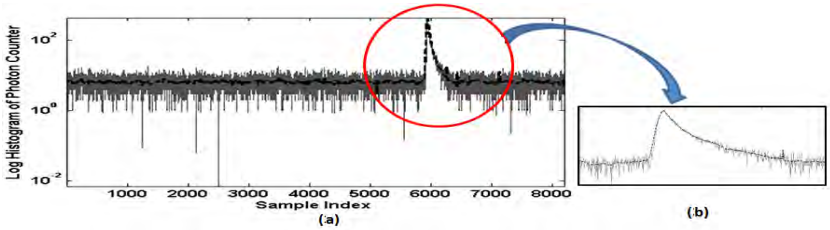



Figure 4. Results obtained with Denoising Algorithm. 4a) Solid lines shows noisy signal and dotted shows de-noised signal. Also it can be observed from 4b that the denoising method has not altered the peak of the signal.

regular data samples utilizes all sampled points and the smoothness of each subset to estimate the best fit. At the final stage, merging of all reconstructed subsets result in giving fully denoised signal. We define $y_{l=1,...N}\ \epsilon Y$

$$y_{i,l}^{k+1} = y_{i,l}^{k} + \lambda_i^k(-\frac{\partial C_i^{MSE,k}}{\partial y_{i,l}}) + \gamma_i^k(-\frac{\partial C_i^{MSO,k}}{\partial y_{i,l}}) \quad (4)$$

Where $i = 1, ..., J$ and $y_{i,l}$ with $l \epsilon 1, ...N_i$ adjacent samples in each subsets. $k$ define the iteration step. $C_i^{MSE}$ is the mean square error estimation of the restored subset with the original signal of respective subset $(y_{oi})$.

Figure 4 shows the obtained signal after denoising. It can be observed from Figure 4b that the signal has maintained its peak after denoising. In the next section, we will use the peak of the signal to extract the useful information about the skin tissue.

## IV. Processing: Varying Distance between Tx and Rx

We had conducted the experiment at a single point on the surface of the phantom by varying the distance between transmitter and receiver. The data from the photon counting device, displayed as a histogram of counts on the provided Picoharp interface [2] as in Figure 5. In order to investigate the effect of distance between Tx and Rx, We had started with the minimum distance of 5mm between the two (Tx and Rx) and then increase with an equal step of 5mm. Trace 0 (in Figure 5) shows minimum distance. The obtained results shown in Figure 6 concludes that, increase in the distance result in reducing the peak count to the half of the previous but more widening of pulse width. By increasing the distance, max counts of photons (Trace 1) reduced to nearly half of the previous. However, reaching time of the peak also slightly increases with the distance but was approximately with in the range of 40ns. At the second stage, we had estimated the width of the pulse at several levels by varying distances. At 3db level, more widening of pulse width was observed by increasing the distance and hence more scattering on the surface. In Figure 6c, curve at respective distance between Tx and Rx shows the pulse width at different levels. More scattering and less absorption was observed with the increase in distance. We can conclude from these findings that by varying the distance between Tx and Rx on the skin surface, we can approximately estimate the depth of the fluid on the level of the skin.

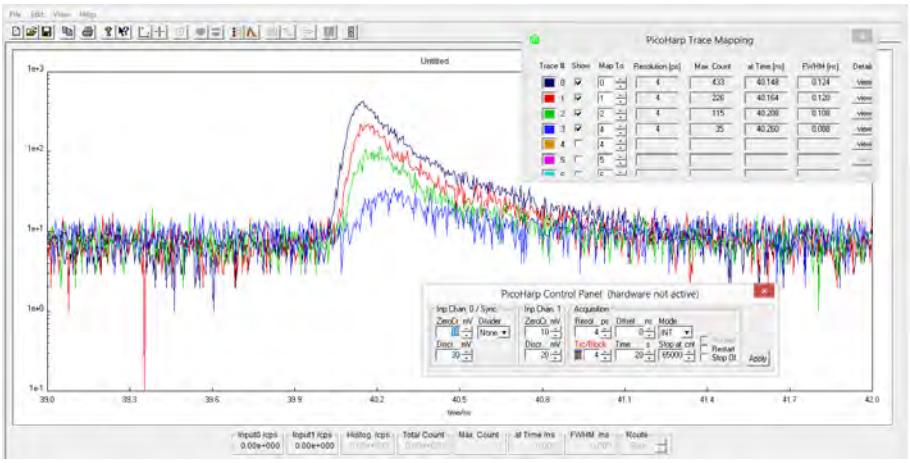



Figure 5. Signal obtained at a point on the Pico harp software. Increase in the distance between Tx and Rx results in reducing the peak counts but increase in the pulse width.

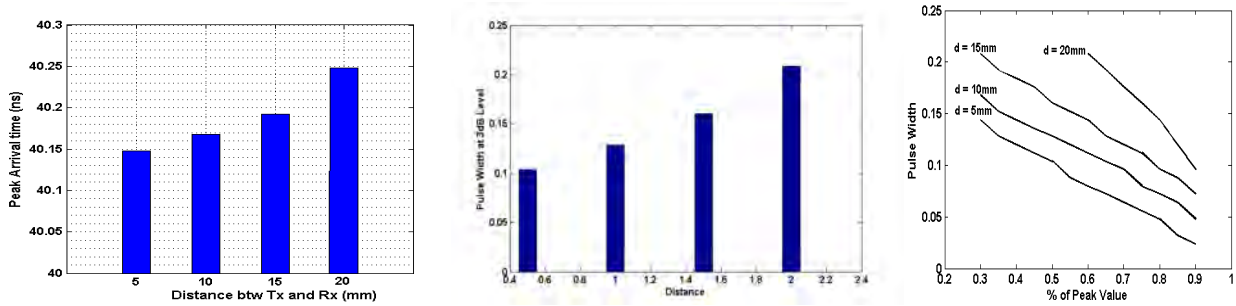



Figure 6. By varying the distance between Tx and Rx. a) Histogram of Peak Arrival time. b) Histogram and Curves at 3dB and c) Pulse width at different levels.

## V. Conclusion

We had worked on optical tomography techniques e.g. time correlated single photon counting methods. We had conducted phantom based experiments to analyze the fluid absorption on the human skin level . The hardware involved in these experiments includes Thor labs Dc motors, Pico quant laser source and corresponding photon counting devices. Pico quant also provides external interface software to better visualize the results. The experiments includes surface scanning and analyzing the signal with variable distance between Tx and Rx. By increasing the distance between Tx and Rx, the max counts of photons keep on reducing but increase in the pulse width. We conclude that the mean penetration depth of diffusively reflected photons depends on the distance between source and the detector.

## VI. Acknowledgement

This work has been supported by 'Regione Toscana, Direzione Generale Competitività e Sviluppo Competenze, Area di Coordinamento della Ricerca'.

## References

[1] Qiang Zhang ; Nan Guang Chen, Pseudo-random single photon counting: the principle, simulation, and experimental results, Proc. SPIE 7170, Design and Quality for Biomedical Technologies II, 71700L (February 23, 2009)

[2] http://www.picoquant.com/

[3] Weirong Mo and Nanguang Chen, Fast time-domain diffuse optical tomography using pseudorandom bit sequences, Optical Society of America, 2008.

[4] A. Antoniadis, J. Bigot, T. Sapatinas, "Wavelet Estimators in Nonparametric Regression: A Comparative Simulation Study ", Journal of Statistical Software, Vol. 6, pp. 1-83(2001)

[5] Eric Fauvet, Olivier Laligant, Bushra Jalil, "Signal Restoration via a SplittingApproach", EURASIP Journal on Advances in Signal Processing 38, 2012.

[6] S. Mallat, "A wavelet Tour of signal processing, Second Ed, Academic Press (1998).

[7] Stein, Charles M., "Estimation of the Mean of a Multivariate Normal Distribution". The Annals of Statistics,.Volume 9, Number 6 (1981), 1135-1151

# *Optimal Facial Areas for Webcam-based Photoplethysmography*

*M. Kopeliovich,*
*Institute for Mathematics, Mechanics, and Computer Science in the name of I.I. Vorovich, Southern Federal University, Rostov-on-Don, Russia,*
*kopeliovich.mikhail@gmail.com*

*M. Petrushan,*
*A.B. Kogan Research Institute for Neurocybernetics, Academy of Biology and Biotechnology, Southern Federal University, Rostov-on-Don, Russia,*
*drn@bk.ru*

***Abstract*—Photoplethysmography is a perspective application of Computer vision. Due to its noninvasiveness and low technical requirements it can be applied for long-term tracking of pulse rate. Mostly, photoplethysmography is based on evaluation of skin color variations within specified areas in facial images. This study proposes certain areas in a facial image that are most appropriate for pulse registration. Viola-Jones algorithm is used to detect a face. Pulse rate is evaluated by using FFT and searching for extreme values of a spectrum. According to proposed criterion, optimal facial areas found are the area near the nose and area on the nose between eyes.***

***Keywords*—photoplethysmography; color analysis; webcam-based pulsometry***

## I. INTRODUCTION

Various methods such as electrocardiography and plethysmography are used for remote pulse registration. However, long-term real-time heart rate tracking should be based on contactless registration to be convenient for applied usage. Moreover, traditional contact pulsometry is not effective for some special applications such as tracking cosmonaut or pilot functional state, detecting pulse rate of neonates with fragile skin and patients with damaged skin [1]. One of the possible decisions for such special issues is video-based photoplethysmography (PPG).

Development of applied webcam-based PPG methods has started only recently [2] [3]. Such methods are based on analysis of color signal - sequence of red, green and blue values extracted from video [2]. Mainly, three steps are used: getting RGB-values from each region of interest (ROI) in a frame, noise removal for color signal and evaluating pulse rate by FFT. Heart rate is expected to correspond to a maximum of spectrum power in a frequency band bounded by physiologically possible pulse extremes.

Regions of video frame for color signal extraction can be defined differently. One way to choose a ROI is to divide frame into smaller parts by a regular grid [2]. Then color signal is collected within each cell of the grid. Another way is to detect a face by Viola-Jones algorithm [4] or Lienhart and Maydt algorithm [5] and then to select areas by its relative coordinates in detected face rectangle. In order to remove noise color signal is collected by averaging the values of red, green and blue over each ROI. Color averaging over time is also used.

Next step of averaged signal processing is usually based on Fourier transform or blind source separation method. For example, in [1] signal is processed by band-pass filter firstly. Then, heart rate is evaluated as a frequency that corresponds to a maximum value of FFT spectrum power. More complex method described in [3] includes independent components analysis (ICA) of color signal. One of source signals found by ICA is transformed by FFT and used for heart rate estimation by searching for a maximum value of spectrum power.

Another approach is based on head motions analysis for pulse rate estimation [1], [6]. First, Viola-Jones algorithm is used for face detection. Second, points of interest are chosen by "Good Features to Track" algorithm [7]. Third, trajectory of every point of interest is tracked by Lucas-Kanade algorithm [8]. The pulse signal is assumed to be a vertical component of tracking trajectories. Noise is removed by trajectory averaging over time. Finally, Discrete Cosine Transform is used to evaluate pulse.

In spite of intensive exploration of video-based PPG, perspective of its applied usage for long-term health monitoring remains unclear. In order to find possibilities and limitations of such PPG several problems need to be studied: optimal visual signal that corresponds to heart rate, optimal regions for signal extraction within facial area, effects of environment factors on color signal. The presented study aims to examine the problems above, focusing on the optimal ROI task. Due to a few number of analyzed regions brute force method was used to find an optimal ROI.

## II. METHODS

### *A. Experimental setup*

Localization of areas for color signal extraction was based on Viola-Jones method. However, even if face was immovable, fluctuations of evaluated face coordinates occurred (from 5 to 20 pixels). This effect considerably complicates a process of collecting color signal from certain zones in the facial image. To overcome this obstacle a construction that fixes camera position relatively to face position was designed (Fig. 1). This system consists of bicycle helmet with a fixed stick 45 cm length and web-cam attached to the end of the stick. To reduce illumination noise 2 LED light sources were attached to constructed system. LEDs have power of 1 Watt and were directed to a face. To balance the construction, a weight of 1 kg was fixed at the opposite side of the stick.

Logitech C920 Webcam was used for video recording. Videos were recorded in color (24-bit RGB) at 30 frames per second (fps) with pixel resolution of 1920x1080. 4 participants (all males) aged from 22 to 32 were enrolled for this study.

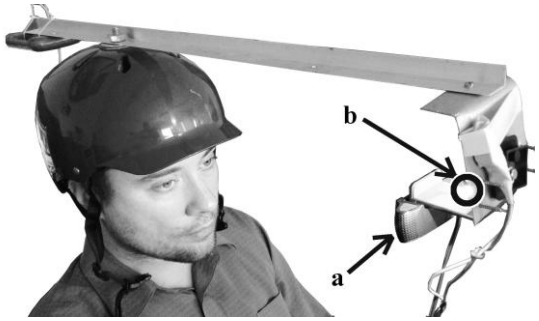



Fig. 1. Designed construction for data collecting. a – web-camera, b – LED.

Experiment comprised three steps. Firstly, video was recorded before physical exertion. Secondly, subject did physical exercises on an exercise bike during 1 minute. Thirdly, another video was recorded while subject was in agitated state (after physical exertion). The length of each recorded video was 2 minutes. Heart rate was manually measured at the beginning and at the end of each video.

### *B. Data collection*

Due to system construction head movements don't change face position on the frame. In order to get face rectangle which would better correspond to real face we use first five frames to detect face rectangle by Viola-Jones algorithm. Coordinates of these rectangles were averaged and then the resulting rectangle was used on each frame of video.

A set of areas of interest was defined as shown in the Fig. 2. Their coordinates are fixed relatively to coordinates of a face rectangle. Head motions caused by pulsation in the vessels (described in [1], [6]) force the webcam to make slight movements relatively to face. It has almost no effect on variations of coordinates of areas due to their size. However, it provides information on how signal caused by head motions can correspond to pulse rate. Therefore one area of interest was placed on helmet zone and contained strong intensity gradients.

Red, green and blue color values were averaged independently within each ROI for each frame of the video.


Fig. 2. Selected areas on facial rectangle.

### *C. Data processing. Optimality criterion.*

Color signal CS(t) for each RGB channel is a parametric function of time. It describes changes of averaged RGB values in time during the experiment. Examples of G (green) and B (blue) color signals for two ROIs are shown in the Fig. 3.

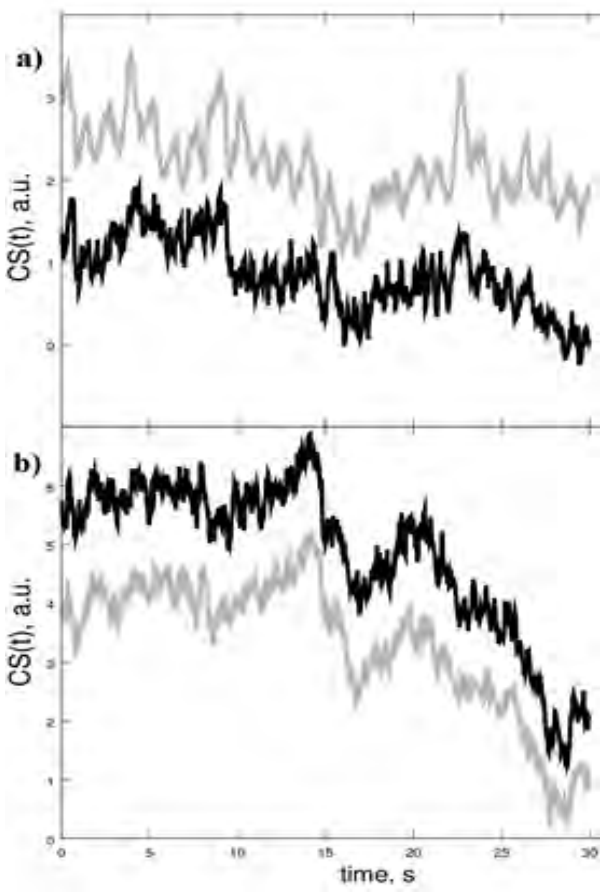



Fig. 3. Examples of green (shown by gray) and blue (shown by black) color signals, extracted from different facial areas of the same person (see Fig 2.) during 30 seconds: a) area №2, b) area №8.

60 second length signal was extracted from total color signal with offset equal to 10 seconds from the beginning of the video record. Fourier spectrum of extracted color signal was calculated: $Pw_{10}(w)=FT(CS(t))$, where FT is Fourier transform. Then, spectrum was calculated again from a signal extracted with offset equal to 11 seconds. 25 Fourier spectra were calculated in such way with different signal offset. Then, averaged spectrum was computed:

$$Pw(w) = \frac{1}{25}\sum_{i=10}^{34} Pw_i\,(w).$$

A maximum of averaged spectrum power in a range $[w_r - b; w_r + b]$ was found, where $w_r$ is the known value of heart rate (measured manually), b – is the auxiliary parameter equal to 4 bpm. Found value $P_s$ is a maximum in a bounded range closest to real heart rate.

$$P_s = \max{}_w(Pw(w): w \in [w_r - b; w_r + b])$$

Optimality criterion for facial areas was defined as the ratio of maximum power with frequency, closest to real heart rate ($P_s$), to the mean value of spectrum power in the range [45, 100] bpm.

$$Rm = \frac{P_s}{\frac{1}{56}\sum_{w=45}^{100} Pw(w)} \quad (1)$$

Relative maximum (1) is higher if spectrum power distribution has a global maximum near real pulse rate and is lower otherwise. Facial areas with maximal value of criterion (1) are expected to be optimal for pulse registration.

## III. Results

Relative maximum of Fourier spectrum power (1) was higher for green channel then for blue channel in the both types of experiments – before and after physical exercise (Fig 4.).

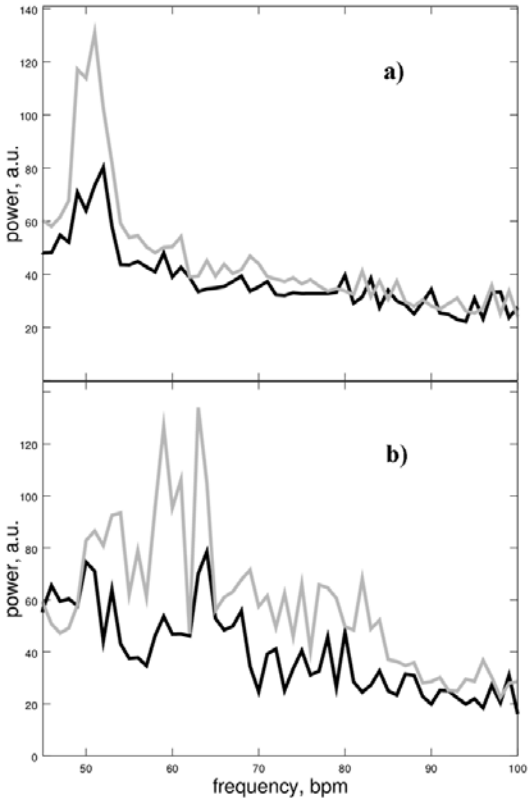



Fig. 4. Examples of Fourier spectra of color signal for green (shown by gray) and blue (shown by black) channels of the same person: a) – before a physical exercise, b) – after a physical exercise.

Due to the fact of the best correspondence of green channel variations to the heart rate, known from previous works [2] and confirmed during described experiments, this (green) channel was used for further analysis.

The values of the optimality criterion for different facial areas were calculated (Table 1). Mean values (with standard error of the mean in brackets) of this criterion are shown for different experiment types: before exercise, after, and for all types. Maximal values are shown in bold.

TABLEI. OPTIMALITY CRITERION VALUE FOR FACIAL AREAS

| Area index | *Before exercise* | *After exercise* | *All experiments* |
|---|---|---|---|
| 0 | 0.98 (0.29) | 1.13 (0.07) | 1.06 (0.14) |
| 1 | **2.12** (0.41) | 1.36 (0.31) | 1.74 (0.28) |
| 2 | 1.95 (0.36) | 1.39 (0.33) | 1.67 (0.25) |
| 3 | 1.95 (0.20) | 1.33 (0.26) | 1.64 (0.19) |
| 4 | **2.11** (0.41) | **1.43** (0.35) | **1.77** (0.28) |
| 5 | 2.00 (0.35) | 1.34 (0.29) | 1.67 (0.25) |
| 6 | 1.86 (0.29) | 1.41 (0.45) | 1.64 (0.26) |
| 7 | 1.74 (0.24) | 1.09 (0.16) | 1.41 (0.18) |
| 8 | 1.27 (0.13) | 0.94 (0.06) | 1.10 (0.09) |

Optimal facial areas satisfying the proposed criterion were found. Areas coordinates are: [$x_1$=0.420, $y_1$=0.346, $w_1$=0.160, $h_1$=0.104], [$x_4$=0.218, $y_4$=0.569, $w_4$=0.138, $h_4$=0.147] (for area 1 and 4 in Fig. 4 respectively), where $x_i$, $y_i$ – area's point of the left up corner and $w_i$, $h_i$ – area's width and height. These coordinates are given relatively to coordinates of face rectangle in fractions of facial width and height respectively. Coordinate system origin is positioned in the top left corner of the facial rectangle. The difference of criterion values for areas 4 and 6 possibly was caused by slight asymmetry of illumination and face position in image.

## IV. Conclusions

The series of experiments was conducted to find optimal facial areas for webcam-based photopletismography. In order to exclude environmental factors camera and light sources were fixed relatively to a head. Coordinates of optimal ROIs were found according to proposed optimality criterion (1). The value of relative maximum of Fourier spectrum is much lower in a second type experiments, possibly due to fast pulse rate relaxation occurred after physical exercise.

## Acknowledgment

The work is supported by The Ministry of Education and Science of the Russian Federation, base project № 213.01-11/2014-30.

## References

[1] G. Balakrishnan, F. Durand and John Guttag, "Detecting Pulse from Head Motions in Video," in Proc. CVPR, pp. 3430-3437, 2013.

[2] W. Verkuysse, L. O. Svaasand and J. S. Nelson, "Remote plethysmographic imaging using ambient light," Opt. Express, Vol. 16, Issue 26, pp. 21434-21445, 2008.

[3] M.-Z. Poh, D. J. McDuff and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," Optics Express 18, pp. 10762-10774, 2011.

[4] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proc. CVPR, pp. 511-518, 2001.

[5] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," Proceedings of IEEE Conference on Image Processing, 2002.

[6] R. Irani, K. Nasrollahi and T. B. Moeslund, "Improved Pulse Detection from Head Motions Using DCT," Proceedings 9th International Conference on Computer Vision Theory and Applications, pp. 118-124, 2014.

[7] J. Shi and C. Tomasi, "Good features to track," in Proc. CVPR, pp. 593-600, 1994.

[8] J.-Y. Bouguet, "Pyramidal Implementation of the Lucas Kanade Feature Tracker," Intel Corporation, Microprocessor Research Labs, Technical Report, 2000.

# Optimization of mutual information-based stochastic gradient ascend algorithm for image registration

*Sergey Voronov, Alexander Tashlinskiy, Ilia Voronov*
Radio Engineering Department
Ulyanovsk State Technical University
Ulyanovsk, Russia
s.voronov@ulstu.ru

***Abstract* — An optimization criterion for stochastic gradient search of registration parameters when mutual information is considered as an objective function is found. The criterion aims to increase the convergence rate and consists in using the most informative points in terms of optimal Euclidian distance between the true point coordinates and it's coordinates calculated using current registration parameter estimates.**

***Keywords — image registration, mutual information, gradient ascend, optimization.***

## I. INTRODUCTION

Digital image registration is a process by which the most accurate match is determined between two images, which may have been taken at the same or different times, by the same or different sensors, from the same or different viewpoints. The registration process determines the optimal transformation, which will align the two images. This has applications in many fields as diverse as medical image analysis, pattern matching and computer vision for robotics, as well as remotely sensed data processing. In all of these domains, image registration can be used to find changes in images taken at different times, or for object recognition and tracking.

Spatial domain methods operate directly on pixels, and the problem of the estimation of registration parameters $\overline{\alpha}$ becomes the problem of searching for the extreme point of a multi-dimensional objective function $\mathrm{J}(\mathbf{Z}, \overline{\alpha})$. The objective function measures the similarity between two images $\mathbf{Z}^{(1)} = \left\{z_{\bar{j}}^{(1)}\right\}$ and $\mathbf{Z}^{(2)} = \left\{z_{\bar{j}}^{(2)}\right\}$, where $\bar{j} \in \Omega$ are nodes of grid mesh $\Omega$ on which the images are defined. There is a wide variety of similarity measures that can be used as objective functions [1]. The decision of which objective function to choose is usually based on the specifics of images, deformation properties and conditions. Recently, objective functions from the theory of information are becoming more popular. Among these functions the most interesting is mutual information. It has been found to be especially robust for multimodal image registration and registration of images with great non-linear intensity distortion [2]. However, mutual information has some drawbacks. One of them is relatively high computational complexity. This fact makes it difficult to implement this objective function in real-time image and video processing systems.

The choice of optimization search technique depends on the type of problem under consideration. Traditional nonlinear programming methods, such as the constrained conjugate gradient, or the standard back propagation in neural network applications, are well suited to deterministic optimization problems with exact knowledge of the gradient of the objective function. Optimization algorithms have been developed for a stochastic setting where randomness is introduced either in the noisy measurements of the objective function and its gradient, or in the computation of the gradient approximation. Stochastic gradient ascend (descend) is one of the most powerful technique of this class. It is an iterative algorithm, where registration parameters can be found as follows:

$$\hat{\overline{\alpha}}_t = \hat{\overline{\alpha}}_{t-1} - \mathbf{\Lambda}_t \overline{\beta}_t(\mathrm{J}(Z_t, \overline{\alpha}_{t-1})),$$

where $\overline{\beta}$ – gradient estimation (noisy gradient) of the objective function $\mathrm{J}$ obtained using not each pixel in the images but a sample $Z_t$ taken randomly on each iteration, $\mathbf{\Lambda}_t$ – positive-definite gain matrix: $\mathbf{\Lambda}_t = \left\|\lambda_{it}\right\|$, $\lambda_{it} > 0$, $i = \overline{1,m}$; $m$ – number of registration parameters.

When images to be registered are corrupted with noise, it has been shown that the implementation of sign function can improve the efficiency of the procedure:

$$\hat{\overline{\alpha}}_t = \hat{\overline{\alpha}}_{t-1} - \mathbf{\Lambda}_t \operatorname{sign}\left(\overline{\beta}_t(\mathrm{J}(Z_t, \overline{\alpha}_{t-1}))\right),$$

where $\operatorname{sign}(x) = -1$ if $x < 0$, $\operatorname{sign}(x) = +1$ if $x > 0$ and $\operatorname{sign}(x) = 0$ if $x = 0$.

The main disadvantages of this optimization algorithm are many local extreme points of the objective function due to the use of small samples and relatively short working range in terms of registration parameters to be estimated. To overcome these problems the number of sample elements can be increased. However, this leads to significant increase in computational efforts. Thus, the problem of optimization of

The work was supported by RFBR 13-01-00555-a

stochastic gradient algorithm for image registration is an important, especially for real-time processing systems.

## II. Optimization method

Mutual information of two images is maximal when these images are perfectly aligned. It can be estimated in terms of entropy [3]:

$$\hat{\mathrm{J}}(\mathbf{Z}^{(1)},\mathbf{Z}^{(2)},\bar{\alpha}) = \hat{H}(\tilde{Z}^{(1)}) + \hat{H}(\mathbf{Z}^{(2)}) - \hat{H}(\tilde{Z}^{(1)},\mathbf{Z}^{(2)}),$$

where $\hat{H}(\mathbf{Z}^{(k)}) = -\sum_i p_{zk}(z_i)\log p_{zk}(z_i)$,

$$\hat{H}(\mathbf{Z}^{(1)},\mathbf{Z}^{(2)}) = -\sum_i\sum_k p_{z1,z2}(z_i,z_k)\log p_{z1,z2}(z_i,z_k)$$ –

are marginal and joint entropies; $p_{z1}$ and $p_{z2}$ – estimations of marginal probability distributions of image intensities $\{\tilde{z}_j^{(1)}\}$ and $\{z_j^{(2)}\}$; $\tilde{Z}^{(1)}$ – continuous image obtained from $\mathbf{Z}^{(1)}$ using an interpolation; $z_i$ – $i$-th intensity value; $p_{z1,z2}$ – joint probability distribution estimation, .

According to [4] estimation of mutual information gradient can be obtained using Parzen-window method for intensity probability distribution restoration and dividing sample into two parts $Z_a$ and $Z_b$, where part $Z_a$ is used for distribution restoration and part $Z_b$ – to entropy estimation. So, one can compute gradient estimation for *i*-th registration parameter to be estimated in the following way:

$$\beta_{it} = \frac{1}{\sigma\mu_b}\sum_{i\in Z_b}\sum_{j\in Z_a}\left(W_z^{(1)} - W_z^{(1,2)}\right)\left(\tilde{z}_i^{(1)} - \tilde{z}_j^{(1)}\right)\times$$
$$\times\left(\frac{d\left(\tilde{z}_i^{(1)} - \tilde{z}_j^{(1)}\right)}{dx}\frac{dx}{d\alpha_i} + \frac{d\left(\tilde{z}_i^{(1)} - d\tilde{z}_j^{(1)}\right)}{dy}\frac{dy}{d\alpha_i}\right), \quad (1)$$

where $W_z^{(1)} = \dfrac{G_\sigma\left(\tilde{z}_i^{(1)} - \tilde{z}_j^{(1)}\right)}{\sum_{z_j\in Z_a} G_\sigma\left(\tilde{z}_i^{(1)} - \tilde{z}_j^{(1)}\right)}$;

$$W_z^{(1,2)} = \frac{G_\sigma\left(\tilde{z}_i^{(1)} - \tilde{z}_j^{(1)}\right)G_\sigma\left(z_i^{(2)} - z_j^{(2)}\right)}{\sum_{z_j\in Z_a} G_\sigma\left(\tilde{z}_i^{(1)} - \tilde{z}_j^{(1)}\right)G_\sigma\left(z_i^{(2)} - z_j^{(2)}\right)},$$

$R_G(x) = \dfrac{1}{\sqrt{2\pi\sigma}}e^{\frac{-x^2}{2\sigma^2}}$ – Gaussian kernel function; $\mu$ – sample size; *x* and *y* – coordinates of points in the sample.

In [5] it has been shown that in order to optimize stochastic gradient estimation procedure (increase the convergence rate) we can on each iteration optimize the sample plan, i.e. we can use priori determined points that are considered as the most informative in terms of deformation, i.e. they have the best Euclidian distance *E* between their true coordinates and coordinates calculated using current registration parameters' estimates (mismatch Euclidian distance). As an optimization criterion the maximum ratio between gradient estimation mean and it's standard deviation can be used [6]:

$$\max\Psi = \max\left|\frac{\mathrm{M}[\beta(E)]}{\sqrt{\mathrm{D}[\beta(E)]}}\right|,$$

where $\mathrm{M}[\beta(E)]$, $\mathrm{D}[\beta(E)]$ – the mean and variance of gradient estimation.

Let us consider that images are corrupted with uncorrelated additive noise and both image and noise has Gaussian distribution with zero mean and standard deviation $\sigma_s$ and $\sigma_\theta$ respectively:

$$z_{jk}^{(2)} = s_j^{(2)} + \theta_j^{(2)},\ \tilde{z}_j^{(1)} = \tilde{s}_j^{(1)} + \tilde{\theta}_j^{(1)},$$

where $\tilde{s}_j^{(1)}$ and $s_j^{(2)}$ – information value; $\tilde{\theta}_j^{(1)}$ and $\theta_j^{(2)}$ – noise value.

Moreover, suppose that we have only 3 points in the sample, precisely $\mu_a = 2$, $\mu_b = 1$. For simplicity, we will consider the situation when the points lie on the circle with radius $\Re$ and make an equilateral triangle (fig. 1).

It can be shown that in this situation the mean and variance of mutual information gradient estimation (1) can be computed as follows:

$$\mathrm{M}[\beta(E)] \approx \frac{3\sigma_s}{4.2\sigma}\left[(R(a)-R(b))(R(a_+)-R(a_-)-R(b_+)+R(b_-)+\right.$$
$$\left.+R(E_-)-R(E_+)+R(E))\right],$$

$$\mathrm{D}[\beta(E)] = \left(\frac{3\sigma_s}{4.2\sigma}\right)^2\left[g^{-1}(2-R(a_+)-R(a_-)-R(b_+)+R(b_-)-\right.$$
$$\left.-R(E)(1-R(E)(1-g^{-1}+R(1)))\right]+\mathrm{M}[\beta]^2,$$

where $R(E_\pm) = R(E\pm 1)$;

$$a = \sqrt{(1.5\Re)^2 + \left(E - \frac{\sqrt{3}\Re}{6}\right)^2},\quad a_\pm = \sqrt{(1.5\Re)^2 + \left(E - \frac{\sqrt{3}\Re}{6}\pm 1\right)^2},$$

$$b = \sqrt{(1.5\Re)^2 + \left(E + \frac{\sqrt{3}\Re}{6}\right)^2},\quad b_\pm = \sqrt{(1.5\Re)^2 + \left(E + \frac{\sqrt{3}\Re}{6}\pm 1\right)^2},$$

$g = \dfrac{\sigma_s}{\sigma_\theta}$ – signal-to-noise ratio.

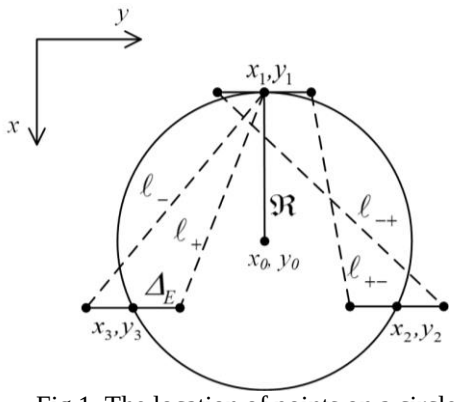
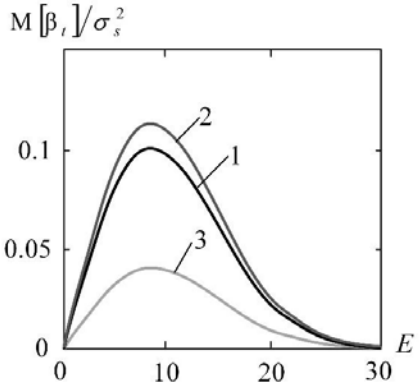
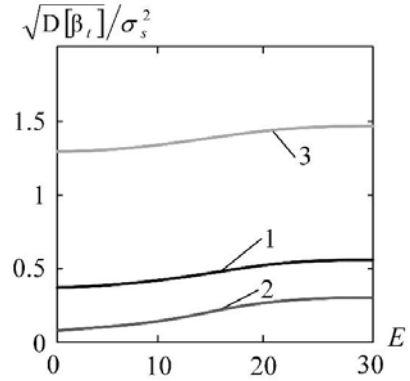



Fig.1. The location of points on a circle.

Fig.2 shows the curves of normalized mean (a) and standard deviation (b) for different circle radius and images with Gaussian correlation function with correlation radius 15. Here curve 1 is for $\Re = 5$, 2 – $\Re = 10$, 3 – $\Re = 15$. We can see that for the mean there is an optimal value of mismatch Euclidian distance and it does not depend on the circle radius.




a)




b)

Fig.2. Normalized mean (a) and standard deviation (b) of mutuall information gradient estimation for different circle radius

Fig.3 shows the curves of optimization criterion for the different circle radius (a) and for the different signal-to-noise ratio (b) and the same image parameters. For fig.3b curve 1 is for $g = 1000$, 2 – $g = 10$, 3 – $g = 2$.

Analyzing fig.3 we can conclude that there is an optimal value of mismatch Euclidian distance and it doesn't depend on the radius of the circle on which the sample points lie but it depends on the signal-to-noise ratio and it increases with the noise strengthen. For $g = 1000$ the optimal value is about 3, for $g = 10 - 4.8$, for $g = 2 - 7.9$.

To define the optimal (or suboptimal) region of samples (where points have the optimal mismatch Euclidian distance) we can use method proposed in [7]. This method is based on the chosen deformation model and its current parameters estimates. It has been shown that the optimal region for example for similarity deformation model is a circle with the radius defined as follows:

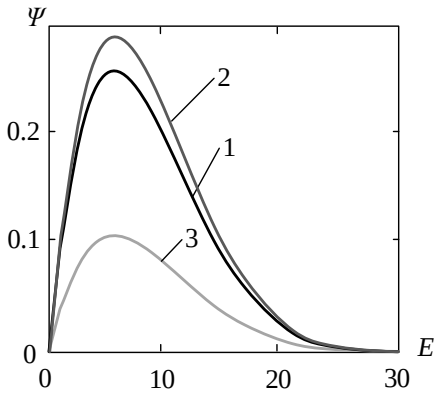
$$r^2 = \frac{(E_{op})^2 - (\varepsilon_x)^2 - (\varepsilon_y)^2}{1 + (1+\varepsilon_k)^2 - 2(1+\varepsilon_k)\cos\varepsilon_\varphi} + a^2 + b^2,$$

where $a = \dfrac{\varepsilon_x - (1+\varepsilon_k)(\varepsilon_x \cos\varepsilon_\varphi + \varepsilon_y \sin\varepsilon_\varphi)}{1 + (1+\varepsilon_k)^2 - 2(1+\varepsilon_k)\cos\varepsilon_\varphi}$,

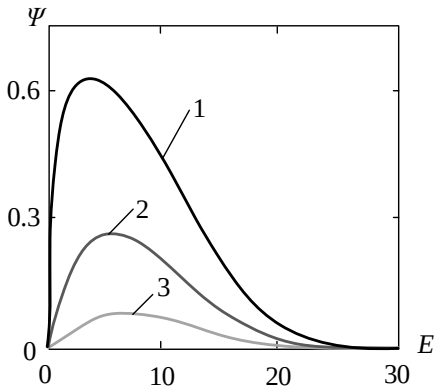
$$b = \frac{\varepsilon_y - (1+\varepsilon_k)(\varepsilon_y \cos\varepsilon_\varphi - \varepsilon_x \sin\varepsilon_\varphi)}{1 + (1+\varepsilon_k)^2 - 2(1+\varepsilon_k)\cos\varepsilon_\varphi},$$

$\bar{\varepsilon} = (\varepsilon_x, \varepsilon_y\, \varepsilon_\varphi, \varepsilon_\kappa)^T$ – vector of current deformation model parameters mismatch.

Let us consider an experiment showing the results of implementing proposed optimization. The experiment conducted on synthesized images with correlation function and probability distribution function of intensities closed to Gaussian. The correlation radius of images was 15 pixels. Additionally, images were corrupted with additive Gaussian noise. The signal-to-noise ratio was 20. The similarity deformation model with parameters $h_x = 10$, $h_y = 10$, $\varphi = 20^\circ$, $\kappa = 1.25$ was implemented. Fig. 4 shows the location of suboptimal regions for the 1, 200 and 400 iterations.




a)




b)

Fig.3. Optimization criterion for the different circle radius (a) and different signal-to-noise ratio (b)

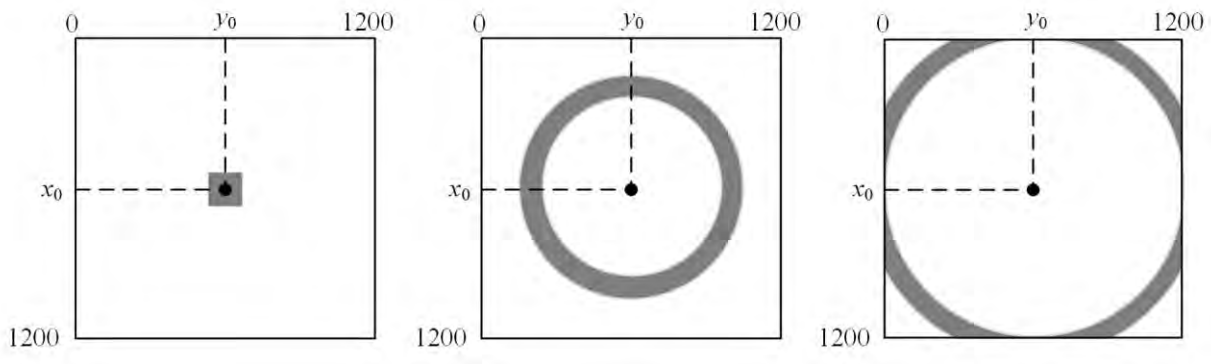



Fig. 4. Location of suboptimal regions of samples for the 1 (a), 200 (b) and 400 (c) iterations

Fig. 5 shows the convergence of mismatch Euclidian distance for this experiment. Here curve 1 – the optimized procedure and curve 2 – not-optimized one.

Experimental results show that optimized procedure on different classes of synthesized and real images increase the convergence rate of registration up to 30 times.

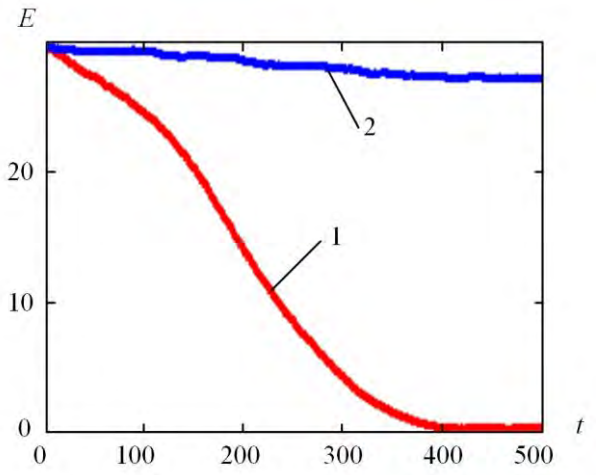



Fig. 5. The convergence of mismatch Euclidian distance

## III. Conclusion

Thus, the optimal mismatch Euclidian distance for stochastic gradient search of registration parameters when mutual information is considered as an objective function is found. It's value depends on the image correlation function, signal-to-noise ration and the circle radius on which the sample points are taken. Experimental results show that optimized procedure on different classes of synthesized and real images increase the convergence rate of registration up to 30 times.

## Acknowledgment

The work was supported by RFBR 13-01-00555-a.

## References

[1] A.A. Goshtasby, "Image registration. Principles, tools and methods", Advances in Computer Vision and Pattern Recognition, 2012, 441 p.

[2] A.G. Tashlinskii, S.V. Voronov, "Specifics of objective functions for recurrent estimation of interframe geometric deformations", The 11th International conference "Pattern Recognition and Image Analysis: New Information Technologies" Conference proceedings, 2013, vol.1, pp. 326-329.

[3] A. Collignon, "Multi-modality medical image registration by maximization of mutual information", PhD thesis, 1998.

[4] P. Viola, W.M. Wiells, "Alignment by maximization of mutual information", International Journal of Computer Vision, 1997, vol. 24, pp. 137-154.

[5] A.G. Tashlinskii, G.L. Safina, S.V. Voronov, "Pseudogradient optimization of objective function in estimation of geometric interframe image deformations", Pattern recognition and image analysis, 2012, vol. 22, №. 2, pp. 386-392.

[6] A.G. Tashlinskii, G.L. Safina, S.V. Voronov, "Optimization of mismatch euclidean distance in evaluating interframe deformations of geometrical images", Pattern recognition and image analysis, 2011, vol. 21, № 2., pp. 335-338.

[7] G.L. Fadeeva, "Optimization of gradient estimation for stochastic gradient image registration", PhD thesis, 2008.

# Parallel Implementation of Roadmap Construction for Mobile Robots using RGB-D Cameras ⋆

**Marco Negrete** * **Jesús Savage** * **Jesús Cruz** * **Jaime Márquez** *

* *Universidad Nacional Autónoma de México*

**Abstract:** This paper describes a method to construct roadmaps for service robots' navigation using vector quantization. Point clouds are acquired from a RGB-D camera and are processed in parallel in a GPU programmed with CUDA. Every point in the cloud is transformed into the canonical horizon in order to obtain the plane that represents the floor. A point is considered to be free space if its Z component is less than a given constant. Vector quantization technique is used to partition the free space into regions and the centroids of such regions become the nodes of a roadmap, that is used by a service robot to navigate. To test the proposed method, several experiments were performed in an indoor environment.

*Keywords:* Mobile robots, vector quantization, clustering, parallel processing

## 1. INTRODUCTION

Service robots must have the capability of navigating in dynamic environments in order to accomplish their purpose: helping humans in everyday domestic tasks (Chen et al., 2014). Navigation in dynamic environments implies the need for a reactive system that allows the robot to avoid obstacles, but, since service robots must execute complex tasks, it is also necessary a world representation for allowing the robot to plan high level tasks. Thus, a service robot's navigation system must be reactive (in real time) to allow a safe navigation and, at the same time, it must be capable of high level task planning.

Roadmaps are useful for mobile robots' navigation in structured environments. If such roadmaps are constructed with a fast enough sampling time and based on information extracted from robot's sensors, they can be used for obstacle avoidance. There are several techniques for roadmap construction when a geometrical representation of the objects in the environment is available, for example, Voronoi diagrams (Latombe, 1991), visibility maps (Lozano-Pérez and Wesley, 1979) or probabilistic roadmap methods (Kavraki et al., 1996). If a geometric representation is not available, it can be obtained by vector quantization (VQ) methods and, based on this representation, it is possible to build a Voronoi diagram.

This paper describes a parallel implementation of the construction of roadmaps using vector quantization techniques. To build the roadmap, we process the point cloud acquired from a Kinect device to separate free and occupied space. Then, both the free and occupied space are clustered. Centroids and sizes of the occupied space clusters are used as a geometrical representation of the objects in the environment. Centroids of the free space are used as nodes of the roadmap. Point clouds are processed in parallel in a GPU programmed with CUDA. We also implemented the whole process in serially, using C++, in order to compare processing times and test the effectiveness of the parallel implementation. The proposed method was tested in an indoor environment in the context of the Robocup @Home tests.

⋆ This work was partly supported by PAPIIT-DGAPA UNAM under Grant IN-107609 and by CONACYT

## 2. THE SERVICE ROBOT JUSTINA

Justina is a service robot built at the Biorobotics Lab of the Engineering School of the National University of Mexico. This robot and its predecessors have been participating in the Robocup@Home league since 2006 performing several task like cleaning up a room, serving drinks and several other tasks that the human beings ask for. It is based on the ViRbot architecture for the operation of mobile robots (Savage et al., 1998). Justina has several sensors: two laser range finders, a Kinect sensor, a stereo camera and a directional microphone. Also, Justina has encoders in each motor (mobile base, head and its two arms). The proposed method uses the Kinect sensor and head encoders. Figure 1 shows the robot Justina and the position of its sensors and actuators.

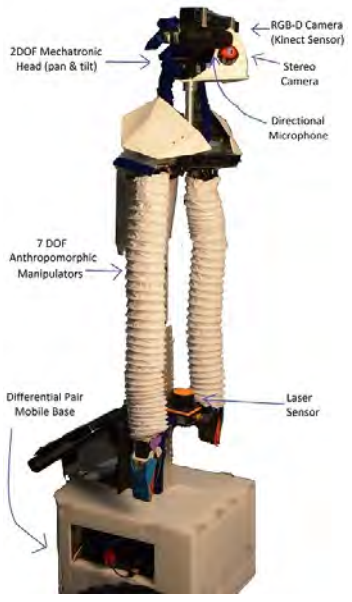



Fig. 1. The service robot Justina

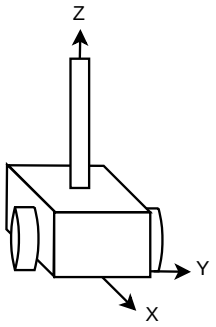



Fig. 2. The robot's frame

## 3. SEPARATION OF FREE AND OCCUPIED SPACE

RGB-D cameras provide information through an RGB image and a point cloud that represents the spatial position of each pixel of the captured image. This research uses only the spatial information which comes as a set $R$ of triplets of the form $r_j = \left(x_{screen_j}, y_{screen_j}, d_j\right)$, where $(x_{screen_j}, y_{screen_j})$ is the pixel location in the image and $d_j$, the distance to the object on the line of sight. Then, the point cloud $S$ of the cartesian positions $s_j = (x_j, y_j, z_j)$ of the objects w.r.t. the Kinect's plane, can be obtained with the transformation

$$s_j = Mr_j \tag{1}$$

with $M$, the matrix of intrinsic parameters of the Kinect camera, provided by the OpenNI libraries (Ope, 2010).

The Kinect sensor is mounted in the robot's mechatronic head. This head has two degrees of freedom: pan and tilt. It is built with Dynamixel servomotors whose encoders allow to measure the head orientation.

To transform the point cloud to the robots frame, the following homogeneous transformation is used:

$$p_j = \begin{bmatrix} cos\phi & 0 & sin\phi & 0 \\ 0 & 1 & 0 & 0 \\ -sin\phi & 0 & cos\phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} cos\theta & -sin\theta & 0 & H_x \\ sin\theta & cos\theta & 0 & H_y \\ 0 & 0 & 1 & H_Z \\ 0 & 0 & 0 & 1 \end{bmatrix} s_j \tag{2}$$

where $s$ is calculated according to (1), and ($\theta$, $\phi$) are the head pan and tilt respectively, with a positive pan when the head is pointing to the left and a positive tilt when head is looking down, and $(H_x, H_y, H_z)$ is the position of the head w.r.t to the robot's frame. Each point of the resulting point cloud $P = p_j$ is expressed w.r.t. the robot's frame, whose position is showed in figure 2.

A point $p_j$ is classified as free space if its $z$ component is less than a constant $K_h$ and, as occupied space, otherwise.

## 4. VECTOR QUANTIZATION

Vector Quantization (Linde et al., 1980) is commonly used for data compression in telecommunications and digital signal processing. In the field of robotics, it is also used to compress data and get a smaller but significative set of data. In this research, we use VQ to cluster the free and occupied space and, based on this clusters, to construct a roadmap.

Given a point cloud $P$, i.e., a set of $N_v$ vectors $p_j = (x_j, y_j, z_j); j = 1, ..., N_v$ that represent the position in space, a set of centroids that represents this vectors is found.

A collection of centroids is called a codebook and it is designed from a long training sequence that is representative of all vectors $p_j$ to be encoded. The codebook is created with the Linde-Buzo-Gray (LGB) algorithm (Linde et al., 1980), as follows:

(1) Initialization: Find an initial coodebook $D_1$, with only one centroid $C_1$ which is obtained by averaging all vectors $p_j$. Let $m = 1$ be the current iteration and $L_m = 1$, the current number of centroids $C_i$ in the codebook $D_m$.
(2) Disturbing centroids: For each centroid $C_i, i = 1, ..., L_m$ in the current codebook $D_m$, obtain two new centroids by adding a disturbance $\pm\psi$ of small magnitude. That is, the new codebook $D_{m+1}$ will contain $L_{m+1} = 2L_m$ new centroids.
(3) Given a codebook $D_m = C_1, ..., C_{L_m}$, assign each vector $p_j$ into the nearest cluster $R_k$, whose corresponding centroid is $C_k$. Determining the nearest cluster is made with some measurement that satisfies the conditions to be a distance function $d_j = d(p_j, C_k)$. In this research we use the Euclidean distance.
(4) For each cluster $R_k$, recompute its centroid $C_k$ by averaging all vectors $p_j$ belonging to $R_k$.
(5) If the difference between the average distance $\overline{d_t} = \frac{1}{N_v} \sum d(p_j, C_k)$, in iteration $t$, between vectors $p_j$ and their corresponding centroids $C_k$, and the previous average distance $\overline{d_{t-1}}$, is greater than a constant, i.e. $\left|\overline{d_t} - \overline{d_{t-1}}\right| > \epsilon$, then go to 3.
(6) If $L_m < L_d$, go to 2.

Where $L_d$ is the desired codebook size (number of regions in the environment) and it is chosen with a tradeoff between computation time limitations for real time operation and the desired precision. In this work, we use $|\psi| = 0.01$, $\epsilon = 0.03$ and $L_d = 64$.

The LGB algorithm is applied to cluster both the free and occupied space obtained as described in the previous section, i.e., a total of 128 centroids is calculated, 64 for the free space and 64 for the occupied space. Figure 3, left and center, shows the resulting clusters.

## 5. ROADMAP CONSTRUCTION

To represent the environment, we consider each centroid of the occupied space as the centroid of a rectangular object of 0.2[m]x0.2[m]. Paths are calculated under this assumption. After the free and occupied spaces are separated and clustered, the roadmap is built following the next algorithm:

Input:

- $N$: Number of nodes in the roadmap (it should be a power of 2)
- $P$: Centroids of the quantized occupied space
- $C$: Centroids of the quantized free space

Output:

- Roadmap G(V, E)
- V: Nodes of the roadmap
- E: Edges of the roadmap

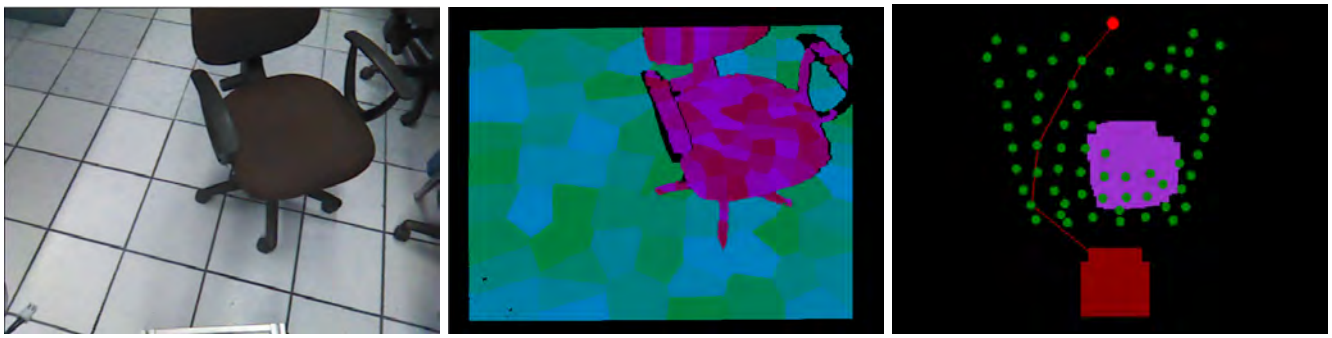


Fig. 3. **Left:** Original RGB image captured by the Kinect. **Center:** Free space clusters are colored in green and occupied space clusters, in purple. The black regions are those points with no depth information. **Right:** Resulting environment representation. Green dots represent the nodes (free space centroids) used to build the roadmap and calculate a path. Purple rectangles represent obstacles and the red lines are the path calculated by Dijkstra algorithm to reach the goal point, also colored in red.

Steps:

(1) $E \leftarrow 0$
(2) $V \leftarrow C$
(3) for all $v \in V$ do
(4) for all $v\prime \in V$ do
(5) if $e(v, v\prime) \notin V$ and $Vis(v, v\prime, p) \neq NV \forall p \in P$
(6) $E \leftarrow E \cup e(v, \prime)$
(7) end if
(8) end for
(9) end for

Where $e(v, v\prime)$ represents the edge between nodes $v$ and $v\prime$ and $Vis(v, v\prime, p)$ is a function that determines if it is possible to reach the node $v$ from node $v\prime$ without crashing with the obstacle whose centroid is $p$. $NV$ means Not Visible and function $Vis$ return this value when node $v$ is not visible from $v\prime$.

Once the roadmap is constructed, a path to the goal point is calculated using the Dijkstra algorithm. Figure 3, right, shows in red lines the resulting path to reach the goal point, also colored in red.

## 6. PARALLEL IMPLEMENTATION

The transformation of the point cloud to the canonical horizon, separation of free and occupied space and clustering were implemented in parallel using CUDA, which is a toolkit designed specifically to develop parallel applications on Nvida GPU's. Calculations of the roadmap edges and the Dijsktra algorithm were implemented in serial.

In this research, we use a GPU Nvida Quadro 2000, which has 192 cores, 1.25 GHz and 1GB RAM. We use the CUDA Toolkit 5.0 for the parallel processing, OpenNI 1.5 to capture data from the Kinect sensor, OpenCV 2.4.8 for basic operations on the image (coloring free and occupied space) and Visual Studio 2010 for the general implementation.

## 7. EXPERIMENTAL RESULTS

Figure 3 shows the results of the whole roadmap construction. The image on the left shows the original image captured with the Kinect device. Since the Kinect is mounted

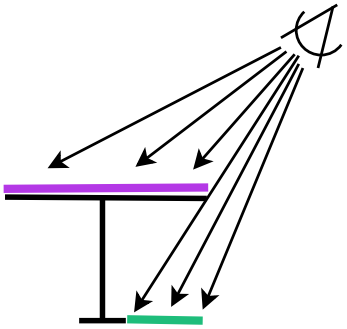


Fig. 4. Schematic of the Kinect's point of view

on the robot's head, this image is in the robot's point of view. The figure on the center shows the free and occupied space clusters. Free space is colored in green and occupied space in purple. Regions in black are those points with no depth information. This image is also in the robot's point of view. Image on the right shows the environment as it is represented by the robot. Since navigation is made only in two dimensions, this image is the 2D projection of the free and occupied space clusters. Nodes of the resulting roadmap are drawn in blue and obstacles (obtained from centroids of the occupied space) are drawn in purple. The red lines are the path calculated by de Dijkstra algorithm to reach the goal point, colored in green.

*Why free space centroids inside occupied space?* Figure 3 shows some green dots inside the purple occupied space. To explain this, see figure 4, which is an schematic of the side view of a table and a Kinect sensor pointing to it. Since the Kinect is not looking from a top view, some infrared rays detect the table surface, represented by the purple line, and some rays can detect the space below the table, represented by a green line. Since the representation of the environment is made only in 2D, the projection of the table and the space below it will result in some nodes inside an occupied space.

Nevertheless, this situation does not affect the roadmap construction, since the function $Vis(v, v\prime, p)$ will determine that nodes inside occupied space are not visible from any other node.

In order to compare the effectiveness of the parallel implementation, the clustering was also implemented in serial.

| Serial Time [s] | Parallel Time [s] |
|---|---|
| 1.341 | 0.078 |
| 1.373 | 0.078 |
| 1.341 | 0.078 |
| 1.388 | 0.077 |
| 1.404 | 0.078 |
| 1.357 | 0.078 |

Table 1. Comparison of processing time for the serial and parallel implementations.

Table 1 shows the times taken to process one video frame, of six frames, both for the parallel and serial clustering. It can be observed that the parallel implementation is significatively faster than the serial one. The mean time for the parallel execution was 0.078 [s] and it was until 18 times faster than the serial one. This allows a safe navigation since the maximum robot speed is 1.0 [m/s].

The proposed method is a part of a more complex navigation system. The roadmap construction is used as a local path planning to reach partial goal points of a global path. The general steps for the robot to reach a goal point are:

(1) A global goal is established by a spoken command such as *Robot go to the living room.*
(2) A global path $P(O_k, N_k, P_g, P_R)$ is calculated by Dijkstra algorithm.
(3) for all nodes $n \in N_k$ do
(4)
- if $NoObstacleInFront$ then try to reach $n$ by potential fields
- else: construct a new roadmap as described in section 5.

where $O_k$ is the set of known obstacles in the environment, generally, static objects such as tables and walls; $N_k$ is the set of nodes to calculate paths considering only the known obstacles; $P_g$ is the global goal point and $P_R$ is the current robot position. The *no obstacle* condition is determined by processing laser readings.

Figure 5 depicts in a very simple manner the Biorobotics Lab at UNAM, where all algorithms are tested. In this figure can be observed the elements above described. Yellow rectangles represent the known obstacles (desks, walls and other static furniture). Blue dots are the known nodes for calculating global paths and colored in red, the current robot position. As mentioned, if an obstacle is detected in front of robot, then a new roadmap is built. In this case, the unexpected obstacle was an office chair, as shown in figure 3. This obstacle is depicted by the little purple rectangles obtained from the occupied space centroids. The new path, colored in red, is calculated considering both the known nodes and those nodes obtained from free space centroids. New roadmaps are built every time an obstacle is detected in front of robot.

The whole navigation system has been tested in the service robot Justina, in the context of the Robocup@Home tests, which involve structured indoor environments as described in Chen et al. (2014).

## 8. CONCLUSIONS

This paper described a method to construct roadmaps for service robots' navigation by clustering the data captured for an RGB-D camera, such as the Kinect sensor.

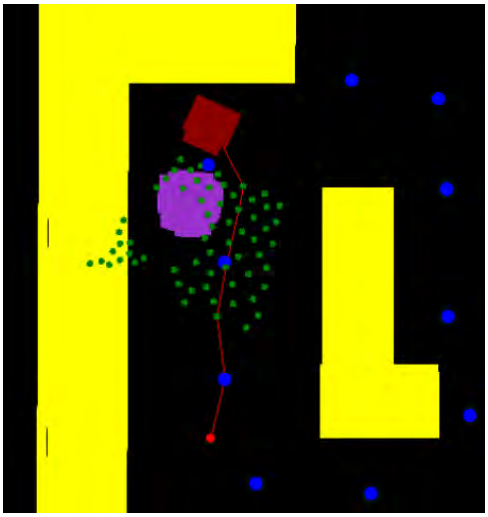


Fig. 5. World Representation

Roadmaps built with this technique have characteristics similar to the ones developed using Voronoi diagrams, but with more flexibility to choose the number and size of the regions in the environment. Commonly, clustering is an expensive process, considering the computation time, and some times it is not suitable to be implemented for obstacle avoidance in robots' navigation. Nevertheless, with the parallel implementation the time taken to process one frame was reduced significatively, allowing the implementation for a safe navigation of a service robot. Finally, the whole process was tested in the service robot Justina in the context of the Robocup @Home tests.

## REFERENCES

(2010). *OpenNI User Guide*. OpenNI organization. URL `http://www.openni.org/documentation`. Last viewed 19-01-2011 11:32.

Chen, K., Holz, D., Rascon, C., Ruiz del Solar, J., Shantia, A., Sugiura, K., Stückler, J., and Wachsmuth, S. (2014). Robocup@home2014: Rules and regulations. `http://www.robocupathome.org/rules/2014_rulebook.pdf`.

Kavraki, L.E., Svestka, P., Latombe, J.C., and Overmars, M.H. (1996). Probabilistic roadmaps for path planning in high-dimensional configuration spaces. *Robotics and Automation, IEEE Transactions on*, 12(4), 566–580.

Latombe, J.C. (1991). *Robot Motion Planning*. Kluwer Academic.

Linde, Y., Buzo, A., and Gray, R.M. (1980). An algorithm for vector quantizer design. *Communications, IEEE Transactions on*, 28(1), 84–95.

Lozano-Pérez, T. and Wesley, M.A. (1979). An algorithm for planning collision-free paths among polyhedral obstacles. *Communications of the ACM*, 22(10), 560–570.

Savage, J., Billinghurst, M., and Holden, A. (1998). The virbot: a virtual reality robot driven with multimodal commands. *Expert Systems with Applications*, 15(3), 413–419.

# PRIAR (Pattern Recognition Image Augmented Resolution) - A TOOL TO COMBINE PATTERN-RECOGNITION WITH SUPER-RESOLUTION

Marco Righi

National Research Council (CNR)
Institute of Information Science
and Technologies (ISTI)
via G. Moruzzi 1, 56124 Pisa, Italy
e-mail: marco.righi@isti.cnr.it

Mario D'Acunto

National Research Council (CNR)
Institute of Information Science
and Technologies (ISTI)
via G. Moruzzi 1, 56124 Pisa, Italy
e-mail: mario.dacunto@isti.cnr.it

Ovidio Salvetti

National Research Council (CNR)
Institute of Information Science
and Technologies (ISTI)
via G. Moruzzi 1, 56124 Pisa, Italy
e-mail: ovidio.salvetti@isti.cnr.it

National Research Council (CNR)
Istituto di Struttura della Materia (ISM)
via Fosso del Cavaliere 100, 00133 Roma, Italy
e-mail: mario.dacunto@ism.cnr.it

***Abstract*—PRIAR (Pattern Recognition Image Augmented Resolution) is a tool that uses the information gained through pattern-recognition to enhance resolution for low quality images, and to allow the end user to explore, recognize and super-resolve low-resolution images. In this paper, we present the basic functionality of the PRIAR algorithms, that have been implemented with Matlab classes and functions. These classes can be easily combined, which has the advantage that one can adapt the simulation programs to various applications. Here, we consider the application of PRIAR to image reproducing cells recorded with scanning probe microscopy.**

## I. Introduction

Image analysis is an important field of computer science application especially to extract meaningful information directly form images and to compute big data. Nowadays a laboratory can automatize or semi-automatize the data acquisition using modern device. The automation generates a high amount of data to be analyzed that requires the aid of the computer and suitable software. PRIAR is a tool born to aid the images analysis, in particular, to enhance low-resolution single frame 2D or 3D input images, recognize specific part of the image and match or substitute these parts with their appropriate models. This enhanced image process is guided by a pattern-recognition algorithm, that takes advantage from the super-resolved image and the knowledge of the model that it explores. In this paper, we describe, for the first time, the basic functionality of the PRIAR algorithm.

