THE RUSSIAN ACADEMY OF SCIENCES

THE MATHEMATICAL SCIENCES BRANCH OF THE RUSSIAN ACADEMY OF SCIENCES

THE NATIONAL COMMITTEE FOR PATTERN RECOGNITION AND IMAGE ANALYSIS

INTERNATIONAL ASSOCIATION FOR PATTERN RECOGNITION (TC 16)

DORODNICYN COMPUTING CENTRE OF THE RAS
(Moscow, The Russian Federation)

S.P. KOROLYOV SAMARA STATE AEROSPACE UNIVERSITY
(NATIONAL RESEARCH UNIVERSITY)
(Samara, The Russian Federation)

IMAGE PROCESSING SYSTEMS INSTITUTE OF THE RAS
(Samara, The Russian Federation)

DEPARTMENT OF INFORMATION TECHNOLOGIES AND COMMUNICATIONS OF THE SAMARA REGION

COMPUTER TECHNOLOGIES, JSC (Samara, The Russian Federation)

INFORMATION RESEARCH AND DEVELOPMENT, LTD. (Moscow, The Russian Federation)

The 11th International Conference «PATTERN RECOGNITION and IMAGE ANALYSIS: NEW INFORMATION TECHNOLOGIES» (PRIA-11-2013)

September 23-28, 2013

CONFERENCE PROCEEDINGS

Volume I

Samara,
The Russian Federation
2013

УДК 004.85+004.89+004.93+519.2+519.7

ISBN 978-5-88940-130-8

The papers are published as they were presented by the authors.

11th International Conference on Pattern Recognition and Image Analysis: New Information Technologies (PRIA-11-2013). Samara, September 23-28, 2013. Conference Proceedings (Vol. I-II), Volume I, Samara: IPSI RAS, 2013. 376 p.

MESSAGE FROM PRIA-11-2013 CHAIRS

It is both an honor and a pleasure to hold the 11th International Conference on PATTERN RECOGNITION and IMAGE ANALYSIS: NEW INFORMATION TECHNOLOGIES (PRIA-11-2013) at the S.P.Korolyov Samara State Aerospace University (National Research University) and the Image Processing Systems Institute of the Russian Academy of Siences (Samara, the Russian Federation), hosted by the National Committee of the Russian Academy of Sciences for Pattern Recognition and Image Analysis and the A.A.Dorodnicyn Computing Centre of the Russian Academy of Sciences (Moscow, the Russian Federation). The conference is supported by the Russian Foundation for Basic Research, Information Research and Development, Ltd. (Moscow, the Russian Federation) and Open Join-Stock "Samara-Informsputnik"(Samara, the Russian Federation).

This is the latest in a long line of the premier international conference in the field of pattern recognition and image analysis, held under the auspices of the National Committee of the Russian Academy of Sciences for Pattern Recognition and Image Analysis and of the International Association of Pattern Recognition (TC 16) in accordance with the List of meetings, conferences, congresses, symposia, workshops and schools of the Russian Academy of Sciences in natural and social sciences for the year of 2013.

PRIA-11-2013 is the main international conference in the Russian Federation and Central and Eastern Europe in the field of pattern recognition and image analysis. The previous PRIAs conferences were held in Minsk, Byelorussia (PRIA-1-91, 1991), Ulyanovsk, the Russian Federation (PRIA-2-95, 1995), Nizhny Novgorod, the Russian Federation (PRIA-3-97, 1997), Novosibirsk, the Russian Federation (PRIA-4-98, 1998), Samara, the Russian Federation (PRIA-5-2000, 2000), Veliky Novgorod, the Russian Federation (PRIA-6-2002, 2002), St.-Petersburg, the Russian Federation (PRIA-7-2004, 2004), Yoshkar-Ola, the Russian Federation (PRIA-8-2007, 2007), Nizhny Novgorod, the Russian Federation (PRIA-9-2008, 2008) and St.-Petersburg, the Russian Federation (PRIA-10-2010, 2010).

PRIA-11-2013 included:

- the conference itself;
- tutorials;
- exhibition of image analysis and pattern recognition software products;
- meeting of the National Committee of the RAS for Pattern Recognition and Image Analysis;
- meeting of the IAPR Technical Committee 16 "Algebraic and Discrete Mathematics Techniques in Pattern Recognition and Image Analysis";
- social program.

The theme of PRIA-11-2013 is "Knowledge Itself is Power: the Russian Academy of Sciences lived, the Russian Academy of Sciences lives, the Russian Academy of Sciences will live eternally". The conference is a multi-track international forum for discussions on recent advances in fundamental and applied researchin the field of prospective information technologies for pattern recognition, image analysis and artificial intelligence, such as follows:

a) formulating, analysing, and solving of mathematical problems arising in the development and implementation of data transformation and estimation methods in pattern recognition, classification, and forecasting, with input data represented in the form of numerical and text arrays, expert data, signals, images, time-domain series, and random multidimensional fields;
b) image processing, analysis, recognition, understanding, and synthesis (including computer graphics, visualization, and virtual reality);
c) speech processing, analysis, recognition, and synthesis;
d) development, investigation, modification, and systematization of mathematical and computational methods providing the algorithmic foundation for information technologies;

e) development of methods for the computer-aided synthesis and testing information technologies;
f) development of dedicated information technologies (for classes of problems and subject domains);
g) development and implementation of general-purpose toolkits, databases, and knowledge bases to support information technologies for pattern recognition, image analysis, speech analysis and synthesis, and signal processing;
h) development, adaptation, and implementation of methods to handle off-standard, critical, and mass-scale tasks of pattern recognition and image, speech, and signal analysis and understanding.

The main topics of PRIA-11-2013 Scientific Program are as follows:

1. Mathematical Theory of Pattern Recognition.
2. Mathematical Theory of Image Processing, Analysis, Recognition and Understanding.
3. Mathematical Theory of Speech Processing, Analysis, Recognition and Understanding.
4. Models, Methods and Tools to Represent the Initial Data for Pattern Recognition, Image and Signal Analysis.
5. Automation of Information and Knowledge Extraction, Processing and Estimation in Solving Pattern Recognition, Image, Speech and Signal Processing, Analysis and Understanding Tasks.
6. Information Technologies and Software for Pattern Recognition, Image, Speech and Signal Processing, Analysis and Understanding.
7. Databases, Knowledge Bases, and Linguistic Tools Supporting Information Technologies for Pattern Recognition, Image, Speech and Signal Processing, Analysis and Understanding.
8. Special-Purpose Architectures, Software and Hardware Tools Supporting Information Technologies for Pattern Recognition, Image, Speech and Signal Processing, Analysis and Understanding.
9. Neural Networks for Data Processing, Analysis and Interpretation.
10. Algorithms, Software and Information Technologies for Intellectual Geographic and Cartographic Information Systems. GIS Technologies.
11. Algorithms, Software and Information Technologies for Biomedical and Biotechnical Systems.
12. Algorithms, Software and Information Technologies in Bioinformatics and Medical Informatics.
13. Video Processing and Analysis.
14. Computer Graphics, Visualization, and Virtual Reality.
15. Optical and Digital-Optical Systems for Image and Signal Processing and Analysis.
16. Applied Problems being Solved via Pattern Recognition, Image, Speech and Signal Processing, Analysis and Understanding Approaches and Techniques.
17. Visual and Speech Perception.
18. Multimedia Technologies.

The scientific Program of PRIA-11-2013 included 210 papers (21 Plenary and invited papers, 2 tutorials, 72 session papers, 112 posters, 3 exhibits). These presentations were organized into four main tracks, some with associated themes:

- Track 1 "Mathematical Methods in Pattern Recognition" (12 session papers, 4 posters);
- Track 2 "Representation, Processing, Analysis, and Understanding of Images" (32 session papers, 28 posters);
- Track 3 "Software and Hardware for Pattern Recognition and Image Analysis" (9 session papers, 19 posters);
- Track 4 "Applied Problems" (19 session papers, 61 posters).

PRIA-11-2013 Proceedings records the contributions presented at the conference. They are a witness to the continuous advances being made by the pattern recognition and image analysis community in fundamental research and to the associated impact on the technological developments that reach the market place.

In putting the Proceedings together, many played a significant role which we would like to acknowledge. First of all our thanks are due to the authors who have contributed their work into the conference. Second, we owe immense gratitude to the Program Committee members and reviewers

for their dedicated effort in evaluating the submitted papers and in providing the necessary decision support information. Third, our gratitude to the staff of the Local Committee for technical implementation of Proceedings preparation and production.

We appreciate contribution of the PRIA-11-2013 sponsors:

- The Mathematical Sciences Branch of the RAS;
- The Open Join-Stock "Samara-Informsputnik"(Samara, the Russian Federation)
- The Information Research and Development, Ltd. (Moscow, the Russian Federation).
- The Russian Foundation for Basic Research.

We hope that the Proceedings, following the tradition of all PRIA conferences, will not only impact on the current research of the readers but will also represent important archival material to which future generations of researchers in pattern recognition and image analysis that will be returning for years to come.

Full texts of PRIA-11-2013 papers presented by the authors at the conference and recommended by the Program Committee will be published in the year of 2014 as special issues of the international journal "Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications" (Pleiades Publishing Ltd./c/o INTERNATIONAL ACADEMIC PUBLISHING COMPANY "Nauka/Interperiodica", Moscow, distributed worldwide by SPRINGER).

Conference Chair	Professor, Full Member of the RAS **Yuri Zhuravlev,** Chair of the National Committee of the Russian Academy of Sciences for Pattern Recognition and Image Analysis, Dorodnicyn Computing Centre of the Russian Academy of Sciences, Moscow, the Russian Federation, zhur@ccas.ru
Program Committee Chair	Professor **Heinrich Niemann**, Friedrich-Alexander-University of Erlangen-Nuremberg, Erlangen, Germany, niemann@informatik.uni-erlangen.de
Conference Vice Chairs	Dr.-Eng. **Igor Gurevich,** Scientific Secretary of the National Committee of the Russian Academy of Sciences for Pattern Recognition and Image Analysis, Dorodnicyn Computing Centre of the Russian Academy of Sciences, Moscow, the Russian Federation, igourevi@ccas.ru
	Professor, corresponding member of the RAS **Victor Soifer,** Image Processing Systems Institute of the Russian Academy of Sciences, Samara, the Russian Federation, ipsi@smr.ru
Local Committee Chair	Professor **Vladislav Sergeev,** Image Processing Systems Institute of the Russian Academy of Sciences, Samara, the Russian Federation, vserg@smr.ru

CONTENTS

INVITED TALKS AND TUTORIALS

SOFTWARE TECHNOLOGY OF IMPACT CRATERS DETECTION ON SPACE IMAGES[1]

A.A. Buchnev[2,3], V.P. Pyatkin[2,4]

[2] Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Academician Lavrentiev Av., 6, Novosibirsk 630090, Russia, Tel. (+7-383)330-73-32, E-mail: [3] baa@ooi.sscc.ru, [4] pvp@ooi.sscc.ru, sgi@ooi.sscc.ru

In geological studies of the Earth from space and in the problem of the Earth space protection there is an urgent need to study the processes of celestial bodies falling to Earth and mapping these places. In article available nonparametric statistical test to detect the circular structures in space pictures representing the impact craters on the Earth's surface is discussed. We also consider issues related to software implementation of the algorithms. The processing results of real space images confirm the effectiveness of the proposed statistical approach to the detection of impact craters on Earth's surface.

Introduction

The program of applied remote sensing of the Earth from space is vast and varied [1]. In geological studies of the Earth from space and in the problem of the Earth space protection there is an urgent need to study the processes of celestial bodies falling to Earth and mapping these places. Materials aerospace survey, which showed widespread circular formations in the structure of the crust (see http://tsun.sscc.ru/nh/edeis.html), led to widespread interest in them on the part of researchers impact craters on the Earth's surface and geologists.

Statistical approach for circular structures detection

When processing of satellite images to detect the objects of interest, for a number of reasons, the preference given to the statistical approach [1-4]. The main reason is that, due to the random nature of the natural processes, the remote sensing data (spectrum superposed image) contain a lot of random variations masking the distinctions in the values of the brightness of the image points of the object area and the background area. The random variables which values are obtained from measurements (observations) in the image points are called the observed random variables. In such a situation, reliable detection algorithms of detection can only be built with the help of the probabilistic (statistical) approach. In this way, the best algorithm may be obtained if the probability distribution of the values of the observed values at the points of the object and the background area are known in advance. However, the distribution of each image may be the same and may vary even within the single image, so they are usually not known to the observer in practice. We only know that they are continuous. Here nonparametric criteria are efficient to detect the objects, as criteria statistics distributions don't depend on distributions of studied values (unknown to an observer), when objects are absent in the field of vision. The detection scheme is as follows. All (or almost all) possible positions of the objects of interest are analyzed sequentially (or in parallel). It is considered, that in the investigations of each possible position of the object, all the observed quantities are measured in the points that are rather far from each other in the image so that they can be considered as statistically independent jointly when there is no object in the field of view. In each situation, to minimize the risk of making a wrong conclusion, we should check the statistical hypothesis that the random variables observed in the tested points of the object and the surrounding background points are identically distributed (homogeneous), indicating the absence of the object.

The detection of poorly visible objects which form is close to a circle (disc) and other objects with the form known to the observer, when the

[1]The work was supported in part by Russian Foundation for Basic Research (grant No. 13-07-00068-a).

homogeneity of the background in diameter (or width) of the object does not take place, can be realized by detecting their edges. If we want to take more simple algorithm (or algorithms) of edges detection of the object which is close to a circle, it is enough to take the observation in the points, located only at two circles with a common center, close to the center of a possible (audited) position of the object, and with such radii that the edge of the object (in the case of the object presence) is located within the band between the two circles. Suppose for $i = 1,...,k$ ζ_i and ξ_i – the values observing in the points of intersection of the line i-th normal to check the position of the object edge with these circles, i.e. at the points of the checked object area and in the closely related points in the surrounding background of the object respectively, and let the observer know from the past experience that the brightness ζ_i at the points of the object area is stochastically more or less than the brightness ξ_i at the points background. Then, to test the hypothesis of homogeneity, one can use the well-known in mathematical statistics criterion of signs (in this case) the differences $\zeta_i - \xi_i$. To make possible to use a more efficient test we should increase the number of observed values for each normal. The first step in this direction is to attract the elements at the circle of a larger radius. Let's denote the values ξ_{i1}, ξ_{i2} observing at the line of i-th normal in this case. According to the results of the works [3-4] we consider the following statistics: $S^+ = \sum_{i=1}^{k} I\{\zeta_i > \max(\xi_{i1}, \xi_{i2})\}$, $S^- = \sum_{i=1}^{k} I\{\zeta_i < \min(\xi_{i1}, \xi_{i2})\}$, $S^0 = k - S^+ - S^-$, where $I\{\cdot\}$ - the indicator function of the event $\{\cdot\}$, equal to 1 or 0 depending whether the event $\{\cdot\}$ took place or not. Then, to test the hypothesis of homogeneity (object is absent) the following statistical test $|S^+ - S^-| > 2\lambda(S^0) - (k - S^0)$ will be the most effective. Here $\lambda(z)$ - the smallest integer λ, such as $\sum_{i=\lambda+1}^{k-z} \binom{k-z}{i} 2^{-(k-z)} \le \alpha/2$, where α - a guaranteed level of significance criterion. Since the left side is increasing with decreasing λ discretely, a true significance level of two-sided criteria, which was introduced, may turn out to be significantly less α. The ring object is a fuzzy ring, close to round, i.e. a closed fuzzy band. These objects can be detected using a criterion similar to the above. The observed values should be taken similarly on the sequence of normals, but now to the midline of the checked position of the band, and the values ξ_{i1} and ξ_{i2} are taken at the points on opposite sides of the checked position of the band, and the values ζ_i - at the points belonging to the area of the band itself. Statistics S^+, S^-, S^0 and criteria are defined using the same formulas.

Software technology of impact craters detection

The implementation of algorithms for the detection of circular objects is a part of the software package PlanetaMonitoring for aerospace images processing [5]. The full set of complex operations PlanetaMonitoring allows building the software technologies aimed at specific applications. In particular, one of the technologies is technology of the circular structures detection in the satellite images. This technology is divided into two steps – the step of the preparing of processed image and the actual step of the required shape structures detection. The first step is required in the cases when the earth area of researchers' interest is covered by several satellite images obtained at different times and with different conditions of the space survey. In the general case, this step requires the following operations:

1. Enhancement of brightness and contrast characteristics of satellite images obtained at different times. In particular, the technique of the reference stretch [6], in which the brightness and contrast characteristics of the images are stretched either the characteristics of a picture which is selected as a reference (base) or the characteristics of the Gaussian distribution with the specified parameters, is used.

2. Build a digital map of the desired part of the earth surface. Map building is the creation of a clean map sheet with the specified parameters (the type of projection, the coordinates of the

northern and southern latitudes, the coordinates of the western and eastern longtitudes, the parameters of coordinate grid, map resolution) and subsequent plotting of the existing digital information about geographic features of the specified area on this sheet. The use of GIS ArcView shape files is preferred as the information contained there is more detailed than in files of other formats.

3. Successive mapping of satellite images on the map which has been made. The mapping is based on the mechanism of control points where identifiable objects of the image and the map are labeled as the same. By the set of control points of the image and the map the polynomial of the given degree (first, second or third) is constructed, with the help of which you are mapping. There may be used one of three types of interpolation of pixel values: no interpolation, bilinear or bicubic interpolation.

4. The construction of the single image from the resulting projections of the images on the map (mosaic).

The following figures illustrate the results of operations of this step. Fig. 1 and 2 show the images of impact craters Kurai obtained by satellite SPOT-4 (resolution 20 m) 08.10.2007 and 14.05.2012 respectively. Note, that all satellite images used for experiments were kindly provided by Siberian Center FGBU "SRC “Planeta". Fig. 3 shows the results of operations of the first step with respect to the original space images.

The content of the second of the above steps is the software implementation of the algorithm of the circular objects detection in satellite images.

For the working of the procedure a set of parameters is given, the main ones are the radius of the circular object (the setting of a range of values is possible), the allowable probability of false detection of the object, the distance from the expected position of the object to measure the values of the pixels and the number of such measurements along the circle. The main part of the procedure is to calculate the previously mentioned statistics S^+ and S^- on the base of the analysis of the values of the image pixels along the normals to the expected position of the object. For each possible position the decision of the object presence is made by the result of comparing of the counted statistics S^+ and S^- values with the predefined threshold values obtained on the basis of input parameters. Fig. 4 shows the result of applying of the second step to the image in Fig. 3.

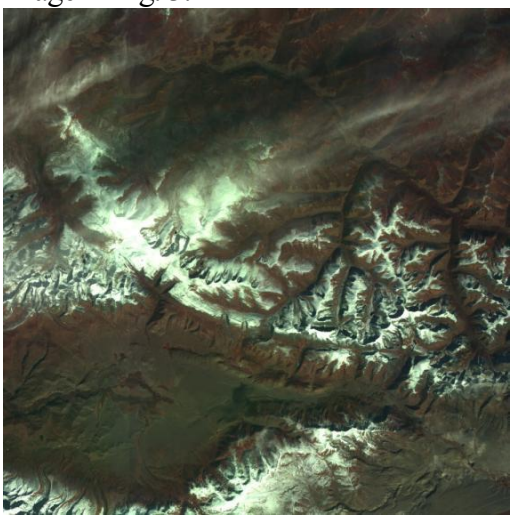


Fig. 1. 08.10.2007

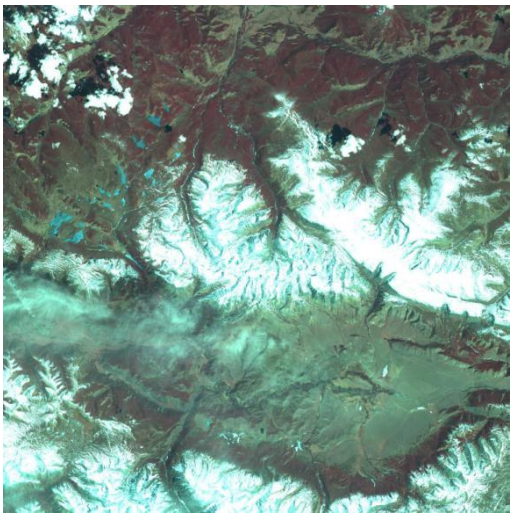


Fig. 2. 14.05.2012

The analysis of pixel values used to build the statistics S^+ and S^- is time-consuming when it’s executed sequentially. Therefore, two parallel implementation of the algorithm of circular structures detection are designed. One of them is made in the parallel programming system MPI (Message Passing Interface) at the process level, which provides portability of software to various multiprocessor systems. The second implementation is achieved by means of the Windows at the level of threads: the number of running threads is equals the number of logical processors. Each process (or thread), having the total number of processes and its own number, determines the horizontal band of the image to be processed. The bands of the neighboring processes overlap and the depth of the overlap depends on the radius of the

detected objects. Besides crater Kurai, numerous experiments with space images that contain other reliable impact structures of the Earth were studied during the work.

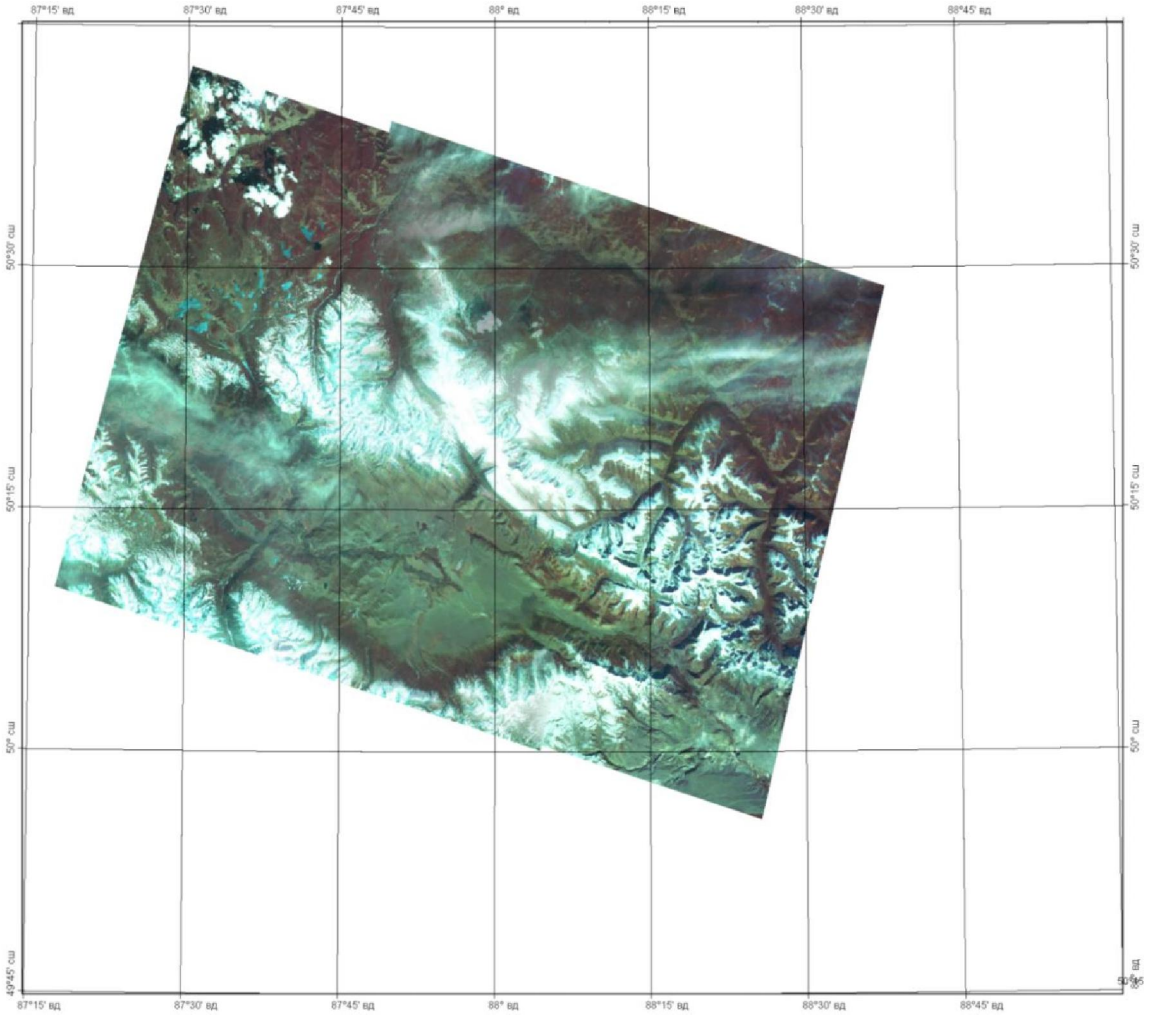


Fig. 3. The result of the operation of the first stage

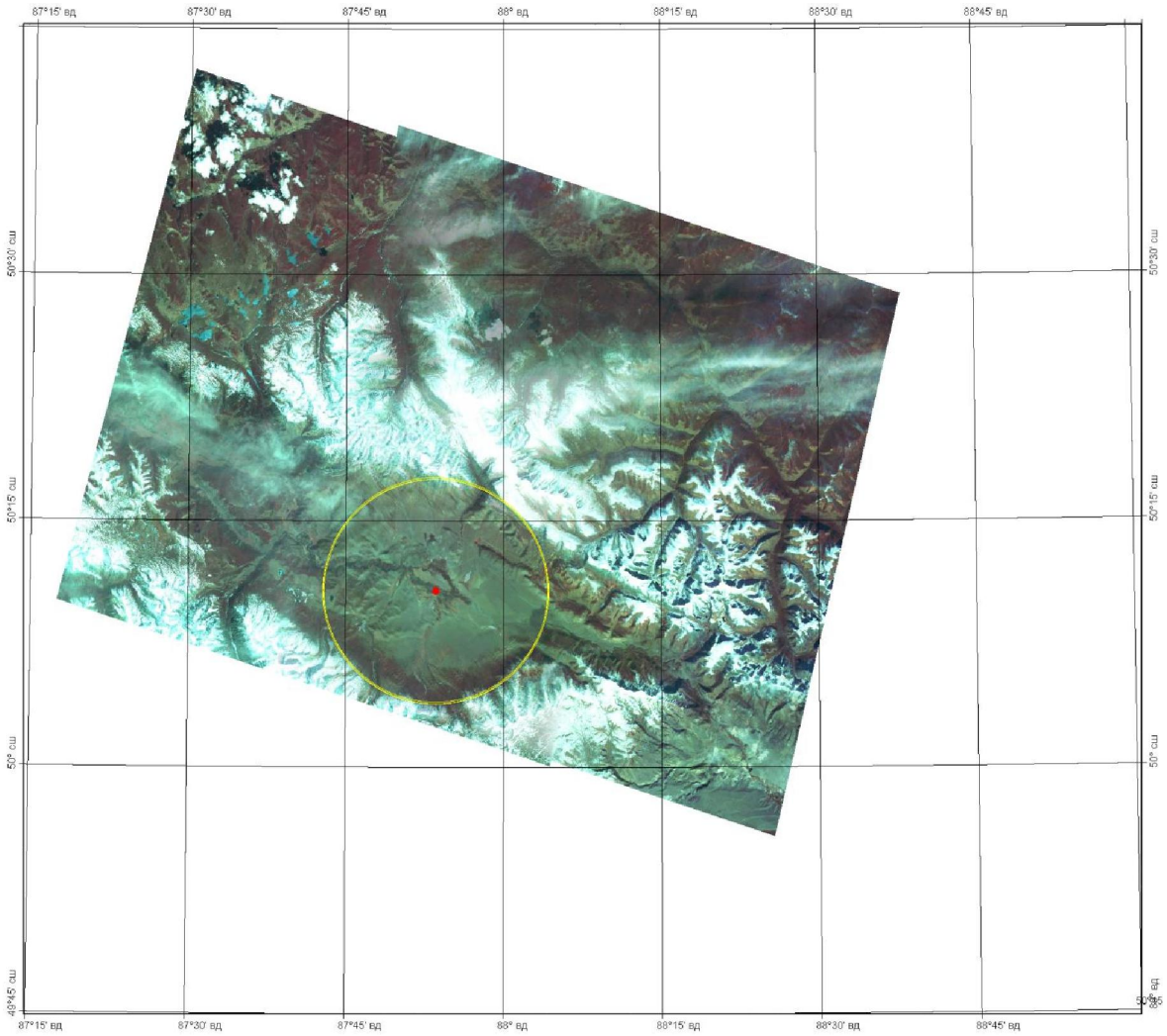


Fig. 4. The result of the detection of impact crater Kurai

Conclusion

Let's notice that for the first time elements of the described program technology (the truth, with other criterion) were used at detection of an impact crater of Zhamanshin in Northern Priaralye [5]. The developed software allows to detect on space images not only circular structures, but also contours of any form. In this case contours are interpolated by chains of the linear elements (lineaments), which construction is made according to the algorithm given above.

References

1. Remote Sensing: The Quantitative Approach. – Edited by P.H. Swain and S.M. Davis. – USA, McGraw-Hill, Inc. – 1979. – 396 p.
2. Alekseev A.S., Pyatkin V.P., Salov G.I. Crater Detection in Aerospace Imagery Using Simple Nonparametric Statistical Tests. //Lecture Notes in Computer Science, Springer-Verlag, V.179, 1993, p.793-799.
3. Salov G.I. On the power of nonparametric tests for the detection of extended objects on a random background. // Avtometriya. 1997. N 3. p. 60-75. (in Russian).
4. Salov G.I. A new statistical test for problems with two and three samples, more powerful than the Wilcoxon and the Whitney tests. // Avtometriya. 2011. N 4. p. 58-70 (in Russian).
5. Asmus V.V., Buchnev A.A., Pyatkin V.P., Salov G.I. Software System for Satellite Data Processing of Applied Tasks in Remote Sensing of the Earth. //Pattern Recognition and Image Analysis, vol. 19, No.3, 2009.
6. Schowengerdt R.A. Remote sensing: models and methods for image processing. - Amsterdam: Elsevier; Burlington: Academic Press, 2007. - 515 p.

DISCRIMINATION OF 3D OBJECT BY ITS SHAPES OF PROJECTIONS

Y.A. Furman[1], R.V. Eruslanov[1], I.L. Egoshina[1] ,

[1]Volga State University of Technology, 424000, Russia, Mari El, Yoshkar-Ola, Lenin sq., 3, Krtmbs@volgatech.net

The problem of distinguishing of 3D objects reconstructed from their projections on the underlying surface are solved. Reconstruction are performed by the information contained in the contours of the projections. Methods of retrieval and detection of conjugated points which are required for calculation of coordinate of 3D object contour points are substantiated. Algorithm of extracting of its mathematical model in the form of spatial contour which are uniquely pass through all vertices of object are created. The optimal maximum-likelihood algorithm of distinguishing the noised reconstructed 3D object has been synthesized and its effectiveness has been determined experimentally.

Introduction

The report examines one of the approaches to the discrimination of the three-dimensional objects located on the underlying surface. The concretization of the problem within the framework of the problem lies in the fact that distinguishing objects are defined by a series of their shadows. The evolution of modern three-dimensional modeling technology stimulates the design and construction of multi-functional machine vision systems for three-dimensional imaging of objects located in the free space and the underlying surface. The distinction between these objects is important for remote sensing space systems, airborne survey of radio and laser-radar and optical earth observation systems installed on the aircraft.

For distinguishing of object in terms of base problem solving approaches [1, 2] it is necessary to have the normalized mathematical models of an etalon objects $\mathbf{v}_m$, $m = 0,1,..., M-1$, and a distinguishing object W, the measure apparatus and the decision-making rules based on their comparison. Although etalon objects $\mathbf{v}_m$ are likely to be available to the developer, but their three-dimensional nature creates obstacles to get adequate and at the same time simple and resistant to noise mathematical models.

Distinguishing object $\mathbf{W}$ the distance is usually much greater than the size of the base which can be implemented with the hardware platform. Therefore, the rays reflected or modulated by an object and received by a space-distributed sensors that have small divergence angle, which leads to a weak stereo effect. But even in those cases when it is possible to obtain detailed three-dimensional image of the object $\mathbf{W}$, the part of its surface is outside the field of view sensors. Therefore, to obtain an adequate mathematical model of the object $\mathbf{W}$ must have a series of its projections $\mathbf{w}_n$, $n = 0,1,...,$ by which one can calculate the coordinates of the points located on its surface.

Distinguishing object $\mathbf{W}$ is formed by the transformation of one of the etalon objects $\mathbf{v}_m$ from alphabet $A = \{A_m\}_{m=0}^{M-1}$. All vectors of the etalon objects have the dimension S. For specificity we consider such an object as object $\mathbf{w}_l$ of a class A_l and we call it a prototype. Transformations of prototype $\mathbf{w}_l$ are the rotation angle ψ around the axis with the direction vector $\boldsymbol{\rho}$, the cyclic shift by the value d, $d = 0,1,..,s-1$, of components its vector, the scale and the noise masking. Parameters l, ψ, $\boldsymbol{\rho}$ and d are considered unknown. Apparatus for determining the similarity measure of the signals obtained on the basis of mathematical models $\mathbf{D}_\mathbf{W}$ and $\mathbf{D}_m$ spatial objects $\mathbf{W}$ and $\mathbf{v}_m$ is based on the calculation of the functional in the form of

a normalized scalar product (NSP) of the vectors $\mathbf{D}_{\mathbf{W}}$ and $\mathbf{D}_m$, $m = 0,1,..., M-1$. *The goal of the paper is to provide an optimal maximum likelihood algorithm for distinguishing the object* $\mathbf{W}$ *from its projections* $\mathbf{W}_n$, $n = 0,1,...$, *under a priori uncertainty of parameters* l, ψ, $\boldsymbol{\rho}$ *and* d. Feature of the algorithm is that the projections of the object $\mathbf{W}$ are the images of his shadows.

Formation of mathematical models of the distinguishing objects

Shadow image of the object is characterized by a constant value of brightness. The samples of the brightness of the shadow points have strong internal correlation and so have little information. Useful information about the shape of the object $\mathbf{W}$ is contained only in the shadow of the contour. The transition from the shadows to the illuminated backdrop of the underlying surface is accompanied by a jump in brightness. Therefore, the contour shadow, seen as a signal characterized by a high signal/noise ratio. On the contours of the two shadows obtained at different angles of projection are conjugate points, allowing to calculate the coordinate of a point on the surface of the object $\mathbf{W}$ in the region of large values of the gradient, i.e., points of its contour. Therefore, the form of shadows fragment contains information about the form the object.

As an adequate mathematical model of three-dimensional object is used considered in [3] wired model representing a three-dimensional polygonal line passing through all the vertices of the object. Sequence of point vertices in contour which are identified with the object **W** of the polyhedron **X** in the wire model is defined by its structure and therefore, the model retains all of the information about the object during its rotation and scaling. The elements of the model are the vector quaternions defining vectors connecting submitted to it the adjacent vertices.

The edges of distinguishable polyhedron object are formed on the basis of the properties of incidence.

The rules of construction wire model:

R1. The numbering starts with the point of the reference face $\mathbf{G}_0$ when traversing its path in a **counter-clockwise direction**, starting from the point $\mathbf{a}_t = \mathbf{a}^{(0)}$:

$$\mathbf{a}_t^{(0)},\ \mathbf{a}_{t+1}^{(1)},\ \ldots,\ \mathbf{a}_{t+l_0-1}^{(l_0-1)},$$

where $\mathbf{a}_t$ – point of face which are select as starting point $\mathbf{a}^{(0)}$ of 3D contour $\mathbf{D}$ of object, l_0 – count of points on face $\mathbf{G}_0$.

R2. Continuous bypassing *upper* edges of the object is performed in a **clockwise direction**. It begins at the point $\mathbf{a}_{t+l_0-1}^{(l_0-1)}$ and terminates at a point $\mathbf{a}_t = \mathbf{a}^{(0)}$. At this time a passing vertices are sequentially numbered.

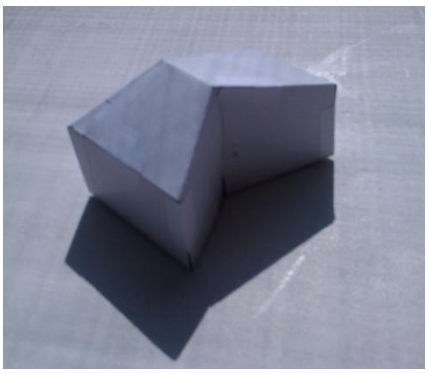
R3. When crawling prohibited deadlock associated with the transition to the points of face $\mathbf{G}_0$ at a time when the top of the object vertices are unnumbered.

R4. Contour line $\mathbf{D}$ obtained must not contain nodes (branches), i.e. it is the only way through all the vertices of the object.

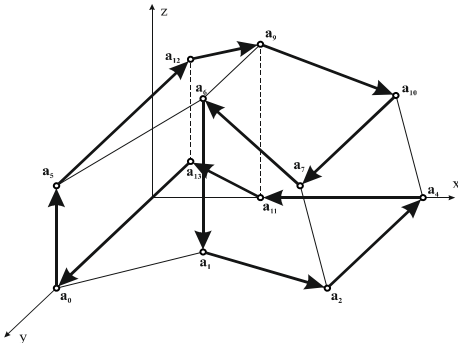
R5. 3D contour lines which are passed through the vertices of the *upper* part of the object closes only at point $\mathbf{a}_t^{(0)}$ of 3D object contour.

Example of wired model shown in Fig 1.

In actual practice the restored object due to the action of coordinate noise differs from the etalon. Fig. 2 shows an image of a polyhedron $a_0a_1a_2a_3$ without noise and polyhedron $b_0b_1b_2b_3$ which are obtained from the first subject by the noising. With centers at the vertices a_0, a_1, a_2 and a_3 the spherical strobes with radius 3σ are constructed. Due to the action of noise vertices of the first polyhedron can be located anywhere in the spherical strobe. If to connect vertices b_0, b_1, b_2 and b_3 accordingly, the resulting polyhedron may differ substantially from the polyhedron $a_0a_1a_2a_3$ (Fig. 2).



a)



b)

Fig. 1. Object **W** (a) and it wired model **D** (b)

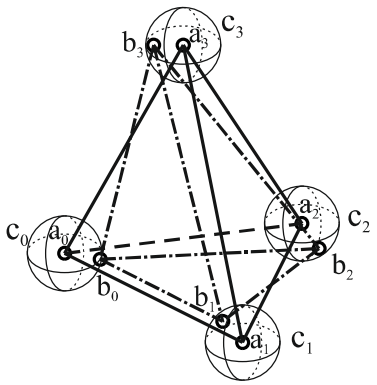


Fig.2. The original $a_0a_1a_2a_3$ and noised $b_0b_1b_2b_3$ polyhedrons

Ordering principle of the noisy vertices of polyhedron is based on the use of a priori information about the location of the vertices of the etalon objects. In general, the mathematical model of distinguishable deterministic 3D object $\mathbf{W} = \mathbf{V}_l + \mathbf{Z}$, $l = 0, 1, 2,... M-1$, is the sum of the prototype (etalon) $\mathbf{V}_l$ and noise spatial object $\mathbf{Z}$ with the same dimensions s. The object $\mathbf{Z}$ consists of a sequence of random 3D vectors $\mathbf{z}(n) = z_1(n)\mathbf{i} + z_2(n)\mathbf{j} + z_3(n)\mathbf{k}$. Components $z_1(n)$, $z_2(n)$ and $z_3(n)$ are independent, centered, normally distributed random variables with variance σ^2.

The device ordering of vertices of the noisy distinguishable object **W** is realized on the basis of 3D selector. The algorithm is the same as the algorithm of the system of second-order radar data processing in the mode selection noised targets in the mathematical 3D strobe[1]. For ordering of vertices distinguishable object **W** on the assumption that it belongs to the class A_m of alphabet , in 3D space the s spherical mathematical strobes with radius 2σ are constructed, center coordinates of which are defined by vertices of the etalon object $\mathbf{V}_m$, $m = 0,...M-1$.

Vector signal as coordinate points of vertices of the distinguishable polyhedral object **W** is transferred to input of *M* 3D-input selectors, each of which is based on the one etalon object $\mathbf{V}_m$, $m = 0,...M-1$. In the *n*-th window of *m*-th selector as the *n*-th point vertices of the object **W**, on the assumption that it belongs to the class A_m, are selected nearest to the center point. The result is a *M* vector signals, each of which is constructed under the assumption that the order of its components corresponding to the etalon $\mathbf{V}_m$. The conventional wired models $\mathbf{D}_{\mathbf{W},m}$ of the distinguish object **W** are constructed on them, $m = 0,...M-1$. Next, the values of the measures of similarity vectors $\mathbf{D}_{\mathbf{W},m}$ and $\mathbf{V}_m$ are calculated , and distinguish objects belong to a class, which corresponded to the maximum value of the measure.

Structure of device for distinction

Device for discrimination of noised object with polyhedral shape composed of *M* parallel channels, each of which contains 3D selector, the constructor of wired model and the computer the similarity measure. The output signals of channels received by a decision unit (Fig. 3).

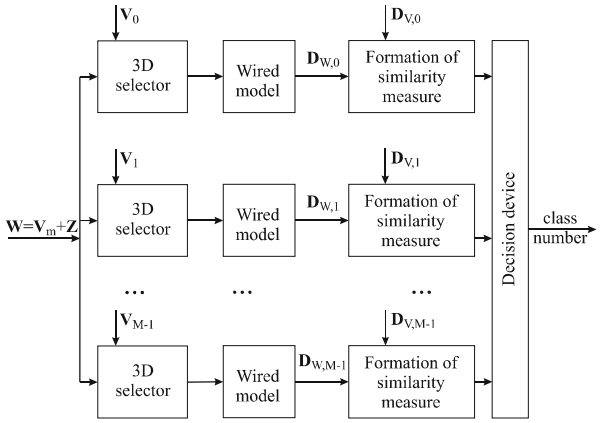


Fig. 3. Structure of device for distinction of reconstructed object

At small deformations of the reconstructed object $\mathbf{W}$, achieved at a high signal / noise ratio, the wired model is constructed on the rules above without the use of 3D selector. In this case, in accordance with the algorithm described in [5], the angular alignment may distinguish objects with each of the reference objects and the coordination number of the component of a vector signal. Upon reaching such agreements is possible to distinguish the reconstructed object with a priori unknown parameters ψ and d values.

References

1. V. A. Vasin, I. B. Vlasov, Yu. M. Egorov, et al., Information Technologies in Radiotechnical Systems: Tutorial, 2nd rev. and enlarged ed. (Izdat. MGTU im. N. E. Baumanna, Moscow, 2004) [in Russian, ed. by I. B. Fedorov]

2. Y.A. Furman, R.V. Eruslanov, I.L. Egoshina. Reconstructing the coordinates of surface points of an object with a set of images of its shadow / Optoelectronics, Instrumentation and Data Processing, 2011, Volume 47, Issue 6, pp 537-546. [Russian in Avtometriya, 2011, Vol. 47, No. 6, pp. 16–27.]

3. Y.A. Furman, K.B. Ryabinin, M.I. Krasilnikov. A wire model of a spatial group point object / Optoelectronics, Instrumentation and Data Processing, 2008, Volume 44, Issue 3, pp 189-199 [Russian in Avtometriya, 2008, Vol. 44, No. 3, pp. 3–16.]

4. Furman, Ya. A. Reconstruction of Images and Recognition of Polyhedral Objects / Ya. A. Furman, R. V. Eruslanov, I. L. Egoshina // Pattern Recognition and Image Analysis, 2012. – Vol. 22. – No. 1. – P. 196-209.

5. Furman, Ya. A. Matching the angular parameters of quaternion signals // Proektirovanie i tehnologija jelektronnyh sredstv. – 2011. – №1. – P. 12-15. [in Russian].

CONSTRUCTION OF ADAPTIVE IDENTIFICATION ALGORITHMS, USING THE ESTIMATES CONFORMITY PRINCIPLE[1]

V. Fursov[2,3]

[2] Image Processing Systems Institute of Russian Academy of Science
151, Molodogvardeyskaya st., 443110 Samara, Russia, fursov@smr.ru
[3] Samara State Aerospace University,
34, Moskovskoye shosse, 443086 Samara, Russia, fursov@ssau.ru

The paper deals with the development of estimation algorithms using a small number of observations. We discuss the feasibility of solving such problems on the basis of the estimates conformity principle. In particular, the approach based on direct exhaustive search of the data is developed. This involves a more realistic formulation of the task as compared with that adopted in the theory of statistical estimation, because in this case the probability models are not reliable.

Introduction

There are at least two approaches to solving the problems of identification [1]: a statistical one and engineering one. In the statistical approach, the problem of parameters estimating is considered as the problem of the distribution parameters determining. Engineering approach is characterized by a variety of empirical algorithms.

Often this problem should be solved by the identification from an extremely small number of measurements. In this case, the theory of statistical estimation is not suitable enough [2], and it is more appropriate to use an adaptive approach [2, 3]. The approach uses a multi-stage process, in which the estimates of the each stage are refined using the information obtained at the previous stages.

Statement of the problem of parametric identification

Generally the problem of the parametric identification is formulated as the problem of finding the most correct estimate $\hat{\mathrm{F}}$ of the operator $\boldsymbol{F}$ in equation [1]:

$$\mathbf{y} = \mathrm{F}\mathbf{x}, \qquad (1)$$

where $\mathbf{x}$ and $\mathbf{y}$ are scalar or vector input and output of an object. In the case of a linear model with one output, the equation (1) is written as [1]

$$\mathrm{y} = \mathbf{c}^{\mathrm{T}}\mathbf{x} + \varepsilon, \quad (2)$$

where y is an output object (scalar), distorted by random noise of measurement ε; $\mathbf{c}$ is $M\times1$-vector of unknown parameters; $\mathbf{x}$ is $Mx1$-vector obtained from measurements or known a priori.

If vector $\mathbf{c}$ can be considered constant for the time needed for N observations, the matrix equation can be written with the use of (2):

$$\mathbf{y} = \mathbf{X}\mathbf{c} + \boldsymbol{\xi}, \quad (3)$$

where $\mathbf{y} = \left[\mathrm{y}_1, \mathrm{y}_2, ...\mathrm{y}_N\right]^{\mathrm{T}}$, $\mathbf{X} = \left[\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_N\right]^{\mathrm{T}}$, $\boldsymbol{\xi} = \left[\varepsilon_1, \varepsilon_2, ...\varepsilon_N\right]^{\mathrm{T}}$.

The problem of identification is to determine the estimate $\hat{\mathbf{c}}$ of the $M\times1$-vector parameter $\mathbf{c}$ in the equation (3) using the observed NxM-matrix $\mathbf{X}$ and $Nx1$-vector $\mathbf{y}$ $(N>M)$, with the $Nx1$-vector of errors $\boldsymbol{\xi}$ being unknown.

Thus, the problem is substantially reduced to solving the overdetermined system of equations (3). Therefore, methods for the examination of errors, covered in this paper, are similar to those which are usually considered in the algebraic theory of perturbations [4]. In particular, considerable attention is paid to influence of the system conditioning on the accuracy of estimates (3). If the source data is generated during the stable

[1]This work was financially supported by the Samara Region Administration and the RFBR (grants № 12-07-00581-a, № 12-07-31208).

state of a dynamic system, where the processes are changing slowly, the problem is ill-conditioned [4, 5]. Therefore the accuracy of the estimates using a small number of observations should be determined taking into account the combined effect of all the factors: errors of the input data, computational errors and the sensitivity to these errors, which depend on the conditioning [3].

It is known that the estimate of the least squares method (LSM)

$$\hat{\mathbf{c}} = \left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} \qquad (4)$$

is unbiased and efficient if the standard statistical assumptions [6, 7] are satisfied. Unfortunately, the statistical justification of the LSM with small number of observations is not appropriate to the content of the problem. Estimates of the mean, variance and conditioning values, which are calculated on a small set of data, can significantly differ from those in other data sets or the statistical population, from which this data is selected.

The main assumptions

In this paper, the assumptions are much more free from a priori probability models that are traditionally used:

1. The matrix **X** and the vector **y** are fixed, i.e. $x_{i,j}, y_i$, $i = \overline{1,N}$, $j = \overline{1,M}$ are known from measurements on the single realization.

2. There is an indefiniteness of the conditioning properties of the matrix **X** and the statistical characteristics of the vector $\boldsymbol{\xi}$.

3. The norm of the error vector $\boldsymbol{\xi} = \left[\xi_1, \xi_2, ..., \xi_N\right]^{\mathrm{T}}$ is limited by: $\|\boldsymbol{\xi}\| \le R_{\xi}$. (5)

4. In the system (3) the most noise-free subsystem is contained, for which the estimates $\hat{\mathbf{c}}$ can be constructed with a desired accuracy**.**

5. There is an exact model of a system corresponding to the Eq. (6):

$$\mathbf{y}^{*} = \mathbf{X}\mathbf{c}, \qquad (6)$$

where $\mathbf{y}^{*} = \mathbf{y} - \boldsymbol{\xi}$,

Thus, we reject the assumptions of the classical regression [6]: $M\{\boldsymbol{\xi}\}=0$ and $\mathrm{cov}\{\boldsymbol{\xi}\} = \sigma^2\mathbf{E}_N$. The rejection of these assumptions is forced, but the stated assumptions are more realistic. In this case, we use LSM as a method for solution of an overdetermined system of equations $\mathbf{X}\hat{\mathbf{c}} = \mathbf{y}$:

$$\hat{\mathbf{c}} = \mathbf{X}^{+}\mathbf{y}, \qquad (7)$$

where $\mathbf{X}^{+} = \left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\mathbf{X}^{\mathrm{T}}$ is the pseudoinverse for **X**. If the solution (7) is obtained, we can write the equation for the residual in accordance with (3) as follows:

$$\hat{\boldsymbol{\xi}} = \mathbf{y} - \mathbf{X}\hat{\mathbf{c}} = \mathbf{y} - \hat{\mathbf{y}}. \qquad (8)$$

The conforming estimate method leans on the assumption 4. In this case the problem consists in searching for the most noise-free subsystem. As an indicator of this subsystem it is offered to use a criterion of mutual closeness of estimates, which are received on subsystems of small dimensionality.

Description of the conforming estimate method

In accordance with the above principle the algorithm of the conforming estimates construction is presented as follows.

The set of so-called lower-level subsystems is formed from the initial system (3), using the rectangular matrixes:

$$\mathbf{y}_k = \mathbf{X}_k\mathbf{c}_k + \boldsymbol{\xi}_k,\ k=1,2,\ldots, \qquad (9)$$

where $\mathbf{y}_k = \mathbf{G}_k\mathbf{y}$, $\mathbf{X}_k = \mathbf{G}_k\mathbf{X}$, $\boldsymbol{\xi}_k = \mathbf{G}_k\mathbf{y}$, $\dim(\mathrm{R}(\mathbf{X}_k)) = \mathrm{rank}\mathbf{G}_k = S_k$.

Elements of the matrixes $\mathbf{G}_k$ can be only 0's and 1's, and each row consists entirely of zeroes except for a single one. Nonzero elements in S_k rows of the matrixes $\mathbf{G}_k$, k=1,2,… are specified so that "to cut out" the subsystems (9) from the initial system (3) with different combinations of rows. If the dimension of the lower-level subsystems is fixed, i.e. $S_k = S \ge M$, the number of the lower-level subsystems is equal to C_N^S (S<N).

By calculating an LSM-estimate for each of these subsystems

$$\hat{\mathbf{c}}_k = [\mathbf{X}_k^{\mathrm{T}}\mathbf{X}_k]^{-1}\mathbf{X}_k^{\mathrm{T}}\mathbf{y}_k,,$$

we obtain a set of all possible estimates on the lower-level subsystems of dimension S:

$$\Xi = \{\hat{\mathbf{c}}_k = [\mathbf{X}_k^{\mathrm{T}}\mathbf{X}_k]^{-1}\mathbf{X}_k^{\mathrm{T}}\mathbf{y}_k \ \ | |\Xi| = C_N^S\}.$$

A set of rectangular $P\times N$ weight matrixes $\mathbf{H}_l$, $l = \overline{1,L}$, $L = C_N^P$, $rank\mathbf{H}_l = P$, (S<P<N) is constructed in a similar way (using 0's and

1's). With the use of these matrixes, the so-called upper-level subsystems are constructed:

$$\tilde{\mathbf{y}}_l = \tilde{\mathbf{X}}_l \mathbf{c}_l + \tilde{\boldsymbol{\xi}}_l, \quad (10)$$

where $\tilde{\mathbf{X}}_l = \mathbf{H}_l \mathbf{X}$, $\tilde{\mathbf{y}}_l = \mathbf{H}_l \mathbf{y}$, $\tilde{\boldsymbol{\xi}}_l = \mathbf{H}_l \mathbf{y}$, $l = \overline{1, L}$, $L = C_N^P$.

It is assumed that for all $k = \overline{1, K_l}$ and $l = \overline{1, L}$ $rank\, \mathbf{X}_{l,i} = M$, i.e. all the lower-level subsystems have a full rank, though conditionality of the matrixes for different l and k may vary.

Each upper-level subsystem has a subset of the lower-level subsystems and a subset of preliminary estimates corresponding to it:

$$\Theta_l = \{\hat{\mathbf{c}}_k \in \Xi \mid \forall k : k = \overline{1, K}, \quad K = |\Theta_l| = C_P^S\},$$

$$l = \overline{1, C_N^P}.$$

To characterize these sets Θ_l, the criterion of mutual closeness is proposed:

$$W(\Theta_l) = W(\hat{c}_{l,1}, ..., \hat{c}_{l,K}), \qquad c_{l,k} \in \Theta_l, \quad k = \overline{1, K}, \quad l = \overline{1, L}. \qquad (11)$$

Let us refer to the set Θ_l with minimal value $W(\Theta_l)$ as the most conformed set of estimates. Finding the most conformed set of estimates actually reduces to searching for an index $\hat{l}$:

$$W(\hat{l}) = \min_l W(\Theta_l). \quad (12)$$

To solve this problem, the criterion of mutual pairwise closeness is further used:

$$W\left[\Theta_l\right] = \sum_{i,j=1}^{K} \left(\hat{\mathbf{c}}_{l,i} - \hat{\mathbf{c}}_{l,j}\right)^2. \quad (13)$$

There are some of the results of the conformity identification, connected with the splitting initial system (3) on subsystems, described below.

Outcome of the experiments

Another example of finding the solution using the criterion (13) discussed in [2] is given below. The initial data was formed using a random number generator. The problem dimension is *N*=10, *M*=4. The most conformed estimates were selected from C_{10}^7=120 samples of the upper-level subsystems, with 7 observations taken for every subsystem. Then, for each upper-level subsystem and for C_7^4=35 lower-level subsystems (4x4), a set of LSM-estimates and the subsequent mutual closeness function (16) were calculated. The resulting set of values of the conformity criterion was arranged in ascending order (Fig. 1). For comparison, the criteria of estimation quality $\|\Delta c\|$ were arranged in the same order (Fig. 2). In Fig. 2, the horizontal line indicates the values of $\|\Delta c\|$ for the LSM-estimate calculated over the entire set of data.

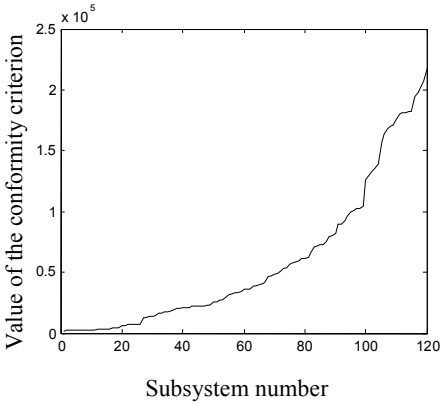


Fig. 1. Mutual closeness function

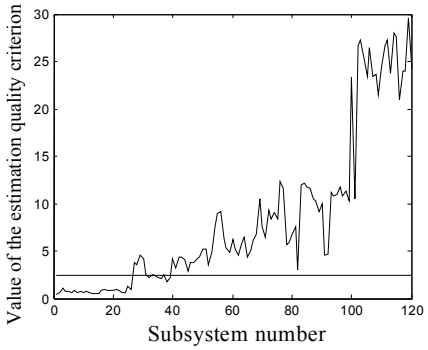


Fig 2. Estimates quality

Comparison of the conformed identification method and the RANSAC algorithm was conducted in order to compare their accuracy in the same operating conditions [8]. The experiments were carried out on the data sets that were modeled as follows.

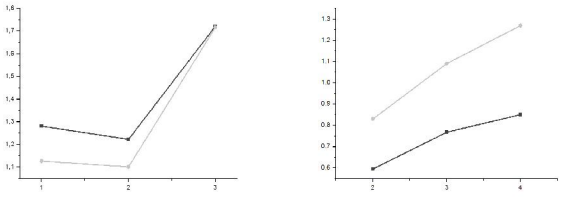
In various systems (2) M = 8 and N = 12 or N = 16. The components of the parameter vector **c** were specified as uniformly distributed random numbers in the range of 1 to 10. The elements of matrix X were calculated in accordance with (2) the corresponding points coordinates, which are modeled as random sequences with given variances.

Components of the error vector were formed in such a manner that the normal error signal-to-noise ratio was in the range of 40-60 dB. For rough errors signal-to-noise ratio was specified in the range of 0-10 dB. Also, the frequency of the

anomalous error was specified 50-60% of the number of degrees of freedom (N-M) of the resulting system.

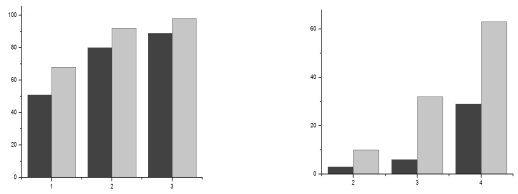
While comparing the accuracy and reliability of the algorithm, the identification is assumed to be correct if the ratio of the error vector norm to the parameter vector norm does not exceed 0.3.

Figure 3 shows a plot of the mean values of the error on the number of correct identifications of anomalous errors for different values of the number of estimated parameters (light - for the method of RANSAC, dark - for the conformed identification method). Figure 4 shows graphs illustrating the number of false identifications to 100 experiments on the number of rough errors for the same values of the estimated parameters. It is obvious that in all implementations, the conformed identification method shows better results both in reliability and accuracy.



a) b)

Fig. 3. Mean error of the correct identifications: a) N=12, M=8; b) N=16, M=8



a) b)

Fig. 4. The number of false identifications for 100 experiments: a) N=12, M=8; b) N=16, M=8.

Conclusion

The given examples prove a principle feasibility of constructing of estimates, using a criterion, other than that based on the statistical stability. This fact is of great significance, for it contains not only a traditional criticism (which, as a rule, causes no objections) of small-samples-based estimating methods but makes it possible to solve the existing problems.

To choose a suitable correction vector or to identify the most noise-free subsystems, which ensure the construction of a sufficiently exact solution of the problem formulated, an exhaustive search of a great number of subsystem samples (both upper- and lower-level) may be required. This is an unavoidable fee for the deficit of a priori information, resulting from the insufficient number of observations. Feasibility to employ simple (and, hence, relatively cheap) statistical processing schemes results from significant efforts connected with the conducting of a great number of observations, that are needed to meet the law of large numbers.