December 1, 2014

## II. Main steps of PRIAR

This algorithm combines single-frame super-resolution and pattern-recognition methods. Here, we describe the main steps of PRAR, the steps are explained in the following paragraphs. We have focused the application of PRIAR to images recorded with scanning probe microscopy applied to biological systems (animal cells). In these microscopic techniques, an image represents the object as a matrix pixels $(x, y)$ and assign to any pixel a z value with a gray scale. Nevertheless, the algorithm has been designed to be a platform applicable to a wide range of imaging techniques.

As first, it classifies the image. Hence, the information acquired during the image classification is used during the reconstruction of the identified object. In fact, it permits to use the appropriate model to build the object in the image space.

The second step regards the improved definition of image details, previously classified or simply detected, by increasing spatial (x-y) and depth (z) resolution. During this step various super-resolution methods are generally applied in order to obtain the best image and the best trade-off between image resolution and calculating time [1] [2] [3] [4] [5] [6] [7].

In turn, the third elaboration gets as input the super-resolved image and searches an object according to the model obtained during the first two previous steps [8] [9]. According to the recognized pattern and the model built consequently, the image is improved by increasing the spatial resolution, color depth and inserting objects that have been identified morphological properties. The algorithm can be summarized as follows:

Algorithm 1: PRIAR

```
Function [path enhanced_image] =
    PRIAR(input_image)
class=classification(input_image);
switch class
  case:'grating'
    segmentation_type='otsu';
  case:'cell'
    segmentation_type='edge_discover';
end
sr_image:=blind_sr(input_image)
initial_path=
  explore(seed,
    super_resolved_image);
edge_discovered=
  edge_discover(sr_image,
    segmentation_type);
path:=
  explore_extend(initial_path,
    segmented_image,
    edge_discovered);
enhanced_image:=
  build_model(sr_image,class,path)
return [path enhanced_image];
end
```

## III. Classifier

The PRIAR classifiers distinguish two kinds of images: the first one represents a regular pattern (that we have named *grating*); the second one represents the object (in our case an animal cell). These two kinds of images represent a benchmark for our model to be built. In fact, the objective of this tool is to classify the kind of image in order to apply the appropriate pattern-recognition method and merge the information acquired by blind super-resolution and pattern recognition to enhance the image, substituting the recognized image components with their models (that can also be provided of meta-attribute that helps to integrate characterization in its contest).

PRIAR is currently using a linear classifier. The classifier performs a segmentation of the image in order to identify each region [8] [9]. It is calculated for each region a set of features $f_i$: mean value along the Z axis, orientation of the segmented region [10] [11] and the distribution of the pixel-color in the segmented image [10] [11]. The algorithm approximates the features as independent-features so the probability that an image is in class $C_i$ can be expressed in Bayesian form:

$$p\left(C_i|f_1,\ldots,f_i\right) = \frac{p\left(C_i\right)\cdot p\left(f_1,\ldots,f_i|S_i\right)}{p\left(f_1,\ldots,f_i\right)} \tag{1}$$

## IV. The super-resolution method

PRIAR focuses on the problem of enriching the information that can be collected from a single input image. Firstly, in a $z = f(x,y)$ image, it is necessary to improve the spatial resolution and the color depth. In other words, it is necessary increase the image resolution and calculate a correct color value for each pixel. This can be made extending the Kim Kwon algorithm [1]. The single image super resolution algorithm uses only the information contained in the input image. In order to obtain the final super-resolved image it is necessary to calculate some intermediate images. The input image is interpolated, the result of the interpolation is an X-image that still contains low-frequency information (generally, it is a blurred image). The high frequency information is calculated by combining the Laplacian of the X-image and other estimators that are calculated on different local observation of input image (each local observation get in output different partial information). A single image with enhanced resolution is finally obtained as a convex combination for each pixel of the set of candidate pixels based on their estimated likelihood. To improve the visual quality, the results are post-processed based on an appropriate estimated prior [2]. To calculate the high-frequency it is used a regression-based method that convex to the super resolved image. For each location $(x,y)$ it is generated a set of patch $Z_i(x,y)$ and it is calculated a vector of differences $d_1(x,y),,d_N(x,y)$ between the output (that is not yet calculated) and each candidates.
Each pixel is calculated using equation 2:

$$H\left(x,y\right) = \sum_{\{i=1,\ldots,N\}} p_i\left(x,y\right)\cdot z_i\left(x,y\right) \tag{2}$$

where $p_i(x,y)$ is the weight given to a certain patch.

## V. Adopted pattern-recognition method

The algorithm goal is recognize a specific area $I_0(x,y) \subseteq I(x,y)$ of the image I. Our algorithm accepts as input a coordinate $(x_0,y_0) \subseteq I_0(x,y)$ or a path that provides to the pattern-recognition method more information to match the structure. The algorithm uses one of the described input coordinates and expand the region using a method based on the gradient value pixel of the neighborhood. 3) All the pixel in a neighborhood region are correlated and a local estimate of the correlation surface is made Assuming $I_0(x,y)$ as the target image of size $I_a \times I_b$, and the $W(x,y)$ is a windowing function (of the same size of $I_0(x,y)$ or less) containing the object of interest, whose gradient goes to zero at the edges (a sort of boundary condition), then the normalized squared correlation is given by (for sake of simplicity in continuous form):

$$C^2\left(\rho,\theta\right) = \frac{\left[\iint I\left(x,y\right)\cdot\left(I_0\cdot W\right)\left(x-\rho,y-\theta\right)dx\,dy\right]^2}{\iint I^2\left(x,y\right)\cdot W\left(x-\rho,y-\theta\right)dx\,dy} \tag{3}$$

with the assumption that $\iint (I^2W)(x,y)dx\,dy$. Taking the derivative of the normalized correlation we find that:

$$\nabla C^2(\rho,\theta) = -2\frac{\iint I(x,y)\cdot(I_0\cdot W)(x-\rho,y-\theta)\,dx\,dy}{\iint I^2(x,y)\cdot W(x-\rho,y-\theta)\,dx\,dy}\cdot \frac{\iint I(x,y)\cdot(I_0\cdot W)(x-\rho,y-\theta)\,dx\,dy}{\iint I^2(x,y)\cdot W(x-\rho,y-\theta)\,dx\,dy} + \frac{\left[\iint I(x,y)\cdot(I_0\cdot W)(x-\rho,y-\theta)\,dx\,dy\right]^2}{\iint I^2(x,y)\cdot W(x-\rho,y-\theta)\,dx\,dy}\cdot \frac{\left[\iint I(x,y)\cdot(I_0\cdot W)(x-\rho,y-\theta)\,dx\,dy\right]^2}{\iint I^2(x,y)\cdot W(x-\rho,y-\theta)\,dx\,dy} \quad (4)$$

that, at the least, leaves four term to be calculated.
These equations can be transformed to the discrete domain, reducing the calculation of the correlation gradient at a central pixel p and the correspondent 8 neighbours [2] [12].

## VI. Adopted pattern-recognition method

During the last step of the algorithm PRIAR uses the computed information by pattern-recognition and classifying to super-resolve the image.
The information of the identified class $C_i$ and the information acquired by patter-recognition algorithm are merged to create a new image where it is reconstructed the identified area. The a-priori knowledge that is contained in the class can be of two different kinds: the first one is a geometrical description and it is used to super-resolve the image, the second one is used to mark the features of the classified object that are not representable by graphics.
In our case the geometry of the objects is like a pipe so we simply minimizing the difference between the pipe and the data presents on the image itself.
In turn, the GUI (Graphic user interface) guides the end-user to use the features of the tool. Fig. 1 shows the main window of the GUI. It is possible use the GUI to manage basic information to choose the image to elaborate, to choose the directory that contains the results, to choose the area of interest, to choose the method used during the super-resolution step, to use the internal classifier or manually classify the image, to choose a single point where the pattern-recognition starts or track a path that guides the pattern-recognition process.

The tool has been developed to be remotely used by command line too. In fact, the core of the program can get the structure containing all parameters or a file that containing the variables to use during the computation. PRIAR tool shows intermediate results and the finale result in the form of images. The images are still saved in the chosen directory. Figure 2 shows the input images, figure 3 the tracked path. Figure 4 shows a rendering 3D of the reconstructed area.

## VII. Conclusion

In this paper, we have presented the features and functionality of the PRIAR algorithm. PRIAR is a tool designed to aid the images analysis, in particular, the improve low-resolution single frame 2D or 3D input images, recognize specific part

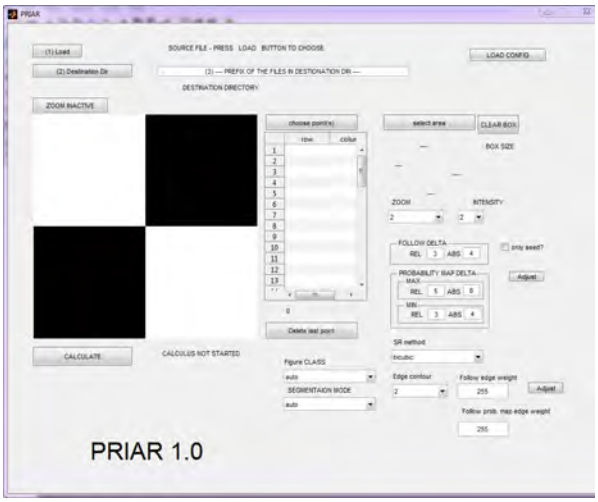



Fig. 1: a screenshot of the PRIAR 1.0 GUI

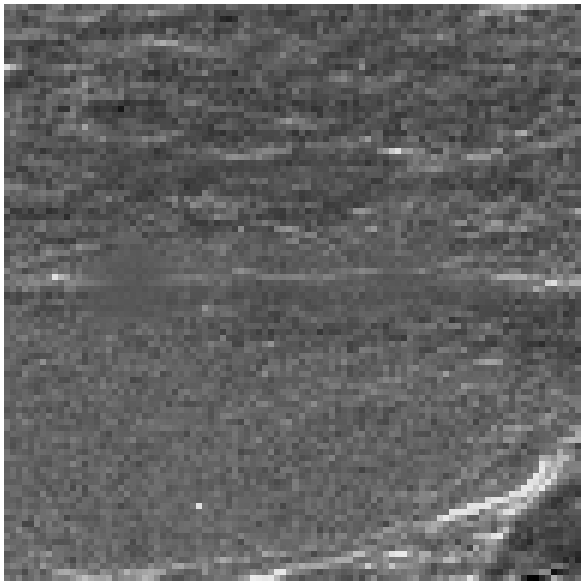


Fig. 2: an input image representing a portion of mesenchimal cell cytoskeleton

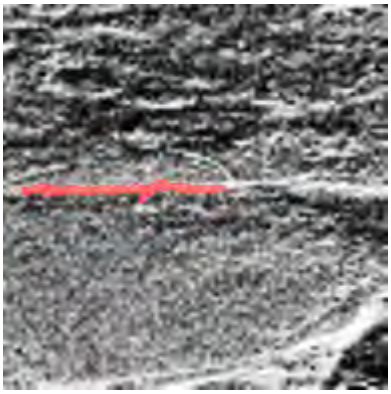


Fig. 3: the same area as in figure 2 with a recognized filament

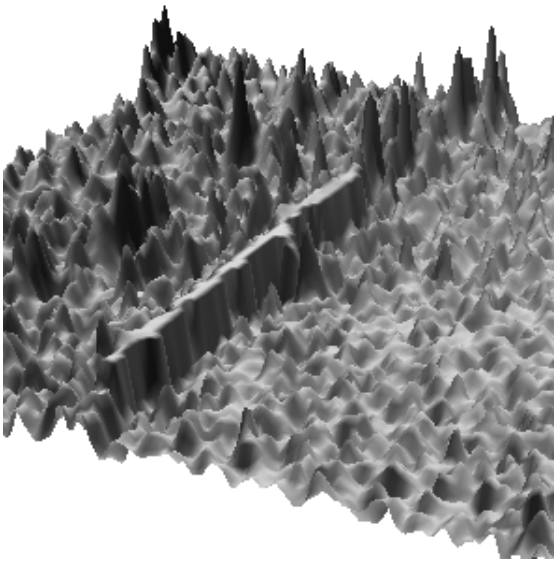


Fig. 4: a 3D visual rendering of figure 3

of the image and match or substitute these parts with their appropriate models. The algorithm combines single-frame super-resolution and pattern-recognition methods. PRIAR works as follows:

- classification and identification of specific objects from single frame images recorded with a microscope.
- Reconstruction of the identified object.
- Super-resolution of single identified objects according to the model obtained during the first two previous steps.

Figures 2-3 summarize the identification and improved resolution of cell components, (filaments, microtubules) of animal cells and correspondent 3D reconstruction.

## ACKNOWLEDGMENT

The authors thank Serena Danti, University of Pisa, for the preparation of mesenchimal cell samples.

## REFERENCES

[1] K. I. Kim and Y. Kwon, *Example-based learning for single-image super-resolution and JPEG artifact removal*, 3rd ed. Technical Report 173, 2008.

[2] M. D'acunto et al., *A methodological approach for combining super-resolution and pattern-recognition to image identification Pattern Recognition and Image Analysis*, N. 24, Vol 2, pp. 209-217, 2014.

[3] M. Sonka et al., *Image Processing, Analysis, and Machine Vision*, 3rd ed. Thomson-Engineering, 2008.

[4] I. Leena and T. Pekka, *Distance and Nearest Neighbor Transforms of Gray-Level Surfaces Using Priority Pixel Queue Algorithm* Springer Berlin Heidelberg, pp. 308-315, 2005.

[5] P. Bourke, *Bicubic Interpolation for Image Scaling*, 2001.

[6] P. Getreuer, *Linear Methods for Image Interpolation*, Image Processing On Line, 2011

[7] E. Ardizzone et al., *Fuzzy-based kernel regression approaches for free form deformation and elastic registration of medical images*, Biomedical Engineering, pp. 347-368, 2009.

[8] M. Sezgin and B. Sankur, *Survey over image thresholding techniques and quantitative performance evaluation*, Journal of Electronic Imaging Vol. 13, N. 1, pp. 146165, 2004.

[9] N. Otsu, *A threshold selection method from gray-level histograms*. IEEE Transactions on Systems, Man and Cybernetics, Vol. 9 N. 1, pp. 6266, 1979.

[10] Rafael C. Gonzalez et al., *Digital Image Processing Using MATLAB*, 2nd ed. USA: Prentice-Hall, 2010.

[11] R. C. Gonzalez and Richard E. Woods, *Digital Image Processing*, 3rd ed. USA:Prentice-Hall, 2006.

[12] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4th ed. USA:Academic Press, 2008.

# Problems of an Image Reducing to a Recognizable Representation

Igor Gurevich and Vera Yashina
Mathematical and Applied Problems of Image Analysis
Dorodnicyn Computing Center of the Russian Academy of Sciences
Moscow, Russian Federation
igoourevi@ccas.ru, werayashina@gmail.com

***Abstract*— The presentation is devoted to the research of mathematical fundamentals for image analysis and recognition procedures being conducted currently in the Dorodnicyn Computing Centre of the Russian Academy of Sciences, Moscow, Russian Federation. The paper presents and discusses the main results obtained using the descriptive approach to analyzing and understanding images when solving fundamental problems of the formalization and systematization of the methods and forms of representing information in the problems of the analysis, recognition, and understanding of images. In particular, that arise in connection with the automation of information extraction from images in order to make intelligent decisions (diagnosis, prediction, detection, evaluation, and identification of patterns). The final goal of this research is automated image mining: a) automated design, test and adaptation of techniques and algorithms for image recognition, estimation and understanding; b) automated selection of techniques and algorithms for image recognition, estimation and understanding; c) automated testing of the raw data quality and suitability for solving the image recognition problem.**

***Keywords—algebraic approach, decpriptive approah, image analysis***

## I. INTRODUCTION

The automation of processing, analyzing, evaluating, and understanding of information provided in the form of images is one of the critical breakthrough problems of theoretical computer science. The image is one of the main means of representing and transmitting information needed to automate intelligent decision making in a variety of application domains.

To date, in the analysis and evaluation of images, there is extensive experience in the application of mathematical methods from different branches of mathematics, computer science, and physics, in particular algebra, geometry, discrete mathematics, mathematical logic, probability theory, mathematical statistics, mathematical analysis, mathematical theory of pattern recognition, digital signal processing, and optics.

On the other hand, the variety of methods used does not replace the need for some regular basis for ordering and selecting appropriate methods of image analysis, a uniform representation of the processed data (images) that meet standard requirements of pattern recognition algorithms to the source data, the construction of mathematical models of images focused on the identification problem, and the general availability of a universal language for the uniform description of images and their transformations

This paper presents the main results on the formalization and systematization of methods and forms of information representation in problems of analysis, recognition, and understanding of images. We have summarized the development of a descriptive approach (DA) to analyzing and understanding images formulated by I.Gurevich [3, 8]. This is a direction of research concerning the formalization and representation of images. Recall that DA is a specialization of the algebraic approach of Yu.Zhuravlev [18] to the case of the representation of information in the form of images.

Axiomatics and formal structures of the DA provide methods and tools for presenting and describing images for their subsequent analysis and evaluation. The theoretical basis of the research is the DA; general algebraic methods; and methods of the mathematical theories of image processing, image analysis, and pattern recognition.

It is established that the overall success and effectiveness of the analysis and evaluation of information provided in the form of images are determined by the possibilities of reducing images to a form suitable for recognition (RIFR).

RIFR processes are crucial for solving applied problems of image analysis and, in particular, to make intelligent decisions based on information extraction from images. The DA provides the ability to solve both problems associated with the construction of formal descriptions of images as objects of recognition and problems of synthesis of procedures of pattern recognition and image understanding. The operational approach to characterizing images requires that processes of analyzing and evaluating information provided in the form of images (the trajectory of problem solving) as a whole could be viewed as a sequence/combination of transformations and computing of a set of interim and final (defining the solution) evaluations. These transformations are defined on the equivalence classes of images and their representations. The latter are defined descriptively, i.e., using a base set of prototypes and corresponding generative transformations that

are functionally complete with respect to the equivalence class of admissible transformations.

Now we outline the goals of theoretical development in the framework of the Descriptive Approach (and image analysis algebraization) ("What for") and necessary steps to finalize the Descriptive Approach ("What to Do or What to be Done") and the global problem of an image reduction to a recognizable form.

## II. DESCRIPTIVE APPROACH TO IMAGE ANALYSIS AND UNDERSTANDING

This section contains a brief description of the principal features of the DA needed to understand the meaning of the introduction of the conceptual apparatus and schemes of RIFR proposed to formalize and systematize the methods and forms of representation of images.

The automated extraction of information from images includes (1) automating the development, testing, and adaptation of methods and algorithms for the analysis and evaluation of images; (2) the automation of the selection of methods and algorithms for analyzing and evaluating images; (3) the automation of the evaluation of quality and adequacy of the initial data for solving the problem of image recognition; and (4) the development of standard technological schemes for detecting, assessing, understanding, and retrieving images.

The automation of information extraction from images requires complex use of all the features of the mathematical apparatus used or potentially suitable for use in determining transformations of information provided in the form of images, namely in problems of processing, analysis, recognition, and understanding of images.

Experience in the development of the mathematical theory of image analysis and its use to solve applied problems shows that, when working with images, it is necessary to solve problems that arise in connection with the three basic issues of image analysis, i.e., (1) the description (modeling) of images; (2) the development, exploration, and optimization of the selection of mathematical methods and tools for information processing in the analysis of images; and (3) the hardware and software implementation of the mathematical methods of image analysis.

The main purpose of the DA is to structure and standardize a variety of methods, processes, and concepts used in the analysis and recognition of images.

The DA is proposed and developed as a conceptual and logical basis of the extraction of information from images. This includes the following basic tools of analysis and recognition of images: a set of methods of analysis and recognition of images, RIFR techniques, conceptual system of analysis and recognition image, descriptive image models (DIM) classes, the descriptive image algebra (DIA) language, statement of problems of analysis and recognition of images, and the basic model of image recognition.

The main areas of research within the DA are (1) the creation of axiomatics of analysis and recognition of images, (2) the development and implementation of a common language to describe the processes of analysis and recognition of images (the study of DIA), and (3) the introduction of formal systems based on some regular structures to determine the processes of analysis and recognition of images (see [3, 4]).

Mathematical foundations of the DA are as follows: (1) the algebraization of the extraction of information from images, (2) the specialization of the Zhuravlev algebra [18] to the case of representation of recognition source data in the form of images, (3) a standard language for describing the procedures of the analysis and recognition of images (DIA) [8], (4) the mathematical formulation of the problem of image recognition, (5) mathematical theories of image analysis and pattern recognition, and (6) a model of the process for solving a standard problem of image recognition.

The main objects and means of the DA are as follows: (1) images; (2) a universal language (DIA); (3) two types of descriptive models, i.e., (a) an image model and (b) a model for solving procedures of problems of image recognition and their implementation; (4) descriptive algebraic schemes of image representation (DASIR); and (5) multimodel and multiaspect representations of images, which are based on generating descriptive trees (GDT) [9].

The basic methodological principles of the DA are as follows: (1) the algebraization of the image analysis, (2) the standardization of the representation of problems of analysis and recognition of images, (3) the conceptualization and formalization of phases through which the image passes during transformation while the recognition problem is solved, (4) the classification and specification of admissible models of images (DIM), (5) RIFR, (6) the use of the standard algebraic language of DIA for describing models of images and procedures for their construction and transformation, (7) the combination of algorithms in the multialgorithmic schemes, (8) the use of multimodel and multiaspect representations of images, (9) the construction and use of a basic model of the solution process for the standard problem of image recognition, and (10) the definition and use of nonclassical mathematical theory for the recognition of new formulations of problems of analyzing and recognizing images.

Note that the construction and use of mathematical and simulation models of studied objects and procedures used for their transformation is the accepted method of standardization in the applied mathematics and computer science.

The creation of the DA was significantly influenced by the following basic theories of pattern recognition: (1) the algebraic approach to pattern recognition of Zhuravlev[18] and their algorithmic algebra and (2) the theory of images of Grenander [1, 2], in particular algebraic methods for the representation of source data in image recognition problems developed in it.

As was already noted, in the DA, it is proposed to carry out the algebraization of the analysis and recognition of images using DIA. DIA was developed from studies in the field of the algebraization of pattern recognition and image analysis carried out since the 1970s. The creation of a new algebra was directly influenced by algorithms of Zhuravlev [18] and the research of

Sternberg [17] and Ritter [15, 16], which identified classic versions of image algebras.

A more detailed description of methods and tools of the DA obtained in the development of its results can be found in [2-10].

## III. What to Do or What to be Done. Basic Steps

The critical points of an image analysis problem solution are: 1) precise setting of a problem; 2) correct and "computable" representation of raw and processed data for each algorithm at each stage of processing; 3) automated selection of an algorithm: a) decomposition of the solution process for main stages; b) indication points of potential improvement of the solution ("branching points"); c) collection and application of problem solving experience; d) selection for each problem solution stage of basic algorithms, basic operations and basic models (operands); e) classification of the basic elements; 4) performance evaluation at each step of processing and of the solution: a) analysis, estimation and utilization of the raw data specificity; b).diversification of mathematical tools used for performance evaluation; c) reduction of raw data to the real requirements of the selected algorithms.

Basic steps of the development of image analysis theory are: a) mathematical settings of an image recognition problem (step 1); b) image formalization space and descriptive image models (step 2); c) generating descriptive trees and multimodel representation of images (step 3); d) image equivalence (step 4); e) image metrics (step 5); f) descriptive image algebras (step 6).

### *A. Mathematical settings of an image recognition problem*

DONE:

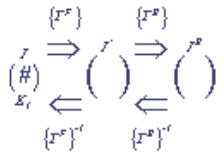
*1) Descriptive Model of Image Recognition Procedures*

*2) Mathematical Setting of an Image Recognition Problem. Image Equivalence Case.*

Image analysis and recognition deal with properties of the object (scene) shown and deformations associated with the way and procedure of obtaining the image. In this case, to formalize image processing, we need to specify three sets (models) of images, on which we postulate the existence of classes of equivalence and sets of admissible transformations given on the classes of equivalence [12, 13]. Introducing classes of equivalence on the sets of image models, we accept that any image possesses some regularity or a mix of regularities of different types. Under this assumption, analysis and recognition problem is reduced to making a difference between images that preserve their own regularity and images, the regularity of which can be broken.

Figure 1 shows the descriptive model of the image recognition problem.

Here, $\{J\}$ is the set of ideal images, $\{J^*\}$ is the set of observable images, $\{J^R\}$ is the set of images obtained as a result of solving the recognition problem, $\{T^F\}$ is the set of admissible transformations to form the image, $\{T^R\}$ is the set of admissible transformations to recognize the image, and $\{K_i\}$ are classes of equivalence.



$$J, J^*, J^R; T^F, T^R, T^R : J \Rightarrow J^*, T^R : J^* \Rightarrow J^R; \{J\} = \bigcup_j K_j.$$

*Fig 1. Decriptive model of the image recognition problem.*

Let $J$ be some true image of the object involved. We can consider processes of obtaining, forming, discretization, etc. (all procedures that make it possible to work with the image) as if the true image were transferred via the noisy channel. As a result, we analyze some real (observable) image $J^*$ rather than the true image. This real image is to be classified in the course of analysis, i.e., we should determine the prototype in the true class of equivalence or find the regularity (regularities) of the given type $J^R$ on the observable image $J^*$. Thus, we can specify the sets $\{J\}$, $\{J^*\}$, and $\{J^R\}$ and transformations to form ($T^F$) and recognize ($T^A$) the image

$$T^F: J \to J^*, \quad (1)$$

$$T^A: J^* \to J^R. \quad (2)$$

To perform image recognition, we need to give algebraic systems of transformations $\{T^F\}$ and $\{T^A\}$ on classes of equivalence of the set $\{J\}$ and apply them to observable images $J^*$ to perform the backward analysis, i.e., classify images according to the nature of their regularity (restore true images, i.e., indicate classes of equivalence they belong to), and the forward analysis, i.e., search the image $J^*$ for regularities of the certain type $J^R$ and localize them.

Stating the analysis problem in such a way, we can give the class of image processing procedures, analysis process of which is of fixed structure, with interpretation (particular implementation) depending on the purposes and type of analysis. There are following main stages of analysis.

Now three mathematical statements of image recognition problems are considered. The first one Z is introduced by Yu. Zhuravlev [18].

When solving real image recognition problems, we deal with the images of objects rather than with the objects themselves. Therefore, we will assume that the whole set of images is somehow divided into equivalence classes. We also assume that there is correspondence between the equivalence classes of images and the objects; however, in the future, we will not mention objects in the statement of the recognition problem. Taking into account the concept of equivalence of images introduced above, we can formulate the image recognition problem as follows.

The difference between problems $Z^2$ and $Z^1$ is that each equivalence class in problem $Z^2$ is replaced by a unique image—a representative of the class—with the number $n_i$ , $1 \leq n_i \leq p_i$, where $i$ is the number of the equivalence class. This replacement is performed by introducing the concept of an admissible transformation.

Problem $Z^1$ differs from Z by the fact that it explicitly uses equivalence classes of images. To reduce the image recognition problem $Z^1$ to the standard recognition problem Z, one should pass from the classification of a group of objects to the classification of a single object. Under certain constraints on admissible transformations, problem $Z^2$, which differs from $Z^1$ in that it contains admissible transformations that do not take the image beyond the equivalence class, allows one to handle a single image for each equivalence class—a representative of this equivalence class.

TO BE DONE:

1) *establishing of interrelations and mutual correspondence between image recognition problem classes and image equivalence classes;*

2) *new mathematical settings of an image recognition problem connected with image equivalency;*

3) *new mathematical settings of an image recognition problem connected with an image multimodel representation and image image data fusion.*

*B. Iimage formalization space and descriptive image models*

DONE:

*1) the conceptualization of a system of concepts that describe the initial information (images) in recognition problems has been carried out;*

*2) descriptive models of images focused on the recognition problem have been defined;*

*3) the image formalization space has been introduced, the elements of which include different forms (states, phases) of representing the image transformed from the original form into the recognizable one, i.e., into the image model;*

*4) the basic axioms of the descriptive approach were introduced.*

Image formalization space (IFS) is the space including sets of an image “states” and sets of image transforming schema for formalization and systematization of techniques and forms of information representations in image analysis, recognition and understanding problems. More detailed description of IFS [14] includes: a) construction of algorithmic schema generating phase trajectories for solving image analysis and recognition problems; b) DIM - mathematical objects providing representation in a form acceptable for a recognition algorithm of information carried by an image and by an image legend (context); c) multiple DIM and multi-aspect image representations; d) topological properties of the Image Formalization Space.

Recall that, in the DA, the processes of analyzing and evaluating the information presented in the form of images are considered as sequences of the transformations and calculations of a set of intermediate and final (defining the solution) evaluations. These estimates are essential characteristics of representations of the source image obtained at each stage of RISR. Final estimates are used in the final stage of solving the problem of recognition/classification of the source image in the application of algorithms for the recognition/classification of the image model created by RIFR.

The descriptive algebraic scheme of the image representation (DASIR), which is a formal scheme designed to produce a standardized formal description of surfaces, point configurations, shapes that form **the image**, and the relations between them, is recorded using the DIA.

DASIR reflect sequential and/or parallel use of transformations from the set of transformations to the initial information from the space of initial data. In [12], there is an example of constructing DASIR to solve the problem of the morphological analysis of blood cells. All steps of the algorithmic scheme for training recognition algorithms for problems of analyzing cytological preparations and classifying the new image using three diagnoses based on a recognition algorithm with adjusted parameters were defined and described using the DIA with one ring.

In the DA, three classes of admissible transformations of images are considered [10], i.e., procedural transformations, parametric transformations, and generative transformations. The basic classes of transformations of images (procedural, parametric, and generating) are defined, as well as related concepts of the structuring element, which generates rules and correct generative transformation.

All transformations for processing and analyzing images are conducted using the DIA record on the transformations of images. It makes it possible to vary the methods for solving the subproblem using different operations of the image analysis of fixed DIA and keeping the whole scheme of the technology of RISF and the extraction of information from images unchanged.

In order to apply of pattern-recognition algorithms to the created formal descriptions of images, it is necessary to implement the created schemes (to set specific transformations from the fixed DAI and parameters of transformations selected in the schemes) and apply them the initial information, i.e., to create models of images.

An example of the DASIR implementation can be found in [10], i.e., at each step of the algorithmic scheme the created DASIR was concretized by the selection of a transformation that belongs to a specific DAI that describes the step.

In the general case, we can say that the use of the convolution of structuring elements and admissible transformations of the image to the initial information about the image leads to the transformation of the initial information

in the image model. Specific allowable image transformations and specific methods of applying them to the initial information are selected based on the set problem of the analysis and recognition of images.

Axiomatization of algebraic image analysis constitutes a base for unification of image analysis algorithms representations and image models representations. The axioms define properties and structure of the Image Formalization Space (IFS).

It was shown that all image representations and procedures of RFSR form a topological space (IFS). The main properties of this space, as well as the conceptual basis of the synthesis of image models, are defined by the following axioms that constitute basic provisions of the DA.

Image models are the results of RIFR (taking into account all the information about the image). On the set of image models, basic DIA are introduced on image models of three classes in accordance with operations used for their construction. Note that these models are descriptive image models (DIM).

TO BE DONE:

1) *Creation of image models catalogue*
2) *Selection and study of basic operations on image models for different types of image models (including construction of bases of operations)*
3) *Use of information properties of images in image models*
4) *Study of multimodel representations of images.*

*C. Generating descriptive trees and multimodel representation of images*

DONE:

*Generating Descriptive Tree (GDT) - a new data structure for generation plural models of an image is introduced*

The introduction of axiomatics of DA and definition of three classes of DIM has led to the introduction of a new mathematical object for structuring representations of images and generation of image models.

Three types of appropriate conversions generating rules and a source image are necessary for constructing the three classes of representations of images (procedural, parametrical, and generating representations of images). The source image is described by means of a set of its implementations and by means of context-sensitive and semantic information.

According to the introduced axiomatics and definitions of various classes of representations of images, in this way, for merging and combination of various properties of image models, it is necessary to introduce the following hierarchies: the hierarchy of possible implementations of images; the hierarchies of semantic and context-sensitive information in images; the hierarchies of parametrical, procedural, and generating conversions; and hierarchies of generating rules. It is suggested to implement such structures in the form of special trees.

Specialization of the concept of a tree on the whole is related to specialization of nodes of a tree. As nodes we will select objects, operations, or rules of image analysis tasks used to construct different image models. Such nodes are called GDT descriptors. The definitions of parent, calculated, fixed, objective, and abstract GDT descriptors have been introduced, but will not be dwelt on in this work.

**Definition 1 [10].** The generating descriptive tree (GDT) is the structure intended for classification and automated generation of image models and it possesses the following properties:(1) GDT descriptors are GDT nodes; (2) Every GDT combines the descriptors of one type; that is, GDTs represent the same type of properties of an image; (3) Each GDT element can be united with another element to generate new partial multiaspect image models; (4) Descriptors are linked among themselves by parent–daughter relationships; (5) Each descriptor has a relationship with a unique parent descriptor and can have some links with derived descriptors. If the descriptor has no parent, it is called a radical GDT. If the descriptor has no derived descriptors, it is called a leaf.

Note that parametrical GDTs are GDTs intended for classification and automation of the generation of parametrical image models. A parametrical GDT, thus, contains GDT descriptors describing the properties of parametrical conversions, leading to an evaluation of features of images. A procedural GDT is a GDT intended for classification and automation of the generation of procedural image models. A procedural GDT, thus, contains the GDT descriptors describing the properties of procedural conversions.

TO BE DONE:

*1) to define and to specify GDT;*
*2) to set up image recognition problem using GDT;*
*3) to define descriptive image algebra using GDT;*
*4) to construct a descriptive model of image recognition procedures based on GDT using;*
*5) to select image feature sets for construction of P-GDT;*
*6) to select image transform sets for construction of T-GDT;*
*7) to define and study of criteria for selection of GDT-primitives.*

*D. Image equivalence*

DONE:

*There were introduced several types of image equivalence: image equivalence based on the groups of transformations; image equivalence directed at the image recognition task; image equivalence with respect to a metric.*

We consider the problem of searching for a correct algorithm for the image recognition problem. We consider various methods for defining the equivalence of images, namely, equivalence on the basis of transformation groups, equivalence oriented to a special statement of the image

recognition problem, and equivalence with respect to metric. In the case of definition of equivalence based on transformation groups, we construct examples of equivalence classes. It is shown that the concept of equivalence is one of the key concepts in image recognition theory. We study the relationship between equivalence and invariance of images.

Using the introduced concept of equivalence of images, we modify the standard mathematical statement of the image recognition problem and formulate an image recognition problem in terms of equivalence classes. We prove that, under certain constraints on the image transformations, the problem of image recognition in the standard statement can be reduced to an abridged problem for which there exists a correct algorithm within the algebraic closure of the class of recognition algorithms for calculating estimates (ACEs).

TO BE DONE:

1) *to study image equivalence based on information properties of the image;*
2) *to define and construct image equivalence classes using template (generative) images and transform groups;*
3) *to establish and to study links between image equivalence and image invariance;*
4) *to establish and to study links between image equivalence and appropriate types of image models;*
5) *to establish and to study links between image equivalence classes and sets of basic image transforms.*

*E. Image metrics*

It is an open problem.

TO BE DONE:

1) *to study, to classify, to define competence domains of pattern recognition and image analysis metrics;*
2) *to select workable pattern recognition and image analysis metrics;*
3) *to construct and to study new image analysis-oriented metrics;*
4) *to define an optimal image recognition-oriented metric;*
5) *to construct new image recognition algorithms on the base of metrics generating specific image equivalence classes.*

*F. Descriptive image algebras*

DONE:

1) *Descriptive Image Algebras (DIA) with a single ring were defined and studied (basic DIA);*
2) *it was shown which types of image models are generated by main versions of DIA with a single ring;*
3) *the technique for defining and testing of necessary and sufficient conditions for generating DIA with a single ring by a set of image processing operations were suggested;*
4) *the necessary and sufficient conditions for generating basic DIAs with a single ring were formulated;*
5) *the hierarchical classification of image algebras was suggested;*
6) *it was proved that the Ritter's algebra could be used for construction DIA's without a "template object".*

This object is studied in developing a mathematical apparatus for analysis and estimation of information represented in the form of images. For a structural description of possible algorithms for solving these problems, we need a formal instrument that allows us to describe and justify the chosen way of solution. As formalization tools, we chose the algebraic approach, which should provide a unique form of procedures for describing the objects–images and transformations of these objects–images.

The need to develop a mathematical language that ensures that solutions of problems of image processing, analysis, and understanding may be uniformly described by structural algorithmic schemes is justified by the following factors:

(1) there are many algorithms (designed and introduced into practice) for analysis, estimation, and understanding of information represented in the form of images;

(2) the set of algorithms is neither structured nor ordered;

(3) as a rule, methods for image analysis and understanding are designed on the basis of intuitive principles, because the information represented in the form of images is hardly formalized;

(4) the efficiency of these methods is estimated (as is usual in experimental sciences) by the success in solving actual problems—as a rule, the problem of rigorous mathematical justification of an algorithm is not considered.

"Algebraization" is one of the most topical and promising directions of fundamental research in image analysis and understanding. The main goal of the algebraic approach is the development of a theoretical basis for representations and transformations of images in the form of algebraic structures that enable one to use methods from different areas of mathematics in image analysis and understanding.

An object that lies most closely to the developed DIA is the image algebra proposed and developed by Ritter [15, 16]. Ritter's main goal in developing the image algebra is the design of a standardized language for description of algorithms for image processing intended for parallel execution of operations. A key difference in the new image algebra from the standard Ritter image algebra is that DIA is developed as a descriptive tool, i.e., as a language for description of algorithms and images rather than a language for algorithm parallelizing.

The conceptual difference of the algebra under development from the standard image algebra is that objects of this algebra are (along with algorithms) descriptions of input information. DIA generalizes the standard image algebra and

allows one to use (as ring elements) basic models of images and operations on images or the models and operations simultaneously. In the general case, a DIA is the direct sum of rings whose elements may be images, image models, operations on images, and morphisms. As operations, we may use both standard algebraic operations and specialized operations of image processing and transformations represented in an algebraic form. In more detail, the definition of the standard image algebra and that of DIA are considered in [8].

To use DIA actively, it is necessary to investigate its possibilities and to attempt to unite all possible algebraic approaches, for instance, to use the standard image algebra as a convenient tool for recording certain algorithms for image processing and understanding or to use Grenander's concepts for representation of input information.

The main attention was given to DIAs with one ring, which form the main subclass of basic DIAs. In future, we are going to consider DIAs based on superalgebras and investigate other possibilities of application of other algebraic concepts in the theory being developed.

TO BE DONE:

*1) to study DIA with a single ring, whose elements are image models;*

*2) to study DIAs with several rings (super algebras);*

*3) to define and study of DIA operation bases;*

*4) to construct standardized algebraic schemes for solving image analysis and estimation problems on the DIA base;*

*5) to generate DIA using equivalence and invariance properties in an explicit form;*

*6) to demonstrate efficiency of using DIA in applied problems;*

*7) to study alternative algebraic languages for image analysis, recognition and understanding.*

## IV. Conclusion

In principle, the success of image analysis and recognition problem solution depends mainly on the success of image reduction to a recognizable form, which could be accepted by an appropriate image analysis/recognition algorithm. All above mentioned steps contribute to the development techniques for this kind of image reduction/image modeling. It appeared that an image reduction to a recognizable form is a critical issue for image analysis applications, in particular for qualified decision making on the base of image mining. The main tasks and problems of an image reduction to a recognizable form are listed below:

1. Formal Description of Images:

1) Study and construction of image models (Step 2);

2) Study and construction of multimodel image representations (Step 3);

3) Study and construction of metrics (Step 5).

2. Description of Image Classes Reducible to a Recognizable Form:

1) Introduction of new mathematical settings of an image recognition problem (Step 1);

2) Establishing and study of links between multimodel representation of images and image metrics (Steps 3, 5);

3) Study and use of image equivalencies (Step 4).

3. Development, Study and Application of an Algebraic Language for Description of the Procedures of an Image Reduction to a Recognizable Form (Step 6).

We hope that after passing through the above mentioned steps we'll be able to formulate the axiomatics of the descriptive (mathematical) theory of image analysis.

## Acknowledgment

This work was supported in part by the Russian Foundation for Basic Research (projects no. 14–01–00881), by the Presidium of the Russian Academy of Sciences within the program of the Department of Mathematical Sciences, Russian Academy of Sciences "Algebraic and Combinatorial Methods of Mathematical Cybernetics and Information Systems of New Generation" ("Algorithmic schemes of descriptive image analysis") and "Information, Control, and Intelligent Technologies and Systems" (project no. 204).

## References

[1] U. Grenander. General Pattern Theory. A Mathematical Study of Regular Structure. Clarendon Press, Oxford, 1993.

[2] U. Grenander. Elements of Pattern Theory. The Johns Hopkins University Press, 1996.

[3] I.B. Gurevich. "The Descriptive Framework for an Image Recognition Problem", Proceedings of the 6th Scandinavian Conference on Image Analysis.- Pattern Recognition Society of Finland, 1989.- vol. 1. – P. 220 – 227.

[4] I.B. Gurevich. "Descriptive Technique for Image Description, Representation and Recognition", Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications in the USSR.- MAIK "Interpreodika", 1991.-vol. 1- P. 50 – 53.

[5] I.B. Gurevich. "The Descriptive Approach to Image Analysis. Current State and Prospects", Proceedings of 14th Scandinavian Conference on Image Analysis.- Springer-Verlag Berlin Heidelberg, 2005.- LNCS 3540.- pp. 214-223.

[6] I.B.Gurevich, I.A. Jernova. "The Joint Use of Image Equivalents and Image Invariants in Image Recognition", Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications. - 2003. - Vol. 13, No.4. - pp. 570-578.

[7] I.B. Gurevich and I.V. Koryabkina. "Comparative Analysis and Classification of Features for Image Models", Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications. - MAIK "Nauka/Interperiodica"/Pleiades Publishing, Inc., 2006. - Vol.16, No.3. - P. 265-297.

[8] I.B. Gurevich, V.V. Yashina. "Operations of Descriptive Image Algebras with One Ring", Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications. Pleiades Publishing, Inc. 2006. - Vol.16, No.3. - pp. 298-328.

[9] I.B. Gurevich and V.V. Yashina. "Computer-Aided Image Analysis Based on the Concepts of Invariance and Equivalence", Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications. - MAIK "Nauka/Interperiodica"/Pleiades Publishing, Inc., 2006. - Vol.16, No.4. – pp.564-589.

[10] I. Gurevich, V. Yashina. “Descriptive Theory of Image Analysis. Models and Techniques”, 8th International Conference “Pattern Recognition and Image Analysis: New Information Technologies (PRIA-8-2007). Conference proceedings. In two volumes. – Yoshkar-Ola, 2007. - Vol.1. - P. 103-112.

[11] I.B. Gurevich and V.V. Yashina. “Descriptive Approach to Image Analysis: Image Models”, Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications. - MAIK "Nauka/Interperiodica"/Pleiades Publishing, Inc., 2008. - Vol.18, No.4. - P. 518-541.

[12] I.B. Gurevich, V.V. Yashina, I.V. Koryabkina, H. Niemann, and O. Salvetti. “Descriptive Approach to Medical Image Mining. An Algorithmic Scheme for Analysis of Cytological Specimens”, Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications. - MAIK "Nauka/Interperiodica"/Pleiades Publishing, Inc., 2008. - Vol.18, No.4. - P. 542-562.

[13] I.B.Gurevich, V.V. Yashina. “Descriptive Approach to Image Analysis: Image Formalization Space”, Pattern Recognition and Image Analysis, 2012, Vol. 22, No. 4, pp. 495-518.

[14] G.X. Ritter, J.N. Wilson. Handbook of Computer Vision Algorithms in Image Algebra, 2-d Edition. CRC Press Inc., 2001.

[15] G.X. Ritter. Image Algebra. Center for computer vision and visualization, Department of Computer and Information science and Engineering, University of Florida, Gainesville, FL 32611, 2001.

[16] S. R. Sternberg. An overview of Image Algebra and Related Architectures, Integrated Technology for parallel Image Processing (S. Levialdi, ed.), London: Academic Press, 1985.

[17] D. Marr. Vision, Freeman, New York, 1982.

[18] Yu.I. Zhuravlev. “An Algebraic Approach to Recognition and Classification Problems”, Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications.- MAIK "Nauka/Interperiodica", vol.8. 1998.-pp.59-100.

# Real-time hand detection using continuous skeletons

Victor Chernyshov
Department of Computational Mathematics
and Cybernetics
Moscow State University
Moscow, Russia
Email: webcreator18@gmail.com

Leonid Mestetskiy
Department of Computational Mathematics
and Cybernetics
Moscow State University
Moscow, Russia
Email: mestlm@mail.ru

***Abstract*—In this paper, a fast and reliable method for hand detection based on continuous skeletons approach is presented. It demonstrates real-time working speed and high detection accuracy (3-5% both FAR and FRR) on a large dataset (50 persons, 80 videos, 2322 frames). These make it suitable for use as a part of modern hand identification systems including mobile ones. Overall, the study shows that continuous skeletons approach can be used as prior for object and background color models in segmentation methods with supervised learning (e.g. interactive segmentation with seeds or abounding box).**

## I. Introduction

Rapid progress in mobile technologies naturally causes the development of personal biometrics systems based on tablets and smartphones. Together with iris and fingerprints, hand is one of the most promising biometrics modalities. Characteristics of modern devices (performance, camera quality, wireless communication capabilities) make it possible to realize various types of architecture of hand authentification/identification application [1]:

1) mobile device based (the full cycle of processing is on board of a mobile device),
2) "truly" client-server architecture (a mobile device only captures images/videos and sends to a server),
3) hybrid (the processing stages are shared by a mobile device and a server).

Apparently, that schemes 1) and 3) suppose the client side of the application to be as fast as possible.

The main objective of this paper is to introduce fast and reliable method for real-time hand detection based on continuous skeletons approach [2], [3].

Another motive is to show the outlook of using skeleton representation for setting appearance models in segmentation methods with supervised learning.

Algorithms described in this work are a part of ongoing project "Mobile Palm Identification System" (MoPIS[1]), earlier introduced in [4]. We remind of the fact that MoPIS system has a client-server architecture where the client is an application for Android-based mobile devices, and the server is implemented with a help of Debian GNU/Linux distributive, Nginx web-server and programming language C++. Android application uses frames from a high-resolution video camera as input.

[1]http://mopis.ru is project's official website (currently in russian)

Let us highlight the main ideas which have led to the development of the proposed method. While processing a frame we usually conduct two consequent subroutines: hand detection in the frame (at client-side) and, in the case of a positive outcome, detailed analysis of this frame with a hand (at server-side). Due to the network restrictions we can send only several frames to the server during the identification session. That's why hand detection procedure should approve for the further analysis as few images guaranteed to be unfit as possible (e.g. we need a low false acceptance rate). And at the same time, it should process a video stream from the camera at wide range of mobile devices in real-time. Thus, the required method has to meet very strict requirements both to the quality of recognition and performance.

There have been some research to detect hand using AdaBoost-based methods with promising results in accuracy and speed [5], [6], [7] suitable for use in unconstrained environments (various kinds of lighting, diverse background, etc.). But these methods generally need exhaustive classifier training and a large dataset including samples with different rotations and scaling. Also such methods don't explicitly utilize any hand geometrics like mutual disposition of fingers or their proportions, making difficult to separate "bad" hands (e.g. with partially "glued" fingers — Fig. 3; such sample is ineligible for the shape analysis procedure) from "good" ones. Another approach is to use skin color based detection [8], [9] but it is unreliable because of sensitivity to lighting conditions and messing with skin-colored objects. Optical flow methods [10] demonstrate good results for stationary cameras and permanently moving objects, so, cann't be directly used in our case without improvements.

The rest of the paper is organized as follows. Equipment and collected dataset are described in Section II. The hand detection procedure is fully presented in Section III. Next, the experiment results are introduced and discussed in Section IV. Finally, Section V shortly concludes the paper.

## II. Data

During the research we collected 80 short videos of hands of 50 different people (1-3 video for each person, the back side of the right hand was captured). All videos were taken using cameras of mobile devices. After that they were decomposed into frames (each 5th frame was used), which were saved as graphic files (*.jpg or *.bmp). As a result, we got 2322 images.

Important notice: in our work we consider the assistance of participants — the reasonable person's intention is to be

correctly and quickly recognized by the identification system. Thus, all cases of cheating (and corresponding videos as well) are excluded from consideration. To form a qualitative dataset (as to get adequate results from using MoPIS) one should follow the recommendations given below while capturing videos:

1) The videos should be recorded using a mobile device with a camera producing video files with a resolution of 640*360 and more and frequency of 15 frames/second or higher. Preference should be given to devices with hardware autofocus support. The optimal duration is 3-6 seconds. For video recording one can utilize an Android application which is a part of MoPIS, and also similar applications. During the experiments we mainly used the smartphone LG G2 with 13Mp camera (supports autofocus and optical stabilization), and recorded HD video (1920x1080 or 1280x720 resolution and 30 frames/sec frequency) with a help of MoPIS Android application. Also, some data was captured using Samsung Galaxy Note 10.1 tablet (5Mp camera; 1280x720 video resolution and 30 frames/sec frequency).
2) The background should be black or dark (homogeneity is not necessary) otherwise there may be problems with binarization by Otsu (and therefore with a hand detection in the frame). This in turn will affect the quality of the further analysis of the hand. So, we used black homogeneous cloth as a background in our experiments.
3) The camera is supposed to be stable, only the tested hand should move. Modern mobile devices have good optical stabilization system, so, there is no need in a tripod or a holder to fix the position. Nonetheless, some of the videos of our dataset were made with a help of a tripod (Fig. 1).
4) To increase the variability of frames obtained from the video the probationer should slowly move his fingers (bring together and separate them) in the horizontal plane. One should avoid sudden movements. To improve the representativeness of the dataset it's highly advisable to record a few videos from each hand in different lighting conditions.
5) Experiments should be carried out in good diffused light (artificial or natural), and in its absence we recommend to turn on the built-in mobile flash.

## III. Detection procedure

The first thing we need to do after getting a frame is perform a rescaling (factor of $1/2$, $1/3$ or even less; the majority of experiment was done with 640x360 images) — hand detection procedure is supposed to be a real-time routine even at mobile devives.

Next, a binarization is performed and the largest contours are marked out for further analysis.

Otsu binarization [11] was chosen as a primary method (Fig. 4). The good balance of speed, accuracy and versatility was proved by the numerous experiments. Of course, one should follow the recommendations described in the Section II to get acceptable quality but really it doesn't significantly limit

Fig. 1. MoPIS experimental setup based on Samsung Galaxy Note 10.1.

the range of applicability of our application. In the Fig. 5 you can find the visualized output of the detection procedure. Though the boundary of hand (i. e. the result of countour traversing after binarization, green line) is a little bit saw-edged, the whole detection procedure worked out correctly.

### A. Skeleton construction

As it was mentioned in Introduction, the detection procedure heavily uses continuous skeleton of a binarized image. Skeleton representation of an object supposed to be a hand is built and afterwards regularized using the same logic as described in [2]. Both internal and external skeletons (see Fig. 5 — they are drawn via red and yellow, blue and turquoise segments respectively) of hand shape are used for the further analysis. The hand detection problem in our case consider low false positive rate, so, we have to develop an extended check "valid hand/not valid" that is introduced in the next subsection. Proposed validation routine significantly improved the one presented in [2] if we consider detection results on collected dataset containing the *mix* of valid and invalid hands.

### B. Hand validation

All the checks provided in this subsection are applied sequentially. If the false detection result is returned then the shape is excluded from the following analysis. If all tests are passed the shape is supposed to be valid hand.

First of all, the internal skeleton is examined to the depth from the vertices of degree 1. It searches the branches starting

at vertices of degree 1 and ending at the *root* (the center of maximal circle with radius $R_{max}$ inscribed in the shape; it's also called *center of hand*). Thus, *finger branches candidates* are obtained. A branch failed to pass some check is eliminated from the following processing. If at a certain point less than 5 branches are found the false detection result is considered. Similarly, if in the end of validation more than 5 branches are left.

Radius of maximum inscribed circle is associated with any point of skeleton. Function $R(x)$ that maps points of skeleton to radiuses of maximum inscribed circles is called the *radial function*. So, the values of the radial function (or its interpolation) in 30 cue points evenly located from the tip to the end of the finger branch candidate are calculated. After it, the linear mapping $X \times R(X) \to [0,1] \times [0,1]$ is applied. The obtained function is named *normalized branch radial function*. Using train dataset makes it easy to calculate the lower and upper bounds of this function in the cue points and than slightly expand this "tube" called *finger boundary corridor*. We consider the general boundary corridor for all fingers. To perform the *boundary coridor check* one just need to sure that normalized branch radial function is located inside the boundary corridor (Fig. 2).

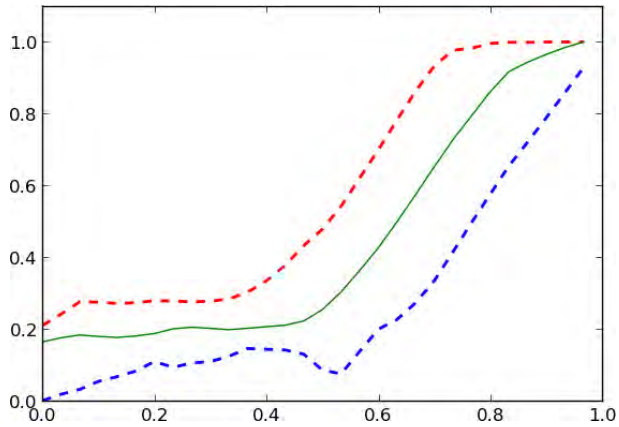



Fig. 2. Frame 1217.0000. Doted lines are bounds of finger boundary corridor. Green solid line — normalized branch radial function of the pointing finger ($id = 1$).

The next step is to determine the branch top and bottom nodes corresponding to the tip and base of a potential finger. This is done similarly to the method given in [2]. The line connecting top and bottom nodes is called the *axis* of the branch. The total length of branch edges between top and bottom nodes is called *length* of a potential finger. After it, several threshold checks (hereinafter, all the values are obtained from train dataset and all the distances are normalized to $R_{max}$) are applied. We examine whether the bottom node of a finger is at distance from the root which is greater than the threshold value $\epsilon_1$. The same checking is applied for the top node (threshold $\epsilon_2$).

Further, for a given potential finger we delete all the fingers which bottom nodes are located inside the bottom node circle of this finger on condition that they are shorter than the given one.

Next, we come to *triples* check. The fingers are arranged in the order of the contour traversal. Triples of adjacent fingers

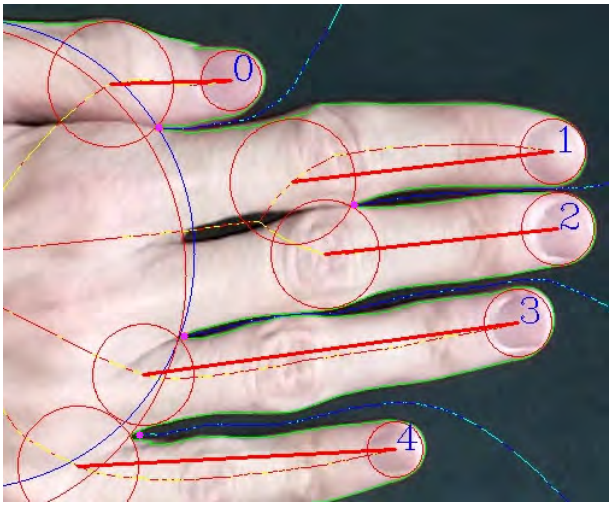



Fig. 3. Frame 1237.0160, cropped. Fingers sticked together.

are examined. For each triple the middles of the first and the third fingers are connected. This segment is supposed to intersect the second finger (the point of intersection should lie within both segments). Such a heuristics successfully works because of blob structure of a hand and small axes between neighboring fingers.

To the present moment the vast majority of non-fingers branches are discarded. We expect that the most distant to the others bottom node belongs to the thumb. So, the fingers are put in the order of the contour traversal starting from the thumb ($id = 0$ is assigned).

Though the validation process above is strict and reliable, it doesn't cope with fingers sticked together (Fig. 3). That's why we implemented *median* check. We consider 4 fingers (without a thumb). The euclidian distances $\rho_i, i = 1 \ldots 4$ from the bottom node of the fingers to the root are ordered by ascending, as a reference value we select the second value $d_2$ from the beginning. After that, for each finger we count normalized deviations from the reference value: $\eta_i = |\rho_2 - \rho_i| / R_{max}, i \in \{1, 3, 4\}$. Its verified that these deviations are less than the threshold $\epsilon_3$.

Fig. 4. Frame 1217.0000. Otsu binarization.

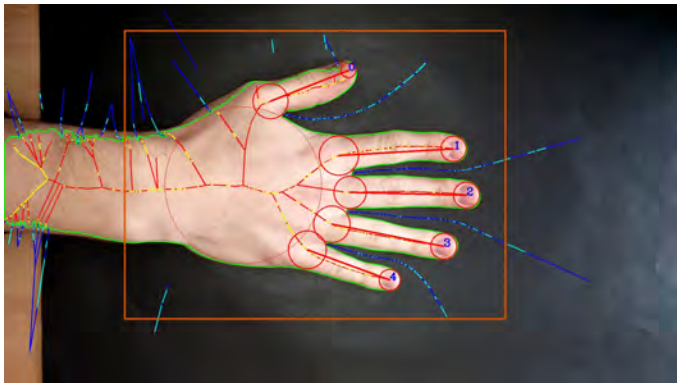


Fig. 5. Frame 1217.0000. After applying detection procedure. Nonlinear Voronoi sites are yellow and turquoise "segments", linear — red and blue segments. Bounding box of the hand is orange.

Fig. 6. Frame 1217.0000. hand (red) and background (blue) seeds.

Fig. 7. Frame 1217.0000. Grabcut segmentation.

## IV. Experiments

All frames from the dataset (see Section II) were manually marked: either containing valid hand or not.

### A. Detection testing

The testing scheme for detection was organized as follows. We made 3 random partitions of our frame dataset into train and test datasets, containing frames of 20 and 30 different people respectively. Train dataset was used to calculate threshold statistics, test was utilized for quality and speed estimation. All frames with resolution 1280x720 and 1920x1080 were resized to 640x360. The experiments were run on a laptop with Intel Core i7 2.4 GHz processor without using multiprocessing routines. According to the time profiler the most time consuming procedure was creation of Voronoi diagram — about 60% of computational time. Hand detection results can be found in Table I. Each row corresponds to a dataset partition. The last column contains time per frame (TPS) values for the whole detection method (summing execution time of binarization, skeleton construction and hand validation).

| # | FAR, % | FRR, % | errors, % | TPF, ms |
|---|---|---|---|---|
| 1 | 3.2 | 4.1 | 3.4 | 40.5 |
| 2 | 2.8 | 4.6 | 3.3 | 40.2 |
| 3 | 3.7 | 5.2 | 4.1 | 40.0 |

TABLE I. Detection results.

Since proposed detection procedure demonstrates low error rates (both false acceptance and false rejection) and real-time processing speed it can be used in hand biometrics systems, e. g. as a part of client-side application. The algorithm is robust to image quality — it works well even with low-resolution images (e. g. with 320x180). The most false acceptance cases are related to hands partially placed in the frame. Also, there are some difficulties with blurred frames though we tried to eliminate quick hand and fingers motions in data acquisition process.

### B. Using detection results for setting appearance models

One of the main purposes of a MoPIS client-server architecture is to remove computationally heavy procedures (like "clever" Markov Random Fields based segmentation methods that are usefull for more accurate shape features extraction) from Android application to a server. At the same time, when the detection procedure is completed the internal and external skeletons representation of a hand are already known. And it can be easily adopted for using as prior for object and background color models in advanced segmentation methods with supervised learning (e.g. interactive segmentation with seeds or abounding box).

As seeds we use circles with centers at the vertices of the skeleton graph (Fig. 6) located inside the bounding box (Fig. 5). For an internal skeleton only fingers branches and the root circle are used. For an external skeleton only 4 branches lying between the fingers axes are considered. Seeds radiuses are the values of the radial functions in the skeleton vertice multiplied by a coefficient (we use 0.9 for internal skeletons, 0.7 for external). Also we limit the minimal external seed radius. Finally, we utilized the seeds as input for graph cut driven method [12] and got really good segmentation results (Fig. 7) that is proven by identification experiments (see the next paragraph). The boundary is correct and smooth — shape is ready to use for features extraction. It should be noted that there is a quantity of supervised segmentation methods

that can be used jointly with seeds; Grabcut segmentation was selected because of its accuracy and speed.

For identification experiments we used the same hand dataset and testing scheme as in detection experiments, extracted only shape-based features and run simple 1NN classifier — utilizing segmentation produced by Grabcut's method gave significantly lower identification error rate then the same system powered with Otsu segmentation (Table II). The second and the third colums contain identification error rates for Otsu and Grabcut based systems correspondively, the last two columns — execution time of the aforementioned segmentation routines.

| # | Otsu, % | Grabcut, % | Otsu, ms | Grabcut, ms |
|---|---|---|---|---|
| 1 | 12.3 | 7.7 | 0.4 | 2002 |
| 2 | 13.6 | 8.1 | 0.4 | 1922 |
| 3 | 13.2 | 8.2 | 0.4 | 2107 |

TABLE II. SEGMENTATION RESULTS.

An average size of colored 640x360 hand images used in segmentation experiments is 30-50kB that meets client-server bandwidth limits. This image quality is sufficient for accurate shape features extraction — using instead source images with 1280x720 or 1920x1080 resolution didn't cause any notable identification results improvements. So, server-side Grabcut segmentation combined with client-side seeds produces solid base for the further hand shape features extraction.

## V. CONCLUSION

In this paper, we have presented a fast and reliable method for hand detection based on continuous skeletons approach, which showed real-time working speed and high detection accuracy (3-5% both FAR and FRR) on a large dataset (50 persons, 80 videos, 2322 frames). These make it suitable for use as a part of modern hand identification systems including mobile ones.

Overall, the study shows that continuous skeletons approach can be used as prior for object and background color models in segmentation methods with supervised learning. This thesis was confirmed by identification experiments.

## REFERENCES

[1] M. Franzgrote, C. Borg, B. Tobias Ries, S. Büssemake, X. Jiang, M. Fieleser, and L. Zhang, "Palmprint verification on mobile phones using accelerated competitive code," in *2011 International Conference on Hand-Based Biometrics (ICHB)*. IEEE, 2011, pp. 124–129.

[2] L. Mestetskiy, I. Bakina, and A. Kurakin, "Hand geometry analysis by continuous skeletons," in *Proceedings of the 8th international conference on Image analysis and recognition - Volume Part II*, ser. ICIAR'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 130–139.

[3] L. Mestetskiy, *Continuous Morphology of Binary Images: Figures, Skeletons and Circulas (in Russian)*. FIZMATLIT, 2009.

[4] V. Chernyshov and L. Mestetskiy, "Mobile machine vision system for palm-based identification," in *Proceedings of the 11th International Conference "Pattern Recognition and Image Analysis: New Information Technologies" (PRIA-11-2013)*, vol. 2, sep 2013, pp. 398–401.

[5] M. Kölsch and M. Turk, "Robust hand detection," in *In International Conference on Automatic Face and Gesture Recognition (to appear), Seoul, Korea*, 2004, pp. 614–619.

[6] Y. Fang, K. Wang, J. Cheng, and H. Lu, "A real-time hand gesture recognition method," in *Proceedings of the 2007 International Conference on Multimedia and Expo (ICME 2007), Beijing, China*. IEEE, 2007, pp. 995–998.

[7] B. Xiao, X.-m. Xu, and Q.-p. Mai, "Real-time hand detection and tracking using lbp features," in *Advanced Data Mining and Applications*, ser. Lecture Notes in Computer Science, L. Cao, J. Zhong, and Y. Feng, Eds. Springer Berlin Heidelberg, 2010, vol. 6441, pp. 282–289.

[8] A. M. Elgammal, C. Muang, and D. Hu, "Skin detection," in *Encyclopedia of Biometrics*, 2009, pp. 1218–1224.

[9] V. Vezhnevets, V. Sazonov, and A. Andreeva, "A survey on pixel-based skin color detection techniques," in *Proceedings of the GraphiCon 2003*, 2003, pp. 85–92.

[10] A. Sobral, "BGSLibrary: An opencv c++ background subtraction library," in *IX Workshop de Vis?o Computacional (WVC'2013)*, Rio de Janeiro, Brazil, Jun 2013.

[11] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[12] M. Tang, L. Gorelick, O. Veksler, and Y. Boykov, "Grabcut in one cut," in *Proceedings of the 2013 IEEE International Conference on Computer Vision*, ser. ICCV '13. Washington, DC, USA: IEEE Computer Society, 2013, pp. 1769–1776.

Proceedings of the 9th Open German-Russian Workshop on Pattern Recognition and Image Understanding
Dec 1-5, 2014, Koblenz, Germany

# Real-time Texture Error Detection on Textured Surfaces with Compressed Sensing

Tobias Böttger
MVTec Software GmbH
Neherstraße 1, 81675 Munich
Germany
Email: boettger@mvtec.com

Markus Ulrich
MVTec Software GmbH
Neherstraße 1, 81675 Munich
Germany
Email: ulrich@mvtec.com

***Abstract*—We present a real-time approach to detect and localise defects in grey-scale textures within a Compressed Sensing framework. Inspired by recent results in texture classification, we use compressed local grey-scale patches for texture description. In a first step, a Gaussian Mixture model is trained with the features extracted from a handful of defect-free texture samples. In a second step, the novelty detection of texture samples is performed by comparing each pixel to the likelihood obtained in the training process. The inspection stage is embedded into a multi-scale framework to enable real-time defect detection and localisation. The performance of compressed grey-scale patches for texture error detection is evaluated on two independent datasets. The proposed method is able to outperform the performance of non-compressed grey-scale patches in terms of accuracy and speed.**

## I. Introduction

Surface inspection is an important field within machine vision inspection systems that has traditionally been covered by trained human inspectors. However, the visual assessment by the human eye is not only time-consuming, it also lacks a sufficient degree of accuracy [1]. In order to increase the accuracy and to decrease the testing time, attempts are being made to replace manual inspection by automatic visual inspection systems.

An important subproblem of surface inspection is the detection of texture defects, which attempts to locate errors in textured images as is displayed in Fig. 1. Texture error detection schemes have a broad variety of industrial applications, one of the most prominent is fabric defect detection, which has received much attention within the last decade [2]. Further applications include the evaluation of 3D seismic data [3], fault detection in the wood industry [4], [5], error and crack detection in marble slates [6], detection of welding defects [7] and the grading of apples into quality categories [8]. Due to the large variety of applications a vast amount of different algorithms and machine-vision based inspection systems has been presented [9], [10].

Texture defect detection schemes differ mostly by the features used to discriminate the texture. Throughout the texture defect detection and texture classification literature many different features exist, the most common are based on filter banks [9]–[13]. Nevertheless, Varma and Zisserman [12] challenge the performance of filter banks in the context of texture classification by using compact pixel neighbourhood patches with sizes as small as $3\times3$ to describe texture. Xie and Mirmehdi [6] incorporate the features in their texture defect detection method named TEXEM, and are able to significantly outperform a similar state-of-the-art approach based on Gabor-filters [11].

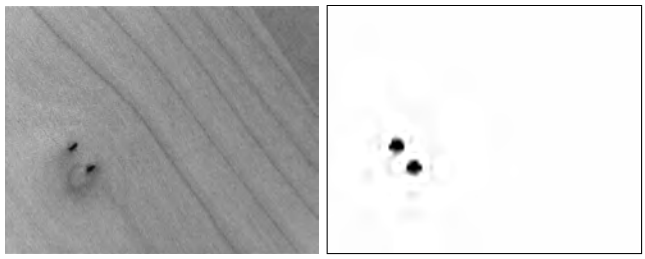


Fig. 1. An example of a texture defect detection application. The left image displays a wooden board with defects. The right image displays a localisation of the texture defect

The key parameter within patch-based texture classification and texture defect detection schemes is the patch size. Although small patches are able to achieve surprisingly good results, they can not capture large-scale structures and are inherently not robust to local changes of the texture. Larger patches are able to overcome these drawbacks, but lead to a quadratic increase of the feature dimension. The observation that an increased patch size leads to increasingly sparse features has motivated the use of Compressed Sensing to decrease the feature size without the loss of feature information. Liu *et al.* [14], [15] extract a small set of random features from locally extracted feature patches using random projection and embed them into a bag-of-words model to perform texture classification. The approach is able to outperform various state-of-the-art texture classification schemes [16].

Motivated by the success of compressed local image patches in texture classification, the main contribution of this paper is to use the same features in a texture defect detection framework. Due to its convincing results, our approach uses the the state-of-the-art TEXEM framework [6] for novelty detection. By adding signal compression in the feature extraction step, we are able to increase the defect detection accuracy of the TEXEM method at a considerably reduced computational complexity.

## II. Related Work

Texture defect detection schemes can be divided into three groups, distinguished by the amount of training that is required beforehand: *supervised texture classification*, *unsu-*

*pervised novelty detection* and *supervised novelty detection*. Since supervised texture classification methods assume that all possible texture defects are known beforehand [17], their application for industrial inspection applications is restricted. On the other hand, unsupervised novelty detection approaches try to detect defects in an automatic framework without prior information on the texture and the possible defects. Although fully automatic schemes would be ideal within an industrial inspection system, the approaches to date are restricted to textures that exhibit a high degree of regularity, such as fabrics [2], [18], [19]. As a further restriction, they often assume that the texture defects only occupy a relatively small area within a defect-free background [20].

Supervised novelty detection schemes use a set of defect-free image samples to construct a model of flawless texture. In the inspection step, the model is used to identify and locate novel regions that do not belong to the texture. When the images within the inspection framework have a certain consistency in terms of their imaging quality, supervised methods have been able to achieve remarkable results [6], [11]. To date, most novelty detection schemes use filter-bank-based features such as Gabor filters [20], [21] or Wavelets [22]. Nevertheless, Xie and Mirmehdi [6] were able to show the superiority of their local patch based method to the state-of-the-art Gabor filter-based approach from Escofet [11]. An approach using Local Binary Patterns [23] was not able to compete with either of the results.

In general, Compressed Sensing measuring systems have received increasing attention within computer vision in the last years. Applications range from face recognition [24], texture classification [15] to object tracking [25]. Their explicit use for texture description has very recently been proposed within a texture classification framework [14]. The compression of local texture patches as texture features was able to considerably outperform many existing state-of-the-art methods [16], [26].

## III. Method

Although the novelty detection results of the TEXEM method are convincing, its use for industrial applications is restricted due to its high computational overhead. The authors themselves state that "the computational needs of the method are somewhat demanding for a real-time factory installation" [6]. To enable a real-time implementation, we propose to adapt their general framework and to incorporate the ideas of Compressed Sensing within the feature extraction.

### A. Compressed Sensing (CS)

The validity of using compressed local texture patches can be evidenced with the theory of CS. The standard finite dimensional CS model by Candès, Romberg, Tao [27], [28] and Donoho [29] sets out with a compressible signal in $\mathbb{R}^n$ and a non-adaptive linear measurement system. The measurement of a local grey-scale texture patch $Z_i \in \mathbb{R}^n$ can be represented as

$$Y_i = AZ_i \ , \tag{1}$$

were $A$ is an $m \times n$ sensing matrix and $Y_i \in \mathbb{R}^m$. The matrix $A$ performs a dimensionality reduction from $\mathbb{R}^n$ to $\mathbb{R}^m$, where $n$ is typically much larger than $m$. Per definition, $A$ destroys information of the input patch $Z_i$ since it has a nullspace.

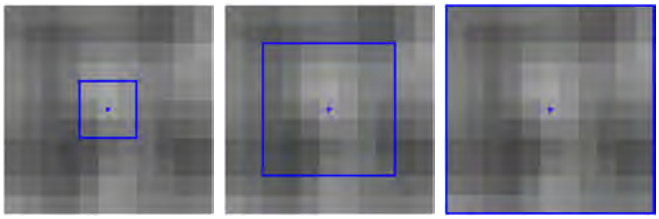


Fig. 2. Local grey value image patches of size 3×3, 7×7 and 11×11, with corresponding feature dimensions of $9, 49$ and $121$ respectively. The assumption is that although larger patches incorporate more texture information, they include an increasingly large amount of *redundant* texture information

In order to perform a reduction and be able to guarantee a possible reconstruction of the input vector, a certain amount of compressibility or sparsity needs to be provided by the input signal $Z_i$. In Fig. 2 it becomes obvious that though larger texture patches have the advantage of increasing the overall feature information, they have a growing amount of redundant information. This observation motivated the use of CS in the texture classification scheme of Liu and Fieguth [14]. In order to ensure that the dimensionally reduction (1) is information preserving, it would be ideal for $A$ to approximately preserve the distance between two input patches $Z_i$ and $Z_j$ in the sense that

$$1 - \epsilon \leq \frac{\|A(Z_i - Z_j)\|_2}{\| Z_i - Z_j\|_2} \leq 1 + \epsilon \ , \tag{2}$$

for small $\epsilon > 0$. Although it might seem surprising at first, Baraniuk *et al.* [30] present a fundamental relationship between CS and the Johnson-Lindenstrauss lemma, showing (2) is satisfied by certain random matrices, specifically including random matrices drawn from Gaussian distributions. Under this theoretical foundation the use of CS to compress local texture patches for texture description is validated [26].

### B. Texture Features

The use of CS to create compressed texture features is straightforward. First an adequate sensing matrix $A \in \mathbb{R}^{m \times n}$, where $n$ is the number of pixels in the extracted texture patches and $m$ the compressed feature dimension, needs to be constructed. It was proven in [31] that matrices where the entries are independent realizations of Gaussian or Bernoulli random variables or of related distributions are a valid choice [30]. The construction scheme for the Gaussian sensing matrix used within this paper is presented in [32].

The dimension reduction of the texture features is achieved by projecting each patch $Z_i \in \mathbb{R}^n$ to $\mathbb{R}^m$ with the sensing matrix $A$, creating the set

$$\mathcal{Y} = \{Y_i \in \mathbb{R}^m\}_{i=1}^P = \{AZ_i \in \mathbb{R}^m\}_{i=1}^P \ , \tag{3}$$

where $P$ is the number of patches extracted from the training images. Liu *et al.* [14] show the texture classification performance of the texture patches to level off for $m \approx n/3$ in their experiments.

### C. Texture Exemplars (TEXEMS)

The novelty detection framework consists of an offline training phase that learns a texture model and an online testing phase, were texture images are tested against the texture model and errors are located. Every pixel constitutes the centre of one

$\sqrt{n} \times \sqrt{n}$ squared image patch $Z_i$ and contributes to the set of compressed features vectors $\mathcal{Y}$ from (3). The patches can be viewed as independent realisations of a Gaussian mixture model (GMM), which is determined by the set of $K$ mixtures, or *texems*, $\mathcal{M} = \{m_k\}_{k=1}^{K}$. Given a texem $m_k$, the probability of a feature $Y_i$ is $p(Y_i|m_k)$. Using the law of total probability, $p(m_k, \Theta) = \alpha_k$ and the fact that each texem is determined by its mean $\mu_k$ and its covariance matrix $\Sigma_k$, we obtain:

$$p(Y_i|\Theta) = \sum_{k=1}^{K} p(Y_i|m_k, \theta_k) p(m_k, \Theta) = \sum_{k=1}^{K} \mathcal{N}(Y_i; \mu_k, \Sigma_k)\alpha_k \ , \tag{4}$$

where $\Theta = \{\theta_k\}_{k=1}^{K} = \{\alpha_k, \mu_k, \Sigma_k\}_{k=1}^{K}$ is the set of parameters containing $\alpha_k$, which is the prior probability of the $k$-th texem, constrained by $\sum_{k=1}^{K} \alpha_k = 1$. To calculate the model with the most probable parameter set $\Theta$, we require the maxima of the Log-Likelihood function

$$\log \mathcal{L}(\Theta|\mathcal{Y}) = \sum_{i=1}^{P} \log \left( \sum_{k=1}^{K} \mathcal{N}(Y_i; \mu_k, \Sigma_k)\alpha_k \right) , \tag{5}$$

which can efficiently be determined using the EM-algorithm [33].

After the GMM has successfully been calculated, a threshold to distinguish between defective and non-defective image patches is determined. The novelty score of each pixel is the negative log-likelihood of the corresponding feature patch:

$$\mathcal{V}(Y_i|\Theta) = -\log \mathcal{L}(\Theta|Y_i) \ . \tag{6}$$

The lower the novelty score, the more likely it is that the patch belongs to the GMM. The one-dimensional distribution of patch probabilities is clustered and the sigma surrounding of the largest cluster is used to identify the upper threshold of defect-free texture.

In the test stage, a patch is extracted for each pixel, compressed, and its novelty score compared to the threshold determined in training. The test setup is embedded into a multi-scale framework, originally proposed by Escofet [11] and also incorporated within the original TEXEM framework [6].

### D. Justification and Computational Complexity

The use of compressed features has two main advantages. First of all, the dimension of the feature $Y_i$ effectively influences the quality of the optimum that can be found by the EM-algorithm. GMMs suffer from the curse of dimensionality, in the sense that for high dimensional data, a very large number of samples is required for training. This leads to the fact that learning GMMs with high dimensional data is not only computationally demanding, with increasing dimension, it is also increasingly difficult to find a good maximum of (5), as was shown by Dasgupta [32].

Secondly, the evaluation of (6) is required for every pixel within the testing stage. The numerical complexity of the evaluation is essentially quadratic in the feature dimension $n$:

$$\sum_{i=1}^{P} \underbrace{\log}_{\mathcal{O}(1)} \sum_{k=1}^{K} \underbrace{\mathcal{N}(Y_i; \mu_k, \Sigma_k)\alpha_k}_{\mathcal{O}(n^2)} \ . \tag{7}$$

Although using CS adds a matrix vector multiplication for each feature patch in the feature extraction step, its computation only requires $m^2 + nm - m$ operations. Hence, the proposed approach is able to reduce the computational complexity of the testing stage from $\mathcal{O}(n^2)$ to $\mathcal{O}(mn)$.

## IV. Experimental Results

The original TEXEM method was presented in three different forms; a grey-level scheme, a scheme that analyses the image colour channels independently and thirdly, an approach using a full colour model [6]. Although the best results were achieved by the full colour model, the computation overhead is around 10 times bigger than the grey-level scheme and thus infeasible for a real-time implementation. The functionality of the other two schemes is very similar, differing mostly in a pre-processing step that converts the RGB channel images of the coloured approach to PCA-based channels. The single channels are then processed by the grey-scale scheme independently. Since we are focusing on a real-time application we compare our method to the grey-scale approach and show its superiority.

We evaluate the performance for two datasets. Firstly, we use 44 different collages of the MIT VisTex database [34], similar to those used by Xie and Mirmehdi [6]. Furthermore, we evaluate the methods on a dataset of over 200 images of 17 different textures and various different texture defects.

In accordance with the evaluation performed in [6], we quantify the testing results by calculating the *specificity*, the *sensitivity*, and the *accuracy* according to:

$$\begin{cases} \text{specificity} & = \frac{F_t \cap F_g}{F_g} \times 100\% \\ \text{sensitivity} & = \frac{D_t \cap D_g}{D_g} \times 100\% \\ \text{accuracy} & = \frac{F_t \cap F_g + D_t \cap D_g}{F_g + D_g} \times 100\% \ , \end{cases} \tag{8}$$

where $D$ is the set of defective pixels, $F$ is the set of flawless pixels and the subscripts $t$ and $g$ denote the testing results and the ground truth, respectively.

In the original publication [6], the novelty detection capabilities of the TEXEM method were compared to similar state-of-the-art methods, such as Escofets Gabor filter-based approach [11] or Ojalas LBP-based scheme [23]. Since the TEXEM method significantly outperformed the other methods, we restrict the evaluation to comparing the performance of the standard TEXEM method to the proposed approach using compressed features.

The fact that every texture patch is supported by the $\sqrt{n} \times \sqrt{n}$ neighbourhood of a pixel leads to an artificial dilation of the resulting novelty regions. To neglect this effect and in order to enable a fair comparability of the different methods, an erosion with a diamond shape with radius 1 is conducted to the standard TEXEM approach (Patch25) and of radius 2 for the approaches using $7 \times 7$ patches (Patch49 and PatchCS15), since they have a larger support. The parameter settings are displayed in Table I.

TABLE I. The parameter settings used within the experiments

| $K$ | Method | Method Settings | Feature Dim. | #Scales |
|---|---|---|---|---|
| 12 | Patch25 | $\sqrt{n} = 5$ | 25 | 4 |
| 12 | Patch49 | $\sqrt{n} = 7$ | 49 | 4 |
| 12 | PatchCS15 | $\sqrt{n} = 7, m = 15$ | 15 | 4 |

In general, the EM-algorithm requires the number of Gaussian mixtures $K$ to be determined beforehand. We use the

TABLE II. *Specificity*, *sensitivity*, AND *accuracy* RATES THROUGHOUT THE 44 DIFFERENT VISTEX COLLAGES FOR THE APPROACH WITHOUT (PATCH25 AND PATCH49) AND WITH (PATCHCS15) CS. THE EVALUATIONS WERE PERFORMED 5 TIMES AND THE MEAN VALUES AND THE CORRESPONDING STANDARD DEVIATIONS ARE DISPLAYED. THE EXECUTION TIMES ARE NORMALIZED WITH RESPECT TO THE PATCH25 APPROACH

| | **spec.** $\pm\sigma$ | **sens.** $\pm\sigma$ | **accu.** $\pm\sigma$ | **time** |
|---|---|---|---|---|
| Patch25 | 91.48 $\pm$0.004 | 52.62 $\pm$0.02 | 71.8 $\pm$0.006 | 1.0 |
| Patch49 | 91.36 $\pm$0.006 | **57.56** $\pm$0.02 | **74.16** $\pm$0.007 | 2.05 |
| PatchCS15 | **93.12** $\pm$0.005 | 48.34 $\pm$0.03 | 70.5 $\pm$0.011 | **0.63** |

fixed number of $K = 12$ in all our experiments[1], although $K$ could equally be determined by an optimization schemes or by determining optimal key figures like the Akaike information criterion (AIC) or the Bayesian information criterion (BIC) [35].

### A. Evaluation Using VisTex Collages

The VisTex database is used to establish the novelty detection capabilities of the TEXEM framework. In a first step, the feature patches extracted from a single texture image are used for the training process. In a second step, the texture model is compared to image collages constructed in such a manner, that $50\%$ of the image is covered by a texture that is different from the texture used within the training process. A handful of collages and the resulting novelty detection results are displayed in Fig. 3.

The average value and the standard deviation of the measures presented in (8) are calculated for 5 independent test runs, which each use a different random projection matrix. The results presented in Table II show that although the compressed sensing approach PatchCS15 has a reduced feature dimension, it is able to compete with the classification accuracy of the regular Patch25 method at a considerably reduced processing time. Furthermore, although a new random matrix is calculated for each test run, the approach only has a marginally increased standard deviation. This outlines the validity of using random matrices for features compression within a novelty detection framework.

### B. Evaluation Using Texture Defect Database

To display the capabilities of the proposed method for texture error detection, we evaluate its performance on observable real-life texture errors. The database contains various textures and their defects, ranging from knotholes in natural texture such as wood, holes in randomly structured texture such as cork to errors in regular texture such as fabrics. To compare the defect detection performance, a ground truth was created by manually marking the defects in almost 200 images.

Again, the evaluations were performed 5 times and the mean and the corresponding standard deviation are displayed in Table III. The results show that the PatchCS15 method is able to significantly decrease the runtime while marginally increasing the accuracy of the defect detection results. Although the original TEXEM approach was specifically constructed to help locate defects in random texture, the database includes textures of periodic and structural nature. Nevertheless, the novelty detection framework using compressed patches is able

[1]In accordance with the observations made in [6]

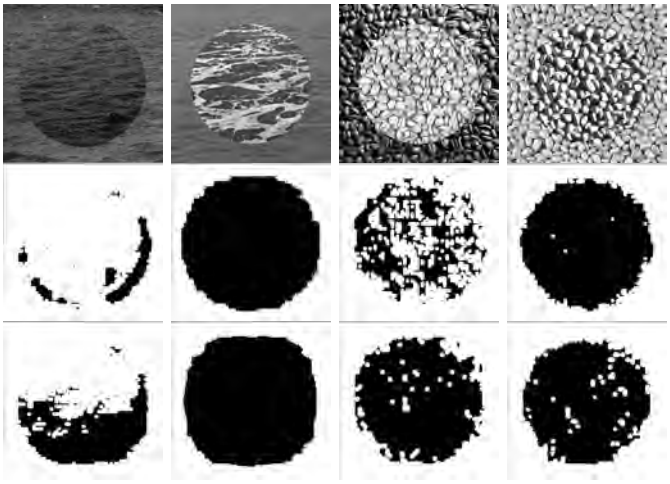


Fig. 3. The R channel of a few collages created from the VisTex database. The detected novelty regions in the middle row are obtained from the Patch25 method and in the bottom row from the proposed PatchCS15 approach

TABLE III. *Specificity*, *sensitivity*, AND *accuracy* RATES OF THE HAND-MARKED TEXTURE IMAGES FOR THE APPROACH WITHOUT (PATCH25 AND PATCH49) AND WITH (PATCHCS15) CS. THE EVALUATIONS WERE PERFORMED 5 TIMES AND THE MEAN VALUES AND THE CORRESPONDING STANDARD DEVIATIONS ARE DISPLAYED. THE EXECUTION TIMES ARE NORMALIZED WITH RESPECT TO THE PATCH25 APPROACH

| | **spec.** $\pm\sigma$ | **sens.** $\pm\sigma$ | **accu.** $\pm\sigma$ | **time** |
|---|---|---|---|---|
| Patch25 | 85.8 $\pm$0.012 | 81.76 $\pm$0.011 | 85.5 $\pm$0.011 | 1.0 |
| Patch49 | 81.92 $\pm$0.006 | **83.8**$\pm$0.014 | 82.02 $\pm$0.006 | 2.0 |
| PatchCS15 | **87.98** $\pm$0.014 | 78.66 $\pm$0.026 | **87.54**$\pm$0.014 | **0.62** |

to correctly classify approximately $88.0\%$ of the pixels within the database. A selection of the defect detection results and the corresponding hand-marked ground truths are displayed in Fig. 4. It should be noted that the parameters where unchanged for all of the experiments. If prior knowledge about the structure of the texture and the expected defects is known beforehand, the detection results can be improved considerably. For example, the horizontal fabric defect in the bottom right image of Fig. 4 is detectable by the lowest scale within the multi-scale framework. Nevertheless, the detection is lost because it does not reappear in higher scales due to its high frequency. Adjusting the number of scales or the image resolution would enable the approach to report the defect.

To test a single image channel, a C implementation of the proposed approach requires around $100$ms on an Intel(R) Core i5-4430 CPU with 3.0GHz for the parameters $n = 49$ and $m = 15$ for $256 \times 256$ sized images. Fortunately, the structure of the algorithm is highly parallelisable and an optimised code should be capable of even lower execution times.

## V. CONCLUSION

We have presented a texture defect detection algorithm combining the TEXEM method with compressed local texture patches as features. The method is able to reduce the computational overhead of the inspection step, as well as the complexity of the training. The accuracy of the novelty detection results are marginally improved when using compressed $7 \times 7$ sized patches compared to non-compressed $5\times5$ sized patches. Thus, the presented modifications open up the possibility for time-

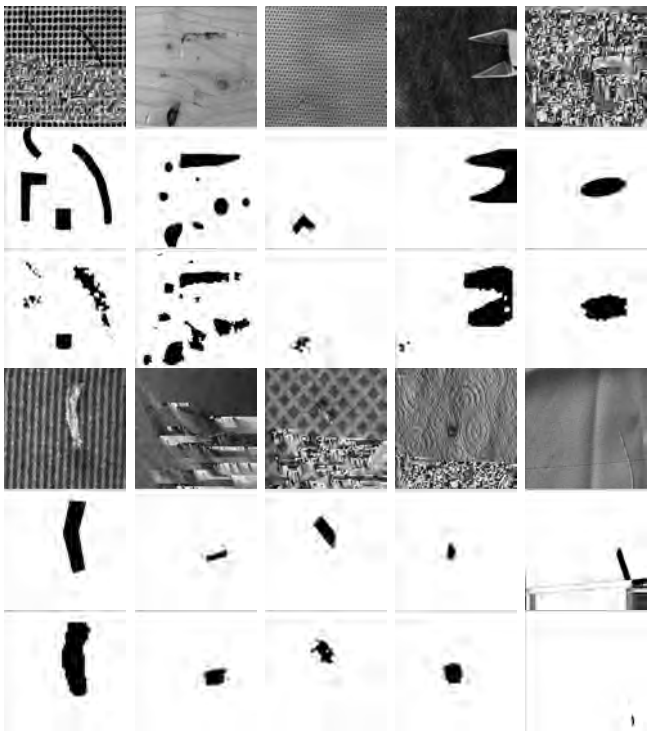


Fig. 4. The defect detection results obtained for a selection of textures within the texture database. The ground truth is displayed in the second and fifth row, respectively. The corresponding results are displayed in the third and sixth row, respectively

critical applications to benefit from the Compressed Sensing TEXEM approach. Furthermore, the compression parameter $m$ enables the user to choose the feature dimension and weigh between accuracy and speed according to the requirements.

To enhance the classification results further, more elaborate compressed features such as sorted random projections or normalized patches could be used. We will investigate the novelty detection for such features within future work.

## REFERENCES

[1] A. D. F. Clarke, "Modelling visual search for surface defects," Ph.D. dissertation, Department of Computer Science, Heriot-Watt University, Edinburgh, 2010.

[2] A. Kumar, "Computer-vision-based fabric defect detection: A survey," *IEEE Trans. on Industrial Electronics*, vol. 55, no. 1, pp. 348–363, 2008.

[3] D. Gibson, M. Spann, and J. Turner, "Automatic fault detection for 3d seismic data," in *Proc. Digital Image Computing: Techniques and Applications*, 2003, pp. 821–830.

[4] C. W. Kim and A. J. Koivo, "Hierarchical classification of surface defects on dusty wood boards," *Pattern Recognition Letters*, vol. 15, no. 7, pp. 713–721, 1994.

[5] O. Silvén, M. Niskanen, and H. Kauppinen, "Wood inspection with non-supervised clustering," *Machine Vision and Applications*, vol. 13, no. 5, pp. 275–285, 2003.

[6] X. Xie and M. Mirmehdi, "TEXEMS: texture exemplars for defect detection on random textured surfaces," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 29, no. 8, pp. 1454–1464, 2007.

[7] D. Mery and M. A. Berti, "Automatic detection of welding defects using texture features," *Insight-Non-Destructive Testing and Condition Monitoring*, vol. 45, no. 10, pp. 676–681, 2003.

[8] V. Leemans and M. F. Destain, "A real-time grading method of apples based on features extracted from defects," *Journal of Food Engineering*, vol. 61, no. 1, pp. 83–89, 2004.

[9] T. Randen and J. H. Husoy, "Filtering for texture classification: A comparative study," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, no. 4, pp. 291–310, 1999.

[10] X. Xie, "A review of recent advances in surface defect detection using texture analysis techniques," *Computer Vision and Image Analysis*, vol. 7, no. 3, pp. 1–22, 2008.

[11] J. Escofet, R. Navarro, M. S. Millan, and J. Pladellorens, "Detection of local defects in textile webs using gabor filters," *Optical Engineering*, vol. 37, no. 8, pp. 2297–2307, 1998.

[12] M. Varma and A. Zisserman, "Texture classification: Are filter banks necessary?" in *IEEE Proc. 2003 Computer Society Conf. on Computer Vision and Pattern Recognition*, vol. 2. IEEE, 2003, pp. II–691–8 vol. 2.

[13] ——, "A statistical approach to texture classification from single images," *International Journal of Computer Vision*, vol. 62, no. 1, pp. 61–81, 2005.

[14] L. Liu and P. Fieguth, "Texture classification from random features," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 34, no. 3, pp. 574–586, 2012.

[15] L. Liu, P. Fieguth, and G. Kuang, "Compressed sensing for robust texture classification," *Computer Vision ACCV 2010*, vol. 6492, pp. 383–396, 2011.

[16] L. Liu, P. Fieguth, G. Kuang, and H. Zha, "Sorted random projections for robust texture classification," in *IEEE Int. Conf. on Computer Vision (ICCV), 2011*. IEEE, 2011, pp. 391–398.

[17] I. Novak and Z. Hocenski, "Texture feature extraction for a visual inspection of ceramic tiles," in *Proc. IEEE Int. Symp. on Industrial Electronics, ISIE 2005*, vol. 3. IEEE, 2005, pp. 1279–1283.

[18] H. Y. T. Ngan, G. K. H. Pang, and N. H. C. Yung, "Automated fabric defect detection–a review," *Image and Vision Computing*, vol. 29, no. 7, pp. 442–458, 2011.

[19] H. Y. T. Ngan, G. K. H. Pang, S. P. Yung, and M. K. Ng, "Wavelet based methods on patterned fabric defect detection," *Pattern Recognition*, vol. 38, no. 4, pp. 559–576, 2005.

[20] M. Ralló, M. S. Millán, and J. Escofet, "Unsupervised local defect segmentation in textures using gabor filters: application to industrial inspection," in *Proc. of SPIE*, vol. 7443, 2009, p. 74431T.

[21] ——, "Unsupervised novelty detection using gabor filters for defect segmentation in textures," *JOSA A*, vol. 26, no. 9, pp. 1967–1976, 2009.

[22] Y. Zhang, C. Yuen, and W. Wong, "A new intelligent fabric defect detection and classification system based on gabor filter and modified elman neural network," *Int. Conf. on Advanced Computer Control (ICACC), 2010*, vol. 2, pp. 652–656, Mar. 2010.

[23] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[24] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[25] X. Mei, H. Ling, Y. Wu, E. Blasch, and L. Bai, "Minimum error bounded efficient 1 tracker with occlusion detection," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2011*. IEEE, 2011, pp. 1257–1264.

[26] L. Liu, P. Fieguth, D. Clausi, and G. Kuang, "Sorted random projections for robust rotation-invariant texture classification," *Pattern Recognition*, vol. 45, no. 6, pp. 2405–2418, 2011.

[27] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[28] ——, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[29] D. L. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[30] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof

of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[31] D. Achlioptas, "Database-friendly random projections," in *Proc. ACM Symp. Principles of Database Systems*, ser. PODS '01. ACM, 2001, pp. 274–281.

[32] S. Dasgupta, "Experiments with random projection," in *Proc. of the Sixteenth conference on Uncertainty in artificial intelligence, UAI '00.* Morgan Kaufmann Publishers Inc., 2000, pp. 143–151.

[33] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal statistical Society*, vol. 39, no. 1, pp. 1–38, 1977.

[34] MIT MediaLab, "VisTex texture database," 1995. [Online]. Available: http://vismod.media.mit.edu/vismod/imagery/VisionTexture/

[35] K. P. Burnham and D. R. Anderson, "Multimodel inference understanding AIC and BIC in model selection," *Sociological methods & research*, vol. 33, no. 2, pp. 261–304, 2004.

Proceedings of the 9th Open German-Russian Workshop on Pattern Recognition and Image Understanding
Dec 1-5, 2014, Koblenz, Germany

# Robust Dynamic Facial Expressions Recognition using LBP-TOP Descriptors and Bag-of-Words Classification Model

Alexey Spizhevoy
Nizhny Novgorod State University and Itseez Inc.
Russian Federation, Nizhny Novogorod
Email: alexey.spizhevoy@itseez.com

***Abstract*—In this work we investigate the problem of robust dynamic facial expression recognition. We develop a complete pipeline that relies on the LBP-TOP descriptors and the Bag-of-Words (BoW) model for basic expressions classification. Experiments performed on the standard dataset such as the Extended Cohn-Kanade (CK+) database show that the developed approach achieves the average recognition rate of 97.7%, thus outperforming the state-of-the-art methods in terms of accuracy. The proposed method is quite robust as it uses only relevant parts of video frames such as areas around mouth, noise, eyes, etc. Ability to work with arbitrary length sequence is also a plus for practical applications, since it means there is no need for complex temporal normalization methods.**

## I. Introduction

The problem of automatic human facial expression recognition by images and/or videos has been attracting growing attention to itself for many years. It's worth paying attention to that the area of dynamic facial expressions recognition hasn't been studied as intensively as the area of analyzing expressions by still images. That's why in our research we intentionally address that problem and compare the proposed approach with the best methods published so far.

Robust solutions able to estimate human expressions accurately by still images are found to be useful in many areas such as human-computer interaction, surveillance monitoring, video content analysis, targeted advertising, entertainment and many others. The same can be said about the methods for recognizing emotions as facial dynamics patterns. Facial expressions are dynamic in nature and can be considered as events consisting of onset, peak and offset phases. That leads to the idea of incorporating facial muscle dynamics into recognition methods to improve their accuracy and robustness. The fact that we're working with human faces narrows and simplifies more general problem of human action recognition. Working with human faces we might pay attention to most informative parts of the face only, i.e. the areas near eyes, nose, mouth, etc. That's also very convenient and useful since facial landmarks detection methods are quite deeply studied and widely used [1], [2].

Our research is application guided mostly and the contribution to the field includes the following points:

1) The proposed approach provides high recognition rate, outperforming on the standard CK+ dataset state-of-the-art methods.
2) The method is a robust solution for automatic dynamic facial expressions recognition that doesn't require any sophisticated normalization techniques that increases the method applicability in practice.

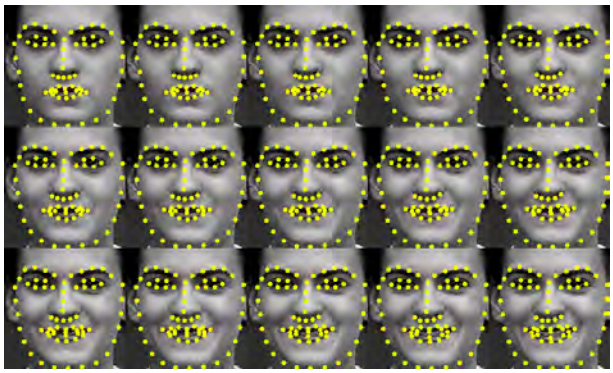


Fig. 1. A sample sequence from the CK+ database with facial landmarks showing the facial expression progression in time.

## II. Related Work

There is a huge variety of methods used for computing descriptors and learning expressions models both static and dynamic. Below we briefly describe recent state-of-the-art works related directly or implicitly to our research.

Bag-of-Words (BoW) models are widely used in text classification, image categorization, object retrieval, and human action recognition [3], [4]. In work [3] the authors propose 3D interest point detector. Descriptors computed by cuboids around detected interest points are used for training BoW model of behavior recognition.

Covariance descriptors proposed in [5] are used together with manifold learning for dimensionality reduction and LogitBoost for classification of dynamic facial expressions. On the CK+ dataset the proposed method achieves accuracy of $92.3\%$.

Temporal Bayesian network is utilized in [6] for capturing facial dynamics and recognizing dynamic facial expressions. On the CK+ dataset the method achieves accuracy of $86.3\%$ only slightly improving the baseline in $83.3\%$ from [7] where AAM features are coupled with SVM model for recognizing expressions.

Latent dynamic conditional random field (LDCRF) is employed in [8] with both shape and appearance features for learning facial expression dynamics. On the CK+ datasets the authors achieve accuracy of $95.79\%$. However according to the paper all the sequences labeled with contempt expression were removed from the original dataset, thus making the problem easier.

The authors of [4] relying on the methods from [3], propose to use LBP-TOP descriptors for human action recognition. We extend their work developing a method that doesn't involves generic interest points detectors. Instead we use the fact we work with faces, so facial landmarks can be incorporated naturally.

In [9] the idea of computing descriptors around facial landmarks is utilized as well. Using AdaBoost model for learning dynamic facial expressions the authors achieve accuracy of $96.32\%$ on the CK+ dataset. In the work only six basic expressions were taken into consideration out of seven expression provided in the CK+ dataset.

In paper [10] the authors propose a method for analyzing facial landmark position and their motion in 3D space. Unfortunately the average area under ROC curve measure employed in that work differs very much from what's widely used and proposed in [7] – the average recognition rate. That's why we can't directly compare performance of that method with our results. However we compare it with results from other recent publications, see table II for details.

## III. LBP-TOP Descriptors

### A. Spatial Descriptors

The local binary pattern on three orthogonal planes (LBP-TOP) descriptor proposed in [11] is an extension of the original LBP operator [12] which has been widely used for computing image descriptors in computer vision. Given an image $I$ the LBP operator is defined as follows:

$$LBP_{P,R_x,R_y} = \sum_{p=0}^{P-1} H(I(x_p, y_p) - I(x, y))2^p, \qquad (1)$$

where $x_p = x + R_x cos(\frac{2p\pi}{P})$, $y_p = y + R_y sin(\frac{2p\pi}{P})$ are the points of neighborhood, and $H(.)$ is step function:

$$H(n) = \begin{cases} 1, & n > 0 \\ 0, & \text{otherwise.} \end{cases} \qquad (2)$$

The LBP operator output is integer number in $[0, 2^P)$ range, where $P$ denotes the number of comparisons, i.e. the number of bits in response value. After the operator is applied to the image $I$, histogram of the response values is computed and used as descriptor. Such procedure discards a lot of spatial information, that's why one often divides the image into blocks that are processed independently. The corresponding histograms are then concatenated into single feature vector, which is used further.

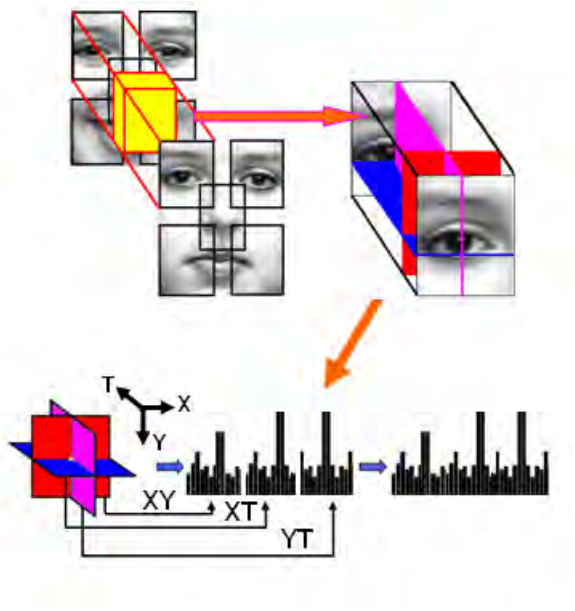



Fig. 2. The LBP operator is applied to three orthogonal planes: XY, XT, and YT. Three corresponding histograms are concatenated into single LBP-TOP descriptor. LBP-TOP descriptors are computed for cuboids around all facial landmarks.

### B. Spatio-Temporal Descriptors

Since the local volumes of video around facial landmarks we're working with are already small enough (see section V-C) we don't split them into blocks. Given a piece of video $V(x, y, t)$, the LBP-TOP descriptor is concatenation of regular LBP descriptors for three orthogonal slices of that volume. We assume that $V$ is a cuboid with center at point $(0, 0, 0)$ and the three orthogonal planes are $V(x, y, 0)$, $V(x, 0, t)$, and $V(0, y, t)$ respectively. Histograms computed on each of those planes using the LBP operator are joined together, forming single feature vector which size $3 * 2^P$.

## IV. Bag-of-Words Classification Model

We employ the Bag-of-Words (BoW) classification model [4], [3] to utilize facial expression dynamics. BoW together with the LBP-TOP descriptors relieves us from involving sophisticated face geometry and illumination normalization methods. We assume that facial landmarks are available for every input frame. While in our case we use the points provided with the CK+ database, facial landmarks can be successfully found in practical applications using such methods, for instance, as Active Shape Models [1], Active Appearance Models [2], shape regression [13].

### A. Vocabulary Construction

During training video cuboids are collected for every frame and every facial landmark. LBP-TOP descriptors computed for those cuboids represent local information used for learning and recognition of facial expressions. Collected descriptors are clustered using the k-means algorithm that finds cluster centers, where the number of the clusters is a parameter of the algorithm. The cluster centers form vocabulary (or bag) of visual words, where each cluster center is just an artificial

descriptor having the same size as the LBP-TOP descriptors computed on cuboids.

In our experiments we construct different visual words for each expression separately and then merge them into single vocabulary $(W_1, W_2, ..., W_{N_{voc}})$. The number of words used for each expression is the same. The final vocabulary size $N_{voc}$ is the total number of visual words constructed for all expressions. Details on how size of per-expression vocabulary affects final accuracy are provided in section V-C.

### B. Features Computation

A video sequence of arbitrary length is converted into fixed length descriptor via matching of the LBP-TOP descriptors extracted from that sequence with visual words from vocabulary. Visual word is considered to be observed if it was picked as closest one from the vocabulary.

Firstly LBP-TOP descriptors $D_i$, $i = 1..N_{descr}$ are computed for input sequence. The total number of descriptors to compute is equal to the number of frames times the number of landmarks per face. Then each descriptor is compared with the visual words from vocabulary and closest ones are chosen:

$$a_i = \underset{j=1..N_{voc}}{\operatorname{argmin}} ||D_i - W_j||_2. \quad (3)$$

Finally histogram of size $N_{voc}$ is constructed as $H(j) = \sum_{i=1}^{N_{descr}} I(a_i = j)$, where $I(.)$ is indicator function, and then the histogram is $L_2$ normalized.

### C. Dimensionality Reduction

The size of histogram is equal to the number of visual words in vocabulary. To achieve high classification accuracy vocabulary must be rich enough to reflect possible local variations. At the same time, having not that many training samples (see section V-A) in the CK+ database, to avoid overfitting related issues one might want to keep only relevant information in BoW features. We achieve that via utilizing Principal Component Analysis (PCA) [14]. The technique is widely used for reducing feature dimensionality via projecting original values onto directions along which variance of data points is highest.

PCA method works as follows. Given $m$ observations $X_i \in \mathbb{R}^N$, $i = 1..m$ of $N$-dimensional random variable $X$ representing feature vector. One can construct sample covariance matrix

$$S = \frac{1}{m-1} \sum_{i=1}^{m} (X_i - \bar{X})(X_i - \bar{X})^T, \quad (4)$$

where $\bar{X}$ denotes sample mean. The eigenvectors $v_j$, $j = 1..n$, where $n << N$ corresponding to largest eigenvalues $\lambda_j$ of matrix $S$, give directions with highest variance. Those directions are used to linearly project original features to subspace of lower dimensionality $n$. In our experiments we kept $95\%$ of original data total energy, i.e. $\sum_{j=1}^{n} \lambda_j \approx 0.95 \sum_{j=1}^{N} \lambda_j$. For more details we refer to [14].

### D. Expression Estimation Algorithm

Normalized histograms as sequence descriptors are used together with non-linear Support Vector Machines (SVM) [15] with Gaussian radial basis function kernel for learning dynamic facial expressions. Given frame sequence the procedure of facial expression estimation follows to the steps provided below:

1) Find facial landmarks in each frame.
2) Compute LBP-TOP descriptors around the landmarks.
3) Match descriptors to visual words from vocabulary.
4) Construct histogram which reflects how often visual words are observed in the sequence.
5) Reduce dimensionality via employing the PCA technique.
6) Estimate facial expression class using a non-linear SVM classifier.

Since each expression has its own "sub-vocabulary", the step of computing histogram can be seen as decomposing or projecting input sequence by basis of expression-specific elements. Intuitively speaking input sequence should give high response for the histogram elements corresponding to the ground truth expression. Accurate relations between the elements of histogram are learned during training SVM classification model.

## V. Experiments

### A. The Extended Cohn-Kanade Dataset

For facial expression validation we used very popular the Extended Cohn-Kanade (CK+) database [7]. The dataset contains 593 face image sequences, from which only 327 are labeled with expression classes. The average sequence length is about 18 frames. Those 327 sequences cover 118 out of 123 persons presented in the database.

TABLE I. Frequency of the stereotypical expressions in the CK+ dataset.

| Expression | # of sequences |
|---|---|
| Anger (An) | 45 |
| Contempt (Co) | 18 |
| Disgust (Di) | 59 |
| Fear (Fe) | 25 |
| Happiness (Ha) | 69 |
| Sadness (Sa) | 28 |
| Surprise (Su) | 83 |

A sample sequence from the dataset is presented in figure 1. Each sequence begins with the neutral face and ends with the peak intensity expression. The database is provided with the 68 face landmarks labeling for each face which we employ in our method. The distribution of facial expression classes in the database is presented in table I.

### B. Results and Comparison with Related Works

For our method we employ 5-fold cross validation (CV) technique to estimate the average recognition rate, i.e. the mean of recognition rates for each expression. Table II shows comparison of our method with the state-of-the-art works that have been published so far. The table demonstrates superiority of the proposed approach over the other methods in terms of the average recognition rate. Moreover some of the methods mentioned in table II were actually evaluated on the subset of the CK+ dataset without the contempt expression sequences which makes the problem easier.

TABLE II. COMPARISON OF OUR APPROACH WITH STATE-OF-THE-ART METHODS IN TERMS OF THE AVERAGE RECOGNITION RATE ON THE CK+ DATASET. THE PROPOSED APPROACH OUTPERFORMS ALL THE CONSIDERED METHODS. THE COLUMN "IS METHOD DYNAMIC?" SHOWS IF THE CORRESPONDING METHOD USES VIDEO SEQUENCE OR ONLY ONE STILL IMAGE.

| Method | Avg. rec. rate, % | Is method dynamic? | Valid. protocol |
|---|---|---|---|
| Baseline [7] | 83.3 | no | LOPO CV |
| Shape+SVM [8] | 84.06 | no | 4-fold CV |
| CSPL [16] | 89.9 | no | 10-fold CV |
| SVM+LBP [17] | 92.6 | no | 10-fold CV |
| ITBN [6] | 86.3 | yes | 15-fold CV |
| Cov3D [5] | 92.3 | yes | 5-fold CV |
| LDCRF [8] | 95.79 | yes | 4-fold CV |
| BCSFMC [9] | 96.32 | yes | 10-fold CV |
| **Ours** | **97.7** | yes | 5-fold CV |

Table II also includes results of state-of-the-art methods that are not dynamic but static, i.e. which work with still images only without taking into account temporal information. Since the CK+ dataset consists of sequences, not independent images, only frames with peak expressions were used to estimate quality of such methods. In average one might conclude that dynamic facial expression recognition methods outperform their static counterparts. That should not be surprising, since incorporating temporal information into features enriches them.

### C. Optimal Parameters Setup

In our experiments we used volumes of $17 \times 17 \times 9$ size in X,Y, and time dimensions correspondingly. The radius in the LBP operator for all dimensions was fixed and set to 3 pixels, the number of comparisons per pixel $P = 8$. The number of clusters in the BoW model vocabulary was set to 400 for each expression (see figure 3 for more details). The k-means clustering was performed for each expression separately.

## VI. CONCLUSION

In this work we developed a complete pipeline for automatic recognition of basic human dynamic facial expressions. Since human expressions are dynamic events in nature, instead of working with still images our method utilizes temporal information. That enriches facial features with additional information compared to single frame methods. The proposed approach works only with relevant areas around mouth, nose, eyes, etc which makes it's less sensitive, i.e. robust to noise and illumination variation in other parts of face and background. Also we employ the LBP-TOP descriptors that itself are quite robust to intensity variations in small cuboids around facial landmarks. Finally using Bag-of-Words model for learning makes our method able to work with arbitrary length sequences, which is quite useful in practical applications. The developed pipeline showed the average recognition rate of $97.7\%$ on the standard CK+ dataset, thus outperforming the state-of-the-art methods published so far.

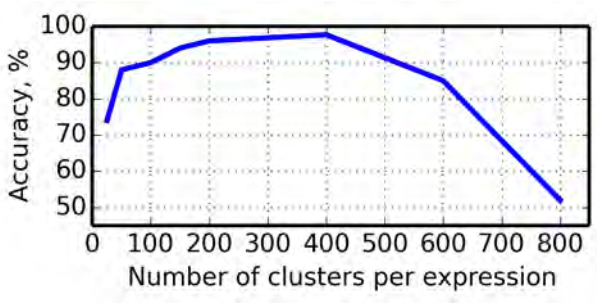



Fig. 3. This figure shows how the number of clusters in the BoW model vocabulary affects the average recognition rate.

## REFERENCES

[1] Timothy F Cootes, Christopher J Taylor, David H Cooper, and Jim Graham, "Active shape models-their training and application," *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[2] Timothy F. Cootes, Gareth J. Edwards, and Christopher J. Taylor, "Active appearance models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, no. 6, pp. 681–685, 2001.

[3] Piotr Dollár, Vincent Rabaud, Garrison Cottrell, and Serge Belongie, "Behavior recognition via sparse spatio-temporal features," in *Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005. 2nd Joint IEEE International Workshop on*. IEEE, 2005, pp. 65–72.

[4] Riccardo Mattivi and Ling Shao, "Human action recognition using lbp-top as sparse spatio-temporal feature descriptor," in *Computer Analysis of Images and Patterns*. Springer, 2009, pp. 740–747.

[5] Andres Sanin, Conrad Sanderson, Mehrtash T Harandi, and Brian C Lovell, "Spatio-temporal covariance descriptors for action and gesture recognition," *arXiv preprint arXiv:1303.6021*, 2013.

[6] Ziheng Wang, Shangfei Wang, and Qiang Ji, "Capturing complex spatio-temporal relations among facial muscles for facial expression recognition," 2013.

[7] Patrick Lucey, Jeffrey F Cohn, Takeo Kanade, Jason Saragih, Zara Ambadar, and Iain Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010, pp. 94–101.

[8] Suyog Jain, Changbo Hu, and Jake K Aggarwal, "Facial expression recognition with temporal modeling of shapes," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1642–1649.

[9] Xiaohua Huang, Guoying Zhao, Matti Pietikäinen, and Wenming Zheng, "Dynamic facial expression recognition using boosted component-based spatiotemporal features and multi-classifier fusion," in *Advanced Concepts for Intelligent Vision Systems*. Springer, 2010, pp. 312–322.

[10] Andras Lorincz, Laszlo Attila Jeni, Zoltán Szabó, Jeffrey F Cohn, and Takeo Kanade, "Emotional expression classification using time-series kernels," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*. IEEE, 2013, pp. 889–895.

[11] Guoying Zhao and Matti Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 6, pp. 915–928, 2007.

[12] Timo Ojala, Matti Pietikainen, and Topi Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.

[13] Shaoqing Ren, Xudong Cao, Yichen Wei, and Jian Sun, "Face alignment at 3000 fps via regressing local binary features," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[14] Hervé Abdi and Lynne J Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[15] Olivier Chapelle, Patrick Haffner, and Vladimir N Vapnik, "Support vector machines for histogram-based image classification," *Neural Networks, IEEE Transactions on*, vol. 10, no. 5, pp. 1055–1064, 1999.

[16] Lin Zhong, Qingshan Liu, Peng Yang, Bo Liu, Junzhou Huang, and Dimitris N Metaxas, "Learning active facial patches for expression analysis," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2562–2569.

[17] Caifeng Shan, Shaogang Gong, and Peter W McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

# Searching for Rotational Symmetries Based on the Gestalt Algebra Operation ∏

Eckart Michaelsen
Fraunhofer-IOSB, Ettlingen, Germany
eckart.michaelsen@iosb.fraunhofer.de

***Abstract*— Among the Gestalt algebra operations the one for rotational symmetries needs a special search strategy. This contribution briefly recalls Gestalt algebra in overview and then concentrates on the rotational "mandalas". Though these are very salient to human perception, they have not been studied much in the classical over hundred years old Gestalt literature. A search procedure is given which can find such patterns in sets of primitive Gestalten extracted from images. Experiments are performed using selected pictures from the 2013 symmetry recognition competition data and SIFT key points as primitive extractor.**

***Keywords—Gestalt perception; combinatorial search; rotational symmetry; visual saliancy;***

## I. GESTALT ALGEBRA

Gestalt Algebra has been introduced as an attempt to capture recursive hierarchies of symmetric patterns in a mathematical setting [1]. Such patterns are obviously salient to the human visual system and probably meaningful for visual recognition tasks such diverse as foreground/background discrimination or automatic understanding of building structure from aerial images of urban terrain. Gestalt Algebra is defined on the Gestalt domain *G* containing 2D position *po*, scale *sc*, orientation *or*, rotational frequency *fr*, and assessment *as*.

$$G = \mathbb{R}^2 \times (0,\infty) \times (\mathbb{R}/\mathbb{Z}) \times \mathbb{N} \times [0,1]$$

Each element g of the domain is called a Gestalt. Figure 1 below shows some example Gestalten displayed on screen. Position and scale are intuitively evident. Assessment is indicated by grey-tone – black=1 being good and white=0 being meaningless (and invisible). Frequency *fr(g)* is indicated as number of spokes of the wheel shape. Thus rotation by *2π/fr(g)* does not change the identity. The orientation attribute gives the angle between the x-axis and the first spoke counterclockwise – value 1 being the maximal angle *2π/fr(g)*. In [1] two operations | and ∑ are defined on the domain one binary and one *n*-ary:

$$|: G^2 \to G: \quad h = g_1 \,|\, g_2$$

$$\Sigma: G^n \to G: \quad h = \sum g_1 ... g_n$$

These are defined for any pair or *n*-tupel of Gestalten respectively, but will yield high assessment values only for Gestalten pairs arranged in mirror-symmetry for |, and rows of Gestalten arranged in good continuation and similar spacing along a row for ∑. Detailed definitions and proofs of algebraic closure can be found in [1] or [2].

## II. RELATED WORK

Related work can be found in textbooks such as [3]. But the topic is being studied for more than a hundred years, with [4] being one of the most known elder examples. In [5] a very interesting approach to a recursive hierarchy for visual objects is given, where emphasis is on assessing the instances (not called Gestalten there) by the minimum description length criterion. Unfortunately, this has not been further pursued. Based on statistical a-contrario tests Gestalt perception is thoroughly investigated in [6]. In that work there are hints on the recursive nature of Gestalten but all examples remain one-step deep. A quite successful example of the published state of rotational symmetry recognition is given with [7]. This approach is based on correlating large numbers of image parts of many scales with other such image parts. It is suitable for mirror- and rotational-symmetries. It is a rather typical example of listing all possible such mappings (discretizing mirror-axes or centers, and scales) and assessing each such hypothesis top-down. Instead the approach given in this contribution constructs a hierarchy of possibilities bottom-up, starting with primitives extracted at interest-point locations in scale-space.

## III. ROTATIONAL SYMMETRIES

For Rotational Symmetries in [2] a third operation ∏ is defined on the Gestalt domain and a proof sketch is given for algebraic closure concerning this operation. Like ∑ it is *n*-ary:

$$\Pi: G^n \to G: \quad h = \prod g_1 ... g_n$$

∏ prefers rotational arrangements with incrementally rising orientations as shown in Fig. 1 below. Fig. 1 also shows an interface meant for testing and understanding Gestalt algebra operations. As mentioned above assessment is displayed as grey tone (black is one = perfect, white is zero = meaningless); the rotational aggregate Gestalt constructed here by operation ∏ of the three smaller parts is not perfect, because the parts are not of equal size, do not perfectly fit into an equilateral triangle, and their orientations do not perfectly increment in $2/3\pi$ steps. Thus it appears in a grey tone. It would be even lighter e.g. if the parts were much further apart (or too close together), or if the orientations would deviate stronger. While there are closed form solutions for | and ∑, the operation ∏ requires an initialization and Newton iteration. For lack of

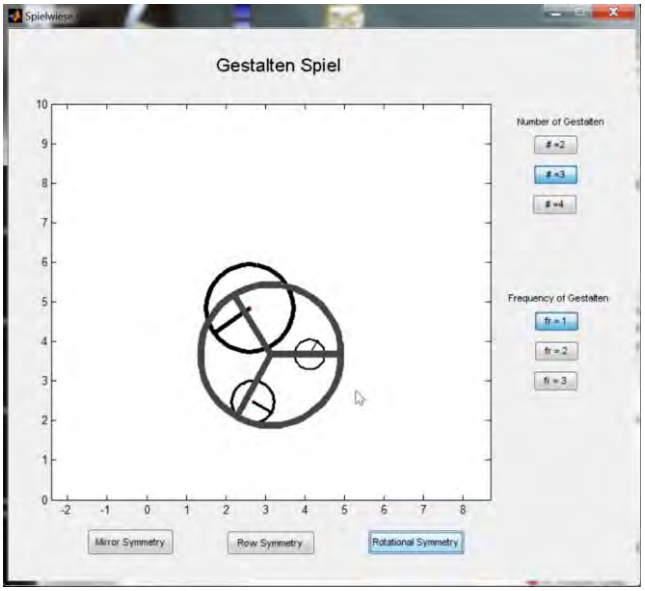



Fig. 1. Interface for interactive Gestalt algebra operations: Three frequency one Gestalten (parts) give a resulting aggregate Gestalt of frequency three

space the technical details – such as the Jacobean used for the iteration – are omitted here with reference to [2].

## IV. Search

Searching for Gestalten with high assessment, with $k$ given primitives from an image, is a task that may demand high computational efforts. Listing all pairs for tests with the mirror operation | is obviously of order $\mathcal{O}(k^2)$. But already for one step using the $n$-ary row operation $\sum$ brute-force exhaustive search would be of order $\mathcal{O}(2^k)$. And that is only for one step. Searching recursively several steps deep for Gestalten with high assessment is more demanding. For practical work up to now heuristics have been used [8] [9]. E.g. for $\sum$ search starts with pairs and tries to prolong the row by appending Gestalten to both ends until the assessment of the aggregate row Gestalt starts dropping. In the following a new rationale for searching for $\prod$ Gestalten is given:

As with $\sum$ search starts with pairs. Preliminarily, only the case $fr(g)=fr(h)=1$ will be handled (e.g. primitive Gestalten). For these the following steps are performed:

- For each pair $(g,h)\epsilon G$ the difference of orientations is investigated in the domain of radians: $d=2\pi(or(h)-or(g))$. In case $d>\pi$ $(h,g)$ will be further investigated instead of $(g,h)$. Given a minimal assessment 1-ϵ for the aggregate Gestalt a number of possible arities $n_1,\ldots,n_m$ is suggested. E.g. for the pair of Gestalten in Fig. 2 $d=2/5\pi$ suggesting arities $n_1=4$, $n_2=5$, and $n_3=6$. A global upper threshold is set for $n_m$ in case $d$ is too small.
- For each arity $n_i$ a center results, indicated as starting point of the polygon in Figure 2. The next two vertices are $po(g)$ and $po(h)$. For $n_i>2$ The following $n_i-2$ vertices give the positions where queries for partner Gestalten are constructed.

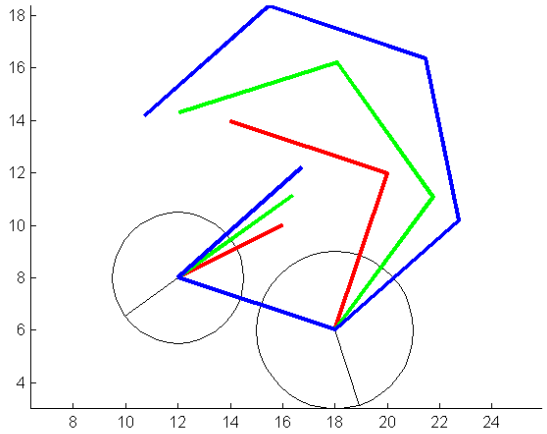



Figure 2: Given a pair (g,h) three orbits are suggested leading to corresponding queries

These queries demand in conjunctive combination: 1) position close to the vertex, 2) orientation in accordance with the orbit run index, and 3) scale compatible with the geometric mean $(sc(g)sc(h))^{0.5}$. Again the minimal assessment $1-\varepsilon$ for the aggregate Gestalt determines the thresholds for the queries. For each of the vertices $2<j\leq n_i$ the Gestalt $f_j$ is chosen that best fits according to three criteria.

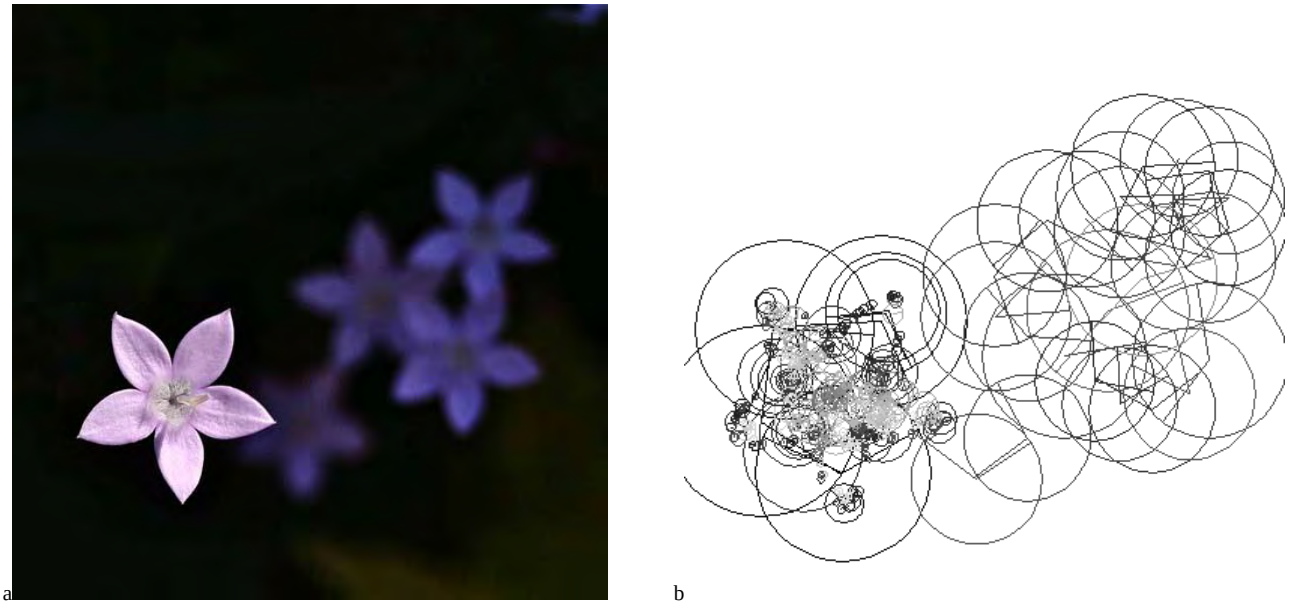
- If the query for one of the vertices $j$ returns an empty set no aggregate will be constructed. Else $\prod gh f_3 \ldots f_{ni}$ is calculated by the mentioned iterative method. This is a greedy search: at most one Gestalt per search region is found. Thus, and because the arities are bounded by $n_m$, the complexity remains polynomial, namely $\mathcal{O}(k^3)$.

Note that arity $2$ is also possible. For this case no search is required. For the case $fr(g)=fr(h)>1$ the search is more complicated. More regular polygons have to be constructed according to the inherent symmetry of such parts.