High computational complexity of the methods based on the estimate conformity principle of course causes dissatisfaction. Nonetheless, the use of this approach is justified when it is required to obtain an exact solution despite the fact that the number of observations is small and conducting additional experiments is, for some reason, impossible.

References

1. Eikhoff, P. The basics of control system identification. Moscow: “Mir” Publisher, 1975.
2. Fursov V.A. Estimates Conformity Principle in the Problems of Identification. Computional Science – ICCS 2003. International Conference Melborne, Australia and St.Petersburg, Russia. June 2003, Proceedings, Part II, p. 463-470.
3. Fursov, Vladimir A. Constructing unified identification algorithms using a small number of observations for adaptive control and navigation systems. Journal: Proc. SPIE Vol. 3087, p. 34-44, Scott A. Speigle; Ed., 1997.
4. Charls L. Lawson, Richard J. Hanson, Solving Least Squares Problems. Prentice – Hall, Inc., Englenood Cliffs, N.J., 1974.
5. Bjorck Ake. Least Squares Methods. Elsevier Science Publishers B.V. (North Holland). 1990.
6. G.A. Seber. Linear regression analysis. John Wiley and Sons. New-York. London. Sydney. Toronto. 1977.
7. P.E. Kalman, Noised systems identification. Advances of Mathematical sciences, v. 40, issue 4(244), 1985.
8. Fursov V. Conformed Identification of the Fundamental Matrix in the Problem of a Scene Reconstruction, using Stereo Images / V. Fursov, Ye. Goshin // Image Mining. Theory and Applications. Proceedings of IMTA-4 2013. – 2013.– pp. 29-37.
9. S.A. Bibikov, V.A. Fursov, A.V. Nikonorov. Desktop supercomputing technology for shadow correction of color images. Proceedings of the International Conference on «Signal Processing and Multimedia Applications» (SIGMAP 2010) (26 – 28 July 2010, Athens, Greece). SciTePress, 2010, p. 124-129.

IMPROVEMENT OF PATTERN RECOGNITION CONCERNED MODELLING THROUGH A MULTIMODAL APPROACH TO KNOWLEDGE PRESENTATION[1]

R. Guadagnin[2]

[2] Catholic University of Brasília, MS Program on Information Technology and Knowledge Management, Catholic University of Brasília, SGAN 916, Módulo B, 70790-160 Brasília, DF, Brazil, renatov@ucb.br

The real world is a repository of objects. Pattern recognition widely works in real world, as it performs a modeling process for providing data, information and knowledge from such objects. It is an instrument of elicitation of reality. Effective decision-making does require good quality knowledge. For a broader use of knowledge generated by Pattern Recognition, this work advocates an extended knowledge focused on human capacity for perception. It considers multimodal display of knowledge to practical feasibility of knowledge presentation adapted to human perception towards decision improvement and inclusion of perception-impaired people.

Introduction

Human beings require the availability of information to support decisions since ancient times. It means value to the decision maker, because it improves the probability to achieve expected results. Often information is not explicit or enough clear for the needs of human beings. Pattern recognition contributes to minimize such shortcoming as a mean of modeling the real world and a means of elicitation of reality.

Both the human and other less skilled beings set languages for information expression through broad understandable signs. Pattern recognition is an instrument for both synthesizing and detailing the original raw data by particular symbols.

Knowledge is an expression of processes eventually described by large volumes of data and information. Knowledge is a value aggregator to decision making. The pressure for good deciding in a short time implies the need for provision of not much information but good quality knowledge. Its usefulness depends primarily on its perception by decision makers. Hence, the need for an adequate knowledge adjusted to the perception features of decision makers emerges. In such context it is remarkable that several systems for text-speech information transformation, such as for car drivers, are available [1].

This paper argues for an extended approach in knowledge exposure, say a multimodal display of knowledge, which considers the practical feasibility of knowledge presentation adapted for vision, hearing, smell, touch and taste. This approach improves the quality of usable knowledge for decision and enables greater social inclusion of perception-impaired people for decision making in a broad sense.

From data to knowledge

Concrete or abstract objects need to be modeled according to a symbology and be contextualized according to ontology for use by both information technology and communication. Shortly the main motivation for this model is of an economic nature, as it represents a reduction in time and space. Data are models in the most rudimentary level, which can be aggregated according to certain criteria, generating new models called information. This new kind of model for object requires a specific purpose for the decision that they will support.

[1] This paper is part of the cooperation between Catholic University of Brasília and German Office for Academic Interchange (DAAD).

A new way of modeling, whose inputs are information refers to knowledge. This model describes knowledge processes, relationships between processes, spatial relationships and temporal trends. Such modeling can be objective or subjective. In the first case, it is possible to use quantitative variables according to a certain widely recognizable symbology. The subjective knowledge requires the adoption of a specific symbology not necessarily widely acceptable. The quality of such symbology is highly relevant to the reduction of loss of knowledge inherent in any modeling process.
Pattern Recognition can generate models of reality in the form of data, information and knowledge.

The cycle of knowledge

The real world is a repository of objects on which knowledge can be generated. This activity is part of the Pattern Recognition. Due to its modification over time, the real world requires ongoing procedures of Pattern Recognition to generate usable knowledge. Human beings use knowledge in making decisions and in performing various procedures. Derived actions can modify the real world. There may be a specific form of knowledge generation from knowledge acquired by humans, so-called insights on new interpretations or ways of solving problems, which are essential for innovation. It is thus the cyclical process of knowledge, as illustrated by Fig. 1.
The actions performed by humans are the result of his perception based on what Pattern Recognition provides. Because of the practical impossibility of having a precise notion of real-world modeling concerning each individual, perception of the real world is actually an abstraction of objects subject to individual factors.

Objective of Pattern Recognition

The real world presents numerous objects with specific features not always available. The instruments of Pattern Recognition look for assessing these features and delivering them to the decision maker. It is a model generator, according to a certain symbology, to meet the needs of decision context. Its instruments can also extract new information from unarticulated information.
The models produced by Pattern Recognition are acceptable, since its intrinsic uncertainty is not above an acceptable limit. Uncertainty acceptability depends upon loses that can be caused by failure.

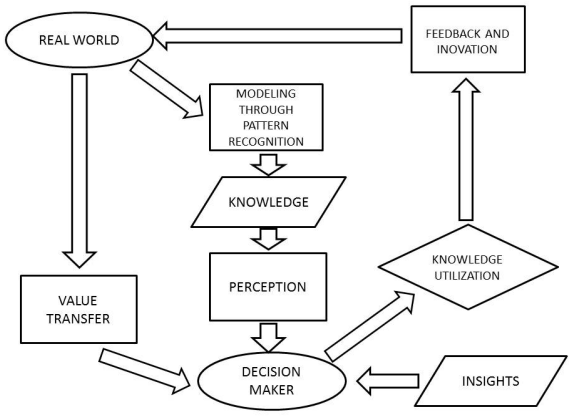


Fig. 1. Real world, knowledge and decision maker

Pattern Recognition as reality modeling

The presently existing large amount of data, information and knowledge cannot be fully utilized. It is an imperative the use of modeling tools. The model acts as an information and knowledge quantity reducer for decision-making and for supporting the implementation of various processes. The quality of this modeling can be optimized in a process of trial and error, with continuous generation and validation of knowledge. The generation of knowledge is vulnerable to distortions arising from specific interests. Validation is important to minimize these distortions.
Like any process modeling, Pattern Recognition implies loss of information. Ratios, trends and relationships between data, which represent a subset from real world can be inferred and presented in a visible way [2]. Efforts to build intelligent agents include learning and use of heuristics and prior knowledge. This way one achieves the best performance, eliminating repetitive work and speeding up the generation of knowledge. A review of prior knowledge by considering the characteristics of the new reality are required

and at the same time risky due to the possible introduction of bias.
A continuous development of new techniques for pattern recognition allows the reduction of uncertainty relevant to knowledge generated by them. BACIC and RAYMOND suggest reducing the uncertainty related to information as a key factor for the implementation of information visualization in the strategic decision making [3]. Automated devices based on intelligent agents, such as automatic vehicles driving, will have wide use only after reaching an acceptable level of accuracy for practical use.

The need for knowledge transforming

Knowledge becomes usable when it becomes appropriate to the capacity of perception of its user. One can understand knowledge user as a human being that wants to succeed through his actions, that is, to extract some value from the result from his decisions.
Obtaining knowledge implies costs that would not be funded if the user could not be subsequently reimbursed by the value provided by his actions. Once the user acquires knowledge to perform activities that return value, he often has no interest in transforming his implicit knowledge to explicit knowledge, due to the risk of later losing of values derivable from its activities.
Knowledge is absorbed by the user at the time of their presentation to his perceptors, according to a specific symbology. So, images, sounds, smells, tastes and tactile impressions can be used for presentation of knowledge from the pattern recognition.
Due to peculiarities of the knowledge domain and the characteristics of the user, it is necessary to develop mechanisms for processing their presentation and portability mechanisms. PISSALOUX et alli developed a visual-tactile interface (intelligent glasses) for providing access to information to visually impaired [4]. YU et alli proposes a walking assistance by transforming images to sounds also concerning visually impaired [5].

A multimodal approach

Visualization, as a way of expression, opens up wide possibilities for condensing the vast amount of knowledge that arises from real world modeling, to a level compatible with the capacity of human perception. Visualization has proven to be an effective strategy for supporting users in coping with complexity in knowledge- and information-rich scenarios [6]. In this context, the system RuleViz enables visualization of knowledge discovery and data mining [7]. EPPLER presents Visualization through the so-called knowledge maps [8]. In addition some disadvantages listed by this author are illegitimate use, possible misinterpretation, information overload, He suggests Visualization oriented preventive actions.
Visualization restricts the absorption of knowledge to the visual ability of the user, requiring the modeling of knowledge in the form of images. Indeed the observation of the real world requires a wide variety of signs, which go beyond the human sensory capacity, e.g., infrared signal, radar waves, ultrasound, micro-oscillations, odorless and tasteless substances.
Modeling the Pattern Recognition can adequate the generated knowledge to human perception, such as magnetic resonance imaging or audible alarms related to equipment operation or local air conditions. Visualization and sonification of brain electrical activity are produced by the experiment performed by JOVANOV et alli [9]. Therefore, it use special interfaces for perceptible signal.
Except in specific cases, the user is subject to the way knowledge is generated. The user cannot, for example, know the result of a blood test in a hearing, get the instructions for troubleshooting a computer problem in a hearing, get a picture explaining the sequence of textual instructions for powering a device.
Exposure of knowledge depends primarily on expectations of its user, which is the quality standard reference. Thus generating a perceptible knowledge in different ways values increases thereto and can minimize the disadvantages of visualization mentioned above. The multimodal perception improves understanding of complex phenomena [9].
There is need to establish protocols equivalence between these forms of expression of knowledge, in order to link knowledge to a

wide portability acceptance. Fig. 2 expands part of Fig. 1.

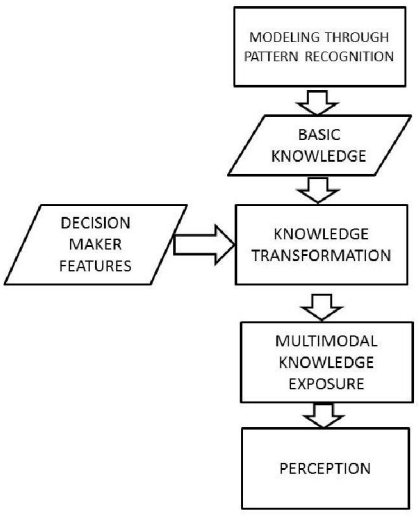


Fig. 2. Multimodal knowledge presentation

Key factors concerning Visualization can also help evaluation of a multimodal instrument. They are:

- perceived usability: the degree in which a person believes that using a particular system would be free of effort;
- perceived usefulness: the degree in which a person believes that using a particular system would enhance his or her job performance, and
- perceived authority: the degree in which a person is confident that using a particular system is a good choice [10].

Conclusion

The here proposed expansion of knowledge forms of presentation involves answers to specific questions, as below.

The loss of information inherent in modeling processes occurs both in the sampling process as in the process of adopting a symbology for their presentation in different ways. What are the possibilities for automatic linking of acceptable criteria for Pattern Recognition requirements of the use of knowledge?

Presentation of knowledge is linked to a process of discretization. Each human receptor has specific characteristics regarding the amount of information according to time and space. To what extent does flexibilization of presentation of knowledge from certain signals to other signals compromise its value?

How can insertion of perception impaired be expanded through multimodal knowledge presentation?

The increasing dissemination of knowledge in various areas runs in direction to a free knowledge. Due to increased portability and assess to the knowledge generated by the Pattern Recognition, does multimodal knowledge presentation effectively contributes to free knowledge expansion?

References

1. Fischer, P., Nurnberger, A. Adaptive and multimodal interaction in the vehicle, IEEE International Conference on Systems, Man and Cybernetics 2008, pp. 1512-1516.
2. YAO, Nathan. The FlowingData Guide for Design, Visualization, and Statistics, USA: Wiley, 2012.
3. Bacic, Dinko, Raymond, Henry. The Role of Business Information in Knowledge Creation, in: Proceedings of the 8th Conference on Information Systems, Washington, USA, August 9-12, 2012.
4. Pissaloux, E., Velazquez, R., Maingreaud, F.. Intelligent Glasses: a Multimodal Interface for Data Communication for the visually impaired, IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, Seoul, Korea, 2008, pp. 120-124.
5. Yu, J., Chung, H.I., Hahn, H. Walking assistance system for sight impaired people based on a multimodal information transformation technique, ICROS-SICE International Joint Conference 2009, Fukuoka, Japan, pp. 1639-1643.
6. Keller, Tanja, Tergan, Sigmar-Olaf. Visualizing Knowledge and Information. An Introduction, LNCS 3426, pp. 1-23, 2005, Berlin Heidelberg.
7. Han, Jianchao, Cercone, Nick. RuleViz: A Model for Visualizing Knowledge Discovery Process, in: ACM, USA, 2000, 244-253.
8. Eppler, Martin J. Making Knowledge Visible through Intranet Knowledge Maps: Concepts, Elements, Cases, Proceedings of the 34th Hawaii International Conference on System Sciences, 2001.
9. Jovanov, Emil, Starcevic, Dusan, Radivojevic, Vlado, Samardzic, Aleksandar, Simeunovic, Vladimir. Perceptualization of Biomedical Data. An Experimental Environment for Visualization and Sonification of Brain Electrical Activity, IEEE Engineering in Medicine and Biology, January/February, 1999, pp. 50-55.
10. Bresciani, Sabrina, EPPLER, Martin J. Beyond Knowledge visualization usability: toward a better understanding of business diagram adoption, 13th International Conference Information Visualization 2009, pp. 474-479.

THE CHALLENGES, THE PROBLEMS AND THE TASKS OF THE DECSRIPTIVE APPROACH TO IMAGE ANALYSIS [1]

I. Gurevich[2,3], Yu. Trusova[2,4], V. Yashina[2,5],

[2] Dorodnicyn Computing Centre, Russian Academy of Sciences,
40 Vavilov str., Moscow, 119333 ,Russian Federation,
[3] igourevi@ccas.ru, [4] ytrusova@ccas.ru, [5] werayashina@gmail.com

The paper is devoted to the state-of-the-art of the Descriptive Approach to image analysis (DA). The main purpose is to estimate the most important results of the DA, to outline the most important problems of the DA and to formulate prospective tasks which are necessary for successful development of fundamentals of the DA and its practical applications. We shall discuss also the hierarchy of the problems and the reasonable sequence of their setting up and solving. The special attention will be given to the opportunities and limitations of algebraic techniques in image analysis.

Introduction

Automation of image processing, analysis, estimating and understanding is one of the crucial points of theoretical computer science having decisive importance for applications
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 is passing 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. 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.
A new approach to analyzing and evaluating information represented in the form of images - the Descriptive Approach to image analysis and understanding (DA) by I.B.Gurevich [2, 6, 14-16] is based on the modification of the "Algebraic Approach" by Yu.I.Zhuravlev [43] for the case when the initial information is represented in the form of images.

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.
Image mining now tends to multiplicity (multialgorithmic and multimodel) and fusion of the results when 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. Yu.I.Zhuravlev obtained the

[1] This work was supported in part by the Russian Foundation for Basic Research (projects nos. 11-01-00990, 12-07-31123) and by the Presidium of the Russian Academy of Sciences within the program "Information, Control, and Intelligent Technologies and Systems" (project no. 204) and the program of the Division of Mathematical Sciences, Russian Academy of Sciences "Algebraic and Combinatorial Methods of New Generation Mathematical Cybernetics and Information Systems".

first and fundamental results in this area in 1970s [43].

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 of 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 and analyze data mathematically and detect hidden regularities.

There are internal scientific problems that have arisen within image recognition. First of all, 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 a recognition object, and some others. These problems form the base of the mathematical agenda of the descriptive theory of image recognition developed using the ideas of the algebraic approach to recognition.

There are three main issues one needs to solve when dealing with images: 1) describe images; 2) develop, study and optimize the selection of mathematical methods and tools of data processing in image recognition; 3) implement mathematical methods of image analysis on a software and hardware basis.

Algebraization of pattern recognition and image analysis

This section contains steps of the algebraization in image analysis fundamentals and the basic theories of pattern recognition, image algebras.

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.

However we still need a regular basis to arrange and choose suitable methods of image analysis, represent in an unified way the processed data (images), construct mathematical models of images designed for recognition problems, and, on the whole, to introduce the universal language for unified description of images and transformations over them.

The necessity to solve complex recognition and specialized image languages generated attention to 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 the substantial achievements in this “descriptive” line of R&D we should mention publications by A.Rosenfeld [34], T.Evans [12], R.Narasimhan [29], R.Kirsh [21], A.Shaw [37], H.Barrow, A. Ambler, and R. Burstall [1], S.Kaneff [20].

In 1970s Yu.I.Zhuravlev proposed “The Algebraic Approach to Recognition and Classification Problems” [43], 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” [18], where he considered methods of data representation and transformation in recognition problems in terms of regular combinatorial structures, leveraging algebraic and probabilistic apparatus. M.Pavel [31] introduced “Theory of Categories Techniques in Pattern Recognition”, to describe pattern recognition algorithms via transforms of initial data preserving its class membership.

The Russian mathematical school also has important original results on algebraic tools for pattern recognition and image analysis in. There are algebras on algorithms, algebraic multiple classifiers, algebraic committees of algorithms, combinatorial algorithms for recognition of 2-D data [1], descriptive image models, 2D formal grammars [34].

In the framework of scientific school of Yu.I.Zhuravlev several essential results were obtained by V.L.Matrosov [26], K.V.Rudakov [35] and V.D.Mazurov [27]. There are significant number of results concerned with algebraic methods of analysis and estimation

of information represented as signals (G.Labunec [22], Ya.A.Furman [13], V.M.Chernov [4]).

Algebraization of pattern recognition and image analysis has attracted and continues to attract the attention of many researchers. First of all – the development of algebraic construction for image analysis and processing – formal grammars, cellular automata, mathematical morphology, image algebras, multiple algorithms, descriptive approach.

The idea of constructing a unified language for concepts and operations used in image processing appeared for the first time in works of Unger [42], who suggested to parallelize algorithms for processing and image analysis on computers with cellular architecture.

Mathematical morphology by G. Matheron [25] and J.Serra [36], became a starting point for a new mathematical wave in image analysis. Serra and Sternberg [39] 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 [28] that it was precisely Sternberg who introduced the term "image algebra" in the current standard sense. The final version of image algebras (IA) was Standard Image Algebra by G.Ritter [32] (algebraic presentation of image analysis and processing operations).

Descriptive Image Algebras (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 [14-16].

In the history of algebraization we should mention: J.von Neumann [30], S.Unger [42] (studies of interactive image transformations in cellular space); M.Duff, D.Watson, T.Fountain, and G.Shaw [10] (a cellular logic array for image processing); A.Rosenfeld [33] (digital topology); H.Minkowski and H.Hadwiger (pixel neighborhood arithmetic and mathematical morphology); G.Matheron, J.Serra, S.Sternberg [25, 36, 39] (a coherent algebraic theory specifically designed for image processing and image analysis - mathematical morphology); S. Sternberg [39] (the first to use the term "image algebra"); P.Maragos [24] (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 [32] (Image Algebra); I.B.Gurevich [15] (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 [7, 8, 10, 11, 19, 40] (recent papers on mathematical morphology and image algebras).

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 synthesis of image models proposed to formalize and systematize the methods and forms of image representation.

By the middle of 1990s, it became obvious that for the development of image analysis and recognition, it is critical to: 1) understand the nature of the initial information – images, 2) find methods of image representation and description that allow constructing image models designed for recognition problems, 3) establish the mathematical language designed for unified description of image models and their transformations that allow constructing image models and solving recognition problems; 4) 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 an unified conceptual structure that helps to develop and implement these models and the mathematical language [14-17]. 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.

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: 1) the description of images; 2) the development, exploration, and optimization of the selection of mathematical methods and tools for information processing in the analysis of images; 3) the hardware and software implementation of the mathematical methods of image analysis.

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 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) [14-16]; 4) the mathematical formulation of the problem of image recognition; 5) mathematical theories of image analysis and pattern recognition; 6) a model of the process for solving a standard problem of image recognition. The main objects and means of the DA are: 1) images; 2) a universal language (DIA); 3) 2 types of descriptive models: a) an image model; b) a model for solving procedures of problems of image recognition and their implementation; 4) descriptive algebraic schemes of image representation (DASIR); 5) multimodel and multiaspect representations of images, which are based on generating descriptive trees (GDT) [14-16].

The basic methodological principles of the DA are: 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 (descriptive image model - DIM); 5) RIRF; 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; 10) the definition and use of nonclassical mathematical theory for the recognition of new formulations of problems of analyzing and recognizing images.

Ontology-based approach to image analysis

The automation of image analysis assumes that researchers and users of different qualifications have at their disposal not only a standardized technology of automation, but also a system supporting this technology, which accumulates and uses knowledge on image processing, analysis and evaluation and provides adequate structural and functional possibilities for supporting the more intelligent choice and synthesis of methods and algorithms. The automated system (AS) for image analysis must combine the possibilities of the instrumental environment for image processing and analysis and a knowledge-based system. Therefore, one of its main components is a knowledge base. Knowledge bases usually contain modules of universal knowledge, which are not related to any subject domain and knowledge modules related to a certain subject domain. The AS must provide software implementation of the hierarchies of classes of the main objects used in image analysis, have a specialized user interface, contain a library of algorithms that allow one to solve the main problems of image analysis and understanding with the help of efficient computational procedures, and provide accumulation and structuring of knowledge and experience in the domain of image analysis and understanding.

The need of efficient knowledge representation facilities can be fulfilled by using a suite of ontologies, For example, in [23], an approach devoted to semantic image interpretation for complex object classification purposes is proposed. The work described in [5] addresses the problem of explicit representation of objectives when developing image processing applications. The proposed

framework demonstrates that ontology-based content representation can be used as an effective way for hierarchical and goal-directed inference in high-level visual analysis tasks.

In [6], a novel knowledge-oriented approach to image analysis based on the use of thesauri and ontologies as tools for representation of knowledge, which are necessary for making intelligent decisions on the basis of information extracted from images, is proposed. The main contribution of this work is the development of a sufficiently detailed and well-structured Image Analysis Ontology (IAO). As a main source of the information about concepts the Image Analysis Thesaurus (IAT) [2] has been used. The important feature of the IAT is a novel hierarchical classification of tasks and algorithms for image processing, analysis and recognition.

Conclusion

Analyzing the existing algebraic apparatus, we came to the following requirements on the language for algorithms for solving problems of image processing and understanding: 1) the new algebra must make possible processing of images as objects of analysis and recognition; 2) the new algebra must make possible operations on image models; 3) the new algebra must make possible operations on main models of procedures for image transformations; 4) it is reasonable 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. The next results will be connected with the Image Formalisation Space and its topological properties.

References

1. H.G.Barrow, A.P.Ambler, R.M.Burstall. Some Techniques for Recognizing Structures in Pictures // Frontiers of Pattern Recognition. Proc. of the Int. Conf. on Frontiers of Pattern Recognition (ed. by Satosi Watanabe). - Academic Press, 1972. – P.1-30.
2. V.N.Beloozerov, I.B.Gurevich, N.G.Gurevich, D.M.Murashov, Yu.O.Trusova. Thesaurus for Image Analysis: Basic Version // Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications. - Vol. 13, No.4. - Pleiades Publishing, Inc., 2003. – P.556-569.
3. S.Bloehdorn et al. Semantic Annotation of Images and Videos for Multimedia Analysis // ESWC 2005, LNCS 3532 (eds. A. Gomez- Perez and J. Euzenat). - Springer, 2005. – P.592–607.
4. V.M.Chernov. Clifford Algebras Are Group Algebras Projections // Advances in Geometric Algebra with Applications in Science and Engineering (eds. E.Bayro-Corrochano, G.Sobczyk). – Boston: Birkhauser, 2001. – P.467-482.
5. R.Clouard, A.Renouf, M.Revenu. An Ontology-Based Model For Representing Image Processing Application Objectives // Int. Journal of Pattern Recognition and Artificial Intelligence. - Vol. 24, no. 8. - 2010. – P.1181-1208.
6. S.Colantonio, I.Gurevich, G.Pieri, O.Salvetti, Yu.Trusova. Ontology-Based Framework to Image Mining // Image Mining Theory and Applications: Proc. of the 2nd Int. Workshop on Image Mining Theory and Applications (in conjunction with VISIGRAPP 2009), Lisboa, Portugal (eds. I.Gurevich, H.Niemann and O.Salvetti). - INSTICC PRESS, 2009. – P.11-19.
7. J.Crespo, J.Serra, R.W.Schaffer. Graph-based Morphological Filtering and Segmentation // Proc. 6th Symp. Pattern Recognition and Image Analysis, Cordoba. - 1995. – P.80 – 87.
8. T.Crimmins, W.Brown. Image Algebra and Automatic Shape Recognition // IEEE Transactions on Aerospace and Electronic Systems. - Vol. 21, No. 1. - 1985. – P.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. - 1993. – P.283–306.
10. M.J.B.Duff, D.M.Watson, T.J.Fountain, G.K.Shaw. A Cellular Logic Array for Image Processing // Pattern Recognition. - Vol.5, No.3. – 1973. – P.229–247.
11. E.R.Dougherty. A Homogeneous Unification of Image Algebra. Part I: The Homogenous Algebra, Part II: Unification of Image Algebra // Imaging Science. - Vol. 33, No.4. – 1989. - P.136-143, P.144-149.
12. T.G.Evans. Descriptive Pattern Analysis Techniques: Potentialities and Problems // Methodologies of Pattern Recognition. Proc. of the Int. Conf. on Methodologies of Pattern Recognition. - Academic Press, 1969. – P.149-157.
13. Ya.A.Furman. Parallel Recognition of Different Classes of Patterns // Pattern Recognition and Image Analysis. - Pleiades Publishing, Ltd., 2009. - Vol.19, No.3. P.380-393.
14. 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. – P.298-328.
15. I.B.Gurevich, 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. - P.564-589
16. I.B.Gurevich, V.V.Yashina. Descriptive Approach to Image Analysis: Image Formalization Space // Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications. - Pleiades Publishing, Inc., 2012. - Vol.22, No.4. – P.495-518.
17. P.D.Gader, M.A.Khabou, A.Koldobsky. Morphological Regularization Neural Networks // Pattern Recognition. - 2000. - Vol.33. – P.935-944.
18. U.Grenander. Elements of Pattern Theory. - The Johns Hopkins University Press, 1996.
19. R.Haralick, L.Shapiro, J.Lee. Morphological Edge Detection //IEEE J. Robotics and Automation. – 1987. - Vol. RA-3, No.1. – P.142-157.
20. S.Kaneff. Pattern Cognition and the Organization of Information // Frontiers of Pattern Recognition. Proc, of the Int. Conf. on Frontiers of Pattern Recognition (ed. Satosi Watanabe). - Academic Press, 1972. – P.193-222
21. R.Kirsh. Computer Interpretation of English Text and Picture Patterns // IEEE-TEC. – 1964. - Vol. EC-13, No. 4.
22. V.G.Labunec. Algebraic Theory of Signals and Systems (Digital Signal Processing). - Krasnoyarsk University, 1984.
23. N.Maillot, M.Thonnat, A.Boucher. Towards ontology-based cognitive vision // Machine Vision and Applications. – 2004. - Vol. 16. – P.33-40.
24. P.Maragos. Algebraic and PDE Approaches for Lattice Scale-Spaces with Global Constraints // Int. Journal of Computer Vision. - Kluwer Academic Publishers, 2003. - Vol.52, No.2/3. - P.121-137.
25. G.Matheron. Random Sets and Integral Geometry. - New York: Wiley, 1975.
26. 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. – 1985. - Vol.24, No.1. – P.79-87.
27. V.D.Mazurov, M.Yu.Khachai. Parallel Computations and Committee Constructions // Journal Automation and Remote Control. - Plenum Press, 2007. - Vol.68, Issue 5. – P.912 – 921.
28. P.Miller. Development of a Mathematical Structure for Image Processing: Optical division tech. report. - Perkin-Elmer, 1983.
29. R.Narasimhan. Picture Languages // Picture Language Machines (ed. S.Kaneff). - Academic Press, 1970. – P. 1-30.
30. J.von Neumann. The General Logical Theory of Automata // Celebral Mechenism in Behavior: The Hixon Symposium. - John Wiley & Sons, 1951.
31. M.Pavel. Fundamentals of Pattern Recognition. - New York: Marcell, Dekker, Inc., 1989.
32. 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.
33. A.Rosenfeld. Digital Topology // American Math Monthly. – 1979. - Vol.86.
34. A.Rosenfeld. Picture Languages. Formal Models for Picture Recognition // Academic Press, 1979.
35. K.V.Rudakov. Universal and local constraints in the problem of correction of heuristic algorithms // Cybernetics. – 1987. Vol.23, Issue 2. – P.181-186.
36. J.Serra. Image Analysis and Mathematical Morphology. - Academic Press, 1982.
37. A.Shaw. A Proposed Language for the Formal Description of Pictures. CGS Memo, 28, Stanford University, 1967.
38. M.Schlesinger, V.Hlavac. Ten Lectures on Statistical and Structural Pattern Recognition // Computational Imaging and Vision. - Dordrecht/Boston/London: Kluwer Academic Publishers, 2002. - Vol.24. –P.520
39. S.R.Sternberg. Grayscale Morphology // Computer Vision, Graphics and Image Processing. – 1986. - Vol.35, No.3. – P.333-355.
40. P.Sussner. Observations on Morphological Associative Memories and The Kernel Method // Neurocomputing. – 2000. - Vol.31. – P.167–183.
41. C.Town. Ontological inference for image and video analysis // Machine Vision and Applications. – 2006. - Vol. 17, No. 2. – P.94-115.
42. S.H.Unger. A Computer Oriented Toward Spatial Problems. // Proc. of the IRE. – 1958. - Vol.46. – P.1744-1750.
43. 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", 1998. - Vol.8. – P.59-100.

AUTOMATIC GENERATION OF IMAGE ANALYSIS PROGRAMS

M. Herrmann[1,2], C. Mayer[1,3], B. Radig[1,4]

[1]Image Understanding and Knowledge-Based Systems
Technical University of Munich, Boltzmann street 3, 85748, Munich, Germany,
[2] herrmmic@in.tum.de, [3] mayerc@in.tum.de, [4] radig@in.tum.de

In this paper, we introduce a system that generates computer vision programs for a given task, which is specified by regions of interest in a collection of example images. The system relies on a database of operators, which are combined by an automated planning approach to create executable programs. We present an early proof-of-concept implementation that relies on a limited database to solve the tasks of finding players on a soccer field or cups on a table. Our experimental evaluation shows that the basic approach is working on relative simple scenarios. Future work will focus on integrating more complex problem descriptions, which requires more sophisticated planning strategies in order to compensate for rapidly increasing search spaces.

Introduction

Traditionally, program code is manually designed and written by a programmer, who selects operators and parameter values and combines them to create an executable program. For that purpose, the programmer relies on their experience and knowledge and on the technical documentation of operators. After the program is composed, the programmer executes it on example data to inspect the program's behavior and to ensure that the program's output matches the task specification. Finding an appropriate parameterization often requires a lot of iterations. Especially in image analysis programs, alteration of parameters may change the complete behavior of a program and determining the optimal values is a crucial and time-consuming task.

In this paper, we propose a different approach that creates the program code automatically given a set of annotated images with positive examples, taking into account that providing the desired output is much more intuitive and less error prone than writing the complete program code and specifying parameter values.

Our systematic approach is based on a database containing technical information about all available operators and how these operators may be combined. The system uses this database to create a large number of executable programs by exploring suitable combinations of operators and parameter values. The output of these programs is compared to the specified result in order to evaluate to which degree they meet the given demands. This idea is close to genetic programming [1] in likewise artificially producing programs that are evaluated by a fitness function. However, in contrast to genetic programming, instead of taking randomized decisions, heuristic search is performed. Currently, an early version of our system exists as a proof-of-concept implementation that includes a limited number of operators and therefore solves only basic computer vision tasks. So far, two example applications have been inspected: The detection of players in a soccer match and the detection of cups on a table. However, these experiments show that the basic approach is successful.

Related Work

Our approach has many similarities to automatic planning, which considers the creation of plans as a combination of actions with preconditions, which have to be true before the action is executed, and postconditions, which model the influence on the world model [2]. In relation to our system,

actions correspond to operators, preconditions correspond to input variables for operators and postconditions correspond to variables, which are generated by operators. Often a heuristic is included that controls the search tree expansion and prevents unsuitable combinations from being inspected [3].

When automatic code generation is considered in literature, mostly easy-to-use programming interfaces are proposed, which assist the user in composing modules into a complex system. Afterwards, the code is generated from these specifications. An example is "Reo", a system that creates a web service by combining other web services [4]. A similar system is presented by Rodrigues et al. An UML diagram is transformed in multiple steps and finally executable code is created [5]. The software "HALCON", distributed by the company MVTec Software GmbH, provides helpful tools for image acquisition, camera calibration etc., which create code snippets to use in larger programs [6]. Although these systems provide automatic code generation on a certain level, they still require a lot of human expertise. Approaches with minimal manual interaction are rare in literature. One approach is presented by Demel et al, where programs for convex optimization are created, given a task description [7]. Another example is presented by Moehrmann et al. Their system generates image analysis software from specifications by training classifiers on manually annotated training data [8].

System Description

The basic idea of our approach is to create combinations of parameterized operators with help of a heuristic search strategy and to apply them to manually annotated images. A fitness function determines to what degree their results match this kind of task specification. Therefore, the system consists of several components:

- A database of available operators. For each operator, the database contains the types of variables that the input operator requires for execution and also the types of output variables that are created by the operator.
- A planner, which creates a search tree with programs in its nodes.
- A system, which checks a single program against the manually specified test data.
- A set of manually annotated example images.

The core component of this approach is the planner, therefore it will be presented in more details.

The planner recursively creates a search tree, whose nodes are programs. Programs contain global variables (either specified at the beginning of the program or created by operators) and the sequence of the applied, parameterized operators.

The planning process starts with the empty program, which contains only the original input image in the variable list and an empty operator list. In a recursive procedure, the planner inspects all available operators and checks whether they are applicable, i.e. whether all input parameters can be assigned by already existent variables of the same type in the variable list of the current recursion path. Here, different types of operator parameters are handled in two ways. Iconic parameters (images or image regions) have to be present in the variable list, whereas numerical parameters are instantiated on-the-fly dependent on some sampling strategy and the operator's specification.

For each combination of operator and parameterization, a new branch in the search tree is created. Please see Fig. 1 for an example. In the first step (a), only the original image is in the parameter list and only operators that work on a single image are applicable.

For instance, a threshold operator is applicable, since it requires only a single image and a numerical value, which is created on-the-fly. In this example the numerical parameter is sampled by 3 values, namely 30, 60 and 90. The planner then creates a tree node for each of these operators and attaches them to the father node (in this example the root). The search tree in step (b) consists of several nodes in which the original image and threshold images, each calculated with a different threshold, are available. As displayed in step (c) these nodes may now be expanded further with operators that use two images (the original image and the threshold image) or again with a single image operator, which now

may choose between the original image and the threshold image. This procedure continues until a predefined depth of the tree is reached. Now each node and each leaf contains a program that will be inspected for its fitness,

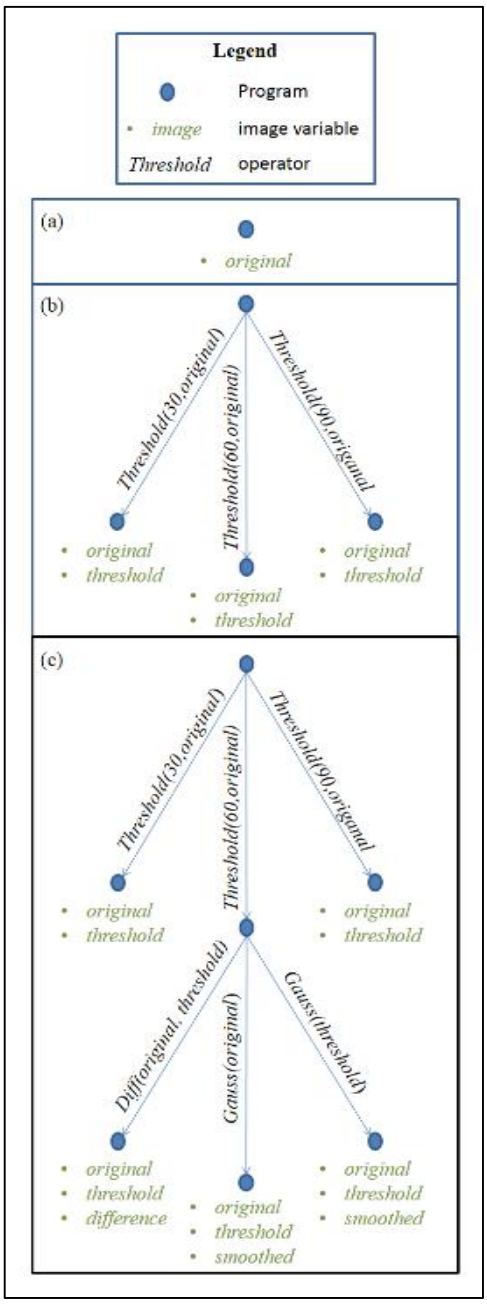


Fig. 1: The planner creates programs by combining and parameterizing single operators.

which requires the definition of some error function. Furthermore, this requires compiling and executing them on the manually annotated image data. To reduce the time required, we apply some simple but effective heuristics, which either directs the search tree expansion and reduces the number of nodes or identify programs that are not considered for evaluation.

The rules are:

- During tree creation an operator with a specific input parameter assignment is only used, if this operator with exact this assignment has not been used in the same program before.
- A program is considered as redundant, if there is an operator, whose output is not used by another operator.
- Only programs that have variables with the same type as the desired output in the variable list (e.g. image regions) are considered in the evaluation step.

In our example the image annotations consist of specified rectangular image regions of interest and the error function e is defined as

$$e(A,N) := (1-w)d(A,N) - wd(N,A), \quad (1)$$

where $0 \leq w \leq 1$ is a regulating parameter and A and N are the sets of actual estimated object regions and of the manual defined nominal regions respectively and

$$d(S_1,S_2) := \frac{\sum_{R_1 \in S_1} \min_{R_2 \in S_2} d_H(R_1,R_2)}{|S_1|}. \quad (2)$$

Thereby, S_1 and S_2 are sets of image regions and the normalized Hamming distance d_H for discrete sets of pixels (image regions) R_1 and R_2 is defined as:

$$d_H(R_1,R_2) := \frac{|R_1 \cap R_2^C| + |R_1^C \cap R_2|}{|R_1| + |R_2|}. \quad (3)$$

The fitness is estimated by means of $1 - e$.

Experimental Evaluation

In our proof-of-concept, we embedded a selection of five basic operators and set the maximum program length to five operators. The operators provided are: Color space conversion from RGB to HSV, a Gaussian image smoothing, histogram-based background estimation, a region-based foreground estimator and creation of derivation images. In the case of cup detection, the planner created 534 programs which took 3 seconds on a 3 GHz CPU. The evaluation of these programs took 59 seconds on four images of size 1024x768 pixel. Please see

Figure 2 for the manually annotated images used in the evaluation step.
Fitness values obtained by automatically created programs range from 0.65 to 0.98 with higher values representing higher fitness. As can be seen in Figure 3 the cups/soccer players are recognized quite well by the program with the highest fitness value.

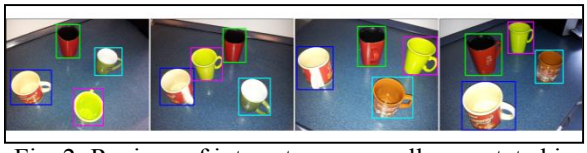


Fig. 2: Regions of interest are manually annotated in four different example images.

This shows that our proof-of-concept implementation is successful.
A current drawback, however, is the time required to execute all programs. Especially the sampling strategy for numerical parameters has a major impact on the number of generated programs and therefore on the runtime of the system.

Fig. 3: Results of the programs with the highest fitness for cup (top) and player detection (bottom).

Future Work

Although our proof-of-concept implementation shows that the basic approach works, there are still many open ends. As already mentioned, the sampling strategy and the heuristics for the tree expansion heavily influence the size of the search space and thus the runtime. Since programs are created from single operators so far, the idea to use complex operator modules or reuse and adapt already existing programs for similar problems seems reasonable. Furthermore, our current approach does not consider control structures like if-clauses or loops. Finally, more complex target specifications than just marking regions of interest on images is a long-time goal, since for most real vision applications the target is expressed in a more complex manner. This is the most challenging research opportunity, because it requires adapting the task specification procedure and the search heuristics, but also requires taking knowledge about the task context into consideration. The ultimate goal is to couple the system with a knowledge base, which allows for complex formal task descriptions and for complex heuristics, which will be required, since more complex tasks lead to more complex programs and to larger search spaces.

Conclusion

We present a system that automatically creates and evaluates computer vision programs for a given task. In the current, early version, the task is specified by rectangular regions of interest in example images. Our evaluation shows that the basic approach is working, but capability and performance have to be improved.

References

1. J. R. Koza. Genetic Programming: On the Programming of Computers by Means of Natural Selection // MIT Press, Cambridge, 1992.
2. R. Jinghai, S. Xiaomeng. A Survey of Automated Web Service Composition Methods // Lecture Notes in Computer Science – Springer, Berlin, 2005. - Vol. 3387 - P. 43-54.
3. F. Rossi, P. van Beek, T. Walsh (ed). Handbook of Constraint Programming, Foundations of Artificial Intelligence // Elsevier, New York, 2006.
4. S.S. T.Q. Jongmans, F. Santini, M. Sargolzaei, F. Arbab, and H. Afsarmanesh. Automatic Code Generation for the Orchestration of Web Services with Reo // Lecture Notes in Computer Science – Springer, Berlin, 2012 – Vol. 7592 - P. 1-16.
5. A. W. O. Rodrigues, F. Guyomarc'h, J.-L. Dekeyser. An MDE Approach for Automatic Code Generation from UML/MARTE to OpenCL // Computing in Science and Engineering - 2013 - Vol. 15, No.1 - P. 46-55
6. http://www.mvtec.com, last visited June 27th 2013
7. J. Demmel, J. Dongarra, V. Eukhout, E. Fuentes, A. Petitet, R. Vuduc, R.C. Whaley, K. Yelick. Self-Adapting Linear Algebra Algorithms and Software // Proceedings of the IEEE - IEEE Computer Society 2005 - Vol. 93, Nr. 2 - P. 293 - 312.
8. J. Moehrmann, G. Heidemann, O. Siemoneit, C. Hubig, U.-P. Kaeppeler, P. Levi: Context Generation with Image Based Sensors: An Interdisciplinary Enquiry on Technical and Social Issues and their Implications for System Design // Proceedings of World Academy of Science, Engineering and Technology - 2010 – Vol. 61 – P. 311-317

HYPERCOMPLEX ALGEBRAS AS UNIFIED LANGUAGE FOR IMAGE PROCESSING AND PATTERN RECOGNITION.

PART 1. IS THE BRAIN A CLIFFORD ALGEBRA COMPUTER?[1]

V. Labunets[2]

[2] Urals Federal University, Ekaterinburg, Russia, vlabunets05@yahoo.com

We present a new theoretical framework for multidimensional image processing using hypercomplex and Clifford algebras. The main goal of the work is to show that commutative hypercomplex algebras and Clifford algebras can be used to solve problems of multi-color image processing and pattern recognition in a natural and effective manner. One can argue that nature has, through evolution, also learned to utilize properties of hypercomplex numbers. Thus, the visual cortex of a brain might have the ability to operate as a Clifford algebra computing device.

Introduction

One of the main and interesting problem of information science is clarification of how animals' eyes and brain process color and multicolor images and recognize objects in the real world. We know that primates and animals with different evolutionary histories have color visual systems of different dimensionality. For example, primates have dichromatic and trichromatic visual systems, and they can use 2D and 3D channel images to recognize color images. Reptiles have multichromatic visual systems. For example, tortoise visual system has five types of color photoreceptors (R,G,B, DC,UV). Shrimps use ten spectral types of photoreceptors in their eyes to recognize fine spectral details. Practice shows that animals' eyes and brain successfully process color images and recognize objects at different locations, of different views and illumination, and with different degrees of blurring. But how is it done by the brain? How do we see? How do we recognize moving and changing objects of the surrounding world? A moving object is fixed in the retina as a sequence of different images. No individual image allows to reach a conclusion about the true shape of the object. This means that a set of sequential images appearing in the retina must contain a constant «something», thanks to which we see and recognize the object as a whole. This constant «something» is called *invariant*. In order for an artificial pattern recognition system to perform in the same way as any biological visual systems, the recognition result should be invariant with respect to various *transformation groups* of the patterns such as translation, rotation, size variation, and change in illumination and color.

For a long time multicolor (MC) and hyperspectral (HS) measurements are used for region classification in satellite images. Here, the term «MC-, HS-image» is used for an image with more than one component

$$\vec{\mathbf{f}}_{Mcol}(x,y)=\left(f_1(x,y),f_2(x,y),...,f_K(x,y)\right)$$

We will interpret multicomponent images as multiplet-valued signals

$$\begin{aligned}&\mathbf{f}_{Mcol}(x,y)=\\&=\left(f_1(x,y),f_2(x,y),\ldots,f_k(x,y)\right)=\\&=f_R(x,y)1+f_G(x,y)\varepsilon+...+f_B(x,y)\varepsilon^{k-1},\end{aligned}\qquad(1)$$

which take values in the multiplet (multicolor) algebra $\mathcal{A}_k^{Mcol}=\left(\mathbf{R}\left|1,\varepsilon^1,\varepsilon^2,..,\varepsilon^{k-1}\right.\right)$, where $\varepsilon^k=1$ and $1,\varepsilon^1,...,\varepsilon^{k-1}$ are hyperimaginary units.

Hypercomplex algebras generalize the algebras of complex numbers, and Clifford algebras generalize the algebras of complex numbers, quaternions and octonions. For MC-

[1]This work was supported by RFBR (grants № 13-07-12168, BR № 13-07-00785) and by MES RF grant № 218-03-167

image processing we use commutative hypercomplex numbers of the multicolor algebra and for recognition of 2D, 3D and nD images we turn the spaces R^2, R^3 and R^n into corresponding Clifford algebras (and call them the space algebras).

Our hypotheses are

1. We suppose that a brain operates with hypercomplex numbers when processing multicolor image and calculates some hypercomplex-valued invariants of an image when recognizing it. In the algebraic-geometrical approach, each color or multicolor pixel is considered not as a kD vector, but as a kD hypercomplex number (k is the number of image spectral channels). We suppose that the human brain can use the spinors and multiplet color algebra for mental motions and for changing color of images (for example, in a dream), which are contained in the brain memory on the so-called «screen of mind».

2. Brains use different algebras on two levels (retina and Visual Cortex-VC). Multicolor images appear on the retina as functions with values in a multiplet kD algebra (k–cycle algebra) where k is the number of image spectral channels. In particular, RGB–color images as they appear on the human retina are represented as triplet–valued functions. But multicolor images in an animals' Visual cortex are functions with values in a 2^k-D Clifford algebra

3. Visual systems of animals with different evolutionary history use different hypercomplex algebras for color and multicolor image processing. One can argue that the Nature has also learned to utilize (through evolution) properties of hypercomplex numbers.

4. Brain uses Clifford algebra for calculating of Clifford-valued invariants during image recognition. Thus, the brain might have the ability to operate as Clifford algebra computer. We don't agree with Kronecker that "Lord created the integers and the rest is the work of man". We think that the Lord knew Clifford algebras, and he was the first engineer who used these algebras for design annimals visual systems

In the algebraic approach, each pixel is considered not as a multi–dimensional vector, but as a hypercomplex (multi–dimensional) number. For this reason, we assume that the human retina and human VC use 3D hypercomplex (triplet) numbers and 8D Clifford numbers, respectively, to process color (RGB)–images. Note that both these assumptions (vector and hypercomplex natures of multicolor images) are only hypotheses. We have no biological evidence in the form of experiments that would verify that the brain actually uses any of the algebraic properties arising from the structures of vector spaces or Clifford algebras. We only know that animals are able to recognize objects in an invariant manner and to process multicolor images effectively.

Algebras and geometries of physical spaces

We suppose that a brain operates with hypercomplex numbers when processing image and calculates some hypercomplex-valued invariants of an image when recognizing it. For recognition of 2D, 3D and nD images we turn the spaces $\mathbf{R}^2$, $\mathbf{R}^3$ and $\mathbf{R}^n$ into corresponding Clifford algebras.

Let «small» nD space R^n be spanned on the orthonormal basis of *n space hyperimaginary units* I_i, $i=1,2,\ldots,n$. We assume

$$I_i^2=\begin{cases}+1 \text{ for } i=1,2,\ldots,p,\\ -1 \text{ for } i=p+1,2,\ldots,p+q,\\ 0 \text{ for } i=p+q+1,2,\ldots,p+q+r=n,\end{cases}$$

and $I_iI_j=-I_jI_i$. Now, we construct the «big» 2^n D hypercomplex space $\mathbf{R}^{2^n}$.

Let $\mathrm{b}=(b_1,b_2,\ldots,b_n)\in \mathrm{B}_2^n$ be an arbitrary n-bit vector, where $b_i\in \mathrm{B}_2=\{0,1\}$ and B_2^n is the nD Boolean. Let us introduce $\mathrm{I}^{\mathrm{b}}:=I_1^{b1}I_2^{b2}\ldots I_n^{bn}$. Then 2^n elements I^{b} form a basis of 2^nD space, i.e., for all $C\in\mathbf{R}^{2^n}$ we have $C:=\sum_{\mathrm{b}\in \mathrm{B}_2^n}\mathrm{c}_{\mathrm{b}}\mathrm{I}_{\mathrm{b}}$. If $C_1,C_2\in\mathbf{R}^{2^n}$, then we can define their product $C_1\cdot C_2$. There are 3^n possibilities for $I_i^2=+1,0,-1,\ \forall i=1,2,\ldots,n$. Every possibility generates an algebra. Therefore, the space $\mathbf{R}^{2^n}$ with 3^n rules of the multiplication forms 3^n different 2^nD algebras, which are called the *space Clifford algebras*. We denote these algebras by

$A_{2^n}^{Sp(p,q,r)}\left(\mathrm{R}\middle|I_1,\ldots,I_n\right)$, $A_{2^n}^{Sp(p,q,r)}$ (or $A_{2^n}^{Sp}$ if $I_1,\ldots,I_n$, p,q,r are fixed).

There is the conjugation operation in Clifford algebras $A_{2^n}^{Sp(p,q,r)}$. It maps every Clifford number $C \coloneqq c_0 I_0 + \sum_{\mathrm{b}\neq 0} c_{\mathrm{b}}\,\mathrm{I}_{\mathrm{b}}$ to the number $\overline{C} \coloneqq c_0 I_0 - \sum_{\mathrm{b}\neq 0} c_{\mathrm{b}}\,\mathrm{I}_{\mathrm{b}}$. The algebras $A_{2^n}^{Sp(p,q,r)}$ are transformed into $2^n\,\mathrm{D}$ pseudometric spaces designed as $\mathbf{G}_{2^n}^{p,q,r}$, if the pseudodistance between two Clifford numbers A and B is defined by $\rho(A,B)=\sqrt{(A-B)\overline{(A-B)}}$. Subspaces of pure vector Clifford numbers $x_1 I_1+\ldots+x_n I_n \in \mathbf{R}^n$ are $n\mathrm{D}$ spaces $\mathbf{R}^n \coloneqq \mathbf{G}_n^{p,q,r}$. The pseudometrics constructed in $\mathbf{G}_n^{p,q,r}$ induce corresponding pseudometrics in $\mathbf{G}_n^{p,q,r}$. Clifford numbers $\mathcal{E}_0 \in A_{2^n}^{\{0\}}$ of unit modulus represent the rotation group of the corresponding space $\mathbf{G}_n^{p,q,r}$ which is called *spinor group* and is denoted by $\mathbf{Spin}\left(A_{2^n}^{Sp(p,q,r)}\right)$. We know that generalized complex numbers and quaternions of unit modulus have the following forms:

$$\mathrm{e}_0 = e^{I\varphi} = \cos\varphi + I\sin\varphi,$$
$$Q_0 = e^{\mathbf{u}_0\varphi} = \cos\varphi + \mathbf{u}_0\sin\varphi,$$

where φ is a rotation angle around vector-valued quaternion u_0 of unit modulus. Clifford spinors $\mathcal{E}_0 \in \mathbf{Spin}\left(A_{2^n}^{Sp(p,q,r)}\right)$ with unit modulus have analogous form

$$\mathcal{E}_0 = e^{\mathrm{u}_0\varphi} = \cos\varphi + \mathrm{u}_0\sin\varphi \in \mathbf{Spin}\left(A_{2^n}^{Sp(p,q,r)}\right)$$

for appropriate bivector u_0.

Theorem 1. *All motions in 2D, 3D and nD spaces* $\mathbf{Aff}\left(\mathbf{G}_2^{p,q,r}\right)$, $\mathbf{Aff}\left(\mathbf{G}_3^{p,q,r}\right)$, $\mathbf{Aff}\left(\mathbf{G}_n^{p,q,r}\right)$ *are represented in the following forms:*

$$\mathbf{z}' = \mathrm{e}_0\,\mathbf{z}\,\mathrm{e}_0 + \mathbf{w},$$
$$\mathbf{x}' = Q_0\mathbf{x}Q_0^{-1} + \mathbf{w}, \quad \mathbf{x}' = \mathcal{E}_0\mathbf{x}\mathcal{E}_0^{-1} + \mathbf{w},$$

where $\mathrm{e}_0 \coloneqq e^{I\varphi/2}$, $Q_0 \coloneqq e^{\mathrm{u}_0\varphi/2}$, $\mathcal{E}_0 \coloneqq e^{\mathrm{u}_0\varphi/2}$, *and* $|\mathrm{e}_0| = |Q_0| = |\mathcal{E}_0| = 1$. *If* $|\mathrm{e}_0|, |Q_0|, |\mathcal{E}_0| \neq 1$, *then the latter transformations form the the «small» affine groups* $\mathbf{Aff}\left(\mathbf{G}_2^{p,q,r}\right)$, $\mathbf{Aff}\left(\mathbf{G}_3^{p,q,r}\right)$, $\mathbf{Aff}\left(\mathbf{G}_n^{p,q,r}\right)$, *respectively).*

Using this result, we can describe geometrical distortions of images on the language of Clifford algebras. These distortions will be caused by: 1) $n\mathrm{D}$ translations $\mathbf{x}\to\mathbf{x}+\mathbf{w}$; 2) $n\mathrm{D}$ rotations $\mathbf{x}\to\mathcal{E}_0(\mathbf{x}+\mathbf{w})\mathcal{E}_0^{-1}$; 3) dilatation: $\mathbf{x}\to\lambda\mathbf{x}$, where $\lambda\in\mathbf{R}^+$. If $\mathbf{f}(\mathbf{x})$ is an initial image and ${}_{\lambda\mathcal{E}_0\mathrm{w}}\mathbf{f}(\mathbf{x})$ is its distorted version, then ${}_{\lambda\mathcal{E}_0\mathrm{w}}\mathbf{f}(\mathbf{x}) \coloneqq \mathbf{f}\left(\lambda\mathcal{E}_0(\mathbf{x}+\mathbf{w})\mathcal{E}_0^{-1}\right)$ where λ is a scale factor, $\mathbf{x},\mathbf{w}\in\mathbf{G}_n^{p,q,r}$.

Clifford-valued invariants

Let us assume that

$$\mathbf{f}_{Mcol}(\mathbf{x}) : \mathbf{R}^n = \mathbf{G}_n^{p,q,r} \to A_k^{Mcol}$$

is an image of a multicolor $n\mathrm{D}$ object. Changes in the surrounding world can be treated in the language of the spatial-multicolor algebra as an action of two groups: the space affine group $\mathbf{Aff}\left(\mathbf{G}_n^{p,q,r}\right)$ acting on the physical space $\mathbf{G}_n^{p,q,r}$ and the multicolor group $\mathbf{MCG}(k)$ acting on A_k^{Mcol}. Let $\mathbf{G}^{SpMcol} = \mathbf{Aff}\left(\mathbf{G}_n^{p,q,r}\right)\times\mathbf{MCG}(k)$ be a spatial-multicolor group, and $\left(\mathrm{g}^{Sp},\mathrm{g}^{Mcol}\right)\in\mathbf{G}^{SpMcol}$, where $\mathrm{g}^{Sp}\in\mathbf{Aff}\left(\mathbf{G}_n^{p,q,r}\right)$, $\mathrm{g}^{Mcol}\in\mathbf{MCG}(k)$. If $\mathbf{x}\in\mathbf{G}_n^{p,q,r}$ is a generalized space Clifford number and $\mathcal{M}\in A_k^{Mcol}$ is multicolor number, then all products of the form $\mathbf{x}\cdot\mathcal{M}$ is called the *spatial-color numbers.* They form a *space-color algebra* $A^{SpMcol} \coloneqq A_{2^n}^{Sp}\otimes A_k^{Mcol}$. Here, we assume that all spatial hyperimaginary units commute with all color units.

Definition 1. *The* A^{SpMcol}*-valued functional* $\mathrm{Jn} = \Phi\left[\mathbf{f}_{Mcol}(\mathbf{x})\right]$ *of the image* $\mathbf{f}_{Mcol}(\mathbf{x})$ *is called the relative* $\mathbf{G}^{SpMcol}$*-invariant if*

$$\mathrm{Jn} = \Phi\left\{\mathrm{g}^{Mcol} \circ \mathbf{f}_{Mcol}\left(\mathrm{g}^{Sp} \circ \mathbf{x}\right)\right\} = \\ = \mathcal{C} \cdot \Phi\left\{\mathbf{f}_{Mcol}\left(\mathbf{x}\right)\right\} \cdot \mathcal{C}^{-1}$$

$\forall \mathrm{g} \in \mathbf{G}^{SpMcol}$, *where* $\mathcal{C}$, $\mathcal{C}^{-1}$ *are left and right* A^{SpMcol}*-valued multiplicators. If* $\mathcal{C} = 1$ *then* Jn *is called the absolute invariant} and denoted by* In.

Let **c** be the centroid of the image $\mathrm{f}_{Mcol}(\mathrm{x})$, then functionals

$$\mathcal{M}_p := \Phi\left\{\mathbf{f}_{Mcol}\right\} = \int_{\mathbf{x} \in \mathbf{G}_n^{p,q,r}} \left(\mathbf{x} - \mathbf{c}\right)^p \mathbf{f}_{Mcol}\left(\mathbf{x}\right) \mathbf{dx}$$

are called the central A^{SpMcol}-valued moments of the nD image $\mathbf{f}_{Mcol}\left(\mathbf{x}\right)$, where $p \in \mathrm{Q}$ are rational numbers. Let us clarify the rules of moment transformations with respect to distortions of color and geometry of the initial images. If $\mathbf{f}_{Mcol}\left(\mathbf{x}\right)$ is initial image, then

$$_{\mathcal{M}_0,\lambda,\mathcal{E}_0,\mathrm{w}}\mathbf{f}_{Mcol}\left(\mathbf{x}\right) = \\ = \mathcal{M}_0\left\{\mathbf{f}_{Mcol}\left(\lambda\mathcal{E}_0\left(\mathbf{x} + \mathbf{w}\right)\mathcal{E}_0^{-1}\right)\right\}\mathcal{M}_0^{-1}$$

denotes its $\mathbf{MCG}(k)$-multicolor and $\mathbf{Aff}\left(\mathbf{G}_n^{p,q,r}\right)$-geometrical distorted copy. Here **x, w** are nD vectors. Summing **x** with **w** brings us to image translation by the vector **w**, two-sides multiplications $\lambda\mathcal{E}_0\left(\mathbf{x}+\mathbf{w}\right)\mathcal{E}_0^{-1}$ by $\lambda\mathcal{E}_0$ and $\mathcal{E}_0^{-1}$ equivalent to both a nD rotation of the vector **x+w** and a dilatation given by factor λ. Here, $\mathbf{f}_{Mcol} \to \mathcal{M}_0\left\{\mathbf{f}_{Mcol}\right\}\mathcal{M}_0^{-1}$ is a multicolor transformation of the initial image.

Theorem 1. *The central moments* $\mathcal{M}_p$ *of the multicolor images* $\mathbf{f}_{Mcol}\left(\mathbf{x}\right)$ *are relative* A^{SpMcol}*-valued invariants*

$$\mathrm{Jn}_p\left\{_{\mathcal{M}_0,\lambda,\mathcal{E}_0,\mathrm{w}}\mathbf{f}_{Mcol}\left(\mathbf{x}\right)\right\} = \mathcal{M}_p\left\{_{\mathcal{M}_0,\lambda,\mathcal{E}_0,\mathrm{w}}\mathbf{f}_{Mcol}\left(\mathbf{x}\right)\right\} = \\ = \left(\lambda^{p+n}\mathcal{M}_0\mathcal{E}_0^p\right) \cdot \mathcal{M}_p\left\{\mathrm{f}_{Mcol}\right\} \cdot \left(\mathcal{E}_0^{-p}\mathcal{M}_0^{-1}\right)$$

with respect to the spatial-multicolor group G^{SpMcol} *with both* A^{SpMcol}*-valued left* $\lambda^{p+n}\mathcal{M}_0\mathcal{E}_0^p$ *and right* $\mathcal{E}_0^{-p}\mathcal{M}_0^{-1}$ *multiplicators, respectively and the normalized central moments*

$$\left\|\mathrm{In}_p\left\{_{\mathcal{M}_0,\lambda,\mathcal{E}_0,\mathrm{w}}\mathbf{f}_{Mcol}\left(\mathbf{x}\right)\right\}\right\| = \\ = \frac{\left[\mathrm{Jn}_p\left\{_{\mathcal{M}_0,\lambda,\mathcal{E}_0,\mathrm{w}}\mathbf{f}_{Mcol}\left(\mathbf{x}\right)\right\}\right] \cdot \left[\mathrm{Jn}_0\left\{_{\mathcal{M}_0,\lambda,\mathcal{E}_0,\mathrm{w}}\mathbf{f}_{Mcol}\left(\mathbf{x}\right)\right\}\right]^{p-1}}{\left[\mathrm{Jn}_1\left\{_{\mathcal{M}_0,\lambda,\mathcal{E}_0,\mathrm{w}}\mathbf{f}_{Mcol}\left(\mathbf{x}\right)\right\}\right]^p}$$

are absolute scalar-valued invariants, with respect to the same group.

Using this theorem we can calculate hypercomplex-valued invariants of grey-level, color, multicolor and hyperspectral 2D, 3D and nD images for Euclidean and non-Euclidean geometries.

Conclusion

Digital computers use Boolean algebra. This algebra is the commutative Clifford algebra $A_{2^n}\left(\mathbf{GF}(2)\right)$ over binary Galois field $\mathbf{GF}(2)$). Are there nature analog computer working in the Clifford algebras?. Now we will try to answer to this question by remembering the question in the title: "*Is the Brain a Clifford Algebra Computer"?* Yes, can be! If this brain belongs to an animal being living in *nD* non–Euclidean space $\mathbf{G}_n^{p,q,r}$ and using multicolor algebra A_k^{Mcol} for recognition of *nD* non–Euclidean multicolor images $\mathbf{f}_{Mcol}\left(\mathbf{x}\right)$.

References

1. V.G. Labunets. Clifford Algebras as Unified Language for Multicolor Image Processing and Pattern Recognition // Computational Noncommutative Algebra and Applications, NATO Science Series.(Editors: J. Byrnes), Kluwer Academic Publishers, London.- 2003. - 2003.- P.197-226.
2. V.G. Labunets, E.V. Rundblad and J.Astola. Is the Brain a «Clifford algebra quantum computer»? // Applied Geometrical Algebras in Computer Science and Engineering. (L.Dorst, C.Doran, J.Lasenby, Editots), Birkhauser. – 2002. - P. 486-495.

HYPERCOMPLEX ALGEBRAS AS UNIFIED LANGUAGE FOR IMAGE PROCESSING AND PATTERN RECOGNITION.

PART 2. COLOR IMAGE PROCESSING[1]

V. Labunets[2,3], D. Gainanov[2,4], D. Berenov[2,5]

[2] Urals Federal University, Ekaterinburg, Russia,
[3] vlabunets05@yahoo.com, [4] damir@dc.ru, [5] berenov@dc.ru

We developed a novel algebraic approach based on triplet algebra to color image processing. A family of discrete color-valued transforms has been presented that can be used in color image processing and recognition.

Introduction

We develop a conceptual framework, design methodologies and software-tool for color, image processing systems with assessment capability. Color images are composed of a series of images in three optical bands at wavelengths $\lambda_R, \lambda_G, \lambda_B$, called RGB-spectral channels. Color image can be characterized as an image consisting of 3D vector-valued pixels:

$$\vec{\mathbf{f}}_{Col}(x,y) = \left(f_R(x,y), f_G(x,y), f_B(x,y)\right),$$

where each pixel is characterized by three scalars representing its red $f_R(x,y)$, green $f_G(x,y)$ and blue $f_B(x,y)$ components.

Our approach to color image processing based on triplet algebra. In the algebraic-geometrical approach, each color is considered not as a 3D vector, but as a 3D hypercomplex number. Real and complex numbers (from signal/image processing) can be extended to the generalized 3D hypercomplex. Among of different types of hypercomplex numbers the sets of so-called *triplet* numbers has a special status, because these algebras have been a fundamental tool in the color image processing theory. In this context, the full machinery of grey-level signal processing theory can be transposed in color image processing theory.

A natural question that arises in our approach is the definition of color transforms that can be used efficiently in color image processing. We propose a wide library of so-called ortho-unitary (color-valued) transforms transforms for using in image processing and pattern recognition applications.

Triplet (color) algebra

The color images are usually interpreted as 3D vectors-valued function.

$$\vec{\mathbf{f}}_{Col}(x,y) = \left(f_R(x,y), f_G(x,y), f_B(x,y)\right).$$

We will interpret color images as triplet-valued signals

$$\begin{aligned}\vec{\mathbf{f}}_{Col}(x,y) &= \\ &= \left(f_R(x,y), f_G(x,y), f_B(x,y)\right) = \\ &= f_R(x,y)1 + f_G(x,y)\varepsilon + f_B(x,y)\varepsilon^2,\end{aligned} \qquad (1)$$

which take values in the triplet (or color) algebra

$$\mathcal{A}_3^{col} = \left(\mathbf{R}\,|1, \varepsilon_{col}, \varepsilon_{col}^2\right) = \mathbf{R}1_{col} + \mathbf{R}\varepsilon_{col}^1 + \mathbf{R}\varepsilon_{col}^2,$$

where 1_{col}, $\varepsilon_{col}^1, \varepsilon_{col}^2$ are color hyperimaginary color units, respectively and $\varepsilon_{col}^3 = 1$. We will denote them by $1, \varepsilon^1, \varepsilon^2$. Numbers of the form $\mathcal{C} = x1 + y\varepsilon + z\varepsilon^2$ $(\varepsilon^3 = 1)$ were considered by Greaves [1]. According to Greaves, these numbers are called the *triplet numbers*. We shall call them the *color numbers*. Greaves considered color number $x + y\varepsilon + z\varepsilon^2$ as a point of 3D space. The addition and product of two triplet numbers $\mathcal{C}_1 = (r_1 + g_1\varepsilon + b_1\varepsilon^2)$ $\mathcal{C}_1 = (r_2 + g_2\varepsilon + b_2\varepsilon^2)$ are given by

[1]This work was supported by RFBR (grants № 13-07-12168, BR № 13-07-00785) and by MES RF grant № 218-03-167

$$\mathcal{C}+\mathcal{C}_2=(r_1+g_1\varepsilon\ +b_1\varepsilon^2)+(r_2+g_2\varepsilon\ +b_2\varepsilon^2)=$$
$$=(r_1+r_2)+(g_1+g_2)\varepsilon\ +(b_1+b_2)\varepsilon^2,$$
$$\mathcal{C}_1\cdot\mathcal{C}_2=(r_1+g_1\varepsilon\ +b_1\varepsilon^2)\cdot(r_2+g_2\varepsilon\ +b_2\varepsilon^2)=$$
$$=(r_1r_2+g_1b_2+b_1g_2)+$$
$$+(r_1g_2+r_2g_1+b_1b_2)\varepsilon\ +(r_1b_2+g_1g_2+r_2b_1)\varepsilon^2.$$

It is easy to see that the triplet product is isomorphic to 3–point convolution

$$\mathcal{C}_1\mathcal{C}_2=(r_1,g_1,b_1)*(r_2,g_2,b_2)=$$
$$=(r_1r_2+g_1b_2+b_1g_2,\ r_1g_2+r_2g_1+b_1b_2,$$
$$r_1b_2+g_1g_2+r_2b_1).$$

The *triplet conjugate* $\bar{\mathcal{C}}$ of a triplet number $\mathcal{C}=(r+g\varepsilon\ +b\varepsilon^2)$ is defined as

$$\bar{\mathcal{C}}=\overline{r+g\varepsilon\ +b\varepsilon^2}=r+g\varepsilon^2\ +b\varepsilon^1.$$

The norm $\|\mathcal{C}\|_2=\mathcal{C}\bar{\mathcal{C}}$ is given by

$$\|\mathcal{C}\|_{\mathcal{A}_3^{col}}=\mathcal{C}\bar{\mathcal{C}}=(r+g\varepsilon\ +b\varepsilon^2)(r+g\varepsilon^2+b\varepsilon)=$$
$$=(r^2+g^2+b^2)-(rg+rb+gb).$$

The color algebra $\mathcal{A}_3\left(\mathbf{R}|1,\varepsilon_{col},\varepsilon_{col}^2\right)$ is the direct sum of the real **R** and complex **C** fields:

$$\mathcal{A}_3^{col}\left(\mathbf{R}|1,\varepsilon_{col},\varepsilon_{col}^2\right)=\mathbf{R}1_{col}+\mathbf{R}\varepsilon_{col}^1+\mathbf{R}\varepsilon_{col}^2=$$
$$=\mathbf{R}\cdot\mathbf{e}_{lu}+\mathbf{C}\cdot\mathbf{E}_{ch}=\mathbf{R}\oplus\mathbf{C},$$

where $\mathbf{e}_{lu}=\left(1+\varepsilon+\varepsilon^2\right)/3$, $\mathbf{E}_{ch}=\left(1+\omega_3\varepsilon+\omega_3^2\varepsilon^2\right)/3$ are orthogonal «real» and «complex» idempotents, respectively, with

$$\mathbf{e}_{lu}^2=\mathbf{e}_{lu},\ \mathbf{E}_{ch}^2=\mathbf{E}_{ch},\ \ \mathbf{e}_{lu}\mathbf{E}_{ch}=\mathbf{E}_{ch}\mathbf{e}_{lu}=0$$

and $\omega_3:=e^{\frac{2\pi}{3}}$. Therefore, every color number $\mathcal{C}=x+y\varepsilon\ +z\varepsilon^2$ is a linear combination $\mathcal{C}=a_{lu}\cdot\mathbf{e}_{lu}+\mathrm{z}_{ch}\cdot\mathbf{E}_{ch}=\left(a_{lu},\mathrm{z}_{ch}\right)$ of the «scalar» $a_{lu}\cdot\mathbf{e}_{lu}$ and «complex» parts $\mathrm{z}_{ch}\cdot\mathbf{E}_{ch}$ in the idempotent basis $\left\{\mathbf{e}_{lu},\mathbf{E}_{ch}\right\}$. We will call the real numbers $a_{lu}\in\mathbf{R}$ the *luminance (intensity) numbers*, and we will call the complex numbers $\mathrm{z}_{ch}=b+jc\in\mathbf{C}$ the *chromaticity numbers*. For this reason, we can consider a color image in the two presentations (formats). The first presentation

$$\vec{\mathbf{f}}_{Col}(x,y)=\left(f_R(x,y),f_G(x,y),f_B(x,y)\right)=\quad(2)$$
$$=f_R(x,y)1+f_G(x,y)\varepsilon+f_B(x,y)\varepsilon^2,$$

is called the (R,G,B)-format and the second presentation

$$\mathbf{f}_{col}(x,y)=f_{lu}(x,y)\mathbf{e}_{lu}+f_{ch}(x,y)\mathbf{E}_{ch}=\quad(3)$$
$$=\left(f_{lu}(x,y),f_{ch}(x,y)\right)$$

is called the "luminance-chrominance" (LC) format. This format defines every pixel in terms of luminance grey-level part $f_{lu}(x,y)$ and complex-valued chrominance part $\mathbf{f}_{ch}(x,y)$, where $|\mathbf{f}_{ch}(x,y)|$ is saturation and $\mathbf{arg}(\mathbf{f}_{ch}(x,y))$ is hue of $f^{col}(x,y)$. In the second form we have separated the color image into two terms: the *luminance* (intensity) term $f_{lu}(x,y)$ and the *chromacity term* $\mathrm{f}_{ch}(x,y)$ (color information). Obviously,

$$\|\mathbf{f}_{col}(x,y)\|_L=\iint_{(x,y)\in\mathbf{R}^2}\|\mathbf{f}_{col}(x,y)\|_{\mathcal{A}_3^{col}}\,dxdy$$

is the norm of a color image $\mathbf{f}_{col}(x,y)$.

One of the primary applications of this work could be in edge detection and color image compression. For the edge detection, we convolve the color (3×3)-masks $M_{col}(i,j)$ with color image $\mathrm{f}_{col}(i,j)$. We use *color Prewitt's-like* masks for detection of horizontal, vertical, and diagonal edges. As entries instead of real numbers these masks have triplet numbers:

$$M_{col}^H=\begin{bmatrix}1&\varepsilon&\varepsilon^2\\0&0&0\\-1&-\varepsilon&-\varepsilon^2\end{bmatrix},\ M_{col}^V=\begin{bmatrix}1&0&-1\\\varepsilon&0&-\varepsilon\\\varepsilon^2&0&-\varepsilon^2\end{bmatrix},$$

$$M_{col}^{RD}=\begin{bmatrix}0&1&\varepsilon\\-1&0&\varepsilon^2\\-\varepsilon&-\varepsilon^2&0\end{bmatrix},\ M_{col}^{LD}=\begin{bmatrix}\varepsilon&\varepsilon^2&0\\1&0&-\varepsilon^2\\0&-1&-\varepsilon\end{bmatrix},$$

We see that triplet color detector is realized without multiplications. Figure 1 shows result of color edge detecting.

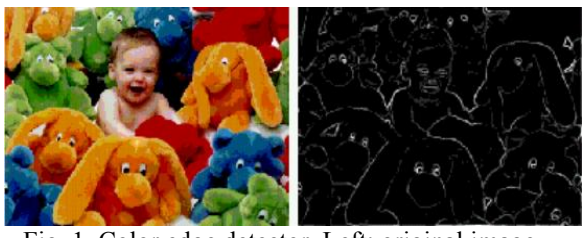


Fig. 1. Color edge detector. Left: original image, right:detected edges.

Ortho-unitary transforms

Generalized Fourier analysis based on orthogonal and unitary transforms plays an important role in DIP. Transforms, notable the classical Discrete Fourier Transform (DFT), are extensively used in digital image filtering and in power spectrum estimation. Other Fourier transforms - e.g. the discrete cosine/sine transforms (DCT/DST), wavelet transforms - are frequently employed in digital image compression. All the above-mentioned transforms are used in single-channel digital image processing. A natural question that arises in our approach is the definition of color and multicolor (RGB-channel and multichannel) transforms that can be used efficiently in color and multicolor image processing.

In our work, we ask the question, «What can we ascertain about image when the range of the image function is viewed as hypercomplex algebra?» Working over past several years, we have begun to develop a rudimentary understanding [2-4]. We have focused on the development of the color wavelet transforms, design of fast serial and parallel algorithms for realizing these transforms, and efficient implementations.

2D discrete color (*N*x*N*)-image $\left[\mathrm{f}^{col}(i,j)\right]_{i,j=1}^{N}$ can be defined as a 2D (*N*x*N*)-array in the (R,G,B) or (LC) formats

$$\mathbf{f}^{col}(i,j):\mathbf{Z}_N^2\to\mathcal{A}_3^{col},$$
$$\mathbf{f}^{col}(i,j):\mathbf{Z}_N^2\to\mathbf{R}\oplus\mathbf{C}.$$

Here, every color pixel $\mathbf{f}^{col}(i,j)$ at position (i,j) is a triplet number in (R,G,B)- or LC-formats, respectively. Generalized Fourier analysis based on orthogonal and unitary transforms plays an important role in digital signal/image processing as currently practiced. A natural question that arises in our approach is the definition of triplet-valued transforms that can be used efficiently in digital color image compression. In this work so-called ortho-unitary transforms are introduced.

We say the operator $L_{2\mathrm{D}}\left[\mathbf{f}_{col}\right]=\mathbf{F}_{col}$

$$L_{2D}:\left(\mathcal{A}_k^{col}\right)^{N^2}\to\left(\mathcal{A}_k^{col}\right)^{N^2}$$

is ortho-unitary if it maintains the norm of color images $\left\|\mathbf{f}_{col}(i,j)\right\|_L$. It should be noted that orthogonal transforms keep the norm of real-valued (gray-level) images, unitary transforms keep the norm of complex-valued (bichromatic) images. For this reason, ortho-unitary transforms are a generalization of orthogonal and unitary transforms for color images. In LC format ortho-unitary transforms can be constructed with help an orthogonal O_{2D} and unitary U_{2D} transforms

$$L_{2D}=O_{2D}\mathbf{e}_{lu}+U_{2D}\mathbf{E}_{ch}=\left(O_{2D},U_{2D}\right)$$

The simplest form of ortho-unitary transform for image processing is a separable 2D transform formed from two 1D transforms by tensor product

$$L_{2D}=L_{1D}\otimes L_{1D}=$$
$$=\left(O_1\otimes O_2\right)\mathbf{e}_{lu}+\left(U_1\otimes U_2\right)\mathbf{E}_{ch},$$

where $\otimes$ is the symbol of tensor product. Using separable transforms reduces the problem of designing efficient ortho-unitary 2D transforms to a one-dimensional problem, and almost all current compression approach employ separable transforms. It is possible to use one pair of orthogonal and unitary transforms, when $O_1=O_2=O$ and $U_1=U_2=U$. In this case we obtain a wide family of ortho-unitary transforms:

$$L_{2D}=\left(O\otimes O\right)\mathbf{e}_{lu}+\left(U\otimes U\right)\mathbf{E}_{ch}$$

using different 1D orthogonal transforms $\mathsf{L}_{1D}=O\mathbf{e}_{lu}+U\mathbf{E}_{ch}$. Every pair (O,U) of an orthogonal O and an unitary U transforms generates ortho-unitary (triplet-valued) transform $\mathsf{L}_{1D}=O\mathbf{e}_{lu}+U\mathbf{E}_{ch}$.

Table 1. Table heading

	F	$\widetilde{\mathbf{W}}$	$\widetilde{\mathbf{Hd}}$	$\widetilde{\mathbf{Ht}}$
W	$(\mathbf{W},\mathbf{F})$	$\left(\mathbf{W},\widetilde{\mathbf{W}}\right)$	$\left(\mathbf{W},\widetilde{\mathbf{Hd}}\right)$	$\left(\mathbf{W},\widetilde{\mathbf{Ht}}\right)$
Hd	$(\mathbf{Hd},\mathbf{F})$	$\left(\mathbf{Hd},\widetilde{\mathbf{W}}\right)$	$\left(\mathbf{Hd},\widetilde{\mathbf{Hd}}\right)$	$\left(\mathbf{Hd},\widetilde{\mathbf{Ht}}\right)$
Ht	$(\mathbf{Ht},\mathbf{F})$	$\left(\mathbf{Ht},\widetilde{\mathbf{W}}\right)$	$\left(\mathbf{Ht},\widetilde{\mathbf{Hd}}\right)$	$\left(\mathbf{Ht},\widetilde{\mathbf{Ht}}\right)$
Hr	$(\mathbf{Hr},\mathbf{F})$	$\left(\mathbf{Hr},\widetilde{\mathbf{W}}\right)$	$\left(\mathbf{Hr},\widetilde{\mathbf{Hd}}\right)$	$\left(\mathbf{Hr},\widetilde{\mathbf{Ht}}\right)$
Wv	$(\mathbf{Wv},\mathbf{F})$	$\left(\mathbf{Wv},\widetilde{\mathbf{W}}\right)$	$\left(\mathbf{Wv},\widetilde{\mathbf{Hd}}\right)$	$\left(\mathbf{Wv},\widetilde{\mathbf{Ht}}\right)$

For example, the table 1 shows some possibilities, where $\mathbf{W}, \mathbf{Hd}, \mathbf{Ht}$, $\mathbf{Hr}, \mathbf{Wv}$ are Walsh, Hadamard, Hartley, Haar, and Wavelet orthogonal transforms, respectively, and $\mathbf{F}, \widetilde{\mathbf{W}}, \widetilde{\mathbf{Hd}}, \widetilde{\mathbf{Ht}}$ are Fourier, Complex Walsh, Complex Hadamard, Complex Hartley transforms. Some examples of basis color functions of color transforms are shown in Figures 2-4.

Conclusion

In this study we define new methods of constructing triplet–valued (color) transforms They are based on triplet algebra. Further work will be concentrated on application aspects of obtained results.

References

1. Ch. Greaves. On algebraic triplets. // Proc. Irish Acad. – 1847. – Vol. 3 - P. 51-54, 57-64, 80-84, 105-108.
2. V. Labunets. Hypercomplex algebras as unified language for image processing and pattern recognition. Part 1. Is the brain a Clifford algebra computer? // In this Proceedings.
3. E. L.-Rundblad, I. Nikitin and V. Labunets. Unified Approach to Fourier-Clifford-Prometheus Sequences, Transforms and Filter Banks // Computational Noncommuta-tive Algebra and Applications, NATO Science Series.(Editors: J. Byrnes), Kluwer Academic Publishers, London.- 2003. - 2003. - P. 389-400.
4. V.G. Labunets, E.V. Rundblad and J.Astola. Is the Brain a «Clifford algebra quantum computer»? // Applied Geometrical Algebras in Computer Science and Engineering. (L.Dorst, C.Doran, J.Lasenby, Editots), Birkhauser. – 2002. - P. 486-495.
5. V.G. Labunets. Clifford Algebras as Unified Language for Multicolor Image Processing and Pattern Recognition // Computational Noncommuta-tive Algebra and Applications, NATO Science Series.(Editors: J. Byrnes), Kluwer Academic Publishers, London.- 2003. - 2003.- P. 197-226.
6. V.G.Labunets, E.V.Rundblad, J.Astola.. Is the brain «Clifford algebra quantum computer? // Proc. of SPIE «Materials and Devices for Photonic Circuits». – 2001. - Vol. 4453. - P 134-145

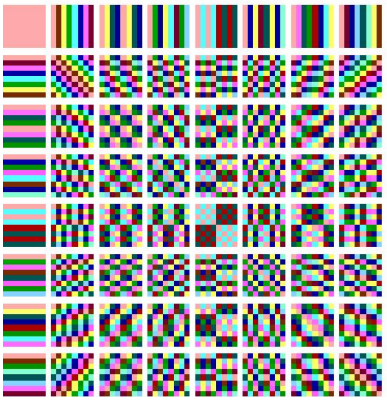


Fig. 2. 2D Color Walsh-Fourier $\mathbf{W} \cdot \mathbf{e}_{lu} + \mathbf{F} \cdot \mathbf{E}_{ch}$ 2D basis functions.

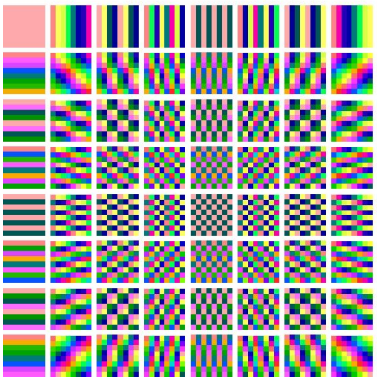


Fig. 3. 2D Color Hartley-Fourier $\mathbf{Ht} \cdot \mathbf{e}_{lu} + \mathbf{F} \cdot \mathbf{E}_{ch}$ 2D basis functions.

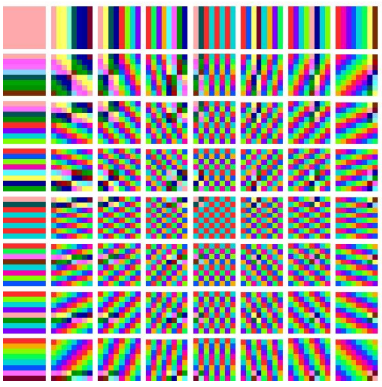


Fig. 4. Color Haar-Fourier $\mathbf{W} \cdot \mathbf{e}_{lu} + \mathbf{F} \cdot \mathbf{E}_{ch}$ 2D basis functions.

HYPERCOMPLEX ALGEBRAS AS UNIFIED LANGUAGE FOR IMAGE PROCESSING AND PATTERN RECOGNITION.

PART 3. HYPERSPECTRAL IMAGE PROCESSING[1]

V. Labunets[2,3], D. Gainanov[2,4], D. Berenov[2,5]

[2] Urals Federal University, Ekaterinburg, Russia,
[3] vlabunets05@yahoo.com, [4] damir@dc.ru, [5] berenov@dc.ru

We developed a novel algebraic approach based on hypercomplex and Clifford algebras to color, multicolor and hyperspectral image processing. A family of discrete Color-valued, multicolor-valued and hyperspectral-valued 2D Fourier and Wavelet transforms has been presented that can be used in color, multicolor and hyperspectral image compression.

Introduction

We develop a conceptual framework, design methodologies and software-tool for color, multicolor and hyperspectral image processing systems with assessment capability. Multi-color and hyper-spectral images are composed of a series of images it different optical bands at wavelengths $\lambda_1, \lambda_2, ..., \lambda_K$, called spectral channels, where K is the number of different optical channels. Hyper-spectral cameras, based on electronically tunable filters (ETF), produce a stack of images at a sequence of channels, forming the familiar "3D image cube" of (x, y, λ)-dimensions (2 spatial (x, y)-dimensions and a third spectral λ-axis). Each pixel in the image (x, y, λ)-cube, therefore, represents the spectrum at the wavelengths $\lambda_1, \lambda_2, ..., \lambda_K$ of the observed scene at point (x, y). Multi-color or hyper-spectral image can be characterized as an image consisting of K dimension (K-D) vector-valued pixels:

$$\vec{\mathbf{f}}_{Mcol}(x, y) = \left(f_1(x, y), f_2(x, y), ..., f_K(x, y)\right)$$

A common example of such a data set is color (3-channel or 3-band) image

$$\vec{\mathbf{f}}_{Col}(x, y) = \left(f_R(x, y), f_G(x, y), f_B(x, y)\right),$$

where each pixel is characterized by three scalars representing its red $f_R(x, y)$, green $f_G(x, y)$ and blue $f_B(x, y)$ components. A second example is hyper-spectral image gathered from remote sensing platforms. These systems gather data from a single area of the earth's surface at K of optical wavelengths (channels). For some systems each pixel $\vec{\mathbf{f}}(x, y)$ can consist of some hundreds spectral components. If the image is composed of few spectral channels (less than 10) it is called multi-spectral while the image consisting of a several tens or hundreds of bands is called hyper-spectral. In rest of the invention we use the term *multi-spectral* (or *K-channel*) *images* to mean both *multi-color* and *hyper-spectral images.*

Our approach to color, multicolor (MC) and hyperspectral (HS) image processing based on hypercomplex and Clifford algebras. In the algebraic-geometrical approach, each color or multicolor pixel is considered not as a k D vector, but as a k D hypercomplex number (k is the number of image spectral channels). Real and complex numbers (from signal/image processing) can be extended to the generalized complex numbers we have named *hypercomplex*. The theory of numbers has always been a quite fascinating and rich subject, especially when these numbers are described in terms of algebra. Along these lines, hypercomplex algebras play a central part in the theoretical physics, quantum computing, signal/image processing, as well as

[1]This work was supported by RFBR (grants № 13-07-12168, BR № 13-07-00785) and by MES RF grant № 218-03-167

in mathematics. And among of different types of hypercomplex numbers the sets of so-called *triplet* (*multiplet*) and Clifford numbers has a special status: mainly because it is a 3D and k D algebras of color and multicolor (hyperspectral) numbers. Thus these algebras have been a fundamental tool in the color, multicolor and hyperspectral image processing theory. In this context, the full machinery of grey-level signal processing theory can be transposed in C-, MC- and HS-image processing theory.

A natural question that arises in our approach is the definition of color and multicolor (RGB-channel and multichannel) transforms that can be used efficiently in color and multicolor image processing. We propose a wide library of so-called multiorthounitary transforms for using in image processing and pattern recognition applications.

Multiplet (multicolor) algebra

The multicomponent (multicolor and hyperspectral) images are usually interpreted as k D vectors

$$\mathbf{f}_{Mcol}(x,y)=\left(f_1(x,y),f_2(x,y),\ldots,f_k(x,y)\right)$$

We will interpret multicomponent images as multiplet-valued signals

$$\begin{gathered}\mathbf{f}_{Mcol}(x,y)= \\ =\left(f_1(x,y),f_2(x,y),\ldots,f_k(x,y)\right)= \\ =f_R(x,y)1+f_G(x,y)\varepsilon+\ldots+f_B(x,y)\varepsilon^{k-1},\end{gathered} \tag{1}$$

which take values in the multiplet (multicolor) algebra

$$\begin{gathered}\mathcal{A}_k^{Mcol}=\left(\mathbf{R}\middle|1,\varepsilon^1,\varepsilon^2,..,\varepsilon^{k-1}\right)= \\ =\mathbf{R}\cdot 1+\mathbf{R}\cdot\varepsilon^1+\mathbf{R}\cdot\varepsilon^2+\ldots+\mathbf{R}\cdot\varepsilon^{k-1},\end{gathered}$$

where $\varepsilon^0,\varepsilon^1,\varepsilon^2,\ldots,\varepsilon^{k-1}$ are hyperimaginary units and $\varepsilon^k=1$. Multiplet numbers are represented in its basic form by

$$\mathcal{M}=a_0+a_1\varepsilon^1+a_2\varepsilon^2+\ldots+a_{k-1}\varepsilon^{k-1},$$

where $a_i\in\mathbf{R},\quad i=1,2,\ldots,k-1$.

The addition and multiplication of muliplet numbers $\mathcal{M}_1$ and $\mathcal{M}_2$ are given by

$$\begin{gathered}\mathcal{M}=\mathcal{M}_1+\mathcal{M}_2= \\ =\left(\sum_{n=0}^{k-1}a_n\varepsilon^n\right)\pm\left(\sum_{n=0}^{k-1}b_n\varepsilon^n\right)=\sum_{n=0}^{k-1}\left(a_n+b_n\right)\varepsilon^n.\end{gathered}$$

and

$$\begin{gathered}\mathcal{M}=\mathcal{M}_1\cdot\mathcal{M}_2= \\ =\left(\sum_{n=0}^{k-1}a_n\varepsilon^n\right)\cdot\left(\sum_{n=0}^{k-1}b_n\varepsilon^n\right)=\sum_{n=0}^{k-1}c_n\varepsilon^n,\end{gathered}$$

where

$$c_l=\sum_{m=0}^{k-1}a_{l\ominus m}b_m$$

It is easy to see that the multiplet product is isomorphic to k–point convolution

$$\begin{gathered}\mathcal{C}_1\mathcal{C}_2=(a_0,a_1,\ldots,a_{k-1})*(b_0,b_1,\ldots,b_{k-1})= \\ =(c_0,c_1,\ldots,c_{k-1}).\end{gathered}$$

The *multiplet conjugate* $\overline{\mathcal{M}}$ of a multiplet number $\mathcal{M}=\sum_{n=0}^{k-1}a_n\varepsilon^n$ is defined as

$$\overline{\mathcal{M}}=\overline{\sum_{n=0}^{k-1}a_n\varepsilon^n}=\sum_{n=0}^{k-1}a_n\varepsilon^{-n}.$$

The norm $\left\|\mathcal{M}\right\|_2=\mathcal{M}\overline{\mathcal{M}}$ is given by

$$\begin{gathered}\left\|\mathcal{M}\right\|_{\mathcal{A}_k^{Mcol}}=\mathcal{M}\overline{\mathcal{M}}=\left(\sum_{n=0}^{k-1}a_n\varepsilon^n\right)\left(\sum_{m=0}^{k-1}a_n\varepsilon^{-m}\right)= \\ =\left(\sum_{n=0}^{k-1}a_n\varepsilon^n\right)\left(\sum_{m=0}^{k-1}a_n\varepsilon^{-m}\right).\end{gathered}$$

The algebra $\mathcal{A}_k^{Mcol}=\left(\mathbf{R}\middle|1,\varepsilon^1,\ldots,\varepsilon^{k-1}\right)$ is the direct sum of the real **R** and complex **C** fields:

$$\begin{gathered}\mathcal{A}_k^{Mcol}\left(\mathbf{R}\right)=\mathbf{R}^{k_{lu}}\oplus\mathbf{C}^{k_{ch}}= \\ =\begin{cases}\mathrm{R}\cdot\mathbf{e}_{lu}^1+\mathrm{R}\cdot\mathbf{e}_{lu}^2+\sum_{j=1}^{\frac{k}{2}-1}\mathbf{C}\cdot\mathbf{E}_{ch}^j, & \text{if } k \text{ even},\\ \mathrm{R}\cdot\mathbf{e}_{lu}^1+\sum_{j=1}^{\frac{k-1}{2}}\mathbf{C}\cdot\mathbf{E}_{ch}^j, & \text{if } k \text{ odd},\end{cases}\end{gathered}$$

where $\mathbf{e}_{lu}^i$ and $\mathbf{E}_{ch}^j$ are "real" and "complex" orthogonal idempotents, respectively, with

$$\left(\mathbf{e}_{lu}^{i}\right)^{2}=\mathbf{e}_{lu}^{i},\ \left(\mathbf{E}_{ch}^{j}\right)^{2}=\mathbf{E}_{ch}^{j},\ \mathbf{e}_{lu}^{i}\mathbf{E}_{ch}^{j}=\mathbf{E}_{ch}^{j}\mathbf{e}_{lu}^{i}.$$

Therefore, every multiplet (multicolor) number $\mathcal{M}$ is a linear combination of k_{lu} «scalar» parts and k_{ch} «complex» parts:

$$\mathcal{M}=\sum_{i=1}^{k_{lu}}\left(a_i\cdot\mathbf{e}_{lu}^{i}\right)+\sum_{j=1}^{k_{ch}}\left(z_j\cdot\mathbf{E}_{ch}^{j}\right),$$

where $k_{lu}=1,2$ and $k_{ch}=\frac{k}{2},\frac{k-1}{2}$ if k is odd or even, respectively.

The real numbers $a_i\in\mathbf{R}$ are called the *multi-intensity numbers* and complex numbers $\mathbf{z}_j=b+ic\in\mathbf{C}$ are called the *multi-chromacity numbers*. For this reason, we can consider a multicolor image in the two presentations (formats). The first presentation

$$\vec{\mathbf{f}}_{Mcol}(x,y)=$$
$$=f_R(x,y)1+f_G(x,y)\varepsilon^{1}+...+f_B(x,y)\varepsilon^{k-1}$$

is called the multiplet-format and the second presentation

$$\mathbf{f}_{Mcol}(x,y)=$$
$$=\sum_{i=1}^{k_{lu}}\left[f_{lu}^{i}(x,y)\cdot\mathbf{e}_{lu}^{i}\right]+\sum_{j=1}^{k_{ch}}\left[\mathbf{f}_{ch}^{j}(x,y)\cdot\mathbf{E}_{ch}^{j}\right]=$$
$$=\left(f_{lu}^{1}(x,y),...,f_{lu}^{k_{lu}}(x,y);\ \mathbf{f}_{ch}^{1}(x,y),...,\mathbf{f}_{ch}^{k_{ch}}(x,y)\right)$$

is called the "multiluminance-chrominance" (MLC) format [1-2]. This format defines every pixel in terms of k_{lu} luminance gray-level parts $\left(f_{lu}^{1}(\mathbf{x}),...,f_{lu}^{k_{lu}}(\mathbf{x})\right)$ and k_{ch} complex-valued chrominance parts $\left(\mathbf{f}_{ch}^{1}(\mathbf{x}),...,\mathbf{f}_{ch}^{k_{ch}}(\mathbf{x})\right)$, where

$$\left|\mathbf{f}_{ch}^{1}(x,y)\right|,\left|\mathbf{f}_{ch}^{2}(x,y)\right|,...,\left|\mathbf{f}_{ch}^{k_{ch}}(x,y)\right|$$

are multisaturations and

$$\arg\left\{\mathbf{f}_{ch}^{1}(x,y)\right\},\arg\left\{\mathbf{f}_{ch}^{2}(x,y)\right\},..,\arg\left\{\mathbf{f}_{ch}^{k_{ch}}(x,y)\right\}$$

are multihues of multicomponent image $\mathbf{f}_{Mcol}(x,y)$. Obviously,

$$\left\|\mathbf{f}_{Mcol}(x,y)\right\|_{L}=\iint_{(x,y)\in\mathbf{R}^2}\left\|\mathbf{f}_{Mcol}(x,y)\right\|_{\mathcal{A}_k^{Mcol}}dxdy$$

is the norm of a color image $\mathbf{f}_{Mcol}(x,y)$.

Ortho-unitary transforms

2D discrete multicomponent (NxN)-image $\left[\mathbf{f}^{Mcol}(i,j)\right]_{i,j=1}^{N}$ can be defined as a 2D (NxN)-array in the multiplet-format or in the multiluminance-chrominance formats

$$\mathbf{f}^{Mcol}(i,j):\mathbf{Z}_N^2\to\mathcal{A}_k^{Mcol},$$
$$\mathbf{f}^{Mcol}(i,j):\mathbf{Z}_N^2\to\mathbf{R}^{k_{lu}}\oplus\mathbf{C}^{k_{ch}}.$$

Here, every multi-spectral pixel $\mathbf{f}^{Mcol}(i,j)$ at position (i,j) is a multiplet-valued number in the multiplet-format or in the multiluminance-chrominance formats, respectively.

We say the operator $\mathcal{L}_{2D}\left[\mathbf{f}^{Mcol}\right]=\mathbf{F}^{Mcol}$

$$\mathcal{L}_{2D}:\left(\mathcal{A}_k^{Mcol}\right)^{N^2}\to\left(\mathcal{A}_k^{Mcol}\right)^{N^2}$$

is ortho-unitary if it keeps the norm $\left\|\mathbf{f}_{Mcol}(x,y)\right\|_{L}$ of color images. It should be noted that 1) orthogonal transforms keep the norm of real-valued (gray-level) images, 2) unitary transforms keep the norm of complex-valued (bichromatic) images and ortho-unitary transforms keep the norm of color images [??]. For this reason, multiortho-unitary transforms are a generalization of orthogonal, unitary and orto-unitary transforms for MC-and HS-images.

The simplest form of ortho-unitary transform for image processing is a separable 2D transform formed from 1D transforms by tensor product $\mathcal{L}_{2D}=\mathcal{L}_{1D}\otimes\mathcal{L}_{1D}$, where

$$\mathcal{L}_{1D}=\left[\bigoplus_{i=1}^{k_{lu}}O_{1D}^{i}\mathbf{e}_{lu}^{i}\right]+\left[\bigoplus_{j=1}^{k_{ch}}U_{1D}^{j}\mathbf{E}_{ch}^{j}\right]=$$
$$=\left(O_{1D}^{1},\ldots,O_{1D}^{k_{lu}};U_{1D}^{1},\ldots,U_{1D}^{k_{ch}}\right),$$
$$\mathcal{L}_{1D}=\sum_{n=0}^{k-1}L_{1D}^{n}\varepsilon^{n}$$

in GLC and multiplet formats, respectively.

In MLC format multiortho-unitary transforms can be constructed with help an orthogonal k_{lu} orthogonal $O_{2D}^{1},,...,O_{2D}^{k_{lu}}$ and k_{ch} unitary $U_{2D}^{1},U_{2D}^{2},...,U_{2D}^{k_{ch}}$ transforms

$$\mathcal{L}_{1D}=\left[\bigoplus_{i=1}^{k_{lu}}O_{1D}^{i}\mathbf{e}_{lu}^{i}\right]+\left[\bigoplus_{j=1}^{k_{ch}}U_{1D}^{j}\mathbf{E}_{ch}^{j}\right]=$$
$$=\left(O_{1D}^{1},\ldots,O_{1D}^{k_{lu}};U_{1D}^{1},\ldots,U_{1D}^{k_{ch}}\right),$$

In this work we use the following 1D *m*-plet valued transforms

$$\mathcal{S}_{1D} = \left[\bigoplus_{i=1}^{k_{lu}} O\Delta^i \mathbf{e}_{lu}^i\right] + \left[\bigoplus_{j=1}^{k_{ch}} O\boldsymbol{\Delta}^j \mathbf{E}_{ch}^j\right] =$$

$$= O\left\{\left[\bigoplus_{i=1}^{k_{lu}} \Delta^i \mathbf{e}_{lu}^i\right] + \left[\bigoplus_{j=1}^{k_{ch}} \boldsymbol{\Delta}^j \mathbf{E}_{ch}^j\right]\right\} = O\tilde{\Delta}$$

where $\left(\Delta^1, \Delta^2, \ldots, \Delta^{k_{lu}}\right)$ are k_{lu} real-valued diagonal matrices:

$$\Delta^s = diag(a_0^s, a_1^s, \ldots, a_{N-1}^s), \quad s = 1, 2, \ldots, k_{lu},$$

$\left(\boldsymbol{\Delta}^{k_{lu}+2}, \ldots, \boldsymbol{\Delta}^{k_{lu}+k_{ch}}\right)$ are k_{ch} diagonal complex matrices:

$$\boldsymbol{\Delta}^t = diag(\mathbf{z}_0^t, \mathbf{z}_1^t, \ldots, \mathbf{z}_{N-1}^t), \quad t = 1, 2, \ldots, k_{ch}.$$

and

$$\tilde{\boldsymbol{\Delta}} = diag(\mathcal{M}_0, \mathcal{M}_1, \ldots, \mathcal{M}_{N-1})$$

is triplet-valued diagonal matrix, where

$$\mathcal{M}_r = \sum_{i=1}^{k_{lu}} \left[a_r^i \cdot \mathbf{e}_{lu}^s\right] + \sum_{j=1}^{k_{ch}} \left[\mathbf{z}_r^j \cdot \mathbf{E}_{ch}^j\right],$$

and $r = 0, 1, \ldots, N-1$.

The more "simplest" form of triplet numbers $\mathcal{M}_r$ is $\mathcal{M}_r = \varepsilon^{i_r}$. In this case $\tilde{\Delta} = diag(\varepsilon^{i_0}, \varepsilon^{i_1}, \ldots, \varepsilon^{i_{N-1}})$. Note that the product of $\mathcal{M}_r = \varepsilon^{i_r}$ with a multicolor pixel $\vec{\mathbf{f}}_{Mcol}(x, y) = \sum_{r=0}^{k-1} f_r(x, y)\varepsilon^r$ is realized without multiplications as right shift of color components

$$\mathcal{M}_r \vec{\mathbf{f}}_{Mcol}(x, y) = \varepsilon^{i_r} \vec{\mathbf{f}}_{Mcol}(x, y) =$$

$$= \sum_{r=0}^{k-1} f_r(x, y)\varepsilon^{r+i_r} = \sum_{s=0}^{k-1} f_{s-i_r}(x, y)\varepsilon^s$$

In this case the situation is the same as for number theoretical transforms. For this reason, multiplication complexity of diagonal matrix $\tilde{\Delta} = diag(\varepsilon^0, \varepsilon^1, \ldots, \varepsilon^{N-1})$ is equal to zero.

Conclusion

We developed a novel algebraic approach based on hypercomplex algebras to hyperspectral image processing. A family of discrete hyperspectral-valued 2D Fourier transforms has been presented that can be used in image compression and pattern recognition.

References

1. V. Labunets. Hypercomplex algebras as unified language for image processing and pattern recognition. Part 1. Is the brain a Clifford algebra computer? // In this Proceedings.
2. V. Labunets, D. Gainanov, D. Berenov. Hypercomplex algebras as unified language for image processing and pattern recognition. Part 2. Color image processing // In this Proceedings.
3. V.G. Labunets, E.V. Rundblad and J.Astola. Is the Brain a «Clifford algebra quantum computer»? // Applied Geometrical Algebras in Computer Science and Engineering. (L.Dorst, C.Doran, J.Lasenby, Editots), Birkhauser. – 2002. - P. 486-495.

MULTIPARAMETRIC WAVELET TRANSFORMS AND PACKETS[1]

V. Labunets[2,3], D. Gainanov[2,4], D. Berenov[2,5]

[2] Urals Federal University, Ekaterinburg, Russia,
[3] vlabunets05@yahoo.com, [4] damir@dc.ru, [5] berenov@dc.ru

The main goal of the paper is to show that wavelet transforms and packets have the *multiparametric representation* in the form of a product of the rotation Jacobi matrices. Each multiparameric wavelet transform (MPWT) depends on several free Jacobi parameters. When parameters are changed multiparametric transform is changed too taking form of all known and unknown orthogonal wavelet transforms. It gives unified approach to describing a wide set of cyclic orthogonal wavelet transforms and endows with adaptive properties of those transforms.

Introduction

The wide class of orthogonal wavelet transforms $\mathcal{WT}$ can be defined by two sets of coefficients 2, [1], [2]: h_0, h_1, …, h_{L-1} and g_0, g_1, …, g_{L-1}, where $L=2D$ is an even number. In fact $\mathcal{WT}$ is determined only by a set of h-coefficients h_0, h_1, …,. h_{L-1}, since the second set of coefficients is usually assigned according to the rule $g_0=h_{L-1}$, $g_1=-h_{L-2}$, …, $g_{L-1}=-h_0$. For this reason we will designate wavelet transform as

$$\mathcal{WT}_{2^n}\left[h_0,h_1,\ldots,h_{L-1}\right]= \\ =\prod_{r=1}^{n-m+1}\left[\mathcal{AT}_{2^{n-r+1}}[h_0,h_1,\ldots,h_{L-1}]\oplus \mathrm{I}_{2^n-2^{n-r+1}}\right], \quad (1)$$

where $m=\rceil \log_2 2D\lceil$ the smallest positive integer such that $2^{m-1}\le 2D\le 2^m$. For example.

$$\mathcal{WT}_{16}\left[h_0,h_1,h_2,h_3\right]=\left[\mathcal{AT}_4\oplus\mathbf{I}_{12}\right]\left[\mathcal{AT}_8\oplus\mathbf{I}_8\right]\left[\mathcal{AT}_{16}\right]=$$

$$=\left[\begin{pmatrix} h_0 & h_1 & h_2 & h_3 \\ h_2 & h_3 & h_0 & h_1 \\ g_0 & g_1 & g_2 & g_3 \\ g_2 & g_3 & g_0 & g_1 \end{pmatrix}\oplus\mathbf{I}_{12}\right]\cdot\left[\begin{pmatrix} h_0 & h_1 & h_2 & h_3 & & & & \\ & & h_0 & h_1 & h_2 & h_3 & & \\ & & & & h_0 & h_1 & h_2 & h_3 \\ h_2 & h_3 & & & & & h_0 & h_1 \\ g_0 & g_1 & g_2 & g_3 & & & & \\ & & g_0 & g_1 & g_2 & g_3 & & \\ & & & & g_0 & g_1 & g_2 & g_3 \\ g_2 & g_3 & & & & & g_0 & g_1 \end{pmatrix}\oplus\mathbf{I}_8\right].$$

$$\begin{pmatrix} h_0 & h_1 & h_2 & h_3 & & & & & & & & & & & & \\ & & h_0 & h_1 & h_2 & h_3 & & & & & & & & & & \\ & & & & h_0 & h_1 & h_2 & h_3 & & & & & & & & \\ & & & & & & h_0 & h_1 & h_2 & h_3 & & & & & & \\ & & & & & & & & h_0 & h_1 & h_2 & h_3 & & & & \\ & & & & & & & & & & h_0 & h_1 & h_2 & h_3 & & \\ & & & & & & & & & & & & h_0 & h_1 & h_2 & h_3 \\ h_2 & h_3 & & & & & & & & & & & & & h_0 & h_1 \\ g_0 & g_1 & g_2 & g_3 & & & & & & & & & & & & \\ & & g_0 & g_1 & g_2 & g_3 & & & & & & & & & & \\ & & & & g_0 & g_1 & g_2 & g_3 & & & & & & & & \\ & & & & & & g_0 & g_1 & g_2 & g_3 & & & & & & \\ & & & & & & & & g_0 & g_1 & g_2 & g_3 & & & & \\ & & & & & & & & & & g_0 & g_1 & g_2 & g_3 & & \\ & & & & & & & & & & & & g_0 & g_1 & g_2 & g_3 \\ g_2 & g_3 & & & & & & & & & & & & & g_0 & g_1 \end{pmatrix}.$$

Coefficients h_0, h_1, …, h_{L-1} depend upon each other. The coefficients, which we can change independently of one another will be called *angle-parameters* in this paper.

We will prove that multiparametric presentation of wavelet transform exists and that any orthogonal wavelet transform (WT) and wavelet packets (WP) depends on D angle-parameters φ_0, φ_1, …, φ_{D-1}:

$$\mathcal{WT}_{2^n}\left[h_0,h_1,\ldots,h_{L-1}\right]=\mathcal{WT}_{2^n}\left[\varphi_0,\varphi_1,\ldots,\varphi_{D-1}\right].$$

Multiparametric presentation of atomic WT

In order to find multiparametric form of WT we will use the Jacobi rotations. For that we should define the $(2^n\times 2^n)$ sparse rotation matrix on an angle φ in the plane spanned on i and j basis vectors:

[1]This work was supported by RFBR (grants № 13-07-12168, BR № 13-07-00785) and by MES RF grant № 218-03-167

$$\mathbf{CS}_{i,j}(\varphi)=\begin{array}{c} \\ i \\ \\ j \\ \\ \end{array}\begin{array}{c} \begin{array}{cccccc} i & & & j & & \end{array} \\ \left(\begin{array}{ccccccc} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & c & \cdots & s & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & -s & \cdots & c & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{array}\right) \end{array},$$

where $c=\cos(\varphi)$ and $s=\sin(\varphi)$.

As we see from (1) the WT $\mathcal{WT}_{2^n}$ is factorized into a product of sparse matrixes $\mathcal{AT}_{2^n}\left[h_0,h_1,\ldots,h_{L-1}\right]$, named stairs-like atomic WTs. We will multiply $\mathcal{AT}_{2^n}\left[h_0,h_1,\ldots,h_{L-1}\right]$ by $\mathbf{CS}_{i,j}(\varphi)$ matrix sequentially with such choice of angles φ that product $\mathbf{CS}_{i_k,j_k}(\varphi_k)\cdot\ldots\cdot\mathbf{CS}_{i_0,j_0}(\varphi_0)\cdot\mathcal{AT}_{2^n}\left[h_0,h_1,\ldots,h_{L-1}\right]$ will be permutation matrix or unit matrix.