## V. Experiments

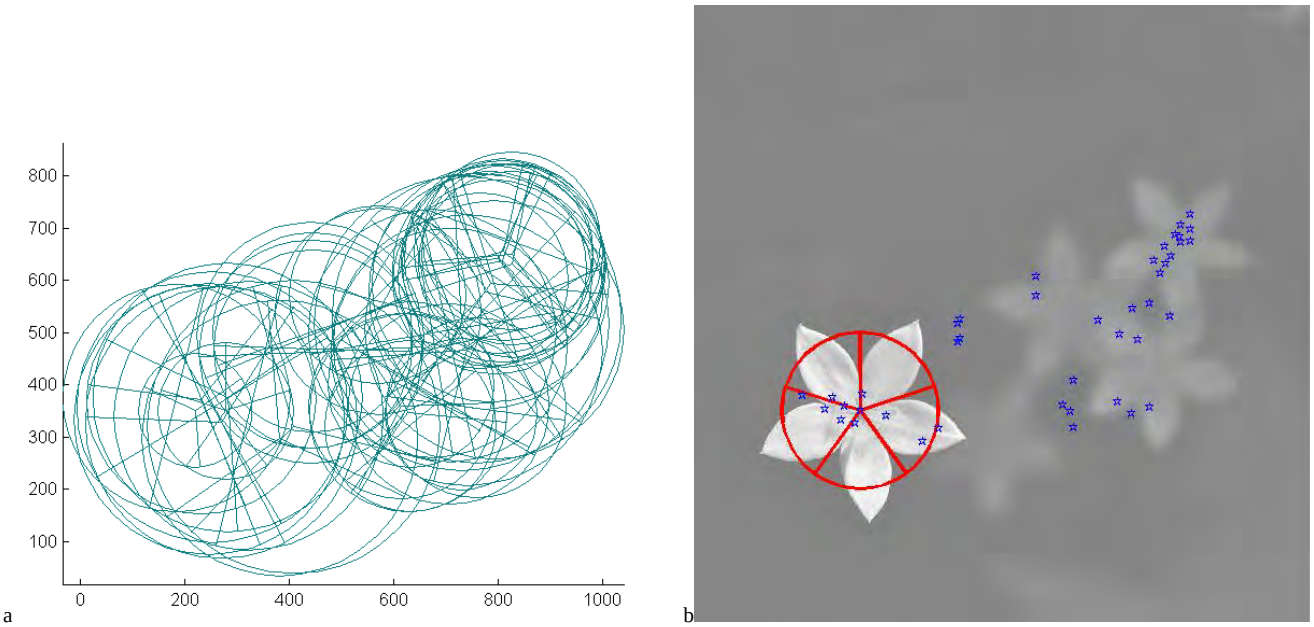
Experiments are made with some selected pictures from the Penn State symmetry recognition data [10]. SIFT key-points are used for the primitives and only to these the search procedure outlined above is applied. Figure 3a shows some example image; and 3b displays the 422 primitive Gestalten of frequency one that are obtained from this image using the standard SIFT key-point extractor with default parameter settings. Figure 4a displays the 41 rotational Gestalten that result from the search procedure outlined above using ε=0.1. In Figure 4b the centers of these Gestalten are overlaid in blue color to a lighter version of the picture.

As usual a cluster analysis has to follow the search. One Gestalt alone is often illusory, but an image structure salient to human observers will often cause a dominant cluster of Gestalten. With rotational Gestalten most of the members found have too low frequency, because only some sub-set of



a b

Fig. 3: Example from the 2013 CVPR competition data: a. original image, b. primitive Gestalten




a b

Fig. 4: Result on the data shown in Figure 3

the correct parts is grouped, or a sub-group in the algebraic sense is established (for non-prime frequencies). Thus we trust most in the maximal frequency found within each cluster. Clustering is seeded by the maximally assessed Gestalt. In the example the corresponding cluster is on the salient flower to the lower left and contains 11 Gestalten. Highest frequency in that cluster is five, and the corresponding Gestalt (draw in red color) almost perfectly fits human perception. After removing this cluster the next best seed Gestalt is to the lower right. It is illusory and draws only few neighbors. The third cluster is located close to the less salient flower in the upper right. It contains only frequencies up to three and is displaced a little. Results on others of the easier images of the set are similar, but most of the images in this collection are too difficult for the Gestalt algebra approach in its current state.

## VI. Conclusion

It may be stated that it is possible to recognize rotational symmetries from natural images using SIFT key-points as primitive Gestalten, the operation $\prod$ as underlying mathematical model, and the indicated new search strategy. However, as already the easy-looking example displayed in Fig. 3 sows, very often some of the *n* parts of a rotational

Gestalt may not manifest as appropriate SIFT-key-points. Carefully looking at Fig. 3b one may notice, that the rightmost primitive of the salient pentagon is actually missing. With the other four parts being arranged perfectly, the search is forced to insert some less well fitting Gestalt in the role of this missing part. The resulting correct Gestalt of frequency five is thus too small in scale and not the best assessed in its cluster. Obviously, strategies will be required for hallucinating missing parts in order to achieve more impressing recognition performance rates.

## REFERENCES

[1] E. Michaelsen and V. Yashina, "Simple Gestalt Algebra," IMTA-4 2013, workshop along with VISAPP 2013 in Barcelona, 2013, pp 38-47.

[2] E. Michaelsen and V. Yashina, "Simple Gestalt Algebra," Pattern Recognition and Image Analysis, 24 (4), (to appear December 2014).

[3] S.J. Dickinson and Z. Pizolo, (eds.), "Shape Perception in Humans and Computer Vision," ACVPR, Springer, London, UK, 2013.

[4] M. Wertheimer, "Untersuchungen zur Lehre der Gestalt," Psychol. Forsch. 4, 1923, pp 301–350.

[5] E. Bienstock, S. Geman, D. Potter, "Compositionality, MDL Priors, and Object Recognition," In: M.C. Mozer, M.I. Jordan, T. Petsche, (Eds.) Advances in Neural Information Processing Systems, MIT Press: Cambridge, MA, USA, 1997, pp 838–844.

[6] A. Desolneux, L. Moisan, J.-M. Morel, "Gestalt theory and computer vision," In: A. Carsetti (Ed.) Seeing, Thinking and Knowing, Kluwer, Dordrecht, The Netherlands, 2004, pp 71–101.

[7] S. Kondra, A. Petrosino, S. Iodice, "Multi-scale kernel operatiors for reflection and rotation symmetry: Further achievements," In: CVPR workshop on Symmetry Detection from Real World Images, 3, 2013.

[8] E. Michaelsen, D. Muench, M. Arens, "Recognition of Symmetry Structure by Use of Gestalt Algebra," In: CVPR workshop on Symmetry Detection from Real World Images, 3, 5, 2013.

[9] E. Michaelsen, "Gestalt Algebra—A Proposal for the Formalization of Gestalt Perception and Rendering," Symmetry, 6, 2014, pp 566-577.

[10] Symmetry Detection from Real World Images—A Competition. Available online (accessed on 6. October 2014).: http://vision.cse.psu.edu/research/symComp13/index.shtml

Proceedings of the 9th Open German-Russian Workshop on Pattern Recognition and Image Understanding
Dec 1-5, 2014, Koblenz, Germany

# Semantic Volume Segmentation with Iterative Context Integration

Sven Sickert, Erik Rodner and Joachim Denzler
Computer Vision Group
Friedrich Schiller University Jena
{sven.sickert, erik.rodner, joachim.denzler}@uni-jena.de

***Abstract*—Automatic recognition of biological structures like membranes or synapses is important to analyze organic processes and to understand their functional behavior. To achieve this, volumetric images taken by electron microscopy or computed tomography have to be segmented into meaningful semantic regions. We are extending iterative context forests which were developed for 2D image data for image stack segmentation. In particular, our method is able to learn high order dependencies and import contextual information, which often can not be learned by conventional Markov random field approaches usually used for this task. Our method is tested for very different and challenging medical and biological segmentation tasks.**

## I. Introduction

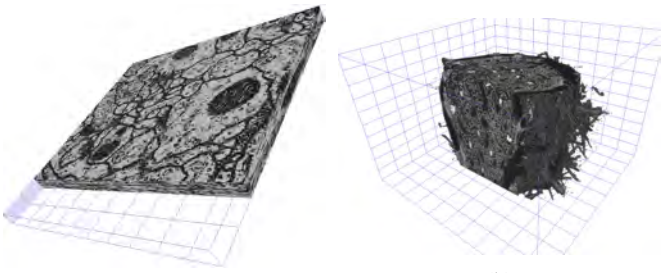
Extracting specific regions in volume images is an important task in medical image processing and a prerequisite for quantitative analysis of biological data (Fig. 1). For example, X-ray computed tomography (CT) is a common tool in medical imaging and the electron microscopy (EM) is used to investigate biological processes or structures in organic and inorganic specimens. However, manual annotation of 3D data is challenging and often requires a huge amount of time and expertise. Therefore, the aim of the paper is to present a method, which is easy to implement, to use, and can be applied to generic volume data by learning from a small number of previously annotated volumes. Being able to automatically segment semantic regions in volume data does not only save time, but it also allows for quantitatively analyzing a large number of volumes, important for providing applied researchers with robust statistics.

Our proposed method labels every single voxel in the volume and is able to capture context information along every axis of the volume. We show that the use of relatively simple feature extraction methods on channels for color and gradient values is sufficient to obtain a decent segmentation of volume images.

The outline of the paper is as follows: In Section II, we give an overview of related work. Section III and IV continue with a description and discussion of the proposed method. In Section V, we present segmentation results for different data sets. We conclude with a discussion of the results in Section VI.

## II. Related Work

The task of semantic volume segmentation for image data provided by CT imagery or EM microscopy is of great interest in computer vision [1], [2], [3]. Related works most often propose segmentation techniques to find and reconstruct objects of a specific class, *e.g.*, synapses or membranes in neural tissue. A popular tool for segmentation is the graph-cut algorithm. The method in [3] uses a gradient flux term in the energy function and incorporates information from different sections by applying the SIFT flow algorithm [4]. In [1] a modified energy function is presented which omits the gradient flux term. Instead, perceptual grouping constraints for contour completion are introduced. It can be useful for the segmentation of thin elongated structures. However, the authors argue that this term may also lead to false positive membrane segmentations due to textures. Instead the final energy function consists of a data term, a directional energy term for smoothness, a penalty for discontinuities, and a term that incorporates information from adjacent sections.



(a) (b)

Fig. 1: Volume images used in our experiments: (a) an electron microscopy stack of neural tissue and (b) a CT scan of a sponge

In [2], the correspondence between nearby areas is transformed into a fusion problem. First, an RDF classifier is trained to obtain the probability of each pixel belonging to the cell boundary. Watershed transformation is applied to obtain segmentations for each section. After that, 3D links between these sections should connect segments of different slices. Finally, a fusion problem of 2D segments and 3D links is defined that identifies each neuron.

CT images typically show objects of a larger scale. As the authors of [5] show, semantic segmentation can also help with localizing organs in a human body. The authors are using so-called entangled-decision forests to model context between nearby organs in CT scans which usually have a fixed composition.

In our approach we do not model the connection between different sections explicitly, but use *auto-context* features [6]

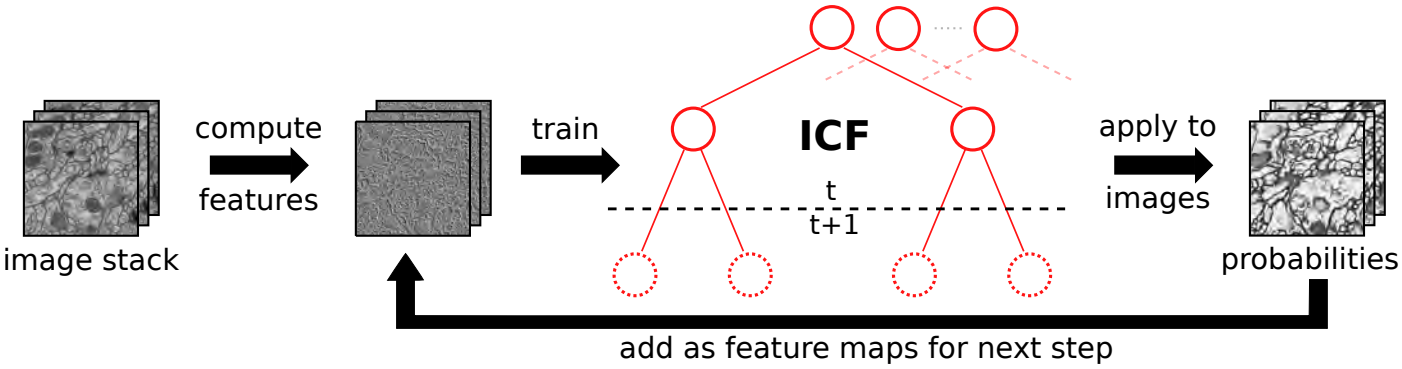



Fig. 2: Framework of our method: First, slices of a image stack are processed in order to compute feature maps. Feature extraction methods applied to these maps are used to train the ICF by finding *good* splits at the current level $t$. The ICF learned so far is applied to the training slices in order to create class probabilities for each pixel. Then, these probability maps are add to the pool of feature maps in the training step $t + 1$.

within an iterative pixelwise labeling approach. Moreover, we are using easy to compute features like gray-values and gradients in a cubic neighborhood. Therefore, we extend the iterative context forests (ICF) of [7] to 3D data which was previously limited to 2D images and multi-channel images occurring in remote sensing applications.

More common applications for semantic segmentation are urban scenes and images of landscape, persons, or animals. In these settings typical constellations like *car on street*, *sky above building* or *sheep on grass* are observable as well. Many methods in this field of application are using either random decision forests (RDF) [8], [9], [10], [7] or conditional random fields (CRFs) [11], [12], [13] which is a common technique to model context. As we show in this paper, in contrast to CRFs, our approach is not restricted to pairwise potentials and is able to learn high order dependencies.

## III. Iterative Context Forests for Image Stacks

The proposed approach for semantic segmentation consists of two fundamental parts: feature extraction and pixel classification. In this section, we focus on the classification framework which is based on the popular random decision forests. Details on the efficient computation of features will follow in the next section.

### A. Random Decision Forests

The core of ICF [7] is a random decision forest (RDF) [14], which has been specifically adapted to semantic segmentation. The concept of decision trees is well known, so we will not elaborate on how they work and refer the interested reader to [14]. In order to explain iterative context forests, we give a short introduction to random decision forests. The RDF approach in general aims at overcoming drawbacks of original decision trees, like over-fitting and long training times, with two concepts of massive randomization during training.

The first concept is known as bagging and trains several decision trees individually with a random subset of the training data. Furthermore, a second randomization concept is applied determining binary splits in inner nodes. Instead of computing all available features in each inner node only a random subset is drawn from the set of all possible features, which we will refer to as *feature pool*. Among them the best feature and split is determined by maximizing the impurity criterion, which in our case is the information gain.

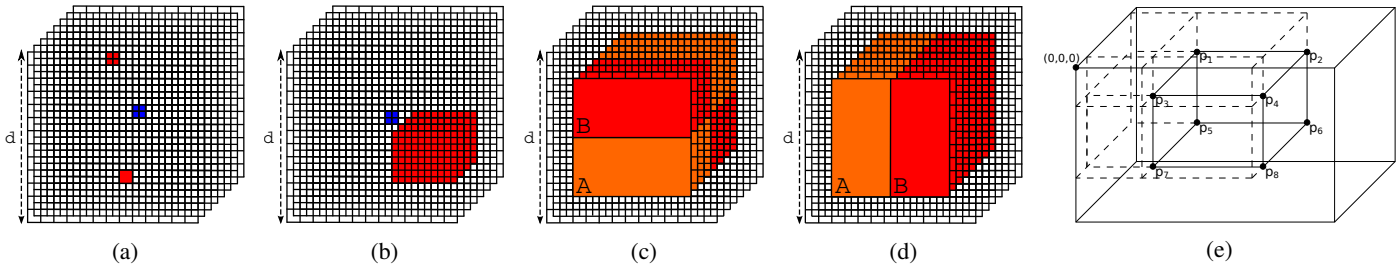
In our case, we are confronted with different types of features (see Sec. IV for more details) with each of them having a number of parameters (*e.g.*, position, used feature map, *etc.*). To sample a specific feature, we first sample the feature type and we then sample the parameters in a second step. This guarantees that the learning is not biased towards feature types with a larger parameter space. The whole random selection process is exactly what renders learning in our case with millions of features tractable.

Although finding a good split in a single tree node makes it necessary to test various randomly chosen splits, the training of random decision forests is still fast. Learning an RDF is a matter of minutes, while convolutional neural networks for instance need several days on a GPU for such a task [15]. The classification of test images is even faster. Each pixel of an example is traversing the trees of the trained RDF until it reaches the leaf nodes. The empirical distributions in all leaves are then combined in order to estimate class-wise probabilities.

### B. Incorporating Context Knowledge

Applying RDFs for pixel- or voxelwise classification directly has two disadvantages: first, for each pixel the tree has to be traversed down to the leafs, and second, feature extraction is limited to a local neighborhood and is unable to integrate high-order dependencies.

To address the second issue, [7] proposed to sequentially traverse the tree level by level for all pixels in each image. This allows for using outputs of the previous level as an additional source for features. At each level of the random forest class probability maps for the current image (or volume) are computed. On a lower level these information allow the extraction of contextual features in order to model relational dependencies like one class is above another. This concept was introduced in [6] and is called *auto-context*. See Fig. 2 for an overview of our our pixel-wise classification framework.




(a) (b) (c) (d) (e)

Fig. 3: Feature extraction methods for a 3D neighborhood of side length $d$ around a voxel (blue): (a) pairs of voxels, (b) a smaller cuboid of arbitrary size or (c-d) different types of 3D Haar-like features where the feature value is the difference of the voxel intensity sums of red and orange regions. (e) Eight references are necessary to compute the sum of an arbitrary cuboid within the integral volume image.

Furthermore, during testing time the structure of ICFs also addresses a trade-off between accuracy and time needed for inference. In some applications it might be sufficient to have a rough approximation of the best obtainable result after a short time. Since ICFs are built in breadth-first manner, a prediction on each level of the trees can be done by returning the empirical class distribution stored at inner nodes.

As observed in our experiments, each depth-level gives a more accurate classification result than the one before until saturation is reached. Stopping at an earlier level will give only rough results depending on fewer features but will also save time. The ability of an algorithm to allow for iterative refinements during testing is usually referred to as *anytime classification* [16].

## IV. Efficient Features in Volumetric Images with ICF3D

In this section, we describe how computation of specific features is achieved. In the first part we broach the issue of creating a large feature pool with millions of possible features. After that, a fast method for computing such features is presented.

### A. Creating a large feature pool

The ICF classifier allows inputs to have an arbitrary number of channels. Depending on the application, the algorithm can make use not only of raw color channels but additional layers like gradient images, resulting probabilities of other pre-learned classifiers, or unsupervised segmentation outputs. The relevance of these *feature maps* is automatically determined by the RDF classifier, since all binary splits are evaluated by an impurity measure and automatically selected.

The extraction of features from these maps is done in our case with simple operations performed in a neighborhood of the current center pixel: (a) Single values extracted from neighboring voxels, (b) sum or difference of two neighboring voxels, (c) sum of values within a cuboid and (d) 3D Haar-like features given by the difference of two or more cuboids. Visual examples can be found in Fig. 3.

With the size of the neighborhood the feature pool grows exponentially and it is not possible to compute all of the possible features in the training step. As a consequence, we only draw a fixed number of features. From this subset of features, the best split in each level of a decision tree is chosen by the impurity criterion.

### B. Fast computation of 3D features

In order to compute 3D features in a fast manner, it is possible to use an intermediate representation. In the case of 2D images the so called *integral image* (or *summed area table*) is used [17], [18]. It contains at location $(x, y)$ the sum of the pixel values above and to the left of the same location in the original image . The computation of rectangle features as for instance Haar-like features benefit from this representation. The sum of a rectangle area of arbitrary size can be computed by only four points in the integral image.

This idea can be easily extended to *integral volumes*. In consequence, an integral volume image $\mathcal{V}$ contains at location $(x, y, z)$ the sum of gray values from the voxels above, to the left and in front of its location in the input image $\mathcal{I}$ as well as the gray value of the voxel $(x, y, z)$ itself:

$$\mathcal{V}(x, y, z) = \sum_{x' \leq x,\ y' \leq y,\ z' \leq z} \mathcal{I}(x', y', z') \quad . \tag{1}$$

With the integral volume of an image, any sum of values in a cuboid can be be computed by eight array references. Analogously to [18], the values of the cuboids to the left, on top and in front of the currently processed volume have to be removed. Otherwise these region would be incorporated twice in the final computation. Consider a cuboid $\mathcal{C}$ with an arbitrary position inside a larger cuboid, *i.e.,* the whole integral volume image $V$. The eight corners of $\mathcal{C}$ are $p_1, ..., p_8$ with $p_i = (x_i, y_i, z_i)$ being the locations in $V$. Location $p_6$ marks the lower right back corner and $p_2$ the upper right front corner of $\mathcal{C}$. Then, the integral volume $\mathcal{V}_\mathcal{C}$ of $\mathcal{C}$ can be computed as:

$$\mathcal{V}_\mathcal{C} = p_1 + p_4 + p_6 + p_7 - p_2 - p_3 - p_5 - p_8 \quad . \tag{2}$$

A corresponding visualization showing the eight reference points can be found in Fig. 3(e). With this method cuboid features and especially the 3D Haar-like features can be computed very efficiently. Now that we have extended the ICF framework with 3D features and added the possibility to analyze volume images, we will henceforth call it ICF3D.

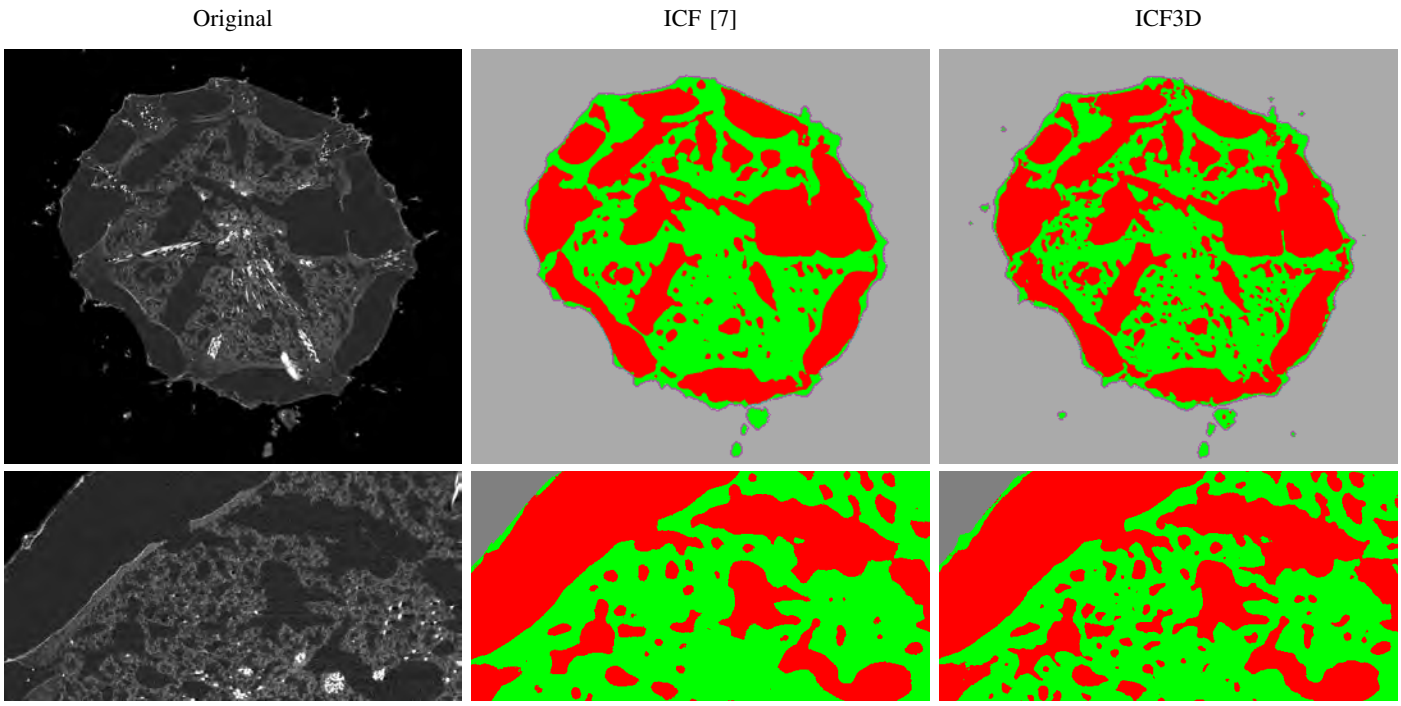



Fig. 4: Results for the sponge CT data: The first row depicts slice #85 of the testing stack and corresponding segmentation results by ICF [7] and ICF3D with labeled classes *canal* (red), *tissue* (green) and *background* (gray). Second row shows an enhanced comparison of slice #259. Our ICF3D approach improves segmentation results of subtle canal structures. Figures are best viewed in color.

## V. Experiments

We evaluate our method on two different datasets. For a qualitative demonstration of ICF3D, we use data of a CT scan. A quantitative comparison with other approaches using an EM stack of neural tissue is done after that.

### A. CT data of a sponge

We are using CT scanning data of a sponge with an isotropic resolution of 3.4 micrometers per pixel. The resolution of each slice is $1536 \times 1536$ pixels. The task is to distinguish between the classes *canals*, *tissue*, and *background*. There is a stack of eleven labeled slices and a stack of 351 unlabeled slices. We used the former stack for training and the second one for testing. All images feature noise and artifacts from the recording procedure. The training slices were accurately labeled by a biologist. In Fig. 4 (a) you can find a typical image. According to the expert, all dark gray structures belong to the class *canals*, even the tiny ones. A magnified part of another slice can be seen in Fig. 4 (d).

Because the resolution of this data is in all dimensions the same, a 2D window of the neighborhood for each pixel can be simply replaced by a cube with the same side length. In consequence, we can run the experiments with ICF [7] and ICF3D with comparable parameter configurations. We use feature maps originating from gray values and gradients. Color information is not available for this data.

Figure 4 demonstrates the potential of our volume segmentation method. Images (c-e) and (f-h) each show a slice of the test stack and the voxel classification results of [7] and ICF3D. As can be seen our approach shows improvements when it comes to subtle structures. Without using 3D information from nearby slices the segmentation is more coarse and fine structures get lost.

### B. EM stack of neuronal tissue

We use the freely available *Drosophila* first instar larva ventral nerve cord (VNC) data [19] of the ISBI 2012 challenge. A stack with 30 slices is labeled to distinguish between the classes *membrane* and *interior* of neurons. The resolution of these stacks is $4 \times 4 \times 50$ nanometers, which is rather coarse in z-direction.

To account for this setting, we adapt the neighborhood in z-direction in an analogous way. The length of the neighborhood cuboid in z-direction is only 10% of the length in the other two directions, which are of higher resolution. For instance, a neighborhood of size $d = 50$ represents a $50 \times 50 \times 5$ cuboid in the image volume. In consequence, only two slices in front of and behind the current slice are used in the process.

We report results for the measures *pixel error* and *rand error* and compare our performance with selected methods of the challenge. The former one is a common measure in binary classification of pixels and is also referred to as accuracy or overall recognition rate. The second measure is related to the well known F-score, which is the harmonic mean of precision and recall. However, in this specific task the so-called *Rand index* is used for the computation. It is a measure of similarity

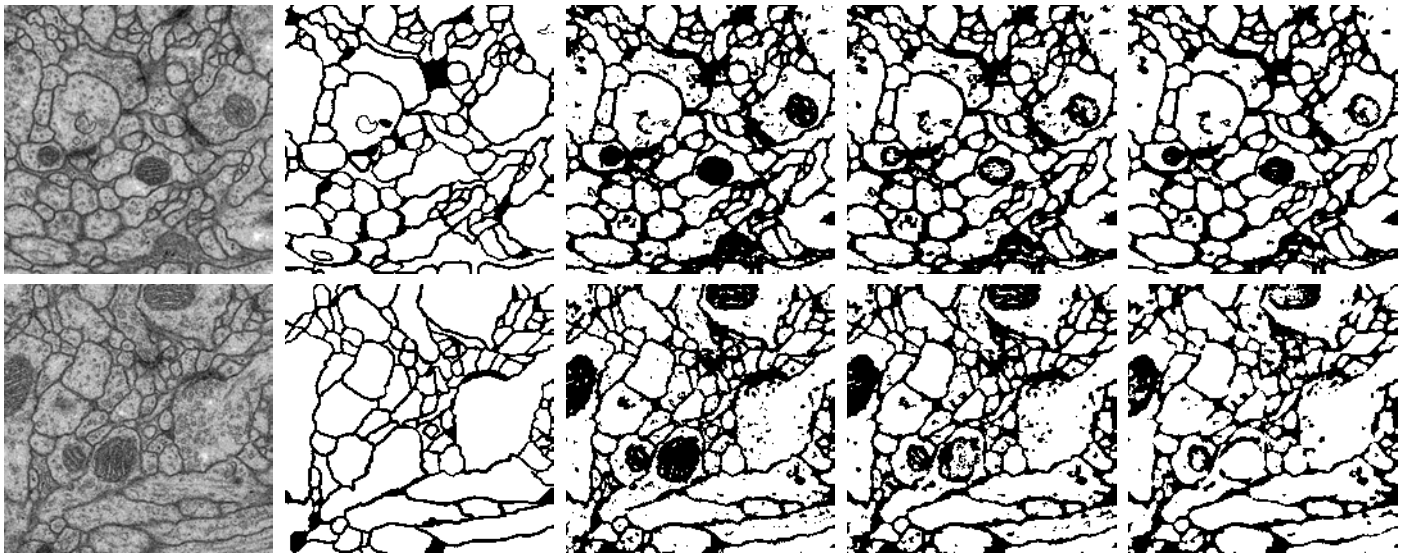



Fig. 5: Qualitative comparison of the neural tissue dataset using different types for the feature extraction. When only using pixel pair features the results appear to be cluttered. Especially in the larger interior areas the labels of nearby pixels are not consistent. Furthermore, synapses are labeled as cell membrane. These issues lessen or even vanish when Haar-like features and context features are added.

between two clusters or segmentations. The most important measure in the ISBI 2012 challenge is the rand error and we therefore also sort our results by this measure (see Table I).

In order to model the textural differences between membrane and mitochondria in the data not visible in the gradients maps, we incorporate local binary patterns (LBP) [20] as additional feature maps. They are a powerful tool in texture classification and increased the performance slightly in our case.

As can be seen in the rand error column, we are not able to produce state-of-the-art results. When taking the actual pixel errors into consideration ICF3D performs decently. It is a typical binary segmentation task, where modeling of context is hardly possible. However, we achieve comparable results to [1] and the patch-based SVM of [21]. The convolutional network approach of [15] shows best performance for this data. Note, that we are not using task specific knowledge (*e.g.* shape information) or applying any post-processing for smoothing.

| **Method** | **Rand error** $[\cdot 10^{-2}]$ | **Pixel error** $[\cdot 10^{-2}]$ |
|---|---|---|
| CNN [15] | 4.8 | 6.0 |
| Dense correspondence [3] | 6.4 | 8,3 |
| Watershed Tree [22] | 8.4 | 13.4 |
| Perceptual Grouping [1] | 8.4 | 15.7 |
| CellProfiler [23] | 9.0 | 10.0 |
| Segment features [24] | 13.9 | 10.2 |
| Two-step class. [25] | 15.3 | 8.8 |
| Contextual Grouping [26] | 16.2 | 10.9 |
| Patch-based SVM [21] | 23.0 | 15.0 |
| ICF [7] | 28.1 | 13.5 |
| **Ours:** ICF3D w/o LBP | 24.1 | 12.4 |
| **Ours:** ICF3D w/ LBP | 22.9 | 12.4 |

**Table I**: Results of some competitors and our generic method on the ISBI 2012 challenge data [19].

### *C. Analysis of feature types*

In a third set of experiments, we analyzed the influence of different feature types (see Sec. IV-A) for the segmentation of neural tissue. For this task, we only used the training data of the ISBI 2012 challenge data, because ground truth labels are available. We split the 30 slices of the training stack into six stacks of five slices each to do a 6-fold cross-validation. This allowed us to analyze performance without using the evaluation server.

In Fig. 5, we show a qualitative comparison of two slices taken from two different cross-validation runs. While the first two columns depict original input images and their corresponding ground truth annotation the other three columns show results with different feature extraction methods. It can be seen that the use of pixel pair differences is sufficient to segment cell membranes. However, synapses are also labeled as membrane and areas of cell interior are not very homogeneous. This is due to the limitation of single pixel values instead of average values in feature maps across whole regions.

When rectangle features and Haar-like features are added, some parts of the synapses are not labeled as membrane anymore. Furthermore, the inhomogeneity in the cell interiors is less. Both issues even improve when context features taken

from probability maps are incorporated and some wrong labeled synapses even vanish.

For a quantitative evaluation we used the provided script of the ISBI 2012 challenge. When only pixel pair differences are allowed the pixel error is 15.6% (average of 6-fold cross-validation). With incorporated rectangle and Haar-like features the error decreases to 14.2%. The best result of 13.5% with respect to the pixel error can be achieved when context features are used.

## VI. Conclusions

In this paper, we presented a fast method for volume image segmentation. We extended an existing semantic segmentation method that incorporates context information. Our proposed algorithm is able to segment even subtle structures in CT data. The method is also applicable to EM stacks, as we have shown for the ISBI 2012 challenge data.

We also did some more experiments with data of EM stacks and CT scans and discovered that incorporating information from nearby slices can also be misleading. Whether segmentation performance improves with these information highly depends on the quality of the data and on the recording method itself.

While CT imagery usually creates volume images with isotropic resolution and similar illumination at each slice, EM stacks often have a coarse resolution in z-dimension as well as illumination and contrast changes. On the other hand, CT scans usually contain artifacts or noise. These things have to be considered when working with such stacks to attain improvements over sequential 2D slice segmentation.

## Acknowledgment

The authors would like to thank Henry Jahn and Jörg U. Hammel of the *Porifera.net Lab* at the *Friedrich Schiller University Jena* for providing pixel-wise labeled CT imagery data of sponges.

## References

[1] V. Kaynig, T. Fuchs, and J. Buhmann, "Neuron geometry extraction by perceptual grouping in sstem images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 2902–2909.

[2] A. Vazquez-Reina, M. Gelbart, D. Huang, J. Lichtman, E. Miller, and H. Pfister, "Segmentation fusion for connectomics," in *Proceedings of the International Conference on Computer Vision (ICPR)*, 2011, pp. 177–184.

[3] D. Laptev, A. Vezhnevets, S. Dwivedi, and J. Buhmann, "Anisotropic sstem image segmentation using dense correspondence across sections," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2012, pp. 323–330.

[4] L. Ce, , J. Yuen, and A. Torralba, "Sift flow: Dense correspondence across scenes and its applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 33, pp. 978–994, 2011.

[5] A. Montillo, J. Shotton, J. Winn, J. Iglesias, D. Metaxas, and A. Criminisi, "Entangled decision forests and their application for semantic segmentation of ct images," in *Proceedings of the International Conference on Information Processing in Medical Imaging (IPMI)*, 2011, pp. 184–196.

[6] Z. Tu and X. Bai, "Auto-context and its application to high-level vision tasks and 3d brain image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 32, no. 10, pp. 1744–1757, 2010.

[7] B. Fröhlich, E. Rodner, and J. Denzler, "Semantic segmentation with millions of features: Integrating multiple cues in a combined random forest approach," in *Proceedings of the Asian Conference on Computer Vision (ACCV)*, 2012, pp. 218–231.

[8] G. Brostow, J. Shotton, J. Fauqueur, and R. Cipolla, "Segmentation and recognition using structure from motion point clouds." in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2008, pp. 44–57.

[9] C. Zhang, L. Wang, and R. Yang, "Semantic segmentation of urban scenes using dense depth maps." in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2010, pp. 708–721.

[10] B. Fröhlich, E. Rodner, and J. Denzler, "A fast approach for pixelwise labeling of facade images," in *Proceedings of the International Conference on Pattern Recognition (ICPR)*, 2010, pp. 3029–3032.

[11] C. Galleguillos, A. Rabinovich, and S. Belongie, "Object categorization using co-occurrence, location and appearance." in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[12] L. Ladicky, C. Russell, P. Kohli, and P. Torr, "Graph cut based inference with co-occurrence statistics," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2010, pp. 239–253.

[13] X. Boix, J. Gonfaus, J. van de Weijer, A. Bagdanov, J. Serrat, and J. GonzÃ lez, "Harmony potentials - fusing global and local scale for semantic image segmentation," *International Journal of Computer Vision (IJCV)*, vol. 96, no. 1, pp. 83–102, 2012.

[14] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[15] D. C. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Advances in Neural Information Processing Systems (NIPS)*, 2012, pp. 2852–2860.

[16] S. Esmeir and S. Markovitch, "Anytime learning of anycost classifiers," *Machine Learning*, vol. 82, no. 3, pp. 445–473, 2011.

[17] F. C. Crow, "Summed-area tables for texture mapping," in *International Conference and Exhibition on Computer Graphics and Interactive Techniques (SIGGRAPH)*, 1984, pp. 207–212.

[18] P. Viola and M. Jones, "Robust real-time object detection," *International Journal of Computer Vision (IJCV)*, vol. 57, pp. 137–154, 2002.

[19] A. Cardona, S. Saalfeld, S. Preibisch, B. Schmid, A. Cheng, J. Pulokas, P. Tomancak, and V. Hartenstein, "An integrated micro- and macroarchitectural analysis of the drosophila brain by computer-assisted serial section electron microscopy," *PLoS Biology*, vol. 8, no. 10, 2010.

[20] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 24, no. 7, pp. 971–987, 2002.

[21] S. Iftikhar and A. Godil, "The detection of neuronal structure using patch-based multi-features and support vector machines algorithm," in *Proceedings of ISBI 2012 EM Segmentation Challenge*, 2012.

[22] T. Liu, M. Seyedhosseini, E. Jurrus, and T. Tasdizen, "Neuron segmentation in em images using series of classifiers and watershed tree," in *Proceedings of ISBI 2012 EM Segmentation Challenge*, 2012.

[23] L. Kamentsky, "Segmentation of em images of neuronal structures using cellprofiler," in *Proceedings of ISBI 2012 EM Segmentation Challenge*, 2012.

[24] R. Burget, V. Uher, and J. Masek, "Trainable segmentation based on local-level and segment-level feature extraction," in *Proceedings of ISBI 2012 EM Segmentation Challenge*, 2012.

[25] X. Tan and C. Sun, "Membrane extraction using two-step classification and post-processing," in *Proceedings of ISBI 2012 EM Segmentation Challenge*, 2012.

[26] E. Bas, M. G. Uzunbas, D. Metaxas, and E. Myers, "Contextual grouping in a concept: a multistage decision strategy for em segmentation," in *Proceedings of ISBI 2012 EM Segmentation Challenge*, 2012.

# SEMI-AUTOMATIC LIVER SEGMENTATION USING TV-L$^1$ DENOISING AND REGION GROWING WITH CONSTRAINTS

A. Nikonorov, P. Yakimov
Samara State Aerospace University
line 2-name of organization, acronyms acceptable
Samara, Russia
artniko@gmail.com, pavel.y.yakimov@gmail.com

A. Kolsanov, S. Chaplygin,
A. Ivaschenko, Y. Yuzifovich
Samara State Medical University
Samara, Russia
sergion163@gmail.com, ivachencoaveg@rambler.ru,
yuriyvyuzifovich@gmail.com, info@samsmu.ru

***Abstract*— We present a technique for liver segmentation to automate liver volumetric analysis, a critical preparation step of hepatic surgeries. To increase the accuracy of segmentation on noisy data, the image is enhanced by TV-L$^1$ filtering. Our modified region growing algorithm accounts for the distance between image points and seed points. Two hepatic surgeries benefited from better procedure planning using our technique.**

***Keywords—liver volumetry, image segmentation, TV-L1 filtering, CT data processing***

## I. INTRODUCTION

Liver volumetry is a critical aspect of safe hepatic surgeries. Various methods exist for automatic segmentation of anatomy structures using CT scanning data. [1]. Liver volume measurement is routinely done by manual tracing of liver boundaries in all axial slices of the CT image. This time-consuming procedure requires a trained professional and can produce inconsistent results by different experts.

Despite recent advances [2], automated liver segmentation in CT scans remains a challenging task. Delineating liver tissue from surrounding organs is particularly difficult. For example, chest muscles have similar intensity in CT scans and located too close to the liver to make correct automatic discrimination (Fig.1). Complex liver shape variations among patients, the presence of tumors and large blood vessels within the liver boundaries, and image noise also affect segmentation accuracy.

Concerns about overexposure to radiation due to CT scanning [7] resulted in using lower radiation doses, producing higher levels of image noise and degrading diagnostic performance. Available techniques for controlling noise in CT can be categorized into three main classes: projection space denoising (PSDN), image-space denoising (ISD), and iterative reconstruction (IR) [6].

We propose a two-stage technique that improves the accuracy of semi-automatic segmentation on noisy data, while reducing computational complexity of the segmentation. Filtering stage is based on TV-L1 minimization while segmentation stage is based on a new algorithm that takes advantage of user defined areas in CT data.

Fig. 1. Noisy CT data.

## II. OPTIMAL FIRST ORDER PRIMAL-DUAL CONVEX OPTIMIZATION FOR CT IMAGES DENOISING

To de-noise 3D CT data, we developed optimization methods based on the optimal first-order primal-dual framework by Chambolle and Pock [5].

Let X and Y be the finite-dimensional real vector spaces for the primal and dual space, respectively. Consider the following operators and functions:

$\mathbf{K}: X \to Y$ is a linear operator from X to Y;

$\mathbf{G}: X \to [0, +\infty)$ is a proper, convex, (l.s.c.) function;

$\mathbf{F}: Y \to [0, +\infty)$ is a proper, convex, (l.s.c.) function;

where l.s.c. stands for lower-semicontinuous.

The optimization framework [5] considers general problems in the following form:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \mathbf{F}(\mathbf{K}(\mathbf{x})) + \mathbf{G}(\mathbf{x}) \qquad (1)$$

This work was supported by sMedX LLC (www.smedx.com)

To solve this problem, the following algorithm is described in paper [5]. During initialization, $\tau, \sigma \in R_+$ are set, $\theta \in [0,1]$, $(\mathbf{x}_0, \mathbf{y}_0) \in \mathrm{X} \times \mathrm{Y}$ is some initial approximation, $\bar{\mathbf{x}}_0 = \mathbf{x}_0$. For 3D CT data, the final result obtained on the previous slice is used as the initial approximation for the next slice. With *n* as the current step number, $n \geq 0$, $\mathbf{x}_n, \mathbf{y}_n, \bar{\mathbf{x}}_n$ are iteratively updated as follows:

$$\mathbf{y}_{n+1} = prox_{\sigma F^*}(\mathbf{y}_n + \sigma \mathbf{K} \bar{\mathbf{x}}_n), \tag{2}$$

$$\mathbf{x}_{n+1} = prox_{\tau \mathbf{G}}(\mathbf{x}_n + \tau \mathbf{K}^* \mathbf{y}_{n+1}), \tag{3}$$

$$\bar{\mathbf{x}}_{n+1} = \mathbf{x}_{n+1} + \theta(\mathbf{x}_{n+1} - \mathbf{x}_n). \tag{4}$$

The proximal operator with respect to **G** in (3), is defined as:

$$prox_{\tau \mathbf{G}}(\tilde{\mathbf{x}}) = (\mathbf{E} + \tau \partial \mathbf{G})^{-1}(\tilde{\mathbf{x}}) = \arg\min_{\mathbf{x}} \frac{1}{2\tau} \|\mathbf{x} - \tilde{\mathbf{x}}\|_2^2 + \mathbf{G}(\mathbf{x}) \tag{5}$$

where **E** is an identity matrix. The proximal operator $prox_{\sigma F^*}$ (2) is defined in similar way.

In the case of 3D CT data, we use approximation for $\partial$ operator which differs from the operator in paper [5]:

$$\nabla(\mathrm{u})_{i,j,k} = \begin{pmatrix} (\nabla(\mathrm{u}))^1_{i,j,k} \\ (\nabla(\mathrm{u}))^2_{i,j,k} \\ (\nabla(\mathrm{u}))^3_{i,j,k} \end{pmatrix}, \tag{6}$$

$$(\nabla(\mathrm{u}))^1_{i,j,k} = \begin{cases} u_{i+1,j,k} - u_{i,j,k}, & i < I \\ 0 & i = I \end{cases}, \tag{7}$$

$$(\nabla(\mathrm{u}))^2_{i,j,k} = \begin{cases} u_{i,j+1,k} - u_{i,j+1,k}, & j < J \\ 0 & j = J \end{cases}, \tag{8}$$

$$(\nabla(\mathrm{u}))^3_{i,j,k} = \begin{cases} u_{i,j,k+1} - u_{i,j,k}, & k < K \\ 0 & k = K \end{cases}. \tag{9}$$

*k* is the index of an axial slice of 3D CT image, *I*, *J*, *K* are boundaries.

The model of denoising is based on the total variance approach [5] and is described by the following functional:

$$\min_{\mathbf{u} \in \mathbf{X}} \|\nabla \mathbf{p}\|_1 + \lambda \|\mathbf{p}_0 - \mathbf{p}\|_1 \tag{10}$$

where $\| \ \|_1$ is the robust $L_1$ norm, $\mathbf{p}_0$ is the source noisy image, $\mathbf{p}_1$ is the target filtered image and $\lambda$ is the weighting parameter, which defines the tradeoff between regularization and data fitting.

In order to apply the described algorithm to (10), we follow the [5]:

$$G(\mathbf{p}) = \|\nabla \mathbf{p}\|_1, \tag{11}$$

$$F^*(\mathbf{p}) = \|\mathbf{p}_0 - \mathbf{p}\|_1, \tag{12}$$

Finally, using (11) and (12) proximal operators for steps (2) and (3) of the algorithm can be obtained. Please refer to [5] for further details. The denoising algorithm based on total variance is able to preserve sharp edges. Also, using the $L_1$ makes it possible to efficiently remove strong outliers.

Fig.2. shows the result of the TV-$L^1$ filtration compared to the filters implemented in the Insight Segmentation and Registration Toolkit (ITK). The implementations of bilateral filter (Fig. 2c) and curvature anisotropic diffusion filter (Fig. 2a) can be found in ITK library [3].

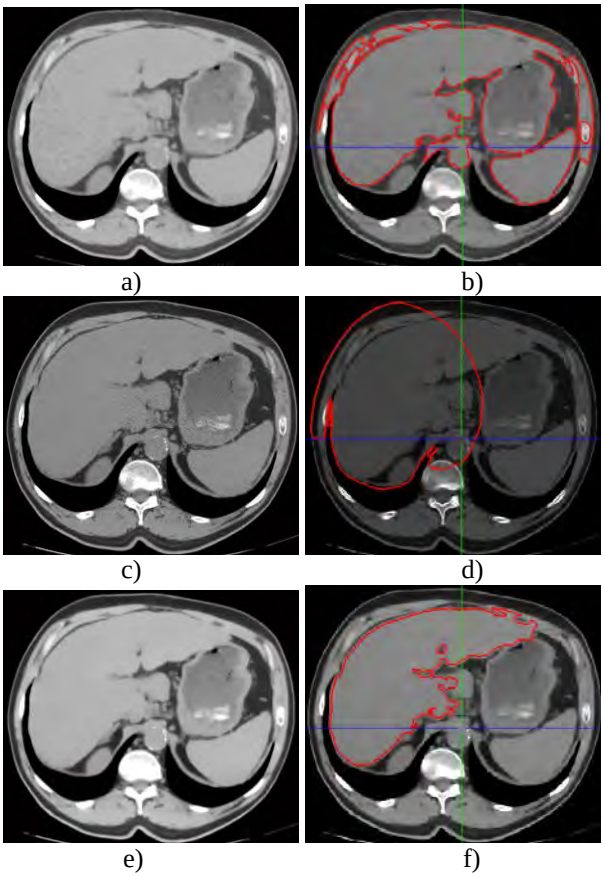



Fig. 2. Filtered images and images with segmentation result using curvature aniotropic diffusion filter (a), (b); bilateral filtering (c),(d); TV-$L^1$ filtering (e), (f).

The following parameters were used for bilateral filter: domain sigma of 7, range sigma of 7; and for curvature anisotropic diffusion [4]: time step of 0.09, 8 iterations and a conductance value of 3.0. The segmentation results in Fig. 2b,

2d and 2f were obtained using fast marching and region growing algorithms from ITK.

As shown in Fig. 2, the total variance algorithm using $L_1$ norm gives smoother result than other denoising algorithms.

## III. Constrained Region Growing Algorithm

After the original data is pre-processed and filtered, one of the most popular methods for liver segmentation is fast marching [3]. While fast marching algorithm works well with small areas, it is computationally intensive, which results in excessive time to process a whole liver, have a variety of parameters that need to be tuned correctly, and is unreliable in limiting segmentation propagation to the neighboring organs with similar CT intensities. Another simple segmentation method is a region growing algorithm, which does not require as many parameters as the fast marching algorithm does. An ability to segment the whole liver with a single seed point makes this algorithm easier to use. However, it can still occasionally propagate beyond the liver into surrounding tissue, including kidney and heart (Fig. 2b).

We propose a modified region growing algorithm for semi-automatic liver segmentation with additional constraint points. The standard region growing algorithm included in ITK does not take into account the distance between image points and a seed point. For example, an image point can be occasionally marked as located within the liver even if the distance is as high as 50 cm. To take into account typical dimensions of the liver, we introduced a so-called potential plane ***P*** (Fig. 3a, 3b). This is a matrix of the same dimensions as the original image with floating-point values between 0 and 1, representing a potential plane to be used when applying a dynamic threshold. We initialize it with round areas of specific radius around each seed point. The dynamic threshold value is then multiplied by the value of the potential plane point at the same point. Resulting constrained region growing is shown in Fig. 3e.

To prevent mis-classification of chest muscle tissue as liver, we apply an additional constraint to the potential plane. The red circle in Fig. 3d makes the potential plane as shown in Fig. 3b. The segmentation result in Fig. 3f shows correct classification for all points. Green and red circles in Fig. 3c and 3d show the position of potential plane points in relation to the original data.

When the algorithm is applied to 3D data, the circles in the potential plane become spheres. The 3D recursive region growing algorithm also uses points from adjacent slices.

## IV. Experimental Results

The results obtained using TV-$L^1$ filtration and the modified region growing algorithm were compared to Sliver 7 database [8]. We used manually segmented images from training sets to compute the following criterion:

$$Q = \frac{V(\mathbf{p}_{AD} \cdot I_s)}{V(\mathbf{p} \cdot I_s)}, \qquad (13)$$

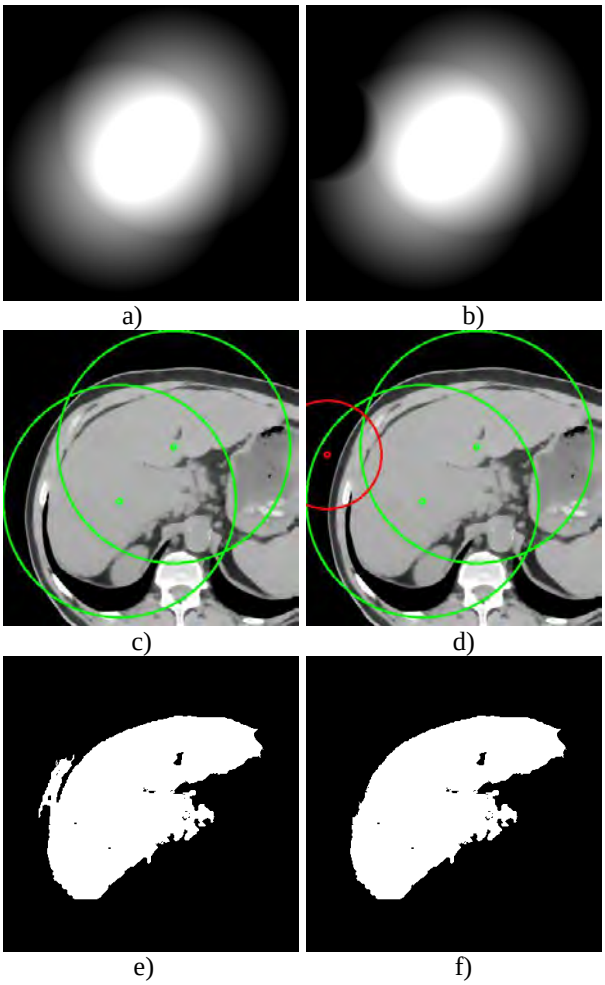



Fig. 3. Potential planes (a),(b); original data with seed points (c), (d); segmentation results with extra points (e) and without extra points (f)

where $V$ is variance estimation, $\mathbf{p}_{AD}$ is anisotropic diffusion filtering result, $\mathbf{p}$ is a result for filtering using (10), $I_s$ is an indicator function which takes its values from inside the pre-segmented area. We obtained $Q$ for all Sliver 7 training sets. The average $Q$ value is 1.4318, the minimum and maximum values are 1.0514 and 2.2864, respectively. Image denoising improved segmentation accuracy of the constrained region growing algorithm by 74%, which was tested using the relative volume difference metric [8]. These experimental results show that our proposed filtering and segmentation algorithms perform better than the algorithms implemented in ITK do [4].

## V. Conclusion

We proposed a two-step filtering and segmentation technique for processing CT data to obtain liver segmentation. TV-$L^1$ filtration used as the first step performs with a better accuracy than standard algorithms implemented in the ITK framework do. A modified region growing segmentation algorithm provides a powerful and easy to use tool for semi-automatic liver segmentation. Using our technique, we created

several 3D liver models that included inner structure. These models were successfully used for visual guidance during two live hepatic resection surgeries performed at Samara State Medical University Hospital.

Further research is needed to refine suggested algorithms. With a more diverse selection of patients, we can test the algorithm robustness to liver variations with and without pathology. To test how hardware-invariant our algorithm is, we can apply it to images taken by various CT scanners. Additional optimization is needed to process internal liver structures, such as blood vessel and bile ducts. Before we can recommend a wide adoption in medical imaging, we also need to improve setup procedures to automate parameter generation to accommodate different scanning resolution and quality.

## REFERENCES

[1] Tibamoso, G. Semi-automatic Liver Segmentation From Computed Tomography (CT) Scans based on Deformable Surfaces / Gerardo Tibamoso, Andrea Rueda // Url: http://sliver07.org/data/2010-07-26-1926.pdf, P. 5

[2] Lim, S. Automatic liver segmentation for volume measurement in CT Images / Seong-Jae Lim, Yong-Yeon Jeong, Yo-Sung Ho // Journal of Visual Communication and Image Representation, Volume 17, Issue 4, August 2006, Pages 860-875

[3] Johnson, H. The ITK Software Guide Third Edition Updated for ITK version 4.5 / Hans J. Johnson, Matt McCormick, Luis Ibanez // Insight Software Consortium, December 17, 2013

[4] Terry S. Yoo Insight Into Images Principles and Practice for Segmentation // Registration, and Image Analysis, Massachusetts, 2004

[5] Chambolle, A. A first-order primal-dual algorithm for convex problems with applications to imaging / Chambolle, A. Pock, T. // J. Math. Imaging Vis. 40, 2011, p. 120–145.

[6] Li, Z. Adaptive nonlocal means filtering based on local noise level for CT denoising / Zhoubo Li, Lifeng Yu, Joshua D. Trzasko, David S. Lake, Daniel J. Blezek, Joel G. Fletcher, Cynthia H. McCollough, and Armando Manduca, // 011908-1 Med. Phys. 41 (1), January 2014.

[7] Brenner, D. Computed tomography: An increasing source of radiation exposure / D. J. Brenner and E. J. Hall, // New Engl. J. Med. 357, 2277–2284 (2007).

[8] Heimann, T. Comparison and Evaluation of Methods for Liver Segmentation from CT datasets // IEEE Transactions on Medical Imaging, volume 28, number 8, pp. 1251-1265, 2009.

Proceedings of the 9th Open German-Russian Workshop on Pattern Recognition and Image Understanding
Dec 1-5, 2014, Koblenz, Germany



# Semiautomatic Quantitative Evaluation of Micro-CT Data

Miroslav Jirik, Jiri Kunes, Milos Zelezny

***Abstract*—Quantitative analysis of histology slides can bring unique knowledge about the investigated sample. Unfortunately this is time consuming procedure. In this paper we sugesst method to overcome this disadvantage. However, everything has its price. Semi-automatic evaluation cannot beat human operator by its precision, but it is able to process big amount of data in short time. In some fields it can be useful property.**

***Keywords*—histology, micro CT, liver**

## I. Introduction

Quantitative analysis of histology slides can bring unique knowledge about the investigated sample. Unfortunately this is time consuming procedure. In this paper we suggest method to overcome this disadvantage. However, everything has its price. Semi-automatic evaluation cannot beat human operator by its precision, but it is able to process big amount of data in short time. In some fields it can be useful property.

This work is motivated by cooperation with the Faculty of Medicine in Pilsen. Our goal is to estimate the possibilities of regenerative potential of liver parenchyma. Crucial step in this process is description of liver microvasculature. Data are obtained from Micro CT of corrosion cast of pig liver. Manual evaluation of small sample (2x2x2cm) take more than one week to the operator.

Based on former work [1] we developed an semi-automatic application for evaluation of Micro CT data. Our software is distribuated with application Lisa and the implementation is freely available [2]

## II. Methods

The original manual methods consist of slowly evaluating data slice by slice. When calculating values like volume density or surface density, every slice of data is marked with set number of evenly distributed points. Then the values can be estimated by counting number of points that are inside of vessels. Its a lot more difficult to calculate the other values because they require detailed tracking of every vessel through many different slices of data. The detailed explanations of these methods can be found in [3] and [4].

M. Jirik is with the Department of Cybernetics of Faculty of Applied Sciences, University of West Bohemia, Pilsen, Czech Republic. e-mail: mjirik@kky.zcu.cz

J., Kunes is with the Department of Cybernetics of Faculty of Applied Sciences, University of West Bohemia, Pilsen, Czech Republic. e-mail: jirka642@students.zcu.cz

M., Zelezny is with the Department of Cybernetics of Faculty of Applied Sciences, University of West Bohemia, Pilsen, Czech Republic. e-mail: mzelezny@kky.zcu.cz

As it was already written, the manual methods are very slow. Thats why we decided to try to automatize this process by using methods of automatic image processing.

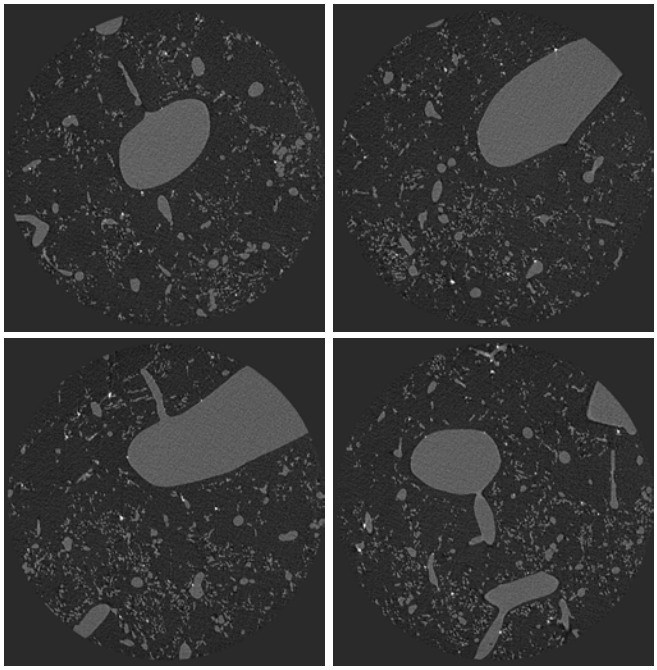


Fig. 1. Sample Micro CT slices of corrosive cast of pig liver with resolution $0.004682 \times 0.004682 \times 0.004682$ [mm] and size $992 \times 1013$

### A. Data segmentation

The input data of our application can be medical DICOM or any image format supported by library SimpleITK [5] which is python wrapper for Insight Toolkit library. After loading them, data can be cropped to the area of interest, subsequently its possible to remove subareas that are shouldnt be counted in statistical results. Finally the median filter is used to remove noise, this filter was chosen because our algorithm requires that the transition between vessels and background is not blurred.

After pre-processing step follows segmentation whose threshold can be set manually or graphically by marking few point that are inside a vessel. Segmentation is based on seeded region growing algorithm [6]. Resulting segmented data are then deprived of holes and other segmentation errors by using adjustable number of dilation and erosion iterations.

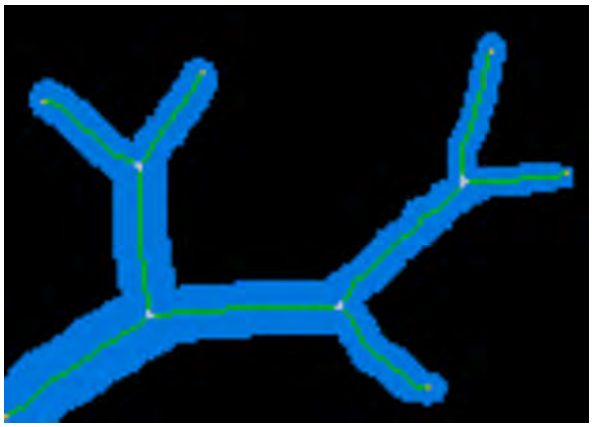


Fig. 2. 2D example of skeletonized vessel

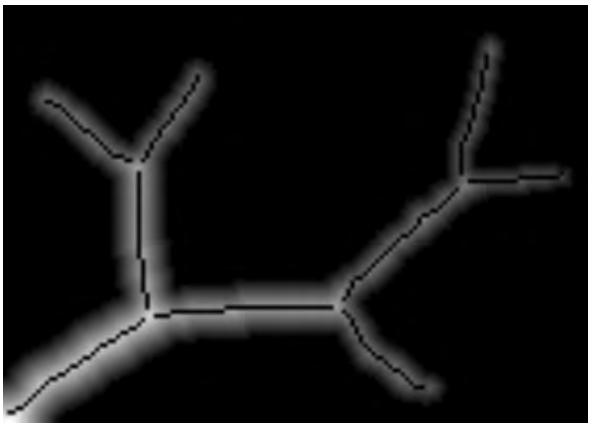


Fig. 3. Distance trasform in vessels and skeleton line (black)

### *B. Skeletonization and skeleton analysis*

The segmented volumetric data that were produced are processed using our thinning3D library ([7] to skeleton. Based on convolution with 3x3x3 kernel an identification of joints and other topological parts of skeleton is performed. Princip of identification is simple. If there is continuous skeleton line with no junction there are only 2 neighbors voxels of skeleton for examined voxel. In 26-connected neighborhood is suma computed by convolution equal to 3. However, if there is is higher suma we can assume that there is jucntion point. On the other hand, if the sume is lower then 3, there is ending point of skeleton. An 2D example of vessel tree with skeleton can be seen on image 2. By removing junctions the skeleton is separated into not connected areas. Each area represents one segment of vessel. This image is labeled with connected-component labeling algorithm. It makes each segment identifable.

Then its possible to to create 1D graph representation of connections between skelets of every vessel. 1D model is describing vessel tree topology and geometry. It can be used for describing blod flow through veins [8]. Finally the graph representation is checked for vessel that are not connected to anything and other errors that are mostly caused by problems with skeleton.

Important feature of vessel segment is its radius. We estimate radius by following algorithm. The volume vessel tree image is inverted. Vessels are labeled as "background" and other space is labeled as "object". Then a distance transform is performed, as can be seen in figure 3. In this image there are shown with black line the vessel segments. There are taken only the values lying on segment skeleton from the distance image and this values. The mean from values along the skeleton segment is used as radius estimation.

### *C. Quantitative evaluation*

After obtaining segmented volumetric data and 1D graphical representation of vessel tree made from its skeleton, we proceed to the actual histological analysis.

The volume density can be calculated from the volumetric data as the number of voxels inside the vessels divided by the total number of image voxels.

Next step is the processing of the statistical values of the individual vessels. Since we have skeleton for every vessel that should look like or be very similar to a generic curve, we are able to easily estimate its length. It is done counting number of pixels with examinated label. This number is then multiplied by size of voxel given by mean of voxel size along all dimensions. It can be done without loss of precision only for data with equal sizes for all dimensions. In our case we had data with cubic voxels. Otherwise this assumption can be satisfied by data resampling.

Tortuosity of vessel is then calculated as length ($L$) divided by distance from end points of vessel ($C$).

$$\tau = \frac{L}{C} \tag{1}$$

Tortuosity that was calculated this way should have very similar value to to tortuosity that was calculated by algorithm Tort3D [9], because the curve we got from skeleton and curve that we would have gained by fitting it to centroids from Tort3D would be very similar. After processing every vessel, overall tortuosity can be calculated as a average tortuosity from every vessel.

Volume fraction can be calculated as fraction of estimated volume of the microvessels $V$ and total volume $V(ref)$

$$V_V = \frac{V}{V(ref)} \tag{2}$$

Length density ($L_V$) is calculated as sum of all lengths divided by total volume of analysed sample (equation 3)

$$L_V = \frac{L}{V(ref)} \tag{3}$$

We developed a graphical user interface based on PyQt4 library. On image 4 it can be found GUI with results visualization.