As an example, we have taken the atomic Daubechies-6 (8×8)-matrix:

$$\mathcal{AT}_8\left[h_0,h_1,h_2,h_3,h_4,h_5\right]=$$

$$=\left(\begin{array}{cccccccc} h_0 & h_1 & h_2 & h_3 & h_4 & h_5 & & \\ & & h_0 & h_1 & h_2 & h_3 & h_4 & h_5 \\ h_4 & h_5 & & & h_0 & h_1 & h_2 & h_3 \\ h_2 & h_3 & h_4 & h_5 & & & h_0 & h_1 \\ h_5 & -h_4 & h_3 & -h_2 & h_1 & -h_0 & & \\ & & h_5 & -h_4 & h_3 & -h_2 & h_1 & -h_0 \\ h_1 & -h_0 & & & h_5 & -h_4 & h_3 & -h_2 \\ h_3 & -h_2 & h_1 & -h_0 & & & h_5 & -h_4 \end{array}\right). \qquad (2)$$

The angle φ_0 can be chosen such a way that the coefficient $h_5=0$ in the zeroth and fourth rows in the left product of matrix (2) by $\mathbf{CS}_{0,4}(\varphi_0)$. In this case coefficient h_4 will be zero in the same rows too:

$$\mathbf{CS}_{0,4}(\varphi_0)\cdot\mathcal{AT}_8\left[h_0,h_1,h_2,h_3,h_4,h_5\right]=$$

$$=\left(\begin{array}{cccccccc} h'_0 & h'_1 & h'_2 & h'_3 & \mathbf{0} & \mathbf{0} & & \\ & & h_0 & h_1 & h_2 & h_3 & h_4 & h_5 \\ h_4 & h_5 & & & h_0 & h_1 & h_2 & h_3 \\ h_2 & h_3 & h_4 & h_5 & & & h_0 & h_1 \\ \mathbf{0} & \mathbf{0} & h'_3 & -h'_2 & h'_1 & -h'_0 & & \\ & & h_5 & -h_4 & h_3 & -h_2 & h_1 & -h_0 \\ h_1 & -h_0 & & & h_5 & -h_4 & h_3 & -h_2 \\ h_3 & -h_2 & h_1 & -h_0 & & & h_5 & -h_4 \end{array}\right),$$

if the angle is chosen such that $c_0h_5-s_0h_0=0$, where $c_0=\cos(\varphi_0)$ and $s_0=\sin(\varphi_0)$. Hence,

$$\mathbf{CS}_{3,7}(\varphi_0)\cdot\mathbf{CS}_{2,6}(\varphi_0)\cdot\mathbf{CS}_{1,5}(\varphi_0)\cdot\mathbf{CS}_{0,4}(\varphi_0)\cdot$$
$$\cdot\mathcal{AT}_8\left[h_0,h_1,h_2,h_3,h_4,h_5\right]=$$

$$=\left(\begin{array}{cccccccc} c_0 & & & & s_0 & & & \\ & c_0 & & & & s_0 & & \\ & & c_0 & & & & s_0 & \\ & & & c_0 & & & & s_0 \\ -s_0 & & & & c_0 & & & \\ & -s_0 & & & & c_0 & & \\ & & -s_0 & & & & c_0 & \\ & & & -s_0 & & & & c_0 \end{array}\right)\cdot$$

$$\cdot\left(\begin{array}{cccccccc} h_0 & h_1 & h_2 & h_3 & \boldsymbol{h}_4 & \boldsymbol{h}_5 & & \\ & & h_0 & h_1 & h_2 & h_3 & \boldsymbol{h}_4 & \boldsymbol{h}_5 \\ \boldsymbol{h}_4 & \boldsymbol{h}_5 & & & h_0 & h_1 & h_2 & h_3 \\ h_2 & h_3 & \boldsymbol{h}_4 & \boldsymbol{h}_5 & & & h_0 & h_1 \\ \boldsymbol{h}_5 & -\boldsymbol{h}_4 & h_3 & -h_2 & h_1 & -h_0 & & \\ & & \boldsymbol{h}_5 & -\boldsymbol{h}_4 & h_3 & -h_2 & h_1 & -h_0 \\ h_1 & -h_0 & & & \boldsymbol{h}_5 & -\boldsymbol{h}_4 & h_3 & -h_2 \\ h_3 & -h_2 & h_1 & -h_0 & & & \boldsymbol{h}_5 & -\boldsymbol{h}_4 \end{array}\right)=$$

$$=\left(\begin{array}{cccccccc} h'_0 & h'_1 & h'_2 & h'_3 & & & & \\ & & h'_0 & h'_1 & h'_2 & h'_3 & & \\ & & & & h'_0 & h'_1 & h'_2 & h'_3 \\ h'_2 & h'_3 & & & & & h'_0 & h'_1 \\ & & h'_3 & -h'_2 & h'_1 & -h'_0 & & \\ & & & & h'_3 & -h'_2 & h'_1 & -h'_0 \\ h'_1 & -h'_0 & & & & & h'_3 & -h'_2 \\ h'_3 & -h'_2 & h'_1 & -h'_0 & & & & \end{array}\right)=$$

$$=\mathtt{AWT}_8\left[h'_0,h'_1,h'_2,h'_3\right].$$

As a result we get a new atomic matrix $\mathcal{AT}_8\left[h'_0,h'_1,h'_2,h'_3\right]$ with only four coefficients. To get the atomic matrix with two coefficients we should iterate foregoing procedure:

$$\mathbf{CS}_{0,7}(\varphi_1)\cdot\mathbf{CS}_{3,6}(\varphi_1)\cdot\mathbf{CS}_{2,5}(\varphi_1)\cdot\mathbf{CS}_{1,4}(\varphi_1)\cdot$$
$$\cdot\mathcal{AT}_8\left[h'_0,h'_1,h'_2,h'_3\right]=\mathcal{AT}_8\left[h''_0,h''_1\right].$$

Reiteration of this procedure on matrix $\mathcal{AWT}_8\left[h''_0,h''_1\right]$ results in:

$$\mathbf{CS}_{1,7}(\varphi_2)\cdot\mathbf{CS}_{0,6}(\varphi_2)\cdot\mathbf{CS}_{3,5}(\varphi_2)\cdot\mathbf{CS}_{2,4}(\varphi_2)\cdot$$
$$\cdot\mathcal{AT}_8\left[h''_0,h''_1\right]=\mathbf{P}_8, \qquad (3)$$

where $\mathbf{P}_8$ is a quasipermutation matrix (there are only +1 or −1 in every row and in every column of it). As the final result we get:

$$\left[\mathbf{CS}_{1,7}(\varphi_2)\cdot\mathbf{CS}_{0,6}(\varphi_2)\cdot\mathbf{CS}_{3,5}(\varphi_2)\cdot\mathbf{CS}_{2,4}(\varphi_2)\right]\cdot$$
$$\cdot\left[\mathbf{CS}_{0,7}(\varphi_1)\cdot\mathbf{CS}_{3,6}(\varphi_1)\cdot\mathbf{CS}_{2,5}(\varphi_1)\cdot\mathbf{CS}_{1,4}(\varphi_1)\right]\cdot$$
$$\cdot\left[\mathbf{CS}_{3,7}(\varphi_0)\cdot\mathbf{CS}_{2,6}(\varphi_0)\cdot\mathbf{CS}_{1,5}(\varphi_0)\cdot\mathbf{CS}_{0,4}(\varphi_0)\right]\cdot$$
$$\cdot\mathcal{AT}_8\left[h_0,h_1,h_2,h_3,h_4,h_5\right]=\mathbf{P}_8.$$

From here we obtain the multiparametric representation of the atomic WT matrix:

$$\mathcal{AT}_8\left[h_0,h_1,h_2,h_3,h_4,h_5\right]=$$

$$= \left[\mathbf{CS}_{3,7}(-\varphi_0)\cdot\mathbf{CS}_{2,6}(-\varphi_0)\cdot\mathbf{CS}_{1,5}(-\varphi_0)\cdot\mathbf{CS}_{0,4}(-\varphi_0)\right]\cdot$$
$$\cdot\left[\mathbf{CS}_{0,7}(\varphi_1)\cdot\mathbf{CS}_{3,6}(\varphi_1)\cdot\mathbf{CS}_{2,5}(\varphi_1)\cdot\mathbf{CS}_{1,4}(\varphi_1)\right]\cdot$$
$$\cdot\left[\mathbf{CS}_{1,7}(\varphi_2)\cdot\mathbf{CS}_{0,6}(\varphi_2)\cdot\mathbf{CS}_{3,5}(\varphi_2)\cdot\mathbf{CS}_{2,4}(\varphi_2)\right]\cdot\mathbf{P}_8 =$$

$$= \begin{pmatrix} c_0 & & & & -s_0 & & & \\ & c_0 & & & & -s_0 & & \\ & & c_0 & & & & -s_0 & \\ & & & c_0 & & & & -s_0 \\ +s_0 & & & & c_0 & & & \\ & +s_0 & & & & c_0 & & \\ & & +s_0 & & & & c_0 & \\ & & & +s_0 & & & & c_0 \end{pmatrix}\cdot$$

$$\cdot\begin{pmatrix} c_1 & & & & & & & -s_1 \\ & c_1 & & & -s_1 & & & \\ & & c_1 & & & -s_1 & & \\ & & & c_1 & & & -s_1 & \\ & +s_1 & & & c_1 & & & \\ & & +s_1 & & & c_1 & & \\ & & & +s_1 & & & c_1 & \\ +s_1 & & & & & & & c_1 \end{pmatrix}\cdot$$

$$\cdot\begin{pmatrix} c_2 & & & & & & -s_2 & \\ & c_2 & & & & & & -s_2 \\ & & c_2 & & -s_2 & & & \\ & & & c_2 & & -s_2 & & \\ & & +s_2 & & c_2 & & & \\ & & & +s_2 & & c_2 & & \\ +s_2 & & & & & & c_2 & \\ & +s_2 & & & & & & c_2 \end{pmatrix}\cdot\mathbf{P}_8 =$$

$$= \mathbf{T}_8^0(-\varphi_0)\cdot\mathbf{T}_8^1(-\varphi_1)\cdot\mathbf{T}_8^2(-\varphi_2)\cdot\mathbf{P}_8,$$

where $c_i = \cos(\varphi_i)$, $s_i = \sin(\varphi_i)$, $i = 0,1,2$ and every matrix $\mathbf{T}_8(\varphi_i)$ is the product of the following sparse rotation $\mathbf{CS}$-matrixes:

$$\mathbf{T}_8^0(\varphi_0) = \mathbf{CS}_{3,7}(\varphi_0)\mathbf{CS}_{2,6}(\varphi_0)\mathbf{CS}_{1,5}(\varphi_0)\mathbf{CS}_{0,4}(\varphi_0),$$
$$\mathbf{T}_8^1(\varphi_0) = \mathbf{CS}_{0,7}(\varphi_0)\mathbf{CS}_{3,6}(\varphi_0)\mathbf{CS}_{2,5}(\varphi_0)\mathbf{CS}_{1,4}(\varphi_0),$$
$$\mathbf{T}_8^2(\varphi_0) = \mathbf{CS}_{1,7}(\varphi_0)\mathbf{CS}_{0,6}(\varphi_0)\mathbf{CS}_{3,5}(\varphi_0)\mathbf{CS}_{2,4}(\varphi_0).$$

This result is general and valid for any $(2^r\times 2^r)$ atomic matrix:

$$\mathbf{P}_{2^r} = \left(\prod_{i=0}^{D-1}\mathbf{T}_{2^r}^i(\varphi_i)\right)\cdot\mathcal{AT}_{2^r}\left[h_0,h_1,\ldots,h_{2D-1}\right]$$

and, therefore,

$$\mathcal{AT}_{2^r}\left[h_0,h_1,\ldots,h_{2D-1}\right] = \left(\prod_{i=D-1}^{0}\mathbf{T}_{2^r}^i(-\varphi_i)\right)\mathbf{P}_{2^r}. \qquad (4)$$

It is the multiparametric representation of the atomic WT matrix.

Multiparametric representations of WT and WP

Let's begin with consideration of (16×16) Daubechies-4 wavelet transform. In the matrix form it is the product of the following atomic matrixes:

$$\mathcal{WT}_{16}\left[h_0,h_1,h_2,h_3\right] = \left[\mathcal{AT}_4\oplus\mathbf{I}_{12}\right]\cdot\left[\mathcal{AT}_8\oplus\mathbf{I}_8\right]\cdot\left[\mathcal{AT}_{16}\right].$$

Every atomic matrix $\mathcal{AT}_4$, $\mathcal{AT}_8$, $\mathcal{AT}_{16}$ can be represented in multiparametric form:

$$\mathcal{AT}_4 = \mathbf{T}_4^0(-\varphi_0)\mathbf{T}_4^1(-\varphi_1)\mathbf{P}_4,$$
$$\mathcal{AT}_8 = \mathbf{T}_8^0(-\varphi_0)\mathbf{T}_8^1(-\varphi_1)\mathbf{P}_8,$$
$$\mathcal{AT}_{16} = \mathbf{T}_{16}^0(-\varphi_0)\mathbf{T}_{16}^1(-\varphi_1)\mathbf{P}_{16}.$$

Therefore,

$$\mathcal{WT}_{16}\left[h_0,h_1,h_2,h_3\right] = \mathcal{WT}_{2^n}\left[\varphi_0,\varphi_1\right] =$$
$$= \left[\mathbf{T}_4^0(-\varphi_0)\mathbf{T}_4^1(-\varphi_1)\mathbf{P}_4\oplus\mathbf{I}_{12}\right]\cdot\left[\mathbf{T}_8^0(-\varphi_0)\mathbf{T}_8^1(-\varphi_1)\mathbf{P}_8\oplus\mathbf{I}_8\right]$$
$$\cdot\left[\mathbf{T}_{16}^0(-\varphi_0)\mathbf{T}_{16}^1(-\varphi_1)\mathbf{P}_{16}\right] =$$

$$= \left[\left(\prod_{i=1}^{0}\mathbf{T}_4^i(-\varphi_i)\right)\mathbf{P}_4\oplus\mathbf{I}_{12}\right]\cdot\left[\left(\prod_{i=1}^{0}\mathbf{T}_8^i(-\varphi_i)\right)\mathbf{P}_8\oplus\mathbf{I}_8\right]\cdot$$
$$\cdot\left[\left(\prod_{i=1}^{0}\mathbf{T}_{16}^i(-\varphi_i)\right)\mathbf{P}_{16}\right].$$

It is two-parametric form of Daubechies-4 WT. It is possible to obtain all the transforms of $\mathcal{WDT}_{16}\left[h_0,h_1,h_2,h_3\right]$-type by changing the angles φ_0 and φ_1.

All the atomic matrices in multiparametric representation of WT are characterized by the same set of angle-parameters. And all the angles have equal values in each atomic matrix and have to be chosen synchronously. Of course, it is possible to use different angles sets in different atomic matrixes and to change them not synchronously, but in this case we will get heterogeneous WT.

The most general expression for multiparametric presentation of WT is the following:

$$\mathcal{WT}_{2^n}[h_0,h_1,\ldots,h_{2D-1}] = \mathcal{WT}_{2^n}\left[\varphi_0,\varphi_1,\ldots,\varphi_{D-1}\right] =$$
$$= \prod_{r=1}^{n-m+1}\left[\left(\prod_{i=D-1}^{0}\mathbf{T}_{2^{n-r+1}}^i(-\varphi_i)\right)\mathbf{P}_{2^{n-r+1}}\oplus\mathbf{I}_{2^n-2^{n-r+1}}\right], \qquad (5)$$

where $\underset{2^{n-r}}{\oplus}$ is addition modulo 2^{n-r}. The last expression presents any WT in multiparametric form. For example, for $\mathcal{WT}_{16}\left[h_0,h_1,\ldots,h_5\right]$ from (5) we have

$$\begin{aligned}\mathcal{WT}_{16}[h_0,...,h_5]&=\mathcal{WT}_{16}[\varphi_0,\varphi_1,\varphi_2]=\\ &=\left[\mathcal{AT}_8[h_0,...,h_5]\oplus\mathbf{I}_8\right]\cdot\mathcal{AT}_{16}[h_0,...,h_5]=\\ &=\left[(-1)\cdot\mathbf{P}_8\cdot\left[\mathbf{T}_8^0(\varphi_0)\,\mathbf{T}_8^1(\varphi_1)\,\mathbf{T}_8^2(\varphi_2)\right]\cdot\mathbf{C}_8^2\oplus\mathbf{I}_8\right]\cdot\\ &\cdot\left[(-1)\cdot\mathbf{P}_{16}\cdot\left[\mathbf{T}_{16}^0(\varphi_0)\,\mathbf{T}_{16}^1(\varphi_1)\,\mathbf{T}_{16}^2(\varphi_2)\right]\cdot\mathbf{C}_{16}^2\right].\end{aligned}\quad(6)$$

The classical WT with coefficients $h_0,h_1,...,h_{2D-1}$ is constructed from atomic WT according to the following rule:

$$\begin{aligned}&\mathcal{WT}_{2^n}[h_0,h_1,...,h_{2D-1}]=\\ &=\prod_{r=1}^{n-m+1}\left[\mathcal{AT}_{2^{n-r+1}}[h_0,h_1,...,h_{2D-1}]\oplus\mathbf{I}_{2^n-2^{n-r+1}}\right].\end{aligned}\quad(7)$$

The atomic transform is used only once within each iteration in (6). In fact, the atomic transform could be repeated not more then $2^n/2^{n-r+1}=2^{r-1}$ times. Let $\mathbf{s}^r=(s_1^r,s_2^r,...,s_t^r,...,s_{2^{r-1}}^r)$ be a binary 2^{r-1}-digital integer. Every binary digit s_t^r controls the t^{th} position of the matrix $\mathcal{AWT}_{2^{n-r+1}}$ in the r^{th} iteration sparse matrix

$$\mathcal{AT}_{2^{n-r-1}}^{s_t^r}=\begin{cases}\mathcal{AT}_{2^{n-r-1}}, & s_t^r=1,\\ I_{2^{n-r-1}}, & s_t^r=0.\end{cases}$$

All such matrices form a packet of atomic matrices

$$\begin{aligned}\mathcal{AP}_{2^n}^{\mathbf{s}^r}&=\bigoplus_{t=1}^{2^{r-1}}\mathcal{AT}_{2^{n-r+1}}^{s_t^r}=\\ &=\mathcal{AT}_{2^{n-r+1}}^{s_1^r}\oplus\mathcal{AT}_{2^{n-r+1}}^{s_2^r}\oplus...\oplus\mathcal{AT}_{2^{n-r+1}}^{s_{2^{r-1}}^r}.\end{aligned}$$

Using atomic packets $\mathcal{AWT}_{2^{n-r-1}}^{s_t^r}$ in (7), we obtain discrete controlled WP

$$\begin{aligned}&\mathcal{WT}_{2^n}^{\mathbf{s}^1,\mathbf{s}^2,...,\mathbf{s}^{n-m+1}}[h_0,h_1,...,h_{2D-1}]=\\ &=\prod_{r=1}^{n-m+1}\mathcal{AP}_{2^n}^{\mathbf{s}^r}=\prod_{r=1}^{n-m+1}\left[\bigoplus_{t=1}^{2^{r-1}}\mathcal{AT}_{2^{n-r+1}}^{s_t^r}\right]=\\ &=\prod_{r=1}^{n-m+1}\left[\mathcal{AT}_{2^{n-r+1}}^{s_1^r}\oplus...\oplus\mathcal{AT}_{2^{n-r+1}}^{s_{2^{r-1}}^r}\right]\end{aligned}\quad(8)$$

with discrete binary parameters

$$\begin{aligned}&\mathbf{s}^1=(s_1^1),\\ &\mathbf{s}^2=(s_1^2,s_2^2),\\ &\mathbf{s}^3=(s_1^3,s_2^3,s_3^3,s_4^3),\\ &.....,\\ &\mathbf{s}^{n-m}=(s_2^{n-m},s_2^{n-m},...,s_{2^{n-m-1}}^{n-m}).\end{aligned}$$

But

$$\mathcal{AT}_{2^{n-r+1}}^{s_t^r}=\left(\prod_{i=D-1}^{0}\mathbf{T}_{2^{n-r+1}}^i(-\varphi_i)\right)^{s_t^r}\mathbf{P}_{2^{n-r+1}}^{s_t^r}.$$

Substituting this expression in (8), we obtain the multiparametric representation of WP

$$\begin{aligned}&\mathcal{WT}_{2^n}^{\mathbf{s}^1,\mathbf{s}^2,...,\mathbf{s}^{n-m+1}}[h_0,h_1,...,h_{2D-1}]=\\ &=\prod_{r=1}^{n-m+1}\left[\bigoplus_{t=1}^{2^{r-1}}\left(\prod_{i=D-1}^{0}\mathbf{T}_{2^{n-r+1}}^i(-\varnothing_i)\right)^{s_t^r}\mathbf{P}_{2^{n-r+1}}^{s_t^r}\right].\end{aligned}\quad(9)$$

For inverse WT we have

$$\begin{aligned}&\mathcal{WT}_{2^n}^{-1}[h_0,...,h_{2D-1}]=\mathcal{WT}_{2^n}^{t}[h_0,...,h_{2D-1}]=\\ &=\prod_{r=m}^{n-1}\left[\mathbf{P}_{2^{r+1}}^t\left(\prod_{i=0}^{D-1}\prod_{k=0}^{2^r-1}\mathbf{CS}_{k\underset{2^{r-1}}{\oplus}i,\,k+2^r}(\varphi_i)\right)\oplus\mathbf{I}_{2^n-2^{r+1}}\right],\end{aligned}\quad(10)$$

where $\underset{2^r}{\oplus}$ is addition modulo 2^r. Obviously inverse wavelet packet is:

$$\begin{aligned}\mathcal{WP}_{2^n}^{-1;(\mathbf{s}^m,\mathbf{s}^{m+1},...,\mathbf{s}^{n-1})}[h_0,h_1,...,h_{2D-1}]&=\prod_{r=m}^{n-1}\mathcal{AP}_{2^n}^{-1;\mathbf{s}^r}=\\ &=\prod_{r=m}^{n-1}\left[\mathcal{AT}_{2^{r+1}}^{-1;s_1^r}\oplus...\oplus\mathcal{AT}_{2^{r+1}}^{-1;s_{2^{n-r-1}}^r}\right]=\\ =\prod_{r=m}^{n-1}\left[\left(\prod_{t=1}^{2^{n-r-1}}\left[\mathbf{P}_{2^n}^{-1,\,s_t^r}\prod_{i=0}^{D-1}\prod_{k=0}^{2^r-1}\mathbf{CS}_{(i\oplus k)+t2^{r+1},\,(2t+1)2^r+k}(\varphi_i)\right]^{s_t^r}\right)\right].\end{aligned}\quad(11)$$

Conclusion

In this paper we defined the new representation of orthogonal wavelet transform, named multiparametric forms of cyclic orthogonal WT (5) and WP (8). These forms are the product of sparse rotation matrixes. Defined representations of WT and WP depend on finite set of free parameters, which could be changed independently of one another. For each set of parameters values we get the unique orthogonal WT and WP.

References

1. I. Daubechies, W. Sweldens. Factoring wavelet transforms into lifting steps // J.Fourier Anal. Appl. – 1998. - Vol. 4, No.3. – P. 247-269.
2. I. Daubechies. Ten Lectures on Wavelets // Society for Industrial and Applied Mathematics, Philadelphia, PA. – 1992. - 68 p.
3. V.G. Labunets. Unified approach to fast algorithms of unitary transforms. // Multi–valued Elements, Structures and Systems, (in Russian), Institute of Cybernetics of Ukranian Academy of Sciences Press.- Kiev. - 1983. - P. 46–58.
4. V.G. Labunets. Fast multiparameter transforms (in Russian) // Proceedings of Radioelectronics.- 1985. - No.8. – P. 89–109.

THE BEST MULTIPARAMETRIC WAVELET TRANSFORMS[1]

V. Labunets[2,3], D. Gainanov[2,4], D. Berenov[2,5]

[2] Urals Federal University, Ekaterinburg, Russia,
[3] vlabunets05@yahoo.com, [4] damir@dc.ru, [5] berenov@dc.ru

In this work we find optimal values of parameters of multiparameric wavelet transform (MPWT) for which spectral wavelet coefficients have minimal entropy. Obviously, wavelet transform with optimal parameters have maximal value of coefficient ratio. Experiments show that there are several global optimal values with different computer complexities. MPWT with minimal computer complexity and maximal coefficient ratio is the best wavelet transform for image compression.

Introduction

The common way to analyze and describe complicated signals and images is to represent them as a superposition of simple signal or images. Most of the currently available approaches to signal/image representations are based on "Single Integral Transform Concept". The integral transforms and signal representations associated with them are the most important of any representation. The best known examples are the Fourier and Wavelet transforms. However, there are many other transforms (representations) which can be interesting for image and signal processing. An important aspect of these representations is the possibility to extract relevant information (which is actually available, but hidden) from signals and images in their complex initial form. The extracting of relevant information is the most important part of the signal/image analysis and interpretation. However, the diversity of problems that face science is so great that there is no available single universal method which would be well suited to all problems simultaneously. We hope that this difficulty can be overcome using so-called "Best Transform Concept", based on multiparametric wavelet transform (MPWT and MPWP) [1-9]. One of the most important challenges is to find the best representation for the signals or images. Our approach to the signal is based on the orthogonal MPWT and MPWP. This approach consists of three main steps:

1. Select a best transform (optimal basis) for the problem at hand from continuous family of transforms belonging to the multiparameteric wavelet transform (5) or packet (8) in [1]. Each MPWT (or MPWP) depends on one or more free parameters. When parameters are changed in some way, the type and form of transform are changed as well.For example, MPWP $\mathcal{WT}_{2^n}\left[\varphi_0,\varphi_1,\ldots,\varphi_{D-1}\right]$ may be the discrete Fourier transform (DFT) for some values of parameters, or discrete Walsh transform (DWT), discrete Hartley transform (DHT), discrete cosine transform (DCT) for other values (see Fig. 1 and Fig.2).

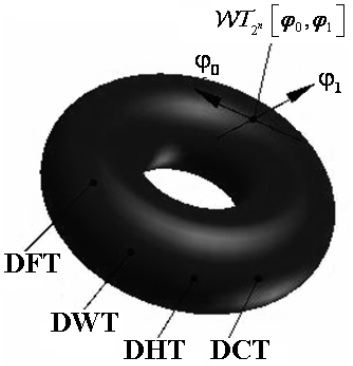
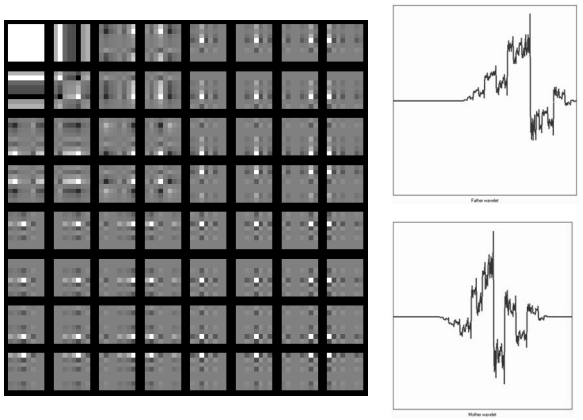


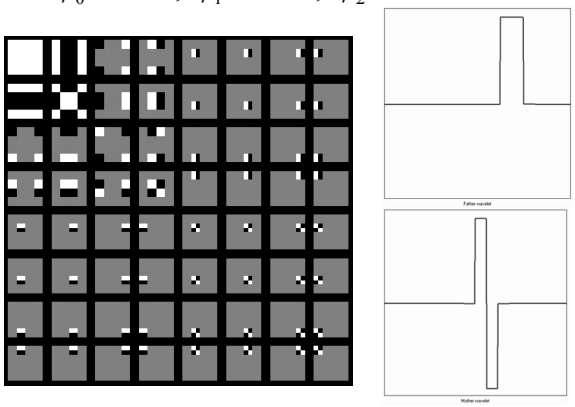
Fig. 4. Space of MPWT $\mathcal{WT}_{2^n}\left[\varphi_0,\varphi_1\right]$.

This property enables us to calculate the spectra of signals and images for a continuous number of transforms. When parameters are changed, MPWT calculates the spectra of an image for all transforms belonging to MPWP and finds the best possible transform by optimizing certain criteria.

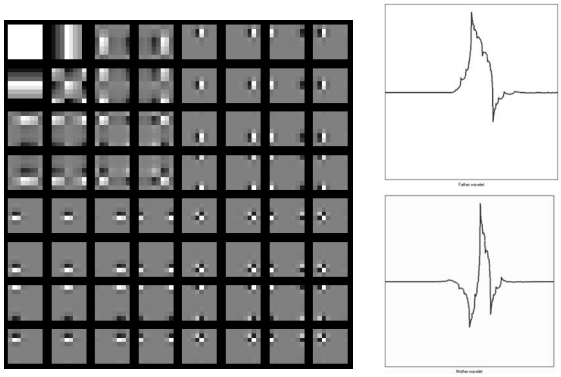
[1]This work was supported by RFBR (grants № 13-07-12168, BR № 13-07-00785) and by MES RF grant № 218-03-167



$\varphi_0 = 138^\circ,\ \varphi_1 = 337^\circ,\ \varphi_2 = 110^\circ$



$\varphi_0 = 45^\circ,\ \varphi_1 = 0^\circ,\ \varphi_2 = 180^\circ$



$\varphi_0 = 182^\circ,\ \varphi_1 = 53^\circ,\ \varphi_2 = 350^\circ$

Fig. 2. Basis pictures of 2D 3-parametric WT $\mathcal{WT}_{2^3}[\varphi_0, \varphi_1, \varphi_2]$ (left). Forms of mother and father functions (right) for different values of $\varphi_0, \varphi_1, \varphi_2$.

While in the classic signal/image processing systems (based on "Single Integral Transform Concept") we observe a single "photo" of spectrum (for example, the spectrum of an image in a concrete Daubechies basis), then in MPWP we observe a "movie" consisting of image spectra (transformed images) for different transforms belonging to MPWP. We can then select the best transform (the best coordinate system). The main reason to use a continuous family of transforms instead of a single transform (for example, Daubechies, Walsh or Haar-wavelet) is that we may lose our flexibility for handling various changes in images or manifolds.

2. Sort the components of transformed domain by "importance" for a specific task and discard "unimportant" components. What is best and "important" clearly depends on the problem. Usually, a best transform for representing a signal or image should give large magnitudes along a few axes and negligible magnitudes along most axes when signal or image is expended into the best basis associated with best transform. For image compression, a basis which provides only a few large components in the transformed domain should be used since we can discard the other components without large signal degradation. A basis through which we can "view" classes as maximally-separated point clouds in the N-dimensional space (where N is the number of pixels in each image) is an option. In this case, the "distances" between classes should be used as a measure of the transform efficiency

3. Use the surviving coordinates to solve the problem at hand using generalized non-harmonic analysis. The output of a computation can be memorized to be used in subsequent numerical computations: compression, interpolation, segmentation, edge detection, filtering, denoising, etc.

Implementation of a best transform selection procedure for a prescribed signal/image (or a family of signals/images) requires the introduction of an acceptable cost function which translates best into a minimization process. In this section we use entropy of wavelet coefficient for cost function.

MPWT and optimal image compression

The wide class of orthogonal wavelet transforms $\mathcal{WDT}$ can be defined by coefficients: h_0, h_1, …, h_{L-1} or Jacobi angles $\varphi_0, \varphi_1, \ldots, \varphi_{D-1}$ 2]:

$$\mathcal{WT}_{2^n}[h_0, h_1, \ldots, h_{2D-1}] = \mathcal{WT}_{2^n}[\varphi_0, \varphi_1, \ldots, \varphi_{D-1}].$$

In order to estimate compression properties of multiparametric orthogonal wavelet transform we have conducted experiments for revealing dependency of spectra's coefficients

entropy $E^D(\varphi_0,\varphi_1,\ldots,\varphi_{D-1})$ on quantity of angle-parameters D and values of angle-parameters φ_i. We use the entropy of spectra's coefficients, quantized to integer values, as the cost function. The form of the dependency $E^2(\varphi_0,\varphi_1)$ (case of two-parametric transform) is schown on Fig. 3.

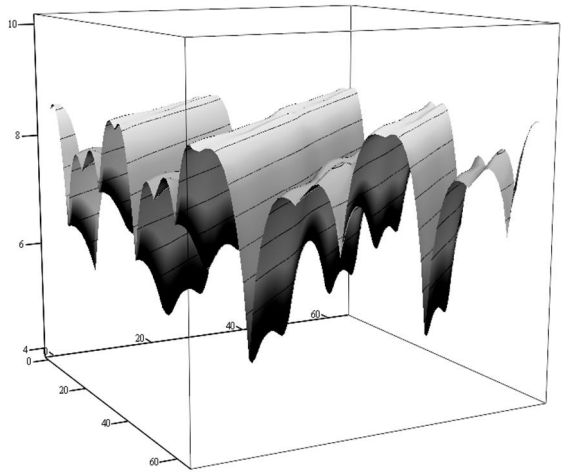


Fig. 3. Entropy of spectra E^2 relative to parameters φ_0 and φ_1 for $\mathcal{WDT}_{2^8}[\varphi_0,\varphi_1]$. (Test image is "Lena")

The experimental gotten result for minimal attainable spectra's entropy $E^D_{\min}(\varphi_0,\varphi_1,\ldots,\varphi_D)$ relative to D (for $m=6$, $D=3$; and $m=6$, $D=4$) for three different images "Baboon", "Lena" and "Lighthouse" is shown on fig 4. Only the minimal attainable spectra's entropy varied from image to image.

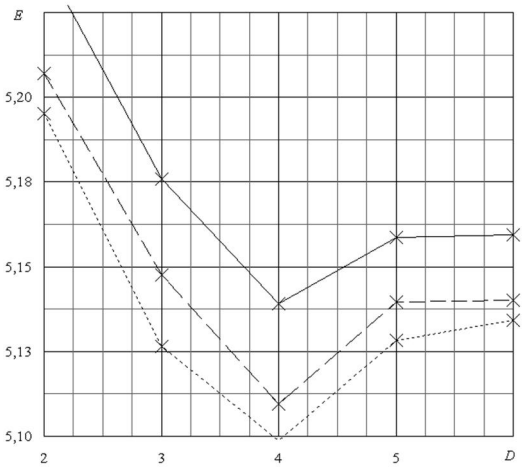
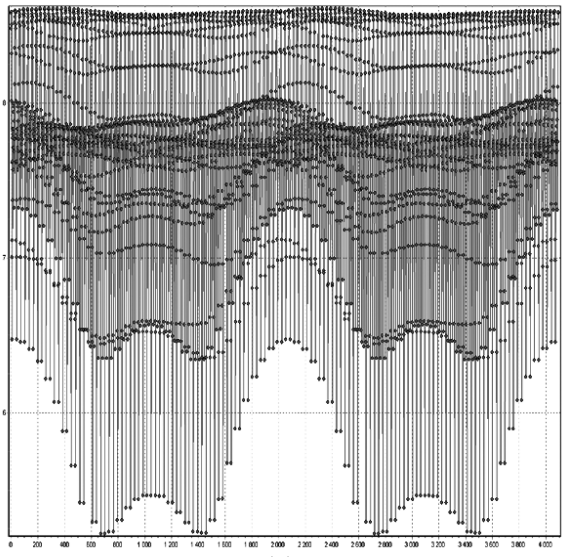
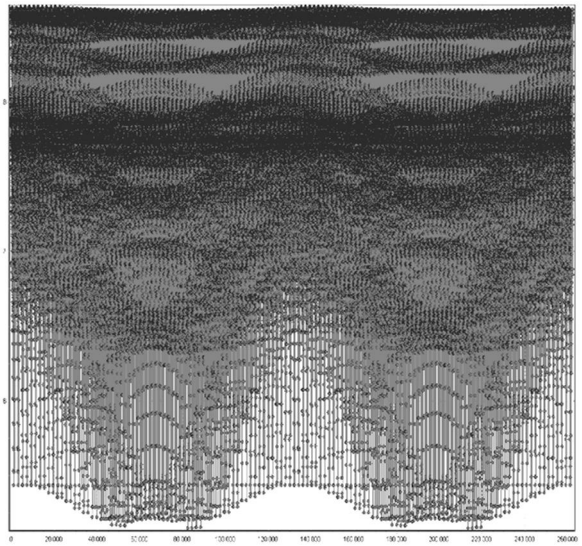


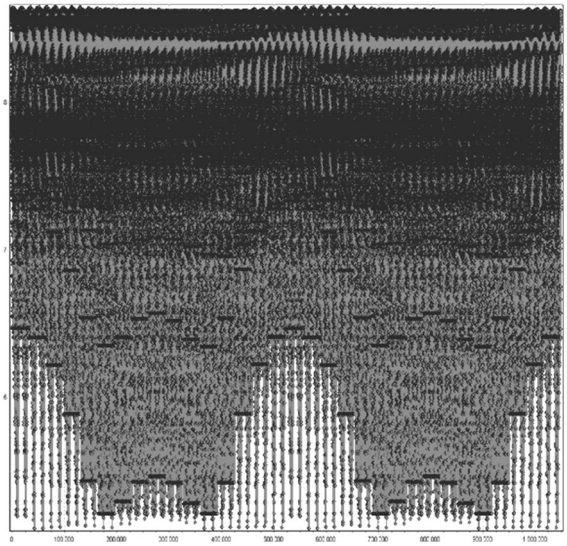
Fig. 4. Minimal attainable spectra's entropy relative to D



(a)



(b)



(c)

Fig. 5. Entropy of spectra relative to P for different values of parameters:
a) $m=6$, $D=2$; b) $m=6$, $D=3$;
c) $m=6$, $D=4$.

Dependency $E^D(\varphi_0,\varphi_1,\ldots,\varphi_D)$ are multi-dimensional if $D>2$. To visualize such dependency we divide the turn-down of each parameter $\varphi_i \in [0,2\pi)$ into 2^m parts. And then we treat the vector $(\varphi_0,\varphi_1,\ldots,\varphi_D)$ as on the D-digit number in Radix-2^m number system $\langle P\rangle_{2^m}=\sum_{i=0}^{D-1} p_{D-i}(2^m)^i$. This number we can rewrite in Radix-10 number system $\langle P\rangle_{10}$. Thus, we will put this number $\langle P\rangle_{10}$ on X-axis and on Y-axis we will put the entropy of obtained spectra.

Fig. 3 and Fig.a ($m=6$, $D=2$) show, that researched dependency has four minimums. Moreover, the form of dependencies and the location of minimums were the same for different images.

In order to compare the classical Daubechies wavelet transforms (for $L=4,6,8$) with multiparametrical wavelet transforms (for $D=2,3,4,5$) we investigated the quality of reconstructed image *PSNR* relative to the percent of not zeroed spectra's coefficients. The results are in Fig.6. The values of the angle-parameters, used in this comparison, conform with the minimums of the function $E^D(\varphi_0,\varphi_1,\ldots,\varphi_D)$ for proper D. The test image «Lena» was used.

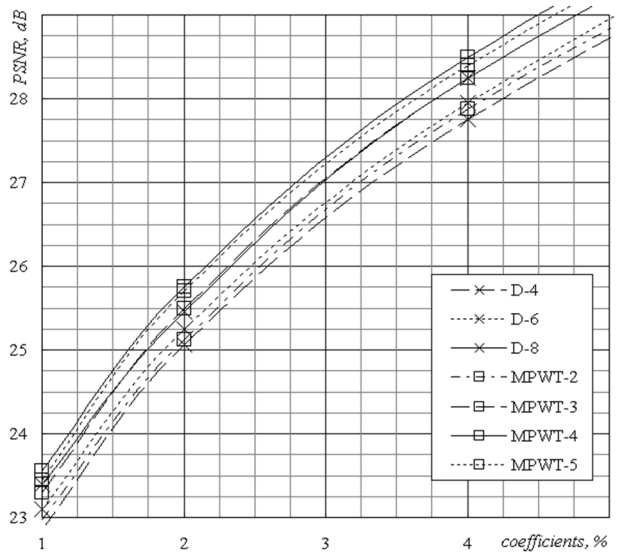


Fig. 6. Reconstruction of image "lena" with different percent of preserved wavelet spectra coefficients.

As shown on Fig., the transforms, found with assistance of multiparametric presentation of wavelet transforms, excels the classical ones with the same carrier length with respect of compression properties.

Conclusion

In this work we researched compression properties of orthogonal wavelet transforms. In more general case, the using of multiparametric presentation of wavelet transform gives us the opportunity of finding the best wavelet basis to solve various particular problems. For example it could be the image filtering problem or special image analysing in medicine.

References

1. V. Labunets, D. Gainanov, D. Berenov. Multiparametric wavelet transforms and packets // In this Proceedings.
2. K. Knothe. Adaptable orthogonal wavelet transformations // Preprints of 7th International Student Olympiad on Automatic control (Baltic Olympiad). S.-Pb. – 1999.- P. 51-55.
3. S.A. Lebedev S.A., S.N. Unzhakov, V.G.Labunets, E.V.Ostheimer-Labunets. Optimal with image compression point of view manyparametric wavelet transforms // International Conference "Telecom-2006". Russia, Ekaterinburg, Ltd. "RealMedia", - 2006. –P. 272-274.
4. S.A.Lebedev, V.G.Labunets, O.A. Gusev. E.V. Cyclic manyparaptric wavelet transforms: How many are there optimal (with image compression point of view) wavelet transforms? // 8th International Conference "Digital Signal Processing and Its Applications", March 29-31. Moscow, Russia. – 2006. - P. 113-116.
5. E.Labunets, V.Labunets. New networks for nonlinear, linear and orthogonal manyparametric transforms // Proceed. V–th Int. Workshop on Parallel Processing by Cellular Automata and Arrays (PARCELLA' 90), Berlin. – 1990. – P. 239–244.
6. V.G.Labunets. Fast nonlinear multiparameter transforms // All–Union Scientific Conference Methods and Microelectronical Devices of Information Transform and Processing, (in Russian), Moscow. – 1987. – P. 64–74.
7. V.G.Labunets. Fast multiparameter transforms (in Russian) //Proceedings of Radioelectronics. – 1985.- No 8. - P. 89–109
8. V.G.Labunets. Unified approach to fast algorithms of unitary transforms // Multi–valued Elements, Structures and Systems (in Russian), Institute of Cybernetics of Ukrainian Academy of Sciences Press, Kiev. – 1983. - P. 46–58.

TIME-FREQUENCY SPONTANEOUS EEG FEATURES OF EARLY STAGE PARKINSON'S DISEASE

Yu.V. Obukhov[1], M.S. Korolev[1], K.Yu. Obukhov[1], O.S. Sushkova[1], R.R. Nigmatullina[2], Z.A. Zaljalova[2], 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, +7(495) 629-7285, obukhov@cplire.ru
[2] Kazan State Medical University, Butlerova, 49, Kazan, 420012, +7(843) 292-7299, razinar@mail.ru
[3] Institute of Higher Nervous Activity and Neurophysiology of RAS, 5A Butlerova St., Moscow 117485, Russia, +7(499)334-2622, agabova@gmail.com
[4] Koltzov Institute of Developmental Biology of RAS, ul. Vavilova 26, Moscow, 119334, Russia, +7(499) 135-8842, mugrumov@mail.ru

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.

Introduction

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.

Methods

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

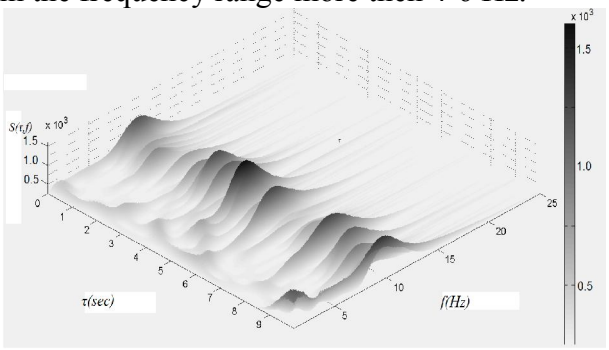
$$Sx(\tau, f) = \left|W(\tau, f)\right|2 , \quad (1)$$

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

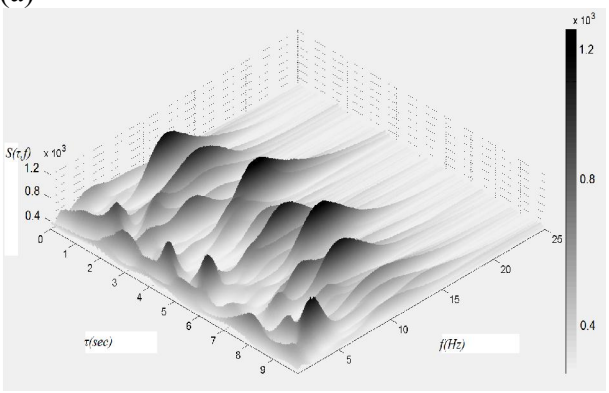
$$\psi(\eta) = \frac{1}{\sqrt{\pi F_b}} e^{2i\pi F_c \eta} e^{-\frac{\eta^2}{F_b}} \quad (3)$$

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

Fig. 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].

Fig. 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.

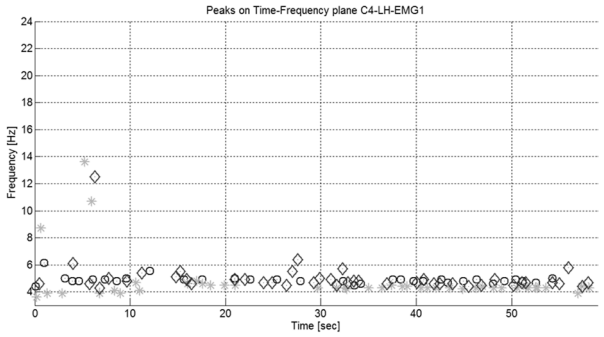
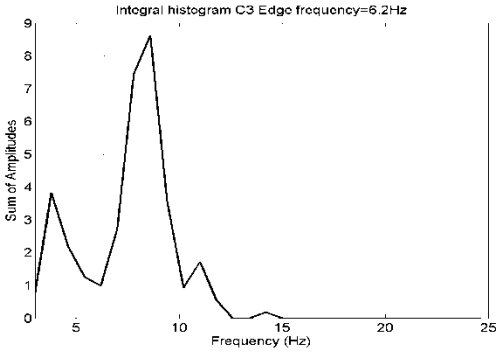
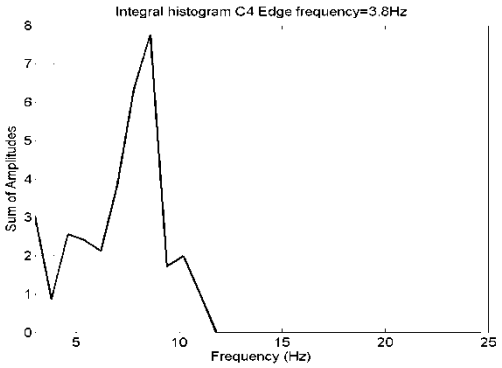


Fig. 2 C4 EEG (circles), left hand electromyogram (rhombs), and left wrist tremor (stars) of scalograms extreme at the 1st stage PD patient

To analyze the scalograms peaks time-frequency distribution we consider the histograms of extreme power sums at (Δf, Δt) rectangles.

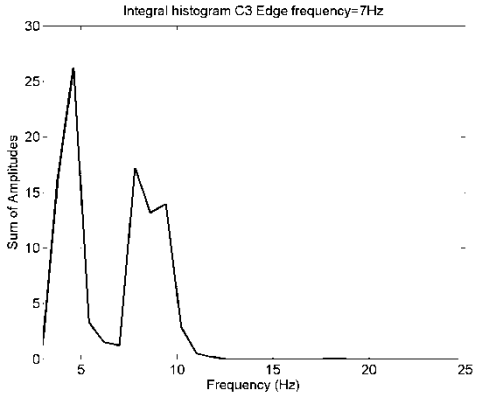


(a)

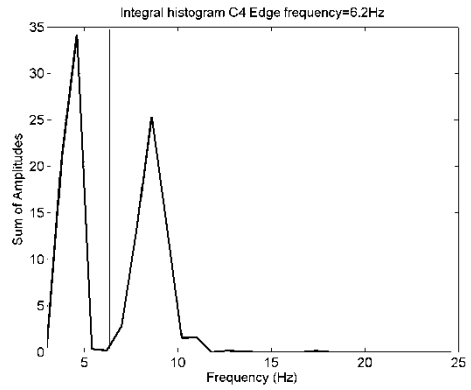


(b)

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



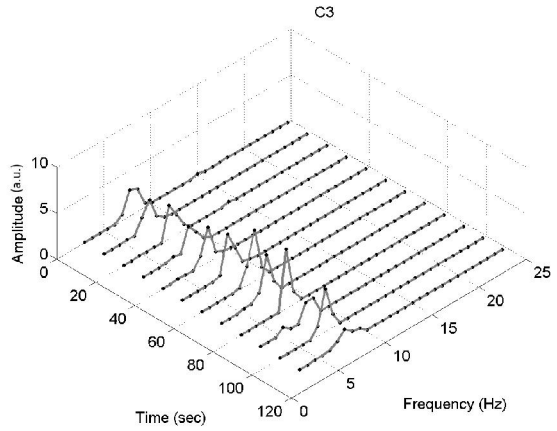
(a)



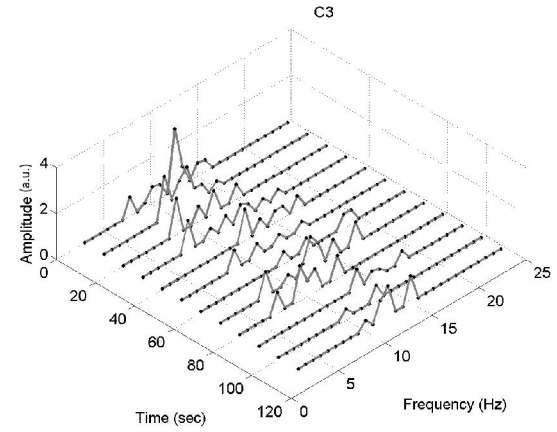
(b)

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

Fig. 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 disappears.



(a)



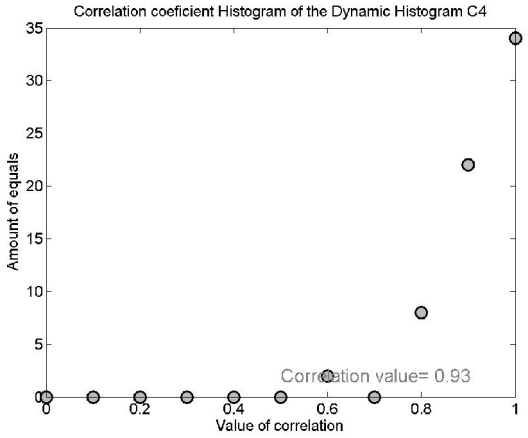
(b)

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

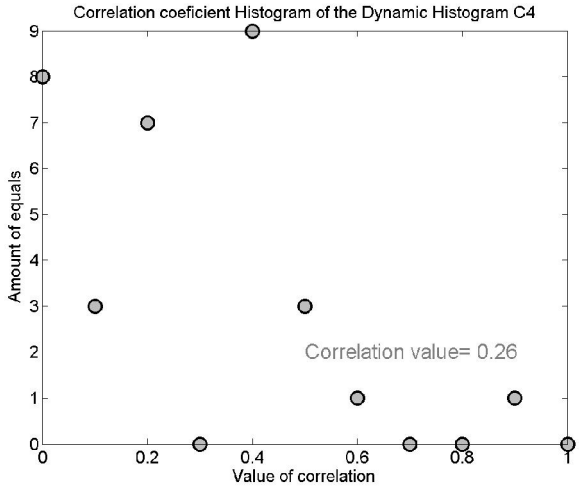
The dynamical histograms calculated at (0.7 Hz, 10 sec) rectangles show the disordering of electrical activity in more then 6 Hz for the 1st stage PD (see fig. 5). Such disordering can be evaluated with the help of dynamical histograms correlation matrixes. This evaluation can be done by histograms of correlation values. Fig. 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 $N(Ri)$ is a quantity of correlation values Ri in correlation matrix. The average correlation value P indicated in fig. 6.

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.

Fig. 6 shows the decreasing feature *P* or increasing alpha rhythm disorganization with increasing PD stage.

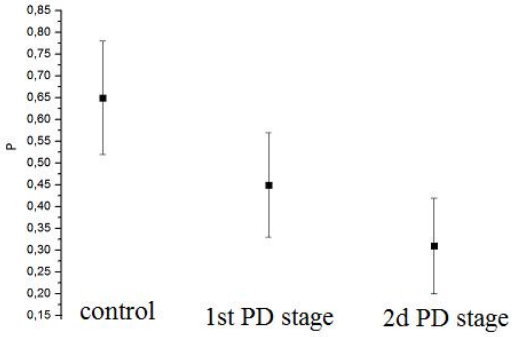


Fig. 7 Average correlation value *P* for investigated normal volunteers, 1st, and 2nd Hoehn-Yahr stages PD patient groups

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. Bernheimer H., Birkmayer W., Hornykiewicz O., Jellinger K., Seitelberger F. 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. Bezard E., Gross C.E. Compensatory mechanisms in experimental and human parkinsonism: towards a dynamic approach. Prog. Neurobiol. 1998. v. 55. № 2. p. 93-116.
4. England A.C., Schwab R.S., Peterson E. The electroencephalogram in Parkinson's syndromee. EEG Clin. Neurophysiol. 1959. v. 11. № 4. p. 723–731.
5. Soikkeli R., Partenen J., Soininen H. Paakkonen A., Riekkinen Sr P. Slowing of EEG in Parkinson's disease. EEG Clin. Neurophysiol.. 1991. v. 79. № 3. p. 159–165.
6. Stoffers D., Bosboom J., Deijen J.B., Wolters E.C., Berendse H.W., Stam C.J. Slowing of oscillatory brain activity is a stable characteristic of Parkinson's disease without dementia. Brain 2007. v. 130. № 7. p. 1847-1860.
7. Sarizawa K., Kamei., Morita A., Hara M., Mazutani T., Yoshihashi H., Yamaguchi M., Takeshida J., Hiravanagi K. Comparison of quantitative EEG between Parkinson disease and age adjusted normal controls. J. Clin. Neurophis. 2008. v. 25. p. 361-366.
8. Moazami-Goudarzi M., Sarnthein J., Michels L., Moukhtieva R., Jeanmonod D. 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. Berendse H.W., Stam C.J. Stage-dependent patterns of disturbed neural synchrony in Parkinson's disease. Parkinsonism and Related Disorders. 2007. v. 13, Suppl. 3. p. 440–445.
10. Obukhov Yu.V, Korolev M.S., Gabova A.V., Kuznetsova G.D., Ugrumov M.V. Method of early stage Parkinson's disease electroencephalography diagnostics // RF patent. - 2484766, 20.06.2013, (in Russian).
11. Hoehn MM, Yahr MD. Parkinsonism: onset, progression and mortality. // Neurology. - 1967, V. 17, pp. 427-442, PMID 6067254.

DESIGN PRINICPLES FOR INTERACTIVE IMAGE INFORMATION MINING SYSTEMS

G.K. Ouzounis[1]

[1] DigitalGlobe Inc., 1601 Dry Creek Drive, Suite 260, Longmont, CO 80503, USA.
georgios.ouzounis@digitalglobe.com

Modern image information mining systems operate on vast amounts of data making the efficient organization and management of the image information content to be of crucial importance. Addressing this challenge, tree-based image representation structures were introduced to allow for the precise definition of image components, to shape up component hierarchies and to provide efficient means for component attribution. Trees are used as interfaces between the image space and some control module often hosting a feature space. Making use of the advanced search and retrieve mechanisms of most tree algorithms a new system architecture and an operation protocol for interactive image information mining are proposed. They are demonstrated using the Alpha-Tree structure on a built-up mining task from very high resolution panchromatic satellite imagery.

Introduction

Image information mining (IIM) systems are found in a wide range of modern applications such as online search engines, intelligence gathering platforms, diagnostic tools in medicine, etc. IIM systems facilitate query mechanisms over an input data space. A query is a supervised procedure in which a user or a system requests the detection of image features that satisfy certain criteria such as semantic labels, thresholds on object attributes or enumerated resemblance to one or more target prototypes from a set of examples [1-3].
The architecture of IIM systems varies depending on the application, however there exist three key requirements common in each design; the definition of a structured search space, the definition of an object attribution or semantic characterization schema, and the definition of a search and retrieve mechanism [2]. All three are crucial in the design of a functional system and each has its own share on the overall performance, especially as the volume and complexity of the input data increases. It is often the case that each requirement drives the design of a separate subsystem. This increases the overall system complexity at the cost of reduced performance. This talk gives an insight into design principles from a single system perspective and introduces a new IIM paradigm based on hierarchical image representation structures.

The Anatomy of an IIM System

The definition of a structured search space requires the organization of the image information content into meaningful entities or components. Shaping up such components for the entirety of the image space compresses the information content and improves the performance of the query mechanism. The set of all elementary components, i.e. those in the initial allocation, is called a base layer. The search space may coincide with the base layer or consist of multiple layers, capturing in this way hierarchical relations between components according to some pre-defined nesting order. Fig.1. shows a test pattern on the white frame and the corresponding hierarchical image representation structure, an Alpha-Tree [4] that coincides with the search space. Components are shown as white dots interlinked by red wires.

The second requirement concerns the way components are represented. Components are attributed during or after the organization of the image information content.

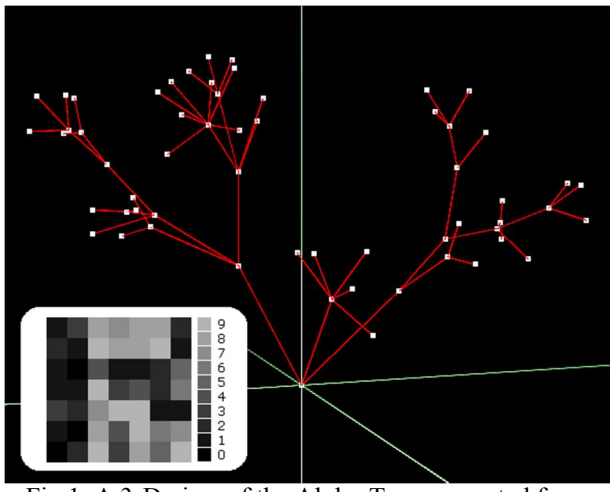


Fig.1. A 3-D view of the Alpha-Tree computed from the 10-level test pattern shown in the white frame. The Alpha-Tree defines the search space for the test pattern.

Attributes are often geometric, radiometric or other statistically derived measures that describe each component in question. The set of attributes associated with each component is used by the search and retrieve mechanism when comparing an image feature against a target prototype. Thus the better the attribution is, the better the output of the query will be.

The query mechanism is often a set of logical predicates applied on each examined component. If the component satisfies the query criteria it remains intact otherwise it is removed in its entirety. Both actions are executed by rearranging spatial and/or hierarchical relations between the components involved.

The Worst Case Scenario

Given a complete system we now touch upon the scenario in which one or more of these requirements is not met or not implemented properly.

The lack of structure or poorly defined structure of the information content and thus of the search space means that components must be formed during the query run-time. This introduces major bottlenecks and makes queries time-consuming, resource-demanding and prone to error as they become sensitive to region-access-order. Moreover, repeated queries for fine tuning the output make the search process prohibitive.
The attribution is mostly responsible for the quality of the query results. A target prototype that is poorly described in terms of attributes may be the cause of severe noise in the output. By contrast, over constrained characterization will often fail to generalize appropriately, limiting the target class variability and thus requiring more same-class prototypes.

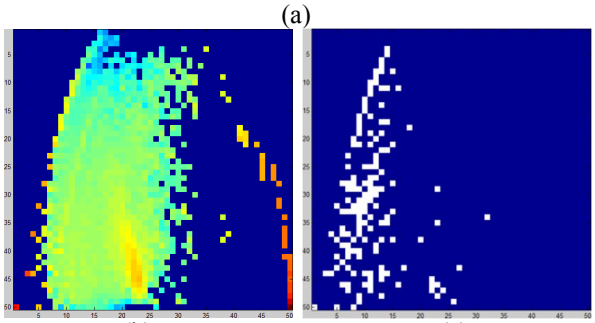
The search and retrieve mechanism is interrelated to the search space organization thus any poor design has an impact primarily on the system performance.

Hierarchical Image Representation

Complying with all three requirements in the design of a complete system, we end up with what is usually referred to in image analysis as a hierarchical image representation data structure or simply a dendrogram. Examples of known dendrograms are the Max-Tree [5], the Alpha-Tree [4], the Binary Partition Tree [6], and others. The key difference between different types of trees is the set of rules defining the base layer and the evolution parameter shaping up the respective hierarchy.

The Alpha-Tree [4] was demonstrated as an IIM system in [2] giving evidence of being a powerful tool. In brief, an Alpha-Tree organizes the image space into a stack of nested partitions, each defined at a given dissimilarity threshold α. Each partition cell is an α-component which is essentially an image region consisting of pixels for which, any two that are adjacent have a difference of less than or equal to α according to some pre-defined dissimilarity metric. The Alpha-Tree maps only the first instance of an α-component from the stack of partitions, eliminating this way redundancies in the representation. The leaves of the tree correspond to components consisting of elements that are fully similar to each other with respect to the dissimilarity metric. The set of leaves defines the base layer in the IIM search space. The root of the tree corresponds to the single α-connected component that coincides with the image definition domain.
Each node of the tree maps to a unique α-component from the stack of partitions and points to its parent; the first superset of the given α-component at α' strictly greater than α.

(a)



(b) (c)

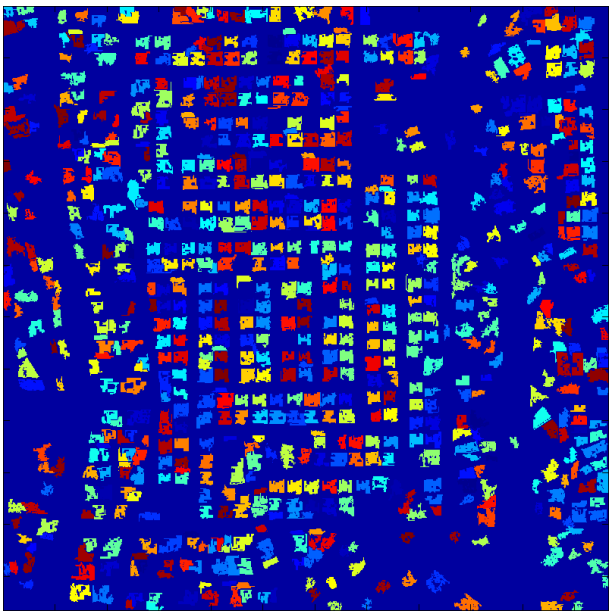
Fig.3. (a) A Quickbird panchromatic image of an urban area at a spatial resolution of 0.6m. ©DigitalGlobe Inc. Green masks indicate marked target prototypes. (b) The size vs. compactness pattern spectrum of the input image; (c) the marked bins associated to the prototypes.

Moreover, each node contains a pool of auxiliary data that are updated incrementally during the tree construction phase. Based on this data, component attributes can be computed in real time and during every query. Nodes are indexed making search and retrieval mechanisms fast and efficient.

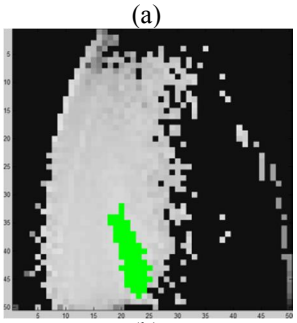
A Protocol for Interactive Image Information Mining

Assume an Alpha-Tree defining an IIM system architecture. Image operators are computed on an Alpha-Tree with a single pass through the structure. Node attributes are computed upon visiting a node from the pool of auxiliary data it contains.

This functionality allows for real-time multi-dimensional feature space computation.



(a)



(b)

Fig.4. (a) The segmentation from a query based on the switch selection (green) on the pattern spectrum in (b).

An example for the input image of Fig.3(a) is shown in Fig.3(b). This is an α-connected pattern spectrum [7] with the origin at its top left corner. The x-axis corresponds to a logarithmic mapping of component size and the y-axis to a square-root mapping of the component compactness measure. The energy of each bin, shown in heat-map color code, corresponds to a normalized contribution of the image information content; i.e. the pixel count of all components (represented by nodes) satisfying the bin extreme conditions multiplied by each node's lifetime.