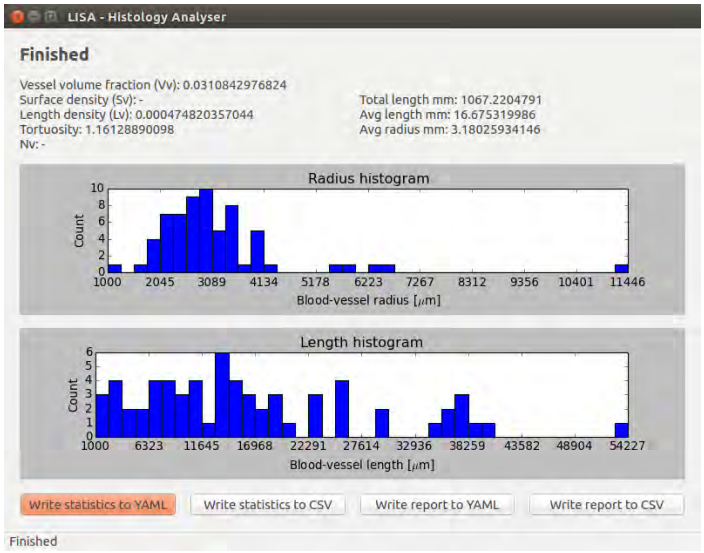



Fig. 4. Histology Analyser: Graphical User Interface

The resulting histological statistics for whole data and each individual vessels can be exported to csv or yaml format. You can see example of yaml output file on image 5.

## III. Results and Discussion

For the algorithm evaluation we organised a comparative experiment. We used corosive cast of pig liver which were scanned by micro-CT machine with resolution $0.004682 \times 0.004682 \times 0.004682$ [mm]. Our dataset contains a set of six 3D images. These images with hepatic lobules were selected by the experts.

In table I can be found measures obtained by standart manual quantitative evaluation of our data. Table II contain measurement obtained by our software.

As can be seen from table above there is correlation between reference and our automatic quantification. Correlation coefficient for Vessel volume fraction(Vv), Length density (Lv) and Tortuosity is 0.9126346766, 0.9682207442 and 0.06632676542 respectively. Average error for Lv and Vv is 15.86 and 19.67 percent. Error of Tortuosity is 2.85 but its correlation is low.

| Data # | Vv (volume) | Lv (length density) | Tortuosity |
|---|---|---|---|
| 1 | 0,025338776 | 61,744621037 | 1,1303067565 |
| 2 | 0,036310204 | 80,5774779501 | 1,1392404180 |
| 3 | 0,025077551 | 100,4722112532 | 1,1614828584 |
| 4 | 0,056685714 | 81,665038706 | 1,2472470664 |
| 5 | 0,013061224 | 46,213109164 | 1,1501091607 |
| 6 | 0,03082449 | 50,8722333122 | 1,1312067130 |

TABLE I. Manually obtained results

| Data # | Vv (volume) | Lv (length density) | Tortuosity |
|---|---|---|---|
| 1 | 0,0251985 | 69,95580 | 1,19787150 |
| 2 | 0,0232613 | 91,87886 | 1,15993166 |
| 3 | 0,0239793 | 131,47988 | 1,16394949 |
| 4 | 0,0611915 | 110,34549 | 1,17642350 |
| 5 | 0,0174375 | 62,97628 | 1,15692600 |
| 6 | 0,0312831 | 61,87116 | 1,16437544 |

TABLE II. Results computed by our algorithm

Due to low contrast in fine vessels of our corrosion casts there is easy to lost accidentally some vessel connection. Skeleton constructed from segmented data can be topologically not perfect. This is weakness of proposed solution. We would like to make segmentation more robust in the future.

## IV. Conclusion

We developed application for quantitative evaluation of the corrosion casts. As the results show the human operator is irreplaceable with our software for this time. Especially for sensitive evaluation in field of quantitative histology. Nevertheless, evaluation 3D histology images is time consuming. For big amount of data can be automatic algorithm the only one sensible solution.

We would like to work on bifurcation angle analysis in the future.

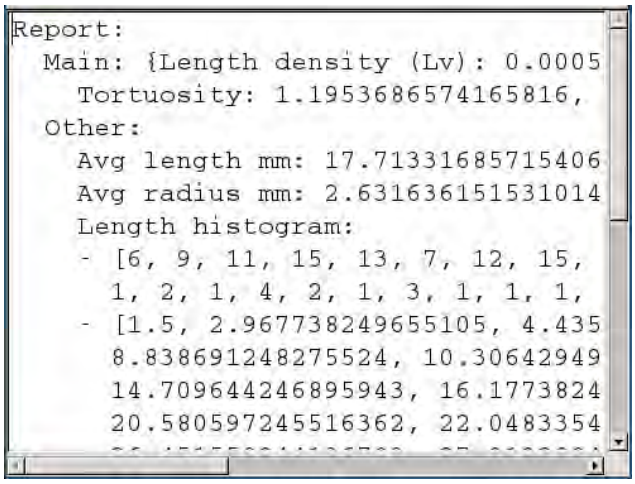



Fig. 5. Example of output YAML report

## Acknowledgment

The work has been supported by the grant of The University of West Bohemia, project number SGS-2013-032 and GRANT IGA MZ R 13326 2012-2015 and postdoctoral project LFP04 CZ.1.07/2.3.00/30.0061

## References

[1] T. Gregor, P. Kochová, L. Eberlová, L. Nedorost, E. Prosecká, V. Liška, H. Mírka, D. Kachlík, I. Pirner, P. Zimmermann, and Others, "Correlating Micro-CT Imaging with Quantitative Histology," 2012.

[2] M. Jirik, V. Lukes, P. Volkovinsky, M. Klima, P. Neduchal, and J. Kunes, "LISA - LIver Surgery Analyser." [Online]. Available: https://github.com/mjirik/lisa

[3] A. Baddeley and E. B. V. Jensen, *Stereology for statisticians / Adrian J. Baddeley and Eva B. Vedel Jensen*. Chapman & Hall/CRC Boca Raton, Fla. ; London, 2005.

[4] C. B. Saper, "Unbiased Stereology: Three-Dimensional Measurement in Microscopy by C.V. Howard and M.G. Reed," *Trends in Neurosciences*, vol. 22, no. 2, pp. 94–95, 1999. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016622369801368X

[5] B. C. Lowekamp, D. T. Chen, L. Ibáñez, and D. Blezek, "The Design of SimpleITK." *Frontiers in neuroinformatics*, vol. 7, no. December, p. 45, Jan. 2013. [Online]. Available: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3874546\&tool=pmcentrez\&rendertype=abstract

[6] R. Adams and L. Bischof, "Seeded region growing," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 16, no. 6, pp. 641–647, 1994.

[7] M. Jirik, "skelet3D," 2014. [Online]. Available: http://github.com/mjirik/skelet3d

[8] A. Jonášová, E. Rohan, V. Lukeš, and O. Bublík, "Complex Hierarchical Modeling of the Dynamic Perfusion Test: Application to Liver," *wccm-eccm-ecfd2014.org*, no. Wccm Xi, pp. 1–12, 2014. [Online]. Available: http://www.wccm-eccm-ecfd2014.org/admin/files/filePaper/p2911.pdf

[9] M. Kutay, "Modeling moisture transport in asphalt pavements," *Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki*, 2005. [Online]. Available: http://scholar.google.com/scholar?hl=en\&btnG=Search\&q=intitle:No+Title\#0http://drum.lib.umd.edu/handle/1903/2911

**Miroslav Jirik** Ing. Miroslav Jiřik was born in Klatovy, Czech Republic in 1984. He received his Bc. and Ing. (similar to M.S.) degrees in cybernetics from the University of West Bohemia, Pilsen, Czech Republic (UWB) , in 2006 and 2008 respectively. As a Ph.D. candidate at the Department of Cybernetics, UWB his main research interests include computer vision, machine learning, medical imaging, image segmentation, texture analysis. He is a teaching assistant at the Department of Cybernetics, UWB.

**Jiri Kunes** Jiří Kuneš was born in Planá, Czech Republic, 1993. He is a student in the Bachelor's program focused on cybernetics and digital data processing at the University of West Bohemia, Plzen, Czech Republic (UWB).

**Milos Zelezny** doc. Ing. Miloš Železný, Ph.D. was born in Plzen, Czech Republic, in 1971. He received his Ing. (=M.S.) and Ph.D. degrees in Cybernetics from the University of West Bohemia, Plzen, Czech Republic (UWB) in 1994 and in 2002 respectively. He is currently a lecturer at the UWB. He has been delivering lectures on Digital Image Processing, Structural Pattern Recognition and Remote Sensing since 1996 at UWB. He is working in projects on multi-modal speech interfaces (audio-visual speech, gestures, emotions, sign language). He is a member of ISCA, AVISA, and CPRS societies. He is a reviewer of the INTERSPEECH conference series.

# Solving Problems of Clustering and Classification of Cancer Diseases Based on DNA Methylation Data

A. Polovinkin, I. Krylov, P. Druzhkov,
M. Ivanchenko, I. Meyerov, A. Zaikin, N. Zolotykh
Computational Mathematics and Cybernetics Department
Lobachevsky State University of Nizhni Novgorod
Nizhni Novgorod, Russian Federation
itlab.bio@cs.vmk.unn.ru

A. Zaikin
Department of Mathematics
University College London
London, Great Britain
alexey.zaikin@ucl.ac.uk

***Abstract*— The article deals with the problem of diagnosis of oncological diseases based on the analysis of DNA methylation data using algorithms of cluster analysis and supervised learning. The groups of genes are identified, methylation patterns of which significantly change when cancer appears. High accuracy is achieved in classification of patients impacted by different cancer types and in identification if the cell taken from a certain tissue is aberrant or normal. With method of cluster analysis two cancer types are highlighted for which the hypothesis was confirmed stating that among the people affected by certain cancer types there are groups with principally different methylation pattern.**

***Keywords*— *cancer, DNA methylation, classification, clustering***

## I. Introduction

In spite of all advances of modern medicine introduction of new diagnosis and treatment methods, cancer disease and mortality rates constantly keep steadily growing all over the world. Unfortunately show signs of clinical symptoms indicate the extensive-stage disease that is why pre-existing cancer detection seems to be the most promising approach. According to the international practices the selection of risk groups and screening study are the most prospective early detection of malignant neoplasms. This article discusses a new approach to the problem of early cancer diagnosis based on searching and analysis of low-level factors which can witness the development and existence of oncological disease in gene domain.

At the moment the genetic risk factors of cancer are practically elusive, however, their identification is supposed to be breakthrough advance which will result in much more effective screening methods and early diagnostics.

Epigenetic changes are DNA modifications resulted not from variations of nucleotide sequence, i. e. the variations occur not in genes but in external factors directly related to gene activities. One type of epigenetic changes is DNA methylation when methyl group (-CH3) joins certain molecule regions.

Aberrant structure of DNA methylation is one of the essential cancer signs, which enables its early diagnosis [7], however the exact role of this data in cancer genesis and clinical prediction remains elusive. Cancer is characterized by both hypermethylation (increase of methylation) and hypomethylation (decrease) of DNA. However, cancer can be witnessed not only by variation of mean level of gene methylation. The hypothesis has been proposed according to which dysregulation of stem cell genes results from aberrant variability (dispersion) of intragenic DNA methylation. This correlates with the fact that not only methylation level but also variability in certain genomic locations may be highly relevant to cancer development [6]. In particular, it has been shown that the increased stochasticity and variability in regions where the methylation level changes with cancer, results in aberrant and modified gene expression, thus explaining tumor heterogeneity [3]. Also some authors have shown that the markers reflecting differential variability of DNA methylation features may provide for better diagnosis and risk assessment of precancer genesis [8, 9].

Though the importance of study of intragenic and intergenic DNA methylation structure is clearly understood at the moment only modifications between different genomes have been studied but the problem how the remodeling of intragenic and intergenic DNA methylation is related to the origin of carcinomas. This article deals with the problem how to identify the gene group, methylation patterns of which significantly change with emerging of cancer disease, and analyses the application efficiency of certain DNA methylation methods to solve problems of classification and identification of essential features. Using methods of cluster analysis the hypothesis is studied which states that among the people affected by the same cancer type there are different groups which might be treated with potentially different methods. Authors analyze accuracy of solving problems of binary and multiclass classification between different cancer types under application of ensembles of decision trees.

## II. Methods and Data

As initial data for supposed study we propose to use inspection results of examinees from the international data base The Cancer Genome Atlas [10], which contain information about methylation level received with TheIlluminaInfinium HumanMethylation450 BeadChip [5]. Data contain circa 485000 loci per genome the intragenic location for 330000 of which is known as well as the name of related gene. Thus 15-

17 loci per gene are available. These data are available for 13 different cancer types (Bladder Urothelial Carcenoma, BLCA; Breast Invasive Carcenomia, BRCA; etc). The number of objects related to each cancer type varies from tens to hundreds.

The following methods and markers are proposed to study intragenic DNA methylation structure. The first marker group does not depend on probe sequence inside the gene or related gene region. This marker group includes the mean value for gene methylation and dispersion. The second group includes markers considering the intragenic probe location. These markers are the mean value of outlier derivative, degree of spatial outlier asymmetry, degree of deviation from line linking methylation levels at the gene ends. Except the computation of their value, for the markers of the first and second groups Z-score is applied that is computation of deviation from the mean value of proper degree for all samples from "normal" selection, measured in mean-square deviation. The value obtained in such manner will be an instability degree for methylation outlier for the corresponding gene.

As classifier we suggest to apply the decision trees [4] and their ensembles (in particular Random Forest [1]). Among the advantages of Random Forest the following features shall be mentioned: high quality of obtained models, similar to SVM and boosting, and better compared with neural nets [2], ability to effectively process data with large number of features and classes, invariance for monotonic transformations of features values, possibility to process both continuous and discrete features, presence of methods to evaluate the importance of specific features in the model. Moreover Random Forest model enables estimation of generalization error on-the-fly during its training (out-of-bag error [1]).

## III. Results of Numerical Experiments

### A. Classification Results

From the practical point of view it is desirable to consider two types of problems: to define if a person is affected or normal using tissue samples of certain organ; to distinguish between different cancer types and normal cells. The developed measures of intragenic methylation are used as a sample description. Classification accuracy is expressed in terms of misclassified samples fraction and estimated using the out-of-bag error of the Random Forest model. As Table 1 shows, the achieved accuracy for problems of binary classification makes up from 93% to 100% and for problem of multiclass classification (Table 2) is 96.5% which enables practical application of these results.

TABLE I. Results of Binary Classification for 13 Cancer Types

| Cancer type | Accuracy | Type I error | Type II error |
|---|---|---|---|
| BLCA | 0.965 | 0.166 | 0.008 |
| BRCA | 0.978 | 0.133 | 0.009 |
| COAD | 0.989 | 0.105 | 0 |
| NHSC | 0.975 | 0.12 | 0.006 |
| KIRC | 1.000 | 0 | 0 |
| KIRP | 0.984 | 0 | 0.023 |
| LIHC | 0.966 | 0.02 | 0.051 |
| LUAD | 1.000 | 0 | 0 |
| LUSC | 1.000 | 0 | 0 |
| PRAD | 0.933 | 0.183 | 0.04 |
| READ | 0.98 | 0.25 | 0 |
| THCA | 0.968 | 0.26 | 0.003 |
| UCEC | 0.983 | 0.139 | 0 |

### B. Clustering Results

The hypothesis has been proposed that within the same type of oncological diseases there are a number of different groups which theoretically speaking might be treated differently. To test this hypothesis the following has been proposed – for each type of cancer to divide classifying description of cells, related to specified type, in groups applying the methods of cluster analysis. The degree of mean level of methylation for the gene has been used as a classifying description, the k-means algorithm has been used for clustering [4] (for practical reasons 2-3 clusters has been supposed to be available). The experiment has shown (see Fig.2) that there is a distinct separation of individuals affected by KIRC and THCA cancer types in clusters. So the further analysis of these results is required from practical point of view.

### C. Selection of Important Features

One of the problems of practical importance is the existence of specific genes responsible for the development of one or another oncological disease. In this article the importance of each gene was analyzed regarding its usefulness to solve problems of binary classification (if a cell of certain tissue is malignant or normal). A set of features has been selected for each type of cancer thereby each feature ensures the quality of binary classification (cross-validation error in decision trees of depth 1) exceeding a certain threshold (within this article this value equals 0.9). The lists of significant genes for all cancer types obtained in this manner have been combined. The Fig. 1 shows the overall statistics (the number of genes simultaneously important for classification of a certain number of cancer types). As the diagram shows there are relatively small sets of genes important for classification of several cancer types. In the future the obtained results shall be analyzed medically.

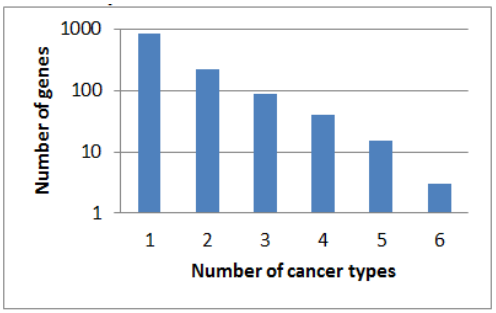



Fig. 1. Number of Genes Important for Classification of Several Cancer Types

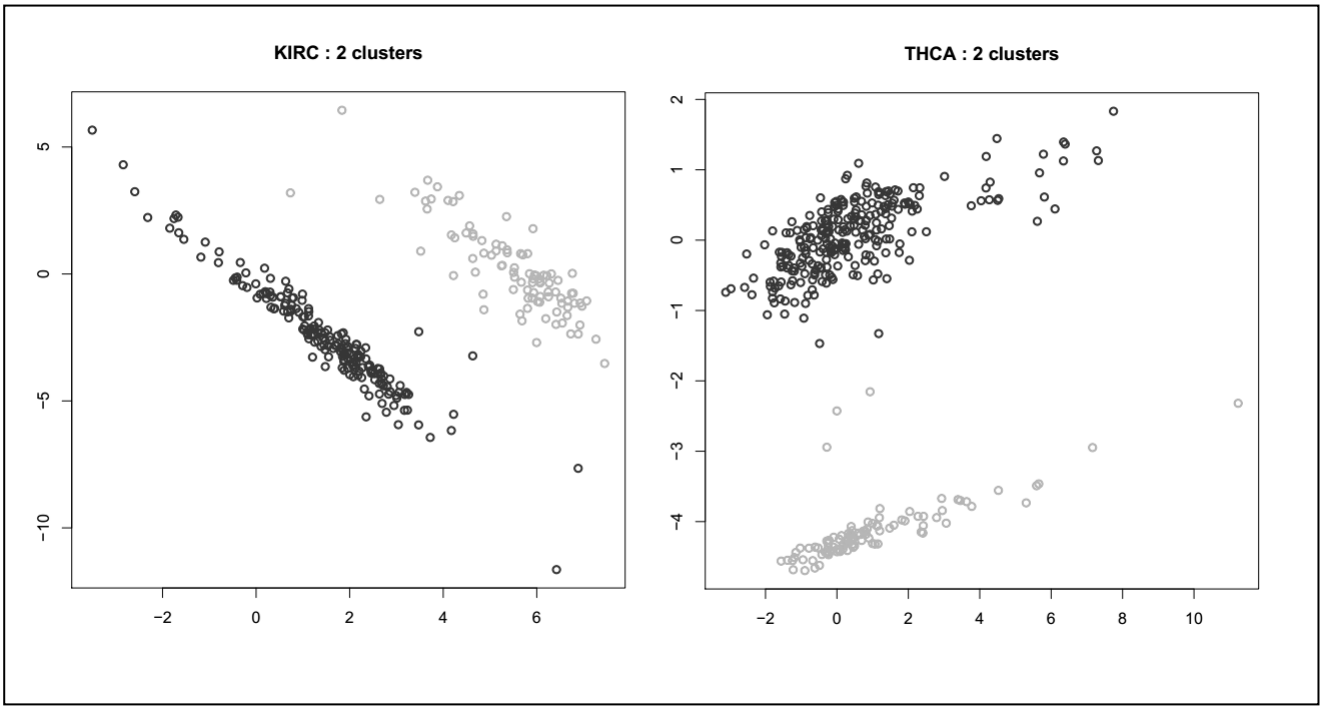



Fig. 2. Clustering Results with k-means Algorithm for KIRC and THCA Cancer Types

TABLE II. MISCLASSIFICATION TABLE TO CLASSIFY INDIVIDUALS AFFECTED BY DIFFERENT CANCER TYPES

| | **HEALTHY** | **BLCA** | **BRCA** | **COAD** | **HNSC** | **KIRC** | **KIRP** | **LIHC** | **LUAD** | **LUSC** | **PRAD** | **READ** | **THCA** | **UCEC** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HEALTHY | 0.96 | 0 | 0.01 | 0 | 0 | 0 | 0 | 0.02 | 0 | 0 | 0 | 0 | 0.01 | 0 |
| BLCA | 0.01 | 0.85 | 0.01 | 0 | 0.09 | 0 | 0 | 0 | 0 | 0 | 0.04 | 0 | 0 | 0 |
| BRCA | 0.01 | 0 | 0.99 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| COAD | 0 | 0 | 0 | 1.00 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| HNSC | 0.01 | 0 | 0 | 0 | 0.97 | 0 | 0 | 0 | 0 | 0 | 0.02 | 0 | 0 | 0 |
| KIRC | 0.03 | 0 | 0 | 0 | 0 | 0.95 | 0.01 | 0 | 0 | 0 | 0.01 | 0 | 0 | 0 |
| KIRP | 0.02 | 0.03 | 0 | 0 | 0 | 0.13 | 0.82 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| LIHC | 0.01 | 0 | 0 | 0 | 0 | 0 | 0 | 0.99 | 0 | 0 | 0 | 0 | 0 | 0 |
| LUAD | 0.07 | 0.01 | 0 | 0 | 0 | 0 | 0 | 0 | 0.92 | 0 | 0 | 0 | 0 | 0 |
| LUSC | 0.01 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.97 | 0.02 | 0 | 0 | 0 |
| PRAD | 0 | 0 | 0 | 0 | 0.08 | 0 | 0 | 0 | 0 | 0.08 | 0.84 | 0 | 0 | 0 |
| READ | 0.23 | 0 | 0 | 0.03 | 0 | 0 | 0 | 0 | 0 | 0.04 | 0 | 0.70 | 0 | 0 |
| THCA | 0.07 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.93 | 0 |
| UCEC | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.00 |

## IV. CONSLUSION

Within this article the gene group methylation patterns of which significantly change with emerging of cancer disease has been identified. Using methods of cluster analysis the hypothesis has been studied which states that among the people affected by the same cancer type there are different groups. Two cancer types have been outlined for which the hypothesis has been confirmed. The obtained solution accuracy for the problems of binary and multiclass classification enables the practical application of the results.

## REFERENCES

[1] L. Breiman. Random Forests. Machine Learning 45 (1), 2011, p 5-32.

[2] R. Caruana, A. Niculescu-Mizil. An Empirical Comparison of Supervised Learning Algorithms Using Different Performance Metrics. ICML '06 Proceedings of the 23rd international conference on Machine learning. p. 161-168.

[3] K.D. Hansen, et al. Increased methylation variation in epigenetic domains across cancer types. Nat Genet, 2011. 43(8): p. 768-775.

[4] Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. — Springer, 2001.

[5] Infinium HumanMethylation450 BeadChip | Illumina [http://products.illumina.com/products/methylation_450_beadchip_kits.ilmn].

[6] A.E. Jaffe, et al. Significance analysis and statistical dissection of variably methylated regions. Biostatistics, 2012. 13(1): p. 166-178.

[7] P.A. Jones, S.B. Baylin. The epigenomics of cancer. Cell, 2007. 128(4): p. 683-692.

[8] A.E. Teschendorff, M. Widschwendter. Differential variability improves the identification of cancer risk markers in DNA methylation studies profiling precursor cancer lesions. Bioinformatics, 2012. 28(11): p. 1487-1494.

[9] A.E. Teschendorff, et al. Epigenetic variability in cells of normal cytology is associated with the risk of future morphological transformation. Genome Med, 2012. 4(3): p. 24.

[10] The Cancer Genome Atlas – Cancer Genome – TCGA [http://cancergenome.nih.gov/].

[11] M. Widschwendter, et al. Epigenetic stem cell signature in cancer. Nat Genet, 2007. 39(2): p. 157-158.

Proceedings of the 9th Open German-Russian Workshop on Pattern Recognition and Image Understanding
Dec 1-5, 2014, Koblenz, Germany

# Stereo EKF Pose-based SLAM for AUVs

Markus Solbach[1], Francisco Bonin-Font[2], Antoni Burguera[2], Gabriel Oliver[2], and Dietrich Paulus[1]

[1]Computational Visualistics Group, University Koblenz-Landau, Koblenz (56070) (Germany)
[2]Systems, Robotics and Vision Group, University of the Balearic Islands (UIB), Palma de Mallorca (07122) (Spain)

***Abstract*—Visual localization is a crucial task in *Autonomous Underwater Vehicles* (AUV) and it is usually complicated by the extreme irregularity of the natural aquatic environments, or by unfavorable water conditions. Visual *Simultaneous Localization and Mapping* (SLAM) approaches are widely used in land and represent the most precise techniques for localization, but applied underwater, they are still an open and ongoing challenge. This paper presents a general approach to visual 3D pose-based SLAM based on *Extended Kalman Filters* (EKF). This approach has a general design being applicable to any vehicle with up to 6 *Degrees of freedom*, so, it is particularly suitable for AUV. It uses only visual data coming from a stereo camera, all orientations involved in the system are represented in the quaternion space in order to avoid the gimbal lock singularities, and the sparsity of the covariance matrix is guaranteed during the whole trajectory since the state vector only includes the vehicle global pose. The vehicle pose is continuously predicted by means of a stereo visual odometer, and eventually corrected with the pose constraints given by a particularization of the *Perspective N-Point* problem (PNP) [1], applied to the registration of images that most likely close a loop. Experimental results show the important pose corrections given by the SLAM approach with respect to a ground truth, compared with the evident trajectory errors present in the visual odometer estimates.**

## I. Introduction and Related Work

Nowadays, *Remotely Operated Vehicles* (ROVs) are commonly used in a variety of scientific or industrial applications, such as surveying, sampling, rescue or industrial infrastructure inspection and maintenance. However, *Autonomous Underwater Vehicles* (AUVs) are being progressively introduced to run highly repetitive, long or hazardous missions, reducing notably the operational costs and the complexity of human and material resources.

The localization task becomes a crucial issue in AUVs since significant errors in pose can lead to the programmed mission failure. The motion of an underwater vehicle with 6 *Degrees of Freedom* (DOF) can be estimated, for instance, (a) using inertial sensors, (b) using odometry, computed via cameras or acoustic sensors, or, (c) fusing all these sensorial data in *Extended Kalman Filters* (EKF) or particle filters, to smooth trajectories and errors [7]. However, all these methods are, to a greater or lesser extent, prone to drift, being necessary a periodical adjustment of the vehicle pose to minimize the accumulated error. *Simultaneous Localization And Mapping* (SLAM) [3] techniques constitute the most common and successful approach to perform precise localization by identifying areas of the environment already visited by the robot. Traditionally, SLAM has been developed using range sensors, but cameras outperform range sensors in temporal and spatial resolutions.

Imaging natural sub-aquatic environments has additional and challenging difficulties not present in land: the completely irregular structures of the bottoms, the light attenuation, flickering, scattering, the lack of man made structured frameworks, and the subsequent difficulty to register images, that is, to identify the same scene visualized from different viewpoints, maybe under different environmental conditions, with partial or total overlap, and taken at different time instants.

The literature is scarce in efficient visual SLAM solutions especially addressed to underwater robots and tested in field robotic systems. Many of them particularize the approach commonly known as EKF-SLAM [3], correcting the odometry with the results of an image registration process in an EKF context. These systems normally include the vehicle pose and the landmarks in the state vector, correcting continuously the vehicle trajectory and the whole map [14], [13]. However, this approach presents two major problems: (a) the computational cost increases significantly with the number of the detected landmarks, and (b) the linearization errors inherent to the EKF. Eustice *et al* [4] adopted a *Delayed State Filter* (DSF) to alleviate both problems, but the used image registration process is still a costly procedure.

EKF-SLAM approaches can be *pose-based*, if each iteration of the filter gives in the state vector a set of successive robot poses with respect to an external fixed global frame, or *trajectory-based*, if the state vector contains the successive robot relative displacements from point to point of the trajectory. The trajectory-based approach reduces the EKF linearization errors with respect to pose-based approaches but, contrarily to the later, it does not scale well for large environments, since the Jacobian of the observation function is non-zero with respect to all intermediate elements between two poses closing a loop [5]. Although the trajectory-based schema can be adopted to abate EKF linearization errors [2], it is more suitable for low and mid scale missions. All these solutions represent the vehicle orientation in the Euler angles space, assuming that the submarine can not adopt singular poses.

This paper presents a stereo pose-based EKF-SLAM approach, with the next relevant characteristics: a) it is a generic solution for vehicles with up to 6DOF ([$x, y, z$,roll, pitch, yaw]), so especially useful in AUV; it is feed with pure 3D data computed only from stereo vision; contrarily to previous underwater visual SLAM approaches, this is a *Multiplicative Extended Kalman Filter* (MEKF) approach, since all orienta-

This work is partially supported by the Spanish Ministry of Economy and Competitiveness under contracts PTA2011-05077 and DPI2011-27977-C03-02 and by the Erasmus European Program.

tions involved in the present approach are represented in the quaternion space to avoid filtering errors due to the gimbal lock singularities; b) the vector state contains only the set of robot global poses, keeping the sparsity of the covariance matrix at each iteration; the computational resources needed are drastically reduced with respect other EKF approaches that include the landmarks in the state vectors; c) it pioneers the adaptation of the well known Perspective N-Point problem (PNP) [1] to the image registration process underwater, framing it in such a stereo EKF SLAM approach; the algorithm performs robustly two tasks in one shot: firstly, it confirms or it rejects the existence of overlap between two stereo pairs (i.e. if both views represent a loop closing) and, in case there is a coincidence, it calculates the camera relative transformation, in translation and orientation, between the two poses at which both views were taken; these transformations are later used as the measurements to correct the predictions in the EKF; d) the implementation has been published in a public repository (https://github.com/srv/6dof_stereo_ekf_slam) to facilitate further research and development in this area.

## II. 3D Transformations

### A. Composition

One of the key targets of this work is modeling, for 6DOF and in the quaternion space, the classical *composition* ($\oplus$) and *inversion* ($\ominus$) transformations, described by Smith *et al* [15] in the context of stochastic mapping, and deriving their Jacobians.

Both operations define a transformation in translation and rotation. The $\oplus$ operation permits accumulating a pose transform $Y$ (translation, $[x^Y, y^Y, z^Y]$ and rotation in roll, pitch and yaw, represented as a quaternion $\hat{q}^Y = [q_w^Y, q_1^Y, q_2^Y, q_3^Y]$) to a current global pose $X$ (position $[x^X, y^X, z^X]$ and its quaternion orientation $\hat{q}^X = [q_w^X, q_1^X, q_2^X, q_3^X]$).

Let us define $X_+$ as the global pose obtained from the composition between $X$ and $Y$. $X_+ = X \oplus Y = [X_+^t, X_+^r]$, where

$$X_+^t = [x^X, y^X, z^X, 1] + A^X \cdot [x^Y, y^Y, z^Y, 1] \tag{1}$$

, being $A^X$ the rotation matrix obtained from $\hat{q}^X$ and, $X_+^r = \hat{q}^X * \hat{q}^Y$, where the operator $*$ denotes the product of quaternions.

The covariance of the composition function $f_\oplus = X \oplus Y$ is:

$$C_+ = J_{1\oplus} \cdot C^X \cdot J_{1\oplus}^T + J_{2\oplus} \cdot C^Y \cdot J_{2\oplus}^T \tag{2}$$

, where $C^X$ and $C^Y$ are the corresponding covariances of $X$ and $Y$, $J_{1\oplus} = \frac{\partial f_\oplus}{\partial X}|_{\hat{X},\hat{Y}}$ and $J_{2\oplus} = \frac{\partial f_\oplus}{\partial Y}|_{\hat{X},\hat{Y}}$, being $\hat{X}$ and $\hat{Y}$ the mean of the $X$ and $Y$ variables.

### B. Inversion

The operation ($\ominus$) returns the *inverse* of a given transformation in position and orientation. Let us denote $X = [t, \hat{q}^X]$, being, $t = (x^X, y^X, z^X)$ and $\hat{q}^X = (q_w^X, q_1^X, q_2^X, q_3^X)$ a global pose with 6DOF. $X$ can also be represented as a matrix,

$$\begin{pmatrix} & \vec{n} & \vec{o} & \vec{a} & \vec{p} \\ & & A & & t \\ & 0 & 0 & 0 & 1 \end{pmatrix} \tag{3}$$

, where $A$ is the $3 \times 3$ rotation matrix obtained from $\hat{q}^X$.

Let us denote the inverse of $X$ as $f_\ominus = \ominus X = [-\vec{n} \circ \vec{p}, -\vec{o} \circ \vec{p}, -\vec{a} \circ \vec{p}, \hat{q}^{X(-1)}]$, where $\circ$ represents the dot product and $\hat{q}^{X(-1)}$ is the quaternion result of inverting $\hat{q}^X$.

The covariance of $f_\ominus = \ominus X$ is:

$$C_- = J_\ominus \cdot C^X \cdot J_\ominus^T \tag{4}$$

, being $J_\ominus = \frac{\partial f_\ominus}{\partial X}|_{\hat{X}}$.

## III. Image Registration

The image registration process is in charge of verifying if two stereo images close a loop, that is, if they have a certain overlap, although they are taken at different time instants, at different view points, at different height, or even with different environmental conditions.

Registering successfully such pieces of information is essential to impose strong pose constraints to the filter, which increases accuracy in the incremental localization process. If two images present overlap, the image registration procedure must estimate the motion of the camera between the points at which both images where taken so that they can be represented with respect to a common coordinate frame.

Algorithm 1 describes the main steps of this process.

**Algorithm 1:** Image Registration

**input** : Current Stereo Image pair $S_l$ (left frame), $S_r$ (right frame) and Recorded Stereo Image $I = (I_l, I_r)$ candidate to close a loop with $S_l$ and $S_r$

**output**: 3D Transformation $[R, t]$

**begin**

1 $[F_l, F_r] \leftarrow$ `stereoMatching` $(S_l, S_r)$;
2 $F_t \leftarrow$ `findFeature` $(I_i)$;
3 **if** `match` $(F_l, F_t)$ == *true* **then**
4 $[F_l, F_r] \leftarrow$ `updateFeature` $(F_l, F_r)$;
6 $P_{3D} \leftarrow$ `calc3DPoints` $(F_l, F_r)$;
7 $[R, t] \leftarrow$ `solvePnPRansac` $(F_t, P_{3D})$;
*return* $[R, t]$
**else**
8 *return error*;

***Line 1*** finds and matches image features between $S_l$ and $S_r$, applying RANSAC to eliminate outliers, and stores them in $F_l$ and $F_r$.

***Line 2*** finds image features in $I_l$ and stores them in $F_t$.

***Line 3*** performs a feature matching between $F_l$ and $F_t$ refined by RANSAC. If the number of features matched

between $F_l$ and $F_t$ is greater than a certain threshold, it is assumed that, most likely, there is a loop closing. Otherwise, it returns an error.

***Line 4*** updates the features in $F_r$ that remain as inliers after the matching between $F_l$ and $F_t$, to be in line with the inliers matching between $F_l$ and $F_t$.

***Line 5*** computes the 3D points coordinates, using the stereoscopy principle, corresponding to the remaining inliers in $F_l$ and $F_r$, and stores them in $P_{3D}$.

***Line 6*** solves the Perspective N-Point problem (PNP), returning a pose transformation $[R, t]$, between $S_l$-$S_r$ and $I_l$-$I_r$, that minimizes the error of reprojecting the 3D points stored in $P_{3D}$ onto the 2D features of the image $I_l$, assuming the existence of a translation and a rotation among them:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} [R|t] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{5}$$

where $(u, v, 1)$ are the homogeneous image coordinates of the re-projected point, $(X, Y, Z, 1)$ defines the 3D world point to be re-projected, $(f_x, f_y)$ is the focal length, $(c_x, c_y)$ is the principal point, and $[R|t]$ is the matrix that describes the camera motion between both scenes.

The selected value of $[R|t]$ in equation 5 is the one that minimizes the total re-projection error:

$$[R|t] = argmin_{R,t} \sum_{i=1}^{N} \|p_i - s_i\|^2 \tag{6}$$

where $p_i$ is the re-projected image point $(u, v)$, $s_i$ is its corresponding original feature on the image $I_l$ and $N$ is the number of inliers stored in $P_{3D}$. Equation 6 can be solved iteratively using Levenberg-Marquardt algorithm (LMA), also known as the damped least-squares (DLS) method [10].

The PNP-problem is widely discussed and can be found in the literature formulated in multiple solutions. This technique is applied in a wide range of applications such as *object recognition* or *structure from motion* [8].

## IV. Stereo Pose-based EKF-SLAM

The localization module performs a pose-based stereo SLAM approach in an EKF context. The Kalman state vector $\chi$ contains a successive set of robot poses expressed with respect to a global static frame, in the form of $X = [t, q_p]$ (position in 3D and a quaternion representing an orientation in 3 axis). The initial state of $\chi = (0, 0, 0, 1, 0, 0, 0)$ (position= (0,0,0), and an orientation of 0 in all axis). The covariance $C$ of the state vector is initially set to a $7 \times 7$ zero-matrix. The approach has 3 main stages, the Prediction step, the State Augmentation step and the Update step.

During the prediction stage, the vehicle motion is estimated by a stereo visual odometer, in the form of $Y_o = [t, q_o]$ (translation in 3D and a rotation in 3D) with a $7 \times 7$ covariance matrix $C_0$. The predicted pose is $X_p = X \oplus Y_o$ with an associated $7 \times 7$ matrix covariance $C_t^+$ calculated as detailed in section II-A. Then, $\chi$ is augmented with $X_p$, giving rise to the prediction function $f_p(\chi, Y_o) = [\chi, X_p]$. The covariance $C$ of the state vector is also augmented according to: $C^+ = J_c C J_c^T + J_o C_o J_o^T$, being $J_c = \frac{\partial f_p(\chi, Y_o)}{\partial \chi}|_{\hat{\chi}}$ and $J_o = \frac{\partial f_p(\chi, Y_o)}{\partial Y_o}|_{\hat{\chi}}$. After $n$ iterations, the length of the state vector will be $n * 7$.

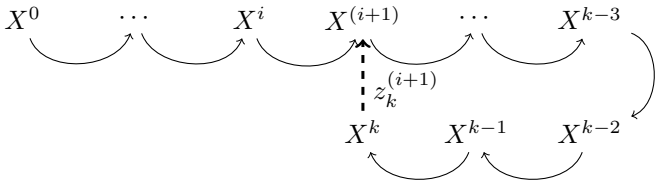



Fig. 1: A loop closing (dashed arrow) in the state vector (black arrows).

The update step is in charge of correcting the predicted motion using the loop closings detected between the image grabbed at the current filter iteration and all the images grabbed previously. When the stereo image grabbed at the current state is registered with an image captured and stored during any other previous state of the covered trajectory, the system is providing an additional pose constraint between both camera positions. This constraint can be compared with the transformation between both positions giving by a pure composition of the corresponding poses stored in the filter state. Figure 1 illustrates the idea. $X^0$, $X^1$, $\cdots$, $X^k$ represent the successive absolute poses of the vehicle along its trajectory stored in the state vector. After $k$ iterations, the image grabbed at iteration $i + 1$ is registered with the current image at $X^k$, so both close a loop. The result of this image registration process is $z_k^{(i+1)}$, a relative transformation from $X^{(i+1)}$ to $X^k$ which depends only on the image registration process. The observation function for one loop closing is defined as $h^k = \ominus X^k \oplus X^{(i+1)}$, which is the relative transformation between both registered states according to the successive filter estimates. The Kalman innovation for one loop closing is defined as $\Upsilon_k = h^k - z_k^{(i+1)}$. The observation vector $h$, the measurements vector and the innovation vector $\Upsilon$ will have as many rows as loop closings are found with the current image. The observation matrix $H = \frac{\partial h}{\partial \chi^+}\Big|_{\hat{\chi}^+}$ will have as many rows as loop closings, as many columns as elements in the state vector, and all positions not corresponding to those states involved in each loop closing will be 0:

$$H = \begin{bmatrix} \mathbf{0} & \frac{\partial h^1}{\partial X^i} & \mathbf{0} & \dots & \mathbf{0} & \frac{\partial h^1}{\partial X^k} \\ \dots & & & & & \\ \mathbf{0} & \mathbf{0} & \frac{\partial h^n}{\partial X^j} & \dots & \mathbf{0} & \frac{\partial h^n}{\partial X^k} \end{bmatrix} \tag{7}$$

where $n$ is the number of loop closings registered with the current image and $X^i, X^j$ represent two of those $n$ registered states.

Due to the nature of the quaternions (completely different quaternions can represent the same orientation and vice-versa), the pure subtraction that defines the innovation might not reflect correctly how is, or how the innovation should be when two orientations are very similar. For this reason, our approach calculates the innovation subtracting the translation

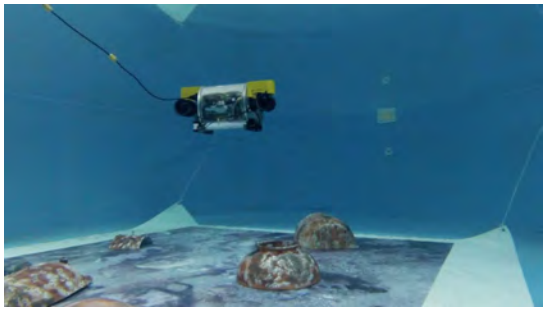


Fig. 2: Fugu-C navigating in the water tank, with the bottom covered by a poster and some containers on it.

vector of $h^k$ and $z_k^{(i+1)}$ and subtracting the modules of the corresponding quaternions to get the difference of orientations: $|q_z i| - |q_h i|$, where $q_z i$ represents the quaternion corresponding to the orientation of the $i_{th}$-measurement and $q_h i$ represents the quaternion of the corresponding $i_{th}$ observation.

From now on, by applying the Kalman equations, one can obtain an updated state vector $\chi^+$ and its updated covariance.

$$S = H \cdot C^+ \cdot H^T + R, \tag{8a}$$
$$K = (C^+ \cdot H^T)/S, \tag{8b}$$
$$\chi^+ = \chi + K \cdot \Upsilon, \tag{8c}$$
$$C_u = (1 - K \cdot H) \cdot C^+, \tag{8d}$$

where $R$ represents the measurements covariance matrix and $C_u$ represents the updated state vector covariance ($C$).

## V. Experimental Results

A first set of experiments were conducted with the Fugu-C platform, a low-cost mini-AUV developed at the University of the Balearic Islands. The sensor suit for this vehicle includes two stereo rigs, one looking forward and another one looking downwards, a MEMS Inertial Measurement Unit and a pressure sensor. The axis of the down-looking camera is perpendicular to the ground. Fugu-C works with ROS [11] as middleware, and thanks to the ROS-bag technology, missions were recorded on-line and reproduced offline with exactly the same conditions as the original mission. A stereo visual odometer based on LibViso2 [6] was used to compute the first estimates of the robot displacement. Visual odometry data was provided at 10Hz and all routes were traveled at a constant depth. The first experiments with the robot were conducted in a water tank 7 meters long, 4 meters wide and 1.5 meters depth, whose bottom was covered with a printed digital image of a real seabed. The trajectory ground truth was computed by registering each image captured online with the whole printed digital image, which was previously known. Figure 2 shows Fugu-C navigating in the water tank with some plastic containers on the bottom to simulate some relieve.

The first example shown in this section corresponds to a sweeping task performed in the tank. In order to assess the performance of the SLAM approach with different levels of error and drift in the visual odometry, the results of the stereo odometer were corrupted with different levels of aditive zero mean Gaussian noise. In total six noise levels were tested 20 times to obtain significant statistical results. The noise covariance ranges from $[\Sigma_x, \Sigma_y, \Sigma_z, \Sigma_{qw}, \Sigma_{q1}, \Sigma_{q2}, \Sigma_{q3}] = [0, 0, 0, 0, 0, 0, 0]$ (noise level 1) to $[\Sigma_x, \Sigma_y, \Sigma_z, \Sigma_{qw}, \Sigma_{q1}, \Sigma_{q2}, \Sigma_{q3}] = [3e-6, 3e-6, 3e-6, 3e-6, 3e-6, 3e-6, 3e-6]$ (noise level 6). In order to have a quantitative measure of the SLAM quality, the trajectory error was defined as the difference between the ground truth and the corresponding 3D estimate given by the odometry and by the EKF, divided by the length of the trajectory. Calculated like this, the obtained error units are meters per traveled meter. This technique permits the direct comparison of results obtained in different experiments.

| *Noise Level* | *0* | *1* | *2* | *3* | *4* | *5* |
|---|---|---|---|---|---|---|
| *Noise Covariance* | *0* | *3e-9* | *9e-9* | *3e-8* | *5e-7* | *3e-6* |
| *Odom. error ∅* | 0.038 | 0.417 | 0.494 | 0.806 | 2.614 | 6.898 |
| *EKF error ∅* | 0.027 | 0.282 | 0.285 | 0.309 | 0.590 | 0.953 |
| *Improv. (%)* | 28.9 | 32.3 | 42.3 | 61.6 | 77.4 | 86.1 |

TABLE I: Experiment 1: odometry and EKF-SLAM trajectory mean errors (∅). Error units are meters per traveled meter.

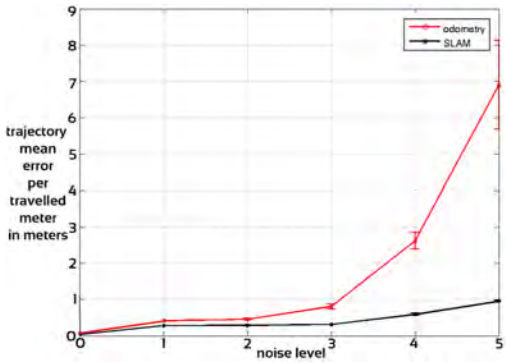



Fig. 3: Evolution of the odometry mean error and the EKF-SLAM mean error, for the different levels of corruptive noise.

Table I shows, for this first experiment, how the presented EKF-SLAM approach improves the odometric estimates since the mean of the trajectory error with respect to the ground truth are always clearly smaller. In the first column, where, in fact, no noise is used, the improvement is $28.9\%$, from an odometric mean error of $0.038m$ down to a SLAM mean error of $0.027m$. When the noise level added to the odometry increases, the correction given by the EKF-SLAM is more evidently reflected in the percentage of improvement. For a noise level of 4, the odometry mean error is $0.806m$ while the EKF mean error is $0.309m$, an improvement of $61.6\%$.

Even with the highest noise level that causes an odometry trajectory mean error of $6.898m$, the EKF-SLAM is able to improve the estimates a $86.1\%$. Figure 3 shows how the trajectory mean error raises very fast up to $7m$ as the noise level added to the odometry increases, whilst the level of the trajectory mean error of the EKF-SLAM estimates is bounded between $0m$ and $1m$. The vertical error bars correspond to $0.1\sigma$ to provide a clearer representation. $y-$axis shows the error per travelled meter in meters and the $x-$axis represents the different noise levels corrupting the odometry.

Figure 4 shows the trajectory of the aforementioned sweeping task, according to the odometry estimates (in black) corrupted with different levels of Gaussian noise, the ground truth (in blue) and the EKF estimates (in red). All units are expressed in meters. The plot corresponding to the noise level 6 shows clearly how the EKF-SLAM approach is able to correct

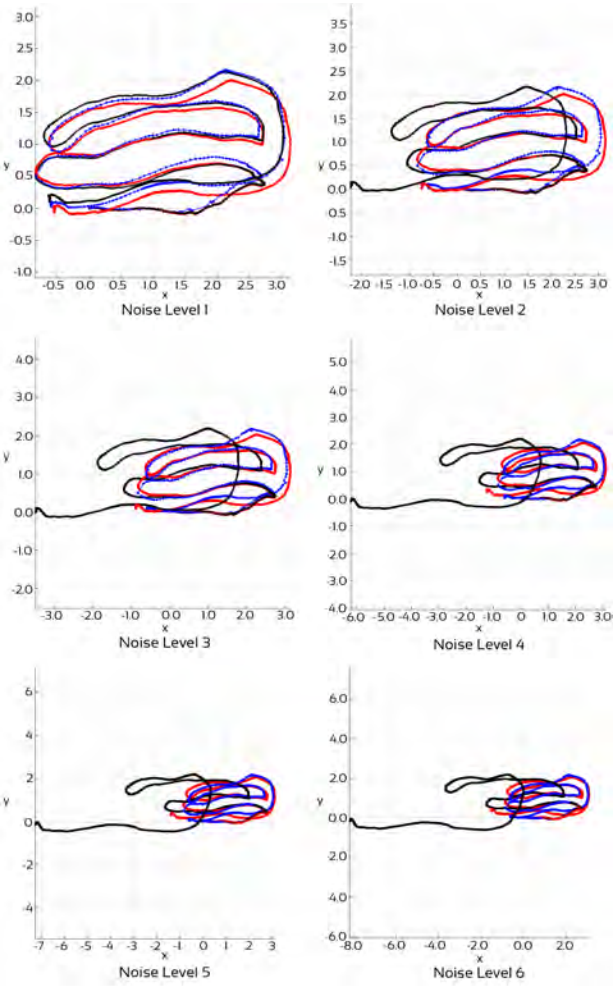



Fig. 4: The sweeping trajectory according to the three different estimates with different levels of corruptive noise.

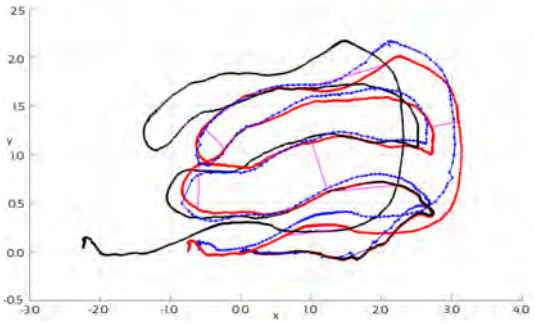


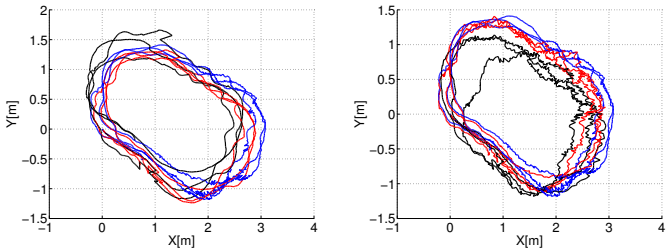
Fig. 5: The trajectory according to the three different estimates plus 8 loop closings (in magenta).

the odometry trajectory which is clearly drifted, setting it close to the ground truth.

Figure 5 shows the trajectory according to the three different estimates, being the odometry corrupted with a noise level 3, and eight loop closings. Each loop closing is shown as an edge in magenta linking the two images involved. Although in this trajectory there are more than 30 loop closings, only 8 have been plotted just to present the figure with enough clarity.

A second set of experiments were run using several datasets grabbed by a stereo camera mounted on the Girona500 AUV


Fig. 6: The Girona500 AUV with the camera circled in red.




(a) (b)

Fig. 7: Experiment in the CIRS water tank. (a) Trajectory without noise corrupting the odometry. (b) Trajectory with noise corrupting the odometry. Noise covariance = *3e-4*.

[12] (looking downwards, with its axis perpendicular to the ground), while moving in a bigger water tank located in the *Underwater Vision and Robotics Research Center* (CIRS) of the University of Girona. The bottom of this water tank was also covered with a poster picturing a real marine context, used also to calculate the trajectory ground truth. Figure 6 shows the Girona500 AUV with the stereo camera mounted in it.

Figure 7 shows the trajectory, in 2D, followed by the robot during one of the experiments conducted in the CIRS water tank. In this trial, Girona500 repeated three times an ellipsoidal motion. Plot 7-(a) shows the trajectory estimated by the uncorrupted odometry (in black), the EKF (in red) and the groud truth algorithm (in blue). Plot 7-(b) shows the trajectory estimated by the odometry corrupted with noise (Noise covariance = *3e-4*) (in black), the EKF output (in red) and the groud truth (in blue).

Table II shows the evolution of the mean error in the EKF-SLAM estimates when the odometry is not corrupted and when the odometry is corrupted with several levels of testing noise. In these experiments, the odometry error, the EKF error and the subsequent percentage of improvement are quite stable up to a certain level of noise affecting the odometry. However, the percentage of improvement provided by the filter starts to decrease when the noise level in higher than *1e-4*, pointing to a saturation in the filter effect.

Figure 8-(a) shows the evolution of the trajectory error according to the uncorrupted odometry and the EKF, during the mission of figure 7. Figure 8-(b) shows the evolution of the trajectory error of the odometry corrupted with a *3e-4* covariance noise.

Figure 9 shows two image pairs, grabbed during the

| *Noise Level* | *0* | *1* | *2* | *3* |
|---|---|---|---|---|
| *Noise Covariance* | *0* | *3e-6* | *3e-5* | *3e-4* |
| *Odom. error ∅* | 0.373 | 0.413 | 0.413 | 0.415 |
| *EKF error ∅* | 0.174 | 0.173 | 0.172 | 0.203 |
| *Improv. (%)* | 53.3 | 58.1 | 58.3 | 51.1 |

TABLE II: Experiment in the CIRS water tank: odometry and EKF-SLAM trajectory mean errors (∅). Error units are meters per traveled meter.

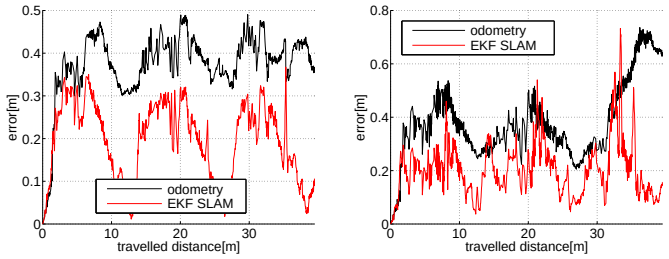



Fig. 8: Experiment in the CIRS water tank. Instantaneous error in the 3D estimates. (a) Without noise corrupting the odometry . (b) With noise corrupting the odometry. Noise covariance = *3e-4*.

experiment in the CIRS water tank, both closing a loop in two different scenes. The loop closing on the top presents a relative rotation of nearby $90°$ and a very slight relative translation in $y$, while the loop closing on the bottom shows relative rotation of approximately $20°$ and an evident translation in $(x, y)$.

## VI. Conclusions

This paper presents an approach to perform pose-based stereo EKF-SLAM to be used as a main localization system in several underwater applications such as surveying and intervention. The remarkable points of this work are: a) it is generally designed for vehicles with 6DOF, thus it is particularly useful for underwater autonomous robots; the approach deals with pure stereo data and the vehicle orientation is always represented in the quaternion space during all the stages of the algorithm, avoiding possible singularities and their affectation in the EKF estimates; b) the filter state vector does not include any landmark, but only the consecutive vehicle global poses; this approach speeds up the system considerably; b) it particularizes the PNP problem to detect loop closings (3D-2D), resulting in a robust and reliable method to find the camera pose transformation between both views that close a loop; c) pose-based approaches generate sparse covariance matrices, scaling much better for long routes than their trajectory-based counterparts; d) experimental results obtained in two different controlled environments, with two different AUVs, show that the approach is highly effective in finding overlapping views and correcting the odometry estimates, even if they are affected by certain levels of corruptive noise; e) the implementation has been published in a public repository (https://github.com/srv/6dof_stereo_ekf_slam) to be used for research and development under GNU license.

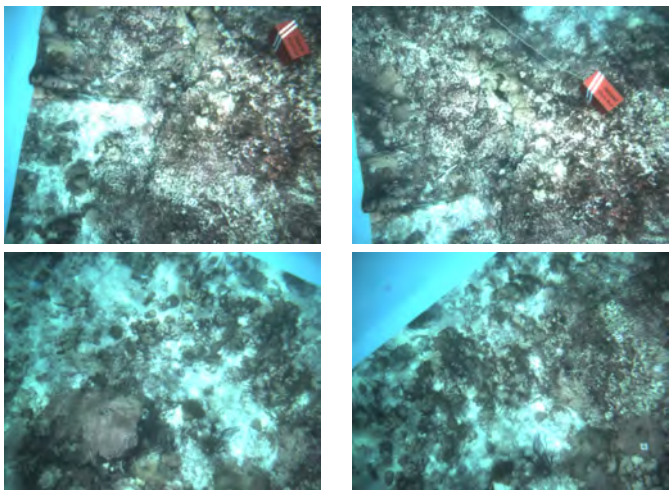


Fig. 9: Two image pairs closing two loops during the experiment in the CIRS water tank.

Future work is focused in comparing the performance of this approach with the performance of a stereo graph-SLAM approach [9], in order to choose the best option to work in real marine scenarios. The output of these localization system is also used to build and map the environment in 3D by meshing conveniently the consecutive stereo point clouds.

## References

[1] M. Bujnak, S. Kukelova, and T. Pajdla, "New Efficient Solution to the Absolute Pose Problem for Camera with Unknown Focal Length and Radial Distortion," *Lecture Notes in Computer Science*, vol. 6492, pp. 11–24, 2011.

[2] A. Burguera, Y. González, and G. Oliver, "Underwater slam with robocentric trajectory using a mechanically scanned imaging sonar," in *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS)*, San Francisco, CA, October 2011.

[3] H. Durrant-Whyte and T. Bailey, "Simultaneous localization and mapping (SLAM): part I," *IEEE Robotics and Automation Magazine*, vol. 13, no. 2, pp. 99–110, June 2006.

[4] R. Eustice, O. Pizarro, and H. Singh, "Visually augmented navigation for autonomous underwater vehicles," *IEEE Journal of Oceanic Engineering*, vol. 33, no. 2, pp. 103–122, April 2008.

[5] R. Eustice, H. Singh, and J. Leonard, "Exactly sparse delayed-state filters for view-based slam."," *IEEE Transactions on Robotics*, vol. 22, no. 6, pp. 1100–1114, December 2006.

[6] A. Geiger, J. Ziegler, and C. Stiller, "Stereoscan: Dense 3d reconstruction in real-time," in *IEEE Intelligent Vehicles Symposium*, Baden-Baden, Germany, June 2011.

[7] M. Hildebrandt and F. Kirchner, "Imu-aided stereo visual odometry for ground-tracking auv applications," in *Proceedings of Oceans*, Sydney, Australia, May 2010.

[8] C. Mei, "Robust and accurate pose estimation for vision-based localisation," *IEEE International Conference on Intelligent Robots and Systems*, pp. 3165–3170, 2012.

[9] P. Negre, F. Bonin-Font, and G. Oliver, "Stereo graph slam for autonomous underwater vehicles," in *Proc. of International Conference on Intelligent Autonomous Systems (IAS)*, 2014.

[10] J. Pujol, "The solution of nonlinear inverse problems and the levenberg-marquardt method," *Geophysics*, vol. 8, no. 72, pp. 1–16, 2007.

[11] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Ng, "ROS: an open source robot operating system," in *ICRA Workshop on Open Source Software*, 2009.

[12] D. Ribas, N. Palomeras, P. Ridao, M. Carreras, and A. Mallios, "Girona 500 AUV: From Survey to Intervention," *IEEE/ASME Transactions on Mechatronics*, vol. 17, no. 1, pp. 46–53, 2012.

[13] J. Salvi, Y. Petillot, and E. Batlle, "Visual slam for 3d large-scale seabed acquisition employing underwater vehicles," in *Proceedings of the International Conference on Intelligent Robots and Sytems*, 2008, pp. 1011–1016.

[14] R. Schattschneider, G. Maurino, and W. Wang, “Towards stereo vision slam based pose estimation for ship hull inspection,” in *Proceedings of Oceans*, Waikoloa, Hawaii, June 2011, pp. 1–8.

[15] R. Smith, P. Cheeseman, and M. Self, “A stochastic map for uncertain spatial relationships.” in *Proceedings of International Symposium on Robotic Research, MIT Press*, 1987, pp. 467–474.

# Synthesis of Quasi-Invariant Regulator for Vibration Damping Multi-Storey Buildings Using Methods of Pattern Recognition

D.V. Balandin
dbalandin@yandex.ru

I.V. Kotel'nikov
neymark@pmk.unn.ru

L.G. Teklina
teklina2010@yandex.ru

The Research Institute for Applied Mathematics and Cybernetics
Lobachevsky State University of Nizhny Novgorod
Nizhny Novgorod, Russia

***Abstract*—The possibilities and peculiarities of using pattern recognition methods for synthesis of dynamical systems with given properties are considered on the example of control system for vibration damping of high-storey buildings.**

***Keywords—pattern recognition, dynamical system, quasi-invariant control***

## I. INTRODUCTION

In view of the authors pattern recognition represents not only methods, but also a new insight into the problem being solved, which makes it possible to single out the primary information from a large volume and suggest a nontrivial solution. This allows extending the scope of pattern recognition to the problems that are not standard for data mining. Synthesis of the control systems with the given properties is one of such problems. General formulation of this problem as a problem of pattern recognition is given in [1].

Distinctive features of the problem solution in such statement are the follows:

- the transition from the classical methods of studying to the new technique in which processing of the experimental results by methods of pattern recognition;

- the solution results are statistical in nature with their assessment of the degree for statistical reliability.

In paper the possibilities and peculiarities of formulating and solving synthesis problem are considered on the example of control system for vibration damping of high-storey buildings. The known approaches to its solution are the following:

- formulation of the problem as an optimal control problem, for example [2-4];

- system control through the selection of the parameters describing the control object, by solving the optimization problem (in particular, by methods of data mining) [5], and others.

The difficulties of their use rise with increasing number of estimated parameters and extending requirements to the control system. Moreover, solution of such problem can be not unique. Using methods of recognition is an attempt to overcome the arising difficulties.

The subject of this paper is the parametric synthesis of a control system with a given property, namely: synthesis of the quasi-invariant control.

## II. FORMULATION OF PROBLEM

We consider the problem of synthesizing a decentralized control system for the vibration damping of the multi-storey buildings which are described by the mathematical model of the form [2]:

$$\begin{cases} \ddot{x}_1 = -2\beta\dot{x}_1 - 2\alpha x_1 + \beta\dot{x}_2 + \alpha x_2 + u_1 - \xi \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \ddot{x}_s = -2\beta\dot{x}_s - 2\alpha x_s + \beta\dot{x}_{s-1} + \alpha x_{s-1} + \beta\dot{x}_{s+1} + \alpha x_{s+1} + u_s - \xi \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \ddot{x}_n = -\beta\dot{x}_n - \alpha x_n + \beta\dot{x}_{n-1} + \alpha x_{n-1} + u_n - \xi \end{cases}$$

where $x_i$ is coordinate of $i$ - th material point relative to the base, $u_i$ is the driving force applied to the $i$ - th material point, $\xi$ is the external disturbance. $\mathbf{x} = col(x_1,...,x_n)$ are the controlled variables, $\mathbf{u} = col(u_1,...,u_n)$ is the control function. Our goal is to construct a quasi-invariant control, i.e. to select such control function that the control error in the steady state should satisfy the condition $|x_i(t)| < \varepsilon$ under $t > T$ ($T$ is duration of the transition process) for all controlled variables with unknown external disturbance.

The works of V.A. Yakubovich, A.V. Proskurnikov are devoted to theoretical study of linear quasi-invariant control

systems [6,7]. In particular, they have proved that the conditions for existence of quasi-invariant control are the following:

- stability of the control system;

- dimension of $\mathbf{u}(t)$ should not be less than the number of the controlled variables.

Relying on the proven conditions for the existence of the quasi-invariant regulators, the paper deals with the problem of parametric synthesis of linear multidimensional systems of quasi-invariant control with the construction of the control functions of the form $\mathbf{C}(p)\mathbf{u}(t) = \mathbf{D}(p)\mathbf{x}(t)$, where $p = \frac{d}{dt}$ and $\mathbf{C}(p)$, $\mathbf{D}(p)$ are the matrix polynomials of dimension $n \times n$. Synthesis of the control function reduces to finding the unknown coefficients of polynomials in the matrices $\mathbf{C}(p)$ and $\mathbf{D}(p)$ that the control system should satisfy the requirement of quasi-invariance for an unknown but limited external disturbance under the given initial conditions. The solution of this problem is not unique. Our goal is finding and description of not all possible values of the unknown parameters but at least a subset of them fairly simple configuration which is determined with a given high degree of statistical reliability and complies with the requirement of adequacy for the measure of robust stability for defined parameters. We propose to use for their search the methods of pattern recognition that work in spaces of large dimension and allow find the desired domain of the unknown parameters with a given degree of statistical reliability.

## III. Solving Problem by Methods of Pattern Recognition

For the formulation of the task as a problem of pattern recognition we select the space of unknown parameters, which are the coefficients of polynomials forming the matrices $\mathbf{C}(p)$ and $\mathbf{D}(p)$, as a space of features $\boldsymbol{\Omega}$. The domain of parameters $\boldsymbol{\Omega}^*$ corresponding to the given properties is taken for the recognizable pattern in the space of features. Obviously, $\boldsymbol{\Omega}^* \subseteq \boldsymbol{\Omega}_0 \subset \boldsymbol{\Omega}$ where $\boldsymbol{\Omega}_0$ is the domain of stability for the synthesized system. In general, the domain of stability is not convex and connected but we aim to choose and describe at least part of this domain, believing it is connected and convex.

Solution of the recognition problem in space $\boldsymbol{\Omega}$ is the construction in this space a local decision rule of rather simple form (a set of parallelepipeds, spheres, ellipsoids, etc.), describing the desired domain of parameters $\tilde{\boldsymbol{\Omega}}^* \subseteq \boldsymbol{\Omega}^*$, under two conditions: to minimize the number of errors of the second kind for recognizable pattern and to maximize the measure of the robust stability for a set of parameters $\tilde{\boldsymbol{\Omega}}^*$. In order to speed up the process of solving we try to find the solution that is not optimal but satisfies the desired reliability of recognition $P_0$ and the necessary measure of the robust stability for defined parameters $R(\tilde{\boldsymbol{\Omega}}^*) \geq R_0$.

Synthesis of the control system consists of the sequential searching domains $\tilde{\boldsymbol{\Omega}}_0$ and $\tilde{\boldsymbol{\Omega}}^*$, and the domain $\tilde{\boldsymbol{\Omega}}_0$ is determined by the control system and does not depend on the size and type of the external disturbance, but the domain $\tilde{\boldsymbol{\Omega}}^*$ depends on the nature and size of the external influence. In according with this the problem solution by methods of pattern recognition is realized in two stages:

I. Construction of the parameters domain satisfying the stability of the control system.

II. Construction of the parameters domain satisfying the goal of control.

Each stage consists of solving three tasks:

1. searching a point in the parameters space that satisfies the aim condition of the stage;

2. forming the training sample in the parameters space on the basis of the found point using the hypothesis of compactness;

3. constructing the decision rule that meets a predetermined degree of statistical reliability.

At the first stage in task 1 $\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}_0$ is searched by solving the minimization problem $\min_{\boldsymbol{\omega}} \Lambda(\boldsymbol{\omega})$, where $\Lambda(\boldsymbol{\omega}) = \max_i(\mathrm{Re}\,\lambda_i)$, $\lambda_i$ are the roots of the characteristic polynomial, and the minimization process ends as soon as $\Lambda(\boldsymbol{\omega}) < 0$. At the second stage in task 1 $\boldsymbol{\omega}^* \in \boldsymbol{\Omega}^* \subseteq \tilde{\boldsymbol{\Omega}}_0$ is selected by solving the problem $\min_{\boldsymbol{\omega} \in \boldsymbol{\Omega}_0} F(\boldsymbol{\omega})$, where $F(\boldsymbol{\omega}) = \max_{t>T} \left\| \tilde{\mathbf{x}}(t, \boldsymbol{\omega}) \right\|$, $\tilde{\mathbf{x}}(t, \boldsymbol{\omega})$ is the solution of the system for the controlled variables, corresponding to a set of parameters $\boldsymbol{\omega}$. Minimization process ends if $F(\boldsymbol{\omega}) < \varepsilon$.

In the solving process of the task 2 the training sequence is formed by means of random selection of parameters on the base of the uniform distribution out of a certain set $G$, which is a neighborhood of the found point. Requirements for the $G$ are that it should include the points with parameters satisfying the aim of stage and the points which do not satisfy the aim of stage. It is not difficult to make by simple extension of the set $G$.

In the solving process of the task 3 the assessment of the reliability for the constructed decision rules is carried out on the independent test set. In the case if the test result is not correspond the given degree of statistical reliability $P_0$ the training sample is updated with new data, and the learning process continues. For construction of the decision rules there were used methods based on application of the optimal irreducible fuzzy tests [8]. Syndromes decision rules are covering the required domain of parameters by the set of parallelepipeds, an each syndrome can be considered as a solution of the problem. Selection of a single solution is determined by specifying the quality criteria as measure of robust stability for a set of parameters $\tilde{\boldsymbol{\Omega}}^*$.

## IV. Experimental Results

Computational experiments were carried out for ten-storey building ($n=10$, the dimension of the phase space $k=20$) with the parameters $\alpha=1$, $\beta=0{,}1$ in the description of the control object. The system is reduced to a canonical form of linear control system $\dot{\bar{\mathbf{x}}}=\mathbf{G}\bar{\mathbf{x}}+\mathbf{H}\mathbf{u}+\mathbf{F}\xi$ ($\mathbf{G},\mathbf{H},\mathbf{F}$ are the real matrices of dimensions $20\times 20$, $20\times 10$ and $20\times 1$) by means of introducing a new vector of variables $\bar{\mathbf{x}}=col(x_1,...,x_{10},x_{11},...,x_{20})$ where $x_{10+i}=\dot{x}_i$. Synthesis of the control system begins with the construction of the control function of the minimum complexity when the matrix $\mathbf{C}(p)$ is diagonal and does not depend on $p$ and with testing hypothesis of the autonomy for the controlled object consisting in the assumption that the motion control of the $i$ - th material point does not depend on the movements of all other material points, i.e. the control function of the form $\mu_i u_i=\dot{x}_i+d_i x_i$ is constructed. In this case we have the 20-dimensional space of features $\mathbf{\Omega}=\{(\mu_1,...,\mu_{10},d_1,...,d_{10})\}$. Construction of the stability domain $\tilde{\mathbf{\Omega}}_0$ in this space gave unexpected results, namely: very large intersection of the change domains for the parameters $\mu_1,\mu_2,...,\mu_{10}$ and $d_1,d_2,...,d_{10}$. This led to testing hypothesis of the more simple form for the control function $\mu u_i=\dot{x}_i+dx_i$ that corresponds two-dimensional space of features $\mathbf{\Omega}=\{(\mu,d)\}$.

Requirements for the control system are the following: with the reliability $P_0=0{,}99$ the control precision $\varepsilon\le 10^{-4}$ must be achieved for $|\xi|\le 100$ and with zero initial conditions. With the selected form of the control function the following results were obtained: under the choice of the stability domain $\tilde{\mathbf{\Omega}}_0$ as a parallelepiped (syndrome)

$$S_1=\begin{bmatrix} 596.966\le d\le 644.445 \\ -0.0006762\le\mu\le -0.00053\end{bmatrix}$$

the set of parameters $\tilde{\mathbf{\Omega}}^*$ that is described by parallelepiped

$$S_2=\begin{bmatrix} 598.014\le d\le 640.496 \\ -0.000616003\le\mu\le -0.000591252\end{bmatrix}$$

corresponds to specified precision and reliability.

When testing these results the external disturbances of various kinds have been considered, in particular:

- $\xi_1=100sign(Sin10t)$, $\varepsilon\le 10^{-4}$;
- $\xi_2=\begin{cases}0 & if \quad t\le 0\\ 100 & if \quad t>0\end{cases}$, $\varepsilon<10^{-4}$;
- $\xi_3=50Sin10t+50|Sin10t|$, $\varepsilon<10^{-4}$, et al.

The graphs of the external disturbance, the control function and the control error for $\xi=\xi_1$ are shown in Fig.1.

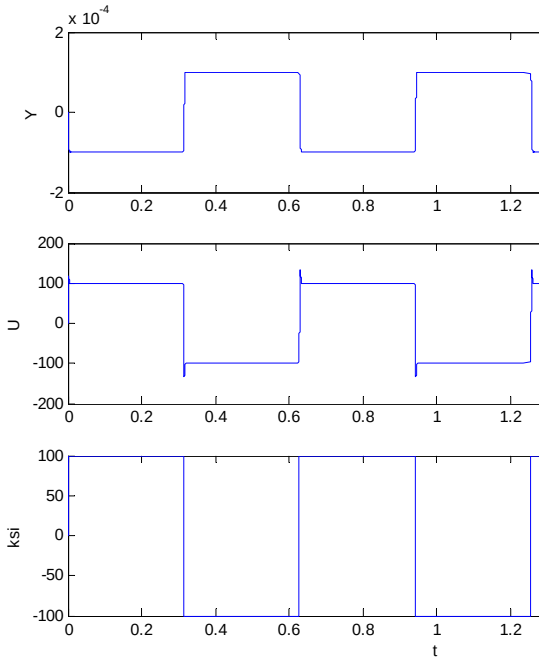



Fig. 1. The control error, the control function, the external disturbance

Note that for lack of control action on the $i$ -th material point the control error $|x_i(t)|$ increased sharply, $|x_{i-1}(t)|$ and $|x_{i+1}(t)|$ changed slightly, all other values remained unchanged.

Completely analogous results were obtained for a mathematical model of a 20-storey building.

## V. Conclusion and Future Work

This paper is intended to show that the problem of synthesis of control systems can be fruitfully viewed as a pattern recognition problem and that in this way one can make significant progress in solving it. In particular, it is assumed that the proposed approach will be used for solving the following problems:

- studying the effect of initial conditions on the final result;

- constructing a robust control system;

- synthesis of a control system by means of selecting the parameters in the description of the control object.

The proposed approach gives an opportunity to synthesize a systems that solve simultaneously all the above problems.

## Acknowledgment

This work was financially supported by the Ministry of Education and Science of the Russian Federation, project №2000.

## REFERENCES

[1] Neimark Yu.I., Teklina L.G. “Statement of the generalized problem for synthesis of the dynamic object as a problem of pattern recognition with an active experiment”, Proceedings of the 15-th Conference "Mathematical Methods of Pattern Recognition", M.: MAKS Press, 2011, pp. 200-202. (In Russian)

[2] D.V. Balandin, M.M. Kogan. Synthesis of the control laws on the base of linear matrix inequalities. M.: Fizmatlit, 2007. (In Russian)

[3] Seyed Amin Mousavi, Amir K. Ghorbani-Tanha “Optimum placement and characteristics of velocity-dependent dampers under seismic excitation”, Earthquake Engineering and Engineering Vibration, 2012, v.11, no.3, pp.403-414.

[4] N. Debnath, S.K. Deb, A. Dutta “Frequency band-wise passiv control of linear time invariant structural sysems with $H_\infty$ optimization”, Journal of Sound and Vibration, 2013, v. 332, pp. 6044-6052.

[5] F. Hejazi, I. Toloue, N.S. Jaafar “Optimization of Earthquake Energy Dissipation System by Genetic Algorithm”, Computer-Aided Civil and Infrastructure Engineearing, 2013, v. 28, pp. 796-810.

[6] Proskurnikov A.V., Yakubovich V.A. “An approximate solution of the invariance problem for the control system”, Reports of the RAS, 2003, v. 392, no.6, pp. 750-754. (In Russian)

[7] Proskurnikov A.V., Yakubovich V.A. “The problem of the invariance of the control system on part of the output variables”, Reports of the RAS, 2006, v. 406, no.1, pp. 30-34. (In Russian)

[8] Kotel’nikov I.V. “A Syndrome Recognition Method Based on Optimal Irreducible Fuzzy Tests”, Pattern Recognition and Image Analysis, 2001, v.11, no.3, pp.553-559.

# The Algebraic Approaches and Techniques in Image Analysis

Igor Gurevich and Vera Yashina
Mathematical and Applied Problems of Image Analysis
Dorodnicyn Computing Center of the Russian Academy of Sciences
Moscow, Russian Federation
igoourevi@ccas.ru, werayashina@gmail.com

***Abstract*— The main task of the tutorial is to explain and discuss the opportunities and limitations of algebraic approaches in image analysis. During recent years there was accepted that algebraic techniques, in particular different kinds of image algebras, is the most prospective direction of construction of the mathematical theory of image analysis and of development an universal algebraic language for representing image analysis transforms and image models. The main goal of the Algebraic Approach is designing of a unified scheme for representation of objects under recognition and its transforms in the form of certain algebraic structures. It makes possible to develop corresponding regular structures ready for analysis by algebraic, geometrical and topological techniques. Development of this line of image analysis and pattern recognition is of crucial importance for automatic image-mining and application problems solving, in particular for diversification classes and types of solvable problems and for essential increasing of solution efficiency and quality. The main subgoals of the tutorial are: a) to set forth the state of the art of mathematical theory of image analysis; b) to consider the algebraic approaches and techniques acceptable for image analysis; c) to present a methodology, mathematical and computational techniques for automation of image mining on the base of Descriptive Approach to Image Analysis; d) to illustrate opportunities of algebraic techniques via an example of biomedical image analysis practical problem.**

***Keywords—algebraic approach, decpriptive approah, image analysis***

## I. INTRODUCTION

Automation of image processing, analysis, estimating and understanding is one of the crucial points of theoretical computer science having decisive importance for applications, in particular, for diversification of solvable application problem types and for increasing the efficiency of problem solving.

The specificity, complexity and difficulties of image analysis and estimation (IAE) problems stem from necessity to achieve some balance between such highly contradictory factors as goals and tasks of a problem solving, the nature of visual perception, ways and means of an image acquisition, formation, reproduction and rendering, and mathematical, computational and technological means allowable for the IAE.

The mathematical theory of image analysis is not finished and passes through a developing stage. It is only recently came understanding of the fact that only intensive creating of comprehensive mathematical theory of image analysis and recognition (in addition to the mathematical theory of pattern recognition) could bring a real opportunity to solve efficiently application problems via extracting from images the information necessary for intellectual decision making. The transition to practical, reliable and efficient automation of image mining is directly dependent on introducing and developing of new mathematical means for IAE.

During recent years there was accepted that algebraic techniques, in particular different kinds of image algebras, is the most prospective direction of construction of the mathematical theory of image analysis and of development of an universal algebraic language for representing image analysis transforms and image models.

Development of this line of image analysis and pattern recognition is of crucial importance for automatic image-mining and application problems solving, in particular for diversification classes and types of solvable problems and for essential increasing of solution efficiency and quality.

It is one of the breakthrough challenges for theoretical computer science to find automated ways to process, analyze, evaluate and understand information represented in the form of images. It is critical for computer science to develop this branch in terms of solving applied problems, in particular, increasing the diversity of classes of problems that can be solved and the efficiency of the process significantly.

Images are one of the main tools to represent and transfer information needed to automate the intellectual decision-making in many application areas. Increasing the efficiency, including automatization, of gathering information from images can help increase the efficiency of intellectual decision-making.

Recently, this part of image analysis called image mining in English publications has been often set off into a separate line of research.

We list the functions of particular aspects of image handling. Image processing and analysis provides for image mining, which is necessary for decision-making, while the very decision-making is done by methods of mathematical theory of pattern recognition. To link these two stages, the information gathered from the image after it is analyzed is transformed so that standard recognition algorithms could process it. Note that

although this stage seems to have an "intermediate" character, it is the fundamental and necessary condition for the overall recognition to be feasible.

At present, automated image mining is the main strategic goal of fundamental research in image analysis, recognition and understanding and development of the proper information technology and algorithmic software systems. In the end, this automatization is expected to help developers of automated systems designed to handle images as well as end users, either in the automated or interactive mode,

- develop, adapt and check methods and algorithms of image recognition, understanding and evaluation;
- choose optimal or suitable methods and algorithms of image recognition, understanding and evaluation;
- check the quality of initial data and whether they can be used in solving the image recognition problem;
- apply standard algorithmic schemes of image recognition, understanding, evaluation and search.

To ensure such automatization, we need to develop and evolve a new approach to analyzing and evaluating information represented in the form of images. To do it, the "Algebraic Approach" of Yu. I. Zhuravlev [109] was modified for the case when the initial information is represented in the form of images. The result is the descriptive approach to image analysis and understanding (DA) proposed and justified by I. B. Gurevich and developed by his pupils [28-38].

By now, image analysis and evaluation have a wide experience gained in applying mathematical methods from different sections of mathematics, computer science and physics, in particular algebra, geometry, discrete mathematics, mathematical logic, probability theory, mathematical statistics, mathematical analysis, mathematical theory of pattern recognition, digital signal processing, and optics.

On the other hand, with all this diversity of applied methods, we still need to have a regular basis to arrange and choose suitable methods of image analysis, represent, in an unified way, the processed data (images), meeting the requirements standard recognition algorithms impose on initial information, construct mathematical models of images designed for recognition problems, and, on the whole, establish the universal language for unified description of images and transformations over them.

In applied mathematics and computer science, constructing and applying mathematical and simulation models of objects and procedures used to transform them is the conventional method of standardization. It was largely the necessity to solve complex recognition problems and develop structural recognition methods and specialized image languages that generated the interest in formal descriptions–models of initial data and formalization of descriptions of procedures of their transformation in the area of pattern recognition (and especially in image recognition in 1960s).

In 1970s, Yu. I. Zhuravlev proposed the so-called "Algebraic Approach to Recognition and Classification Problems" [109], where he defined formalization methods for describing heuristic algorithms of pattern recognition and proposed the universal structure of recognition algorithms. In the same years, U. Grenander stated his "Pattern Theory" [24-26], where he considered methods of data representation and transformation in recognition problems in terms of regular combinatorial structures, leveraging algebraic and probabilistic apparatus. Both approaches dealt with the recognition problem in its classical statement and did not touch upon representation of initial data in the form of images.

Then, up to the middle of 1990s, there was a slight drop in the interest in descriptive and algebraic aspects in pattern recognition and image analysis.

By the middle of 1990s, it became obvious that for the development of image analysis and recognition, it is critical to:

- understand the nature of the initial information – images,
- find methods of image representation and description that allow constructing image models designed for recognition problems,
- establish the mathematical language designed for unified description of image models and their transformations that allow constructing image models and solving recognition problems, and
- construct models to solve recognition problems in the form of standard algorithmic schemes that allow, in the general case, moving from the initial image to its model and from the model to the sought solution.

The DA gives a single conceptual structure that helps develop and implement these models and the mathematical language [28, 29]. The main DA purpose is to structure and standardize different methods, operations and representations used in image recognition and analysis. The DA provides the conceptual and mathematical basis for image mining, with its axiomatic and formal configurations giving the ways and tools to represent and describe images to be analyzed and evaluated.

In this work, we give a brief review of the main algebraic methods and features. The work consists of seven main sections (along with Introduction and Conclusions).

"State of the art of mathematical theory of image analysis" is the section that describes modern trends in developing of mathematical tools for automation of image analysis, in particular in image mining.

The section "Steps of the algebraization" presents leading approaches of mathematical theory for image analysis oriented for automation of image analysis and understanding.

The section "The basic theories of pattern recognition" consists of following parts: "Pattern Theory by Grenander", "Theory of categories techniques in pattern recognition", "The algebraic approach to recognition classification and forecasting problems by Zhuravlev", "Contribution of the russeian mathematical school".

The section "Image Algebras" consists of brief description of different image algebras.

The section "Descriptive approach to image analysis" presents a methodology, mathematical and computational techniques for automation of image mining on the base of Descriptive Approach to Image Analysis.

In conclusion, there are some words about opportunities of algebraic techniques via an example of biomedical image analysis practical problem and discussion the prospects of the mathematical image analysis development.

Below there is the list of used abbreviations:

AEC - class of recognition algorithms for calculating estimates;

DA – descriptive approach to image analysis and understanding;

DASIR - descriptive algebraic schemes of image representation;

DIA – descriptive image algebra;

DIM – descriptive image model;

GDT - generating descriptive tree;

IAE – image analysis and estimation;

IM – image model;

RIFR - reducing images to a form suitable for recognition.

## II. STATE OF THE ART OF MATHEMATICAL THEORY OF IMAGE ANALYSIS

To automate image mining, we need an integrated approach to leverage the potential of mathematical apparatus of the main lines in transforming and analyzing information represented in the form of images, viz. image processing, analysis, recognition and understanding.

Done by pattern recognition methods, image mining now tends to multiplicity (multialgorithmic and multimodel) and fusion of the results, i.e., several different algorithms are applied in parallel to process the same model and several different models of the same initial data to solve the problem and then the results are fused to obtain the most accurate solution.

Multialgorithmic classifiers and multimodel and multiple-aspect image representations are the common tools to implement this multiplicity and fusion. Note that it was Yu. I. Zhuravlev who obtained the first and fundamental results in this area in 1970s [109].

From 1970s, the most part of image recognition applications and considerable part of research in artificial intelligence deal with images. As a result, new technical tools emerged to obtain information that allow representing recorded and accumulated data in the form of images and the image recognition itself became more popular as the powerful and efficient methodology to process, analyze data mathematically, and detect hidden regularities. Various scientific and technical, economic and social factors make the application domain of image recognition experience grow constantly.

There are internal scientific problems that have arisen within image recognition. First, these imply algebraizing the image recognition theory, arranging image recognition algorithms, estimating the algorithmic complexity of the image recognition problem, automating the synthesis of the corresponding efficient procedures, formalizing the description of the image as the recognition object, making the choice of the system of representations of the image in the recognition process regular, and some others. It is the problems that form the basis of the mathematical agenda of the descriptive theory of image recognition developed using the ideas of the algebraic approach to recognition [109] to create a systematized set of methods and tools of data processing in image recognition and analysis problems.

There are three main issues one need to solve when dealing with images–describe (simulate) images; develop, study and optimize the selection of mathematical methods and tools of data processing in image recognition; and implement mathematical methods of image analysis on a software and hardware basis.

What makes image analysis and recognition problems peculiar, complex and thus difficult and catching is the necessity to find a compromise between rather contradictory factors. These factors are the requirements imposed on the analysis, the nature of visual perception, the ways to obtain, form and reproduce images and the existing mathematical and technical ways to process them. The main contradiction is between the nature of the image and the analysis based on formal description (a model, in essence) of the object, which lies in the fact that to leverage the fact that information is represented in the form of images, it is necessary to make this information non-depictive since the corresponding algorithms can only process certain symbolic descriptions.

Most methods of image processing are purely heuristic, with their quality essentially given by the degree to which they are successful in coping with the "depictive" nature of the image using the "non-depictive" tools, i.e., in employing procedures that do not depend on the fact that the information to be processed is organized in the form of images.

When we solve an image recognition problem, it is very important that we are able to choose the right recognition algorithm in a great number of known algorithms, i.e., we need to choose the best in some sense algorithm in the particular situation. It is obvious that both in image recognition and in solving recognition problems with standard teaching information [109], to make the choice of the best algorithm systematic, we need to introduce and formalize the corresponding objects of mathematical theory of image recognition, in particular, the concept of image recognition algorithm.

It is known that the necessity to state and solve the problem of choosing the algorithm with respect to the recognition quality functional led to introducing the concept of the model of recognizing algorithm. To choose optimal or acceptable procedure to solve the particular problem, one needed to fix the class of algorithms somehow. This is the first reason that led to the necessity to synthesize models of recognition algorithms.

With the concept of the model of recognizing algorithm, we can apply strict mathematical methods to study the sets of incorrect recognition procedures (i.e., heuristic procedures that are not justified mathematically but were experimentally tested in solving real recognition problems). Analyzing the totality of incorrect recognition algorithms as they are accumulated, we can select and describe particular algorithms as well as principles to form them. Acting over subsets of algorithms and first formed in a poorly formalized form, these principles can then become accurate mathematical descriptions. At this stage, principles are chosen on a heuristic basis while algorithms generated according to it can be constructed in a standard way. It is in this sense that formalization of different principles of constructing recognizing algorithms results in models of recognizing algorithms.

To construct the model of recognizing algorithm, we need to describe sets of incorrect procedures that nevertheless are efficient in solving practical problems in a uniform way. To give such set, we specify variables, objects, functions, and parameters and their exact variation area, thus introducing the sought model of the algorithm. Given some set of the corresponding variables, objects, parameters and types of functions, we can single out some fixed algorithm from the model we consider.

To construct the model of an image recognition algorithm and determine the proper class of recognition algorithms, it is not enough to transfer the concept of the model of recognizing algorithm developed in the mathematical recognition theory automatically to the image domain and directly use formal representations of a number of known recognition models studied in classical recognition theory [109]. As noted above, the nature and matter of image recognition problems differ from that of the mathematical recognition theory in its classical statement. When we move from classical recognition problems to image recognition problems, there arise mathematical problems due to formal description of the image as the object to be analyzed.

To obtain formal descriptions of images as objects to be analyzed and form and choose recognition procedures, we study the internal structure and content of the image as the result of the operations that can be performed to construct it of sub-images and other objects of simpler nature, i.e., primitives and objects singled out on the image during different stages of handling it (depending on the aspect, morphological and/or scale level used to form the image model). Since this way of characterizing the image is operational, we can consider the whole process of image processing and recognition, including construction of formal description –model of the image, as a system of transformations implemented on the image and given on the equivalence classes that represent ensembles of admissible images [31, 34]. Hence, we operate with the hierarchy of formal descriptions of images, i.e., image models used in recognition relate to different aspects and/or morphological (scale) levels of image representation. In essence, these are multiple-aspect and/or multilevel models that allow choosing and changing the necessary degree of detail of description of the recognition object in the course of solving the problem. This approach to formal description of images forms the basis for the multimodel representation of images in recognition problems.

Note that the idea to create a single theory that embraces different approaches and operations used in image and signal processing has a history of its own, with works of von Neumann continued by S. Unger, M. Duff, G. Matheron, G. Ritter, J. Serra, S. Sternberg and others [106, 10, 52, 78, 91, 100] playing an important role in it.

The main stages of algebraization are:

- Mathematical Morphology (G. Matheron, J. Serra [1970's])
- Algorithm Algebra by Yu.I.Zhuravlev (Yu. Zhuravlev [1970's]
- Pattern Theory (U. Grenander [1970's]
- Theory of Categories Techniques in Pattern Recognition (M.Pavel [1970's])
- Image Algebra (Serra, Sternberg [1980's]
- Standard Image Algebra (Ritter [1990's])
- Descriptive Image Algebra (DIA) (Gurevich [1990-2000])
- DIA with one ring (Gurevich, Yashina [2001 to date])

## III. Steps of the Algebraization

The section presents leading approaches of mathematical theory for image analysis oriented tor automation of image analysis and understanding. First, there is the history of developing algebraic construction for image analysis and processing – formal grammars, cellular automata, mathematical morphology, image algebras, multiple algorithms, descriptive approach

Algebraization of pattern recognition and image analysis has attracted and continues to attract the attention of many researchers. Appreciable attempts to create a formal apparatus ensuring a unified and compact representation for procedures of image processing and image analysis were inspired by practical requirements for effective implementation of algorithmic tools to process and analyze images on computers with specialized architectures, in particular, cellular and parallel.

The idea of constructing a unified language for concepts and operations used in image processing appeared for the first time in works by Unger [106], who suggested to parallelize algorithms for processing and image analysis on computers with cellular architecture.

Mathematical morphology, developed by G. Materon and Z. Serra [52, 91], became a starting point for a new mathematical wave in handling and image analysis. Serra and Sternberg [98-100] were the first to succeed in constructing an integrated algebraic theory of processing and image analysis on the basis of mathematical morphology. It is believed [62] that it was precisely Sternberg who introduced the term "image algebra" [99] in the current standard sense. (We note that U.

Grenander used this concept in the 1970s; however, he was talking about another algebraic construction [24-26]). Within the limits of this direction, an array of works continues to be written, devoted to the development of specialized algebraic constructions implementing or improving upon methods of mathematical morphology.

From that time, until 1990's the interest to descriptive and algebraic aspects of image analysis is failing. The final view of idea of IA has become Standard Image Algebra by G.Ritter [77, 78] (algebraic presentation of image analysis and processing operations).

DIA is created as a new IA provided possibility to operate with main image models and with basic models of procedure of transforms, which lead to effective synthesis and realization of basic procedures of formal image description, processing, analysis and recognition. DIA is introduced by I.B.Gurevich and developed by him and his pupils [28-38].

The history of algebraization:

- J.von Neumann [68, 69], S.Unger [106] (studies of interative image transformations in cellular space)
- M. Duff, D. Watson, T. Fountain, and G. Shaw [10] (a cellular logic array for image Processing)
- A. Rosenfeld [81, 82] (digital topology)
- H. Minkowski [63] and H.Hadwiger [39] (pixel neighborhood arithmetics and mathematical morphology)
- G.Matheron, J.Serra, S.Sternberg [52, 91, 98-100] (a coherent algebraic theory specifically designed for image processing and image analysis - mathematical morphology)
- S. Sternberg [99] (the first to use the term "image algebra")
- P. Maragos [53-56] (introduced a new theory unifying a large class of linear and nonlinear systems under the theory of mathematical morphology)
- L. Davidson [9] (completed the mathematical foundation of mathematical morphology by formulating its embedding into the lattice algebra known as Mini-Max algebra)
- G.Ritter [74-80] (Image Algebra)
- I.B. Gurevich [28-38] (Descriptive Image Algebra)
- T.R. Crimmins and W.M. Brown, R.M. Haralick, L. Shapiro, R.W. Schafer, J. Goutsias, L. Koskinen and Jaako Astola, E.R. Dougherty, P.D. Gader, M.A. Khabou, A. Koldobsky, B. Radunacu, M.Grana, F.X. Albizuri, P. Sussner [2, 8, 11, 12, 13, 14, 23, 73, 74, 79, 80, 101, 102, 103] (recent papers in mathematical morphology and image algebras)

## IV. THE BASIC THEORIES OF PATTERN RECOGNITION

### *A. Pattern Theory by U.Grenander*

"Pattern Theory" (U.Grenander) [24-26] – techniques for pattern recognition data representation and transformation on the base of regular combinatorial structures and algebraic and probabilistic means.

The most general approach to the algebraic description of information for recognition algorithms is Grenander's general pattern theory [28-30], which unites metric theory with probability theory for certain universal algebras of combinatorial type. The main attention is given to the investigation of the structure of recognizing elements. The idea that underlies Grenander's theory is that knowledge about patterns may be expressed in terms of regular structures. Regular structures are structures constructed by certain rules.

The theory is based on three principles, namely, atomism, combinatory, and observability. By atomism, we mean that the structures are composed of certain basic elements. Combinatory means that explicit rules are formulated for definition of admitted and prohibited structures. The third principle is related to the search for identification rules for determining equivalence classes. It should be noted that Grenander used the notion of image algebra: however, he was dealing with a different algebraic construction.

The search for patterns in nature an in the man-made world has generated a huge literature. Grenander tries to formalize the very concept of a pattern in terms of a mathematical framework, a pattern theory.

The subject of Grenander books [24-26] is order, patterns, regularity – concepts that imply that the world we live in has structure making it possible for us to understand it, at least to some extent. Without presupposing such a structure, we would have no hope of comprehending the phenomena that we observe and the logical relation between them.

Grenander presents a catalogue of patterns. One extreme is a completely regular pattern – for example, a crystal – which can be explained through simple rules of generation. Another extreme is complete disorder in terms of pure randomness. Also he considers intermediate situations, in which the phenomena can be analyzed partially through notions of typical structure that may be obscured by a high degree of variability, as, for example, in many instances of biomedical images. Such patterns not only appear complex – they are complex – and therefore they differ in essence from, for example, fractals and various patterns in chaos theory that seem complicated apthough they may have been generated by comparatively simple rules.

Pattern theory is a way to approach patterns through a mathematical formalism, a way of reasoning about patterns. This approach is shown by analytical tools and by employing computational methods.

Patterns are divided into two groups: open patterns and closed patterns.

Open patterns are patterns whose internal structure is as simple as possible. By "simple" Grenander means patterns whose logical architecture – its connectivity – does not involve recurrences (loops) but is straightforward. Such patterns are "ornaments", "language patterns", "a motion pattern", "time recordings", "tracks" and "behavior". So their structure shares a one-dimensional flavor, not in the sense of dimension in

geometry, but in terms of the dependencies, the logical couplings in their structure. They are linear arrangements, one part following another. They are no closed loops in their topology that will later on be described via graphs: they are open, meaning the absence of cycles.

In contrast to the open patterns the closed ones can possess intricately woven systems of dependencies. This induces a topology, an information architecture, that is often hidden from the observer. The pattern analyst faces a serious challenge in finding meaningful representations for closed patterns since their surface appearance may not give a clear indication of what deep (regular) structure supports them. Their mathematical treatment will require more thought. Such patterns are “difficult ornaments”, “weaving”, “textures”, “shapes”, “inside structure”, “connections”, “internal patterns”, “multiple object patterns”, “pattern interference”, “pattern of speculation”.

To understand patterns and analyze their structure it is necessary to introduce a mathematical pattern formalism: a pattern algebra.

Representations of patterns are built from simple building blocks that will be referred to as generators. Then after gluing generators together, it appears a configuration and the bonds of configuration will tell us what combinations will hold together. The configuration space is the first of the regular structures that we are building. The next one consists of images, a concept that formalizes the idea of observables. In other words, a configuration is a mathematical abstraction, which typically cannot be observed directly, but the image can. An ideal observer, with perfect instrumentation so that the sensor used has no observational errors, will be able to see some object, called image, that may carry less information than configuration that is being observed. The loss of information is not caused by noise in the sensors but is more fundamental and it takes some care to formalize the concept of image in order to get a suitable algebraic structure.

Then from the notion “image” and properties of configuration spaces we could define patterns.

Once the representations of pattern are constructed in the form of regular structures, then we should use them for many purposes. Actually, constructing the representations will turn out to be the hardest part in endeavor; once they have been built their use will be derived by applying general mathematical and computational principles. The tasks are divided into two categories: synthesis (or simulation) and analysis (or inference).

### *B. Theory of Categories Techniques in Pattern Recognition by M.Pavel*

“Theory of Categories Techniques in Pattern Recognition” (M.Pavel) [70, 71] – formal describing of pattern recognition algorithms via transforms of initial data preserving its class membership.

Recognizing patterns consists of associating a name or canonical pattern or prototype to a given image. The aim of categories techniques is to give mathematical general meanings to the terms “patterns” and “recognition”, and to unify the different possible approaches.

Patterns can be described by their primitive components and their composition, and/or be defines axiomatically by their invariant properties. Recognizing a pattern generally means detecting some equivalence between two images, given a collection of images and some rule of isomorphism or equivalence, deciding whether or not a given image and some prototype of a set of canonical patterns (i.e. representatives from the equivalence classes) are equivalent. This job is done by a recognition function and the standard way of solving the problem is to establish this equivalence by computing and comparing probes relative to a number of transformation-invariant features, which the image and the pattern, have in common, and which no other figure possesses.

The algebraic formalism leaves open the problems of intrinsically characterization of signs images and synonymy.

In order to answer these questions a discrete topology and a discrete analogous of topological homeomorphism can be introduced. The first is obtained by using the homeomorphism of Euclidean spaces with finitely generated torsion-free abelian groups. For a suitable topologization of a lattice of points in n-space, connected sets are sets of consecutive integers.

The discrete analogous of topological homeomorphism can be defined such that for spaces of dimension 2 one can exhibit classes of equivalence using local connectivity properties of images comparing and computing only connectivity and order of connectivity one can provide algorithms which associate to an arbitrary image an equivalent one displaying certain features of regularity.

Other definitions (e.g. reducibility) of topological homeomorphism are stronger equivalence relations in the sense that they preserve more of the topological invariants of figures homotopy, homology, and cohomology groups. These different approaches coincide for connected figures in discrete spaces of dimension 2.

In this formalism isomorphic images are homeomorphic ones, local connectivity properties are the invariants which allow to check whether or not two objects of the category are isomorphic. In this settings invariants of the category forget all but connectivity equivalence. If one considers reducibility as defining isomorphism, then one gets “group valued” invariants (homotopy and homology groups).

The algebraic formalism of pattern generation and recognition needed and has in recent years partially been completed by topological intrinsical definitions of the notions involved, as well as a certain number of results. However, the link between the two points of view missed. This paper presents for the first time a general formalism of pattern definition and recognition using category theory, which unifies the two (algebraic and topological) convergent themes of this field. It states and proves the condition under which a recognition category can be associated to a category of images, and also defines algebraic and topological invariants and recognition functions in their general sense.

### C. *The Algebraic Approach to Recognition, Classification and Forecasting Problems by Yu.Zhuravlev*

"The Algebraic Approach to Recognition, Classification and Forecasting Problems" (Yu.Zhuravlev) [109] is mathematical set-up of a pattern recognition problem, correctness and regularity conditions, multiple classifiers.

One of the topical problems in image recognition is searching for an algorithm that would provide a correct classification of an image by its description (i.e. the algorithms that produce zero errors on a control set of objects). The approach to image recognition that is developed by the present authors is a specialization of the algebraic approach to recognition and classification problems originally designed by Yu.I. Zhuravlev [109]. The relational for this approach is the fact that there are no accurate mathematical models for weakly formalized fields such as geology, biology, medicine, and sociology. However, in many cases, inexact methods based on heuristic considerations are practically effective. Therefore, it is sufficient to construct a family of such heuristic algorithms for solving appropriate problems and then construct the algebraic closure of this family. The existence theorem has been proved, which states that any problem among the set of problems associated with the study of poorly formalized situations is solvable in this closure [109].

Suppose we give a certain set of admissible patterns described by n-dimensional vectors of features. The set of admissible patterns is covered by a finite number of subsets, called classes. Let there exist l classes $K_1, \ldots, K_l$. There is a recognition algorithm A that constructs an l-dimensional information vector by an n-dimensional description vector. Recall that an information vector is the vector of membership of an object in the classes in which the values of elements of the information vector 0, 1, Δ are interpreted, according to [109], as "an object does not belong to the class," "an object belongs to the class," and "the algorithm cannot determine whether or not an object belongs to the class." We will assume that each recognition algorithm $A \in \{A\}$ can be represented as a sequential execution of algorithms B and C, where B is the recognition operator that transforms learning information and the description of an admissible object into a numerical vector, called the estimate vector, and C is the decision rule that transforms an arbitrary numerical vector into an information vector.

The operation of the recognition algorithm can be schematically represented as follows.

Feature description of an object $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$

↓ Recognition algorithm B

Vector of estimates for a class $\beta = (\beta_1, \beta_2, \ldots, \beta_l)$

↓ Decision rule C

Information vector $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_l)$.

Thus, during the solution of a recognition problem, the object of recognition, i.e., an image, is described by three different vectors: the n-dimensional vector of features, the l-dimensional vector of estimates for a class, and the l-dimensional information vector.

Let us briefly recall the pattern recognition problem in the standard statement that was formulated by Zhuravlev [135].

$Z(I0, S_1, \ldots, S_q, P_1, \ldots, P_l)$ is a recognition problem, where $I_0$ is admissible initial information; $S_1, \ldots, S_q$ is the set of admissible objects described by feature vectors; $K_1, \ldots, K_l$ is a set of classes; and $P_1, \ldots, P_l$ is a set of predicates on the admissible objects, $P_i = P_i(S)$, $i = 1, 2, \ldots, l$. Problem $Z$ consists in finding the values of the predicates $P_1, \ldots, P_l$.

**Definition**. An algorithm is said to be correct for problem $Z$ if the following equality holds:

$$A\left(I, S_1, \ldots, S_q, P_1, \ldots, P_l\right) = \left\| \alpha_{ij} \right\|_{q \times l},$$ where $\alpha ij = Pj(Si)$.

One of the main tasks of pattern recognition is searching for an algorithm that correctly solves the image recognition problem. Zhuravlev proves the existence theorem for such an algorithm stating that the algebraic closure of AECs for the image recognition problem is correct [135]. AECs are based on the formalization of the concepts of precedence or partial precedence: an algorithm analyzes the proximity between the parts of descriptions of earlier classified objects and the object to be recognized.

Suppose we are given standard descriptions of objects $\left\{\tilde{S}\right\}, \tilde{S} \in K_j$ and $\left\{S'\right\}, S' \notin K_j$, and a method for determining the degree of proximity between certain parts of the description of $S$ and the corresponding parts of the descriptions $\left\{I\left(\tilde{S}\right)\right\}, \left\{I\left(S'\right)\right\}$; $S, j=1,2,\ldots,l,$ is the object of recognition. Calculating estimates for the proximity between the parts of the descriptions $\left\{I\left(\tilde{S}\right)\right\}$ and $\left\{I\left(S'\right)\right\}$ and, respectively, between $I\left(S\right)$ and $I\left(S'\right)$, one can construct a generalized estimate for the proximity between S and the sets of objects можно построить обобщённую оценку близости между $S$ и множествами объектов $\left\{\tilde{S}\right\}, \left\{S'\right\}$ (in the simplest case, the generalized estimate is equal to the sum of estimates for the proximity between the parts of descriptions). Then, using the set of estimates, one forms a general estimate of an object over a class, which is precisely the value of the membership function of the object in the class.

For the algebraic closure of the AECs, the following existence theorem for an AEC is proved, which correctly solves recognition problem Z.

**Theorem**. Suppose that natural assumptions on the difference between the descriptions of classes and recognition objects hold for the vectors of features in recognition problem Z. Then the algebraic closure of the class of AECs is correct for problem Z.

The image recognition problem is one of the classical examples of problems with incompletely formalized and partially contradictory data. This suggests that the application of an algebraic approach to image recognition may lead to important results; hence, the "algebraization" of this field is the most promising approach for development of the required

mathematical apparatus for the analysis and estimation of information represented by images.

When the recognition objects are images, this theorem cannot be directly applied. There are several reasons for this. First, representation of an image by a vector of features (as in the case of a standard recognition object) often leads to the loss of a considerable part of the information about the image and, consequently, to an incorrect classification. Second, the existence of equivalence classes is an essential difference between the image recognition problem and the recognition problem in the classical formulation.

Passage from the algebra of pattern recognition algorithms to an algebra of image recognition algorithms requires a choice, first, of algorithms used as elements of algebra, and second, of algebraic representations of images that make it possible to formalize the task of choosing descriptors. It is expedient to select representations taking into account the possibility of combining the initial information and algorithms of different types. For the first time, the idea of a combination of qualifiers with optimization of their operation by algebraic correction was suggested and justified by Yu.I. Zhuravlev [109]. The complex of mathematical methods related to synthesis and research of such qualifiers is known under the common title “Algebraic Approach to Tasks of Recognition, Classification, and Prediction.” In the English-Language literature for the designation of qualifiers, the term Multiple Classifiers [109] is used. Recently, quite interesting results have been achieved in the field of theoretical-informational analysis of combined qualifiers [27], developments of specific strategies for merging algorithms [45], and usage of methods of code theory in tomography [105].

Image analysis and understanding have a certain peculiarity, due to which the use of the Zhuravlev algebraic approach in the general form is inconvenient. The reasons are the following:

- the character of the considered problem is not taken into account if algebraic methods are applied to the information represented in the form of images;
- the results of application of the theory cannot always be simply interpreted;
- there are many natural transformations of images which are easily interpreted from the user’s point of view (for instance, rotation, contraction, stretching, color inversion, etc.) but are hardly representable by standard algebraic operations.

The necessity arises of using algebraic tools to record natural transformations of images. Moreover, the algebraization of image analysis and understanding must include the construction of algebraic descriptions of both the images themselves and algorithms for their processing, analysis, and recognition.

Analyzing the publications related to applications of algebraic methods to image analysis and understanding, we distinguish the following advantages of unified representation of images and algorithms for their processing and analysis:

- construction of unified representations for descriptions of images;
- efficiency of transition from input data in the form of images to different formal models of the images;
- naturalness of uniting the algebraic representation of the information with the developed algebraic tools for pattern recognition, which has been successfully
- employed;
- the possibility of using the methods of mathematical modeling employed in applied domains to which the processed images belong;
- the possibility of using the image descriptions in the form of group-theoretic representations;
- naturalness of uniting the methods of structural analysis of images with tools of probabilistic analysis;
- the possibility of a formalized description for problems of parallelizing with due regard for the specifics of particular computational architectures.

## V. IMAGE ALGEBRAS

Mathematical morphology [2, 5, 8-14, 23, 40, 54-56, 73-76, 91, 96-103,], proposed by Minkowski and Hadwiger and developed by Matheron and Serra, seems to be the first attempt to create a theoretical apparatus that allows one to describe many widespread operations of image processing in the composition of a rather small set of standard simple local operations. Such representations allow one to formalize the choice of procedures for image processing and are convenient for implementation on parallel architectures. It might have been the success of mathematical morphology that initiated numerous attempts of algebraization both in the domain of algorithm representations and in closed domains. Mathematical morphology is an efficient tool for uniform representation of local operations of image processing, analysis, and understanding in terms of algebras over sets. It makes it possible to describe algorithms for image transformations in terms of four basic local operations, namely, those of erosion, dilatation, opening, and closing; moreover, any two of these operations form a basis, in terms of which the other two operations may easily be expressed. This is very convenient for the development of software systems, in which the user can quickly design particular algorithms from basic blocks.

On the basis of mathematical morphology, Sternberg [98-100] introduced the concept of an image algebra.

The image algebra made it possible to represent algorithms for image processing in the form of algebraic expressions, where variables are images and operations are geometrical and logical transformations of the images. It is known that the possibilities of mathematical morphology are very limited. In particular, many important and widely used operations of image processing (feature extraction based on the convolution operation, Fourier transforms, use of the chain code, equalization of a histogram, rotations, recording, and nose elimination), except for the simplest cases, can hardly (if ever) be realized in the class of morphological operations.

The impossibility of constructing a universal algebra for tasks of image processing on the basis of the morphological

algebra may be explained by the limitation of the basis consisting of the set-theoretical operations of addition and subtraction in Minkowski's sense.

It is known that this basis has the following drawbacks [62]:

- complicated realization of widely used operations of image processing;
- impossibility of establishing a correspondence between the operations of mathematical morphology and linear algebra;
- impossibility of using mathematical morphology for transformations between different algebraic structures, in particular, sets including real and complex numbers and vector quantities.

These problems have been solved in the standard image algebra (IA) by G. Ritter [77, 78] on the basis of a more general algebraic representation of operations of image processing and analysis. Standard Image Algebra by G.Ritter is a unified algebraic representation of image processing and analysis operations. Image algebra generalizes the known local methods for image analysis, in particular, mathematical morphology, and provides the following advantages as compared with mathematical morphology:

- it makes it possible to work with both real and complex quantities;
- it allows one to include both scalar and vector data into the input information;
- it makes image-algebra structures consistent with linear structures;
- it provides a more accurate and complete description of its operations and operands;
- with the help of a special structure "template," composite operations of image processing are divided into a number of parallel simplest operations.

The bottleneck in applications of methods of image algebra to image recognition is the choice of the sequence of algebraic operations and templates for representation of composite operations of image processing.

At present, this choice is based, as a rule, on general representations of the character of images and tasks. Deficiencies of this approach are obvious: first, it is subjective and its success depends to a great extent on the user's experience and, second, it is intended to solve a specific narrow class of problems. Image algebra generalizes the known local methods for image analysis, in particular, mathematical morphology.

Investigations in the area of algebraization and image analysis of the 1970–1980s represent a source of development of the descriptive image algebra (DIA). DIA by I.Gurevich is a unified algebraic language for describing, performance estimating and standardizing representation of algorithms for image analysis, recognition and understanding as well as image models.

An object that lies most closely to the developed mathematical object is the image algebra proposed and developed by Ritter [78]. Ritter's main goal in developing the image algebra is the design of a standardized language for description of algorithms for image processing intended for parallel execution of operations. A key difference in the new image algebra from the standard Ritter image algebra is that DIA is developed as a descriptive tool, i.e., as a language for description of algorithms and images rather than a language for algorithm parallelizing.

The conceptual difference of the algebra under development from the standard image algebra is that objects of this algebra are (along with algorithms) descriptions of input information. DIA generalizes the standard image algebra and allows one to use (as ring elements) basic models of images and operations on images or the models and operations simultaneously. In the general case, a DIA is the direct sum of rings whose elements may be images, image models, operations on images, and morphisms. As operations, we may use both standard algebraic operations and specialized operations of image processing and transformations represented in an algebraic form. To use DIA actively, it is necessary to investigate its possibilities and to attempt to unite all possible algebraic approaches, for instance, to use the standard image algebra as a convenient tool for recording certain algorithms for image processing and understanding or to use Grenander's concepts for representation of input information.

In the 1980s, Sternberg formalized the notion "image algebra" and introduced the following definition.

Image algebra is the representation of algorithms for image processing on a cellular computer in the form of algebraic expressions whose variables are images and whose operations are procedures for constructing logical and geometrical combinations of images.

This image algebra is described on the basis of mathematical morphology and is identified by the author with mathematical morphology. In 1985, Sternberg [99] noted that the languages for image processing were being developed for each processor architecture and none of them has been created for one computer and run on another. However, there are explicit language structures that satisfy the same principles. It is for description of these structures that image algebra (or mathematical morphology) appeared. Ritter's image algebra generalizes mathematical morphology, unites the apparatus of local methods for image analysis with linear algebra, and generates more complex structures. Examples of such structures are templates and morphological algorithms. In [77], various operations and operands of standard image algebra are described, as well as applications of these structures to actual problems. Since the standard image algebra does not just generalize mathematical morphology, but is a wider and more convenient structure, the language of image algebra admits both implementation of known algorithms and design of new algorithms. The structure of the standard image algebra may be extended by introducing new operations. Hence, it may be successfully applied in the cases where a satisfactory result

cannot be obtained with the help of morphology and linear algebra.

A standard image algebra is a heterogeneous (or multivalued) algebra with a complex structure of operands and operations if the basic operands are images (sets of points) and values and characteristics related to these images (sets of values related to these points).

Analyzing the existing algebraic apparatus, we came to the statement of the following requirements on the language designed for recording algorithms for solving problems of image processing and understanding:

- the new algebra must make possible processing of images as objects of analysis and recognition;
- the new algebra must make possible operations on image models, i.e., arbitrary formal representations of images, which are objects and, sometimes, a result of analysis and recognition; introduction of image models is a step in the formalization of the initial data of the algorithms;
- the new algebra must make possible operations on main models of procedures for image transformations;
- it is convenient to use the procedures for image modifications both as operations of the new algebra and as its operands for construction of compositions of basic models of procedures.

An algebra is called a descriptive image algebra if its operands are either image models (for instance, as a model, we may take the image itself or a collection of values and characteristics related to the image) or operations on images, or models and operations simultaneously.

It should be noted that, due to the variety of "algebras", we should indicate which algebra is meant in definition of DIA. For the generality of the results and extension of the domain of applications of the new algebra, to define DIA with one ring, we use the definition of the classical algebra of Van der Waerden [107].

Thus, a DIA with one ring must satisfy the properties of classical algebras. A DIA with one ring is a basic DIA, because it contains a ring of elements of the same nature, i.e., either a ring of image models or a ring of operations on images.

To design efficient algorithmic schemes for image analysis and understanding, it is necessary to investigate different types of operands and different types of operations applicable to the chosen operands, which generate the DIA.

## VI. CONTRIBUTION OF THE RUSSIAN MATHEMATICAL SCHOOL

This section presents the most important original results on algebraic tools for pattern recognition and image analysis including algebras on algorithms, algebraic multiple classifiers, algebraic committees of algorithms, combinatorial algorithms for recognition of 2-D data, descriptive image models, 2-D formal grammars.

### *A. Zhuravlev's school*

The "algebraic approach" to solve the tasks of classification and/or pattern recognition was developed in the school of Yu. Zhuravlev starting from 1960s as means to build the correct algorithms (i.e. the algorithms that produce zero errors on a control set of objects) over specified sets of features. Within the framework of the algebraic approach, the algorithms are built as compositions of type $A = C \circ B$ where A is the entire algorithm, B is an operator "base classifier" that maps the feature space into a matrix of estimates of the assignments of the objects' classes, C is the "decision rule" operator that maps the matrix of estimates into binary matrix of the answers of the entire algorithm *A.*

In the framework of scientific school of Yu.I.Zhuravlev several essential results were obtained in algebraic direction by V.L.Matrosov [48-51], by K.V.Rudakov [83-90] and V.D.Mazurov [58-61].

### *B. Category Theoretic Approach by K.V.Rudakov*

Within the category theoretic approach to the algebraic approach, developed by K.V. Rudakov [83-90], the composition scheme of the entire algorithm is complemented by corrective operations ("aggregative function") that are built over the space of the cartesian products of the answers of the B-operators. The aggregative functions allow more flexibility in achieving the correctness of algorithms over arbitrary selection of the training/control sets of objects.

The application of the category theoretical apparatus the constructions of the algebraic approach allowed to demonstrate universal nature of the constructions of the type $A = C \circ F\left(B_1,...,B_p\right)$ which thus guarantees existence of a correct solution at any non-contradictory sets of objects ($B_1,\ldots,B_p$ – base classifiers).

This approach allows significantly increase the accuracy of classification and was applied in a number of fields: Monitoring trade markets at MICEX (Moscow Interbank Currency Exchange) and other tasks of the analysis and prognosis of the time series, the tasks of text analysis ("Antiplagiat" system), problems of bioinformatics, etc.

The purpose of system "Antiplagiat" is detecting citations in documents: a) protecting intellectual property from unauthorized copying, b) finding duplicates and similar documents in vast storages. This system is already used at Higher School of Economics, Moscow Institute of Economics Management and Law, Moscow State Pedagogical University, Academy of Budget and Treasury of Finance Ministry of Russian Federation. A problem-oriented formalism for describing the problem of protein secondary structure recognition was developed. Experiments were based on 165000 precedents found in the Protein Data Bank (www.rcsb.org). The most informative feature values were effectively selected via solvability analysis that ensure more than 95% solvability of the recognition problem on an arbitrary set of objects of sufficient size.

*C. The Method of Committee by V.D.Mazurov*

Recognition and forecasting are the fundamental concepts of mathematical modeling in wide range of economic, social, natural (e.g. geophysical) phenomena. These concepts lie in the basis of decision-making. Committee constructions represent a class of discrete generalizations of the notion of solution for problems that can be both feasible and infeasible (contradictory). The first result in this area belongs to C.Ablow and D.Kaylor [1], which formulated the committee solution concept for a system of linear inequalities in explicit terms.

A finite sequence ($q$-tuple of vectors in $\mathbb{R}^n$) $Q = (x_1, \dots, x_q)$ is called *a committee (generalized) solution* (or just a committee) of the system

$$a_j^T x < b_j \quad (j = 1,2, \dots, m) \qquad (1),$$

if for any $j$ the major part of elements of $Q$ satisfy the $j$-th inequality. The number $q$ is called *a number of the elements (length)* of the committee $Q$. Finally, the committee $Q$ with the minimum for system (1) length $q$ is called *a minimum committee*.

V. D. Mazurov (and his followers) [58-61, 42, 43] have started from this simple definition and developed the elegant mathematical theory of discrete approximations for infeasible systems of constraints and collective learning algorithms in pattern recognition. Now this theory is known now as *The Method of Committees.*

There are existence theorems:

**Theorem 1** [59]. System (1) has a committee solution if and only if any its 2-inequalities subsystem is feasible.

**Theorem 2.** Let any subsystem of rank $k$ (of system (1)) has a committee solution with at most $q$ elements. Then the system itself also has a committee solution of at most $2q\left\lceil \frac{\lfloor (m-1)/2 \rfloor}{k} \right\rceil + 1$ elements.

**Theorem 3**. If any $(k+1)$-subsystem of system(1) is feasible then this system has a committee solution of at most $2\left\lceil \frac{\lfloor (m-1)/2 \rfloor}{k} \right\rceil + 1$ elements.

These results can be extended to infinite dimensional spaces. Let $f_1, \dots, f_m$ be real-valued functionals over some Banach space $B$. Consider a system of inequalities

$$f_j(x) > 0, \quad (j = 1,2, \dots, m) \qquad (2)$$

**Theorem 4.** Let $f_1, \dots, f_m$ be Frechet differentiable at the point $x_0 = 0$ such that

1) $f_j(x_0) = 0, \quad j = 1,2, \dots, m,$
2) rank $r$ of the system of linear functionals $f_j'(0)$ is positive.
3) for a system

$$f_j'(0)x > 0 \quad (j = 1,2, \dots, m),$$

any $(k+1)$-subsystem for $0 \le k < r$ be feasible.

Then system (2) has a committee solution.

Let $X$ be a real vector space (e.g., $\mathbb{R}^n$), which is called *feature space* and can be interpreted as a space of measurements over objects to be recognized, and $Y = \{0,1, \dots, K-1\}$ be a finite set of *patterns.* Any function $f: X \to Y$ is called *a decision rule (or classifier)* and can be used to classify an object $o$ using a vector $x = x(o)$ obtained during some measurements over it. Denote a family of the feasible decision rules by $\mathcal{F}$.

A decision rule $F[f_1, \dots, f_q]$, $f_i \in \mathcal{F}$ is called *committee decision rule* if for any $x \in X$, $F(x) = y$ if and only if $y$ is the maximum element of $Y$, for which the most part of $f_i(x) = y$.

Learning in the class of committee decision rules is just searching for the most admissible number $q$ and functions $f_1, \dots, f_q$ for training sample and closely related to the concept of *separating committee*. Introduce it for the simplest case of $K = 2$.

A finite sequence $Q = (f_1, \dots, f_q)$ is called *a separating committee* for finite subsets $A, B \subset X$ if the major part of $f_i(a) = 1$ for any $a \in A$ and, conversely, the major part of $f_j(b) = 0$ for any $b \in B$.

The most part of results have been obtained in the field of learning in the class *affine* decision rules, where $f_i(x) = w_i^T x + b_i$ for some vector $w_i$ and bias $b_i$.

**Theorem 5.** Let $A, B \subset X$. Affine separating committee for the sets $A$ and $B$ exists if and only if $A \cap B = \emptyset$.

**Theorem 6.** Let $\mathcal{F}_q$ be family of linear committee decision rules over $\mathbb{R}^n$, then $VCD(\mathcal{F}_q) = O(nq)$.

As stated in Theorem 6, any time during the learning procedure (by virtue of the well known Vapnik-Chervonenkis theory) it is important to construct a committee decision rule with the least possible $q$ (for the given training sample). This fact motivates studying the following combinatorial optimization problem and related problems.

**Minimum affine separating committee (MASC) problem**. For finite $A, B \subset \mathbb{R}^n$ it is required to find an affine separating committee $Q = (f_1, \dots, f_q)$ with the least possible $q$.

This is a list of the selected results.

1. MASC problem is strongly NP-hard, and remains intractable even under the following additional constraints:
   a. Dimensionality n>1 is fixed
   b. $A, B \subset \{-1,0,1\}^n$
   c. $A, B$ are in the general position
2. MASC problem is solvable in a polynomial time if:
   a. n=1
   b. Sets $A, B$ are induced by sets that are uniformly distributed (by D.Gale) on unit sphere.
3. MASC problem is poorly approximable. Particularly, it does not belong to APX approximability class (unless $P = NP$) and is MaxSNP-hard for any fixed $n > 1$.
4. Several polynomial approximation algorithms were developed. The best know approximation guarantee is $O(n)$.

There are some other results:

1. Concept of *Hypergraph of Maximal (by Inclusion) feasible subsystems (HMFS)* of the system in question. Characterization theorem for HMFS-s. Classification of the minimal committee generalized solution in terms of their HMFS. Existence theorems of committee solutions in terms of HMFS.
2. Antagonistic game against nature on the basis of committee solutions existence is studied. Equilibrium conditions were obtained.
3. Computational complexity and approximability of several combinatorial optimization problems related to affine separating committees were investigated.

### *D. Algebraic method of analysis and estimation of information represented as signals*

Apart from basic researches of Yu.I.Zhuravlev scientific school there are significant number of papers concerned with algebraic methods of analysis and estimation of information represented as signals, in partially V.G.Labunec [46], Yu.P.Pityev [72], I.N.Sinicyn [92], Ya.A.Furman [19-22], V.M.Chernov [6, 7, 17].

For example, Ya.A.Furman [19-22] learns methods and tools for handling complex and hypercomplex signals. He considers the methodology of vector signal processing theory: the basis of information, the signal and its mathematical model, the mathematical apparatus of the theory of signal processing, the vector-geometrical representation of signals, the random vector signals, the scalar multiplication in problems of processing of the vector signals, the vector product of vectors, the cartesian reference system are considered.

New types of signals (complex and quaternionic) are introduced: these signals are used for recognition of boundary points (contours) of the image and the image analysis. The properties of the scalar multiplication, the orthogonal basis, the questions of spectral and correlation analysis, the questions of matched filtering are described for the each signal.

Furman considered the discrete complex signals. The assignment of complex signals, the complex numbers as elements of complex linear space, the spectral analysis of complex digital signal, the correlation functions of complex digital signal, the contour matched filtering are considered.

And the most interesting part of his research is devoted to the discrete quaternion signals. The hypercomplex numbers, the association of quaternions with complex numbers, the scalar product of the quaternion, the rotation of vectors in three dimensions space, the quaternion discrete signals, the orthogonal basis in the quaternionic space, the spectral representation of the discrete quaternion signals, the decomposition of discrete quaternion signals, the correlation functions of discrete quaternion signals, the matched filtering of discrete quaternion signals, the conjugate-matched filtering of discrete quaternion signals are considered.

## VII. Descriptive Approach to image analysis

It was largely the necessity to solve complex recognition problems and develop structural recognition methods and specialized image languages that generated the interest in formal descriptions–models of initial data and formalization of descriptions of procedures of their transformation in the area of pattern recognition (and especially in image recognition in 1960s).

As for the substantial achievements in this “descriptive” line of study, we mention publications by A. Rosenfeld [81, 82], T. Evans [15, 16], R. Narasimhan [64-67], R. Kirsh [44], A. Shaw [93, 94], H. Barrow, A. Ambler, and R. Burstall[4], S. Kaneff [41]. In 1970s, Yu. I. Zhuravlev proposed the so called “Algebraic Approach to Recognition and Classification Problems” [109], where he defined formalization methods for describing heuristic algorithms of pattern recognition and proposed the universal structure of recognition algorithms. In the same years, U. Grenander stated his “Pattern Theory” [24-26], where he considered methods of data representation and transformation in recognition problems in terms of regular combinatorial structures, leveraging algebraic and probabilistic apparatus. Both approaches dealt with the recognition problem in its classical statement and did not touch upon representation of initial data in the form of images.

Then, up to the middle of 1990s, there was a slight drop in the interest in descriptive and algebraic aspects in pattern recognition and image analysis.

The main intention of DA is to structure different techniques, operations and representations being applied in image analysis and recognition. The axiomatics and formal constructions of DA establish conceptual and mathematical base for representing and describing images and its analysis and estimation. The DA provides a methodology and a theoretical base for solving the problems connected with the development of formal descriptions for an image as a recognition object as well as the synthesis of transformation procedures for an image recognition and understanding. The analysis of the problems is based on the investigation of inner structure and content of an image as a result of the procedures “constructing” it from its primitives, objects, descriptors, features and tokens, and relations between them.

This section contains a brief description of the principal features of the DA needed to understand the meaning of the introduction of the conceptual apparatus and schemes of synthesis of image models proposed to formalize and systematize the methods and forms of representation of images.

The automated extraction of information from images includes (1) automating the development, testing, and adaptation of methods and algorithms for the analysis and evaluation of images; (2) the automation of the selection of methods and algorithms for analyzing and evaluating images; (3) the automation of the evaluation of quality and adequacy of the initial data for solving the problem of image recognition; and (4) the development of standard technological schemes for detecting, assessing, understanding, and retrieving images.

The automation of information extraction from images requires complex use all the features of the mathematical apparatus used or potentially suitable for use in determining transformations of information provided in the form of

images, namely in problems of processing, analysis, recognition, and understanding of images.

Experience in the development of the mathematical theory of image analysis and its use to solve applied problems shows that, when working with images, it is necessary to solve problems that arise in connection with the three basic issues of image analysis, i.e., (1) the description (modeling) of images; (2) the development, exploration, and optimization of the selection of mathematical methods and tools for information processing in the analysis of images; and (3) the hardware and software implementation of the mathematical methods of image analysis.

The main purpose of the DA is to structure and standardize a variety of methods, processes, and concepts used in the analysis and recognition of images.

The DA is proposed and developed as a conceptual and logical basis of the extraction of information from images. This includes the following basic tools of analysis and recognition of images: a set of methods of analysis and recognition of images, reducing images to a form suitable for recognition (RIFR) techniques, conceptual system of analysis and recognition image, DIM classes, the DIA language, statement of problems of analysis and recognition of images, and the basic model of image recognition.

The main areas of research within the DA are (1) the creation of axiomatics of analysis and recognition of images, (2) the development and implementation of a common language to describe the processes of analysis and recognition of images (the study of DIA), and (3) the introduction of formal systems based on some regular structures to determine the processes of analysis and recognition of images (see ([28, 29]).