If target prototypes are to drive a query, they must be first defined. Prototypes can be imported from an external data-base or collected directly from the image space. In both cases a binary mask is used to mark all maximal components that are fully included in the mask's foreground extent. For each external prototype an individual tree is constructed from which component attributes are computed. For prototypes collected from the input image space, a visit to each node

associated with a detected maximal component allows for direct attribute computation.
The attributes of all components identified as target prototypes are then projected onto the feature space indicating which regions are of interest. Fig.3(c) shows this for all targets within the green masks in Fig.3(a). The dominant cluster at the left of the plot that stretches along the y-axis shows the contribution of side components to our targets, which are small in size, of varying compactness and captured under the mask. Ignoring them and focusing at the remaining points, a second smaller cluster is found at the bottom of the plot and in the middle of the x-axis. This cluster marks a wider region of interest according to the heat map in Fig.3(b), suggesting the presence of a persistent pattern. Using the switchboard interface [2] a pilot query is launched by activating the switches associated with the bins in this area of interest. Fig.4(b) shows the activated switches in green. A pass through the tree collects and projects the attributes of each node onto the feature space. If the switch of the corresponding bin is set on the associated component is retained in the query output, otherwise it is ignored. This generates a segmentation and an example for the image in Fig3.(a) is shown in Fig.4(a).

Discussion

Revising the proposed architecture of an IIM system we observe the following facts: (i) The tree construction is fast and offers a compact and explicit representation of the image information content. (ii) The feature space can be computed rapidly [8] and further be used for training a classifier. (iii) Classifiers can be computed directly on the tree structure [3]. (iv) The search and retrieve mechanism comes with each tree algorithm and is very efficient.

The most important feature of this architecture is the ability to launch queries on adjustable number of target prototypes without the need of system recalibration. Adding or removing prototypes translates into projecting their attributes onto the feature space and since the query is controlled via the switchboard the system performance remains unaffected.

Conclusions

The adaptation of a tree data structure for implementing an IIM system complies with all the design requirements discussed. The system benefits from all the additional tree functionalities and supports rapid queries with varying numbers of target prototypes.

References

1. L. Gueguen, G. K. Ouzounis. Tree-based representations for fast information mining from VHR images. // ESA-EUSC-JRC Proc. 8th Int. Conf. Image Information Mining. - 2012. P. 15-20.
2. G. K. Ouzounis. V. Syrris. L. Gueguen. P. Soille. The switchboard platform for interactive image information mining. // ESA-EUSC-JRC Proc. 8th Int. Conf. Image Information Mining. - 2012. P. 26-30.
3. L. Gueguen, G. K. Ouzounis. Hierarchical data representation structures for interactive image information mining. Taylor and francis, Int. J. Image and data Fusion. – 2012. Vol.3, No. 3. – P221-241.
4. G.K. Ouzounis, P. Soille. The Alpha-Tree algorithm. //Publications Office of the European Union. – 2012.
5. P. Salembier. A. Oliveras. L. Garrido. Anti-extensive connected operators for image and sequence processing. // IEEE TIP. - 1998. - Vol.7, No. 4. - P. 555-570.
6. P. Salembier. L. Garrido. Binary partition tree as an efficient representation for image processing, segmentation and information retrieval. IEEE TIP. – 2000. – Vol.9, No. 4, – P 561-576.
7. G.K. Ouzounis, P. Soille. Pattern spectra from partition pyramids and hierarchies. // Springer, LNCS: Proc. 10th Int. Symposium on Mathematical Morphology. - 2011, - Vol. 6671. - P. 108–119.
8. M. H. F. Wilkinson, U. Moschini. G. K. Ouzounis, M. Pesaresi. Concurrent computation of connected pattern spectra for very large image information mining. // ESA-EUSC-JRC Proc. 8th Int. Conf. Image Information Mining. - 2012. P. 21-25.

About the author

Dr. Georgios K. Ouzounis specializes in mathematical modeling and algorithm development/optimization for geospatial image analysis. From September 2009 to September 2012 he worked as a researcher at the Global Security and Crisis Management Unit of the Joint Research Centre of the European Commission in Ispra, Italy. He is with DigitalGlobe since October 2012.

Georgios K. Ouzounis has invented three major algorithms: the dual-input Max-Tree algorithm, the Area Zone Decomposition for computing Differential Area Profiles (DAPs), and the Alpha-Tree. He was the architect of all computational engines related to the DAP vector fields that were used by the JRC Image Query (IQ) system to populate the first VHR Global Human Settlement Layer (GHSL).

GEOMETRICAL FEATURES IN POINT CLOUDS FOR ROBUST REGISTRATION OF LASER SCANS

D. Paulus[1, 2], A. Mützel[1, 3], F. Neuhaus[1, 4]

[1]Computervisualistik, Universität Koblenz-Landau, Germany
[2]paulus@uni-koblenz.de, [3]amuetzel@uni-koblenz.de, [4]fneuhaus@uni-koblenz.de

We examine the question of whether geometric point features are suitable for real-time 3D mapping in outdoor scenes using high-resolution laser scans. Features are extracted from keypoints in the 3D point clouds. We inspect two fundamental problems: pair-wise registration of 3D laser scanning and recognition of places. We claim that it is possible in principle to create 3D maps using the above characteristics, even if other approaches to registration applied to our data sets do not provide sufficient results. As two adjacent scans in our data set have a considerable offset, local methods based on ICP will fail without a good initial estimate of the pose. This estimate could originate from robot odometry for example, which, unfortunately, is not always available. We present a system that can register various standard data sets as well as our own scans in real time, without requiring an initial guess. For place recognition we use a bag-of-words descriptor. Visual basis words are selected after clustering the training data. The method allows for loop closing and localization in existing maps without using any external sensors such as GPS.

Introduction

The Simultaneous Localization and Mapping (SLAM) problem is a current area of research in the robotics context. This problem describes the task of incrementally building a map of the environment and incorporating new information into it.
Many different types of sensors can be used for this task, for example cameras, laser scanners, inertial sensors or GPS sensors, either alone or in combination. This work will approach the problem using only geometric information provided by 3D laser scanners. In addition, techniques from the areas of object reconstruction and object recognition are applied to the SLAM task.

Related Work

Our system consists of the following components, for which we outline related work: SLAM frontend, SLAM backend, interest points, descriptors, and place recognition.
If a robot providing odometry information is used for mapping, this information is typically used as a starting point for a local laser scan registration algorithm such as Iterative Closest Point (ICP). It results in a relative pose between the last laserscans, a quantity that is needed in the *SLAM frontend*. The latter constructs a graph from these computed relative poses, which can then later be optimized to find a globally optimal configuration of robot poses. An example related to our work was proposed by Steder et al. [1], where the incremental registration of laser scans is done this way and the map is later used for place recognition. The correlative scan matching by Olson [2] provides an alternative to this method, though it is designed to work in 2D only.

The task of the *SLAM backend* is to find the optimal configuration of the robot poses with respect to the constraints defined by the observations. Being essentially a least-squares optimizer, one wrong hypothesis can lead to a catastrophic failure where the complete pose-configuration gets inconsistent. Olson and Agarwal [3] present a solution to this problem by proposing to model edges in the graph as mixtures of different Gaussian distributions, so that inconsistent loop edges can be effectively disabled.

Interest Points, or *keypoints*, are distinct points in a point cloud that mark recognizable areas. Desirable properties are listed in [4]. The NARF (Normal Aligned Radial Feature) keypoint detector and descriptor to work on 2D range images is introduced in [1]. Zaharescu et al. present the MeshDoG keypoint detector [5], an approach similar to the 2D SIFT operator but useful for 3d point clouds.

A comparative study of different 3D *descriptors* that use geometry and optionally color information is done by Alexandre [6].

place recognition.Many mMore approaches can be found in the literature to any of these topics, but the ones mentioned above are most relevant to our work.

Feature Extraction

We examined a number of geometry-only keypoint detectors and descriptors, e.g.. the NARF, DoG and Harris3D detectors and the PFH, FPFH, SHOT and NARF descriptors.

Additionally, we propose two modifications to the local reference frame of the SHOT descriptor, resulting in more repeatable reference frames and, thus, more correct correspondences. In the first version, called SHOT-NA (SHOT Normal Aligned), the disambiguation of the axes of the reference frame is performed so that the axes are oriented to be consistent with the majority of the surface normals in the neighborhood, instead of the location of the neighborhood points. In the second version, the local reference frame is computed using the global z-axis and the surface normal, as it is done in the NARF descriptor. This results in a not rotation invariant descriptor with a more stable reference frame; the resulting descriptor will be denoted SHOT-U (SHOT Upright).

Another important addition to the existing features is the introduction of a multiscale approach. For each scale, the resolution of the point cloud was divided by two, while the keypoint and feature radii were doubled.

Keypoints far from the sensor were rejected in the multiscale variant due to the fact that sensor noise increases greatly for distant points. For correspondence estimation, only features in the same scale are considered, since the absolute scale is known.

To select the best combination of keypoint detectors and descriptors, we performed a quantitative evaluation of the keypoint detectors in combination with different feature descriptors.

For the evaluation, our data set with inter-scan distances of up to 11m (Fig. 2, below) was down-sampled to a resolution of 10cm using the VoxelGrid filter found in the open source PointCloud Library (PCL), registered with human assistance and the result refined by the ICP algorithm to obtain a ground truth map. Features were computed for each scan and correspondences determined for each incremental scan matching edge in the SLAM graph.

$\chi 2$ norms were examined. This experiment showed that iIn he single scale as well as in the multiscale cases, the DoG detector combined with the SHOT-U descriptor performed best, resulting in 4.8% and 4.4% of correct correspondences. Note that the single-scale features have slightly more correct correspondences in this data set; however, the multiscale features are necessary to register all available data sets.

Another observation is that the FPFH, PFH and NARF features never resulted in more than 1.3% of correct correspondences.

From this experiment, it is clear that existing geometry-only features are not particularly descriptive when used in the context of outdoor scan registration with large distances between scans, which is most likely due to the fact that depth information tends to have a very low frequency in comparison to color information for instance.

Pairwise Registration

The goal of the pairwise registration procedure is to align 3D laser scans quickly that have been captured in a stop-and-go fashion. When capuring scans this way, it is desirable to maximize inter-scan distances to reduce the time needed to scan a given area. Typical examples for the application of this work are the Freiburg360° data set published by Bastian Steder and the data sets recorded for this work on the campus of the University of Koblenz-Landau. In the data sets described

in the results below, distances of up to 11m between scans are common and ambiguities let other algorithms taken from the literature converge to wrong local optima of the optimized cost function. As described in the last section, tThe registration has to handle a large proportion of outliers in the set of all correspondences. We propose optimizing the algorithm of Steder et al [1], i.e. a registration algorithm based on single point correspondences and a robust sensor model for the task of a reliable registration for a simple reason: With only 4.8% of good correspondences in the average case, the chance of randomly selecting three correct correspondences as required by SAC-IA is very low, and, thus a large number of iterations has to be performed.

When using the local reference frames of the feature descriptors of NARF or SHOT, a single correspondence is sufficient to determine all six degrees of freedom of the transformation between the scans.

Sensor Model

The basic idea of the sensor model in [1] is to estimate a transformation **T** from the coordinate system of one scan to another, to render the keypoints of one scan into the next scan, to compare them and to determine different possible problems or situations resulting from e.g. occlusion. The distances of each matched points are weighted according to their properties and a score is determined for the estimated transformation. Our extension to the sensor model of [1] is shown in Fig. 1 where we propose a new lower weight *w* for points classified as occluded obstacles.

In contrast to [1] we use all correspondences from scan A to B and vice versa and we select those with lowest error to do the final estimate.

The BoW approach to place recognition of Steder et al. [1] is based on classifying features into geometric words via a k-means clustering and then using histograms of word counts to compute similarities between scans. In contrast to the original work we re-used the computed features for the incremental matching step instead, due to the focus on real-time registration. Re-using the features reduces the time needed to process each scan. To build the dictionary, all extracted features from one or multiple data sets are used. When registration is performed in a multi-scale way, only features at the coarsest scale are used because these are the only ones calculated at the whole scan and not just close the the sensor. These features are then clustered with a k-means clustering and the cluster centers used as the dictionary words.

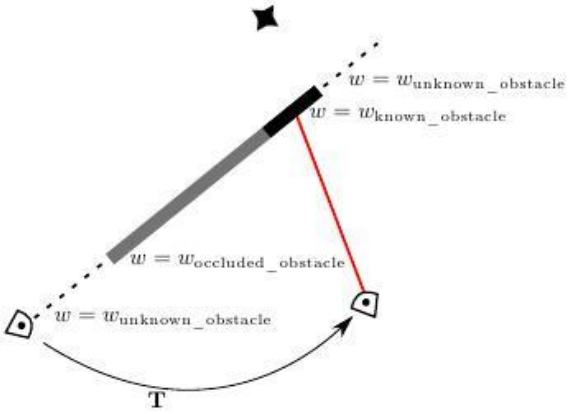


Fig. 1. Extended sensor model.

GPU acceleration

We identified keypoint detection and transformation scoring as bottlenecks which can easily be reduced in computation time using the GPU. The corresponding parts of the algorithms have been coded in CUDA.

Results

(FB) published and used in [1] and 7 own data sets (KO) recorded on our campus. The FB data consist of 77 laser scans with approximately 170.000 points per scan and inter-scan distances of about 10m while our data sets consist of mostly scans with about 750.000 points, while some are recorded at a resolution of 1.500.000 points. The offset between the scans vary between 5m and 12m and each sub-map generated from a single data set is connected to each other sub-map with at least a small area of overlap.

For both data sets, the DoG keypoints (detector radius: 0.3m) were combined with SHOT-U descriptors (descriptor radius of 2.1m) at two scales.

All incremental registrations in the single data sets were performed successfully when using the multiscale registration algorithm. In the

FB data, the multiscale registration was not necessary for this result, while there was a single failed incremental registration in the KO data with the single-scale registration.
The place recognition component using the BoW descriptor was evaluated with the same criteria as used by Steder et al [1]: Two scans that are less than 10m apart in the ground truth map and are correctly registered are defined to be a true positive.
In the FB data, 100% precision at 100% recall were possible by evaluating at least 15 BoW hypotheses per scan. With only 5 evaluated hypotheses as feasible in online procesing, still a recall rate of 89.7% could be achieved.
The KO data contain many ambiguities, such as the long corridor-like environments shown in Fig. 2. Because of these, only a recall rate of 50.5% was possible without false positives when matching all scans against all other, and a rate of 38.2% when only evaluating the 5 best hypotheses. Note that some false positives can be tolerated when using a robust backend such as the one proposed by Olson et al [3], resulting in higher recall rates.
Even when only evaluating the 5 best hypotheses, all loops inside and between the data sets have been closed correctly, resulting in the full map of Fig. 2 that is generated fully automatically.
On average, the feature computation took 640ms (810ms without GPU acceleration), while each registration took 110ms (430ms) and the final ICP step took 90ms.

Conclusions

We designed and tested as system that can register and fuse 3d laser scans with considerable inter-scan distance, for which other algorithms fail. The main challenge was to find robust features for matching. A new multi-scale implementation of DoG keypoints in conjunction with SHOT descriptors provided the best results. Using place recognition we were able to generate complex maps from large data sets. The system works in real-time and has been evaluated on standard data sets.

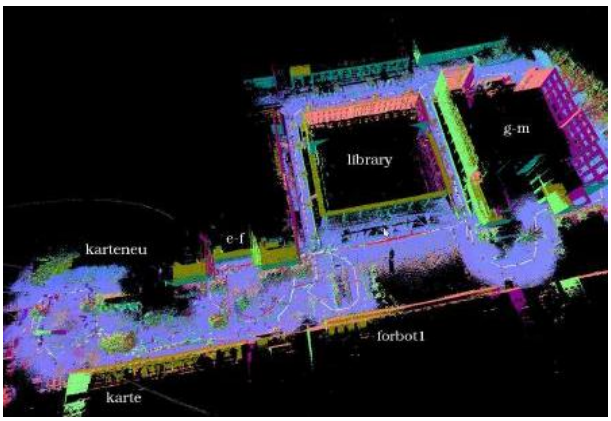


Fig. 2. Combined 3D map generated from a total of 7 data sets.

References

1. Bastian Steder, Michael Ruhnke, Slawomir Grzonka, and Wolfram Burgard. Place recognition in 3d scans using a combination of bag of words and point feature based relative pose estimation. In Proc. of the International Conference on Intelligent Robots and Systems (IROS), 2011.
2. Edwin Olson. Real-time correlative scan matching. In Proc. of the IEEE International Conference on Robotics and Automation, 2009., p. 4387–4393. IEEE, 2009.
3. Edwin Olson and Pratik Agarwal. Inference on networks of mixtures for robust robot mapping. In Proc. of Robotics: Science and Systems (RSS), Sydney, Australia, 2012.
4. Stefan Holzer, Jamie Shotton, and Pushmeet Kohli. Learning to efficiently detect repeatable interest points in depth data. In Proc. of the 12th European conference on Computer Vision (ECCV), ECCV'12, p. 200–213, Springer, Berlin, 2012.
5. Andrei Zaharescu, Edmond Boyer, Kiran Varanasi, and Radu Horaud. Surface feature detection and description with applications to mesh matching. In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 373–380, Miami Beach, Florida, 2009. IEEE.
6. Luís A. Alexandre. 3d descriptors for object and category recognition: a comparative evaluation. In Workshop on Color-Depth Camera Fusion in Robotics at the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vilamoura, Portugal, 2012.

REGRESSION BASED ON CONVEX COMBINATIONS AS A VARIANT OF REGULARIZED LEAST SQUARES REGRESSION[1]

O.V. Senko[2,3], A.A. Dokukin[2,4]

[2] Institution of Russian Academy of Sciences Dorodnicyn Computing Centre of RAS, e-mail: [3] senko@ccas.ru, [4] email alex_dok@mail.ru

A new multiple regression method based on optimal convex combinations of simple univariate regressions is discussed, where simple regressions are searched with an ordinary least squares technique. Convex combination is considered optimal if it correlates with the response variable in the best way. It is shown that the developed approach is equivalent to a least squares technique variant regularized by constraints on signs of regression parameters.

Introduction

Standard multiple regression task is considered. Response variable Y is predicted by variables $X_1,\ldots,X_n$ with the help of linear regression function $\beta_0+\sum_{i=1}^{n}\beta_i X_i$. It is assumed that the vector $\boldsymbol{\beta}=(\beta_0,\ldots,\beta_n)$ is chosen from $\Re^{n+1}$ by observations $(y_j,x_{1j},\ldots,x_{nj})$, where $j\in\{1,\ldots,m\}$. Ridge regression, lasso [1] and elastic net [2] are now popular methods for searching regression coefficients in task with high dimensional data. These methods are based on solving optimization task of the type

$$\min_{\boldsymbol{\beta}\in\Re^{n+1}}\{\sum_{j=1}^{m}(y_j-\beta_0-\sum_{i=1}^{n}\beta_i x_{ij})^2+\theta P(\boldsymbol{\beta})\}$$

where $\theta>0$ and $P(\boldsymbol{\beta})=P(\beta_0,\ldots,\beta_n)$ is penalty function. At that the lasso penalty is $\sum_{i=1}^{n}|\beta_i|$, ridge penalty is $\sum_{i=1}^{n}\beta_i^2$, the elastic net penalty is

$$(1-\alpha)\sum_{i=1}^{n}|\beta_i|+\alpha\sum_{i=1}^{n}\beta_i^2$$

where parameter $\alpha\in[0,1]$ characterizes compromise between ridge and lasso penalty. Extremely high efficiency of regularization path based on the elastic net regularization path was shown in variety of tasks. However it is not known if elastic net achieves upper efficiency limits. Let's note that another regularization way exists that is based on solving optimization task with constraints

$$\begin{gathered}\min_{\boldsymbol{\beta}\in\Re^{n+1}}\{\sum_{j=1}^{m}(y_j-\beta_0-\sum_{i=1}^{n}\beta_i x_{ij})^2\}\\ C_1(\beta_0,\ldots,\beta_n)\ge 0\\ \ldots\ldots\ldots\ldots\\ C_k(\beta_0,\ldots,\beta_n)\ge 0\end{gathered}\qquad(1)$$

This way of regularization is used for example in lasso variant described in [3]. Further we shall consider constraints based on demand that sign of regression coefficient for variable X_i is equal to sign of Pearson correlation coefficient between Y and X_i. This coefficient will be denoted as $K(Y,X)$.

Let's consider now regression method that is based on optimal convex combination of simple regressions that are built by each of variables $X_1,\ldots,X_n$. Convex combinations are widely used in pattern recognition. The bagging and boosting techniques [4,5] may be mentioned as an example, as well as methods based on collective solutions by sets of regularities [6]. Convex correction is used in regression tasks also. Thus, neural networks ensembles based on optimal balance between individual forecasting ability of predictors and divergence between them are discussed in [7]. In works [8] a method was discussed that is

[1]Support of the RFBR grants 11-07-00715-a and 12-01-90012-Бел_а

based on searching optimal convex combinations of simple one-dimensional linear regressions $R_1,\ldots,R_n$ calculating Y from variables $X_1,\ldots,X_n$ with the minimal error at training set. Efficient method for calculating such optimal convex combinations was developed. Experiments with simulated data demonstrated rather high forecasting ability of this method. It was also shown that variances of convex combinations usually are small. This fact may be related to decomposition of variances that is true for convex combinations. Let $\tilde{Z}=\{Z_1,\ldots,Z_l\}$ be some set of random variables defined at probability space Ω and $\hat{Z}=\sum_{i=1}^{l}c_iZ_i$ is arbitrary convex combination of variables from . Then decomposition

$$V(\hat{Z})=\sum_{i=1}^{l}c_iV(Z_i)-\tfrac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}c_ic_j\rho(Z_i,Z_j) \qquad (2)$$

is true for variance of $V(\hat{Z})$, where

$$\rho(Z_i,Z_j)=E_\Omega(Z_i-E_\Omega Z_i-Z_j+E_\Omega Z_i)^2$$

and $V(Z_i)$ is the variance of Z_i The decomposition was discussed in [8]. It follows from (2) that $V(\hat{Z})\le\sum_{i=1}^{l}c_iV(Z_i)$. But low variance leads to a decrease of forecasting ability. So an additional linear transformation is necessary to improve it. It is natural to search such a transformation as a simple linear regression calculating Y from $\hat{R}=\sum_{i=1}^{n}c_iR_i$. But forecasting error of simple regression $\beta_0+\beta_1X$ monotonly depends on correlation coefficient between Y and X: $K(Y,X)$. So, the search of convex combination with the best forecasting ability is naturally reduced to a search of convex combination that in the best way correlates with the response. A regression method calculating convex combination of simple linear regressions $\hat{R}_b$ with maximal $K(Y,\hat{R}_b)$ was discussed in [9,10] . It was shown at that efficiency of this method is close to efficiency of elastic net method in simulated tasks where target Y and predictors X were calculated as linear combination of several latent variables. For some scenarios performance of method based on searching of optimal convex combination was even better. This paper is aimed to show that regression method calculating convex combination of simple linear regressions $\hat{R}_b$ with maximal $K(Y,\hat{R}_b)$ is equivalent to calculating linear regression coefficients as solution of optimization task of the type (1).

Optimal convex combination as a way of regularization

Let's consider a method of optimal convex combination that consists of three stages. At the first stage we construct with the help of *OLS* technique a set of *n* simple linear regression models for predicting response Y from each of covariates $X_1,\ldots,X_n$:

$$\tilde{R}=\{R_1=\beta_{01}^u+\beta_{11}^uX_1,\ldots \\ \ldots,R_n=\beta_{0n}^u+\beta_{1n}^uX_n\}$$

At the second stage such convex combination of regression models $R_1,\ldots,R_n$ is searched that correlation coefficient between this combination and response Y is maximal. In other words we search such $c_1^b,\ldots,c_n^b$ in a standard simplex that for $\hat{R}_b=\sum_{i=1}^{n}c_i^bR_i$ inequality $K(Y,\hat{R}_b)\ge K(Y,\hat{R})$ is true for arbitrary convex combination $\hat{R}=\sum_{i=1}^{n}c_iR_i$ where $(c_1,\ldots,c_n)$ belongs to a standard simplex. At the third stage a simple linear regression function $\beta_0^c+\beta_1^c\hat{R}_b$ calculating response Y from convex combination $\hat{R}_b$ is searched from observations with the help of OLS. The described method will be further referred to as Convex Regression (CR) and the regression function

$$\beta_0^c+\beta_1^c(\sum_{i=1}^{n}c_i^b\beta_{0i}^u)+\beta_1^c(\sum_{i=1}^{n}c_i^b\beta_{1i}^uX_i)$$

will be called a CR solution. Let $I_0=\{1,\ldots,n\}\setminus I_p$ and $I_p=\{i\,|\,K(Y,X_i)>0,\ i\in I_0\}$. It is evident that coefficient $\beta_{1i}^u=0$ is equal 0 if $K(Y,X_i)=0$. Suppose

that some CR solution is built by convex combination $\sum_{i=1}^{n} c_i R_i$ where $\exists i \in I_0$ such that $c_i > 0$. Then identical convex combination exists where $c_i > 0$ only when $i \in I_p$. Let's try to show that CR is equivalent to the following variant of LS technique with constraints of the type (1). A vector of regression coefficients is a solution of the following optimization task:

$$\min_{\boldsymbol{\beta} \in \mathbf{R}^n} \left\{ \frac{1}{m} \sum_{j=1}^{m} (y_j - \beta_0 - \sum_{i=1}^{n} \beta_i x_{ij})^2 \right\}$$

$$\beta_i K(Y, X_i) \geq 0, \; i \in I_p$$

$$\beta_i = 0, \; i \in I_0$$

Two statements are true.

Statement 1 *Let $\hat{R}$ be some convex combination with positive variances of regression functions $\{R_i \mid i \in I_p\}$ and β_0^c, β_1^c are coefficients of linear OLS regression calculating response Y from the convex combination Then the constraints from the task (3) are satisfied for coefficients $\beta_1, \ldots, \beta_n$ of a CR solution.*

Indeed, we must show that $sign(\beta_i) = sign[K(Y, X_i)]$. For each linear regression $R_i = \beta_{0i}^u + \beta_{1i}^u X_1$ equality $sign(\beta_{1i}^u) = sign[K(Y, X_i)]$ is true. So, $K(Y, R_i) \geq 0, i = 1, \ldots, n$. Correlation coefficient $K(Y, \hat{R}) \geq 0$ because $\hat{R}$ is sum of functions that are positively correlated with Y. So $\beta_1^c \geq 0$. It follows from equality $\beta_i = c_i \beta_1^c \beta_{i1}^u$ that two possibilities exist: $sign(\beta_i) = sign(\beta_{i1}^u)$ or $\beta_i = 0$. Q.E.D.

Statement 2 *For any regression coefficients β_0 and $\beta_i, (i \in I_p)$ satisfying constraints of the task (3) there are such points $c_1, \ldots, c_n$ in a standard simplex and numbers β_0^c and β_1^c that the following equality is true*

$$\beta_0 + \sum_{i \in I_p} \beta_i = \beta_0^c + \beta_1^c \sum_{i \in I_p} c_i R_i \qquad (4)$$

Indeed, let $\hat{c}_i = \frac{|\beta_i|}{|\beta_{i1}^u|}$ when $i \in I_p$. Equality (4) is true when $c_i = \hat{c}_i / \sum_{j \in I_P} \hat{c}_j$, $\beta_1^c = \sum_{j \in I_p} \hat{c}_j$ $\beta_0^c = \beta_0 - \sum_{j \in I_p} \hat{c}_j \beta_{i0}^u$.

Theorem describing relation between two types of solutions follows from statements 1 and 2.

Theorem 1 *Any solution of the task (3) is a CR solution. Any CR solution is also a solution of the task (3).*

Suppose that vector $(\beta_0^*, \ldots, \beta_n^*)$ is a solution of the optimization task (3). Let's note that according to the Statement2 coefficients $(\beta_0^*, \ldots, \beta_n^*)$ can be received from regressions $R_1, \ldots, R_n$ by use of corresponding convex combination $\hat{R}^* = \sum_{i=1}^{n} c_i^* R_i$ and linear transformation $\beta_0^{c*} + \beta_1^{c*} \hat{R}^*$. The latter must be OLS regression calculating Y from $\hat{R}^*$. Let's suppose that β_0^{c*}, β_1^{c*} are not coefficients of OLS regression. Then squared error for $(\beta_0^*, \ldots, \beta_n^*)$ is less than squared error for coefficients $(\beta_0^{**}, \ldots, \beta_n^{**})$ that are received from convex combination $\hat{R}^*$ with the help of OLS regression. Besides according the statement 1 constraints (3) are satisfied for $(\beta_0^{**}, \ldots, \beta_n^{**})$. So vector $(\beta_0^*, \ldots, \beta_n^*)$ is not a solution of task (3).

Let's suppose that $K(Y, \hat{R}^*)$ is not maximal. In other words there is such a point $(c_1', \ldots, c_n')$ in a standard simplex that for $\hat{R}' = \sum_{i \in I_p} c_i' R_i$

$$K(Y, \hat{R}^*) < K(Y, \hat{R}') \qquad (5)$$

Suppose that $\beta_0^{c'}$, $\beta_1^{c'}$ are coefficients of OLS regression calculating Y from $\hat{R}'$. Let's $(\beta_1', \ldots, \beta_n')$ are such coefficients that the equation

$$\beta_0' + \sum_{i \in I_p} \beta_i' X_i = \beta_0^{c'} + \beta_1^{c'} \hat{R}' \quad \text{is true.}$$

It follows from the statement 1 that constraints of task (3) are satisfied for coefficients $(\beta_0',\ldots,\beta_n')$. But it is known that squared error of OLS regression calculating Y from some variable Z monotonly depends on $K(Y,Z)$. So squared error for $(\beta_0',\ldots,\beta_n')$ is less than squared error for $(\beta_0^*,\ldots,\beta_n^*)$. Thus, there is a contradiction and any solution of task (3) is also CR solution. The first statement of theorem is proved.

Let $(\beta_0^*,\ldots,\beta_n^*)$ be CR solution. In other words there is such a convex combination $\hat{R}^*$ that the inequality $K(Y,\hat{R}^*)\geq K(Y,\hat{R})$ is true for any convex combination $\hat{R}$ and equation $\beta_0^*+\sum_{i\in I_p}\beta_i^* X_i=\beta_0^{c*}+\beta_1^{c*}\hat{R}^*$ is true for coefficients β_0^{c*}, β_1^{c*} of OLS regression calculating Y from $\hat{R}^*$. Let's suppose that $(\beta_0^*,\ldots,\beta_n^*)$ is not a solution of the task (**3**). Then there are such coefficients $(\beta_0',\ldots,\beta_n')$ that squared error of Y prediction for function $\beta_0^*+\sum_{i\in I_p}\beta_i^* X_i$ is greater than squared error for function $\beta_0'+\sum_{i\in I_p}\beta_i' X_i$.

It follows from the statement **2** that there is such convex combination $\hat{R}'$ and numbers $\beta_0^{c'}$ and $\beta_1^{c'}$ that equation $\beta_0'+\sum_{i\in I_p}\beta_i' X_i=\beta_0^{c'}+\beta_1^{c'}\hat{R}'$ is true.

Suppose that β_0^{ls} and β_1^{ls} are coefficients of OLS regression calculating Y from $\hat{R}^*$. It is evident that squared error of Y prediction for linear function $\beta_0^{c'}+\beta_1^{c'}\hat{R}'$ is less than squared error for $\beta_0^*+\sum_{i\in I_p}\beta_i^* X_i$. But then equation $K(Y,\hat{R}')>K(Y,\hat{R}^*)$ holds. So, there is a contradiction and $(\beta_0^*,\ldots,\beta_n^*)$ is a solution of the task (**3**). Thus, the second statement is true and the theorem is proved.

Conclusion

Thus equivalence of two methods was shown in this paper:

convex regression based on search of convex combination of simple linear regression that in the best way correlates with target;

LS model with regularization by constraints on signs of regression coefficients.

In other words convex regression is equivalent to linear regression variant when regression coefficients are searched as minimum point of function $\sum_{j=1}^{m}(y_j-\beta_0-\sum_{i=1}^{n}\beta_i x_{ij})^2$ at set of points from $\Re^n$ satisfying constraints: $sign(\beta_i)=sign[K(Y,X_i)]$, $\beta_i=0$ when $K(Y,X_i)=0$, $i=1,\ldots,n$. So sufficiently high efficiency of CR may be ascribed to commonly high efficiency of different regularization ways in machine learning models. But in case of CR constraints are put not on size but on signs of regression coefficients.

References

1. Tibshirani, R. Regression shrinkage and selection viathe lasso. J. Roy. Stat. Soc., 58:267–288, 1996.
2. Zou, H. and Hastie, T. Regularization and variable selection via the elastic net. J. Roy. Stat. Soc., 67
3. Efron, B., Hastie, T., Jonnstone, I., and Tibshirani, R. Least angle regression. Annals of Statistics, 32 (2):407–499, 2004.
4. Breiman, L. Random forests - random features. Technicalreport, Statistics department, University ofCalifornia, Berkley, 1999.
5. Kuncheva, L.I. Combining Pattern Classifiers. Methods and Algorithms. Wiley Interscience, New Jersey,2004.
6. Zhuravlev, Yu. I., Kuznetsova, A. V., Ryazanov, V. V. ,Senko, O. V., and Botvin, M. A. The use of patternrecognition methods in tasks of biomedical diagnostics and forecasting. Pattern Recognition and Image Analysis, 58(2):195–200, 2008.
7. Brown, G., Wyatt, J. L., and Tino, P. Diversity in regression ensembles. Journal of Machine Learning,6:1621–1650, 2005.
8. Senko, O. V. An optimal ensemble of predictors in convex correcting procedures. Pattern Recognition and Image Analysis, 19(3):465–468, 2009.
9. Senko, O. V. and Dokukin, A. A. Optimal forecasting based on convex correcting procedures. In New Trends in Classification and Data Mining, pp. 62–72. ITHEA, Sofia, Bulgaria, 2010.
10. Oleg Senko, Alexander Dokukin Correlation Maximization in Regression Models Based on Convex Combinations. International Journal "Information Theories and Applications",v.18, N3, pp. 224-231.

TRACK 1 "MATHEMATICAL METHODS IN PATTERN RECOGNITION"

ON UNIMODALITY AND BIMODALITY OF A TWO-COMPONENT GAUSSIAN MIXTURE WITH DIFFERENT VARIANCES

N.N. Aprausheva[1,2], V.V. Dikusar[1,3], S.V. Sorokin[1,4]

[1] Dorodnicyn Computing Centre, Russian Academy of Sciences, Moscow, Russia
[2] plat@ccas.ru, [3] dikussar@yandex.ru [4] www2013@ccas.ru

A few sufficient conditions of unimodality and bimodality of a two-component Gaussian mixture with different variances is formed and a boundary equation of domains of unimodality and bimodality of this mixture is given. The proofs are based on principle of contractive mapping, analytical and experimental methods.

Introduction

Gaussian mixtures are widely used as general approximators in many fields of science: mathematical modelling, physics, biology, geography and others [1-4]. Popularity of the mixtures needs solutions of such problems as the definition of mode number and boundary of unimodal and multimodal domains. In general the problems have not been solved. In the particular case, the exhaustive results have been obtained for two-component mixtures with equal variances [5, 6]..

In this work sufficient conditions of unimodality and bimodality of the two-component Gaussian mixture with different variances are given. Probability density of the mixture is represented as:

$$f(x) = (2\pi)^{-\frac{1}{2}} \sum_{i=1}^{2} f_i(x, \pi_i, \sigma_i, \mu_i), \qquad (1a)$$

$$f_i(x, \pi_i, \sigma_i, \mu_i) = \pi_i \sigma_i^{-1} \exp\left[-(x - \mu_i)^2 (2\sigma_i^2)^{-1}\right], \qquad (1b)$$

π_i is weight of *i*-th component, $\pi_i \in (0,1)$, $\pi_1 + \pi_2 = 1$, σ_i^2 is its variance $\sigma_1^2 \neq \sigma_2^2$, μ_i is its expectation value, $\mu_1 \neq \mu_2$, $x \in (-\infty, \infty)$. The distribution parameters of the mixture μ_1, μ_2, σ_1^2, σ_2^2, π_1 are known.

Probability density mode $f(x)$ is the point of its local maximum, which is a root of the equation [7]

$$f_x'(x) = 0, \qquad (2)$$

$$f_x'(x) = (2\pi)^{-\frac{1}{2}} \sum_{i=1}^{2} \sigma_i^{-2} (\mu_i - x) f_i(x).$$

All roots of the equation of (2) are on the segment $[\mu_1, \mu_2]$, $\mu_1 < \mu_2$, this equation is equivalent to the equation [8]

$$x = \varphi(x),$$

$$\varphi(x) = \left(\sum_{i=1}^{2} \sigma_i^{-2} \mu_i f_i(x)\right) \left(\sum_{i=1}^{2} \sigma_i^{-2} f_i(x)\right)^{-1}.$$

It was proved that the function $\varphi(x)$ is a contracting one, and probability density is unimodal if one of two inequalities is true [8]:

1) $\rho_{max}^2 \leq 4$,

$$\rho_{max}^2 = (\mu_2 - \mu_1)^2 \sigma_{min}^{-2}, \qquad (3)$$

$$\sigma_{min} = \min(\sigma_1, \sigma_2), \quad \rho_{min}^2 = (\mu_2 - \mu_1)^2 \sigma_{max}^{-2}; \qquad (4)$$

$$2)\quad \rho_{max}^2 - 2 < \beta e^{-\frac{\rho_1^2}{2}} + \beta^{-1} e^{-\frac{\rho_2^2}{2}}, \rho_{max}^2 > 4, \tag{5}$$

$$\rho_1^2 = (\mu_2 - \mu_1)^2 \sigma_1^{-2},$$

$$\rho_2^2 = (\mu_2 - \mu_1)^2 \sigma_2^{-2}, \tag{6}$$

$$\beta = (\pi_1 \sigma_2^3)(\pi_2 \sigma_1^3)^{-1}.$$

If for $\rho_{max}^2 > 4$ the condition (5) is not true, then the investigated mixture can be either unimodal or bimodal. The domain boundary of unimodality and bimodality is the set of the degenerated critical inflection points of probability density $f(x)$ [9].

The search of other sufficient conditions of unimodality for $\rho_{max}^2 > 4$ is based on the postulate: if for all values of $x \in [\mu_1, \mu_2]$ the function value $f_i(x, \pi_i, \sigma_{min}, \mu_i)$, $i \in \{1, 2\}$, with minimal variance σ_{min}^2 is not less than the value of the function $f_j(x, \pi_j, \sigma_{max}, \mu_j)$ with the maximum variance σ_{max}^2, $j \neq i$, $j \in \{1, 2\}$, then the probability density of $f(x)$ is unimodal.

For definiteness we assume $\sigma_{min} = \sigma_1$, $\sigma_{max} = \sigma_2$, we denote by $\pi_{(min)}$ the component weight, which has a minimal variance σ_{min}^2, by $\pi_{(\max)}$ — the weight of component which has a maximum variance σ_{max}^2; there are possible variants: $\pi_{(min)} = \pi_{max}$ or $\pi_{(min)} = \pi_{min}$. Then our affirmation of unimodality of the function $f(x)$ is expressed by the inequality ($x = \mu_2$):

$$f_1(\mu_2, \pi_{(min)}, \sigma_{min}, \mu_1) \geq f_2(\mu_2, \pi_{(max)}, \sigma_{max}, \mu_2),$$

which by virtue of (1b), (3) is converted into

$$\pi_{(min)} \sigma_{min}^{-1} e^{-\frac{\rho_{max}^2}{2}} \geq \pi_{(max)} \sigma_{max}^{-1}. \tag{7}$$

Having performed elementary algebraic operations in inequality (7), we obtain

$$ln\left(\frac{\pi_{(min)}}{\pi_{(max)}}\right) \geq ln\left(\frac{\sigma_{min}}{\sigma_{max}}\right) + \frac{\rho_{max}^2}{2}. \tag{8}$$

Inequality (8) is a «hard» condition of unimodality of the investigated mixture [10]. For obtaining «softer» condition we use the results of the work [6]: for $\sigma_1 = \sigma_2$, $\pi_1 \neq \pi_2$ and $\rho^2 > 4$, $\rho^2 = (\mu_2 - \mu_1)^2 \sigma^{-2}$ the mixture is unimodal if this inequality takes placeL

$$\ln\left(\frac{\pi_{max}}{\pi_{min}}\right) \geq \frac{\rho^2}{2}\sqrt{1 - 4\rho^{-2}} - 2\ln\left(\frac{\rho + \sqrt{\rho^2 - 4}}{2}\right). \tag{9}$$

In this case the equation

$$\ln\left(\frac{\pi_{max}}{\pi_{min}}\right) = \frac{\rho^2}{2}\sqrt{1 - 4\rho^{-2}} - 2\ln\left(\frac{\rho + \sqrt{\rho^2 - 4}}{2}\right) \tag{10}$$

is an equation of a domain boundary of unimodality and bimodality of the mixture.

For «softening» the inequality (8) we introduce correction functions into its right part by using the one (9):

$$ln\left(\frac{\pi_{(min)}}{\pi_{(max)}}\right) \geq ln\left(\frac{\sigma_{min}}{\sigma_{max}}\right)\theta(\rho_{max}) + \frac{\rho_{max}}{2}\sqrt{\rho_{max}^2 - 4} - 2\ln\left(\frac{\rho_{max} + \sqrt{\rho_{max}^2 - 4}}{2}\right). \tag{11}$$

For $\sigma_1 = \sigma_2$ and $\pi_{(min)} > \pi_{(\max)}$ the inequality (11) coincides with inequality (9). An expression for the function $\theta(\rho_{max})$ is defined experimentally by a trial-and-error method.

$$\theta(\rho_{max}) = 2 + \frac{\sqrt[3]{\rho_{max}}}{2} + \frac{\sqrt[9]{\rho_{max}}}{2 \cdot 3} + \frac{\sqrt[27]{\rho_{max}}}{2 \cdot 3 \cdot 4}. \tag{12}$$

The formula (12) provides the precision of the boundary equation to ε, $|\varepsilon| \leq 0.122$. The value ε was calculated in each experiment for

different values $\tau = \sigma_{min}\sigma_{max}^{-1}$ and ρ_{max}, $\rho_{max} > 2$, $\tau = 0.9, 0.8, \ldots, 0.1$, $\rho_{max} = 2.5, 3.0, 3.5, 4.0, 4.5$.

The equation of the searched boundary is defined in the form:

$$ln\left(\frac{\pi_{(min)}}{\pi_{(max)}}\right) = ln\left(\frac{\sigma_{min}}{\sigma_{max}}\right)\theta(\rho_{max}) + \frac{\rho_{max}}{2}\sqrt{\rho_{max}^2 - 4} - 2\ln\left(\frac{\rho_{max}+\sqrt{\rho_{max}^2-4}}{2}\right) + \varepsilon_i.. \quad (13)$$

In all experiments the equality (13) is executed with accuracy within ε_i, $i = 1, 2, \ldots, n$, $n>500$, $|\varepsilon_i| < 0.122$.

For our accommodation in expression (13) we introduce the notation

$$\zeta(\rho_{max}) = \frac{\rho_{max}}{2}\sqrt{\rho_{max}^2 - 4} - 2\ln\left(\frac{\rho_{max} + \sqrt{\rho_{max}^2 - 4}}{2}\right) \quad (14)$$

and formulate sufficient conditions of unimodality and bimodality of the investigated mixture.

Affirmation. For $\rho_{max}^2 > 4$ and $\pi_{(min)} \neq \pi_{(max)}$, $\pi_1 \neq \pi_2$ the investigated mixture is unimodal if

$$ln\left(\frac{\pi_{(min)}}{\pi_{(max)}}\right) > ln\left(\frac{\sigma_{min}}{\sigma_{max}}\right)\theta(\rho_{max}) + \zeta(\rho_{max}) + 0.122, \quad (15)$$

and it is bimodal if

$$ln\left(\frac{\pi_{(min)}}{\pi_{(max)}}\right) < ln\left(\frac{\sigma_{min}}{\sigma_{max}}\right)\theta(\rho_{max}) + \zeta(\rho_{max}) - 0.122. \quad (16)$$

The functions $\theta(\rho_{max})$, $\zeta(\rho_{max})$ are defined by the formulae (12), (14).

Conclusion

For a two-component Gaussian mixture with different variances a few sufficient conditions of its unimodality and bimodality have been formulated and the boundary equation of its domains of unimodality and bimodality have been given.

References

1. Harris N. and Smith S. A. B. The sib-sib age of on set correlation among individuals suffering from a hereditary syndrome produced by more than one gene. // Annals of Eugenics. London, 1949. V. 14. Part 4. P. 309-318.
2. Aprausheva N. N., Gorlach I. A., Zhelnin A. A., Sorokin S. V. An experiment on Automated Statistical Recognition of Clouds. // J. Computational Mathematics and Mathematical Physics. 1998. V. 38. № 10. P. 1715-1719.
3. Carreira-Perpiñán N. A. Mode-finding for mixture of Gaussian distributions. // IEEE Trans. On Pattern Analys. and Mach. Intell. 2000. 22. № 11.P. 1318-1323.
4. Di Crescenzo A., Martinucci B. On a symmetric nonlinear birth-death process with bimodal transition probabilities // Symmetry, 2009, T. 1, N 2,pp.201-214.
5. Aprausheva N. N., and Sorokin S. V. On the Unimodality of a Simple Gaussian Mixture. // J. Computational Mathematics and Mathematical Physics, 2004. V. 44. №. 5. P. 785-793.
6. Aprausheva N. N., Sorokin S. V. The exact equation of a boundary of unimodal and bimodal domains of the simplist Gaussian Mixture. // 8th Open German-Russian Workshop. Pattern Recognition and Image Understanding. Nizhy Novgorod, The Russian Federation, 2011.
7. Cramér H. Mathemalical Methods of Statistics. Princeton University Press, Prince-ton, N.J., 1946.
8. Aprausheva N.N. On sufficient conditions of unimodality and bimodality tw0-component Gaussian mixture. Advanced science, №2 (2), Kirov, 2012, pp. 22-31 (in Russian) (http://www.vyatsu.ru/nash-universitet/nauchnyiy-elektronnyiy-zhurnal-advanced-science.html).
9. Arnold V. I., Varchenko A. N., and Gusein-Zade S. M., 1982. Singularities of Differitiable Maps, Nauka, 1982, Moscow ; Birkhäuser, Boston, 1985.
10. Arnold V. I. «Hard» and «soft» mathematical models. Moscow, MTsNMO, 2011, P. 32. (in Russian).

LOGICAL CORRECTORS IN RECOGNITION PROBLEMS[1]

E.V. Djukova[2,4], M.M. Lyubimtseva[3], P.A. Prokofjev[2,5]

[2] Dorodnitsyn Computing Centre, Russian Academy of Sciences, Vavilov st. 40, Moscow, 119333 Russia, e-mail: [4] edjukova@mail.ru, [5] p_prok@mail.ru
[3] Faculty of Computational Mathematics and Cybernatics, Moscow State University, Moscow, 119234 Russia, e-mail: m.lyubimtseva@gmail.com

A problem of constructing for logical correctors (recognition algorithms based on incorrect elementary classifiers) is considered. A new special type of logical corrector is developed. This model is tested on the real tasks.

Introduction

The algebraic-logical approach to construct the correct recognition procedures is considered [2, 3, 7]. The correctness of the recognition procedure is ensured by existence of a correcting Boolean function for a set of elementary classifiers (EC). It is not necessary that all EC are correct. First the idea of the algebraic-logical approach is suggested in [3]. The practical applications of this method are studied in [1, 4]. The object is classified by voting on so called correct sets of EC. A monotonic Boolean function is used in [1, 4] as a correcting function. In this case a good result is obtained. To reduce a computational complexity of searching for relevant correct sets of EC the genetic approach is applied [6].

In this work a new special type of logical corrector is suggested and tested on many real recognition tasks. Also several well-known recognition algorithms are tested on the same recognition tasks.

Main results

We consider the discrete approach to learning recognition algorithms by analyzing a set of precedents (the training sample). It is impotent to construct the correct recognition procedures. A recognition algorithm is called correct if it makes no errors on the training sample.

The classical logical recognition algorithms are based on construction of the fragments of the sample descriptions. The constructed fragments must reflect certain regularities in a set of precedents. Such fragments serve as EC and allow one to classify the objects of unknown origin. Any EC H is called correct if there is no pair of the training objects S_i, S_j from different classes such that H meet both S_i and S_j. The logical data analysis is aimed at determining the most relevant correct EC

The problem of construction for correct recognition procedures is solved by the algebraic approach [7] as following. A heuristic data-processing model of the algorithms is chosen (for example, a version of the model of algorithms for the calculation of estimates). A family of correcting functions is chosen (for example, a set of polynomials, functions with a bounded derivative, monotonic Boolean functions). Then certain set of basis algorithms is selected within the framework of chosen heuristic data-processing model by analyzing the training sample. Finally a corrector (an operation over selected set of basis algorithms) is constructed such that the resulting algorithm is correct.

In [3] any set of EC is called correct if there exists a Boolean function as correcting function. Enumeration of correct sets of EC is reduced to enumeration of covers for the Boolean matrix formed in a special way by the training sample. The algebraic-logical data analysis is aimed at determining the most relevant correct sets of EC. The object is

[1] Russian Foundation for Basic Research (project no. 13-01-00787-a) . Grant of the President of the Russian Federation in Support of Leading Scientific Schools (NSh 4652.2012.1)

classified by voting on the correct sets of EC constructed (the estimates for each class are calculated).

A monotonic Boolean function is used in [4] as a correcting function. In this case obtained result was better than in non-monotonic case. All correct sets of EC consist of single feature-value EC. To reduce a computational complexity of searching for relevant correctors the genetic approach from [6] is applied. The logical corrector constructed in [4] is called monotonic logical corrector (MON).

In [1] the logical corrector LOBAGA with monotonic Boolean correcting function is constructed. The correct sets of EC consist of EC with no limit on feature-value number. Any set of EC $\mathfrak{U}$ is called a local basis if abundant family of correct set of EC is embedded in $\mathfrak{U}$. The learning stage is iterative. At each iteration local basis $\mathfrak{U}$ is constructed. The most relevant correct sets of EC are searched within $\mathfrak{U}$ by the genetic algorithm from [6]. For each correct set of EC the weight is calculated. The weight of correct set is the characteristic of its relevance. The estimates for the training sample are calculated. The constructed correct sets of EC strive to improve the estimates obtained at previous iterations. The number of iteration is the parameter of the algorithm. At recognition stage the objects are classified by the weighted voting on the constructed correct sets of EC.

In this work, following concept of the monotonic corrector we consider a new type of the logical corrector that called *antimonotonic* logical corrector (AMON). The algorithms MON, AMON, LOBAGA are tested on the real recognition tasks from the system "Recognition" [8] and compared with several well-known algorithms.

The results are following. The logical correctors have better performance than the classical logical recognition algorithms (such that the algorithm for the calculation of estimates, test recognition algorithm) and decision trees (C5.0, binary decision trees, full decision trees [5]). The performance of the logical correctors MON and AMON is similar to that of all other algorithms tested but worse than Neuron network. The best performance has been obtained by the logical corrector LOBAGA.

Conclusions

1. Main concepts of the algebraic-logical approach is considered.
2. A new type of corrector is suggested that called an antimonotonic logical corrector.
3. It is shown that the logical correctors have good performance.

References

1. Dyukova E.V., Prokofjev P.A. Models of Recognition Procedures with Logical Correctors // Pattern Recognition and Image Analysis, 2013, Vol. 23, No. 2, pp. 235–244
2. E.V. Djukova, Yu.I. Zhuravlev. Discrete Methods of Information Analysis in Recognition and Algorithm Synthesis // Pattern Recognition and Image Analysis. 1998. Vol.7. No.2. Pp.192–207.
3. Dyukova, E. V.; Zhuravlev, Yu. I.; Rudakov, K. V. On the algebro-logical synthesis of correct recognition procedures based on elementary algorithms. Zh. Vychisl. Mat. i Mat. Fiz. 36 (1996), no. 8, 215-223 [in Russian].; translation in Comput. Math. Math. Phys. 36 (1996), no. 8, 1161–1167 (1997).
4. Dyukova E.V., Zhuravlev Yu. I., Sotnezov M.R. Construction of an Ensemble of Logical Correctors on the Basis of Elementary Classifiers // Pattern Recognition and Image Analysis, 2011, Vol. 21, No. 4, pp. 599–605.
5. Genrikhov I. E. Synthesis and analysis of recognizing procedures on the basis of full decision trees // Pattern Recognition and Image Analysis. – 2011. – Vol. 21, no. 1. – Pp. 45-51.
6. Sotnezov R.M. Genetic Algorithms for Problems of Logical Data Analysis in Discrete Optimization and Image Recognition // Pattern Recognition and Image Analysis, 2009, Vol. 19, No. 3, pp. 469–477
7. Zhuravlev Yu.I. An Algebraic Approach to Recognition or Classifications Problems // Pattern Recognition and Image Analysis. 1998. Vol. 8, № 1. P. 59–100.
8. Zhuravlev Yu.I., Ryazanov V.V., and Sen'ko O.V., "Recognition". Mathematical Methods. Program System. Practical Applications (Fazis, Moscow, 2006) [in Russian].

ON THE FORMALIZATION OF SYNTHESIS TASK IN PATTERN RECOGNITION[1]

A. Dokukin[2]

[2] CC RAS, 40 Vavilova Str. 119333 Moscow, Russian Federation, dalex@ccas.ru

The paper is devoted to a possible formalization of the synthesis task which arises in connection to protein synthesis. It contains problem statement for some important cases, description of proposed solving methods and their theoretical substantiation.

Introduction

The task which is considered in the present article arises in connection to a general theme of definite structure protein synthesis. Some results in that field have been already achieved [2] in assumption that structure forming properties of proteins are bound to polypeptide subsequences of fixed length. The researchers were able to distinguish structure forming 5-element sequences from the rest by pattern recognition methods with 90% quality. If we keep the assumption and work with it further we come to some mathematical problems that we believe to be interesting themselves. Thus, we will focus on mathematical aspect of it and avoid stating any biophysical conclusions.

Let we have a polypeptide sequence with some gaps that we want to fill according to the desired profile of properties. The properties themselves are given as a set of precedents, i.e. examples of their presence. Furthermore, the properties are bound to subsequences of constant length that is much lesser than that of the given template. The next section will be devoted to formalization of this task.

Elementary synthesis task

Two major cases of synthesis task can be distinguished. They are determined by the length of the target template that can either equal or exceed precedents length. We will call the first one elementary and the second general. In this article we consider binary synthesis task in which target property is represented by some binary quality.

In general the binary elementary task can be described as following. Let we have a sample of objects which do or don't have some quality. Let we also have some set of restrictions. The problem is finding an object that satisfies the restrictions and has the quality. It is assumed that the initial sample can be strongly incomplete and inconsistent that eliminates simple enumeration.

In addition it is considered desirable to find an object that has the target quality "in uppermost degree" for reliability of the solution.

Let's write down the above-said more formally.

Definition 1. A set $P = \{P_1, \ldots, P_k\}$ will be called the alphabet and its elements will be referred to as letters.

Definition 2. By a word of length t a vector $\beta = (\beta_1, ..., \beta_t)$ will be considered, such that $\beta_i \in P$.

Definition 3. A set $B_1 \times \ldots \times B_t$, where $B_i \subset P$ and either $|B_i| = 1$ or $B_i = P$ will be called a template of length t.

The meaning of a template is quite simple. It is a set of words restricted to a single specific letter in some positions and unrestricted in others.

Let us have a training sample $\{\beta_1, ..., \beta_t\}$ of words of length n. Let this sample be split into two classes K_1 and K_0, that is words having and not having some μ-quality correspondingly. Finally, let us have a template T of length n.

Definition 4. By elementary synthesis task we will call a search for the word $\beta' \in T$, such that

$$\beta' = \arg\max_{\beta \in T} d(\beta),$$

[1] The work is supported by RFBR, grant No. 12-01-31166.

where $d(\beta)$ is an estimate of presence of μ-quality in a word β.

The key quality of the task is that the specific of the application domain doesn't allow defining the functional $d(\beta)$. That is why it is proposed using the training sample for its calculation. Hereat it seems natural using recognition algorithms for the purpose.

Definition 5. The functional $d(\beta)$ is defined as $d(\beta)=\Gamma_1(\beta)-\Gamma_0(\beta)$, where $\Gamma_1(\beta)$ and $\Gamma_0(\beta)$ are estimates of the object β for classes K_1 and K_0 by some recognition operator correspondingly.

As recognition method became part of synthesis task definition it is important to give some information about it. A special template based variation of logical regularities method [3] has been proposed for the task mostly because of its ability to handle great amounts of data fast. The regularities will be sought in form of templates (see Definition 3). Their significance will be determined by Fisher's test [1] using the approximate formula [4].

Definition 6. Let's consider a template T. We will call it logical regularity of class K_1 with significance $P(T)$ if

$$P(T)=-\ln\left(\frac{C_X^x C_Y^y}{C_{X+Y}^{x+y}}\right)>0,$$

where $X=\left|\{\beta_1,\ldots,\beta_m\}\bigcap K_1\right|$, $x=\left|T\bigcap K_1\right|$, $Y=\left|\{\beta_1,\ldots,\beta_m\}\bigcap K_0\right|$, $y=\left|T\bigcap K_0\right|$. Class K_0 regularities are defined the same way.

Let R_1 and R_0 be regularities of classes K_1 and K_0 correspondingly.

Definition 7. We will call an estimate $\Gamma_s(\beta)$ of the word β for the class K_s the value

$$\Gamma_s(\beta)=\sum_{U\in\{T_k\in R_s\mid\beta\in T;T_i\not\subset T_j,i\neq j\}}P(U),$$

As an estimate for the class a word gets sum of significances of all satisfied regularities of that class minus sum of significances of satisfied regularities of the other one. Hereat if the word satisfies two regularities one of which contains the other the latter's significance is not taken into account.

Template sizes in elementary synthesis task allow searching the solution by complete enumeration. But having in mind possible generalizations we must take into account its extreme growth. That is why we consider important achieving an algorithm less dependent of the template size even for elementary synthesis.

Let's consider a template $T=B_1\times\ldots\times B_n$. Let $B_i=P$, i.e. all letters are allowed in i-th position.

Definition 8. Let's denote by $T[i]\gamma$ the refining of template T in the position i,

$$T[i]\gamma=\{\beta\in T\mid\beta_i=\gamma\},$$

Where γ is some letter from the alphabet, $\gamma\in P$.

We propose the following synthesis algorithm. For each position i of the template in which a letter is not constricted we determine it by the formula

$$\gamma_i'=\arg\max_{\gamma\in P}\left(\sum_{U\in\{B_k\in R_1\mid T[i]\gamma\in B_k\}}P(U)-\sum_{V\in\{B_k\in R_0\mid T[i]\gamma\in B_k\}}P(V)\right).$$

It means that each letter will be independently refined by the value that achieves the highest estimate as refinement of the initial template. The estimate of the refined template is calculated as difference of significances of templates containing it.

Theorem 1. Let the functional d be determined by Definitions 5 and 7. If the proposed algorithm finds a unique solution for each unrestricted letter of template T then it calculates the optimal solution of the elementary synthesis task (see Definition 4).

Proof. Let's prove the theorem ad absurdum. The following is the definition of the optimal solution of the synthesis task β':

$$\beta'=\arg\max_{\beta\in T}\left(\sum_{U\in\{B_k\in R_1\mid\beta\in B_k,B_k\subset B\}}P(U)-\sum_{V\in\{B_k\in R_0\mid\beta\in B_k,B_k\subset B\}}P(V)\right)$$

Let there be a position h in which $\gamma_h'\neq\beta_h'$, where γ_h' is achieved by the synthesis algorithm. Let's consider a set of regularities generating estimates of the optimal solution for the class K_s

$$R_s'=\{B_k\in R_s\mid\beta'\in B_k;B_i\not\subset B_j,i\neq j\}.$$

The set consists of two parts

$$R_s'=\{U\mid U_h=\beta_h'\}\bigcup\{U\mid U_h=P\}.$$

Hereat the second item takes part in estimation regardless the letter in position h. The first one consists of the unique element U such

that $T[h]\beta'_h \subset U$. Indeed one such regularity must exist by the theorem condition because the algorithm comes to a solution. Refinements of the regularity in turn do not take part in estimating.

Since the regularity U must be used in search of γ'_h by the synthesis algorithm definition, we come to a contradiction.

The theorem is proved.

General synthesis task

The goal of this section is to generalize the results of the latter one for the case of longer template.

Let's consider a template T of length N, $N > n$.

First of all, since only a precedent length object can have the quality the target property transforms to a property profile.

Definition 9. Let's call subtemplate at position i of the template T a template $T(i) = \{B_i, \ldots, B_{i+n-1}\}$ of length n.

In a same fashion a subword $\beta(i) = \{\beta_i, \ldots, \beta_{i+N-1}\}$ of word $W = \{\beta_1, \ldots, \beta_N\}$ can be defined that we will call element.

Definition 10. Let's call a property profile for the template T a set $Q = \{q_1, \ldots, q_{N-n+1}\}$, $q_i \in \{K_1, K_0\}$, where q_i designates class for i-th element of target word.

Second, each missing letter can belong to more than one element, all of which should be taken into account during optimization.

Definition 11. By general synthesis task we will call a search for a word $\beta' \in T$, such that

$$\beta' = \arg\max_{\beta \in T} D(\beta),$$

where $D(\beta)$ is an estimate of quality profile Q in a word β

$$D(\beta) = \min_{i=1,\ldots,N-n+1} d_{q_i}(\beta(i)),$$

where $d_{K_1}(\beta) = d(\beta)$, $d_{K_0}(\beta) = -d(\beta)$.

(There is certain freedom in choosing the functional as for example

$$D(\beta) = \sum_{i=1}^{N-n+1} d_{q_i}(\beta(i))$$

can be used. Later in this article we will use the first one, knowing that all the results hold for this one too. The final choice should be made after experiment series.)

Again, an algorithm will work independently with each missing letter.

$$\gamma'_i = \arg\max_{\gamma \in P} \left(\min_{j=1,\ldots,N-n+1} \left(\sum_{U \in \{B_k \in R_1 | T[i]\gamma(j) \in B_k\}} P(U) - \right. \right.$$

$$\left. \left. - \sum_{V \in \{B_k \in R_0 | T[i]\gamma(j) \in B_k\}} P(V) \right) \right).$$

Theorem 2. If the proposed algorithm finds a unique solution for each unrestricted letter of template T then it calculates the optimal solution of the general synthesis task.

Proof in most part repeats the proof of the theorem 1. Instead of a single regularity U we get a set $\{U_1, \ldots, U_n\}$ of regularities corresponding to each element containing the target letter. Nevertheless, each of them has the same properties as U by design.

Conclusion

In the present paper a synthesis task has been described, i.e. the task of searching for an object of definite quality that is described by a set of precedents.

An approach for solving the task has been proposed for an important task family.

At the same time the question of formalizing the method quality estimate remains and consequent testing with real world data are the most obvious ways for development of these results.

References

1. Fisher R. A. On the Interpretation of X2 from Contingency Tables, and the Calculation of P // Journal of the Royal Statistical Society. — 1922. — Vol. 85, No. 1. — Pp. 87–94.
2. Senko O.V., Nekrasov A.N., Ryazanov V.V., Dokukin A.A. Prediction of Structure Forming Properties of Protein with Help of Pattern Recognition Methods // Proceedings of 8-th Open German-Russian Workshop "Pattern Recognition and Image Understanding" OGRW-8-2011. — Pp. 38–40.
3. Ryazanov V.V. Logical regularities in pattern recognition (parametric approach) // Computational Mathematics and Mathematical Physics. — 2007. — Vol. 47, No. 10. — Pp. 1720–1735.
4. Vorontsov K.V. Lections on classification algorithms (in Russian). — 2007. http://www.ccas.ru/voron/download/LogicAlgs.pdf.

LOCALIZATION OF THE NEGATIVE EIGENVALUES FOR A NON-POSITIVELY DEFINITE MATRIX OF PAIRWISE COMPARISONS[1]

S.D. Dvoenko[2,3], D.O. Pshenichny[2,4]

[2] Tula State University, Department of Automation and Remote Control, 300012, Tula, Lenin Ave., 92, (4872)35-36-37,
[3] dsd@tsu.tula.ru, [4] denispshenichny@yandex.ru

In data mining and machine learning, the experimental data are often presented as results of mutual pairwise comparisons of objects. Because of absence of the original feature space, the condition for correct immersion of a set of objects in a metric space is a non-negative definiteness of a matrix of pairwise comparisons of its elements. In this case, pairwise comparisons can be interpreted as the scalar products. This paper proposes an approach to correction of the violated metric by minimal correction of the given matrix of pairwise comparisons using its rows and columns permutations and some values modifications.

Introduction

Let experimental data be presented in the form of pair-wise comparisons between elements of the limited set.

In this case, the experimental results can be presented as a square matrix. If this matrix is positively definite, this means it has positive determinant and positive set of eigenvalues.

The latter means that many systems of orthogonal coordinates with common origin exist and can be transform one in another by rotation of its coordinate axes.

The set elements as vectors are presented in such coordinate spaces and similarities or dissimilarities between them can be evaluated as scalar products or distances, usually Euclidean ones.

In this case based on cosine theorem distances are evaluated for known scalar products and vice versa.

Therefore, if the matrix of pairwise comparisons is positively definite, then we consider the set elements are immersed in unknown (for us) metric space, for example Euclidean, with dimensionality not more than the rank of the matrix.

Therefore, after pairwise similarities have been measured usual transformations are produced to get symmetric and normalized matrix with the main diagonal of units.

Specifically, a matrix of weighed scalar products (correlations) represents experimental results in factor and correlation analysis when elements of the set are features.

Because of such a usual way to get data, the matrix of pairwise comparisons can be non-positively definite and can have negative eigenvalues.

It means that in general case all transformations need to allow for immersion of the specified set of elements in the appropriate metric space.

It is evident this is a problem to be solved before the problem of the data analysis itself, and more importantly, before the measurement process.

This is the modern problem of so-called multi-kernel similarity measures in data analysis [8] and multi-modal image registration in medical applications [10].

Let the non-positively definite matrix of pairwise similarities be given by usual way.

It can be made positively definite by decomposition of it relative its eigenvalues and removing from elements of the matrix contributions of the negative eigenvalues [7].

In general, such a transformation is achieved by the space dimensionality reducing based on discrete Karhunen-Loève expansion [2, 9] to remove, for example, contributions of about 20% of lowest eigenvalues including negative ones.

[1] Supported by RFBR grant № 13-07-00010

It needs to note, first, we cannot control the scale of corrections in data and, second, we actually reduce initial information.

It may be natural for features in Karhunen-Loève problem, but it is not so for the set elements as objects for learning to recognize them.

Indeed, we really need to recognize all elements in the given set and do it mathematically correct by specially developed recognizing and clustering algorithms [3].

Therefore, for the non-positively definite matrix of pairwise similarities we need to correct it minimally for recovering the metric and get positively definite matrix.

Mathematical framework

Let the matrix $S(n,n)$ given. It has diagonal of units, its elements are symmetric with respect to main diagonal, and their absolute value is less than 1.

So S is the matrix of some quadratic form.

To see whether it is positively definite we can use Sylvester's criterion: the matrix of a quadratic form is positively definite if and only if its principal minors are positive [5].

According to the consequence of Sylvester's inertia law, the number of negative eigenvalues ν in S equals to the number of sign changes in the sequence

$S_0 = 1,\ S_1, \ldots,\ S_n$.

Here the notation S_k means k^{th} principal minor of S [5].

One can prove that simultaneous permutation of two rows and two corresponding columns in S does not change its eigenvalues [5].

This transposition means transposition of the corresponding two elements of the set.

As it was shown in [4], the values of the principal minors of S are descending from 1 and once some minor S_k can become negative.

Using the method of matrix correction described in [4], most of all minors from S_k to S_n should be corrected.

Therefore, it is better to have the first negative minor as far as possible.

Inertia law says that k can not be more than $n-\nu$, so we want to have k as close as possible to $n-\nu$.

Method of solution

The idea of approaching k to $n-\nu$ is in following.

We want to get such the ranging of the set elements so every following set element causes minimal descending of the principal minor value.

This ranging can be built like this. Let us have n set elements and the corresponding matrix $S(n,n)$ of pairwise comparisons with ν eigenvalues.

Then we remove such set element so derivable matrix $S(n-1,n-1)$ has $\nu-1$ negative eigenvalues and its determinant has the maximal absolute value.

We assign n^{th} position in our ranging to this set element.

Because of changing of evenness of the negative eigenvalues determinant changes its sign.

If it is not possible to find such element that brings changing of the determinant sign then we remove the element that brings the maximal absolute value of the determinant, and this element obtains $n^{\text{th}`}$ position in the ranging.

Then we repeat this procedure for the matrix $S(n-1,n-1)$.

Each iteration decrements dimension of the matrix S and increments the number of set elements that obtained their position in the ranging.

So when we get the matrix $S(1,1)$ all the set elements will be ranged.

If we choose the set elements in some other order, the determinant of the corresponding matrix of pairwise comparisons will not be changed.

Therefore, our ranging is invariant under the order of the set elements.

Experiments

Let us test this method for two given matrices.

The first one describes the influence between noises and vibrations in a tractor cab and tractor driver's work-rate and state of health by correlation of 33 parameters [6].

The second one describes data that Dr. San-Ho Kim from the Lawrence National laboratory, Berkley, USA, obtained in his research in the molecular biology.

He compared the set of nuclear sequences by Fasta program and selected 420 ones by minimal similarity to get the pair-wised similarity matrix of the size 420x420 [1].

There were two couples of repeating values, so we removed them and we obtained matrix 418 by 418.

First matrix $S(33, 33)$ has two negative eigenvalues.

Scanning through its principal minors, we can see that the first negative minor is 22nd one.

Correcting this matrix with the correction method, we get the following result: there were 12 set elements whose pairwise comparisons should be corrected.

Total deviation of the corrected matrix from the initial one is 47.2026388245.

Now let us firstly find the optimal permutation.

After we have built it, let us scan through the principal minors of the corresponding matrix.

We can see that the first negative principal minor is 32nd.

Due to the fact of having two negative eigenvalues, we can conclude that this permutation is ideal and we need to correct the pairwise comparisons of also two set elements.

Correcting it we obtain the matrix whose elements are differs by only 5.3354921875 from the elements of the initial matrix.

Now let us correct the second matrix. Its size is 418 by 418 and it has 5 negative eigenvalues.

Correcting this matrix without searching the optimal ranging we have the following results.

First negative principal minor is 375th and we have to correct 43 rows and columns of the initial matrix.

The corrected matrix deviates by 3109.3063326680 from the initial one.

Using correction with the search of the best permutation, we can see that it gives also better results. The first negative principal minor is 411st.

Trice we could not find the element that changes the sign of the determinant so we have to correct 8 rows and columns.

Total deviation after correction is 576.1138339844.

Conclusion

Based on the available method of correction of a non-positively definite matrix of pairwise comparisons, we offer an improvement of it.

It consists in the initial set of elements rearranging which gives the optimal ranging of it in the sense of minimal violation of metrics.

This ranging can be used in correction of the non-positively definite matrix of pairwise comparisons.

References

1. Dubchak I., Muchnik I., Mayor C., Dralyuk I., Kim S.-H. Recognition of a protein fold in the context of the SCOP classification // Proteins: structure, function, and genetics – 1999. - №35. – p. 401-407.
2. R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification. Wiley, N.Y., 2001.
3. S.D. Dvoenko, Clustering and Separating of a Set of Members in Terms of Mutual Distances and Similarities. Transactions on Machine Learning and Data Mining. IBaI Publishing, 2009. V. 2. No. 2. P. 80-99.
4. Dvoenko S.D., Pshenichny D.O., On negative eigenvalues removing from matrixes of pairwise comparisons. Intellectualization of information analysis: 9th international conference. Montenegro, Budva, 2012.: Proccedings. – M.: Torus-Press, 2012, pp. 13-16. (in Russian)
5. Gantmacher F. R. The Theory of Matrices. — AMS Chelsea Publishing: Reprinted by American Mathematical Society, 2000. — 660 p.
6. V.Ya. Lumelskii, Parameter Grouping on the Basis of the Square Coupling Matrix. Automation and Remote Control. Pleiades Publishing, Ltd., 1970, V.1. P. 133-143. (in Russian)
7. E. Pekalska, R.P.W. Duin, The Dissimilarity Representation for Pattern Recognition. Foundations and Applications. World Scientific: Singapore, 2005.
8. B. Schölkopf, A.J. Smola, Learning with Kernels. MIT Press, Cambridge, 2002.
9. J.T. Tou, R.C. Gonzalez, Pattern Recognition Principles, London-Amsterdam-Dom Mills, Ontario-Sydney-Tokyo. Addison-Wesley Publishing Company, 1974.
10. L. Zollei, J. Fisher, and W. Wells, A unified statistical and information theoretic framework for multi-modal image registration. Proc. of IPMI Conf., 2003. P. 366–377.

RESEARCHING THE STABILITY OF SOUND SIGNAL DESCRIPTION

V.E. Gai[1]

[1]R.Y. Alekseev Nizhny Novgorod State Technical University,
603950, N. Novgorod, K. Minina street, 24, Russia, iamuser@inbox.ru

The paper describes the method of allocation of sound signal sections the most resistible to distortions. Preliminary signal description is created by means of U-transformation. Results of computing experiments are adduced.

Introduction

The decision of any problem of digital processing of a sound signal starts with a signal description creation. There are various methods of creation of a sound signal description. They are Fourier transforms, wavelet analysis, *U*-transformation. The question of stability of a sound signal description arises while solving a signal processing problem at noise conditions. One of such tasks is a task of media data analysis which consists in search of a set sound fragment in the sound stream or in the file.

The present work is devoted to researches of stability to description distortion created by means of *U*-transformation [1, 2]. U-transformation is the basic transformation used in the theory of active perception (TAP). Within TAP the sound signal is considered as system unit. For detection of system elements integral transformation is used, and for identification of communications between elements – spatial derivation. Spectral signal description is a result of differential structure identification.

Suppose $f(t)$ – an analyzed sound signal observed on a finite amount of time. The result of U-transformation application to f signal is multilevel (coarse-fine) spectral representation of $D = \{d_{ij}\}$, $i = \overline{1,K}$, $j \in \overline{1,M_i}$, where K is a number of decomposition levels, M_i – number of signal segments on i decomposition level, d_{ij} – spectrum including L spectral coefficients (number of filters used), $d_{ij}\{k\}$ – k spectral coefficient ($k = \overline{1,L}$), f_{ij} – a waveform segment of f on which d_{ij} range is calculated. *U*-transformation is realized by means of a set of filters $\{F_i\}$. Below is a basic set of filters:

$$\begin{aligned} F_0 &= \{1, 1, 1, 1\}, \\ F_1 &= \{-1, -1, 1, 1\}, \\ F_2 &= \{-1, 1, 1, -1\}, \\ F_3 &= \{1, -1, 1, -1\}. \end{aligned} \qquad (1)$$

This set of filters is equivalent to the Walsh system filters Hartmuth [3].

Let us assume for filter sets $\{F_i\}$ that ($+1 \rightarrow 1$) and ($-1 \rightarrow 0$). We obtain binary operators sets $\{V_i\}$. Set-theory operations of integration (addition) and intersection (multiplication) are admissible for these operators. Result we have in algebra – $A_V = <\{V_i\}: +, \times>$ [1].

In A_V algebra there are algebraic groups:

1) P_{ni} – groups on three elements (are called as complete), are formed on three operators (V_i,V_j,V_k) which have correlation as follows: $V_i + V_j + V_k \equiv e_1$ – unit. The complete group admits two descriptions – for intersection operation of $V_iV_jV_k$ (multiplication operation, number of inversions is even) and integration operations (addition operations, number of inversions is odd). The number of possible patterns of complete groups, taking into account inversions of operators, is equal to four. Complete group pattern is a compact of four connected elements.

2) P_{si} – groups on four elements (are called as closed), are formed on four operators (V_i,V_j,V_n,V_m), where $(V_i,V_j,V_k) \in P_{ni}$, $(V_n,V_m,V_k) \in P_{nj}$, with $V_iV_j + V_n\overline{V}_m$ description (where desired number of operators inversions is odd) and unit – $V_i + V_j + V_n + \overline{V}_m \equiv e_{1\ \text{pattern}}$. The number of possible closed group patterns, taking into account operator inversions is equal to eight.

Complete group pattern is a compact of eight connected elements.

Sets $\{P_{ni}\}$, $\{P_{si}\}$ are finite and have capacity of 35 and 105 respectively (these sets are formed on 16 operators). Sets $\{V_i\}$, $\{P_{ni}\}$, $\{P_{si}\}$ – the sets of standards used to solve a problem of recognition of the object of research in standard space. Notation of complete group in addition operation is P_{nia}, in multiplication operation is P_{nim}.

Formation of sound signal description within U-transformation is based on use of algebra of groups. Spectral and correlation analysis is carried out by means of closed and complete groups. Complete groups make possible to explore correlation communications between operators. Closed groups make possible to explore correlation communications between complete groups. Drawing an analogy between algebra of groups and language, it is possible to apply the following compliances: operator is a language alphabet, complete group is a word and closed group is a phrase.

Closed group formed on the basis of two full groups, and closed group – on the basis of the three operators.

In [1] it is noted that the greatest resistance to distortions belongs to descriptions created on the basis of closed groups, the smallest one - on the basis of operators.

Signal research algorithm

Let us define indicators which can be used for allocation of stable segments of a signal A. For this purpose we will make analysis of a signal A:

1) single-level U-transformation of a signal A is calculated, the signal is divided into M segments. For each signal segment the set of complete (P_{ni}) and closed (P_{si}) groups is calculated. The signal description represents set of values of operators, and also the groups calculated on each signal segment:

$$D_A = < \bigcup_{i=\overline{1,M}} V_i, P_{ni,i}, P_{si,i} > \qquad (2)$$

2) under signal A by means of operator of distortion B and C signals are formed (distortion level of a signal A has to be equal, however, considering that signal A is imposed with random noise, signals B and C won't be equivalent):

$$\begin{aligned} B &= \text{NOISE}[A], \\ C &= \text{NOISE}[A], \end{aligned} \qquad (3)$$

where NOISE[•] – an operator of imposing noise to a signal;

3) descriptions of signals B and C are calculated: D_B, D_C. The structure of descriptions of D_B and D_C is equivalent to D_A structure;

4) comparison of descriptions is carried out: D_A-D_B and D_A-D_C; it results in arrays generating $C_{B,V}$, $C_{B,ni}$, $C_{B,pi}$, $C_{C,V}$, $C_{C,ni}$, $C_{C,pi}$ ($C_{B,V}[i]$ – number of operators which coincided in signs for i segment of descriptions D_A-D_B; $C_{B,ni}[i]$ – number of complete groups which coincided for i segment of descriptions D_A-D_B; $C_{B,pi}[i]$ – number of closed groups which coincided for i segment of descriptions D_A-D_B).

By results of formation of the above mentioned arrays it is possible to draw a conclusion about stability of segments in accordance with the following rule: If $C_B[i] > T$ and $C_C[i] > T$, then i segment of a signal is distortion-resistant (T – threshold value).

Let us use this method to define characteristics of distortion-resistant signal segments.

After a number of experiments it was established that for allocation of stable signal segments only closed groups can be used. Analyzing the descriptions of distorted and undistorted signals generated by means of closed groups, it was managed to allocate the indicators making possible to divide signal segments into stable and unstable:

1) «P1» indicator: the more difference between maximum and minimum mass of groups in a set, describing a signal segment S, the much more likely that segment description of a signal won't change while distorting:

$$\begin{aligned} D_1 = \text{MAX}(\text{MASS}(P, A)) - \\ - \text{MIN}(\text{MASS}(P, A)), \end{aligned} \qquad (4)$$

where P is a set of groups, calculated by means of a signal A, MAX [•] – an operator calculating the maximum value among numbers in the array, MIN [•] – an operator calculating the minimum value among

numbers in the array, MASS [•] – the operator calculating masses of groups including in the array *P*;

2) «P2» indicator: the more sum of group masses including into a set of groups *P*, describing a signal segment *A*, the much more likely that segment description of a signal won't change while distorting:

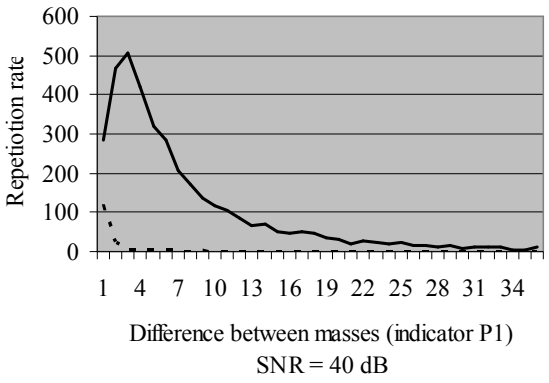
$$D_2 = \mathrm{SUM}(\mathrm{MASS}(P, A)), \qquad (5)$$

where SUM[•] – operator of calculation of the sum of the array elements.

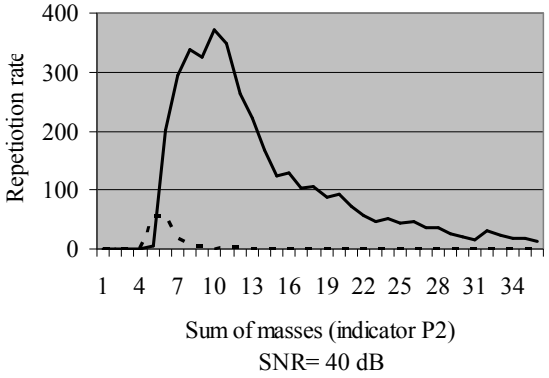
Computing experiment

Let us consider the pictures which prove possibility to use the mentioned herein indicators for dividing signal segments into stable and unstable. In the fig.a and fig.b it is given the statistics of existence in a signal stable and unstable segments at low noise level (relation signal/noise is equal to 40 dB). This statistics is calculated by means of C_B and C_C arrays. When forming C_B and C_C arrays the first *N* by weight of closed groups were taken out. In fig.a the full line represents the diagram of frequency change of stable segments appearance depending on a difference between the mass of groups calculated on *i* signal segment, the dash line defines the same diagram but only for unstable segments. In fig.b the full line represents the diagram of frequency change of stable segments appearance depending on the sum of mass of groups calculated on *i* signal segment, the dash line defines the same diagram but only for unstable segments. The range of values of difference of masses and also the sums of masses are standardized to a segment [0; 100].

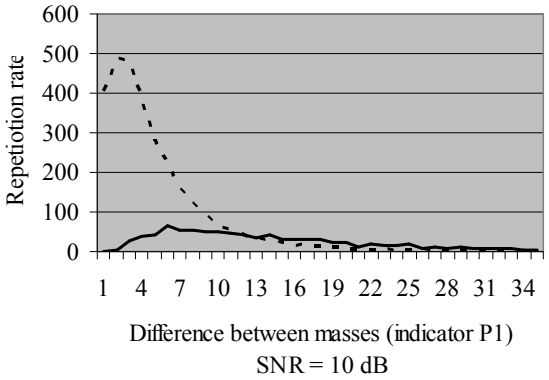
The conducted experiments showed that the peak position in the figure (it is marked with a triangle) can be used as a threshold for dividing distortion-resistant and distortion-labile segments. In the fig.c and fig.d the same diagrams are depicted but only for the signal/noise ratio equal to 10 dB. Comparing fig.a and fig.c we can note that increase of noise level causes the decrease of quantity of stable segments meanwhile peak position is practically not changed (it is slightly moved to the left).



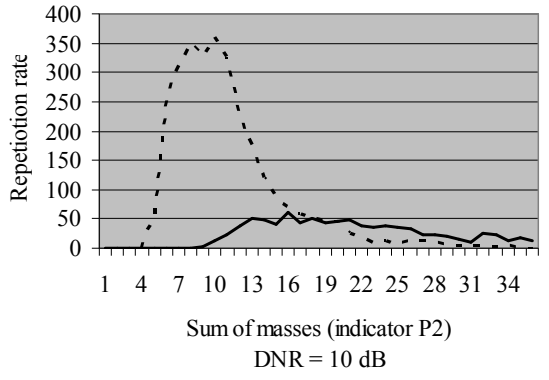
a.



b.



c.



d.

Fig.1. Threshold identification

Table 1. Statistics of signal segment stability (uniform noise)

Noise 40 dB	Noise 20 dB	Noise 10 dB	Noise 0 dB
Segment length: 64			
9640/ 3610	7851/ 5399	6385/ 6865	1570/ 11680
9212/ 4038	6606/ 6644	4747/ 8503	613/ 12637
Segment length: 128			
2588/ 1404	2543/ 1449	1728/ 2264	728/ 3264
2305/ 1687	2355 / 1637	1316/ 2676	653/ 3339
Segment length: 256			
1363/ 485	1363/ 485	1042/ 806	497/ 1350
1255/ 593	1255/ 593	843/ 1005	566/ 1281
Segment length: 512			
959/ 401	959/ 401	820/ 540	251/ 1109
908/ 452	908/ 452	908/ 452	133/ 1227
Segment length: 1024			
462/ 245	364/ 343	364/ 343	195/ 512
425/ 282	425/ 282	547/ 160	85/ 622

Table 2. Statistics of signal segment stability (normal noise)

Noise 40 dB	Noise 20 dB	Noise 10 dB	Noise 0 dB
Segment length: 64			
9640/ 3610	6385/ 6865	2880/ 10370	107/ 13143
9212/ 4038	5584/ 7666	1849/ 11401	179/ 13071
Segment length: 128			
2588/ 1404	2161/ 1831	916/ 3076	536/ 3456
2305/ 1687	1954/ 2038	1042/ 2950	53/ 3939
Segment length: 256			
1363/ 485	1042/ 806	622/ 1226	171/ 1677
1255/ 593	843/ 1005	566/ 1282	57/ 1791
Segment length: 512			
959/ 401	685/ 675	470/ 890	251/ 1109
908/ 452	908/ 452	537/ 823	133/ 1227
Segment length: 1024			
462/ 245	364/ 343	234/ 473	37/ 670
425/ 282	323/ 384	170/ 537	11/ 696

This fact can be used to divide stable and unstable segments. It should be noted that the structure of diagrams for different noise levels is similar.

At table 1 and table 2 presented statistics about steady / non steady signal's segments in different conditions (for different noise levels and for different segments length). At first string presented data for first indicator (P1) and in second string – for second (P2).

Conclusion

This paper presents the method making possible to calculate the sections of a sound signal resistant to distortions. For allocation of such sections two indicators are introduced. The offered method can find practical application to complete tasks of analysis of media data that will make possible to increase the resistance of the description created to different distortions.

References

1. V. A. Utrobin Physical interpretations of image algebra elements // Success of physical sciences, T. 174, № 10, 2004, P. 1089–1104.
2. Utrobyn V.A. Computer processing of images. Decision-making area of standards. N.Novgorod: NNTU, 2004. 221 p. [in Russian]
3. Henning F. Harmuth Application of the methods of information theory in physics. M.: Mir, 1975. 344 P. [in Russian]

COMPLEXITY DECOMPOSITIONS IN PROBLEMS OF COMPARISON OF SYMBOLIC SEQUENCES[1]

V.D. Gusev[2], L.A. Miroshnichenko[2]

[2] Institute of Mathematics, Siberian Branch, Russian Academy of Sciences, pr. Koptyuga 4, Novosibirsk, 630090 Russia, tel. +7-3833634671, luba@math.nsc.ru

We consider basic principles of comparison of symbolic sequences (texts of any language nature). We give examples of definitions of similarity measures and distances between texts. Special attention is paid to the relative complexity of texts: a measure generalizing the Lempel-Ziv complexity approach to the case of pairs (groups) of texts with addition of copying operations that are characteristic for a concrete subject area. We compare the relative complexity with other known measures and show examples of its use.

Introduction

Symbolic sequences (words, strings, texts) as an object of study are encountered in various areas of knowledge: mathematics, informatics, biology, linguistics, music. Many problems related to analysis of texts are of classification nature. A key point in solution of such problems is the choice of similarity measures between texts. This choice is based on defining a set of admissible transformations (operations) characterizing the variability of a given class of objects. For example, at the level of words of a natural language, such operations are substitutions, insertions, omissions, permutations of individual symbols, which give rise to an error or a change of the meaning of the word. The same operations characterize evolution of DNA sequences at the lower level (point mutations). At the higher (chromosome) level, the basic evolution events are large-block rearrangements: inversions, transpositions, etc. An inversion implies breaking a sequence at two points, cutting out the fragment between them, and inserting this fragment in the same place but in the reverse order. A transposition is moving a fragment from one place on the chromosome to another. Specific operations for music compositions are sequential transitions and filling an interval.

Results of comparison of sequences for a concrete subject area essentially depend on the adequacy of the choice of admissible operations, their weights, and similarity measures. A fairly universal idea for calculation of the similarity of two sequences is estimating the complexity of transforming one of them into the other by using a fixed set of admissible operations. A transformation is considered to be optimal if it is performed in the minimal number of steps (one step corresponds to a single use of any operation in the set). This indicator gives a quantitative estimate for the similarity of sequences. The purpose of this paper is comparative analysis of approaches implementing the strategy described above, with a more detailed discussion of one of them that generalize the notion of complexity decomposition of a symbolic sequence [1] to the case of two (or more) texts. We present examples of applications of this strategy to real problems.

1. Distances and similarity measures

Let Σ be a finite alphabet; S a finite sequence composed of elements of Σ (text); $N = |S|$ the length of a text S; let a^m be the m-fold repeat of a symbol a. Let $S = S_1S_2$ be the concatenation of sequences S_1 and S_2.

1.1. The edit distance between texts S_1 and S_2 in the simplest case is defined as the minimum number of "edit" operations (substitution, insertion, and deletion of a symbol) that are necessary for transforming S_1 into S_2. This metric was introduced by Levenstein in connection with problems arising in coding theory. Various modifications of it were used in computational linguistics (edit distance),

[1] Support of the grant RFBR No. 13-07-00400

molecular biology (evolutionary distance), in problems of speech recognition.

The process of transforming S_1 into S_2 is usually represented as "alignment". Alignment is a matrix of two rows S_1' and S_2' obtained from S_1 and S_2 by adding the symbol "–". In this matrix, the identical (and being replaced) elements are situated one under another, and the columns containing the symbol "–" indicate symbols that are being deleted or inserted.

If the operations have "weights", then $d(S_1, S_2)$ is defined as the minimal "cost" of transforming S_1 into S_2. The edit distance is calculated by using a scheme of dynamic programming with the running time being $O(N_1 \times N_2)$. The scheme can be easily modified for detecting local similarity.

Construction of alignment plays an important role in bioinformatics. For reducing running time when comparing long texts, first "synchronization points" are sought in the form of long repeats, which are then extended by using dynamic programming. Here, the minimum of the "cost" of transforming S_1 into S_2 is not guaranteed.

Weights are assigned to the edit operations by using accumulated statistics of occurrence of various kinds of substitutions, insertions, deletions in the alignments.

1.2. The Kolmogorov approach. Kolmogorov defined the complexity of an object (in this case, of a symbolic sequence) as the length of the shortest description $K(S)$ by which the original object can be uniquely reconstructed. When there are two objects (texts S_1 and S_2), we can speak about the quantity of information about S_1 contained in S_2. This quantity can be regarded as the relative complexity $K(S_1/S_2)$ of reconstruction of S_1 by S_2. For comparison of texts one can use both the relative complexity itself and a normalized characteristic, for example, of the form $d_K(S_1, S_2) = 1 - \frac{K(S_1) - K(S_1 / S_2)}{K(S_1 S_2)}$ (see [2].) It is well known that the Kolmogorov complexity is not explicitly computable. Therefore in practice, instead of $K(S)$, $K(S_1/S_2)$, their estimates are used obtained by means of some compression algorithm. Our version of obtaining these estimates is described below.

1.3. Generalization of the Lempel-Ziv approach. In [1] Lempel and Ziv defined the complexity of S as the minimum number of steps required for its synthesis. The admissible operations are the operation of "generating a new symbol" and of copying a "ready-made prototype" from the prehistory (that is, the already synthesized part of the text). The ordering of fragments synthesized at each step of the process is called the "complexity decomposition" of S. In [3] we generalized the Lempel-Ziv approach by substantially extending the spectrum of admissible copying operations and adapting them towards analysis of DNA sequences. Namely, along with direct copying, which corresponds to identical repeats in the ordinary sense, we allow for symmetric copying corresponding to inversions, as well as direct and symmetric copying in combination with a predefined re-naming of elements of the alphabet. This copying models the relation of complementarity specific for genetic texts (direct and symmetric complementary repeats are detected). The number of steps $c(S)$ of the process of synthesis of S, using the extended spectrum of operations, can be a good approximation for the Kolmogorov complexity $K(S)$.

In the case of two sequences S_1 and S_2 we represent one of them, for example, S_1, as a concatenation of fragments from S_2. The ordering of these fragments that are inter-text repeats of various types is called the complexity decomposition of S_1 with respect to S_2, and the number of components in it $c(S_1 / S_2)$ - the complexity of reconstruction of S_1 by S_2. This characteristic serves as an approximation for $K(S_1/S_2)$.

Tiling S_1 by fragments of S_2 is conducted from beginning to end (from left to right). At each step we copy the longest (over all ways of copying and positional tie-ins) fragment of S_2 that coincides with a prefix of a yet uncovered part of S_1. The main advantages of the algorithm are the detailed registration of large-block rearrangements and a linear dependence of the running time on $(|S_1| + |S_2|)$.

An important complement to the scheme of comparison of sequences described above is the possibility of making it universal, not oriented towards a concrete language system. This is achieved by replacing the DNA-

oriented copying operations by others that are not fixed beforehand but manifest themselves in a text by anomalously long hidden repeats identical with respect to re-naming of elements of the alphabet.

1.4. The transformation distance was defined in [5] as the minimal length of a description of the process of assembling S_1 in the presence of S_2 by using large-block operations such as duplication, inversion, and insertion of segments. In contrast to what was described above, insertion of a segment is used not only when S_1 contains symbols that do not occur in S_2 but also when a symbol-by-symbol generation of a segment costs less (with respect to the number of bits required for that) than copying. The second difference is in the order of assembling S_1. Calculation of the transformation distance $tr(S_1, S_2)$ reduces to choosing a path with smallest weight in a graph of special type. The running time of construction of the graph and the choice of a path with smallest weight is estimated by the authors of [5] as $O(|S_1|^6)$.

1.5. The inversion distance. The inversion distance $d_I(\pi,\sigma)$ between sequences π and σ is defined as the minimum number of inversions required to transform one of them into the other. Obviously, not every sequence can be transformed into another by using only inversions. Therefore the initial problem is usually reduced to comparison of permutations of the integers. The problem of calculating the inversion distance for them is NP-complete, but in the case of "signed" permutations there exist polynomial solutions [9]. "Signed" permutations were introduced for solving the problem of comparison of the order of genes in genomes. The permutation gives the ordering of the genes along the chromosome and the sign ("+" or "-") indicates the orientation of the genes (or, equivalently, which strand of DNA they are on). An inversion changes not only the order of elements between the break points but also the signs of these elements. For example, if in the signed permutation +1 [+2 –4 –5] +3 +6 we invert the fragment in brackets, then we obtain the permutation +1 +5 +4 –2 +3 +6.

1.6. The breakpoint distance between two permutations is a more general and easily computable measure as compared to the inversion distance. Despite a radical simplification of the computational procedure, the results obtained on real data strongly correlate with the results of inversion analysis. Moreover, one can see prospects of using this measure for taking onto account other chromosome rearrangements -transpositions and translocations.

To determine the break points we add an arbitrary permutation π by the elements $\pi_0 = 0$ and $\pi_{N+1} = N + 1$. When comparing π with the identity permutation $e = 0\ 1\ 2\ \ldots\ N\ N + 1$, we detect a break between the elements π_i and π_{i+1} if $|\pi_{i+1} - \pi_i| \neq 1$ $(0 \leq i \leq N)$. The total number of breaks in π can be treated as a measure of dissimilarity between π and e.

In the case of arbitrary permutations π and σ, a break between elements $\pi_i = a$ and $\pi_{i+1} = b$ (where a and b are arbitrary numbers from 0 to N+1) is detected when the numbers a and b do not occur in σ consecutively, i.e. σ does not contain either the bigram ab or ba.

2. Complexity decompositions in comparison with other approaches

2.1. Relative complexity and edit distance. The complexity decomposition of a sequence reveals its large-block structure in the form of a concatenation of repeats of various types. The edit distance does not refer to the structure at all (and to the type of repeats in particular); therefore many repeats are simply not registered. The complexity decomposition $H(S_1 / S_2)$ of S_1 with respect to S_2 is independent of the order of the arrangement of the blocks in both texts; the edit distance is very critical to this factor. For example, let $S_1 = t^{17} a^5$ and $S_2 = a^5 t^{17}$. Then $H(S_1 / S_2) = t^{17} * a^5$, where * is a separator between the decomposition components, i.e. $c(S_1 / S_2) = 2$, whereas $d(S_1, S_2) = 10$ (one has to delete 5 occurrences of the symbol "a" from S_2, and then to insert the same symbols after t^{17}). One can show that $c(S_1 / S_2) \leq 2d(S_2, S_1) + 1$ in the general case.

As for the running time of calculating $c(S_1 / S_2)$ and $d(S_2, S_1)$, in the first case it is linear $O(|S_1| + |S_1|)$, whereas in the second case we can speak merely about the quasi-linearity of nuclear algorithms achieved at the expense of giving up the exact (but quadratic) algorithm of dynamic programming.

2.2. Relative complexity and transformation distance. The idea of transformation distance (extension of the set of admissible large-block rearrangements) is conceptually close to the approach that we develop (see Section 1.3). However, in our approach, as in [1], the number of operations used is counted, while in [5] – the cost of each operation in bits. In combination with a more complicated scheme of assembly of one sequence by the other, an optimization problem arises the running time of which is estimated by the authors of [5] as $O(N^6)$ and which makes this approach hardly suitable to real applications.

2.3. Inversion distance, breakpoint distance, and complexity distance. All these measures are compared on ordinary (not signed) permutations under the assumption that there exists a way of transforming the original sequence to a numerical permutation. Since one inversion can remove two break points at once, the inversion distance $d_I(\pi,\sigma)$ and the number of break points $r(\pi,\sigma)$ are connected by the relation $r(\pi,\sigma) \le 2d_I(\pi,\sigma)$. In turn, the number of breaks of π with respect to σ is uniquely connected with the value of $c(\pi/\sigma)$ if direct and symmetric copying are used as admissible operations: $r(\pi/\sigma) = c(\pi/\sigma) - 1$ [6]. Thus, for calculation of $r(\pi/\sigma)$ it is sufficient to obtain the complexity decomposition of π with respect to σ (running time $O(|\pi|)$). The role of complexity decompositions in calculation of inversion distances is more modest: they make it possible to reduce the dimensionality of the problem and to move away from the original permutations to "signed" ones.

3. Complexity distance in applications

The approach to comparison of texts described above, that is based on construction of complexity decompositions, was tested in solution of problems of data analysis in various language systems. In particular, we mention the following problems: finding unintentional adoptions in song melodies [7]; construction of phylogenetic trees from sequences of discs characterizing polytene chromosomes of mosquitos *Chironomus* [6]; translation of chants written in neume form into modern note form (reconstruction of neumes' equivalents for the basic structural units of neume chants for V.M. Metallov's collection) [8].

Conclusion

We consider methods for calculation of similarity of symbolic sequences based on estimating the complexity of transformation of one of them into another using a fixed set of admissible operations. The main attention is devoted to generalization of the Lempel-Ziv concept of complexity of a sequence to the case of two (or more) texts and to the choice of admissible operations. The measure implemented adequately reacts to all possible manifestations of repetitivity in texts and is oriented towards comparison of long texts. The examples of its application in various language-like systems is given.

References

1. Lempel, A. and Ziv, J. On the Complexity of Finite Sequences, IEEE Trans. Inform. Theory, 1976, vol. IT-22, no. 1, pp. 75–81.
2. M. Li, X. Chen, X. Li, B. Ma, P.M.B. Vitányi: The Similarity Metric // SODA.- 2003- P. 863-872
3. V.D. Gusev, L.A. Nemytikova, N.A. Chuzhanova. On the Complexity Measures of Genetic Sequences // Bioinformatics, Vol.15, № 12, 1999, 994–999.
4. Gusev, V.D., Nemytikova, L.A., and Chuzhanova, N.A. Rapid Method for Identification of Interconnections between Functionally and/or Evolutionarily Related Biological Sequences.// Mol. Biol. (Russia), 2001, vol. 35, no. 6, pp. 867-873
5. J.-S.Varré, J.-P.Delahaye, E. Rivals: Transformation Distances: a Family of Dissimilarity Measures Based on Movements of Segments. // Bioinformatics 15(3): 194-202 (1999)
6. L. A. Miroshnichenko, V. D. Gusev, I. I. Kiknadze, L. I. Gunderina, and A. G. Istomina Complexity Decompositions in the Problem of Comparison of Polytene Chromosome Banding Sequences // Pattern Recognition and Image Analysis, 2012, Vol. 22, No. 4, 607-613
7. Bakhmutova I.V., Gusev V.D., Titkova T.N. The Search for Adaptations in Song Melodies // Computer Music Journal, Santa Barbara, 1997, Vol. 21, No.1, 58-67
8. Bakhmutova I.V., Gusev V.D., Titkova T.N. Computational Analysis of Neume Component in the Collection of V.M. Metallov. // Siberian Musical Almanac. Novosibirsk, 2010, 62-81.
9. Hannenhalli, S. and Pevzner, P. Transforming Cabbage into Turnip (Polynomial Algorithm for Sorting Signed Permutation by Reversals). Proc. 27th Ann. ACM Symposium on the Theory of Computing, 1995, pp. 178–189.

ONE APPROACH TO DATA ANALYSIS IN PATTERN RECOGNITION

N.N. Katerinochkina[1]

[1] Dorodnicyn Computing Centre, Russian Academy of Science,
ul. Vavilova 40, Moscow, 119333 Russia,
e-mail: nnkater@yandex.ru

One approach to data analysis in recognition problems on precedents is investigated. The search task for logical regularities of the classes is considered. The concept of elementary predicate is introduced. This predicate determines the belonging of the object to any half-space in the space of features. The logical regularities, which are the disjunctive forms from elementary predicates, are examined. The search methods for these logical regularities are proposed. These methods are based on the constructing convex hulls of subsets of the training sample.

Introduction

One approach to data analysis in recognition problems is presented. The standard recognition tasks on precedents are examined. For these tasks the search methods for logical regularities are proposed.

The concept of logical regularity was introduced in [1]. In this work the predicates of the special form were determined. These predicates determine system of neighborhoods of training objects and have the form of parallelepipeds in the feature space. We call these predicates the logical regularities of the first type. In works [2, 3] the methods of constructing these logical regularities were developed. Here we examine logical regularities of other types.

Formulation of the problem

We consider the following recognition task.

Suppose that a set $\{S_1, S_2,\ldots, S_m\}$ of object descriptions is given. This set is divided into l nonintersecting classes $K_1, K_2,\ldots, K_l$. The set $\{S_1, S_2,\ldots, S_m\}$ is called a training sample, and its elements are called precedents or templates. We assume that the objects are defined by the collections of the numerical features:

$x_1, x_2,\ldots, x_n$.

Denote by $\mathbf{R}$ the set of real numbers. We will denote the object that we will recognize, by $S = S(\mathbf{x})$ or by $\mathbf{x} = (x_1,...,x_n)$, where $x_i \in \mathbf{R}$.

For any object it is necessary to define to what class it belongs.

Questions of data analysis are investigated for this task. For an arbitrary class K_λ we will build logical regularities.

Definition 1. *A logical regularity* for the class K_λ is a predicate $P(\mathbf{x})$ such that $P(\mathbf{x}) = 1$ for many templates of this class and $P(\mathbf{x}) = 0$ for all templates which do not belong to the class K_λ.

Thus each predicate describes a part (subset) of the objects of the class, and the totality of all such predicates describes the entire class K_λ.

Definition 2. The predicate $P_i(\mathbf{x})$ is called *an elementary predicate* if it has the following form:

$$P_i(\mathbf{x}) = (\sum_{j=1}^{n} a_{ij}x_j - b_i \le 0). \qquad (1)$$

We will write it also in the form:

$$P_i(\mathbf{x}) = ((\mathbf{a}_i, \mathbf{x}) - b_i \le 0). \qquad (2)$$

Here $(\mathbf{a}_i, \mathbf{x})$ is scalar product of the vectors $\mathbf{a}_i, \mathbf{x} \in \mathbf{R}^n$, and $b_i \in \mathbf{R}$.

Definition 3. The predicate is called *a logical regularity of the second type* if it has the following common form:

$$P(\mathbf{x}) = \bigvee_{j=1}^{p} \&_{i=1}^{q} P_i^j(\mathbf{x}), \qquad (3)$$

where $P_i^j(\mathbf{x})$ are some elementary predicates.

We assume that the objects of each class belong to some connected closed domain in the space $\mathbf{R}^n$. Let us exclude the cases, where

different classes are very mixed, for example, rational and irrational numbers.

So we will assume that each class can be separated by a certain boundary from the objects of other classes. Our purpose is the constructing the predicate of form (3) that realizes this separation.

The proposed search method for logical regularities is based on the constructing the convex hulls of subsets of the training sample.

Let us introduce a number of designations.

Let $S(K_i)$ be the subset of the training sample that consists of the objects of class K_i.

By $V(K)$ denote the convex hull of arbitrary final set K.

By $V(K_i)$ we denote the convex hull of the precedents of class K_i. So we have

$$V(K_i) \equiv V(S(K_i)).$$

Special cases of the recognition task

Let us consider the recognition task for two classes K_1, K_2. In the optimal case, the convex hulls of the precedents of these classes do not intersect:

$$V(K_1) \bigcap V(K_2) = \varnothing.$$

It is known, that if the sets M and N are convex polyhedrons with the empty intersection, then they are divided by affine function. Consequently, the sets $V(K_1)$ and $V(K_2)$ can be divided by the hyperplane of following form:

$$l(\mathbf{x}) \equiv \sum_{j=1}^{n} a_j x_j - b = 0.$$

On this hyperplane it is easy to construct the elementary predicates for both classes.

In more complicated cases, the intersection mentioned above is not empty. Then the situation is possible, where

$V(K_1) \bigcap S(K_2) = \varnothing$ and $V(K_2) \bigcap S(K_1) \neq \varnothing$.

In this case, the class K_1 can be described by the convex hull of its precedents, or by the predicate of form:

$$P^1(\mathbf{x}) = \&_{i=1}^{q}((\mathbf{a}_i, \mathbf{x}) - b_i \le 0).$$

The condition of the belonging some object to the class K_2 can be defined as the negation of its belonging to the class K_1. Namely, this condition may be written as follows:

$$P^2(\mathbf{x}) = \vee_{i=1}^{q}(-(\mathbf{a}_i, \mathbf{x}) + b_i + \varepsilon \le 0)$$

with some $\varepsilon > 0$.

Let us examine the more general case:

$V(K_1) \bigcap S(K_2) \neq \varnothing$ and $V(K_2) \bigcap S(K_1) \neq \varnothing$.

Let us construct the set

$$W_2 = V(V(K_1) \bigcap S(K_2)).$$

This set is the convex hull of those precedents of the class K_2, which got into the set $V(K_1)$.

If $W_2 \bigcap S(K_1) = \varnothing$, then the condition $\mathbf{x} \in K_1$ is equivalent to the condition

$$(\mathbf{x} \in V(K_1)) \,\&\, (\mathbf{x} \notin W_2).$$

The last condition is described by the predicate of following form:

$$P(\mathbf{x}) = \&_i((\mathbf{a}_i, \mathbf{x}) - b_i \le 0) \,\&\, (\vee_j((\mathbf{c}_j, \mathbf{x}) - d_j \le 0))$$

This predicate is obviously reduced to form (3).

In the case, when the number of classes is more than two ($l > 2$), the good situations may be also possible. Thus, the relations analogous described above may be fulfilled for all pairs of classes. Then for given class K_i, we can construct $l-1$ predicates $P_j(\mathbf{x})$, $j \in \{1,...,l\}$, $j \neq i$, such that $P_j(\mathbf{x})$ divides the class K_i from the class K_j. The conjunction of these predicates is considered as the logical regularity for the class K_i.

In the more complicated cases, "strange" objects can be not compact allocated in the set $V(K_i)$. In these cases, the boundaries between the precedents for the different classes can be too twisting. Then the above indicated methods are not appropriate. It is necessary to use more precision methods.

Description of the general method

for the plane

Suppose that $n = 2$. On the plane the training sample S_1, S_2,…, S_m of objects from the classes K_1, K_2,…, K_l is given. For the arbitrary class K_λ we will construct the logical regularities separating objects of this class from all other objects. By CK_λ denote a set of all objects that do not enter into the class K_λ. Let $S(K_\lambda) = \{S_1,...,S_r\}$ be the set of the precedents for the class K_λ, where $S_i = (x_i, y_i)$.

First let us construct the convex hull of the set $S(K_\lambda)$. We apply for this purpose one of the known methods: T. Chan, R. Graham, D. Kirkpatrik's algorithms, etc. (see for example [4]).

The obtained convex hull $V(K_\lambda)$ is determined by its extreme points (apexes of boundary broken line). Let us denote them by $\mathbf{v}_0, \mathbf{v}_1, ..., \mathbf{v}_t$.

After the constructing the set $V(K_\lambda)$ it is necessary to verify, whether "strange" objects (i.e. objects from the set CK_λ) got to it. If $V(K_\lambda) \bigcap CK_\lambda \neq \varnothing$, then we begin the construction of logical regularity in the common form. For this purpose, we find at first the center of masses $\mathbf{p} = (\bar{x}, \bar{y})$ for the set of precedents of the class K_λ. We have

$$\bar{x} = \frac{1}{r}\sum_{i=1}^{r} x_i, \quad \bar{y} = \frac{1}{r}\sum_{i=1}^{r} y_i.$$

Let us select from the set $S(K_\lambda)$ the point $\mathbf{q} = (x_q, y_q)$ nearest to the point $\mathbf{p}$. Afterwards, we connect the point $\mathbf{q}$ with all apexes of the boundary of the constructed convex hull. Then the set $V(K_\lambda)$ will be represented in the form of association of triangles with the apexes $\mathbf{q}, \mathbf{v}_i, \mathbf{v}_{i+1}$. We will investigate these triangles in turn, beginning with $\Delta\mathbf{q}\mathbf{v}_0\mathbf{v}_1$ and finishing by $\Delta\mathbf{q}\mathbf{v}_t\mathbf{v}_0$.

Let us examine $\Delta\mathbf{q}\mathbf{v}_i\mathbf{v}_{i+1}$. First we verify, whether "strange" objects got to it. For this purpose it is sufficient to check the templates from the set CK_λ with the coordinates satisfying the following relations:

$$\min(x_q, x_i, x_{i+1}) \le x \le \max(x_q, x_i, x_{i+1}),$$

$$\min(y_q, y_i, y_{i+1}) \le y \le \max(y_q, y_i, y_{i+1}),$$

where $\mathbf{v}_i = (x_i, y_i), \mathbf{v}_{i+1} = (x_{i+1}, y_{i+1})$.

If "strange" points got into $\Delta\mathbf{q}\mathbf{v}_i\mathbf{v}_{i+1}$, then it is necessary to exclude them from our set. Denote them by $\mathbf{w}_1^i, ..., \mathbf{w}_s^i$. We will construct the set $V_i = V(\{\mathbf{w}_1^i, ..., \mathbf{w}_s^i\})$, which is the convex hull of objects from the set CK_λ got into $\Delta\mathbf{q}\mathbf{v}_i\mathbf{v}_{i+1}$. After investigating all triangles, we obtain the collection of the sets $V_0, V_1 ..., V_k, \quad k \le t+1$. Suppose that

$$S(K_\lambda) \bigcap (\bigcup_{j=0}^{k} V_j) = \varnothing.$$

Then the belonging condition of object $\mathbf{x}$ to the class K_λ can be written down as follows:

$$(\mathbf{x} \in V(K_\lambda)) \,\&\, (\mathbf{x} \notin \bigcup_{j=0}^{k} V_j).$$

Here the condition $\mathbf{x} \notin \bigcup_{j=0}^{k} V_j$ is equivalent to the condition $\&_{j=0}^{k}(\mathbf{x} \notin V_j)$. Let the set $V(K_\lambda)$ be described by a predicate:

$$\&_{i=1}^{q}((\mathbf{a}_i, \mathbf{x}) - b_i \le 0).$$

The condition $\mathbf{x} \notin V_j$ can be described by the predicate of the following form:

$$\vee_{t=1}^{u_j}((\mathbf{c}_t^j, \mathbf{x}) - d_t^j \le 0).$$

Then the logical regularity for the class K_λ will take the form:

$$\&_{i=1}^{q}((\mathbf{a}_i, \mathbf{x}) - b_i \le 0) \,\&_{j=0}^{k} \vee_{t=1}^{u_j}((\mathbf{c}_t^j, \mathbf{x}) - d_t^j \le 0).$$

It is obviously reduced to the form (3).

The obtained logical regularity describes precisely the class K_λ in that case, when the sets $V_0, V_1 ..., V_k$ do not contain the precedents of this class. Otherwise, we consider this logical regularity as the approximate description of the class. In any case, the condition of this predicate is sufficient (but not always necessary) for the belonging to this class.

The proposed approach is based on the representing the class in the form of difference of a convex set and a polyhedral set. By a polyhedral set we call the association of finite number of convex sets. (About the task of polyhedral separability, see for example [5].)

Some methods of simplification of obtained logical regularities are also developed.

Estimations of the complexity of represented methods are received.

Thus, the complexity of constructing logical regularity of the form (3) is estimated as $O(m \cdot \log m + mt)$. Here m is the number of precedents in the training sample and t is the number of extreme points in the convex hull of the precedents of the class.

The proposed method consists of several stages:

- constructing the convex hull of the precedents of given class;

- finding the center of masses for the set of precedents of this class;
- partitioning the constructed convex hull into triangles;
- constructing the convex hull of "strange" objects in each triangle;
- definition of logical regularity in form (3) for the class.

We will denote the described algorithm by **A2** .

The case *n*>2

Further, propose the approach, which allows to use the algorithm **A2** for *n*>2.

All possible pairs of features (x_i, x_j), $i \neq j$, are examined. For each pair, the projections of the objects from the training sample on the plane of axes *i, j* are investigated. For the class K_λ and the pair (x_i, x_j), we attempt to separate the projections of the precedents of class K_λ from the projections of the set CK_λ.

We introduce the following notations.

Let $\Pi(K_\lambda, i, j)$ be the set of projections of the precedents for the class K_λ on the plane of axes *i, j*.

Let $\Pi(CK_\lambda, i, j)$ be similar set for CK_λ.

Let us construct the predicate of the form (3) for the set $\Pi(K_\lambda, i, j)$ with the help of algorithm **A2.** But here, unlike the case $n = 2$, the domains can be, where the sets $\Pi(K_\lambda, i, j)$ and $\Pi(CK_\lambda, i, j)$ are mixed. Then the obtained predicate together with the objects from the set CK_λ separates many elements of the class K_λ. For the pair (x_i, x_j), denote by ρ_{ij} the number of isolated precedents from the class K_λ.

We select the pair such that the value ρ_{ij} is minimal. For the precedents that do not satisfy the obtained predicate, it is possible to add a third feature or to consider other pairs.

Finely, we select the association of pairs or triples of the features such that the corresponding predicates leave the minimum of the unrecognized precedents.

The disjunction of these predicates is considered as the logical regularity for this class.

Conclusion

In this article one approach to data analysis in recognition tasks on precedents is presented. The search methods for logical regularities of the special form are developed. These methods are based on the constructing convex hulls of subsets of the training sample. A row of special cases of the recognition task is considered. The general algorithm **A2** for a plane ($n = 2$) is described. Estimation of the complexity of represented method is received. The approach, which allows to use the algorithm **A2** for *n*>2, is proposed. The represented methods can be revised and refined.

References

1. Ryazanov V.V., Senko O.V., and Zhuravlev Yu.I. Methods of recognition and prediction based on voting procedures // Pattern Recognition and Image Analysis. – 1999. – Vol. 9, No. 4. – P. 713-718.
2. Ryazanov V.V. Logical regularities in pattern recognition (parametric approach) // Computational Mathematics and Mathematical Physics. – 2007. – Vol. 47, No. 10. – P. 1720-1735.
3. Kovshov N.V., Moiseev V.L., Ryasanov V.V. Algorithms for finding logical regularities in pattern recognition // Computational Mathematics and Mathematical Physics. – 2008. – Vol. 48, No. 2. – P. 314-328.
4. Polovinkin E.S., Balashov M.V. Elements of convex and strongly convex analysis. – Moscow: Fizmatlit, 2004. (In Russian)
5. Eremin I.I. Linear optimization and systems of linear inequalities. – M.: Publishing center "Academy", 2007. (In Russian)

FOURIER-MELLIN TRANSFORM ON A COMPLEX DISCRETE TORUS[1]

V. Mnukhin[2]

[2] Southern Federal University, 44 Nekrasovskiy Lane, 347928 Taganrog, Russia, mnukhin.valeriy@mail.ru

An algebraic method for digital images analysis and processing is considered. It is based on the presentation of digital images of size $p \times p$, where $p \equiv 3 \bmod 4$ is a prime, as functions on a complex discrete torus. For such images *complex rotations* are introduced and a new invertible linear transform, called the *modular Mellin transform,* is presented. The discrete modular Fourier-Mellin transform, invariant under circular shifts, scaling and complex rotations, is also defined.