Mathematical foundations of the DA are as follows: (1) the algebraization of the extraction of information from images, (2) the specialization of the Zhuravlev algebra [109] to the case of representation of recognition source data in the form of images, (3) a standard language for describing the procedures of the analysis and recognition of images (DIA) [31, 36], (4) the mathematical formulation of the problem of image recognition, (5) mathematical theories of image analysis and pattern recognition, and (6) a model of the process for solving a standard problem of image recognition. The main objects and means of the DA are as follows: (1) images; (2) a universal language (DIA); (3) two types of descriptive models, i.e., (a) an image model and (b) a model for solving procedures of problems of image recognition and their implementation; (4) descriptive algebraic schemes of image representation (DASIR); and (5) multimodel and multiaspect representations of images, which are based on generating descriptive trees (GDT).

The basic methodological principles of the DA are as follows: (1) the algebraization of the image analysis, (2) the standardization of the representation of problems of analysis and recognition of images, (3) the conceptualization and formalization of phases through which the image passes during transformation while the recognition problem is solved, (4) the classification and specification of admissible models of images (DIM), (5) RIFR, (6) the use of the standard algebraic language of DIA for describing models of images and procedures for their construction and transformation, (7) the combination of algorithms in the multialgorithmic schemes, (8) the use of multimodel and multiaspect representations of images, (9) the construction and use of a basic model of the solution process for the standard problem of image recognition, and (10) the definition and use of nonclassical mathematical theory for the recognition of new formulations of problems of analyzing and recognizing images.

Note that the construction and use of mathematical and simulation models of studied objects and procedures used for their transformation is the accepted method of standardization in the applied mathematics and computer science.

A more detailed description of methods and tools of the DA obtained in the development of its results can be found in [30, 37, 38].

## VIII. CONCLUSION

Practical application of the algebraic instruments DA was demonstrated: we have shown how to build, by means of DIA, the model of a technology for automating diagnostic analysis of cytological preparations of patients with tumors of the lymphatic system. This model has been used for the creation of software for application of this technology, its testing, and comparison of results.

The main contribution is construction of a model for a method ensuring a unified representation of the technology, instead of development of a method for solving a medical task. This work, thus, solves a dual task: first, it represents a technology in the form of a well-structured mathematical model and, second, shows how DIA can be used in an image analysis task.

In the future, DA and its main instruments— DIA, DIM and GDT—will be applied to constructing models of an information technology for automation of diagnostic analysis of medical images in other areas of medicine.

## ACKNOWLEDGMENT

This work was supported in part by the Russian Foundation for Basic Research (projects no. 14–01–00881), by the Presidium of the Russian Academy of Sciences within the program of the Department of Mathematical Sciences, Russian Academy of Sciences “Algebraic and Combinatorial Methods of Mathematical Cybernetics and Information Systems of New Generation” (“Algorithmic schemes of descriptive image analysis”) and “Information, Control, and Intelligent Technologies and Systems” (project no. 204).

## REFERENCES

[1] Ablow C.M., Kaylor D.J. “Inconsistent Homogeneous Linear Inequalities”, Bull. Amer. Math. Soc., 1965, vol. 71, no. 5, p. 724.

[2] C.S. Araujo. “Novel Neural Network Models for Computing Homothetic Invariances: An Image Algebra Notation”, /Journal of Mathematical Imaging and Vision, Vol. 7, Kluwer Academic Publishers. Manifactured in The Netherlands, 1997.- pp. 69-83.

[3] K.E. Batcher. “Design of a massively parallel processor”, IEEE Transactions on Computers, 29(9):836-840, 1980.

[4] H.G. Barrow, A.P. Ambler, and R.M. Burstall. “Some Techniques for Recognizing Structures in Pictures”, Frontiers of Pattern Recognition (The Proceedings of the International Conference on Frontiers of Pattern Recognition, ed. Satosi Watanabe), Academic Press, New York, London.- 1972.-pp. 1-30.

[5] G. Birkhoff, J.D. Lipson. “Heterogeneous Algebras”, Journal of Combinatorial Theory, Vol.8, 1970. - pp. 115-133.

[6] V.M.Chernov. “Clifford algebras are group algebras projections”, /E.Bayro-Corrochano, G.Sobczyk (Eds) Advanches in Geometric Algebra with Applications in Science and Engineering.-Birkhauser, Boston, 2001.- pp.467-482.

[7] V.M. Chernov. “On defining equations for the elements of associative and commutative algebras”, D.Pavlov, Gh.Atanasiu, V. Balan (Eds) «Space-Time Structure. Algebra and Geometry. Lilia Print, 2007, pp.182-188.

[8] T.Crimmins, W. Brown. “Image algebra and automatic shape recognition”, IEEE Transactions on Aerospace and Electronic Systems, vol. 21, no. 1, January 1985.-pp. 60-69.

[9] J.L. Davidson. “Classification of lattice transformations in image processing”, Computer Vision, Graphics, and Image Processing: Image Understanding, vol. 57, no.3, May 1993.-pp. 283–306.

[10] M.J.B. Duff, D.M. Watson, T.J. Fountain, and G.K. Shaw. “A cellular logic array for image processing”, Pattern Recognition, vol.5, no.3, June 1973.-pp. 229–247.

[11] E.R. Dougherty and D.Sinha. “Computational Gray-scale Mathematical Morphology on Lattices (A Comparator-based Image Algebra). Part 1: Architecture”, Real-Time Imaging, vol. 1, Academic Press Limited, 1995.- pp. 69-85.

[12] E.R. Dougherty and D.Sinha. “Computational Gray-scale Mathematical Morphology on Lattices (A Comparator-based Image Algebra). Part 2: Image Operators”, Real-Time Imaging, vol. 1, Academic Press Limited, 1995.-pp. 283-295.

[13] E.R. Dougherty. “A homogeneous unification of image algebra. Part I: the homogenous algebra”, Imaging Science, vol.33, no.4, 1989 - pp.136-143.

[14] E.R. Dougherty. “A homogeneous unification of image algebra. PartII: unification of image algebra”, Image Science, Vol.33, no.4, 1989.-pp. 144-149.

[15] T.G. Evans. “A Formalism for the Description of Complex Objects and ist Implementation”, Proceedings of the Firth International Conference on Cybernetics, Namur, Belgium, September, 1967.

[16] T.G. Evans. “Descriptive Pattern Analysis Techniques: Potentialities and Problems”, Methodologies of Pattern Recognition (The Proceedings of the International Conference on Methodologies of Pattern Recognition).- Academic Press, New York, London, 1969.-pp. 149-157.

[17] M.Felsberg, Th.Bulov, G.Sommer, V.M.Chernov. “Fast Algorithms of Hypercomplex Fourier Transforms”, G.Sommer (Eds) Geometric Computing with Clifford Algebras. Springer Verlag, 2000, pp. 231-254.

[18] K.S. Fu. “On syntactic pattern recognition and stochastic languages”. Frontiers of Pattern Recognition (S.Watanabe, ed.), Academic Press, New York, 1972.

[19] Ya.A.Furman. “Parallel Recognition of Different Classes of Patterns”, Pattern Recognition and Image Analysis, Pleiades Publishing, Ltd., Vol.19, No.3, pp.380-393, 2009.

[20] Ya.A.Furman. “Recognition of Vector Signals Represented as a Linear Combination”. Journal of Communications Technology and Electronics, 2010. Vol. 55, No.6. – pp. 627-638.

[21] Ya.A.Furman, I.L.Egoshina. “Inverse problem of rotation of three-dimensional vector signals”, Optoelectronics, Instrumentation and Data Processing. Vol. 46, No. 1, 2010. – pp. 37-45.

[22] Ya.A.Furman, R.V.Eruslanov, and I.L.Egoshina. “Recognition of Images and Recognition of Polyhedral Objects”, Pattern Recognition and Image Analysis, Pleiades Publishing, Ltd., vol.22, no.1, pp.196-209, 2012.

[23] P.D. Gader, M.A. Khabou, A. Koldobsky. “Morphological regularization neural networks”, Pattern Recognition, Vol. 33, 2000,- pp. 935-944.

[24] U. Grenander. Lectures in Pattern Theory. N.Y.: Sprinder-Verlag, 1976 V.1; 1978 V.2; 1981 V.3.

[25] U. Grenander. General Pattern Theory. A Mathematical Study of Regular Structure. Clarendon Press, Oxford, 1993.

[26] U. Grenander. Elements of Pattern Theory. The Johns Hopkins University Press, 1996.

[27] J. Grin, J. Kittler, P. Pudil, P. Somol. “Information Analysis of Multiple Classifier Fusion”, Multiple Classifier Systems. Second International Workshop, MCS 2001, Cambridge, UK, July 2001. Proceedings. Springer - Verlag, 2001.- pp. 168 - 177.

[28] I.B. Gurevich. “The Descriptive Framework for an Image Recognition Problem”, Proceedings of the 6th Scandinavian Conference on Image Analysis.- Pattern Recognition Society of Finland, 1989.- vol. 1. – P. 220 – 227.

[29] I.B. Gurevich. “Descriptive Technique for Image Description, Representation and Recognition”, Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications in the USSR.- MAIK “Interpreodika”, 1991.-vol. 1- P. 50 – 53.

[30] I.B. Gurevich. “The Descriptive Approach to Image Analysis. Current State and Prospects”, Proceedings of 14th Scandinavian Conference on Image Analysis.- Springer-Verlag Berlin Heidelberg, 2005.- LNCS 3540.- pp. 214-223.

[31] I.B.Gurevich, I.A. Jernova. “The Joint Use of Image Equivalents and Image Invariants in Image Recognition”, Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications. - 2003. - Vol. 13, No.4. - pp. 570-578.

[32] I.B. Gurevich and I.V. Koryabkina. “Comparative Analysis and Classification of Features for Image Models”, Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications. - MAIK "Nauka/Interperiodica"/Pleiades Publishing, Inc., 2006. - Vol.16, No.3. - P. 265-297.

[33] I.B. Gurevich, V.V. Yashina. “Operations of Descriptive Image Algebras with One Ring”, Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications. Pleiades Publishing, Inc. 2006. - Vol.16, No.3. - pp. 298-328.

[34] I.B. Gurevich and V.V. Yashina. “Computer-Aided Image Analysis Based on the Concepts of Invariance and Equivalence”, Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications. - MAIK "Nauka/Interperiodica"/Pleiades Publishing, Inc., 2006. - Vol.16, No.4. – pp.564-589.

[35] I. Gurevich, V. Yashina. “Descriptive Theory of Image Analysis. Models and Techniques”, 8th International Conference “Pattern Recognition and Image Analysis: New Information Technologies (PRIA-8-2007). Conference proceedings. In two volumes. – Yoshkar-Ola, 2007. - Vol.1. - P. 103-112.

[36] I.B. Gurevich and V.V. Yashina. “Descriptive Approach to Image Analysis: Image Models”, Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications. - MAIK "Nauka/Interperiodica"/Pleiades Publishing, Inc., 2008. - Vol.18, No.4. - P. 518-541.

[37] I.B. Gurevich, V.V. Yashina, I.V. Koryabkina, H. Niemann, and O. Salvetti. “Descriptive Approach to Medical Image Mining. An Algorithmic Scheme for Analysis of Cytological Specimens”, Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications. - MAIK "Nauka/Interperiodica"/Pleiades Publishing, Inc., 2008. - Vol.18, No.4. - P. 542-562.

[38] I.B.Gurevich, V.V. Yashina. “Descriptive Approach to Image Analysis: Image Formalization Space”, Pattern Recognition and Image Analysis, 2012, Vol. 22, No. 4, pp. 495-518.

[39] H. Hadwiger, “Uber Treffanzahlen bei translationsgleichen Eikorpern”, Arch. Math. 8 (1957), 212–213.

[40] R.M.Haralick, S.R. Sternberg, X. Zhuang. “Image Analysis Using Mathematical Morphology”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-9, No. 4, 1987. - pp.532-550.

[41] S.Kaneff. “Pattern Cognition and the Organization of Information”, Frontiers of Pattern Recognition (The Proceedings of the International Conference on Frontiers of Pattern Recognition, ed. Satosi Watanabe).- Academic Press, New York, London.- 1972.-pp. 193-222.

[42] M.Yu. Khachai. “On the Computational Complexity of the Minimum Committee Problem”, J. of Math. Model. and Algor. 2007. Vol. 6, no. 4. P. 547-561.

[43] M.Yu.Khachai. “Computational complexity of recognition learning procedures in the class of piecewise-linear committee decision rules”, Automation and Remote Control. 2010. Vol. 71, no. 3. P. 528-539.

[44] R. Kirsh. “Computer Interpretation of English Text and Picture Patterns”, IEEE-TEC, Vol. EC-13, No. 4, August, 1964.

[45] J. Kittler, F.M.Alkoot. “Relationship of Sum and Vote Fusion Strategies”. Multiple Classifier Systems. Second International Workshop, MCS 2001, Cambridge, UK, July 2001. Proceedings. Springer - Verlag, 2001. –pp. 339 - 348.

[46] V.G. Labunec. Algebraic theory of signals and systems (digital signal processing), Krasnoyarsk University, Krasnoyarsk, 1984.

[47] A. I. Malcev. Algebraic Systems. Nauka, Moscow, 1970; Springer-Verlag, Berlin, 1973.

[48] V.L. Matrosov. “Pair isomorphism of permissible objects in recognition problems”, USSR, Comput.Maths.Math.Phys., Printed in Great Britain, Vol.23, No.1, pp.123-127, 1983.

[49] V.L. Matrosov. “On the incompleteness of a model of algorithms for computing estimates”, USSR, Comput.Maths.Math.Phys, Printed in Great Britain, Vol.23, No.2, pp.128-136, 1983.

[50] V.L. Matrosov. “Lower bounds of the capacity of L-dimensional algebras of estimate-computing algorithms”, USSR, Comput. Maths.Math.Phys., Printes in Great Britain, vol.24, no.6, pp.182-188, 1984.

[51] V.L. Matrosov. “The capacity of polynomial expansions of a set of algorithms for calculating estimates”, USSR, Comput.Maths.Math.Phys., Printed in Great Britain, Vol.25, No.1, pp.79-87, 1985.

[52] G. Matheron. Random Sets and Integral Geometry, Wiley, New York, 1975.

[53] P. Maragos. “Algebraic and PDE Approaches for Lattice Scale-Spaces with Global Constraints”, International Journal of Computer Vision, vol.52, no.2/3, Kluwer Academic Publishers. Manifactured in The Netherlands, 2003.-pp.121-137.

[54] P. Maragos, R. Schafer. “Morphological skeletons representation and coding of binary images”, IEEE Transactions on Acoustics, Speech, and Signal Processing, vol.34, no.5, October 1986,-pp. 1228-1244.

[55] P. Maragos and R.W. Schafer. “Morphological filters Part I: Their set-theoretic analysis and relations to linear shift-invariant filters,” IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-35, August 1987.-pp. 1153–1169.

[56] P. Maragos and R.W. Schafer. “Morphological filters Part II : Their relations to median, order-statistic, and stack filters,” IEEE Transactions on Acoustics, Speech, and Signal Processing, ASSP-35, August 1987.-pp. 1170–1184.

[57] D. Marr. Vision, Freeman, New York, 1982.

[58] V.D. Mazurov. “A committee of a system of convex inequalities”, Siberian Mathematical Journal. 1998. Vol. 9, no. 2. P. 354-357.

[59] V.D. Mazurov. “Committees of inequalities systems and the recognition problem”, Cybernetics. 1971, Vol. 7, no. 3. P. 559-567.

[60] V.D. Mazurov and M.Yu.Khachai. “Committees of Systems of Linear Inequalities”, Automation and Remote Control. 2004. Vol. 65, no. 2. P. 193-203.

[61] V.D. Mazurov and M.Yu.Khachai. “Parallel computations and committee constructions”, Automation and Remote Control. 2007. Vol. 68, no. 5. P. 912-921.

[62] P. Miller. “Development of a mathematical structure for image processing”, Optical division tech. report, Perkin-Elmer, 1983.

[63] H. Minkowski. Geometrie der Zahlen (2 vol.), Teubner, Leipzig, 1896/1910.

[64] R. Narasimhan. “Syntax-Directed Interpretation of Classes of Pictures”, Community ACM, Vol.9, No.3, March.- 1966.

[65] R. Narasimhan. “Labeling Schemata and Syntactic Descriptions of Pictures”, Information and Control, Vol. 7, No. 2, June 1967.

[66] R. Narasimhan. “On the Description, Generalization and Recognition of Classes of Pictures”, NATO Summer School on Automatic Interpretation and Classification of Images, Pisa, Italy, Aug. 26-Sept.7, 1968.

[67] R.Narasimhan. “Picture Languages”, Picture Language Machines (ed. S.Kaneff).- Academic Press, London, New York.-1970.-pp. 1-30.

[68] J. von Neumann. “The general logical theory of automata”, Celebral Mechenism in Behavior: The Hixon Symposium, John Wiley & Sons, New York, NY, 1951.

[69] J. von Neumann. Theory of Self-Reproducing Automata. University of Illinois Press, Urbana, IL, 1966.

[70] M. Pavel. “Pattern Recognition Categories”, Pattern Recognition, 1976, Vol.8, No.3.- pp. 115-118.

[71] M. Pavel. Fundamentals of Pattern Recognition, New York, Marcell, Dekker, Inc., 1989.

[72] Yu.P. Pytiev. Method of mathematical modeling of measuring and computing systems, MAIK Nauka, Moscow (sec.ed), 2004 [in Russian].

[73] B. Radunacu, M.Grana, F.X. Albizuri, “Morphological Scale Spaces and Associative Morphological Memories: Results on Robustness and Practical Applications”, Journal of Mathematical Imaging and Vision, vol. 19, Kluwer Academic Publishers. Manifactured in The Netherlands, 2003.-pp. 113-131.

[74] G.X. Ritter, P. Sussner, and J.L. Diaz-de-Leon, “Morphological associative memories,” IEEE Trans. on Neural Networks, Vol. 9, No. 2, 1998.- pp. 281–292.

[75] G.X.Ritter, P.Sussner. “Introduction to Morphological Neural Networks”, Proceedings of ICPR 1996, IEEE, 1996.-pp. 709-716.

[76] G.X. Ritter, J.L. Diaz-de-Leon, and P. Sussner. “Morphological bidirectional associative memories”, Neural Networks, vol. 12, 1999.-pp. 851–867.

[77] G.X. Ritter, J.N. Wilson. Handbook of Computer Vision Algorithms in Image Algebra, 2-d Edition. CRC Press Inc., 2001.

[78] G.X. Ritter. Image Algebra. Center for computer vision and visualization, Department of Computer and Information science and Engineering, University of Florida, Gainesville, FL 32611, 2001.

[79] G.X.Ritter, P.D. Gader. “Image Algebra techniques for parallel image processing”, Parallel Distributed Computers, Vol.4, no.5, 1987.-pp.7-44.

[80] G.X. Ritter, J.N. Wilson, and J.L. Davidson. “Image Algebra: An Overview”, Computer Vision, Graphics, and Image Processing, vol.49, 1990.- pp.297-331.

[81] A. Rosenfeld. Picture Languages. Formal Models for Picture Recognition.-Academic Press, New York, San Francisco, London, 1979.

[82] A. Rosenfeld. “Digital topology”, American Math Monthly, vol. 86, 1979.

[83] K. V. Rudakov. “Universal and local constraints in the problem of correction of heuristic algorithms,” Cybernetics. March–April, 1987, Volume 23, Issue 2, pp 181-186.

[84] K. V. Rudakov. “Completeness and universal constraints in the correction problem for heuristic classification algorithms,” Cybernetics. May–June, 1987, Volume 23, Issue 3, pp 414-418.

[85] K. V. Rudakov. “Symmetric and functional constraints in the correction problem of heuristic classification algorithms,” Cybernetics. July–August, 1987, Volume 23, Issue 4, pp 528-533.

[86] K. V. Rudakov. “Application of universal constraints in the analysis of classification algorithms,” Cybernetics. January–February, 1988, Volume 24, Issue 1, pp 1-6.

[87] K. V. Rudakov and K. V. Vorontsov, “Methods of Optimization and Monotone Correction in the Algebraic Approach to the Recognition Problem,” Dokl. Akad. Nauk 367, 314–317 (1999) [Dokl. Math. 60, 139–142 (1999)].

[88] K. V. Rudakov and Yu. V. Chekhovich. “Completeness Criteria for Classification Problems with Set-Theoretic Constraints,” Computational Mathematics and Mathematical Physics. Vol. 45, No. 2, February 2005, pp. 329-337.

[89] K. V. Rudakov, A. A. Cherepnin, Yu. V. Chekhovich. “On metric properties of spaces in classification problems,” Doklady Mathematics. October 2007, Volume 76, Issue 2, pp 790-793.

[90] K. V. Rudakov, I. Yu. Torshin. “Selection of informative feature values on the basis of solvability criteria in the problem of protein secondary structure recognition,” Doklady Mathematics. December 2011, Volume 84, Issue 3, pp 871-874.

[91] J. Serra. Image Analysis and Mathematical Morphology, London, Academic Press, 1982.

[92] I.N. Sinicyn. Calman and Pugachev Filters, Logos, Moscow (sec.ed.), 2007 [in Russian].

[93] A. Shaw. “A Proposed Language for the Formal Description of Pictures”, CGS Memo. 28, Stanford University, 1967.

[94] A. Shaw. “The Formal Description and Parsing of Pictures”, Ph.D. Thesis, Computer Sciences Department, Stanford University, December 1967 (also Tech. Rept CS94, April 1968).

[95] M. Schlesinger, V. Hlavac. “Ten Lectures on Statistical and Structural Pattern Recognition,” Computational Imaging and Vision, Vol. 24. Kluwer Academic Publishers - Dordrecht / Boston / London, 2002, 520 p.

[96] P. Soille. “Morphological partitioning of multispectral images”, Journal of Electronic Imaging, July 1996,-vol.5, no.3,-pp. 252-265.

[97] P. Soille. Morphological Image Analysis. Principles and Applications (Second Edition). -Springer-Verlag Berlin Heidelberg, New York, 2003 и 2004.

[98] S.R. Sternberg. “Language and Architecture for Parallel Image Processing,” Proceedings of the Conference on Pattern Recognition in Practice, Amsterdam, 1980.

[99] S. R. Sternberg. An overview of Image Algebra and Related Architectures, Integrated Technology for parallel Image Processing (S. Levialdi, ed.), London: Academic Press, 1985.

[100] S.R. Sternberg. “Grayscale morphology,” Computer Vision, Graphics and Image Processing, vol.35, no.3, 1986.-pp. 333-355.

[101] P. Sussner. “Observations on morphological associative memories and the kernel method,” Neurocomputing, vol. 31, 2000.- pp. 167–183.

[102] P.Sussner, G.X.Ritter. “Rank-Based Decompositions of Morphological Templates“, IEEE Transactions on image processing, vol. 09,no.8, august 2000,-pp. 1420-1430.

[103] P. Sussner. “Generalizing Operations of Binary Autoassociative Morphological Memories Using Fuzzy Set Theory,” Journal of Mathematical Imaging and Vision, vol. 19, Kluwer Academic Publishers. Manifactured in The Netherlands, 2003. -pp. 81-93.

[104] D.M.J. Tax, R.P.W. Duin. “Combining One-Class Classifiers”. Multiple Classifier Systems. Second International Workshop, MCS 2001, Cambridge, UK, July 2001. Proceedings. Springer - Verlag, 2001.

[105] A.Toet. “A morphological pyramidal image decomposition,” Pattern Recognition Letters, vol. 9, 1989. - pp.255-261.

[106] S.H. Unger. “A computer oriented toward spatial problems”, Proceedings of the IRE, vol.46, 1958, - pp. 1744-1750.

[107] B.L. Van Der Waerden. Algebra I, Algebra II, Springer-Verlag, Berlin Heidelberg New York, 1971.

[108] D. Winbridge, J. Kittler. “Classifier Combination as a Tomographic Process,” Multiple Classifier Systems. Second International Workshop, MCS 2001, Cambridge, UK, July 2001. Proceedings. Springer - Verlag, 2001.- pp. 248 - 258.

[109] Yu.I. Zhuravlev. “An Algebraic Approach to Recognition and Classification Problems”, Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications.- MAIK "Nauka/Interperiodica", vol.8. 1998.-pp.59-100.

# *The Development and Research the Digital Image Processing Algorithm on Television Picture for Indoor Positioning*

Alexander Tyukin, Ilya Lebedev
Dynamic of Electronic Systems
P.G. Demidov Yaroslavl State University
Yaroslavl, Russian Federation
tyukin.alexx@gmail.com

***Abstracts – This paper describes a procedure of the indoor positioning system for mobile platform, which uses the digital image processing algorithm on input television image. Described system includes navigation algorithm for mobile platform and its obstacle determination algorithm. The system uses one single video camera and navigation algorithm recognizes in video stream a color beacons placed on the environmental objects for position determination. Obstacle detection algorithm uses vision-based logic. The system has been tested in a variety of environments.***

***Keywords—digital image processing, indoor positioning, obstacle detection, computer vision, color beacons.***

## I. INTRODUCTION

Nowadays the global navigation tasks are successfully solved but some problems may arise with mobile robot indoor orientation. It occurs due to the fact that working indoors is characterized by multiple barriers - from illumination unevenness to problems of radio signals reflection [1].

The modern scientific and technical literature analysis shows that in many cases the most reliable communication channel is the optical one [2]. In such navigation systems beacons with color code are used. They represent a passive device with three fields of different color. Unlike ultrasonic, infrared and laser ones, beacons with color code are simple to be made and do not require any power-supply sources what lets them stay operable during indefinite time. A usual color digital video camera can serve as a beacon optical signal receiver.

Obstacle detection is an important task for modern mobile robots. Nowadays there are many robots relying on transducers and sensors to measure distances to obstacles. However, none of these sensors is ideal [3, 4]. For this reason there was developed and researched the computer vision algorithm that distinguishes the surface color attributes using one single camera.

## II. NAVIGATION ALGORITHM

### *A. The essence of the navigation algorithm*

The color beacons represent a combination of three color areas allocated in-line in close proximity to each other (Fig. 1).

The main steps of the navigation algorithm are to filter input image in video stream, to recognize three-color beacons; to locate the beacons in space relative to the video camera and to find the robot coordinates knowing the coordinates of beacons are set a priori.

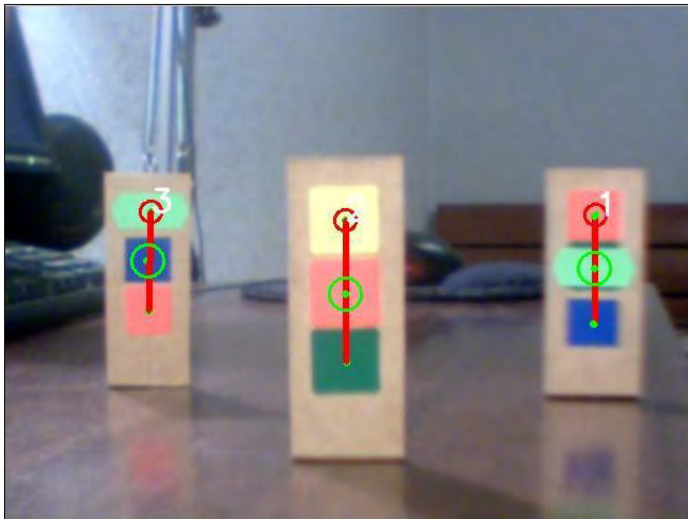


Fig. 1. Example of beacons.

We should detect adjusted number of three-color beacons in the video stream. It is also necessary to locate the beacons in space (relative to the matrix camera.)

In this case, the beacons have some limitations: all three colors at the beacon should be visually identifiable; color areas centers should be located on the same line and must be equidistant from each other, the surface of beacons should be matt.

### *B. The stages of the navigation algorithm*

The following stages can illustrate the work of the color recognition algorithm (Fig. 2):

1. At first, each stream video frame is filtered to smooth image defects and eliminate distortions.
2. The image is converted from the RGB (red, green, blue) color model into the HSV (Hue, Saturation, Brightness) model.

3. For recognition colors on input image for each HSV channel smooth analog function is used [5].

4. Then these three images are multiplied pixel by pixel with the cube root extraction. This action produces a "color mask" for one particular color.

5. In the "color mask" the pixel with maximum intensity value is detected and some area of pixels around is "filled".

6. The coordinates of centers of each filled area are calculated and written into the array.

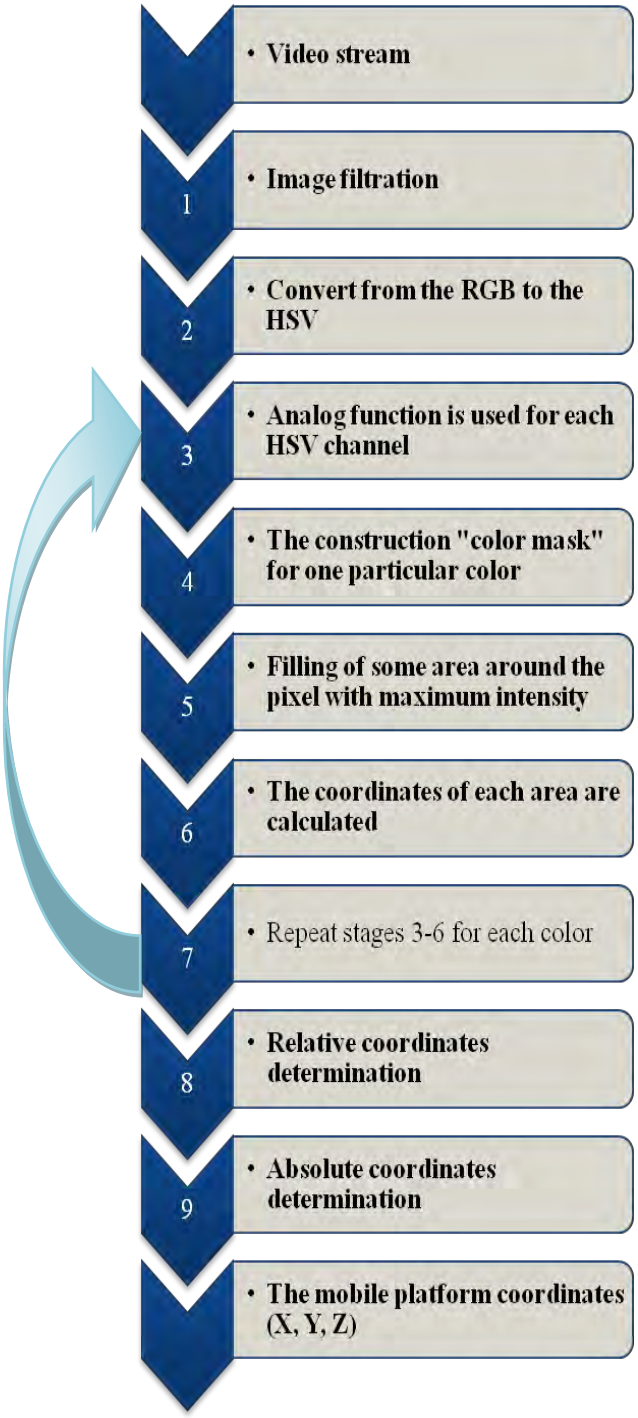



Fig. 2. Stages of the indoor navigation algorithm.

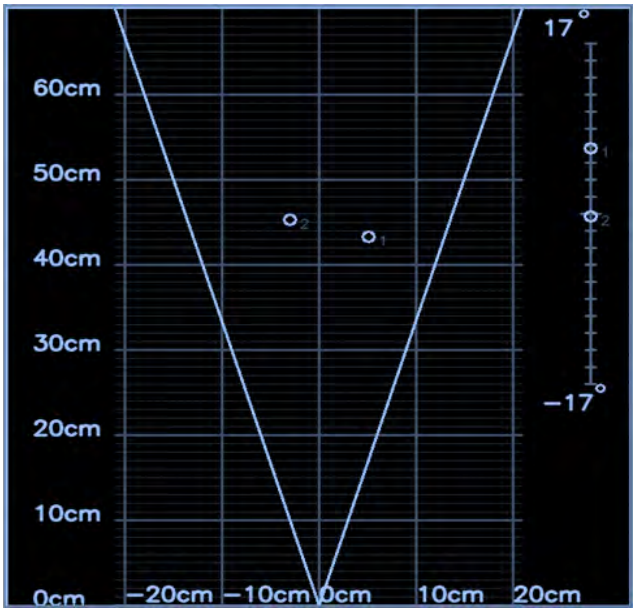



Fig. 3. Beacon location on the relational map.

7. Stages 3) – 6) are repeated for every recognized color.

8. After recognition of each beacon the relational map of the location of beacons relatively mobile platform is constructed. This map is constructed from the relative size of beacons on the image. For the calculation of the beacon three-dimensional coordinates several values should be set a priori: the beacon height and the video camera lens aperture [6]. The height equality of all beacons that are simultaneously present in the frame is an essential condition. The relational map is a top view and vertical level in relation to the lens visual axis (Fig. 3).

9. The mobile platform coordinates are calculated and its location is indicated in the absolute map. It is achieved due to the three-dimensional affine transformations usage. The absolute map is an image of a room or space plan where the robot moves. Due to three coordinates set a priori for each beacon, the video camera coordinates are defined [7].

### *C. The research of the navigation algorithm*

Analysis of the system in the real conditions was conducted and the performance borders of the algorithm were defined. The color recognition procedure forms the basis of the algorithm work. The color of the area the video camera lens points to, does not depend on the reflecting surface physical aspects only, but also on the impinging light spectral structure. That is why at first the algorithm work quality with reference to the external illumination type was researched.

We can see on the histogram (Fig. 4) that the worst results were obtained using a mercury quartz lamp, fluorescent lamp or direct sunlight. This is due to the limitations of the lamp light spectral structure and glare of intensive sunlight. The best results were received incandescent and natural lighting. These lights give steady luminance and wide range of emitted light, which gives high values of the saturation of colors. But for all the light sources dispersion values are not high, and the algorithm preserves its working capacity.

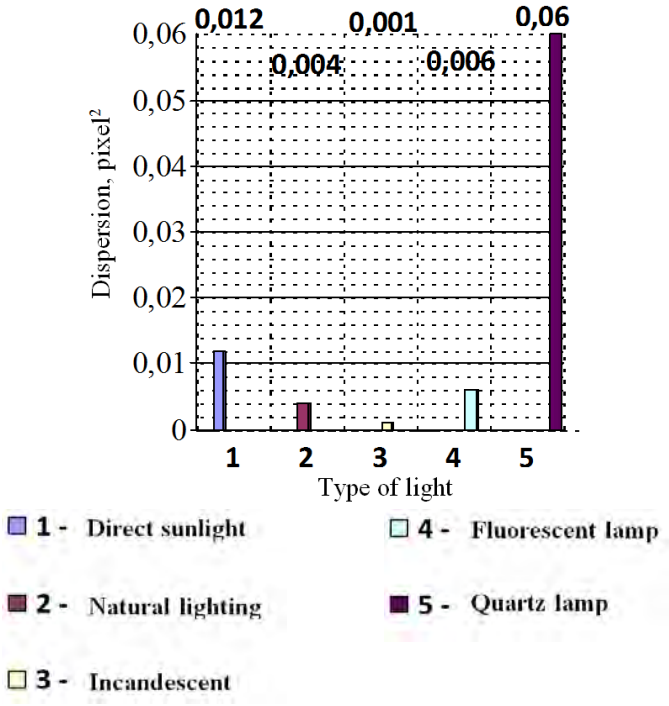



Fig. 4. Depending on the type of ambient light.

Also, since experimental platform is mobile and autonomous, we have investigated depending of beacons recognition on the distance "camera – beacon" (Fig. 5), and also depending on the angle of rotation of the beacon relative to the camera (Fig. 6).

The distance between the camera and the beacon is very important, because when the distance will increase, then the relative area of the beacon will decrease, which it takes in the image. The graph shows that at a distance of more than 1.5m dispersion quickly increases and at distances of more than 1.7m algorithm is not able to recognize the beacon. Consequently, the height of the beacon should be selected on the basis of characteristic distances used in the task.

In the graph of dispersion of the rotation angle we can see, that the best situation is when beacon is perpendicular to the optical axis of the camera lens. In this case the area of the colored areas is maximized. If the value is greater than 40° angle dispersion quickly increases and at angles greater than 60°, the algorithm is not able to detect a beacon. In this experiment beacons used with the color scheme printed on the plane. As an improvement to enhance the reliability of detection can be used cylindrical beacons or beacons with a convex surface.

The output parameters are an absolute coordinates of video camera. These data define platform location in the space. Therefore, determination of the localization accuracy measures is the main part of these researches.

Deviation and dispersion of the absolute camera coordinates was taken like the main parameter of algorithm work stability. The parameter of accuracy was taken deviation of the results from the true values. The true coordinates of video camera was measured with accuracy to the nearest 1 mm.

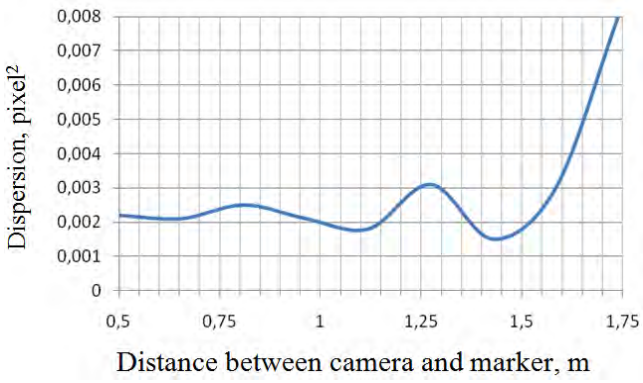



Fig. 5. Depending of dispersion on the distance between camera and beacon.

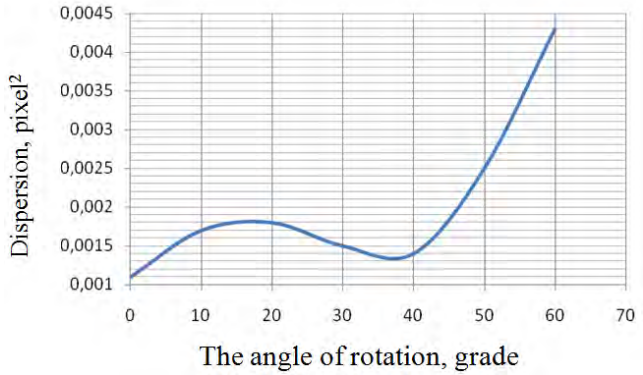



Fig. 6. Depending on the angle of rotation of the beacon relative to the camera.

Five different experiments are conducted with different location of two beacons. In each experiment video stream with beacons and the duration equal 160 sec was passed through the algorithm. The image was static (beacons and video camera stay put), and scene luminance was constant as well.

This video stream was passed through the algorithm and the absolute coordinates was calculated due to two different methods usage: the three-dimensional affine transformations and gradient descent. The results of one experiment are shown on Fig. 7 and Fig. 8.

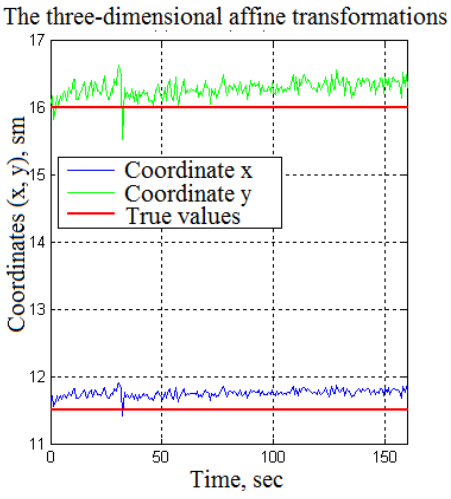



Fig.7 Determination of the absolute coordinate camera using two beacons.

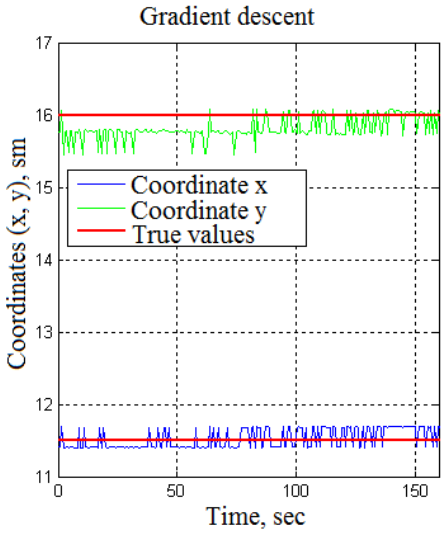



Fig.8 Determination of the absolute coordinate camera using two beacons.

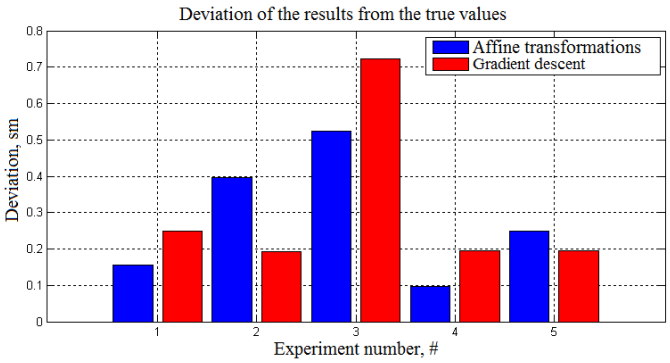



Fig.9 Deviation of the results from the true values using two beacons.

From results of five experiments the histogram of absolute coordinates deviation was being built using two methods: method of affine transformations and gradient descent (Fig. 9).

The histogram shows that in different experiments the deviations of absolute coordinates are different, but all values are within an acceptable range. If compare these determination coordinates methods, then cannot uniquely identify the best method.

This research of the localization accuracy for detection of two beacons showed that deviation of the calculated coordinates from the true values is 4 mm.

This algorithm allows the robot to navigate in space, but to make it move more confidently the obstacle detection ability is needed. For this there was developed a new system that discerns surface color attributes using one single camera in real-time.

## III. OBSTACLE DETECTION

### A. *The discription of the obstacle detection algorithm*

The core of the obstacle detection algorithm is discerning pixels that differ by color from the bottoming surface and classifying them as obstacles. The algorithm works on a real-time basis in different conditions providing images of high definition at the output. The method is based on three suppositions that are reasonable for different internal and external conditions:

1. Obstacles differ from the ground by their outside appearance.
2. The ground is relatively flat.
3. There are no overhanging obstacles.

For many applications it's important to estimate the distance from the camera to the pixel that represents an obstacle. With monocular vision the general approach to the distance estimation consists of the supposition of the ground being relatively flat and no overhanging obstacles present. If these two suppositions are right the distance is a monotonically increasing function of the pixel height in the image.

To start work we need to receive an incoming image: it can be video sequence or just a static image.

The next step is image filtration. Filter is a usual scheme of decimation and interpolation (lower pass filter implementation, odd information removal, signal enhancing by blank readings and interpolation using lower pass filter). The usage of this filter is conditioned by the fact that after its implementation the areas similar in color will be supremely homogeneous and it is essential for the clustering tasks. [15]

The filtered image can be transferred to the HSV color system. The way this transition influences the algorithm work will be shown later.

For the analysis of the bottoming surface with no obstacles in the image the trapezoid is drawn. This procedure is shown in the Fig. 9.

The trapezoid area is divided into 3 clusters using the K-means algorithm. On the basis of each cluster we create a taught model (on its basis the system will be taught for the following environment exploration). This model will include:

- a number of trapezoid pixels that got into the given cluster;
- the percentage of the pixels of the cluster in relation to the total number of pixels;
- cluster covariance matrix;
- the average value of the pixels of the cluster by corresponding color components.

For all image pixels we calculate the distance of the Mahalanobis d to the closest researched model (researched model is a color model of the bottoming surface that has been accumulated during several frames). For each cluster of the trapezoid we calculate the average distance of the Mahalanobis d' for all dots belonging to the cluster. Then the pixels not forming obstacles, are marked. If the following condition is fulfilled the pixel is considered bottoming surface, otherwise – an obstacle:

$$d - d' < \tau \qquad (1)$$

Then the researched models are renewed using training models.

*B. The analysis of two obstacle detection methods*

The survey concerned two modifications of the algorithm of the obstacle detection at the bottoming surface: in RBG and HSV color systems. The dispersion of the number of the dots discovered as bottoming surface was used as the work criteria. The experimental algorithm research has shown the less this value is the more reliable obstacle detection will be.

The algorithms were analysed by the following parameters:

1. Dependence on the Mahalanobis distance.
2. Dependence on the input image contrast ratio.
3. Dependence on the input image distortions.

The dependence of the results on the Mahalanobis distance is represented in the Fig. 10. As it's shown the function has absolute minimum and local minimums. Subsequently, to define the Mahalanobis distance that gives the best discernment automatically it's necessary to plot the whole chart and it takes a lot of time (to define the dispersion not less than 15 frames are needed that equals to 1 sec at the layout speed FPS = 15, so plotting a chart will take 10 sec.). It should be noted that the algorithm of image processing in RGB is of a less dispersion at the minimum point.

We have evaluated the distortion influence on the detection correctness. The results dispersion dependence on the peak signal to noise ratio is shown in the table 1.

The experiment data show that at moderate distortion the algorithms work at the quite same level. But at the higher distortion we see a big difference in the dispersions HSV and RGB (the RGB dispersion is a sequence higher).

## IV. CONCLUSION

In the present work the algorithm of the autonomous mobile platform navigation has been developed. Beacons with color coding were chosen as "benchmarks" for the navigation system. The computer vision system discerns the beacons mentioned above basing only on the color component of the incoming video sequence. The preferable light sources are the following:

- filament lamp;
- daylight, without direct sunlight.

While studying the camera – beacon distance influence it was found out that the algorithm can detect the beacon only at the distance not bigger than 1,7 m. but the experiment was carried out using the beacons 32 mm high and the distance mentioned above is only a relative value. To enlarge the system operation range it's necessary to enlarge the beacon size. Subsequently the beacon height should be chosen with reference to the assigned task scales. Also as an improvement to enhance the reliability of detection can be used cylindrical beacons or beacons with a convex surface.

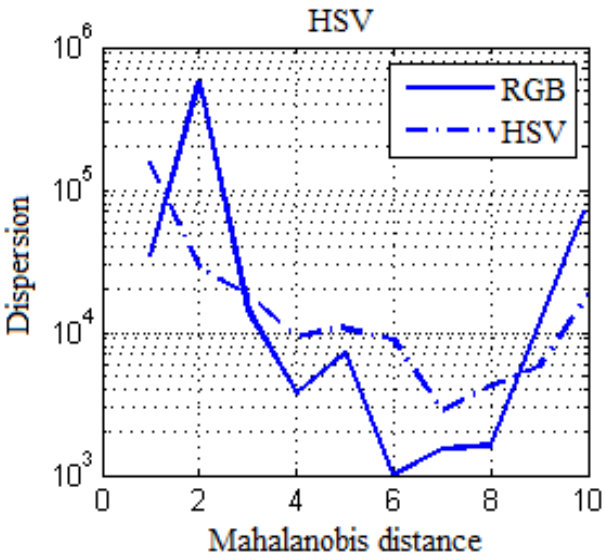



Fig 10. Results dispersion dependence on Mahalanobis distance.

TABLE 1. RESULTS DISPERSION DEPENDENCE ON IMAGE NOISE LEVEL

| № | PSNR(dB) | D (RGB) | D (HSV) |
|---|---|---|---|
| **1** | ∞ | $1,5 * 10^3$ | $2,6 * 10^3$ |
| **2** | 47 | $5,6 * 10^3$ | $7,3 * 10^3$ |
| **3** | 38 | $2,6 * 10^4$ | $1,8 * 10^4$ |
| **4** | 30 | $2,7 * 10^5$ | $1,2 * 10^5$ |
| **5** | 25 | $2,4 * 10^6$ | $1,5 * 10^5$ |
| **6** | 20 | $6,2 * 10^6$ | $2,1 * 10^5$ |
| **7** | 14 | $3,4 * 10^7$ | $1,5 * 10^6$ |

The research of the localization accuracy for detection two beacons showed that the average deviation of the calculated coordinates from the true values is 4 mm.

While developing and studying the obstacle detection algorithm the following steps were taken:

1. Implementation of the algorithms of obstacle detection and avoidance by mobile platform on the basis of bottoming surface discernment using RGB and HSV images.
2. Creation of the specific virtual environment for the computer vision algorithm analysis without hardware tools.
3. Analysis of the outcoming data of the implemented computer vision algorithms, comparative analysis of the implemented algorithms.

During the obstacle detection algorithm analysis it was shown that there exists such a Mahalanobis distance at which the system works optimally. The research of the noise influence on the algorithm has shown that at PSNR > 38 dB the algorithm using the HSV color scheme is preferable and at PSNR < 38 dB it's preferable to use RGB.

## REFERENCES

1. A.V. Ivanov “Navigation systems of mobile surface facilities. Algorithms of information processing in the angle channel”, Radiotekhnika, 2013, №4 (In Russian).

А.В. Иванов Навигационные системы подвижных наземных объектов. Алгоритмы обработки информации в угловом канале // Радиотехника, 2013, №4

2. B. A. Alpatov, A.A. Selyaev and A.I. Stepashkin “Digital image processing in the task of moving object tracking”, Izvestiya vuzov. Priborostroenie, 1985, №2, p.39-43. (In Russian)

Алпатов Б. А., Селяев А. А. Степашкин А.И. Цифровая обработка изображений в задаче отслеживания движущегося объекта. - Изв. вузов. Сер. Приборостроение, 1985, № 2, с. 39-43.

3. N.J. Nilsson Shakey the Robot. TechnicalNote: 1984.

4. H.R. Everett Sensors for Mobile Robots: Theory and Applications. Massachusetts. 1995.

5. O. Faugeras, L. Robert, S. Laveau, G. Csurka, C. Zeller, C. Gauclin and Zoghlami. 3-D reconstruction of urban scenes from image sequences. Comput. Vision and Image Understanding. 1998. p. 292-309.

6. B.A. Alpatov, “Evaluation of moving object parameters in the sequence of changing two-dimensional images”, Avtometria, 1991, №3, p. 21-24. (In Russian)

Алпатов Б.А. Оценивание параметров движущегося объекта в последовательности изменяющихся двумерных изображений. - Автометрия, 1991, №3, с. 21-24.

7. Lebedev I.M., Priorov A.L., Tyukin A.L. Analysis of Algorithms navigation and unimpeded movement of autonomous mobile robots in a confined space. – The 16th International Conference reports DSPA-2014, 2014, Vol. 1, p. 614-618

И.М. Лебедев, Приоров А.Л., Тюкин А.Л., Анализ алгоритмов навигации и беспрепятственного передвижения автономных мобильных роботов в ограниченном пространстве, DSPA-2014 Доклады, Серия: Цифровая обработка сигналов и её применение (выпуск XVI-2), М. – 2014, С. 614-618.

8. Babayan P.V., Alpatov V.A. Image processing in on-board detecting and tracking system. – Digital signal processing. 2006, №2, p. 45-51 (inRussian)

Бабаян П.В., Алпатов Б.А. Методы обработки и анализа изображений в бортовых системах обнаружения и сопровождения объектов // Цифровая обработка сигналов. 2006. № 2. С. 45 51.

9. Fisher R.B. From surface to object. Computer vision and 3D scene analysis.– Moscow.: Radio and communications, 1993. (inRussian)

Фишер Р.Б. От поверхностей к объектам. Машинное зрение и анализ трехмерных сцен.¬– М.: Радио и связь, 1993.

10. Gonzalez R., Woods R. Digital image processing. – Prentice Hall., 2002.

11. Gridin V.N, Titiov V.S., Trufanov M.I. Adaptive system of computer vision SPb.: Science, 2009. (inRussian)

Гридин В.Н, Титов В.С., Адаптивные системы технического зрения СПб.: Наука, 2009.

12. Gruzman I.S., Krivchuk V.C. and other. Digital image processing in information systems: Tutorial. - Novosibirsk, Press NGSU, 2002. (inRussian)

13. Shapiro L., Stockman G. Computer Vision. – Prentice Hall., 2001

Грузман И.С., Киричук В.С. и др. Цифровая обработка изображений в информационных системах: Учебн. Пособие. / Новосибирск, Изд-во НГТУ, 2002.

# The method for effective clustering the dendrite crystallogram images

R.A. Paringer
Samara State Aerospace University (SSAU)
Samara, Russia
E-mail: RusParinger@gmail.com

A.V. Kupriyanov
Image Processing Systems Institute of the RAS
Samara, Russia

***Abstract*— This paper presents a clustering method for the multiscale image of dendritic crystallogram. K-means clustering algorithm was used. As features are three geometric characteris-tic and four feature which as shape factors obtained spatial spectrum. For the new features formation the discriminant analysis algorithm was used. Of these, one new feature was formed. The error clustering decreased from 0.37 to 0.14.**

***Keywords—Dendrite crystallogram, geometrical features, shape factors, discriminant analysis, k-means clustering***

## I. INTRODUCTION

Image analysis is an important part of crystallograms task in medical diagnostics. Automating the process of crystallograms processing will improve the quality of diagnosis and reduce the time required. Crystallogram correspond to a certain type of characteristic values of diagnostic features. It is necessary to select the features that allow us to determine the type of a particular image crystallogram more accurately. There are many different methods for analysis the dendrite crystallogram images [1-3]. The estimating geometrical features and method for calculating form factors were selected. For clustering the k-means method is selected.

## II. GEOMETRICAL PARAMETERS

The dendrite model is presented on Fig.1. A, B, C, D are «top» key elements, E, F, G are «root» key elements, EF, FG are distances between branches. EF with FG is a stem of dendrite, AE, BE, CF, DG are dendrites' branches.

The method consists in the following sequence of actions: threshold and median filters [4], skeletonization, discrimination of key elements, building a «map of dendrites», calculation of geometrical features [5].

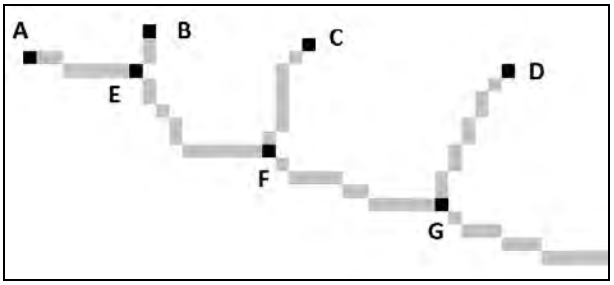



Fig. 1. Dendrite model.

«Growth» factor $F_g$ is the ratio of the lengths of all branches to the sum of the distances between them. For model calculated as follow:

$$F_g = \frac{AE+BE+CF+DG}{EF+FG} \quad (1)$$

«Angle» factor $F_a$ is the ratio of the sum of the angles of all branches the total number of them.

«Symmetry» factor $F_s$ is the ratio of the number of «top» key elements to the number of «root» key elements. For model $F_s$ =4/3.

## III. FORM FACTORS

The information contained in the crystallogram is structurally redundant. Visually, these images are perceived as a set of contour lines, subject to certain rather complex sequence. If we consider the spatial spectrum of such images, it would be located in a sufficiently narrow frequency band in a characteristic spatial frequency, which can be called carrier frequency. It is known that if in the original image collinear stripes of a certain direction prevail, then in Fourier transform of the original image stripes of the same direction will prevail.

Threshold the image spectrum can clearly distinguish the shape of the spectrum and calculate the form factors.

Consider the various geometric features of the spectral shape. This group includes those features, the calculation of which is based on the geometric characteristics appearing on the image objects. This features were called form factors. To calculate the form factors, it is necessary to calculate a perimeter and an area of the object. If we take an image pixel per unit area, then the area of the object is equal to the number of pixels. Calculation of the perimeter is not a trivial task and requires a specific approach.

The general algorithm can be represented by the following sequence of actions: calculation of the image spectrum, threshold processing, transformation of the domain into a closed domain, extraction of the object contour, calculation of the form factors [6].

S is the number of pixels in an image of a closed region of the spectrum. P is the number of pixels equal to the perimeter of the closed contour of the spectrum.

«Blair-Bliss» coefficient $F_b$:

$$F_b = \frac{S}{\sqrt{2\cdot\pi\cdot\sum_i r_i^2}}, \quad (2)$$

i – number of object's pixel;

$r_i$ – distance of the object's pixel from the object's center of gravity.

«Malinowska» coefficient $F_m$:

$$F_m = \frac{P}{2\sqrt{\pi\cdot S}} - 1, \quad (3)$$

«Haralick» coefficient $F_h$:

$$F_h = \sqrt{\frac{(\sum_i d_i)^2}{n\cdot\sum_i d_i^2 - 1}}, \quad (4)$$

$d_i$ – distance of the outline's pixels from the object's center of gravity;

i – number of outline's pixel;

n – number of objective's outline pixels.

«Compact» coefficient $F_c$:

$$F_c = \sqrt{\frac{1}{N}\sum_i \left(r_i - \frac{1}{N}\sum_i r_i\right)^2}. \quad (5)$$

N – total number of objective's pixels;

i – number of object's pixel;

$r_i$ – distance of the object's pixel from the object's center of gravity.

## IV. Discriminant analysis

In the analysis of diagnostic features by classification efficiency of greatest interest are the discriminant analysis methods , including methods of interpretation of intergroup differences , using which it is possible to obtain the following information:

- Assessment of the informative signs to split objects into classes.
- An assessment of separability of classes by using some feature set.
- The optimal parameters of the object separation into classes.

In discriminant analysis - separability criterion classes formed by using the scattering matrix within classes and between classes of scattering matrices.

Scattering matrix shows the scatter within classes of objects in the vectors expectations classes (6).

$$S_w = \sum_{l=1}^{L} P(w_l) E\{(X - M_l)(X - M_l)^T / w_l\}, \quad (6)$$

Scattering matrix between the classes can be determined by (7).

$$S_b = \sum_{l=1}^{L} P(w_l) E\{(M_l - M_0)(M_l - M_0)^T / w_l\}, \quad (7)$$

$P(w_l)$ - the probability of the l-th class;

$X$ - objects belonging to class, which is supposed to be independent $l = \overline{1, L}$.

Use separability criteria (8).

$$J = \mathrm{tr}\left(S_w^{-1} S_b\right). \quad (8)$$

Algorithm steps of forming features that maximize separability criterion.

1. Find the eigenvalues of the matrix $S_w^{-1} S_b$ with $n$ to $n$ size, $n$ - number of characters in the original vector features.

2. Sort results eigenvalues in descending order and select the first $m$ values $m \leq n$.

3. Find the eigenvectors corresponding to the selected eigenvalues.

4. Of the found eigenvectors form the transformation matrix $A$, which allows to generate new diagnostic features (9).

$$Y = AX, \quad (9)$$

$Y$ – new features;

$X$ – source features.

## V. Cluster analysis

For clustering k-means method was selected. Input data contain 468 images with 256 on 256 pixels resolution. Two types of images.

The first type includes dendritic crystallogram 100 times magnification as present on Figure 2. Total 234 images.

Fig. 2. Dendrite crystallogram 100x zoom.

The second type includes dendritic crystallogram 200 times magnification as present on Figure 3. Total 234 images.

The experiment was performed at fixed values of the parameters for all types of crystallograms. Estimation of the geometrical parameters was carried out using an adaptive threshold with a window size of 64 and a median filtering parameter equal to two. For estimation of spectral parameters is performed by a threshold selection probability weighting factor with a value equal to 20. The values were chosen experimentally. Error value is calculating with (10).

$$\varepsilon = \frac{k}{n} \cdot 100\%, \quad (10)$$

k – the number of clustering errors;

n – total number of images.

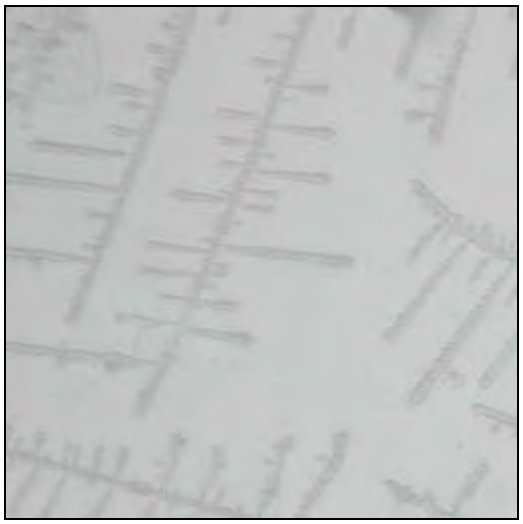


Fig. 3. Dendrite crystallogram 200x zoom.

Error value with seven base features without discriminant analysis is 0.37. Discriminant analysis performed on all the combination of all signs showed that the best separability criteria are achieved by using all seven features and the formation of them one new feature. In this case, the error value is 0.14.

## ACKNOWLEDGMENT

This work was finically support of the RFBR grant (#12-01-00237-a), the ONIT RAS program #6 “Bioinformatics, modern information technologies and mathematical methods in medicine” 2012, the Ministry of Education and Science of the Russian Federation.

## REFERENCES

[1] N.Yu. Ilyasova, A.V. Kupriyanov and A.G. Khramov, “Analysis of Features of Texture Images for Crystallogram Identification and Classification,” *Optical Memory & Neural Networks*, vol. 11, no. 11, pp. 19-28, 2002.

[2] A.V. Kupriyanov, N.Yu. Ilyasova and A.G. Khramov, " Ophthalmic Pathology Diagnostics Using Textural Features of the Lachrymal Fluid Crystal Images," *Pattern Recognition and Image Analysis*, vol. 15, no 4, pp.657-660, 2005.

[3] A.V. Kupriyanov, A.G. Khramov and N.Yu. Ilyasova, " Statistical Features of Image Texture for Crystallogram Classification," *Pattern Recognition and Image Analysis*, vol. 11, no 1, pp.180-183, 2001.

[4] V.A. Soifer, Computer image processing, part 2. Methods and algorithms, VDM Verlag Dr. Müller, 2010.

[5] R.A. Paringer and A.V. Kupriyanov, " Methods For Estimating Geometric Parameters of The Dendrite’s Crystallogramms," in Proc. *8th Open German-Russian Workshop “Pattern Recognition and Image Understanding” OGRW-8-11*, Russia, 2011, pp. 226-229.

[6] N.Yu. Ilyasova, A.V. Kupriyanov and A.G. Khramov, *Information technologies of image analysis for purposes of medical diagnosis*, Radio and communication, 2011.

Proceedings of the 9th Open German-Russian Workshop on Pattern Recognition and Image Understanding
Dec 1-5, 2014, Koblenz, Germany

# The study of features of expert signature for left ventricle in ultrasound images

A.O. Bobkova*, S.V. Porshnev*, V.V. Zyuzin*, V.V. Bobkov†
*Ural Federal University
Yekaterinburg, Russia
Email: zvvzuzin@gmail.com
†Medical information technologies
Verchnyaya Pishma, Russia
Email: btow@yandex.ru

***Abstract*—The article presents the study result of signature of left ventricle (LV) contours which are built by experts. The result is a part of a task of automatic contouring area of LV on ultrasound frames with apical four-chamber view. The signature is LV contour curve built in polar coordinates. An optimal (effective) center point of the LV base line was found during the studies. The resulting signature was approximated by three polynomials of second and third degree on left and right portions of the curve and at the top. The piecewise approximation results has been qualitatively and quantitatively better than the whole curve approximation.**

***Keywords*—echocardiography, left ventricle (LV), contouring.**

## I. Introduction

The ultrasound is a non-invasive method which does not produce ionizing radiation, relatively inexpensive and quite easy to use. For these reasons, it is widespread in different fields of medicine, including cardiology. In this area it is called echocardiography. One of the most important moments in the studies of global and local left ventricle contractility is an apical four-chamber view (Fig. 1). For calculating of the heart functions quantitative indicators doctors contour the LV at echocardiographic images (Fig. 1). It is necessary to mention that at present this procedure is performed either manually or semi-automatically because fully automatic solutions do not exist.

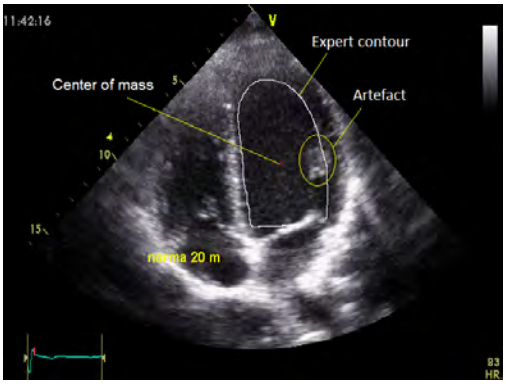



Fig. 1. Example echocardiography image with expert contour, artifact inside the area and the center of mass of the LV contour.

One of the difficulties which appear in the development of automatic LV contouring methods is the presence of artifacts which obstruct the correct contour construction. For example, there are white color areas inside the cavity of the LV on echocardiography images of some patients. While contouring the LV manually, the expert doctor ignores these areas and draws creates the contour as if they don't exist (Fig. 1). It is necessary to select the class of functions for the expert contour approximation and offer a way to identify their parameters for automating of the contouring LV process. In this regard, the research topic is relevant.