Introduction

When developing algorithms for image analysis, it is quite common to proceed on the assumption of the continuity of images. Then powerful tools of continual mathematics, such as complex analysis and integral transforms, can be efficiently used. At the same time, formal application of gained results to digital images could be complicated by systematic sampling errors. As an example, we refer to the well-known Fourier-Mellin transform [1,2], used extensively for image registration [3, p. 674]. It is based on a log-polar representation of the 2D-Fourier spectrum of an image, so that rotations and scaling of *continuous* images yield shifts of the transformed spectrum, therefore producing invariants under affine transformations. Unfortunately, it was pointed out [1] that formal adaptation of continuous log-polar coordinates to *digital* images brings up crucial diffculties and results in substantial errors. So new transforms, originally designed for digital images, could be useful.

In this paper we present a new invertible linear discrete *modular Mellin transform* based on an algebraic interpretation of log-polar coordinates in the complex plane. It is applicable to digital images of sizes $p \times p$, where p is a prime such that $p \equiv 3 \bmod 4$ (so are, for example, 71, 127, 251, 257 and 509). Such images are considered as functions on "finite complex fields" or *complex discrete tori.* Then *complex rotations* as continuous rotations through rational tangent angles, combined with appropriate scaling, are introduced. Note that complex rotations do not preserve distances, angles and areas, so that distortions of images may occur. At the same time transformed images keep all the information about originals and are just permutations of their pixels. Moreover, complex rotations preserve orthogonality and can be used, for instance, for symmetry analysis of digital images [4,5]. Finally, we define a *modular Fourier-Mellin* transform, which is invariant under circular shifts, scaling and complex rotations.

A complex discrete torus

Let $\mathbb{Z}_n = \mathbb{Z} / n\mathbb{Z}$ be a ring of integers modulo $n \geq 2$. The direct products $\mathbb{Z}_m \times \mathbb{Z}_n$ are called *discrete tori*. For a prime p, the torus $\mathbb{Z}_p^2 = \mathbb{Z}_p \times \mathbb{Z}_p$ is a 2-dimensional vector space over a finite field with p elements. Moreover, if p is prime such that $p \equiv 3 \bmod 4$, the torus $\mathbb{Z}_p^2$ may be considered as a "finite complex field". Indeed, following the standard construction [6, p. 515] for the complex field $\mathbb{C}$, consider the factor ring

$$\mathbb{C}_p = \mathbb{Z}_p[x] / (x^2 + 1).$$

It is well-known [7, p. 176], that -1 is a quadratic nonresidue of $p \equiv 3 \bmod 4$. Hence,

[1] The work was supported by RFBR grants 11-07-00591 and 13-07-00327.

the polynomial x^2+1 is irreducible over $\mathbb{Z}_p$ and so

$$\mathbb{C}_p = \{a+ib : a,b \in \mathbb{Z}_p\}$$

is a field with p^2 elements [6, p. 293]. (Note that i stands for the congruence class $\overline{x}$.) The evident bijection $(a,b) \leftrightarrow a+ib$ transfers the field structure to $\mathbb{Z}_p^2 \simeq \mathbb{C}_p$.

By analogy with Gaussian integers [6, p. 229], we call elements of the field $\mathbb{C}_p$ by *discrete complex numbers* and use for them the standard terminology of complex analysis. In particular, the *norm* of $z = a+ib \in \mathbb{C}_p$ is $N(z) = a^2+b^2 \in \mathbb{Z}_p$ (Note that the modulus $|z| = \sqrt{N(z)}$ is undefined in $\mathbb{C}_p$.)

Definition 1. When p is a prime number such that $p \equiv 3 \bmod 4$, the torus $\mathbb{Z}_p^2 \simeq \mathbb{C}_p$ is called the *complex discrete torus*.

Complex Rotations of Digital Images

A *digital image* of size $m \times n$ may be considered as a function $f : \mathbb{Z}_m \times \mathbb{Z}_n \to \mathbb{C}$ on a discrete torus. There are several natural transforms of such images. For example, a *circular shift* $T_{(a,b)}$ is defined by

$$T_{(a,b)}[f(x,y)] = f(x-a, y-b).$$

Further, in a torus $\mathbb{Z}_p^2$, where p is a prime, an invertible transformation S_k of *scaling* is defined by

$$S_k[f(x,y)] = f(kx,ky), \text{ where } k \in \mathbb{Z}_p.$$

At the same time *rotations* of digital images are undefined in general. Note that *continuous rotations* in the complex plane $\mathbb{C}$ can be defined by $R_\alpha[f(z)] = f\left(e^{i\alpha}z\right)$. So we consider the following transform of functions f on a complex discrete torus:

$$R_w[f(z)] = f(wz), \text{ where } 0 \neq w \in \mathbb{C}_p.$$

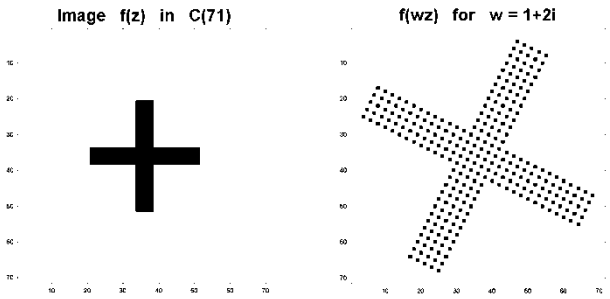


Fig. 1. A complex rotation in $\mathbb{C}_{71}$.

Fig. 1 shows $R_w[f]$ of a 71×71-image f, when $w = 1+2i \in \mathbb{Z}_{71}$. The resulting image may be considered as a continuous rotation of f through the angle $\alpha = \operatorname{arctg}(2)$ combined with the scaling at $\sqrt{5}$.

Definition 2. The transformation $R_w[f]$ of a digital image f on a complex discrete torus $\mathbb{C}_p$ is called the *complex rotation by* $w \in \mathbb{C}_p$.

The complex rotations do not preserve distances, angles and areas, so that distortions of images may occur, as Fig. 2 shows. (Note that the Lenna image from the Brodatz database has been reduced to 251×251 size.) Nevertheless, transformed images keep all the information about originals and are just permutations of their pixels. Moreover, complex rotations preserve orthogonality and so may be used, in particular, for symmetry analysis of digital images [2,4,5].

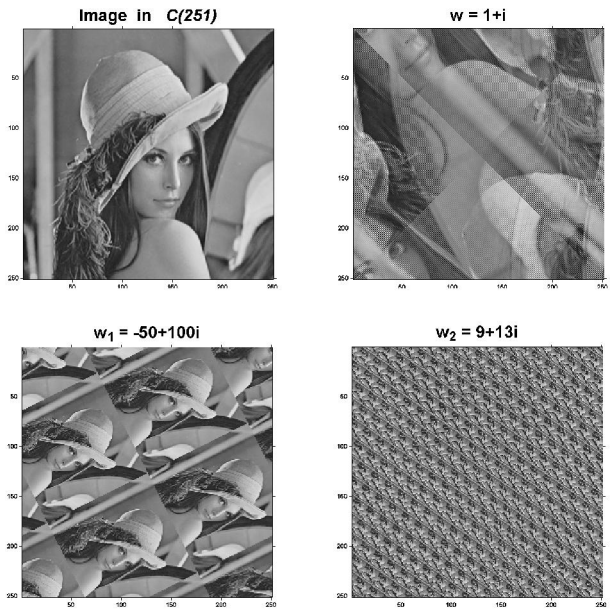


Fig. 2. Complex rotations in $\mathbb{C}_{251}$.

Logarithms and Polar Domains

The purpose of this section is to introduce log-polar coordinates on a complex digital torus. So let $\mathbb{C}_p^* = \mathbb{C}_p \setminus \{0\}$ be the multiplicative group of a finite field $\mathbb{C}_p$. Then $\mathbb{C}_p^*$ is cyclic of order $p^2 - 1$, generated by any primitive element g [6, p. 314]. For instance, one can choose $g = 2 + 7i$ for $p = 71$, and $g = 1 + 5i$, when $p = 251$.

Note [6] that the multiplicative group $\mathbb{C}^*$ of the complex field can be decomposed as

$$\mathbb{C}^* \simeq \mathbb{R}^+ \times \mathrm{U}(1),$$

where $\mathbb{R}^+ = \{r^2 : r \in \mathbb{R}^*\}$ is the multiplicative group of positive real numbers, and $\mathrm{U}(1) = \{z \in \mathbb{C} : N(z) = 1\}$ is the one-dimensional unitary group. Geometrically, this decomposition is just the polar coordinate system in the complex plane. It occurs that a similar decomposition holds for a complex discrete torus also.

Proposition 1. *Let $\mathbb{C}_p$ be a complex discrete torus. Then its multiplicative group $\mathbb{C}_p = \langle g \rangle$ can be decomposed as*

$$\mathbb{C}_p^* \simeq \mathbb{Z}_m \times \mathbb{Z}_n,$$

where $m = (p-1)/2$, $n = 2(p+1)$. *Moreover,*

$$\mathbb{Z}_m = \langle g^n \rangle \simeq \{a^2 : 0 \neq a \in \mathbb{Z}_p\},$$
$$\mathbb{Z}_n = \langle g^m \rangle \simeq \{z \in \mathbb{C}_p^* : N(z) = \pm 1\}.$$

The discrete torus $\mathbb{Z}_m \times \mathbb{Z}_n$ is called *the polar domain* for $\mathbb{C}_p$. If $(l, \theta) \in \mathbb{Z}_m \times \mathbb{Z}_n$ corresponds to $z \in \mathbb{C}_p^*$, then l behaves like the "logarithm of modulus" and θ — like the "argument" of the discrete complex number z.

The observation above can be used to define complex logarithms in $\mathbb{C}_p^*$. First, define

$$\mathrm{Exp}_g : \mathbb{Z}_m \times \mathbb{Z}_n \to \mathbb{C}_p^*$$

by $\mathrm{Exp}_g(l, \theta) = g^{nl + m\theta}$. Note that Exp_g is an isomorphism between the additive group of the ring $\mathbb{Z}_m \times \mathbb{Z}_n$ and the multiplicative group $\mathbb{C}_p^*$. Then the inverse mapping

$$\mathrm{Exp}_g^{-1} = \mathrm{Ln}_g : \mathbb{C}_p^* \to \mathbb{Z}_m \times \mathbb{Z}_n$$

is called the *modular logarithm* to base g. It follows immediately from the definition that

$$\mathrm{Ln}_g(z_1 z_2) = \mathrm{Ln}_g(z_1) + \mathrm{Ln}_g(z_2).$$

Modular logarithms share a lot of common features with principal values of ordinary complex logarithms. Here we only mention the *change of base* property:

Proposition 2. *Let g and h be different primitive elemens for $\mathbb{C}_p^*$. Then for any $z \in \mathbb{C}_p^*$ holds*

$$\mathrm{Ln}_g(h) \odot \mathrm{Ln}_h(z) = \mathrm{I}_p \odot \mathrm{Ln}_g(z),$$

where $\odot$ stands for the multiplication in the ring $\mathbb{Z}_m \times \mathbb{Z}_n$, and $\mathrm{I}_p = (4, m)^{-1} \in \mathbb{Z}_m \times \mathbb{Z}_n$ is a constant, that depends on p only. (For example, $\mathrm{I}_{71} = (9, 107)$ and $\mathrm{I}_{251} = (94, 125)$.)

We associate with a digital image f in $\mathbb{C}_p$ the image ψ in the polar domain $\mathbb{Z}_m \times \mathbb{Z}_n$ such that

$$\psi\big(\mathrm{Ln}_g(z)\big) = f(z).$$

Definition 3. The image $\psi \in \mathbb{Z}_m \times \mathbb{Z}_n$ is called the *polar image* for f, and $\mathrm{P}_g[f] = \psi$ is called the *polar transform* to base g.

The following proposition shows that P_g behaves just like the regular continuous log-polar transform.

Proposition 3. *Complex rotations in $\mathbb{C}_p$ correspond to circular shifts in the polar domain: if $\mathrm{P}_g\big[f(z)\big] = \psi(l, \theta)$, then*

$$\mathrm{P}_g\big[f(wz)\big] = \psi(l - l_0, \theta - \theta_0),$$

where $w \in \mathbb{C}_p^*$ and $(l_0, \theta_0) = \mathrm{Ln}_g(w)$.

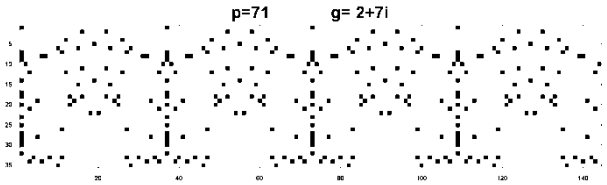


Fig. 3. The polar transform of the image in Fig. 1

Example 1. Let $f \in \mathbb{C}_{71}$ be as shown in the left side of Fig. 1. Its polar image $\mathrm{P}_g[f]$ has size 35×144 and for $g = 2 + 7i$ is shown in Fig. 3. It is easy to check that

$$\mathrm{Ln}_g(1 + 2i) = (14,22) \in \mathbb{Z}_{35} \times \mathbb{Z}_{144},$$

and so the polar image of the complex rotation of f by $w = 1 + 2i$ can be obtained by the circular shift of $\mathrm{P}_g[f]$ by 22 pixels right and 14 pixels down.

Finally, note that the polar transform is "invertible up to a single pixel" (0,0). Appending formally this extra pixel to the polar domain, we may say that P_g^{-1} exists in the "extended" polar domain.

The Modular Mellin Transform

Evidently, the discrete Fourier transform [8] is defined on a discrete torus $\mathbb{Z}_m \times \mathbb{Z}_n$:

$$\mathrm{F}[f] = F(u,v) = \sum_{(x,y) \in \mathbb{Z}_m \times \mathbb{Z}_n} f(x,y)\, \mathrm{e}^{-2\pi i \left(\frac{ux}{m} + \frac{vy}{n}\right)},$$

where $i \in \mathbb{C}$ stands for the regular imaginary unit, and x, y, u, v are representatives of the corresponding residue classes.

Definition 4. Let f be a digital image on a complex digital torus $\mathbb{C}_p$. Its *modular Mellin transform* base g is the discrete Fourier transform of the polar image of f:

$$\mathrm{M}_g[f] = M_g(u,v) = \mathrm{F}\left[\mathrm{P}_g[f]\right].$$

It is easy to note that $\mathrm{M}_g[f]$ is linear and invertible in the extended polar domain. The next result follows from Proposition 2 and the Shif Theorem [8]:

Corollary 1. *The spectrum of the modular Mellin transform of a digital image is invariant under complex rotations of the image:*

$$\left|\mathrm{M}_g[f(wz)]\right| = \left|\mathrm{M}_g[f(z)]\right|.$$

In a similar way, such transformations as $\mathrm{P}^{-1}\mathrm{FP}$, FPF^{-1}, and $\mathrm{P}^{-1}\mathrm{FPF}^{-1}$ could be introduced. In particular, the transformation

$$\overline{\mathrm{H}}_g[f] = \left|\mathrm{M}_g\left[\left|\mathrm{F}^{-1}[f]\right|\right]\right|$$

is invariant under circular shifts, scaling and complex rotations, and so may be considered as a discrete analog of the Fourier-Mellin transform.

Conclusion

A method for digital image analysis, based on a presentation of digital images of size $p \times p$, where $p \equiv 3 \bmod 4$ is a prime, as functions on a complex discrete torus, has been considered. A new invertible linear transform of such images, called the modular Mellin transform, has been presented and its properties have been studied. The modular Fourier-Mellin transform has also been introduced.

References

1. S. Derrode, F. Ghorbel. Robust and efficient Fourier-Mellin transform approximations for gray-level image reconstruction and complete invariant description // Computer Vision and Image Understanding. – 2001.– Vol. 83, No. 1. – P. 57-78.
2. S. Derrode, F. Ghorbel. Shape analysis and symmetry detection in gray-level objects using the analytical Fourier-Mellin representation // Signal Processing. – 2004. – Vol. 84, No. 1. – P. 25-39.
3. W.K. Pratt. Digital Image Processing. – John Wiley and Sons. – 2007. – 782 p.
4. A.N. Karkishchenko, V.B. Mnukhin. Symmetry recognition in the frequency domain (in Russian). // 9th International Conference on Intelligent Information Processing. – 2012. – TORUS Press, Moscow. – P. 426-429.
5. A.N. Karkishchenko, V.B. Mnukhin. Continuous symmetry transform of periodic patterns in the frequency domain (in Russian). // 15th All-Russian Conference on Mathematical Methods in Image Recognition.–2011.–MAKS Press, Moscow.–P.386-389.
6. D.S. Dummit, R.M. Foote. Abstract Algebra. – John Wiley and Sons. – 2004. – 932 p.
7. D.M. Burton. Elementary Number Theory. – McGrow Hill. – 2007. – 434 p.
8. A.D. Poularikas. The Transform and Applications Handbook. – CRC Press. – 2010. – 1336 p.

THE OSCILLATORY NEURAL NETWORKS BASED ON KURAMOTO MODEL FOR CLUSTER ANALYSIS

A.V. Novikov[1,2], E.N. Benderskaya[1,3]

[1] St.-Petersburg State Polytechnical University.
Russia, 195251, St.Petersburg, Polytechnicheskaya, 29.
Phone: +7 (812) 297-4218.
E-mails: [1] spb.andr@ya.ru, [2] helen.bend@gmail.com

This abstract presents the research of synchronization processes in oscillatory neural networks with different structures, estimates of the local and global synchronization of oscillators. Considered models of oscillator are based on the Kuramoto equation. Practical ways to use oscillatory networks for solving the problems of clustering of N-dimensional data and also presented ways to use mentioned networks for a special case of clustering analysis – coloring problem.

Introduction

Neural Networks widely used in different areas for solving various problems such as clustering, pattern segmentation and recognition [9]. Moreover neural networks ensure parallel process of computing through the implementation of mathematical algorithms in its own structure.

Supposedly synchronization among neurons in the brain is used to implement the cognitive functions, for example, vision, motion, memory [3, 6, 9]. Thereby research problem of synchronization in oscillatory neural networks is relevant since it also provides biological methods of solutions that are used by nature. Although simulation of oscillatory network is expensive computationally, their implementation using specific hardware is able to ensure faster solution and maximum performance than traditional mathematic methods.

One of the successful models of synchronization was proposed by Kuramoto. The model attracted attention of many researches due to flexibility and simplicity of model. The present increase of interest in this model is related to the growing interest in the processes of synchronization in oscillatory neural networks.

Synchronization models

Kuramoto model of synchronization makes it possible to simulate the processes of synchronization between oscillators that may have different internal parameters and communication structures [1, 2]. Dynamic of the model described by following equation [10]:

$$\dot{\theta}_i = \omega_i + \frac{K}{N} \cdot \sum_{j=1}^{N} \sin\left(\theta_j - \theta_i\right). \qquad (1)$$

Phase of oscillatory θ_i is basic state variable and is disposed in the range from 0 to 2π. Frequency ω_i can be regarded as offset that occur during synchronization in presented model. Coupling strength K between oscillators is one of the most important parameter influences process of synchronization. N is total number of oscillators in the network.

States of global synchronization, de-synchronization and partial synchronization in the network can be ensured by coupling strength K. The high value of coupling strength provides a high rate of global synchronization. Choice of weak coupling strength K that less than critical coupling strength K_c will be cause of de-synchronization. Before explanation the way to ensure partial synchronization we should

consider the estimate of synchronization and role of critical value of coupling strength.

The degree of synchronization between oscillators can be evaluated by estimate r that helps to define state of synchronization [10]:

$$r = \left| \frac{1}{N \cdot e^{i \cdot \varphi}} \cdot \sum_{j=1}^{N} e^{i \cdot \theta_j} \right|, \quad \varphi = \frac{1}{N} \cdot \sum_{j=1}^{N} \theta_j \ . \quad (2)$$

The state of global synchronization occurs when $r \to 1$, global de-synchronization occurs when $r \to 0$. Partial synchronization occurs in case:

$$r \to \sqrt{1 - \frac{K_c}{K}} \ . \quad (3)$$

The degree of partial synchronization depends on value of critical coupling strength K_c that is depends on the width of the frequency distribution of oscillator: $K_c = 2\gamma$.

An important feature of the Kuramoto model is possibility to provide synchronization processes in networks with different communication structures. We performed experimental study using numerical simulations and found that states of global and partial synchronization can be successfully sets in the oscillatory networks with communication structures such as grids, stars, bidirectional list and unidirectional circular list.

Also research confirms that oscillatory neural networks that are based on Kuramoto model are robust [7, 12]. We have used model with noise, biases and random coupling strength between oscillators:

$$\dot{\theta}_i = \omega_i + \zeta + \sum_{j=1}^{N} \frac{K_{ij}}{N} \cdot \sin\left(\theta_j - \theta_i + A_{ij}\right) . \quad (4)$$

The oscillators have a tendency to be synchronized despite the disorder parameters seek to disrupt the synchronization. The similar results were obtained in several other publications [7].

Cluster analysis of N-dimensional data

Synchronization states in oscillatory neural networks can be used for solving various problems [4, 6, 12]. The main subject of this article is applying oscillatory neural networks based on Kuramoto model for cluster analysis.

The clustering problem is the problem of portioning a given sample of objects into disjoint subsets called clusters. Each cluster should consist of maximally similar parameters of objects. In other words, the solution of the clustering problem that uses the principles of neural network is to provide states of partial synchronization. Thus each ensemble synchronous oscillators corresponds to one cluster of data.

Consider the clustering problem of N-dimensional data using an oscillatory neural network based on a modified model Kuramoto [11]:

$$\frac{\partial \theta_i}{\partial t} = \frac{K}{N_i} \cdot \sum_{j \in N_i} \sin\left(\theta_j - \theta_i\right) . \quad (5)$$

The principal difference modified model from basic Kuramoto model (1) is lack of oscillator frequency and using amount of neighboring oscillators instead of the total number of oscillators. The lack of frequency contributes to accuracy. The set of oscillator neighbors N_i is defined by special radius ϵ. Radius ϵ is the key parameter of clustering that specified by user and should be chosen in line with input data.

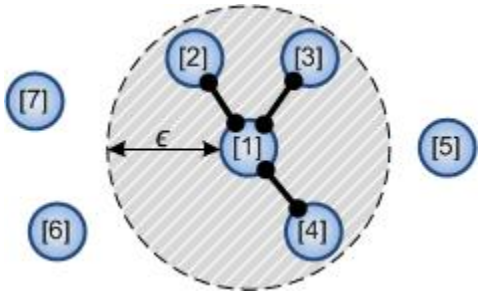


Fig. 1. Illustration of how connections are established for oscillator in line with radius ϵ.

The structure of the oscillatory network for solving clustering problem is defined by a set of input data. Each oscillator is placed in accordance to one object from the set of input data and coordinates of the object correspond to the coordinates of the oscillator. Connections are established between neighboring oscillators if distance between them is less than radius ϵ. Illustration of the formation of the network structure is shown in figure 1.

We propose to evaluate the end of the process clustering by degree of partial synchronization r_c that is described as follows:

$$r_c = \left| \sum_{i=1}^{N} \frac{1}{N_i} \sum_{j \in N_i} e^{\theta_j - \theta_i} \right|. \quad (6)$$

Ending process synchronization (clustering) is indicated when $r_c \to 1$.

We have investigated possibilities of the network to solve problem of clustering using the widespread data set FCPS [13]. Experiments have shown that clusters successfully allocated for all samples, except data with title "EngyTime" where two clusters have equal density and intersect in the two dimensional space. Example of clustering is presented in figure 2. The reason for the failure is because the clusters are formed by local oscillator networks that are based on the input data. In this case we have one oscillatory neural network with strongly coupling and as a result one allocated cluster.

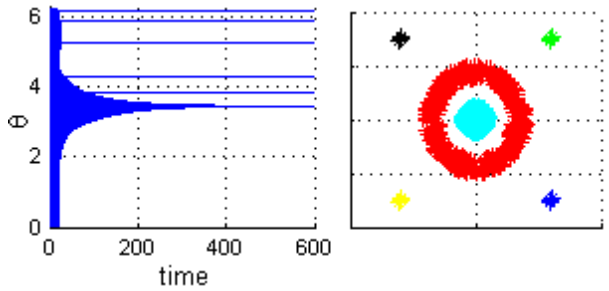


Fig. 2. Dynamics of oscillators (left) and clusters that have been allocated $r_c = 0.99$ (right).

An important issue is the choice of the right radius for successful clustering. In the publication [5] was offered a new approach avoids using the radius. The main idea of this approach can be divided by three general steps. In the first step, average distance should be calculated for connecting at least three oscillators using for example k-means algorithm. Second step, average distance should be increased for connecting four oscillators when process of local synchronization is finished $r_c \to 1$. Second step should be repeated until global synchronization will not be reached $r \to 1$. In the third step, dendrogram of the dynamics of the phases of oscillators should be analyzed to identify true clusters. Of course this approach requires more execution time than approach with radius, but advantage is lack of the mentioned radius.

Graph coloring problem

Synchronization processes in oscillatory neural networks based on Kuramoto models also can be applied for graph coloring problems. In graph theory, graph coloring is a special case of graph labeling. But this problem can be also viewed as a special case of clustering. Consider following case of graph coloring: vertices should be colored by minimum possible number of colors and besides adjacent vertices that have a connection must be colored by different colors.

Modified model Kuramoto for solving described graph coloring problem [14]:

$$\dot{\theta}_i = \omega_i + \frac{1}{K_{MAX}} \cdot \sum_{j=1}^{N} K_{ij} \cdot \sin\left(\theta_j - \theta_i\right), \quad (7)$$

$$K_{ij} = \begin{cases} K_p > 0, & e_{ij} \in \bar{E} \\ K_n < 0, & e_{ij} \in E \end{cases}. \quad (8)$$

The principal difference in presented model (7) is negative connections between the oscillators that can be named as inhibitory connections. Scaling value K_{MAX} is defined as maximum degree of the inverse graph that is used for forming the network structure.

The number of oscillators corresponds to the number of nodes in a graph where each node is matched to only one oscillator. The negative coupling is established between the oscillators whose vertices have connection in the graph. The positive coupling is established between oscillators whose vertices don't have connection in the graph. This architecture allows to oscillators be de-synchronized between themselves if their vertices have connection and be synchronized if their vertices have no connection. The ensemble of synchronized oscillators corresponds to one color.

Positive K_p and negative K_n coupling are key parameters for solving graph coloring problem. Choice of these parameters significantly influences quality of coloring.

The accuracy of the model (7) can be improved by removing the frequency of oscillator ω. The result of graph coloring isn't always optimal as shown in figure 3 where minimum number of colors sufficient for graph coloring must be equal to three. Deviation can be explained by fact that formed clusters are trying to pull shared vertices and as a result distinguishing them into a distinct cluster. Despite the shortcoming this algorithm allows to avoid polynomial complexity due to

the principles of parallelism that are inherited from the concept of oscillatory neural networks.

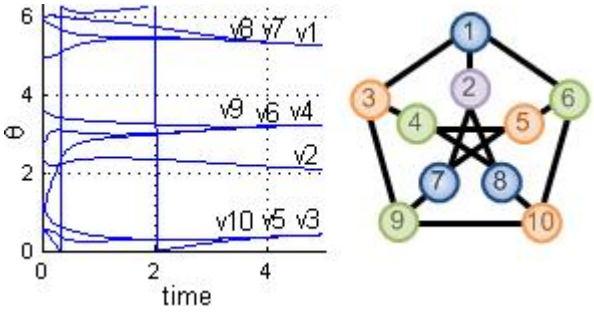


Fig. 3. Dynamics of oscillators without frequency $\omega = 0$ (left) and result of graph coloring $r_c = 0.99$ (right)

We extend architecture of oscillatory network for solving total coloring problem where vertices and edges must be colored. Total coloring assumes that adjacent vertices, adjacent edges and edge and its vertices must be assigned by different colors. In this case number of oscillators equal to total number of vertices and edges. One oscillator can correspond to vertices or to edge. The negative coupling is established between oscillators that correspond to adjacent vertices or adjacent edges or edge and its vertices. The model of oscillator is remained the same (7).

Experiments have shown that described approach successfully colored graph and edges but also inherits the same shortcoming from model of oscillator.

Conclusions

The various synchronization models of phase oscillator were analyzed. Oscillatory neural networks based on Kuramoto model are robust and able to ensure various states of synchronization that depend on parameters or even structures of networks. Thus these oscillatory networks can be successfully applied for solving practical problem.

Investigated possibilities of applying models of synchronization for clustering analysis and have shown general advantages and disadvantages of them. Results of clustering depend on the structures of the networks that depend on the input data. The estimate of clustering results has been proposed.

Also have investigated possibilities of applying model with negative and positive coupling for a special case of clustering – graph coloring problem. Have proposed approach of forming structure of oscillatory network for solving problem of total graph coloring.

References

1. Arenas, A. Diaz-Guilera, J. Kurths, Y. Moreno, Z. Changsong. Synchronization in complex networks // Physics Reports 469. - 2008. - P. 93-153.
2. J.A. Acebron, L.L. Bonilla, C.J.P. Vincente, F. Rotort, R. Spigler. The Kuramoto Model: A Simple Paradigm for Synchronization Phenomena // Rev. Mod. Phys. - 2005. - Vol. 77, P. 137-185.
3. E. Basar. Brain function and oscillations / Springer-Varlag, New York. - 1998. - 363pp.
4. E.N. Benderskaya, S.V. Zhukova. Large-Dimension Image Clustering by Means of Fragmentary Synchronization in Chaotic Systems // Pattern Recognition and Image Analysis. - 2009. - Vol. 19, No.2, P. 306 - 316.
5. Bohm, C. Plant, J. Shao, Q. Yang. Clustering by synchronization // KDD '10 Proceeding of the 16th ACM SIGKDD international conference of Knowledge discovery and data mining. - 2010. - P. 583-592.
6. Cumin, C.P. Unsworth. Generalizing the Kuramoto Model for the Study of Neuronal Synchronisation in the Brain // Report University of Auckland School of Engineering 638. - 2006.
7. H. Daido. Quasientrainment and slow relaxation in a population of oscillators with random and frustrated interactions // Phys. Rev. Lett. 73. - 1992. - P. 1073-1076.
8. I.B. Gurevich. Image recognition problem // Recognition, classification, prediction. Mathematical methods and their application: Yearbook / Edited by Y. Zhuravlev. Moscow: Science. - 1988. - Vol. 1, P. 280. (in Russian)
9. H. Haken. Brain Dynamics / Springer-Verlag Berlin Neidelberg. - 2007. - P. 238.
10. Y. Kuramoto. Chemical Oscillations Waves, and Turbulence / Springer-Verlag Berlin Neidelberg New York Tokyo. - 1984 – P. 157.
11. T. Miyano, T. Tsutsui. Data Synchronization as a Method of Data Mining // International Symposium on Nonlinear Theory and its Applications. - 2007.
12. P.S. Skardal, E. Ott, J.G. Restrepo. Cluster synchrony in systems of coupled phase oscillators with higher-order coupling // Phys. Rev. E84, 036208. - 2011.
13. A.Ultsch. Clustering with SOM: U*C // Workshop on Self Organizing Feature Maps. - 2005. - P. 31-37.
14. J. Wu J, L. Jiao, W. Chen. Clustering dynamics of nonlinear oscillator network: Application to graph coloring problem // Physica D, 240. - 2011. - Vol. 20, Mo. 2, P. 1972-1978.
15. Yu.I Zhuravlev. An algebraic approach to recognition or classification problems // Pattern Recognition and Image Analysis. (Advances in Mathematical Theory and Applications). - 1998. - No.8, P. 59-100.

CASCADE OF CLASSIFIERS BASED ON A HIERARCHICAL SEQUENCE OF GAUSSIAN MIXTURES[1]

N.A. Novikov[2]

[2] Dorodnicyn Computing Centre, Russian Academy of Sciences, Moscow, Russian Federation, e-mail: nikknovikov@gmail.com

A binary classification method based on a cascade of Gaussian mixture models is suggested. An iterative algorithm of splitting the mixture components is applied in order to reduce classification time. An efficiency of the method is evaluated on model data in terms of relation between classification time and error rate. A ratio of false reject and false accept penalties as well as a ratio of sizes of two classes that provide an advantage of the proposed cascade Gaussian mixture-based classifier as against single Gaussian mixture-based classifiers are discussed.

Introduction

The real-time classification problem occurs in many tasks such as object detection, speech recognition, brain-machine interfaces and the others. An idea to solve complex classification problems on the basis of a sequence of simple decision rules was used in a concept of decision trees [1, 2]. An application of degenerate decision trees for a fast binary classification problem was proposed by P. Viola and M. Jones in [3]. These trees (cascades) consist of classifiers with progressively increasing complexity so the classification time can be strongly reduced by rejecting most of objects with using simple classifiers.

The layers of a cascade in the original paper and in most of related works are represented by boosted sets of one-dimensional threshold rules. For this purpose, the other classifiers such as SVM can be used [4, 5]. In this paper, we propose a classifier based on a sequence of Gaussian mixture (GM) models with simplified decision rules. In order to reduce classification time, we use a greedy scheme of cascade training which results in a sequence of GMs that share the most of the mixture components. For this purpose, we perform an iterative component splitting using a partial EM algorithm [6, 7].

Statement of the problem

Let us consider the binary classification problem in which we need to distinguish "positive" objects belonging to a class ω_{pos} and "negative" objects belonging to a class ω_{neg}. Let us assume that these two classes are separated in feature space and the objects in ω_{pos} are distributed in a compact manner while the objects in ω_{neg} do not satisfy such distribution. In this case, a classification algorithm can be defined using a statistical distribution model of ω_{pos} added by a threshold rule.

Classification quality can be estimated by the following criterion:

$$Q = \beta \cdot FRR + (1-\beta) \cdot FAR, \quad (1)$$

where FRR (false reject rate) is a relative number of objects taken from ω_{pos} and assigned to ω_{neg}, FAR (false accept rate) is a relative number of objects taken from ω_{neg} and assigned to ω_{pos}, β is a task-dependent parameter that yields a relative penalty for false rejects with respect to false accepts.

Let $\gamma = N_{neg} / N_{pos}$ be a ratio of numbers of the objects in ω_{neg} and ω_{pos}, respectively, submitted to classification. Then the mean

[1] This work is supported by the RFBR, project 12-01-00920-a

time of object classification can be expressed as follows:

$$T = (t_{pos} + \gamma t_{neg}) / (1 + \gamma),$$

where t_{pos} and t_{neg} denote the mean time required for making a decision about object from ω_{pos} and ω_{neg} respectively.

Various classification algorithms can be compared using the dependencies of T on Q. Points of the curve $T(Q)$ correspond to different values of algorithm parameters such as statistical model complexity or threshold. The better algorithms provide lower values of T in the range of acceptable values of Q. In this work, we propose a classification method based on using a cascade of decision rules and we discuss the values of parameters β and γ that provide a computational profit of the proposed cascade decision rule classifier as against single decision rule classifiers.

Description of the method

Given feature space, a probability distribution density in ω_{pos} can be approximated by a Gaussian mixture (GM) using EM algorithm [6]. Let us denote the classification rule as $c(x)$ so that $c(x) = 1$ if an object x is assigned to ω_{pos} and $c(x) = 0$ otherwise. Then, the decision rule based on k-component GM model of the class ω_{pos} can be defined as follows:

$$c_k(x) = \max_{i=1}^{k} [-\ln \alpha_i + \frac{1}{2} \ln((2\pi)^d |H_i|) + \frac{1}{2}(x - \mu_i)^T H_i^{-1} (x - \mu_i) < \theta], \quad (3)$$

where α_i, μ_i, H_i are the weight, mean, and covariance matrix of the i-th component respectively, d is a dimensionality of the feature space, θ is a threshold value.

In (3), each condition defines an ellipsoid that corresponds to one of GM components. The object x is assigned to class ω_{pos} if it falls inside one of the ellipsoids.

Objects of the class ω_{neg} that are located far from the border of ω_{pos} can be classified using more simple decision rules than the objects located near the border. The main idea is to use a sequence of the classifiers of the form (3) that yields the cascade $\{c_m, m = 1 \ldots k\}$. Each object is submitted to the next classifier in the cascade if this object is assigned to the class ω_{pos} by the previous classifiers. For the cascade of classifiers, the decision rule is defined by

$$c_k^{seq}(x) = \min_{m=1}^{k} c_m(x) \quad . \quad (4)$$

Computational complexity of object classification depends on a number of the ellipsoids that are checked on falling the object into the ellipsoid or not. In the case (3), the number of these ellipsoids is equal to k for each object. In the case (4), the minimum number of the checked ellipsoids is 1 and the maximum number is $k(k+1)/2$. If the threshold θ is properly tuned, the most of the objects of the class ω_{pos} require $k(k+1)/2$ decisions. It means that a computational profit of the rule (4) as against the rule (3) is possible for sufficiently large values $\gamma = N_{neg} / N_{pos}$

Computational complexity can be reduced if the GMs in classifiers c_m share some components. It can be done by splitting one component into a couple of new components each time we increase the order of mixture from m to $(m+1)$ instead of building a $(m+1)$-component mixture with a new run of EM-algorithm.

Let X_{pos}^l is a training set of objects from ω_{pos}. We start with describing X_{pos}^l with a single Gaussian, then we split it into two new components and continue this process iteratively. At each step, we choose a component, splitting of which increases likelihood of X_{pos}^l most strongly. Some component i can be split into components j_1 and j_2 using partial EM-algorithm. The difference from the classical EM is that the

sum of child component weights $\alpha_{j_1} + \alpha_{j_2}$ equals not to 1 but to the parent component's weight α_i, and the sum of posterior probabilities of each object given child components $P(x \mid j_1) + P(x \mid j_2)$ equals to its posterior probability given parent component $P(x \mid i)$.

The procedure of iterative component splitting performed k times constructs a sequence of k Gaussian mixtures with increasing number of components. Applying the decision rule (3) to these mixtures we can obtain a set of classifiers $\{\tilde{c}_m\}$ $(m = 1 \ldots k)$ which can be combined into a cascade $\tilde{c}_k^{seq}$ using a rule similar to (4). If some object x is already classified by $\tilde{c}_m$ then in order to classify it with $\tilde{c}_{m+1}$ we only need to consider 2 additional ellipsoids and make a decision whether x belongs to them or not. So if we use $\tilde{c}_k^{seq}$ instead of c_k^{seq} then the maximum number of considered ellipsoids is reduced from $k(k+1)/2$ to $2k-1$.

Experiment

In order to test the proposed method, we have performed a number of tests on a model 2-dimensional data. We have modeled the distribution p_{pos} of class ω_{pos} with random 6-component Gaussian mixtures. Centers of GM components were normally distributed around $\mu_{cen} = 0$ with std. dev. $\sigma_{cen} = 6$, eigenvalues of component covariance matrices were normally distributed around $\mu_{\text{cov}} = 1.5$ with std. $\sigma_{\text{cov}} = 0.2$. The class ω_{neg} was modeled with a distribution p_{neg} which has a support in form of a circle and takes large values only when p_{pos} is low:

$$p_{neg}(x) = \\ = K \cdot \left[\|x\| < R\right] \Big/ \left(1 + q \cdot p_{pos}(x)\right), \quad (5)$$

where K is normalizing constant, R is the radius of the support, and q defines how strongly are the objects from ω_{pos} and ω_{neg} intermixed. We used values $R = 30$, $q = 1$.

We have generated p_{pos} and p_{neg} 50 times and performed 5 trials for each realization. In each trial, we randomly sampled $N_{pos}^l = 1000$ objects from p_{pos} and used them to train classifiers $c_1, c_2, \ldots, c_6, c_6^{seq}, \tilde{c}_6^{seq}$. Then we randomly sampled N_{pos}^t objects from p_{pos} and N_{neg}^t objects from p_{neg} to evaluate the performance of classifiers. We have examined two different cases: $\gamma = 2$ $(N_{pos}^t = 500, N_{neg}^t = 1000)$ and $\gamma = 10$ $(N_{pos}^t = 100, N_{neg}^t = 1000)$.

In each case, we calculated the error rate Q of the form (1) and the time T of the form (2) for the considered classifiers with different threshold values. The error Q was evaluated for three different values of β: $\beta = 0.1$, $\beta = 0.5$, $\beta = 0.9$. The time T was calculated in the arbitrary units given the assumption that the time needed to make decision about one object and one ellipsoid equals to 1. The values of Q and T were averaged over the realizations of model distributions and over the sampling trials.

The results are shown in the figure. For each classifier $c_1, \ldots, c_6$ we have minimized the error Q over all threshold values and evaluated the time T. Six obtained pairs (Q, T) are represented in the figure by the solid line. For the classifiers c_6^{seq} and $\tilde{c}_6^{seq}$ we have plotted T versus Q for different threshold values. These relations are represented by dashed and dotted lines respectively. Only the parts of graphs which correspond to the range of θ from maximum to the optimal value were plotted. In the case $\gamma = 2$, it makes no sense to use cascade classifiers because the same or lower error Q always can be obtained in less time by using one of the classifiers $c_1, \ldots, c_6$. In the case $\gamma = 10$, classifier c_6^{seq} is still unusable but there are ranges of Q values in which $\tilde{c}_6^{seq}$ is less time-consuming than $c_1, \ldots, c_6$ and provides the

same result. The classifier $\tilde{c}_6^{seq}$ should be used in the range of high Q values when β is low and in the range of intermediate Q values for higher values β.

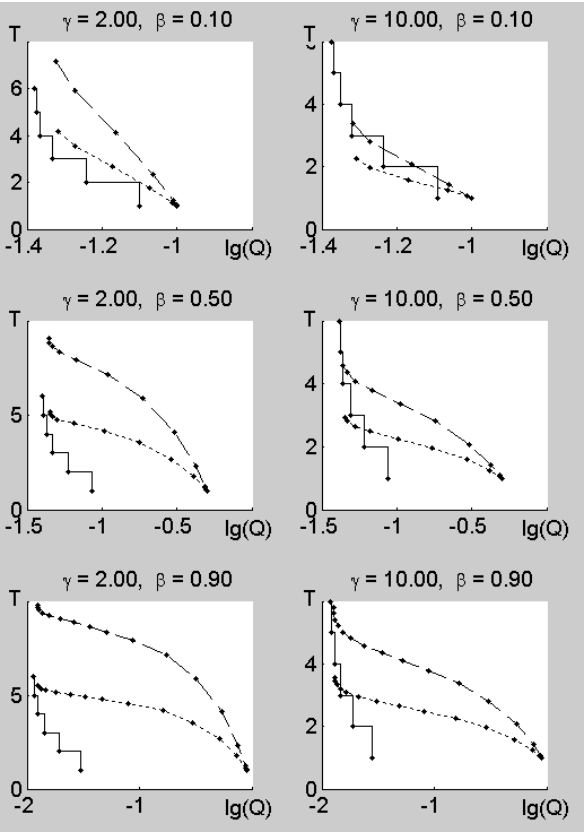


Fig.1. Classification time versus error rate $c_1 - c_6$ (solid line), c_6^{seq} (dashed line) and $\tilde{c}_6^{seq}$ (dotted line).

Conclusion

In this paper we have investigated the binary classification method based on the application of a cascade of decision rules which use GM-models of initial distribution with increasing number of components. We have proposed a way to decrease the computational cost of classification by using the sequence of Gaussian mixtures obtained from single Gaussian by recursive component splitting to construct individual classifiers in a cascade. We have compared the performance of the classifiers based on single Gaussian mixtures, on the cascade of independent mixtures and on the cascade of mixtures constructed by recursive component splitting. The experiments were performed on a model data in the cases of different ratios between numbers of positive and negative objects and different costs of false accepts and false rejects. In terms of relation between time and error rate, the proposed cascade GM-based classifier provides an advantage with respect to single GM-based classifiers when the number of negative objects is sufficiently larger than the number of positive ones. The range of error rates in which the proposed method provides a profit in computational time depends on a relative penalty of two error types.

In further work, we are planning to improve our method by tuning the threshold values independently for each layer of the cascade. Also, we are going to explore the decision trees based on the structured sets of Gaussians. In addition to building these structured sets by recursive GM component splitting, we are planning to apply methods based on merging GM components.

References

1. Quinlan, J. R., Induction of Decision Trees // Machine Learning – Kluwer Academic Publishers, 1986 – Vol.1, No.1 – P. 81-106.
2. W. Buntine. Learning classification trees // Statistics and Computing – Kluwer Academic Publishers, 1992 – Vol. 2, No. 2 – P. 63-73.
3. P. Viola, M. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features // Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition – 2001 – Vol. 1 – P. 511–518.
4. M. Rätsch, S. Romdhani, Th. Vetter. Efficient Face Detection by a Cascaded Support Vector Machine Using Haar-Like Features // Pattern Recognition. 26th DAGM Symposium, Tübingen, Germany, August 30 - September 1, 2004. Proceedings – 2004 – P. 62-70.
5. W-C. Cheng, D-M. Jhan. A cascade classifier using Adaboost algorithm and support vector machine for pedestrian detection // Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Anchorage, Alaska, USA, October 9-12, 2011 – P. 1430-1435.
6. Paalanen, P. Bayesian Classification using Gaussian Mixture Model and EM Estimation: Implementations and Comparisons // Lappeenranta University of Technology, Department of Information Technology, Lappeenranta, Finland, Tech. Rep. – 2004.
7. N. Ueda, NTT Commun. Sci. Labs., Kyoto, Japan, R. Nakano, Z. Ghahramani, G.E. Hinton. Split and merge EM algorithm for improving Gaussian mixture density estimates // Neural Networks for Signal Processing VIII, 1998. Proceedings of the 1998 IEEE Signal Processing Society Workshop – 1998 – P. 274 – 283.

THE USE OF EXTREMA DISTRIBUTION AS A FEATURE VECTOR FOR SPEECH PATTERNS RECOGNITION[1]

I.V. Ognev[2], A.I. Ognev[2], P.A. Paramonov[2,3], N.A. Sutula[2]

[2] National Research University "MPEI", ul. Krasnokazarmennaya 14, Moscow, 111250 Russia
e-mail [3] pa.pawka@gmail.com

In this paper we introduce a new time-domain approach to feature detection based on extrema distribution of speech signal. Finding extrema is a very simple procedure with linear complexity, what is faster than commonly used FFT-based methods. The availability of such distributions is demonstrated on Russian vowels recognition using Artificial Neural Networks (ANN).

Introduction

Feature detection is the first stage of speech recognition process which aims to construct so called feature vector – compact and yet as comprehensive as possible description of the input signal [1, 2]. Thus developers gain double benefit: firstly, omitting meaningless information increases recognition accuracy, secondly, this leads to significant data compression. The most commonly used methods of feature detection are based on spectral analysis of speech, such as Linear Predictive Coding (LPC) [3] or Mel-Frequency Cepstral Coefficients (MFCC) [4]. The most basic procedure of spectral approaches, FFT, has computational complexity $O(n \cdot \log(n))$.

In order to find faster way of deriving feature vector, different time-domain approaches were proposed, which exploit signal's zero-crossing rate [5]. In this paper we introduce a new time-domain method of feature detection based on extrema distribution of speech signal. Finding extrema is a very simple procedure with linear complexity. We demonstrated availability of such distributions on Russian vowel phonemes recognition using ANN.

Speech representation with extrema distribution

Consider a part of digitalized speech depicted in Fig. 1. Samples 1, 4, 5, 9, 10, and 11 are extrema. Let overall number of extrema be M. Any extremum e is limited by ADC resolution k:

$$e \in [-2^{k-1}, 2^{k-1} - 1] \qquad (1)$$

For instance, in this work all speech patterns were digitalized with $k = 16$, and thus $2^{k-1} = 32768$.

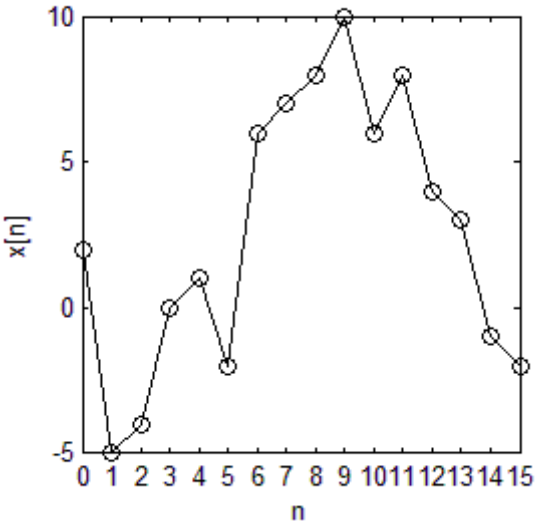


Fig. 1. Segment of digitalized speech with six extrema.

Then we define E_j as the j-th interval of extrema values, so that each interval has width $L = e_{j\max} - e_{j\min}$:

$$E_j = [e_{j\min}, e_{j\max}], j = 0, \ldots, \frac{2^k}{L} \qquad (2)$$

[1] This work was partially supported by the Council on grants of the President of the Russian Federation for State support of young scientists (grant no. MK-3281.2013.9)

Let D_j be the number of extrema that lay in E_j:

$$D_j = \sum_{i=0}^{M} \gamma(e_i, E_j), j = 0, \ldots, \frac{2^k}{L} \quad (3)$$

The γ function indicates whether extremum e_i lays in the j-th interval E_j:

$$\gamma(e_i, E_j) = \begin{cases} 1, e_i \in E_j \\ 0, e_i \notin E_j \end{cases} \quad (4)$$

Finally we can find a portion of extrema that lays in the j-th interval E_j:

$$P_j = \frac{D_j}{M}, j = 0, \ldots, \frac{2^k}{L} \quad (5)$$

Thus it is possible to find distribution vector $P = (P_0, \ldots, P_j, \ldots, P_{\frac{2^k}{L}})$ for each speech pattern, where P_j denotes probability of discrete random variable e laying in the interval E_j, so that $\sum_j P_j = 1$.

Smoothing low-amplitude fluctuations of speech using enveloping

While studying extrema distributions of different speech patterns, it has been noticed that smoothing low-amplitude fluctuations, such as depicted in Fig. 2 (samples from 4 to 10), increases recognition accuracy.

In order to do that we build two enveloping signals $e^{\max}$ and $e^{\min}$ which consists only of maxima and minima respectively. Their length is the same as for original signal; let it be N. Obviously, $M < N$, so between extrema we find sample values using linear interpolation. Then we build the average signal s, where every sample s_i lays between maxima and minima of the original signal:

$$s_i = \frac{e_i^{\max} + e_i^{\min}}{2}, i = 0, \ldots, N \quad (6)$$

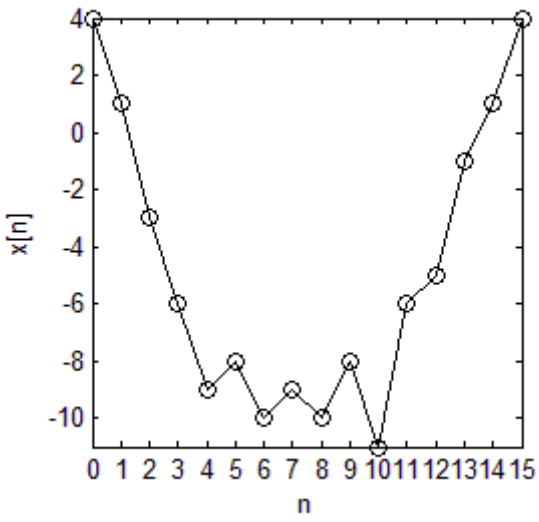


Fig. 2. Segment of speech with low-amplitude fluctuations.

Such approach smoothes undesirable low-amplitude fluctuations of speech (Fig. 3), what leads to both data compression (since the number of extrema decreases significantly) and accuracy enhancement.

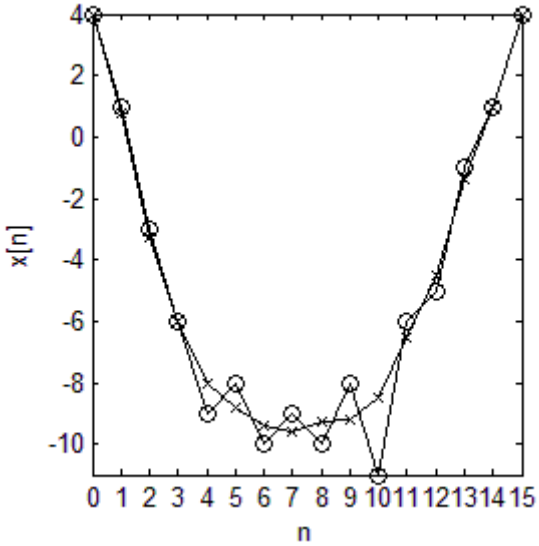


Fig. 3. Smoothed signal s, depicted with crosses.

Extrema distributions of Russian vowels

We studied extrema distributions for six Russian vowels, namely /a/, /и/, /о/, /у/, /ы/, /э/. Our database included 200 examples of each phoneme. Every pattern was preprocessed as follows: firstly, spectral cleaning from microphone noise, secondly, normalization. Figure 4 contains distributions built using enveloping for 100 examples of all six vowels.

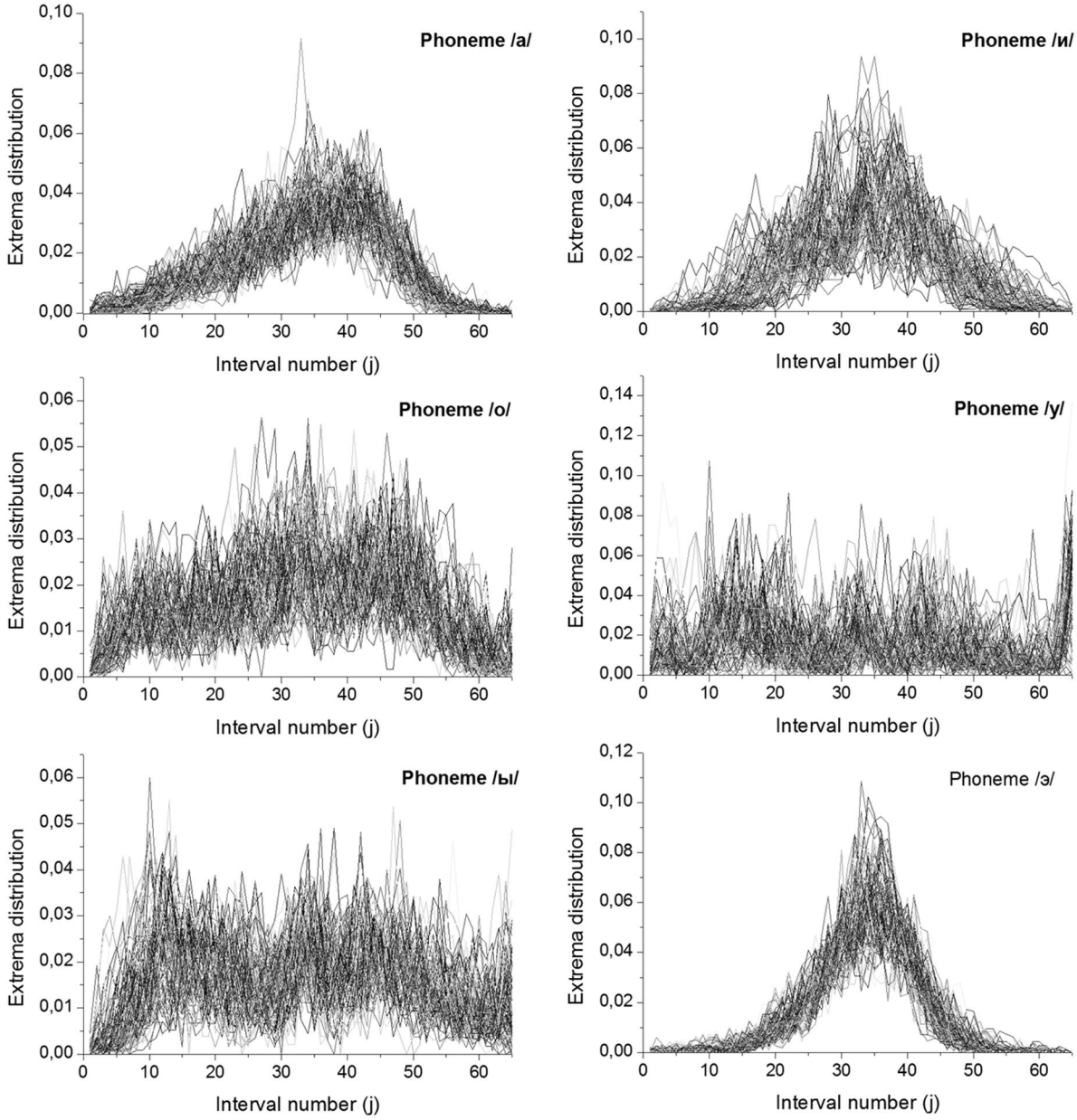


Fig. 4.Extrema distributions for six Russian vowels.

Russian vowels recognition using ANN

In order to demonstrate the use of extrema distribution, we used multilayer feed forward ANN with 66 input neurons, 33 hidden neurons and 6 output neurons. 100 examples of each phoneme were used to train ANN, and the other 100 examples were used on recognition stage. The result of experiment was evaluated as recognition accuracy: a portion of correct ANN outputs. We compared both straightforward approach to detecting extrema and enveloping method.

Table 1. Recognition accuracy for six Russian vowels

Phoneme	Recognition accuracy, %	
	Without enveloping	With enveloping
/а/	53	72.0
/и/	59	71.0
/о/	64	86.0
/у/	87	85.0
/ы/	78	63.0
/э/	78	89.0
Average	69.8	77.7

Average accuracy for enveloping method reached 77.7%. The detailed results for each phoneme are listed in Table 1

Conclusion

A new time-domain approach to feature detection based on extrema distribution of speech signal has been proposed. Availability of such distributions has been studied on Russian vowels recognition using ANN. The proposed approach attracts with linear computational complexity. As future work, more robust extrema detection techniques will be developed, which omit meaningless data and enhance recognition accuracy.

References

1. L. Rabiner, B.-H. Juang. Fundamentals of speech recognition. – 1993.
2. X. Huang, A. Acero. Spoken language processing: a guide to theory, algorithm, and system development. – 2001.
3. J. Benesty, M. Mohan Sondhi, Y. Huang. Springer Handbook of Speech Processing. – 2008.
4. I. McLoughlin. Applied Speech and Audio Processing. – 2009.
5. R.K. Sunil Kumar, V.L. Lajish. Phoneme recognition using zerocrossing interval distribution of speech patterns and ANN // International Journal of Speech Technology. – 2013. – Vol. 16, No.1. – P. 125-131

COMPUTATIONAL INTELLIGENCE AND GENERALIZED CATALAN NUMBERS IN PATTERN RECOGNITION PROBLEMS[1]

A. Reznik[2,3], V. Efimov[2,4], A. Soloview[2,5], A. Torgov[2,6]

[2] Institute of Automation and Electrometry, Siberian Branch, Russian Academy of Sciences, pr. Akademika Koptyuga 1, Novosibirsk, 630090 Russia,

[3] reznik@iae.nsk.su, [4] efimov_vit@mail.ru, [5] solowey@rambler.ru, [6] torgov@iae.nsk.su

An original approach to solving difficult probabilistic problems arising in studying the readout of random discrete fields is proposed (there are no exact analytical solutions at the moment). Several partial-solution algorithms for direct, iterative, and combinatorial-recursive calculations are presented (these solutions are further used to search for the common closed analytical regularities). The huge volume of necessary calculations forced us to formalize algorithms for computer systems. Calculations helped us to establish (and to prove later) a number of new earlier unknown probabilistic formulas responsible for random division of an interval.