## II. The choise of the polar coordinate system origin for signature polynomial approximation

The first stage of the automatic algorithm is image processing which has been described in previous works [1,2,3]. The second stage is objects detection and LV contouring. Informative parts of the LV contour are the left wall or heart partition, the right wall and the upper part. The lower LV part or base, located in valve fixing point, is always an easy approximated segment. There is no problem in automatic mode [1]. Approximation of the remaining contour details is, in contrast, a non-trivial task. Two-dimensional binary frames of ultrasound movies with $640 \times 480$ pixels resolution were contoured by grip expert and then used in our research. For the remaining LV parts approximation we applied a polynomial functions set in the polar coordinate representation. The resulting approximating curve is called signature. An example of the signature of the expert contour is presented in Fig. 2.

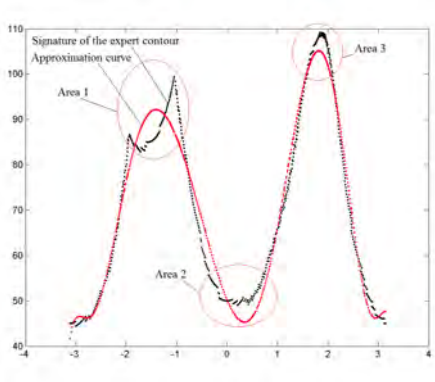



Fig. 2. $M$-shaped signature.

The polar coordinate origin was at the center of mass of the contour. Signature built from the center of mass of the contour is called $M$-shaped signature.)

Fig. 2 shows that the signature in the polar coordinate system can be approximated with a tenth-order polynomial. However, the approximation quality of the considered contour parts: area 1 (base), area 2 (middle of the right wall), area 3 (top) is unsatisfactory (poor). Besides, the shape of the signature, and consequently the order of the approximating functions, as it turned out, were strongly depended from the polar coordinate origin (Fig. 3).

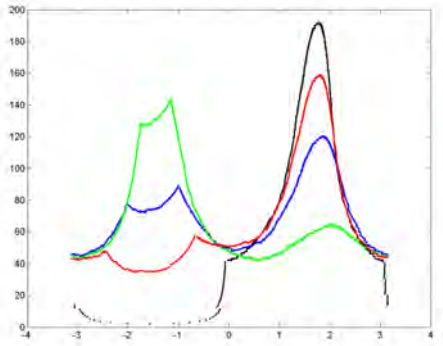


Fig. 3. Signatures built at four different positions of the coordinate system origin.

Fig. 3 shows that the signature started from the middle of LV base is more convenient in terms of approximation, because the curve shape is simplest, since it is easier to allocate the main signature details as base and the rest parts of LV. (Signature built from the middle of the LV base is called the Λ-shaped signature.) In addition, the procedure of finding fixed points of the left ventricle (left and right ends of the base segment) can be automated [4].

## III. The Choise of the Approximating Function Class and Approximation Technique

At the first stage we have studied the possibility of polynomial approximation of a Λ-shaped signature. Typical results of this approximation are presented in Fig. 4.

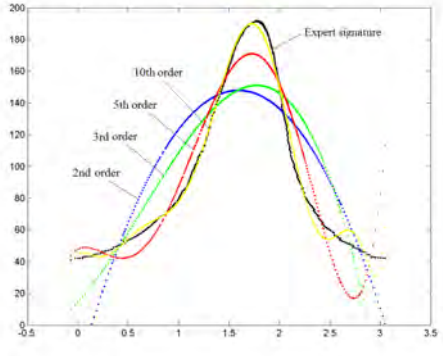



Fig. 4. The Λ-shaped signatures approximation of different order polynomials.

Fig. 4 shows that one polynomial for Λ-shaped signatures approximation cannot provide acceptable quality. In this regard, the piecewise signature approximation was studied. The whole LV contour is divided into several simple parts and then for each of them it need to choose approximating function class and to identify their parameters. We divided Λ-shaped signature into three parts (Fig. 5) and approximated them by 2nd and 3rd order polynomials.

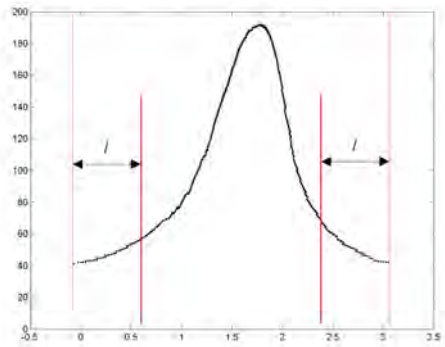



Fig. 5. Separation of the Λ-shaped signature into three parts.

Due to the fact that approximation quality of the selected method depends on borders positions, which divide Λ-shaped signature, total approximation error was calculated as:

$$Err(l) = \frac{1}{N}\sum_{i=1}^{N} Y_i^{sign} + Y_i^{apprx} \qquad (1)$$

here $Y_i^{sign}$ is value of the expert signatures in point with index $i$, $Y_i^{apprx}$ is value of the approximating polynomial in point with index $i$, $N$ is the total number of points in the signature. The calculation results for the polynomials of the 2nd and 3rd order at different positions of borders are presented in Figures 6 and 7.

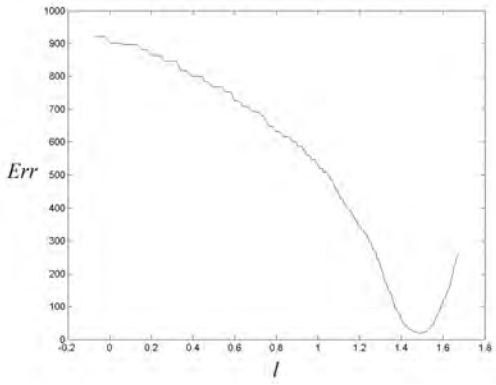



Fig. 6. Total error for 2nd order polynomial approximation versus borders position.

Fig. 6 and 7 show that the best approximation result was achieved at $l = 1.5$ for the 2nd order and at $l = 1.4$ for the 3rd order polynomials. The results of piecewise approximation of the expert contour are presented in Fig. 8.

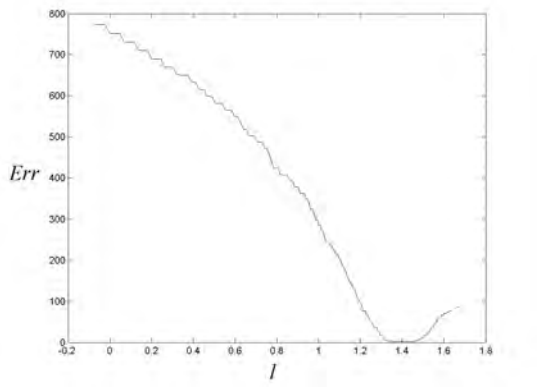



Fig. 7. Total error for 3rd order polynomial approximation versus borders position.

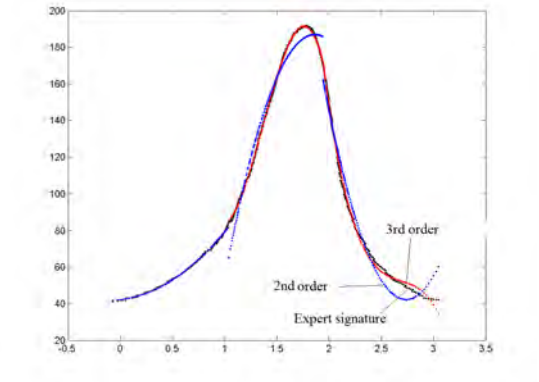



Fig. 8. The result of piecewise approximation with the 2nd and 3rd order polynomials.

Fig. 8 shows that the proposed method of piecewise approximation provides acceptable quality of approximation of the expert contour. Meanwhile, it should be noted, that at conjugation points of the approximated signature parts kinks can occur, but this can be resolved, for instance, using the moving average method. This result can be used to automatic contouring of ultrasound LV images. Fig. 9 and 10 present

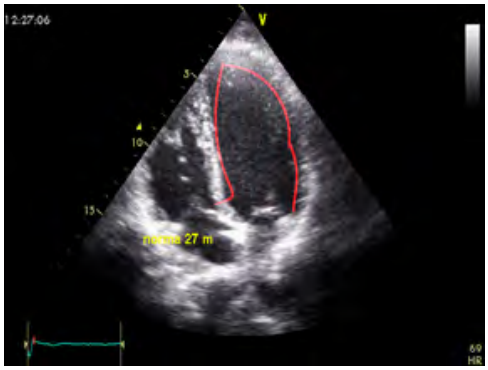


Fig. 9. Edges of LV contour based on 2nd order polynomial.

examples of expert contours approximation with 2nd and 3rd order polynomials. One can see the 3rd order polynomials approximation is more preferable because this signature is

Fig. 10. Edges of LV contour based on 3rd degree polynomial.

more accurate and smoother than 2nd order sample. Note the similar results were obtained for our data base of ultrasound images and expert contours, which contained the ultrasound records for 30 patients. Each record contains 25 frames. Total number of processed frames was about 750.

## IV. Conclusion

A comparison between signatures built from the center of LV mass and from the center-point of the LV base was analyzed. As result, we found some advantages to the LV signature creation from the central point of the LV base. The best approximation is achieved when $l = 1.5$ for the 2nd order polynomial and $l = 1.4$ for the 3rd order polynomial Description of automated algorithm of the left ventricle contouring and their application results for processing of ultrasound images of patients without pathologies and patients with pathologies are subjects for further publications.

## Acknowledgment

The work was carried out with financial support of Federal State-Financed Establishment The Fund of Advancement of Small Enterprises in Scientific and Technical Sphere within the framework of state contract 11475/20975.

## References

[1] A.O. Bobkova, S.V. Porshnev, V.V. Zyuzin, V.V. Bobkov. *Issledovanie metodov udalenija spekl-shumov na ul'trazvukovyh izobrazhenijah*, The 23rd International Conference on Computer Graphics and Vision, GraphiCon2013: Conference proceedings, Vladivostok, 2013, pp. 244-246. (in Russian)

[2] A.O. Bobkova, S.V. Porshnev, V.V. Zyuzin, V.V. Bobkov. *Analysis of methods for removing noise and artifacts on echocardiographic images*, The 11th International Conference PATTERN RECOGNITION and IMAGE ANALYSIS: NEW INFORMATION TECHNOLOGIES (PRIA-11-2013): Conference proceedings, Volume 2, Samara, 2013, pp. 525-528.

[3] A.O. Bobkova, S.V. Porshnev, V.V. Zyuzin, V.V. Bobkov. *Sposob poluavtomaticheskogo okonturivanija levogo zheludochka serdca cheloveka na jehograficheskih izobrazhenijah*, Sciemtific journal Fundamental research, 8, part 1, 2013, pp. 44-48. (in Russian)

[4] A.O. Bobkova, S.V. Porshnev, V.V. Zyuzin, V.V. Bobkov. *Retrieval of base points for left ventricle contouring*,Trudy XI Mezhdunarodnoj nauchno-tehnicheskoj konferencii Pod obshh. Red. Ju.E. Mitel'mana. Ekaterinburg: Izd-vo Ural'skogo universiteta, 2012. S. 361-363. (in Russian)

# Traffic Sign Detection and Recognition Using Modified Generalised Hough Transform

P. Yakimov
Samara State Aerospace University
Samara, Russia
pavel.y.yakimov@gmail.com

***Abstract*— Traffic Signs Recognition (TSR) systems can not only improve safety, compensating for possible human carelessness, but also helps to reduce tiredness, helping drivers keep an eye on the surrounding traffic conditions. This article proposes an efficient algorithm for TSR. The article considers the practicability of using HSV color space to extract the red color. An algorithm to remove noise to improve the accuracy and speed of detection was developed. A modified Generalized Hough transform is then used to detect triangular signs. Finally, the detected objects are being recognized. The developed algorithm has been tested on real scene images.**

***Keywords*— *Scene understanding; Image and video analysis and understanding; 2D/3D object detection and recognition***

## I. INTRODUCTION

Traffic Sign Recognition system is designed to provide the driver with relevant information about road conditions. There are several similar systems: 'Opel Eye' of Opel, 'Speed Limit Assist' from the company Mercedes-Benz, 'Traffic Sign Recognition', Ford and others. Most of them are aimed at the detection and recognition of road signs limiting the velocity of movement [1].

Traffic signs recognition is typically executed in two steps: sign detection and subsequent recognition. There are a lot of different methods of detection [2], [3], [4]. In fact, the recognition of a small size object does not cause any difficulties in the presence of the samples or patterns of possible traffic signs. However, such algorithms have significant computational complexity.

The performance of existing portable computers is not always enough for the real time detection of traffic signs. Most detection algorithms are based on Hough transform that allows you to effectively detect parameterized curves in an image, but this algorithm is very sensitive to the quality of digital images, especially in the presence of noise. The more noise in the image, the longer it will take to detect objects.

Thus, the possibility of detecting traffic signs in real time strongly depends on the quality of the preparation. In this paper, an effective algorithm for extracting high quality prepared images with a low noise level from the input image is proposed. This prepared image is then used for the detection and recognition of road signs.

This paper briefly describes the whole technology of traffic sign detection and recognition. The section with experimental results shows processed real scene images.

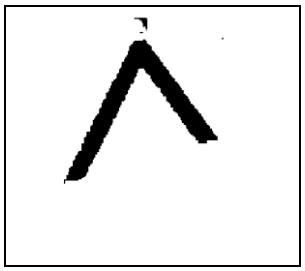
This work was supported by the Ministry of Education and Science of the Russian Federation

## II. TRAFFIC SIGN RECOGNITION

### A. Color analysis

Some specific light conditions significantly affect the ability of correct perception of the color in a scene. When taking the actual traffic situation, there are a number of different lighting conditions on the signs.

(a) Traffic sign with shadowing



(b) Extracting the red in the RGB color space

Fig. 1. Example of color extraction in RGB

The signs detection process becomes much more complicated due to such effects as direct sunlight, reflected light, shadows, the light of car headlights at night. Moreover, the various distorting effects may occur on one road sign at the same time (Fig. 1a).

Thus, it is not always possible to identify an area of interest in the real images by simply applying a color threshold filter directly in the RGB (Red, Green and Blue) color space. Fig. 1b shows an example of applying a threshold filter to the red color channel.

To extract the red color from the input image it is necessary to use the color information of each pixel, regardless of uncontrolled light conditions. For this purpose, the color space HSV (Hue, Saturation and Value) was selected.

Most digital sensors obtain input images in the format of RGB. Conversion to HSV color space is widely described in [5]. Between the three components of H, S and V there are certain dependencies. H component will not matter if the S or V components are represented by values that are close to zero.

The display color will be black if V is equal to 0. Pure white color is obtained when V = 1 and S = 0 [6], [7].

The 'ideal' red (R = 255, G = 0, B = 0) in the HSV color space is defined by the following values $H = 0.0°, S = 1, V = 1$. The experimental method was used to determine the optimal threshold values to extract the red color of road signs in the space of HSV:

$$(0.0° \le H < 23°) \vee (350° < H < 360°) \quad (1)$$

$$0.85 < S \le 1 \quad (2)$$

$$0.85 < V \le 1 \quad (3)$$

Fig. 2 shows the result of image processing of a road sign (Fig. 1a) with threshold values (1) - (3) in HSV.

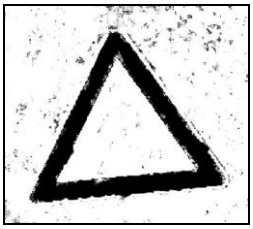


Fig. 2. Threshold color filtering in HSV.

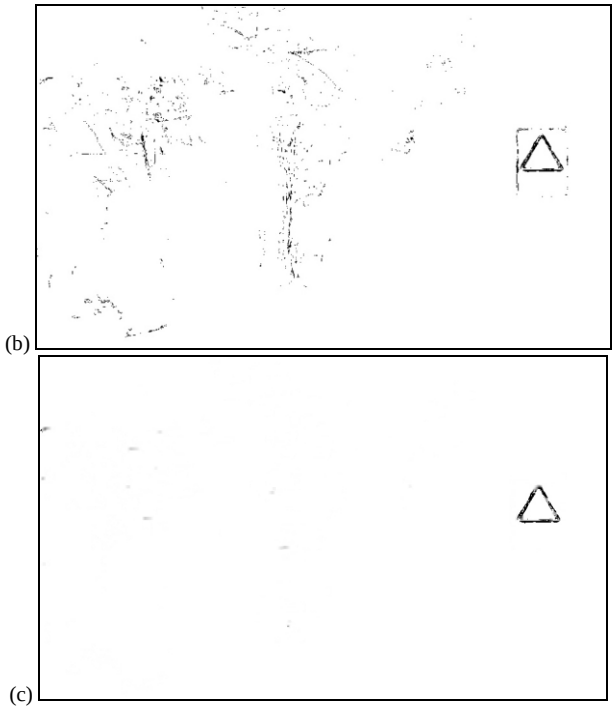
*B. Denoising*

The binary image represented in Fig. 2 satisfies the conditions of many algorithms of traffic signs detection. Such algorithms as the Haar wavelet [8], the Hough transform method [9], the regular expressions [2] and others will work efficiently on the image. However, one can easily notice the presence of noise in the image. The picture in Fig. 2 is well prepared for further processing, but the situation with the frames captured from a real video sequence is completely different.

The image in Fig. 1a was obtained by a camera with high resolution (8.9 megapixels), and shooting conditions were significantly better than when using a built-in car video sensor. Fig. 3a shows a fragment of a frame from the video sequence obtained during the experiments containing a road sign.

Noise in Fig. 3b appears after thresholding to extract the red color. It not only reduces the performance of the system, but also affects the quality of detection. This can lead to false detection of road signs.

(a)



(b)

(c)

Fig. 3. (a) A frame from video sequence; (b) Binary image with extracted red; (c) Result of image denoising.

In order to avoid this point-like noise a modified algorithm based on the results obtained in [10] - [12] was applied. These articles describe the algorithm of detection and retouching of point-like glares on the reproductions of works of art. These artifacts appear on digital photographs because of the irregularities in the painting's surface, i.e. the light is reflected differently in different parts of a picture and the artifacts appear on a relief surface. In order to detect these glares the sliding windows algorithm was used. The bigger window contains a nested smaller one and several special features for detecting artifacts were introduced. According to the algorithm, a pixel is considered as a part of a glare if the following conditions are satisfied:

$$\max_{(x,y)\in W_1} f(x,y) \ge f_1 \quad (4)$$

$$\frac{M_{(x,y)\in W_1} f(x,y)}{M_{(x,y)\in W_0} f(x,y)} \ge M_1 \quad (5)$$

$$\sum_{(x,y)\in W_1} I(x,y) \le S_1 \quad (6)$$

$$D(W_1) \le D_1 \quad (7)$$

Here $f(x,y)$ is a pixel's value in the point with coordinates $x$ and $y$, $W_1$ and $W_0$ are respectively the sizes of large and nested small windows, $M\, f(x,y)$ is the mean value of brightness within the windows, $I(x,y)$ is the number of 'bright' pixels in a certain area, $D(W_1)$ is the diameter of an area, i.e. the value that defines the degree of compactness of an area. $f_1, M_1, S_1, D_1$ are experimentally selected threshold values. Inequality (4) imposes a limit on the minimum value of the absolute brightness, and (5) limits the relative value of the mean brightnesses in the windows $W_1$ and $W_0$.

This algorithm gave effective results in the processing of digital reproductions. The condition (4) was removed, when configuring the algorithm to the problem of detection and elimination of noise in the images with extracted red color,

since the image is binary, and this feature is no longer informative in the new task. Despite the different origin of the noise in the images of road signs, the algorithm performed a high level of efficiency applied to the removal of artifacts (noise). The result of processing the image from Fig. 3b is shown in Fig. 3c.

Paper [12] shows the effective implementation of the denoising algorithm in the massively multi-threaded environment CUDA. The resulting acceleration on the GPU relative to the CPU reached 60-80 times. Frame size in the video sequence is 1920x1080 pixels. Image processing execution time on the CPU is 0.7-1 sec. Using CUDA on NVIDIA GeForce 335m has reduced the execution time to 7-10 ms, which satisfies the requirement of processing video in real time.

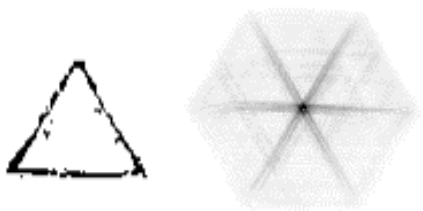
Thus, the modified algorithm for detecting and removing artifacts has significantly improved the quality of the images. The total time required to process the input image is 10-20 ms.

### *C. Detection*

Detecting traffic signs is implemented using a modification of Generalized Hough transform (GHT) [9]. Implementing classic GHT in Full HD 1080p images leads to enormous execution time. One of the main objectives of the TSR system is to operate in real time. Therefore, there are maximum 100 milliseconds for processing one frame on the detection step.

Most TSR systems are designed to detect only circular signs. There is no difficulty to detect circles using an implementation of Hough transform, and using CUDA makes it possible to implement it in real time. All processing takes no more than 40 ms including steps of color extraction, denoising, detection and recognition. Other systems use various machine-learning techniques such as Viola-Jones [13] or Support Vector Machine [14], which do not always suit the execution time limitation.

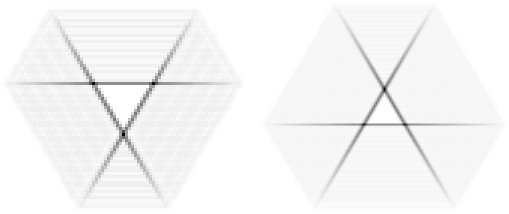
In this paper, we consider detection and recognition of triangular signs in real time. The main difference from the original GHT is in using some other accumulator space (Fig. 4b) and avoiding the R-table construction. After applying a special triangular template to the binary image in Fig.3a, the point with the maximum value in Fig. 4b is the central point of the sought-for object. The case shown in Fig. 4 is for equal scales of a template object and object in the real scene. The colors in the pictures are inverted in comparison to the images used in the algorithm.



(a) (b)

Fig. 4. (a) An example of a triangular sign after color extraction; (b) Accumulator space after implementing the developed algorithm

In case of different scales, we receive some more extremum points in the accumulator, three points when implementing the algorithm using a triangular template (Fig. 5).



(a) (b)

Fig. 5. The accumulator space in case of
(a) template is smaller than the objec in real scene;
(b) template is bigger than the object in real scene

Assuming that the size of a sign is up to 150 pixels, we have found that the distance between two of these points is up to 20 pixels. It allows computing the difference in scales of the template and an object in real scene. Using this value, we can precisely define the area of sign and then pass it to the recognition step. The middle location of these bright points in the accumulator is the coordinates of an object's center.

### *D. Recognition*

For the recognition step, the algorithm uses special prepared binary etalon images, which are actually inner areas of traffic signs. Fig. 6 shows such etalons.

Fig. 6. Etalon images for recognition

After detection, the algorithm obtains images of detected objects, which are quite similar to etalons since they are previously resized to the constant size of etalon images. To determine the type of a found object, we can use any recognition method. However, in case of successful object detection, and due to the execution time limitation, it is expedient to use a simple image subtraction and choose the lowest value, which will point to the most similar etalon. In case of large value, the algorithm gives a false detection message, since no similar etalon images were found.

The execution time of such recognition is 1-2 ms in average with 32 types of image etalons. This performance allows using several etalon images of each type to increase the recognition efficiency and reliability.

## III. EXPERIMENTAL RESULTS

The developed algorithm was tested on the video frames obtained on the streets of the city of Samara using a camera GoPro Hero 3 Black Edition built in to a car.

Fig. 7 shows the fragments of the original images with marked road signs on them.

Fig. 8 shows the result image of the detection algorithm without the prior application of the noise reduction algorithm presented in this paper. It shows false detection. In this noisy image, the accumulator space collected more votes for noisy part of the image than for road sign is in the shade.

Note that it took 80 ms to apply the detection algorithm on the noisy image. While it took almost half as much time time (41 ms) to process the denoised image.

Fig. 7. Frames with detected signs

Fig. 8. False detection.

In order to evaluate the detection algorithm accuracy, we used the German Traffic Sign Detection database [15]. It contains more than 50,000 images with traffic signs registered in various conditions. To assess the quality of the sign detection, we counted number of images with correctly recognized traffic signs. When testing the developed algorithms, we used only 9,987 images containing traffic signs of the required shape and with red contours. The experiments showed 97,3% of correctly detected and recognized prohibitory and warning signs.

## IV. Conclusion

This paper proposes an algorithm for image preprocessing, detecting, and recognition of traffic signs. The HSV color model was approved as the most suitable one for the extraction of red color in the images, and thanks to the experimentally chosen threshold values, it is possible to get the red color regardless of the lighting conditions. The modified algorithm for removing noise helped not only to avoid false detection of signs, but also accelerated the processing of images. The developed algorithm can improve the quality and increase the reliability of automotive traffic sign recognition systems, and reduce the time required to process one frame, which brings the possibility to carry out the detection and recognition of signs in Full HD 1920x1080 images from the video sequence in real time.

An algorithm for detection of triangular signs is considered in the paper. It is based on the Generalized Hough Transform and is optimized to suit the time limitation. The developed algorithm shows efficient results and works well with the preprocessed images. Recognition of detected signs makes sure that the whole procedure of TSR is successful.

In this paper, we consider triangular traffic signs. The developed detection algorithm makes it possible to detect signs of any shape. It is only needed to replace the template image with a sought-for shape.

The use of our TSR algorithms allows processing of video streams in real-time with high resolution, and therefore at greater distances and with better quality than the similar TSR systems have.

## References

[1] M. Shneier, “Road sign detection and recognition,” Proc. IEEE Computer Society Int. Conf. on Computer Vision and Pattern Recognition, 2005, pp. 215–222.

[2] A. Nikonorov, P. Yakimov, P. Maksimov, “Traffic sign detection on GPU using color shape regular expressions,” VISIGRAPP IMTA-4 2013, 2013, Paper Nr 8.

[3] A. Ruta, F. Porikli, Y. Li, S. Watanabe, H. Kage, K. Sumi, “A New Approach for In-Vehicle Camea Traffic Sign Detection and Recognition,” IAPR Conference on Machine vision Applications (MVA), Session 15: Machine Vision for Transportation, May 2009.

[4] R. Belaroussi, P. Foucher, J. P. Tarel, B. Soheilian, P. Charbonnier, N. Paparoditis, “Road Sign Detection in Images,” A Case Study, 20th International Conference on Pattern Recognition (ICPR), 2010, pp. 484-488.

[5] M. Tkalcic, J. Tasic, “Colour spaces - perceptual, historical and applicational background,” In The IEEE Region 8 EUROCON 2003 proceedings, 2003, pp. 304-308.

[6] A. Koschan, M. A. Abidi, “Digital Color Image Processing,” ISBN 978-0-470-14708-5, 2008, p. 376.

[7] D. Travis, “Effective Color Displays Theory and Practice,” Academic Press, ISBN 0-12-697690-2, 1991, p. 328.

[8] Sin-Yu Chen, Jun-Wei Hsieh, “Boosted Road Sign Detection And Recognition,” International Conference on Machine Learning and Cybernetics, 2008, vol.7, pp. 3823-3826.

[9] A. Ruta, Y. Li, X. Liu, “Detection, Tracking and Recognition of Traffic Signs from Video Input,” Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems, Beijing, China, 2008.

[10] P. Yakimov, S. Bibikov, R. Zakharov, A. Nikonorov, V. Fursov, “Detection and color correction of artifacts in digital images,” Optoelectronics, Instrumentation and Data Processing, vol. 47, issue 3, 2011, pp. 226-232.

[11] P.Y. Yakimov, S.A. Bibikov, A.V. Nikonorov, V.A. Fursov, “Investigation of the efficiency of CUDA technology in the problem of distributed prepress of digital images,” conference Science in the Internet: scalability, parallelism, efficiency, 2009, pp. 21-26.

[12] P.Y. Yakimov, V.A. Fursov, “Software for image processing using massively multithreaded CUDA environment,” conference "Conducting research in the field of information and telecommunication technologies", 2010, pp. 119-120.

[13] A. Møgelmose, M. Trivedi, M. Moeslund, “Learning to Detect Traffic Signs: ComparativeEvaluation of Synthetic and Real-World Datasets,” In 21st International Conference on Pattern Recognition, 2012, pp. 3452-3455, IEEE.

[14] S. Lafuente-Arroyo, S. Salcedo-Sanz, S. Maldonado-Basc´on, J. A. Portilla-Figueras,R. J. L´opez-Sastre, “A decision support system for the automatic management of keep-clear signs based on support vector machines and geographic information systems,” Expert Syst. Appl., vol. 37, pp. 767–773, January 2010.

[15] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, C. Igel, “Detection of Traffic Signs in Real-World Images: The {G}erman {T}raffic {S}ign {D}etection {B}enchmark,” International Joint Conference on Neural Networks, 2013

# Using Bit Representation for Generalized Precedents

Alexander Vinogradov
Dorodnicyn Computing Center
Russian Academy of Sciences
Moscow, Russian Federation

Yuriy Laptin
Glushkov Institute of Cybernetics
Ukrainian National Academy of Sciences
Kiev, Ukraine

***Abstract*—In this paper the role of intrinsic and introduced data structures in constructing efficient recognition algorithms is analyzed. We investigate the concept of generalized precedent and based on its use methods of reduction of the dimension of tasks. A new approach to the problem based on the use of positional data representation is proposed.**

***Keywords—generalized precedent; logical regularity; positional representation; bit slice; hyper-cube; correct decision rule***

## I. INTRODUCTION : INHERENT AND INJECTED STRUCTURES IN DATA

Structurization of data pursues two main objectives in solving problems of recognition:

a) identification in the feature space objective clusters of density in which vectors of precedents are grouped;

b) minimization of computation expenses necessary for creation of the decision rule on the basis of structures in training information, and for subsequent calculations with the use of this rule.

The choice of decision method in many respects is determined by asserted priorities for a) and b). Currently a huge number of approaches, algorithms and methods, that are more or less successful, is developed for achieving both purposes [7], [9].

In what further, the close relationship between concepts "precedent" and "cluster" is considered. We investigate mobility of the border between realizations of these concepts in computation environment, and discuss opportunities of usage this mobility in achieving compromise between a) and b).

Parametric approximation of empirical distribution in the form of uniform normal mix

$$\sum C_i = \sum \mu_i \, exp(-0.5(x_i-x)^T \boldsymbol{\sigma}(x_i-x)) \qquad (1)$$

with constant covariance matrix $\boldsymbol{\sigma}$ can serve here as an illustration. Each component $N(x_i,\boldsymbol{\sigma})$ is a compact spatial cluster $C_i$ with center $x_i$, and it is uniquely described by the pair $(x_i, \mu_i)$.

Natural interpretation of (1) is that each cluster $C_i$ is composed of vectors corresponding to random deviations from parameters of the central object $x_i$. Recognized object $x_0$ can also be considered as a single realization of the distribution of possible locations of the true center, which also form a cluster $C_0$ with center $x_0$ and with the same form of distribution $\mu_0 exp(-0.5(x-x_0)^T \boldsymbol{\sigma}(x-x_0))$, where coordinates of the center $x_0$ and the variable $x$ are interchanged according to the Bayes rule. Thus, the inherent structure of the sample has got simple representation, but this simplicity is achieved via time-consuming search of (1) as the solution of hard inverse problem, as well as difficulties in assigning cluster $C_0$ to one of classes, each of which is represented by several clusters of the form $N(x_i,\boldsymbol{\sigma})$. Of course, we show here an exceptional example, but it adequately reflects the close relationship between the two concepts.

Alternative example, in which useful structures in data are injected forcedly, can be found in IP approaches, wherein introducing strict quad-tree hierarchy of clusters in the plane $R^2$ provides high processing efficiency during training and recognition, but the hierarchy is always the same, and in orthodox approaches doesn't obey to the internal structure of the training sample [6]. Forms and places of clusters are fixed, and meaningful information is encoded only by the density of filling at different levels.

Further in this paper we consider a new approach to recognition problems, in which the balance between the accuracy of representation and computational efficiency can be achieved due to the special structural reduction within the pair "precedent-cluster".

## II. GENERALIZED PRECEDENTS

Model (1) involves the use of Euclidean norm $||x||=(\sum x_i^2)^{1/2}$ for evaluation and comparison of vectors in $R^N$. Normalizing of this kind binds together values of different features that can be incomparable in their numeric form. It is convenient at times, but causes doubts at meaningful interpretation of results.

On the contrary, for hierarchy of clusters in quad- or octree, valuations in different dimensions do not interact. In case of IP this is the main drawback of quad-tree-type models, which limits their use [4], [5].

In fact, in the case of images or scenes spatial directions are usually assumed to be equal in treatment. At the same time, models of this type are not invariant to rotations in $R^2$ and $R^3$, and therefore hard reproducible after rotations of the basis.

On the contrary, in abstract feature space the assumption of equality of axes arises seldom. Moreover, invariance of the model to any scaling of dimensions (in general, independent

This work was done under support of grants of RFBR, Russia, and NAS, Ukraine.

nonlinear changes of scale on axes) becomes important advantage.

One of the successful approaches based on use of this invariance is related with logical regularities [1]. In this approach clusters in the form of hyper parallelepipeds in $R^N$ are used, each of clusters is described by conjunction of the form

$$L = \&R_i\ ,\ R_i=(A_i<x_i<B_i), \qquad (2)$$

and is interpreted as a repeated regular joint manifestation of the values of features $x=(x_1,x_2,\ldots,x_N)$ on intervals $(A_i<x_i<B_i)$. The principle of proximity to each other of precedents of the same phenomenon is embodied here in the requirement of filling the cluster by objects of the same class.

Thus the form of cluster gains a special importance, and repeated joint emergence of values of features on certain intervals is considered in this approach as separate phenomenon that is called elementary logical regularity. The approach investigated in this work lies in the same course and deals with elementary logical regularities of a special kind.

In all mentioned approaches just limited number of parameters is used for description of spatial arrangement of the cluster and its filling. In case of quad-tree a cluster is coded by one integer and one real parameter, for normal mix (1) it is dimension of the couple $(x_i,\ \mu_i)$, in case of logical regularities – $2N$ border marks $A_i$, $B_i$ on feature axes, and also the weight of regularity $L$.

V. V. Ryazanov has proposed idea of reducing dimension of the task on the basis of use of essential specific clusters in data, such as hyper-parallelepipeds or components $N(x_i,\boldsymbol{\sigma})$ with considerable aprioristic weight, as new training objects. Each integrated object is considered as a single precedent of some regularity in data. Such *generalized precedents* just are proposed to be objects in new training sample. The generalized precedents are described by sets of parameters of the corresponding clusters.

So, dimensions of the new feature spaces in the examples given above are $1+1$, $N+1$ and $2N+1$. The dimension can change as towards increase, and decrease, but aprioristic information gains significantly more compact representation as a result.

We apply this idea for positional data representation, in which structural elements are a special case of logical regularities of Type I [2], [8].

Positional data representation is a development of quad-tree model in dimensions higher than 2. Main advantage of structuring data in the form of positional hierarchy consists in the fact that positional structure is automatically injected in numerical data while recording, and it is ready at once for usage. Also the independence of scaling of the main axes in models of this type was noted above.

These facts give prospects of use of the proposed approach for various tasks, including tasks with incomparable numerical features

## III. Bit layers and detailing the decision rule

Further it is supposed that final sets $X_k$ in $R^N$ represent classes of the training sample $X=\bigcup X_k$., $k=1,2,\ldots, K$.

Positional representation [3] of data in $R^N$ is defined by a bit grid $D^N \subset R^N$ where $|D| = 2^d$ for some integer $d$. We don't fix parameter $d$ in advance. As we will see, its value is determined by results of the analysis of relative positioning of classes.

Each grid point $x=(x_1, x_2, \ldots, x_N)$ corresponds to effectively performed transformation on bit slices in $D^N$, when the $m$-th bit in binary representation $x_n \in D$ of $n$-th coordinate of $x$ becomes $p(n)$-th bit of binary representation of the $m$-th digit of $2^N$-ary number that represents vector $x$ as whole. It's supposed that $0<m\le d$, and function $p(n)$ defines a permutation on $\{1,2, .., N\}$, $p \in S_N$. The result is a linearly ordered scale $S$ of length $2^{dN}$, representing one-to-one all the points of the grid in the form of a curve that fills the space $D^N$ densely.

For chosen grid $D^N$ an exact solution of the problem of recognition with $K$ classes results in $K$-valued function $f$, defined on the scale $S$. As known, $m$-th digit in $2^N$-ary positional representation corresponds to $n$-dimensional cube of volume $2^{N(m-1)}$. This cube is called $m$-point. For each $m$ the entire set of $m$-points is called $m$-slice. Thus we have

**Lemma 1.** There are just one $d$-point, $2^N$ $(d-1)$-points, and $2^{dN}$ $1$-points on the scale $S$. □

Each of $m$-points, $0<m<d$, can be regarded as separate cluster in $D^N$. If it's non-empty and filled with data of certain class only, we have got generalized precedent.

Further for each $k$, $k=1,2,\ldots, K$, we look for a set of $m$-points, $0<m<d$, representing the generalized precedents, i.e., elementary logical regularities of a class $k$. We will describe the scheme of algorithm **A** realizing such search on hierarchy of $m$-points of the grid $D^N$ from top to bottom.

1. The search is carried out for all classes $k$, $k=1,2,\ldots, K$, simultaneously. Thus data of the training sample $X=\bigcup X_k \subset R^N$ are transformed into $2^N$-ary indices of the grid $D^N$.

2. All objects of the sample are processed in turn. Each next object $x \in X_k$ marks with index $k$ all $m$-points, $m>1$, of the own branch in hierarchy $D^N$. Notice that for $m>1$ there are no more than $\sum_{m=2}^{d} 2^{N(d-m)}$ different $m$-points. For dimensions $N>3$ this number is negligible in comparison with the total number of $1$-points of the grid $D^N$.

3. Upon termination of search in each marked point of hierarchy $D^N$ the final attributing is carried out: if $(m+1)$-point was marked with indexes of various classes (i.e., isn't the generalized precedent), and all $m$-points subordinated to it are generalized precedents, then all of the last are included in the decision rule. Further specification and attributing of subordinated $(m-1)$-points aren't required.

As for all classes $k$, $k=1,2,\ldots, K$, the analysis began with the same $d$-point as the top of hierarchy, we have

**Lemma 2.** Algorithm **A** finds all generalized precedents of specified kind in the sample $X=\bigcup X_k$. □

Now we can explain why parameter $d$ wasn't fixed in advance. It is implicitly supposed above that transform of real-number representation of $X=\bigcup X_k$, $k=1,2,\ldots, K$, to its image on discrete grid $D^N \subset R^N$ doesn't lead to hashing of some classes.

Of course, this condition can be met, having found the minimum gap between objects of different classes in $X$, but necessary parameter $d$ can turn out too big, and hierarchy $D^N$ – inaccessible to processing. Otherwise the truncation errors occur, and objects of different classes can occupy the same lower node $x^*$ in hierarchy $D^N$. But $D^N$ is tree-like hierarchy, and algorithm **A** can be applied without changes to $x^*$ that is considered as new top.

Thereby an iterative process of creation of the decision rule, when parameter $d$ is chosen several times for reasons of computing efficiency, can be realized. Since the number of $m$-points is final, any $m$-point that hashes classes will be certainly resolved, and thus we have got

**Lemma 3.** Iterative process on the basis of algorithm **A** provides creation of exact decision rule that is correct on the training sample $X=\bigcup X_k$. □

At last, we note one else important feature of the approach presented above. Since data of the training sample $X=\bigcup X_k$ are analyzed by algorithm **A** consecutively, further retraining of any recognition algorithm constructed on this way, will demand investigation of objects no more than inside one generalized precedent for each new object.

## IV. Conclusion

In this paper we have considered the use of some inherent and injected structures in data, and after that interrelations between concepts ‘precedent’ and ‘cluster’ were discussed. We analyzed opportunities arising from the use of positional data representation for creation of detailed decision rule in the space of generalized precedents. It is shown that in this case the feature space $R^N$ can be reduced to two-dimensional space where training data become represented by compact clusters. Reduced representation realizes the one-dimensional scan of $R^N$, which is loaded with weights of generalized precedents. Arguments are presented justifying the advantage of this representation in case when recognized objects are described by qualitatively different characteristics with incomparable numerical values. A scheme for an iterative process is proposed that yields to construct exact solutions that are correct on the training data.

## References

[1] Yu.I.Zhuravlev, V.V.Ryazanov, O.V.Senko, RASPOZNAVANIE. Matematicheskie metody. Programmnaya sistema. Prakticheskie primeneniya. Izdatelstvo ”FAZIS”, Moscow, 168 p., 2006 (Russian).

[2] Ryazanov V.V., “Logicheskie zakonomernosti v zadachakh raspoznavaniya (parametricheskiy podkhod)”, Zhurnal vychislitelnoy matematiki i matematicheskoy fiziki, T. 47/10 () pp.1793-1809, 2007 (Russian).

[3] Aleksandrov V.V., Gorskiy N.D., Algoritmy i programmy strukturnogo metoda obrabotki dannykh. L. Nauka, 208 p., 1983 (Russian).

[4] H. Samet, R. Webber, “Storing a Collection of Polygons Using Quadtrees”, ACM Transactions on Graphics, July 1985, pp. 182-222.

[5] H. Eberhardt, V. Klumpp, U. D. Hanebeck, “Density Trees for Efficient Nonlinear State Estimation”, Proceedings of the 13th International Conference on Information Fusion, Edinburgh, 2010.

[6] M. de Berg, M. van Kreveld, M. Overmars, O. Schwarzkopf, Computational Geometry (2nd revised ed.). pp. 291–306, 2000.

[7] T. Lindeberg, Scale-Space Theory in Computer Vision. Kluwer Academic Publishers, 440 p., 1994.

[8] Vinogradov A., Laptin Yu., “Usage of Positional Representation in Tasks of Revealing Logical Regularities”, Rroceedings of VISIGRAPP 2010, Workshop IMTA-3, pp.100-104, 2010.

[9] J. Berman, Principles of big data: preparing, sharing, and analyzing complex information, Elsevier, 2013.

# Vehicle Video Detection and Tracking Quality Analysis*

V.D. Kustikova
Computational Mathematics and Cybernetics Department
Lobachevsky State University of Nizhni Novgorod
Nizhni Novgorod, Russian Federation
itlab.ml@cs.vmk.unn.ru

***Abstract*—This paper considers the problem of vehicle video detection and tracking. A solution based on the partitioning a video into blocks of equal length and detecting objects in the first and last frames of the block is proposed. Matching of vehicle locations in the first and last frames helps detect pairs of locations of the same object. Reconstruction of vehicle locations in the intermediate frames allows restoring separate parts of motion tracks. Combination of consecutive segments by matching makes it possible to reconstruct a complete track. Analysis of detection quality shows a true positive rate of more than 75% including partially visible vehicles, while the average number of false positives per frame is less than 0.3. The results of tracking of separate vehicles show that objects are tracked to the final frame. For the majority of them the average overlapping percent is not less efficient than the currently used Lucas-Kanade and Tracking-Learning-Detection methods. The average tracking accuracy of all vehicles makes about 70%.**

***Keywords—computer vision; object detection; tracking; feature extraction; detector; descriptor.***

## I. Introduction

This paper considers the practically important problem of vehicle video detection and tracking. The problem occurs in the course of analysis of qualitative and quantitative composition of transport flow. In comparison with [1] this paper proposes new modifications of the solution method for the specified problem allowing to increase the accuracy of detection and tracking.

This paper is organized as follows. First an overview of existing methods is given. Then the problem of vehicle video detection and tracking is formulated. A scheme of the proposed solution method is provided according to [1]. Principal modifications of the method are given. Details of implementation and experimental results are discussed.

## II. Related Works

A review and classification of the existing methods of video-based object detection problem are provided in [1]. The problem includes object detection in frames and their subsequent tracking. The object tracking methods fall into several categories [2]:

- *Feature points tracking* [3 – 9]. Objects are represented in consecutive frames by sets of corresponding feature points. Deterministic methods [3] reduce the problem to the minimization of point descriptor compliance function, probabilistic – use an approach based on the concept of state space. Typical examples are methods based on the Kalman [4 – 6] and particle filters [7 – 9].
- *Kernel tracking* – tracking the shape of an object or its appearance described by a geometrical primitive (a template of a rectangular or oval shape, a projection of a three-dimensional model). As a rule, methods of this group are applied, if motion is determined by an ordinary shift, turn or affine transformation. In practice, tracking of components is performed using mean shift and its continuous modification (CAM Shift) [10].
- *Silhouette tracking* – tracking a contour or a set of interconnected simple geometrical primitives limiting tracked regions. There are separate methods for matching and tracking segments containing an object [11], and methods of tracking of a contour. Tracking of fragments is carried out by calculation of an optical flow for inner points of a region [12, 13].

This research was done at Intel-UNN Lab "Information Technologies" at CMC department in the Lobachevsky State University of Nizhni Novgorod.

## III. Video-Based Vehicle Detection Problem

The method of video detection deals with a sequence of video frames. Let us assume that the object location is defined by the bounding box placement [1]. Then the problem consists in mapping each frame into a set of objects locations and finding relevant vehicle location in pairs of consecutive frames to reconstruct tracks of vehicles. Thus, a track is an ordered sequence of locations of the same object in a corresponding set of video frames. As a vehicle can be overlapped completely by other traffic participants, vehicle location is not necessarily seen in consecutive frames. A formal description of the mathematical problem definition is provided in [1].

## IV. Video-Based Vehicle Detection Method

The idea of the proposed method is to divide a video into blocks of images of equal length and then to execute processing of each block. Let us assume that the set of all vehicle locations in the first frame of a block is constructed during the previous iteration. It includes a subset of locations seen in the previous frames, and a subset of objects locations found by the detection algorithm for the first time (could contain false positives). Then it is necessary to detect vehicles

in the last frame of the block, match the sets of locations in the first and last frames of the block and reconstruct the vehicle locations in intermediate frames. As a result of reconstruction existing tracks are continued or new ones are created. A more detailed description of the method is provided in [1]. Here we will dwell on the modifications that were made.

The processes of matching sets of vehicle locations constructed in the first and last frame of a block and further reconstruction of locations in the intermediate frames extensively use the operation of matching pairs of images. The operation is intended for building sets of feature points and their SURF descriptors [14] in every image, and for complete matching of the descriptors with the following cutoff of outliers by RANSAC [15]. If in [1] it is assumed that locations are of the same object with a maximal number of inliers, here we suggest using the maximal relative number of inliers. It represents the ratio of the absolute number of inliers to the total number of feature points in the first image of the pair considered. This modification will provide method stability for vehicles of different classes. For example, images of trucks may contain over 200 inliers in the case of full visibility, while images of cars – not more than 100. At the same time, relative numbers are approximately equal, which is also confirmed experimentally.

## V. Implementation

Implementation is based on OpenCV computer vision library [16]. Latent SVM [17] is used as detection algorithm. Unlike [1], the vehicle classifier (CAR class) has been trained using images from PASCAL Visual Object Challenge 2007 base [18] and one of the videos (hereinafter – ***track_10_5000-7000***) at the ratio of 50% to 50%: 1650 objects, 4250 images not containing objects. The model consists of two components, each of which defines a foreshortening (a view point). The source code and the model are available for downloading [28].

## VI. Vehicle Detection Quality

For analysis of vehicle video detection quality a few videos (frame rate – 25 FPS, resolution – 720x405 pixels) were collected:

- ***track_10_5000-7000*** (2000 frames = 80 seconds, ~3000 bounding boxes, 58 tracks) – video with only CAR class vehicles, which move in 4 lanes in one direction.
- ***track_10_7000-8000*** (1000 frames = 40 seconds, ~1000 bounding boxes, 30 tracks). Contains objects of the CAR and BUS classes.
- ***track_10_9000-11000*** (2000 frames = 80 seconds, ~2300 bounding boxes, 48 tracks). Contains objects of the CAR and BUS classes. The principle difference is a large number of trucks.

The marking included all vehicle locations with partially visible objects (up to 2% of visibility).

In addition, synthetic video sequences 2000 frames long containing stationary vehicles were generated. The video sequences were received as a result of repeated copying of one frame (fig.1).

- ***track_10_5000-7000_1044x2000***. Contains 3 objects: 2 partially visible, and 1 is seen less than by 50% (a car).
- ***track_10_5000-7000_1192x2000***. Contains 4 objects: 3 fully visible, 1 object is visible more than by 50% (a car entering review region the camera).
- ***track_10_5000-7000_656x2000***. Contains 3 objects: 2 completely visible, 1 object is visible by more than 50% (a truck entering the region of interest, the cabin of the driver is fully visible).
- ***track_10_9000-11000_206x2000***. Contains 2 fully visible objects (a truck and a car).


Fig. 1. Test video frames.

This work uses the following measurements for assessment of detection quality: average precision (*AP*) [18]; the true positive rate (*TPR*); the false detection rate (*FDR*); the average false positives per frame (*FPF*) [19 – 21]. An object is considered to have been detected correctly, if percentage of overlapping of detected and marked bounding boxes exceeds a threshold (for *TPR*, *FDR* and *FPF* it was selected as 50%).

The experimental results (table 1) show that application of the described method modifications allows improving detection quality for all test videos, but for the synthetic ones. This is explained by that in case vehicles are not moving the final result is determined only by the choice of detection algorithm. We note that for ***track_10_5000-7000_1044x2000*** an object seen less than by 50% is considered to have been found incorrectly due to inaccuracy of bounding box detecting (the intersection region of constructed and marked boxes ranging from 40 to 50%). For other videos the true positive rate increased by 6–11.5%, false positives rate decreased by 10.8–16.4%, and the average false positives per each 10 frames decreased by 1.4 and 2.1 objects respectively.

TABLE I. The vehicle detection results
(column 1 of each measurement corresponds to the results of [1], column 2 corresponds to the achieved results)

| Video | *AP* | | *TPR* (%) | | *FDR* (%) | | *FPF* | |
|---|---|---|---|---|---|---|---|---|
| *track_10_5000-7000* | 0.68 | **0.80** | 74.8 | **84.2** | 19.9 | **4.4** | 0.27 | **0.06** |
| *track_10_5000-7000_1044x2000* | 0.64 | **0.64** | 66.7 | **66.7** | 33.3 | **33.3** | 1 | **1** |
| *track_10_5000-7000_1192x2000* | 1 | **1** | 100 | **100** | 0 | **0** | 0 | **0** |
| *track_10_5000-7000_656x2000* | 1 | **1** | 100 | **100** | 0 | **0** | 0 | **0** |
| *track_10_7000-8000* | 0.68 | **0.80** | 71.3 | **82.8** | 32.4 | **16** | 0.38 | **0.17** |
| *track_10_9000-11000_206x2000* | 1 | **1** | 100 | **100** | 0 | **0** | 0 | **0** |
| *track_10_9000-11000* | 0.62 | **0.68** | 69.8 | **75.9** | 39.9 | **29.1** | 0.44 | **0.30** |

Let us analyze the consistence of the true positive rate and the false detection rate. We will remove vehicles from the marking that are visible more than the certain threshold and

compute appropriate measurements. Threshold changes from 0 (corresponds to the full marking) to 100% (corresponds to the subset of fully visible vehicles). Experimental results show that the true positive rate becomes greater than 90% if the marking contains objects visible more than 20%. When the marking contains only fully visible vehicles this measurement is about 96–98% for all test videos (fig. 2).

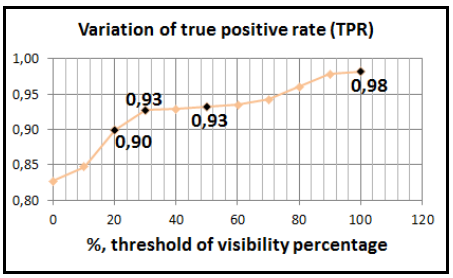



Fig. 2. Variation of true positive rate (*TPR*) for ***track_10_7000-8000*** while changing the visibility percentage of markup vehicles.

Obviously that at the same time the false detection rate increases (fig. 3) because of some partially visible vehicles detected correctly by the algorithm will be charged to the set of false positives.

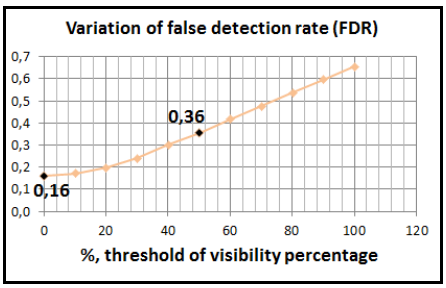



Fig. 3. Variation of false detection rate (*FDR*) for ***track_10_7000-8000*** while changing the visibility percentage of markup vehicles.

If to analyze a set of vehicles that were not detected by the algorithm we will find about ¾ objects visible less than 50% (entering/leaving into the frame), and ¼ objects visible more than 50% (substantially trucks which have invisible cabs).

Let us consider detection quality for the complicated video (fig. 4): resolution – 640x480, 3456 frames = 2 min 18 s, 25 FPS, 20292 bounding boxes, 117 tracks.

Fig. 4. Complcated traffic situation (another camera's point of view, shadows, high density of traffic flow, traffic signs, trolleybus lines).

The proposed method demonstrates the next results: *TPR*=73.7%, *FDR*=13.2%, *FPF*=0.63. True positive rate is worse by 2.2% in comparison with ***track_10_9000-11000***. False detection rate is better in average because this video contains more vehicles detected correctly than test. Average false positives per frame twice more, it is explained by the fact of existence of traffic signs and trolleybus lines located above the road (new overlapping situations which are nonstandard for the vehicle model). If to consider only fully visible vehicles in the marking the true positive rate achieves 93%.

## VII. Single Object Tracking Quality

For analysis of tracking quality tracks of separate objects (fig. 5) of the test videos were selected:

- ***track_10_5000-7000_103-121*** contains fully visible moving car.
- ***track_10_5000-7000_757-780*** contains fully visible moving car of another appearance.
- ***track_10_5000-7000_852-889*** contains fully visible moving truck.
- ***track_10_5000-7000_1712-1727*** contains a moving car overlapped less than 50%.
- ***track_10_5000-7000_1569x100*** contains a stationary car, no other objects in the frame.
- ***track_10_5000-7000_1041x100*** contains a stationary car is overlapped by another object and is visible approximately by 40%.
- ***track_10_9000-11000_1563-1591*** contains a car entering the frame, visibility increased from ~50% to 85%, then decreased to ~30% when another object (bus) appears in the foreground.


Fig. 5. The initial frames of test sequences for assessment of single vehicle tracking quality.

The tracking was carried out using the following methods:

- The proposed method of vehicle video-based detection.
- The Lucas-Kanade algorithm based on optical flow calculation [12, 13]. Pyramidal implementation from OpenCV [22] library was used.
- Median flow algorithm (Predator or Tracking-Learning-Detection, TLD) [23, 24]. An open-source implementation of the algorithm authors [25] was used.

We note that each implementation uses detection results of Latent SVM for CAR objects: the proposed method – in every fifth frame, the other two algorithms – only in the initial frame of the sequence.

To compare the tracking quality of a single object two metrics [26] were used:

- *k* – the ratio of the number of frames, in which the object was tracked until it was lost to the total number of frames containing a track.
- *AvIP* – the average percentage of bounding boxes overlapping in marked and built tracks.

The final values of measurements for the selected set of tracks and methods are shown below (table 2). It can be easily seen that in all the test sequences the developed method tracks the object without losing it in the intermediate frames (column 2). And in most cases the proposed method is not less effective than the TLD (columns 3 and 7, the shadowed cells). The difference (1–10%) in sequences ***track_10_5000-7000_757-780*** and ***track_10_5000-7000_852-889*** (lines 2 and 3) is caused by the fact that the sizes of a bounding box change, when the detection algorithm of the proposed method works at another time. It is noteworthy that small losses of simplicity in the scene lead to a decrease in quality of tracking for the TLD (columns 6 and 7, test sequences ***track_10_5000-7000_1041x100*** and ***track_10_5000-7000_1563-1591***). Partial overlapping of a car by a moving bus (***track_10_5000-7000_1563-1591***) causes losing of the object by the algorithm in the last frames (*k*=0.86), and in the initial frames the average percentage of overlapping with marking differs almost twofold in comparison with the proposed method. In practice, such overlapping occurs quite often, which makes application of this tracking algorithm more complicated. It should be noted that the Lucas-Kanade algorithm does not change the size of a bounding box. In every next frame the location is reconstructed based on the mutual location of feature points and the bounding box in the previous frame. As the scale of an object is slightly changed in motion, the algorithm loses on this measurement of quality in all tests containing moving vehicles (lines 1, 2, 3, 4 and 7). And in cases with stationary objects (line 5 and 6) the possible maximum is reached.

TABLE II. SINGLE VEHICLE TRACKING QUALITY

| Test track | Proposed Method | | Lucas-Kanade | | Tracking-Learning-Detection | |
|---|---|---|---|---|---|---|
| | ***k*** | ***AvIP*** | ***k*** | ***AvIP*** | ***k*** | ***AvIP*** |
| ***track_10_5000-7000_103-121*** | 1 | **0.773** | 1 | 0.640 | 1 | 0.731 |
| ***track_10_5000-7000_757-780*** | 1 | 0.690 | 1 | 0.662 | 1 | **0.794** |
| ***track_10_5000-7000_852-889*** | 1 | 0.688 | 1 | 0.651 | 1 | **0.698** |
| ***track_10_5000-7000_1712-1727*** | 1 | **0.832** | 1 | 0.709 | 1 | 0.732 |
| ***track_10_5000-7000_1569x100*** | 1 | 0.776 | 1 | 0.776 | 1 | 0.776 |
| ***track_10_5000-7000_1041x100*** | 1 | **0.493** | 1 | 0.493 | 1 | 0.479 |
| ***track_10_9000-11000_1563-1591*** | 1 | **0.614** | 1 | 0.396 | **0.86** | **0.328** |

## VIII. ALL OBJECTS TRACKING QUALITY

For assessment of tracking quality for all vehicles in test videos one of the most representative metrics was chosen – average tracking accuracy [27]. This indicator reflects the percentage of marked and built tracks overlapping. The results (table 3) show that the average tracking accuracy does not get lower than 0.7 in all test videos with the exception of the last on. It means that the percentage of overlapping of bounding boxes of the marked and constructed tracks is not less than 70%. It should be noted that it is a high result as the location of a vehicle is determined by a bounding box, which inevitably contains parts of the background or other overlapping objects in its corner segments. The value of the measurement is 2% lower for ***track_10_9000-11000*** – 67.6%. This is due to inaccurate building of bounding boxes. The video contains a large number of trucks, and the detection algorithm, as a rule, finds only the driver's cabin, sometimes with a small part of the body.

TABLE III. AVERAGE TRACKING ACCURACY

| Video | Average tracking accuracy (%) |
|---|---|
| ***track_10_5000-7000*** | 72.1 |
| ***track_10_5000-7000_1044x2000*** | 72.9 |
| ***track_10_5000-7000_1192x2000*** | 77.8 |
| ***track_10_5000-7000_656x2000*** | 74.7 |
| ***track_10_7000-8000*** | 70.4 |
| ***track_10_9000-11000_206x2000*** | 74.6 |
| ***track_10_9000-11000*** | 67.6 |

## IX. CONCLUSION

This paper proposes a modification of the method described in [1]. It was proved that the modification allows improving vehicle detection and tracking quality. The true positive rate grows by 6–11.5% depending on the video compared to [1] and made over 75% provided that the marking contains objects with visibility of up to 2%.

The results of tracking of separate vehicles show that the proposed method is not less effective than the Lucas-Kanade and TLD methods, and for some test objects the overlapping of vehicle locations in marked and built tracks is 10% higher, than for existing methods. The tracking average accuracy of all objects that reflects the percentage of tracks overlapping in a video as a whole makes around 70%.

## REFERENCES

[1] V. Kustikova, I. Meyerov, N. Zolotykh, "Vehicle video detection method," Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications, vol. 24, no. 4, 2014, in press.

[2] A. Yilmaz, O. Javed, M. Shah, "Object tracking: A survey," ACM Computing Surveys, vol. 38, no. 4, article 13, 2006.

[3] C. Veenman, M. Reinders, E. Backer, "Resolving motion correspondence for densely moving points," IEEE Trans. Pattern Analysis and Machine Intelligence", vol. 23, no. 1, pp. 54-72, 2001.

[4] A. Salarpour, A. Salarpour, M. Fathi, D. MirHossein, "Vehicle tracking using Kalman filter and features," Signal & Image Proc.: An Int. J. (SIPIJ), vol. 2, no. 2, 2011.

[5] S. Dan, Zh. Baojun, T. Linbo, "A tracking algorithm based on SIFT and Kalman filter," In Proc. of the 2nd Int. Conf. on Computer Application and System Modeling, pp.1563-1566, 2012.

[6] N. Li, "Corner feature based object tracking using adaptive Kalman filter," In Proc. of the 9th Int. Conf. on Signal Processing (ICSP 2008), pp. 1432-1435, 2008.

[7] M. Isard, A. Blake, "Condensation – conditional density propagation for visual tracking," International Journal of Computer Vision, vol. 29, no. 1, pp. 5-28, 1998.

[8] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, P.J. Nordlund, "Particle filters for positioning, navigation and tracking ," IEEE Trans. on Signal Processing, vol. 2, Issue 2, pp. 425-437, 2002.

[9] Particle filter object tracking [http://blogs.oregonstate.edu/hess/code/particles].

[10] D. Exner, E. Bruns, D. Kurz, A. Grundhofer, "Fast and robust CAMShift tracking," In Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshops, pp. 9-16, 2010.

[11] K. She, G. Bebis, H. Gu, R. Miller, "Vehicle tracking using on-line fusion of color and shape features," In Proc. of the 7th Int. IEEE Conf. on Intelligent Transportation Systems, pp. 731-736, 2004.

[12] J. Shi, C. Tomasi, "Good features to track," In Proc. of the Conf. on Computer Vision and Pattern Recognition, pp. 593-600, 1994.

[13] Pyramidal implementation of the Lucas Kanade feature tracker [http://robots.stanford.edu/cs223b04/algo_tracking.pdf].

[14] H. Bay, A. Ess, T. Tuytelaars, L.V. Gool, "SURF: speed up robust features," Computer Vision and Image Understanding (CVIU), vol. 110, no. 3, pp. 346-359, 2008.

[15] J. Ponce, D.A. Forsyth. Computer vision. A modern approach. 2004.

[16] OpenCV Library [http://opencv.org].

[17] P.N. Druzhkov, V.L. Erukhimov, N.Yu. Zolotykh, E.A. Kozinov, V.D. Kustikova, I.B. Meerov, A.N. Polovinkin, "New object detection features in OpenCV library," Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications, vol. 21, no. 3, pp. 384-386, 2011.

[18] PASCAL Visual Object Challenge 2007 [http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2007].

[19] Z.W. Kim, J. Malik, "Fast vehicle detection with probabilistic feature grouping and its application to vehicle tracking ," In Proceedings of the International Conference on Computer Vision, vol. 1, pp. 524-531, 2003.

[20] S. Sivaraman, M.M. Trivedi, "A general active-learning framework for on-road vehicle recognition and tracking," IEEE Transactions on Intelligent Transportation Systems, vol. 11, no. 2, pp. 267-276, 2010.

[21] X. Song, R. Netavia, "A model-based vehicle segmentation method for tracking ," In Proceedings of the International Conference on Computer Vision, vol. 2, pp. 1124-1131, 2005.

[22] OpenCV library functions for calculation of an optical flow using the Lucas Kanade algorithm [http://docs.opencv.org/modules/video/doc/motion_analysis_and_object_tracking.html#id1].

[23] Z. Kalal, K. Mikolajczyk, J. Matas, "Tracking-Learning-Detection," IEEE Trans. on PAMI, vol. 34, no. 7, pp. 1409-1422, 2012.

[24] A description of the Predator algorithm [http://robot-develop.org/archives/4463].

[25] Implementation of the Predator algorithm (Tracking-Learning-Detection) [https://github.com/zk00006/OpenTLD].

[26] R.T. Collins, X. Zhou, S.K. The, "An open source tracking testbed and evaluation web site," In IEEE Int. Workshop on PETS, 2005, [http://www.cs.cmu.edu/~rcollins/Papers/opensourceweb.pdf].

[27] R. Kasturi, D. Goldgof, P. Soundararajan, V. Manohar, J. Garofolo, M. Boonstra, V. Korzhova, J. Zhang, "Framework for performance evaluation of face, text, and vehicle detection and tracking in video: data, metrics, and protocol," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pp. 319-336, 2009.

[28] Developed vehicle video-based detection system [http://ml.vmk.unn.ru/index.php/en/resources-en].