Introduction

Our researches on the reliability of readout of random discrete-point fields led us to the following very simple (in definition) probabilistic problem:

«Let n points $x_1,x_2, \ldots ,x_n$ be randomly dropped on an interval $(0,1)$, i.e., there are n independent tests of a random variable uniformly distributed in the interval $(0,1)$. It is required to determine the probability $P_{n,k}(\varepsilon)$ of an event that there is no subinterval $\Omega_\varepsilon \subset (0,1)$ of length ε containing more than k points».

The apparent simplicity of this problem is illusory, and its analytic solution is known [12-13] only for k=1:

$$P_{n,1}(\varepsilon) = (1-(n-1)\varepsilon)^n, \quad (0 \le \varepsilon \le 1/(n-1)). \tag{1}$$

It should be noted that many problems that involve random division of an interval [10] are simple in statement, but have no exact analytical solution. It is an example we have here.

One way to find solution (1) is to represent the probability $P_{n,1}(\varepsilon)$ as a repeated integral

$$P_{n,1}(\varepsilon) = n! \int_{(n-1)\varepsilon}^{1} dx_n \left\{ \int_{(n-2)\varepsilon}^{x_n-\varepsilon} dx_{n-1} \cdots \left\{ \int_{2\varepsilon}^{x_4-\varepsilon} dx_3 \left\{ \int_{\varepsilon}^{x_3-\varepsilon} dx_2 \left\{ \int_{0}^{x_2-\varepsilon} dx_1 \right\} \right\} \right\} \right\}. \tag{2}$$

Successive integration of Eq. (2) with respect to the variables $x_1,x_2,..,x_n$ yields relation (1). Unfortunately, for k>1, the probability $P_{n,k}(\varepsilon)$ cannot be reduced to a single repeated integral. Today there is no exact analytic solution of this problem even for k=2. The present paper describes possible approaches to solving the above-discussed problem.

Computer-aided analytical calculation of probabilistic formulas

In general, the probability $P_{n,k}(\varepsilon)$ can be represented in the form

$$P_{n,k}(\varepsilon) = n! \int \ldots \int_{D_{n,k}(\varepsilon)} dx_1 \ldots dx_n,$$

where the domain of integration $D_{n,k}(\varepsilon) \subset R^n$ is described by a system of linear inequalities

[1] This work was supported by the Russian Foundation for Basic Research (Project No. 13-01-00361), by the Presidium of the Russian Academy of Sciences (Project No. 11/2012), and by the Siberian Branch of the Russian Academy of Sciences (Integration Project SB RAS and NASB No. 16/2012).

$$\begin{cases} 0 < x_1 < x_2 < \ldots < x_{n-1} < x_n < 1, \\ x_{k+1} - x_1 > \varepsilon, \\ x_{k+2} - x_2 > \varepsilon, \\ \vdots \\ x_n - x_{n-k} > \varepsilon. \end{cases} \quad (4)$$

Further, integral (3) can be written in equivalent form:

$$\begin{aligned} P_{n,k}(\varepsilon) = n! \int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} 1[x_1] 1[x_2] \ldots \\ \ldots 1[x_n - x_{n-1}] 1[1 - x_n] 1[x_{k+1} - x_1 - \varepsilon] \\ \times 1[x_{k+2} - x_2 - \varepsilon] \ldots 1[x_n - x_{n-k} - \varepsilon] dx_1 \ldots dx_n, \end{aligned} \quad (5)$$

where $1[z] = \begin{cases} 0, & z \le 0, \\ 1, & z > 0. \end{cases}$ Then, the n-dimensional integral (5) is converted to a set of repeated integrals where integration limits are already set by means of successive application of the relation

$$\begin{aligned} \left(\prod_{j=1}^{l} 1[x_r - \alpha_j] \right) \left(\prod_{i=1}^{m} 1[\beta_i - x_r] \right) = \\ = \sum_{j=1}^{l} \sum_{i=1}^{m} 1[x_r - \alpha_j] 1[\beta_i - x_r] 1[\beta_i - \alpha_j] \left(\prod_{\substack{q=1 \\ q \ne j}}^{l} 1[\alpha_j - \alpha_q] \right) \left(\prod_{\substack{s=1 \\ s \ne i}}^{m} 1[\beta_s - \beta_i] \right). \end{aligned} \quad (6)$$

(The form of Eq. (6) means that the expressions α and β do not contain the variable x_r). The algorithm (3-6) allows us to proceed a constructive calculation of the formulas for $P_{n,k}(\varepsilon)$ for fixed values of n and k. The main difficulty of application of this procedure lies in the fact that it is practically impossible to perform all necessary calculations manually (settling the limits of integration, verifying the correctness of all intermediate systems of inequalities, and direct calculating of the repeated integrals for n>4). For these reasons, we created a program package based on algorithm (3)-(6) and performing all analytical calculations [6].

We also proposed another combinatorial approach (using an absolutely different mathematical technique) to find particular solutions of the initial problem. By analogy with the known solution (1) valid for k=1, we tried to find a general solution $P_{n,2}(\varepsilon)$ for k=2. We compose a recursive algorithm where the formulas for $P_{n,2}(\varepsilon)$ as functions of the continuous argument ε are obtained from a discrete-combinatorial scheme. We used the following discrete-combinatorial model. The interval (0,1) is considered to be divided into r equal quantums. Random throwing of n points on the interval (0,1) is interpreted as random throwing of n indistinguishable balls into r boxes. A set of l adjacent quantums serves as an analog of a subinterval of length ε. The outcome of throwing when none of such l-subintervals inside the initial r-interval (0,1) has more than two points is considered as "successful," and the ratio of the total number of "successful throws" to the overall number of possible outcomes of throwing is taken as a discrete-combinatorial analog for the probability $P_{n,2}(\varepsilon)$. Using that approach we obtained new relations for k=2.

Our third software package is based on multiple differentiation of the initial integral (3) with respect to the parameter ε and further reconstruction of the formulas for $P_{n,k}(\varepsilon)$ with the values of all derivatives $\frac{d^{(j)} P(\varepsilon)}{d\varepsilon^{(j)}}$, ($j = 0, 1, \ldots, n$) at zero (i.e., at ε=0). The main advantage of this algorithm is the fact that its implementation allows us to replace labor-consuming procedures of determining the integration limits and subsequent multidimensional integration by elementary operations of substitution and replacement of variables.

Using three above-mentioned program systems, we evaluated the formulas for $P_{n,k}(\varepsilon)$ for particular values of n and k (k<n) up to n=14 and for all ranges of variation of the parameter ε. These computer analytical calculations helped us to establish and then to prove strictly new probabilistic regularities earlier unknown relating to a random division of an interval.

Proof of "computer" formulas with the use of the Catalan numbers

The analysis of the formulas for $P_{n,k}(\varepsilon)$ calculated on a computer allows us to establish a number of new previously unknown analytical dependences. In particular, for even values of n=2m and k=2, we established the formula

$$P_{2m,2}(\varepsilon)=\frac{1}{m}C_{2m}^{m-1}(1-(m-1)\varepsilon)^{2m},$$

which is valid if $1/m<\varepsilon<1/(m-1)$. The coefficients $(1/m)C_{2m}^{m-1}$ in Eq. (7) are the classical Catalan numbers known from Leonard Euler's works, but they are still of interest [1,4,9] because they form the basis of enumerative combinatorics [14]. It is curious to note that relation (7) was "prompted" by a computer and was published as a scientific hypothesis more than thirty years ago [5], and a strict mathematical proof of this formula was obtained rather recently [7]. Thereby, we realized in practice the classical advice of John von Neumann: if you cannot find a straightforward solution of a difficult scientific problem, try to perform laborious auxiliary calculations on a computer. If you are lucky, these auxiliary computer calculations can "prompt" you the right answer, which you will prove further.

Recently we managed to prove that the probability $P_{n,k(\varepsilon)}$ for k=2 and odd values of n=2m+1 is represented as

$$\begin{aligned}P_{2m+1,2}(\varepsilon)&=C_{2m+1}^{m+1}(1-m\varepsilon)^{m+1}(1-(m-1)\varepsilon)^{m}-\\&-2C_{2m+1}^{m+2}(1-m\varepsilon)^{m+2}(1-(m-1)\varepsilon)^{m-1}+\\&+C_{2m+1}^{m+3}(1-m\varepsilon)^{m+3}(1-(m-1)\varepsilon)^{m-2},\end{aligned}\qquad(8)$$

if 1/(m+1)<ε<1/m.

It turned out that it is much more difficult to find and mathematically substantiate relation (8) than to prove formula (7). Thus, one stage of this proof forced us at first to introduce a new concept of "three-dimensional generalized Catalan numbers" and next to calculate their explicit form. In our investigations, these numbers appeared when we had to find the total number of specific permutations of elements of three subsets, with each of these subsets being represented as a ranked sequence of uniformly distributed random variables.

If we put aside the researches dealing with random division of an interval, then the problem leading to three-dimensional generalized Catalan numbers can be formulated in the following most transparent form: "It is necessary to find $Q_{l,m,n}$, which is the exact number of different words of length $(l+m+n)$ that can be formed from l letters "a", m letters "b", and n letters "c" with two simultaneous conditions: 1) the number of the letters "b" never exceeds the number of the letters "a" if the word is viewed from left to right; 2) the number of the letters "c" never exceeds the number of the letters "a" if the word is viewed from right to left".

Having reduced this problem with three-letter words to a geometrical problem of searching for paths on a three-dimensional discrete lattice, we managed to show that

$$\begin{aligned}Q_{l,m,n}&=\frac{(l+m+n)!}{l!m!n!}\times\\&\times\frac{(l+1)(l+2)-(m+n)(l+2)+mn}{(l+1)(l+2)}.\end{aligned}\qquad(9)$$

A detailed proof of this equality can be found in [8].

We named the numbers $Q_{l,m,n}$ as "three-dimensional generalized Catalan numbers" bearing in mind that these numbers extend the traditional Catalan sequence known from many applications (see e.g. [1-2]) and obtained from Eq. (9) at n=0 and l=m. Equation (9) is not only useful in solving applied probabilistic and statistical problems, but also has independent theoretical interest.

Conclusion

We proposed several program systems for labor-consuming analytical calculations to find exact analytical solutions of the challenging probabilistic problems arising in studying reliability of the readout of random discrete-point structures. It became possible to calculate a wide set of particular solutions with developed software packages. The further analysis of the obtained "computer" formulas allowed us to establish (and later on to prove strictly) a number of new earlier unknown analytical equalities. These formulas are necessary to solve many problems dealing with random division of an interval.

One more distinctive feature of the presented work is introduction of a new concept of

"three-dimensional generalized Catalan numbers" for the investigations to be successful. Having this as a separate subtask, we managed to find a simple and transparent interpretation of this generalization of the classical Catalan sequence (as a solution of a problem with special three-letter words). We found the explicit form of the three-dimensional generalized Catalan numbers by means of reducing this new "linguistic" subtask to a geometrical problem of searching for special paths on a discrete three-dimensional lattice under certain constraints. The subtask with the three-dimensional generalized Catalan numbers also has an independent theoretical significance.

References

1. Chamberland M, French C (2007) Generalized Catalan Numbers and Generalized Hankel Transformations. Journal of Integer Sequences 10: 1-7.
2. Gardner M (1976) Mathematical Games, Catalan numbers: an integer sequence that materializes in unexpected places. Scientific American pp 120-125.
3. Hilton P, Pedersen J (1991) Catalan numbers, their generalization, and their uses. Math Int 13: 64–75.
4. Koc C, Guloglu I, Esin S (2010) Generalized Catalan numbers, sequences and polynomials. Turk J Math 34: 441–449.
5. Reznik A (1981) Computer modeling of continuous readout of random discrete-structural images. Avtometriya 6 pp 3-6. (in russian)
6. Reznik A, Efimov V (2003) Analytical Computer Calculations in Analysis of Discrete-Point Images. Pattern Recognition and Image Analysis 10 (1) pp 158-160.
7. Reznik A, Efimov V, Solov'ev A (2011) Computer-analytical calculation of the probability characteristics of readout of random point images Optoelectronics, Instrumentation and Data Processing 47 (1) pp 7-11.
8. Reznik A, Efimov V, Solov'ev A, Torgov A (2011) Generalized Catalan Numbers in Problems of Processing of Random Discrete Images. Optoelectronics, Instrumentation and Data Processing 47 (6) pp 11-15.
9. Stanimirovic S, Stanimirovic P, Ilic A (2012) Ballot matrix as Catalan matrix power and related identities. Discrete Applied Mathematics 160 (3) pp 344-351.
10. David HA and Nagaraja HN (2003) Order Statistics. John Wiley New York.
11. Feller W (1968) An introduction to probability theory and its applications (1) 3ed Wiley.
12. Parzen E (1960) Modern Probability Theory and Its Applications. John Wiley and Sons Inc New York-London.
13. Wilks SS (1962) Mathematical Statistics. J Wiley and Sons New York-London.
14. Stanley RP (1999) Enumerative Combinatorics. vol 2 Cambridge Studies in Advanced Mathematics 62 Cambridge University Press.

STATISTICAL ESTIMATORS OF MULTIPLE-SITE SIMILARITY MEASURES[1]

B.I. Semkin[2,3,4], M.V. Gorshkov [2,5]

[2]Pacific Institute of Geography, Vladivostok, ul. Radio 7, 690041, Russia
[3]Far Eastern Federal University, Vladivostok, ul. Sukhanova 8, 690001, Russia
e-mail: [4] zuevyf@poi.dvo.ru, [5] gorshkov_mv@mail.ru

Aspects of the use of weighted arithmetic relations of values and confidence intervals for statistical estimators fractional-linear multiple-site measures of similarity for type of $K(T,C_\Delta)$ are considered.

Introduction

In the elementary theory of similarity [17] can be identified two groups of measures of similarity – deterministic and stochastic. By deterministic measures are based on finite sets or descriptive sets, and stochastic measures are based on probability and categorized data. For the first group to determine the statistical estimators does not make sense [2, 5, 13], but for the second group to determine the statistical estimators successfully [1, 11, 16]. However, methods for the statistical estimators of weighted arithmetic mean have been developed [6, 10, 14, 15] and it can be used for the statistical estimators of some multiple-site (three and more sites) measures of similarity. The estimators can be represented as the weighted arithmetic mean of two-site measures of similarity. This problem can be solved by means of fractional-linear multiple-site measures of similarity of type $K(T,C_\Delta)$ [18].

Multiple-site measures of similarity of type $K(T,C_\Delta)$ and representation as weighted arithmetic mean of two-site measures of similarity

Consider a multiple-site measure of similarity $K_{(0)}$ of subtype $K_{(\tau)}$ of type $K(T,C_\Delta)$:

$$K_{(0)}=\frac{2C_\Delta}{(n-1)T},$$

where $T=m(a^{(1)})+\ldots+m(a^{(n)})$;
$C_\Delta=\sum_{i<j}^{n}m(a^{(i)}\wedge a^{(j)})$; $a^{(i)}$ and $a^{(n)}$ – descriptive sets; $i=1,\ldots,n$; $j=1,\ldots,n$.

Let's consider the derivation of the multiple-site measure of similarity $K_{(0)}$. Suppose we have $N=\frac{n(n-1)}{2}$ two-site measures of similarity $K_0(a^{(i)},a^{(j)})$, $i<j$, $i=1,\ldots,n$; $j=1,\ldots,n$; $\frac{2m(a^{(1)}\wedge a^{(2)})}{m(a^{(1)})+m(a^{(2)})}$, $\frac{2m(a^{(1)}\wedge a^{(3)})}{m(a^{(1)})+m(a^{(3)})}$, ..., $\frac{2m(a^{(1)}\wedge a^{(n)})}{m(a^{(1)})+m(a^{(n)})}$, ..., $\frac{2m(a^{(n-1)}\wedge a^{(n)})}{m(a^{(n-1)})+m(a^{(n)})}$.

We denote consequentially this relations as $\frac{x_i}{y_i}$, $i=1,\ldots,\frac{n(n-1)}{2}$. Where, $x_1=2m(a^{(1)}\wedge a^{(2)}),\ldots,x_n=2m(a^{(n-1)}\wedge a^{(n)})$, $y_1=m(a^{(1)})+m(a^{(2)}),\ldots,y_n=m(a^{(n-1)})+m(a^{(n)})$, $x_N=2m(a^{(n-1)}\wedge a^{(n)})$, $y_N=m(a^{(n-1)}\wedge a^{(n)})$, $\bar{x}=\frac{1}{N}\sum_{i=1}^{N}x_i=\frac{4C_\Delta}{n(n-1)}$, $\bar{y}=\frac{1}{N}\sum_{i=1}^{N}y_i=\frac{2T}{n}$.
Then $\bar{t}=\frac{\bar{x}}{\bar{y}}=\frac{4C_\Delta}{n(n-1)}\Big/\frac{2T}{n}=\frac{2C_\Delta}{(n-1)T}$.

[1] This work was supported by the grant, FEB RAS, project KPFI 12-06-023 (2012-2014).

Estimators for $K_{(0)}$

We define the standard deviation for $\bar{t}$ and the 95% confidence interval, usung the formulas of the Schwartz's [12], Crane's and Lemoine's works [3]. The method of calculation of the confidence interval for the average of the ratio in the classic version consists of the following steps:

1. Statistics are calculated:

- $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $\bar{t} = \frac{\bar{x}}{\bar{y}}$;
- $S_{11} = \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{1}{n(n-1)}\left(\sum_{i=1}^{n} x_i\right)^2$;
- $S_{22} = \frac{1}{n-1}\sum_{i=1}^{n} y_i^2 - \frac{1}{n(n-1)}\left(\sum_{i=1}^{n} y_i\right)^2$;
- $S_{12} = \frac{1}{n-1}\sum_{i=1}^{n} x_i y_i - \frac{1}{n(n-1)}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i\right)$;
- $S = S_{11} - 2\bar{t}S_{12} + 2\bar{t}^2 S_{22}$.

2. Confidence interval (use 95% confidence level) are calculated:

$$\bar{t} \pm \frac{1.96}{\bar{y}\, n^{1/2}}.$$

The final results of the method are the same in both works. However, the Crane-Lemoine's method are given too large values for the variances S11 and S22 and covariance S12.

These formulas joined us in the general formula to simplify the calculations and to liquidate of the high values of the variance with the formulas of Crane- Lemoine [6, 10, 14]:

$$S_{\bar{t}} = \frac{1}{\bar{y}\sqrt{N}} S_d \quad ; \quad S_d = \sqrt{\frac{\sum_{i=1}^{N}(x_i - y_i\,\bar{t})^2}{N-1}}.$$

$S_{\bar{t}}$ and S_d can be calculate by the above formulas or after substituting the values of x_i, y_i, $\bar{y}$, $N = \frac{n(n-1)}{2}$, $\bar{t} = K_{(0)}$ by the formula:

$$S_{K_{(0)}} = \frac{1}{T}\sqrt{\frac{n}{2(n-1)}} S_d,$$

$$S_d = \sqrt{\sum_{i=1}^{N} \frac{\left[2m\left(a^{(i)} \wedge a^{(j)}\right) - K_{(0)}\left(m\left(a^{(i)}\right) + m\left(a^{(j)}\right)\right)\right]^2}{N-1}}.$$

Also S_d can be defined as follows:

$$S_d = \sqrt{2\sum_{i<j}^{N} \frac{\left[2m\left(a^{(i)} \wedge a^{(j)}\right) - K_{(0)}\left(m\left(a^{(i)}\right) + m\left(a^{(j)}\right)\right)\right]^2}{n^2 - n - 2}}.$$

The standard deviation for the class of the equivalence $K_{(\tau)}$ can be defined by the formula:

$$S_{K_{(\tau)}} = \frac{(1+\tau)\, S_{K_{(0)}}}{\left(1 + \tau - \tau\, K_{(0)}\right)^2},\ \tau > -1.$$

The last formula are derived from the ratio:

$$S^2_{K_{(\tau)}} = \left(\frac{dK_{(\tau)}}{dK_{(0)}}\right)^2 S^2_{K_{(0)}} = \frac{(1+\tau)^2}{\left(1 + \tau - \tau\, K_{(0)}\right)^4} S^2_{K_{(0)}}.$$

Example of calculation

For the calculation of a multiple-site measure of similarity and confidence intervals take data from four partial floras subnival and nival zones of the Central Caucasus: Kazbegi (1) Mamison Pass (2), Elbrus (3) and Upper Svaneti (4) [4]. The main source of information for the calculation is the intersection matrix (matrix of absolute measures of similarity, see table). According to this matrix can be calculated as a two-site measure of similarity, and multiple-site measure of similarity [7-9].

The intersection matrix is symmetric, so it is enough to have $\frac{n(n+1)}{2}$ of its values, i.e. main diagonal of the matrix and the values of above-diagonal (or below-diagonal) elements [7-9].

Table 1. The intersection matrix of the species lists for the subnival and nival zones of the Central Caucasus (denotation in the text)

	1	2	3	4
1	**84**	68	78	69
2	68	**73**	71	65
3	78	71	**85**	73
4	69	65	73	**74**

For the calculation of a measures of similarity necessary to use non-difference lists of the number of species.

We make the calculation of the measure of non-different lists on the formula [7]:

$$R(a_1,\ldots,a_n)=\sum_{i=1}^{n}\min\left(p_i,\frac{1}{n}\right),$$

where $a_i>0$, $i=1,\ldots,n$, $p_i=\dfrac{a_i}{a_1+\ldots+a_n}$, $\sum_{i=1}^{n}p_i=1$. In this case:

$$R(84,73,85,74)=\min(0.265;0.250)+$$
$$+\min(0.231;0.250)+\min(0.260;0.250)+$$
$$+\min(0.234;0.250)=0.250+0.231+0.250+$$
$$+0.234=0.965.$$

The degree of non-difference of the lists is high, that lets you find on this matrix, the measures of similarity. We define the following parameters of a series of four lists:

$$C_\Delta=68+78+69+71+65+73=424,$$

$$T=84+73+85+74=316.$$

Then $K_{(0)}=\dfrac{2\times424}{3\times316}\approx0.895$.

Calculate the standard deviation $\delta_{K_{(0)}}$:

$$S_d^2=\frac{D}{5},\ D=[2\times68-0.895\times(84+73)]^2+$$
$$+[2\times78-0.895(84+85)]^2+$$
$$+[2\times78-0.895(84+85)]^2+$$
$$+[2\times69-0.895(84+74)]^2+$$
$$+[2\times71-0.895(73+85)]^2+$$
$$+[2\times65-0.895(73+74)]^2+$$
$$+[2\times73-0.895(85+74)]^2=71.0212;$$
$$S_d^2=14.2042;\ S_d\approx3.7688;$$
$$\frac{1}{T}\sqrt{\frac{n}{2(n-1)}}=\frac{1}{316}\sqrt{\frac{2}{3}}\approx0.0026.$$
$$S_{K_{(0)}}=0.0026\times3.7688=0.0097.$$

Confidence interval (95% confidence level):

$$0.8945\pm1.96\times0.0097=0.8945\pm0.0191.$$

We define the arithmetic mean of two-site measures of similarity K_0 as $M_0=\dfrac{1}{C_n^2}K_0\left(a^{(i)},a^{(j)}\right)$ ($M_0=0.8928$). This measure is similar to the measure $K_{(0)}=0.8945$. However, estimators of the measure M_0 is not defined.

$$K_{(1)}=\frac{K_{(0)}}{2-K_{(0)}}=\frac{0.8945}{2-0.8945}\approx0.8091.$$

$$S_{K_{(1)}}=\frac{2\delta_{K_{(0)}}}{\left(2-K_{(0)}\right)^2}=\frac{2\times0.0097}{(2-0.8945)^2}\approx0.0159.$$

Confidence interval (95% confidence level):

$$0.8091\pm1.96\times0.0159=0.8091\pm0.0312.$$

Conclusion

The estimators for the equivalent multiple-site measure of similarity $K_{(1)}$ and $K_{(0)}$ provide for the first time. It should be noted that the variance of the measure $K_{(0)}$ is less than $K_{(1)}$, at that the values of measure $K_{(1)}$ much less than $K_{(0)}$. Probably, the selection the required multiple-site measure of similarity can be justified by means of estimation of variance or coefficient of variation.

References

1. Basargin D.D., Gorovoi P.G., Semkin B.I. Taxonomic characterization of the sizes if the mericarpium at Hogweed Heracleum L. The Far East and North America // Russian Bot. J. 1978. V.63. №12. P. 1766-1774. [in Russian]
2. Vasilevich V.I. Statistical methods in geobotany. – Leningrad: Nauka, 1969. – 232 p. [in Russian]
3. Crane M.A., Lemoine A.J. An Introduction to the Regenerative Method for Simulation Analysis. New York: Springer-Verlag, 111 p.
4. Nakhutsrishvili G.S., Gamtsemlidze Z.G. Plant life in the extreme conditions of high mountains: the case of the Central Caucasus, Leningrad: Nauka, 1984. – 123 p. [in Russian]
5. Pesenko Y.A. The principles and methods of quantitative analysis of faunal studies. – Moscow: Nauka, 1982. 287 p. [in Russian]
6. Semkin B.I., Gorshkov M.V., Varchenko L.I. On changes of water content in annual shoots of coniferous plants in the temperate climate zone // Siberian Journal of Ecology. 2008. №4. V.15. P. 537-544. [in Russian]

7. Semkin B.I., Klochkova N.G., Gusarova I.S., Gorshkov M.V. Discretness and continuity of macrophyte algae flora of the Far Eastern seas of Russia. I. General species composition // Izv. TINRO. 2010. V.160. P. 57-70. [in Russian]
8. Semkin B.I., Klochkova N.G., Gusarova I.S., Gorshkov M.V. Discretness and continuity of macrophyte algae flora of the Far Eastern seas of Russia. II. Green (Chlorophyta), Brown (Phaeophyta), Rdd (Rhodophyta) // Izv. TINRO. 2010. V.162. P. 104-112. [in Russian]
9. Semkin B.I., Klochkova N.G., Gusarova I.S., Gorshkov M.V. Discretness and continuity of macrophyte algae flora of the Far Eastern seas of Russia. III. Taxonomic spectrum // Izv. TINRO. 2010. V.163. P. 217-227. [in Russian]
10. Semkin B.I., Gusarova I.S., Gorshkov M.V. On the invariance of the averages (for some of the morphological features of the thalli of Japanese kelp (Laminaria japonica Aresch.) from the subtidal zone of northern Primorye) // Izv. TINRO. 2012. V. 171. P. 313-320. [in Russian]
11. Semkin B.I., Petropavlovsky B.S., Kislov D.E., Brizhataya A.A. On the use of bioinformatic technologies in the environmental mapping // The forests of the Russian Far East: Monitoring of forest dynamics in the Russian Far East. – Vladivostok: LAINS, 2012. P. 201-204. [in Russian]
12. Schwartz G. Selective method. Guide on the application of statistical methods of estimation. – Moscow: Statistica, 1978. – 213 p. [in Russian]
13. Shmidt V.M. Statistical methods in the comparative floristics. – Leningrad: LSU, 1980. – 176 p.
14. Semkin B.I., Gorshkov M.V., Varchenko L.I. Variations of water content in annual shoots of coniferous trees in temperate zone // Contemporary Problems of Ecology. 2008. V.1. №4. P. 414-419.
15. Semkin B.I., Petropavlovsky B.S., Kislov D.E., Zuev Y.F. A study on the use bioinformatic techniques for environmental mapping // Pattern Recognition and Image Understanding. 8th Open German-Russian Workshop. – Nizhny Novgorod: 2011. P. 262-264.
16. Semkin B.I., Petropavlovsky B.S., Kislov D.E., Zuev Y.F. A study on the use bioinformatic techniques for environmental mapping // Pattern Recognition and Image Analysis. 2012. [in print].
17. Semkin B.I. Elementary theory of similarities and its use in biology and geography // Pattern Recognition and Image Analysis. 2012. V.22. № 1. P. 92-98.
18. Semkin B.I. The elementary theory of similarity and its use in biology and geography. Multiple-site measures of similarity and dissimilarity // Pattern Recognition and Image Analysis. 2013 [in print].

THE ELEMENTARY THEORY OF SIMILARITY AND ITS APPLICATION IN BIOLOGY AND GEOGRAPHY. MULTIPLE-SITE OF THE SIMILARITY AND DISSIMILARITY MEASURES

B.I. Semkin[1,2]

[1]Pacific Institute of Geography, Vladivostok, ul. Radio 7, 690041, Russia
[2]Far Eastern Federal University, Vladivostok, ul. Sukhanova 8, 690001, Russia
e-mail: zuevyf@poi.dvo.ru

The system of multiple-site similarity measures axioms, certain on a series of finite descriptive sets is considered. The concept of equivalence of multiple-site similarity is offered. Classes of equivalence of multiple-site similarity measures are allocated. For fractional-linear multiple-site similarity measures types and subtypes are allocated. The new concept of a coordination multiple-site similarity measures equivalence classes on unequally in power sequences of descriptive sets series. Illustrative examples of multiple-site similarity measures are taken from biology and geography.

Introduction

Earlier [8, 9] we considered mainly two-site measures of similarity and dissimilarity of descriptive sets. In work [9] the system of axioms of multiple-site similarity measures is resulted. It has appeared, that the multiple-site measures similarity theory is much more complex than the two-site similarity measures theory and demands special consideration.

The system of axioms of multiple-site similarity measures in a primary kind has been offered by the author in 1972 [2]. In a number of works [3-5, 9] this system has been added by new axioms. The review on use of multiple-site similarity measures in biology and geography is presented in work [6].

System of axioms of multiple-site similarity measures

Multiple-site measure of similarity $K\left(a^{(i)},\ldots,a^{(n)}\right)$ for a series from n descriptive sets $\left\{a^{(i)}\right\}$, $i=1,\ldots,n$; $n\geq 3$ is defined by following system of axioms [9].

MSM 1. $0\leq K\left(a^{(1)},\ldots,a^{(n)}\right)\leq 1$ (an axiom of restriction);

MSM 2. $K\left(a^{(1)},\ldots,a^{(n)}\right)=K\left(a^{(i1)},\ldots,a^{(in)}\right)$, were $i1$, …, in - any rearrangement of numbers $i=1,\ldots,n$ (an axiom of symmetry);

MSM 3a. If $\underbrace{a^{(1)}\wedge a^{(2)}=\varnothing,\ldots,a^{(n-1)}\wedge a^{(n)}=\varnothing}_{\frac{n(n-1)}{2}}$, then $K\left(a^{(1)},\ldots,a^{(n)}\right)=0$ (a direct axiom of the minimal similarity);

MSM 3b. If $K\left(a^{(1)},\ldots,a^{(n)}\right)=0$, then $\underbrace{a^{(1)}\wedge a^{(2)}=\varnothing,\ldots,a^{(n-1)}\wedge a^{(n)}=\varnothing}_{\frac{n(n-1)}{2}}$ (a return axiom of the minimal similarity);

MSM 4a. If $a^{(1)}=\ldots=a^{(n)}$, then $K\left(a^{(1)},\ldots,a^{(n)}\right)=1$ (a direct axiom of the maximal similarity);

MSM 4b. If $K\left(a^{(1)},\ldots,a^{(n)}\right)=1$, then $a^{(1)}=\ldots=a^{(n)}$ (a return axiom of the maximal similarity);

MSM5. $K\left(\lambda a^{(1)},\ldots,\lambda a^{(n)}\right)=K\left(a^{(1)},\ldots,a^{(n)}\right)$, $\lambda>0$ (an axiom of uniformity).

The system of axioms MSM 1-MSM 5 is consistent. For example, to it the multiple-site similarity measure satisfies [2, 5, 7]:

$$K\left(a^{(1)},\ldots,a^{(n)}\right)=\frac{n\left(T-S\right)}{\left(n-1\right)T},$$

$$T\left(a^{(1)},\ldots,a^{(n)}\right)=m\left(a^{(1)}\right)+\ldots+m\left(a^{(n)}\right),$$

$$S\left(a^{(1)},\ldots,a^{(n)}\right)=m\left(a^{(1)}\vee\ldots\vee a^{(n)}\right).$$

Properties of multiple-site dissimilarity measures can be defined from system of axioms MSM 1-MSM 5 and parities $F\left(a^{(1)},\ldots,a^{(n)}\right)=1-K\left(a^{(1)},\ldots,a^{(n)}\right)$.

Types of fractional-linear multiple-site similarity measures

In biology and geography often define «average similarity» series' from several descriptions, and as a measure of similarity take an average arithmetic of set $\frac{n(n-1)}{2}$ of pairs two-site measures' [5, 6]:

$$M_\tau = C_n^2 \sum_{i<j}^{n} K_\tau\left(a^{(i)}, a^{(j)}\right) =$$

$$= \frac{1}{C_n^2} \sum_{i<j}^{n} \frac{K_0\left(a^{(i)}, a^{(j)}\right)}{1+\tau-\tau\, K_0\left(a^{(i1)}, a^{(j)}\right)},$$

$$M_0 = \frac{1}{C_n^2} \sum_{i<j}^{n} K_0\left(a^{(i)}, a^{(j)}\right),$$

$$K_0\left(a^{(i)}, a^{(j)}\right) = \frac{2m\left(a^{(i)} \wedge a^{(j)}\right)}{m\left(a^{(i)}\right) + m\left(a^{(j)}\right)},$$

$\tau > -1$, $C_n^2 = \frac{n(n-1)}{2}$, $n \geq 3$.

The measure M_τ satisfies to system of axioms MSM 1-MSM 5 for everyone $n \geq 3$ [5]. For a series from three, four, etc. descriptive sets lacks are inherent in multiple-site similarity measures M_τ. First, there is no functional dependence between $M_0\left(a^{(i)}, \ldots, a^{(j)}\right)$ and M_τ, $\tau > -1$. Hence, recalculation of values of measures M_τ on values of a measure M_0 is impossible. In this connection it is impossible to determine of the equivalent multiple-site measures classes. Secondly, there is no opportunity to compare the measures specified on unequally in power sequences of the descriptive sets series. In this connection it is necessary to consider types and subtypes of the multiple-site fractional-linear similarity measures deprived these lacks. Following types and subtypes are allocated:

Type of the fractional-linear simila-rity measures K (M_0).

The measure of similarity of this type looks like

$$K_{\{\tau\}} = \frac{K_{\{0\}}}{1+\tau-\tau\, K_{\{0\}}}, \quad K_{\{0\}} = M_{\{0\}}, \quad \tau > -1, \quad \tau$$

does not depend from *n*.

Type of the fractional-linear simila-rity measures K(T,C_Δ)

$$K_{(\tau)} = \frac{K_{(0)}}{1+\tau-\tau\, K_{(0)}}, \quad \tau > 0, \quad K_{(0)} = \frac{2C_\Delta}{(n-1)T},$$

$$T\left(a^{(1)}, \ldots, a^{(n)}\right) = m\left(a^{(1)}\right) + \ldots + m\left(a^{(n)}\right)$$

$$C_\Delta\left(a^{(1)}, \ldots, a^{(n)}\right) = \sum_{i<j}^{n} m\left(a^{(i)} \wedge a^{(j)}\right),\ i = 1, \ldots, n.$$

In this type is allocated three subtypes:

1. $K_{(\tau)}$, $\tau > -1$, τ does not depend from *n*.

2. $K_{\langle(\varepsilon)\rangle}$, $\tau = \frac{(1+\varepsilon)n}{2} - 1$, ε does not depend from *n*.

3. $K_{\langle(\varpi)\rangle}$, $\tau = \frac{2(1+\varpi)}{n} - 1$, ϖ does not depend from *n*. Multiple-site measures of the specified subtypes look like:

$$K_{(\tau)} = \frac{K_{(0)}}{1+\tau-\tau\, K_{(0)}}, \quad \tau > -1, \quad \tau \text{ does not depend from } n;\ K_{(0)} = \frac{2C_\Delta}{(n-1)T};$$

$$K_{\langle(\varepsilon)\rangle} = \frac{K_{\langle(0)\rangle}}{1+\varepsilon-\varepsilon\, K_{\langle(0)\rangle}}, \quad \varepsilon > -1, \quad \varepsilon \text{ does not depend from } n;$$

$$K_{\langle(0)\rangle} = \frac{2K_0}{n-(n-2)K_{(0)}} = \frac{4C_\Delta}{n(n-1)T - 2(n-2)C_\Delta};$$

$$K_{[(\varpi)]} = \frac{K_{[(0)]}}{1+\varpi-\varpi\, K_{[(0)]}}, \quad \varpi > -1, \quad \varpi \text{ does not depend from } n.$$

$$K_{[(0)]} = \frac{n\, K_{(0)}}{2+(n-2)K_{(0)}} = \frac{n\, C_\Delta}{(n-1)T + (n-2)C_\Delta}.$$

Type of the fractional-linear similarity measures K(T, S).

The fractional-linear similarity measures of the type $K(T,S)$, satisfying to system of axioms MSM 1-MSM 5, look like:

$$K_\tau = \frac{K_0}{1+\tau-\tau\,K_0},\quad \tau>0,\quad K_0=\frac{n\,(T-S)}{(n-1)\,T},$$
$$T = m\left(a^{(1)}\right)+\ldots+m\left(a^{(n)}\right),$$
$$S = m\left(a^{(1)}\vee\ldots\vee a^{(n)}\right)$$

In this type it is allocated three subtypes:

K_τ, $\tau>-1$, τ does not depend from n;

$K_{\langle\varepsilon\rangle}$, at $\tau=\frac{(1+\varepsilon)\,n}{2}-1$, $\varepsilon>-1$, ε does not depend from n;

$K_{[\varpi]}$, at $\tau=\frac{2(1+\varpi)}{n}-1$, $\varpi>-1$, ϖ does not depend from n. Multiple-site measures of the specified types look like:

$K_\tau=\frac{K_0}{1+\tau-\tau\,K_0}$, $\tau>-1$, τ do not depend from n; $K_0=\frac{n\,(T-S)}{(n-1)\,T}$;

$K_{\langle\varepsilon\rangle}=\frac{K_{\langle 0\rangle}}{1+\varepsilon-\varepsilon\,K_{\langle 0\rangle}}$, $\varepsilon>-1$, ε do not depend from n; $K_{\langle 0\rangle}=\frac{2(T-S)}{T+(n-2)\,S}$;

$K_{[\varpi]}=\frac{K_{[0]}}{1+\varpi-\varpi\,K_{[0]}}$, $\varpi>-1$, ϖ do not depend from n; $K_{[0]}=\frac{n^2\,(T-S)}{(n^2-2)T-(n-2)\,S}$.

Coordination classes of equivalences of multiple-site similarity measures

Let's define as a coordination for the multiple-site similarity measures a subtype K_τ of types $K(T,S)$.

Definition. Classes of equivalence of the multiple-site similarity measures of type $K(T,S)$ a subtype K_τ are coordinated on unequally in power sequence of descriptive sets' series $\left\{A_{g_e}^{(e)}\right\}$, $e=1,\ldots,N$, $g_e\in\{3,\ldots,n\}$ in only case when, when any two classes of the multiple-site similarity measures equivalence are coordinated.

Two classes of multiple-site similarity measures equivalence of types $K(T,S)$ a subtype K_τ, $K_\tau\left(A_{g_{e1}}^{(e1)}\right)$ and $K_\tau\left(A_{g_{e2}}^{(e2)}\right)$, $g_{e1}\neq g_{e2}$, $\tau>-1$; on series of descriptive sets $A_{g_{e1}}^{(e1)}$ and $A_{g_{e2}}^{(e2)}$ are coordinated in only case when, when from unstrict inequality $K_0\left(A_{g_{e1}}^{(e1)}\right)\geq K_0\left(A_{g_{e2}}^{(e2)}\right)$ unstrict inequality $K_\tau\left(A_{g_{e1}}^{(e1)}\right)\geq K_\tau\left(A_{g_{e2}}^{(e2)}\right)$, $\tau\neq 0$, follows, and equality if it is reached in both inequalities simultaneously. We shall result three theorems of multiple-site similarity measures relations from different types and subtypes.

The theorem 1. The fractional-linear similarity measures of types $K(M_0)$, $K(T,C_\Delta)$, $K(T,S)$, specified on equally in power series of descriptive sets, can be only one of classes of equivalence for each type.

The theorem 2. Multiple-site measures of similarity K_τ, $K_{\langle\varepsilon\rangle}$, $K_{[\varpi]}$, $K_{(\tau)}$, $K_{\langle(\varepsilon)\rangle}$, $K_{[(\varpi)]}$, $K_{\{\tau\}}$, M_τ ($\tau>-1$, $\varepsilon>-1$, $\varpi>-1$, τ,ε,ϖ do not depend from n), are generalization of one class of two-site similarity measures equivalence:

$$K_\tau\left(a^{(1)},a^{(2)}\right)=\frac{K_0\left(a^{(1)},a^{(2)}\right)}{1+\tau-\tau\,K_0\left(a^{(1)},a^{(2)}\right)},$$

at $\tau>-1$, $K_0\left(a^{(1)},a^{(2)}\right)=\frac{2m\left(a^{(1)}\wedge a^{(2)}\right)}{m\left(a^{(1)}\right)+m\left(a^{(2)}\right)}$.

The theorem 3. Classes of multiple-site similarity measures equivalence of descryptive sets from each subtype K_τ, $K_{\langle\varepsilon\rangle}$, $K_{[\varpi]}$ type's $K(T,S)$ (accordingly subtypes $K_{(\tau)}$, $K_{\langle(\varepsilon)\rangle}$, $K_{[(\varpi)]}$ type's $K(T,C_\Delta)$) are coordinated on unequally in power sequence of descriptive sets series $\left\{A_{g_e}^{(e)}\right\}$, $e=1,\ldots,N$, $g_e\in\{3,\ldots,n\}$, and classes of multiple-site similarity measures equivalence from different subtypes of type $K(T,S)$ (accordingly different subtypes of type $K(T,C_\Delta)$) are not coordinated on the same of descriptive sets sequence.

Conclusion

Use of the axiomatic approach for construction of the elementary similarity theory in a multivariate case has led, on the one hand, to construction of a continuum of similarity and dissimilarity measures, and with another to allocation of types and subtypes of its fractional-linear multiple-site similarity measures. Owing to introduction

of such important concepts as the equivalence of multiple-site similarity measures and a coordination of the equivalence classes of multiple-site similarity measures on unequally and equally in power sequences of the series of descriptive sets, has appeared an opportunity to the allocate types and subtypes of multiple-site similarity measures and to establish between them certain relations. In the further it is necessary to consider statistical estimations for the multiple-site measures of similarity for type $K(T, C_{\Delta})$ and to develop techniques on use of the multiple-site similarity measures in biological and geographical researches.

References

1. Pesenko Y.A. The principles and methods of quantitative analysis of faunal studies. – Moscow: Nauka, 1982. 287 p. [in Russian]
2. Semkin B.I. About the axiomatic approach to definition of dissimilarity and quasidissimilarity measures on families of sets // Information methods in systems of direction, measurement and the control. T. 1. Vladivostok: DVNC AS the USSR, 1972. P. 23-26. [in Russian]
3. Semkin B.I., Gorshkov M.V. System of two symmetric functions' variables' axioms and the measures, measuring the relations of similarity and dissimilarity, compatibility and dependence for components of a biodiversity // Bulletin TGEU. 2008.№4. P. 31-46. [in Russian]
4. Semkin B.I., Gorshkov M.V. Axiomatic introduction of measures of similarity, dissimilarity, compatibility and dependence for components of a biodiversity in a multivariate case // Bulletin KrasGAU. 2009.12. P. 18-24. [in Russian]
5. Semkin B.I., Gorshkov M.V. About an estimation of similarity and dissimilarity in a series floristic and phytocenotic descriptions // Komarovsky's readings. 2010. Issue LVII. P. 203-220. [in Russian]
6. Semkin B.I., Gorshkov M.V. Ecological interpretation of multiple-site similarity and dissimilarity measures at research of a differentiating diversity of cenosis // Woods of the Russian Far East: Monitoring of dynamics of woods of the Russian Far East: Mat. V All-Russian. konf. On September, 18-20th 2012 - Vladivostok: LAINS, 2012. P. 198-201.
7. Diserud O.H., Sdegaard F. A. Multiple-site similarity measure // Biol. Lett. 2007.3. P. 20-22.
8. Semkin B.I. The axiomatic approach to introducing measures' for ordering and classification of descriptive sets // Pattern Recognition and Image Analysis. 2011. V.21. .2. P. 164-166.
9. Semkin B.I. Elementary theory of similarity and its use in biology and geography // Pattern Recognition and Image Analysis. 2012. V.22. 1. P. 92-98.

AGGREGATIONALLY CORRECT OPERATIONS ON ALGORITHMS [1]

Z.M. Shibzukhov[2]

[2] Institute of Applied Mathematics and Automation KBSC RAS, Nalchik, Russian Federation, szport@gmail.com

A class of recognition algorithms which are correct with respect to aggregation functional of algorithm's quality. A class of pointwise / aggregationally correct operations on such algorithms, which transforms any finite tuple of pointwise / aggregationally correct algorithms to new pointwise / aggregationally correct algorithm, is discussed. By this way one can extend classes of basic recognition algorithms to extended classes, which are preserve correctness property in pointwise / aggregate manner.

We discuss a new class of *correct operations* on recognition algorithms. A general concept of correct operations was introduced in [1], a class of pointwise correct operations was discussed in [2−3]. In this paper we discuss *correct algorithms* with respect to quality functionals of algorithms that was built as aggregation function of the values of quality function of algorithm's responses or estimates on finite set of samples.

Let $y:\mathbf{X}\to Y$ be unknown functional relationship between descriptions of objects from $\mathbf{X}$ and answers from Y with respect to given question, $a:\mathbf{X}\to Y$ is some algorithm from given class $\mathbf{A}$ that "approximates" y.

Pointwise correct algorithms

Let $Q(a\mid\mathbf{x})$ be a function $\mathbf{A}\times\mathbf{X}\to\mathbf{R}_+$ that estimates quality of response of the algorithm $a\in\mathbf{A}$ on input $\mathbf{x}\in\mathbf{X}$. Let $Y\subset\mathbf{R}$. Let's define for each $\widetilde{y}\in Y$ a subset $U_{\widetilde{y}}\subset Y$ that corresponds to correct answers.

Definition. Response $y=a(\mathbf{x})$ is *correct* if $Q(a\mid\mathbf{x})\in U_{\widetilde{y}}$, where $\widetilde{y}=y(\mathbf{x})$.

One way to estimate the quality of the responses is based on the *loss function* $\ell(y,\widetilde{y})$ that evaluates the "cost" of difference between $y=a(\mathbf{x})$ and $\widetilde{y}=y(\mathbf{x})$.

Definition. $\ell:Y\times Y\to\mathbf{R}_+$ is a loss function if the following conditions holds: 1) if $y=\widetilde{y}$ then $\ell(y,\widetilde{y})=0$; 2) if $y_1\geq y_2\geq\widetilde{y}$ or $y_1\leq y_2\leq\widetilde{y}$ the $\ell(y_1,\widetilde{y})\geq\ell(y_2,\widetilde{y})$.

So response y is correct with respect to "true" answer $\widetilde{y}$ if $\ell(y,\widetilde{y})=0$ and $U_{\widetilde{y}}=\{y\in Y:\ell(y,\widetilde{y})=0\}$.

By the definition $Q(a\mid\mathbf{x})=\ell\big(a(\mathbf{x}),y(\mathbf{x})\big)$.

Definition. Algorithm a is *pointwise correct* on $\mathbf{X}^0$ if for each input $\mathbf{x}\in\mathbf{X}^0$ response $y=a(\mathbf{x})$ is correct.

Pointwise correct operations on algorithms by responses

Let's try to define pointwise correct operations by response on the basis of *aggregation functions* [4−6].

Let M be an aggregation function on Y, i.e. for all $m=1,2,\ldots$:

$\forall y_1,\ldots,y_m\in Y:\ \mathsf{M}(y_1,\ldots,y_m)\in Y$;

$\mathsf{M}(y_1,\ldots,y_m)$ is non−decreasing on Y^m (i.e. for each $y\in Y:\ \mathsf{M}^{-1}(y)$ is connected set)

Criterion. M is correct operation by response if for each m and each $\widetilde{y}\in Y$: $\mathsf{M}(U_{\widetilde{y}},\ldots U_{\widetilde{y}})\subseteq U_{\widetilde{y}}$.

In particular if $U_{\widetilde{y}}=\{\widetilde{y}\}$ and M is idempotent aggregation function (i.e. $\mathsf{M}(y,\ldots,y)=y$) then M is correct operation.

Example. *Weighted g−means by Kolmogorov* [7−8].

[1] Research is supported by RFFI № 12−01−00162−a

$$\mathsf{M}_g\{y_1,...,y_m\} = g^{-1}\left(\sum_{i=1}^{m} w_i g(y_j)\right),$$

Where $w_1,...,w_m > 0$ and $w_1 + \cdots + w_m = 1$, g – strictly monotone continuous function $Y \to \mathbf{R}$.

M_g has well known examples:

power mean: $\mathsf{M}_p\{y_1,...,y_m\} = \left(\sum_{i=1}^{m} w_i y_j{}^p\right)^{1/p}$;

soft max:

$$\mathsf{M}\ \{y_1,...,y_m\} = \frac{1}{p}\ln\left(\sum_{i=1}^{m} w_i e^{py_j}\right);$$

geometric mean: $\mathsf{M}_G\{y_1,...,y_m\} = \prod_{i=1}^{m} y_i{}^{w_i}$.

There is also more general Kolmogorov–like (non–symmetrical) form:

$$\mathsf{M}_{g_1,...,g_m}\{y_1,...,y_m\} = g^{-1}\left(\sum_{i=1}^{m} g_i(y_j)\right),$$

Where $g(y) = g_1(y) + \cdots + g_m(y)$, $g_1,...,g_m$ – are strict monotone functions $Y \to \mathbf{R}$ that are all increasing or all decreasing and has same domain of values.

Condition: If $U_{\tilde{y}}$ is connected interval then $\mathsf{M}(U_{\tilde{y}},...U_{\tilde{y}}) \subseteq U_{\tilde{y}}$, i.e. M_g is correct operation.

Pointwise correct operations on algorithms by estimates

When Y is finite set algorithm a represents by composition: $a = R \circ A$, where $A : \mathbf{X} \to \mathbf{U} \subseteq \mathbf{R}^q$ – algorithm that calculates estimates, $R : \mathbf{U} \to Y$ is decision rule.

In this case we evaluate quality of response $y = a(\mathbf{x})$ by evaluating of the quality of estimate $\mathbf{u} = A(\mathbf{x})$.

Let $\ell : \mathbf{U} \times Y \to \mathbf{R}_+$ be quality function. By definition $Q(a \mid \mathbf{x}) = \ell\big(A(\mathbf{x}), y(\mathbf{x})\big)$. Let $\mathbf{U}_{\tilde{y}} = \{\mathbf{u} \in \mathbf{U} : \ell(\mathbf{u}, \tilde{y}) \in \mathbf{L}_Q\}$.

Let $\mathbf{L}_Q$ be subset of quality values that corresponds to correct responses.

Definition. Estimation $\mathbf{u} = A(\mathbf{x})$ is correct for response $\tilde{y}$ if $\mathbf{u} \in \mathbf{U}_{\tilde{y}}$.

Let $\mathbf{M}$ is aggregation mapping on $\mathbf{U}$, i.e. for each $m = 1,2,...$:

for each $\mathbf{u}_1,...,\mathbf{u}_m \in \mathbf{U}$: $\mathbf{M}\{\mathbf{u}_1,...,\mathbf{u}_m\} \in \mathbf{U}$;

$\mathbf{M}\{\mathbf{u}_1,...,\mathbf{u}_m\}$ is non–decreasing (i.e. for each $\mathbf{u} \in \mathbf{U}$ is connected).

Criterion. $\mathbf{M}$ is correct operation by response if for each m and each $\tilde{y} \in Y$: $\mathbf{M}(\mathbf{U}_{\tilde{y}},...\mathbf{U}_{\tilde{y}}) \subseteq \mathbf{U}_{\tilde{y}}$.

Example. *Weighted multivariate Kolmogorov* $\mathbf{g}$*-means*.

$$\mathbf{M}_\mathbf{g}\{\mathbf{u}_1,...,\mathbf{u}_m\} = \mathbf{g}^{-1}\left(\sum_{i=1}^{m} \mathbf{w}_i \mathbf{g}(\mathbf{u}_j)\right),$$

Where $\mathbf{w}_1,...,\mathbf{w}_m > 0$ and $\mathbf{w}_1 + \cdots + \mathbf{w}_m = \mathbf{1}$, $\mathbf{g}$ – continuous and invertible mapping $\mathbf{U} \to \mathbf{R}^Q$.

Aggregationally correct algorithm

Let Q be some aggregation function. In oreder to evaluate a quality of algorithms on finite set of examples $\mathbf{X}^0$ let's define aggregation functional:

$\mathsf{Q}(a \mid \mathbf{X}^0) = \mathsf{Q}\{Q(a \mid \mathbf{x}) : \mathbf{x} \in \mathbf{X}^0\}$

Let $\mathbf{L}_\mathsf{Q} \subset \mathbf{U}$ be given subset of values of functional Q. Values from $\mathbf{L}_\mathsf{Q}$ (and only they) corresponds to algorithms that are correct in given sense on $\mathbf{X}^0$.

Definition. Algorithm a is aggregationally correct on $\mathbf{X}^0$ with respect to functional Q if $\mathsf{Q}(a \mid \mathbf{X}^0) \in \mathbf{L}_\mathsf{Q}$.

Aggregationally correct operations on algorithms by answers

Let F be aggregation function on Y.

Definition. Operation $a = \mathsf{F}\{a_1,...,a_m\}$ is *aggregationally correct* if $\mathsf{F}\{\mathbf{L}_\mathsf{Q},...,\mathbf{L}_\mathsf{Q}\} \subseteq \mathbf{L}_\mathsf{Q}$.

Let $I \subseteq \mathbf{R}$ be connected interval, $f : I \to I$ – real function $f(x)$, M is idempotent aggregation function. Let's define M–convex functions.

Definition. $f(x)$ is M–*convex* if $f\big(\mathsf{M}\{x_1,...,x_m\}\big) \le \mathsf{M}\{f(x_1),...,f(x_m)\}$

Let's consider Kolmogorov's g–mean M_g.

Fact. Let g be monotone and increasing, $g \circ f \circ g^{-1}$ is convex function $\mathbf{R} \to \mathbf{R}$. Then f is M–convex on I.

For example:

$\ell(y,\tilde{y}) = \left|y^{<p>} - \tilde{y}^{<p>}\right|^{<1/p>}$, where $y^{<p>} = \text{sign}(y)\,|y|^p$ is convex with respect to power mean.

$\ell(y,\tilde{y}) = \frac{1}{p}\ln\left|e^{py} - e^{p\tilde{y}}\right|$ is convex with respect to exponential mean.

Let $f(x)$ be M−convex on I, $\tilde{\mathsf{M}}$ is idempotent aggregation function on I.

Definition. $\tilde{\mathsf{M}}$ is *not dominate* over M if

$$\tilde{\mathsf{M}}\{\mathsf{M}\{u_{11},...,u_{1m}\},...,\mathsf{M}\{u_{N1},...,u_{Nm}\}\} \le \mathsf{M}\{\tilde{\mathsf{M}}\{u_{11},...,u_{N1}\},...,\tilde{\mathsf{M}}\{u_{1m},...,u_{Nm}\}\}$$

For example:

if $\mathsf{M} = \tilde{\mathsf{M}}$ and M is bisymmetrical then M is not dominate over themselves;

$\min$ is not dominate over $\max$.

Theorem. Let $\ell(y,\tilde{y})$ be M−convex on Y, Q is defined on the base of $\tilde{\mathsf{M}}$ and $\tilde{\mathsf{M}}$ is not dominate over M. Then M is aggregationally correct with respect to $\tilde{\mathsf{M}}$.

For example, let $\mathsf{M} = \tilde{\mathsf{M}} = \mathsf{M}_g$. Then

$$\mathsf{Q}(a \mid \mathbf{X}^0) = g^{-1}\left(\sum_{k=1}^{N} w_k \cdot g\left(\ell\left(a(\mathbf{x}_k), y_k\right)\right)\right).$$

In this case aggregationally correct operations has the form:

$$\mathsf{F}\{a_1,...,a_m\} = g^{-1}\left(\sum_{i=1}^{m} w_i g(a_i)\right).$$

References

1. Z.M. Shibzukhov Correct extensions of correct $\Sigma\Pi$ −algorithms // Reports of 15th All−Russian conference “Mathematical methods of pattern recognitions”. M.: Max−Press, 2011. PP. 116−119 (in russian).
2. Z.M. Shibzukhov. Poinwise correct operations on algorithms // Report of International conference IIP−10. M: Torus−Press. 2012. PP. 90−93 (in russian).
3. Z.M. Shibzukhov. Pointwise Correct Operations on Recognition and Prediction Algorithms // Doklady Mathematics, Pleiades Publishing, Ltd. 2013, Vol.87, No.3, PP.1−4.
4. Beliakov G., Pradera A., Calvo T.. Aggregation Functions: A Guide for Practitioners. Springer, Heidelberg, Berlin, New York, 2007.
5. Mesiar R., Komornikova M., Kolesarova A., Calvo T.. Aggregation functions: A revision. In H. Bustince, F. Herrera, and J. Montero, editors, Fuzzy Sets and Their Extensions: Representation, Aggregation and Models. Springer, Berlin, Heidelberg, 2008.
6. Grabich M., Marichal J.−L.,PaP E. Aggregation Functions. Series: Encyclopedia of Mathematics and its Applications, No.127. Cambridge University Press. 2009.
7. Kolmogoroff A.N. Sur la notion de la moyenne. (in French). Atti Accad. Naz. Lincei, 12(6):388–391, 1930
8. Nagumo M. Uber eine klasse der mittelwerte. (in German). ¨ Japanese Journ. of Math., 7:71–79, 1930.

BIO-INSPIRED SOLUTION FOR THE HOMOGRAPHY PROBLEM

Z. Talai[1,2], Y. Mohamed Ben Ali[1,3],

[1] Computer Science Department, University of Badji Mokhtar, P.O. Box No.12, Annaba 23000, Algeria.
[2] talai_zoubir@yahoo.com, [3] benaliyam2@yahoo.fr

In this paper, a Particle Swarm Optimization proposition is presented to solve the Homography problem. We tested our technique with several datasets. Also a comparison with existing technique (SVD) is done. PSO shows a significant superiority comparing to classic approach (SVD). The main gain of this technique is the accuracy of results and easiness of implementation.

Introduction

Lately, significant interest is growing for 3D reconstruction. Lots of researchers are trying to propose an optimal solution for this problem. We call 3D reconstruction, every process that creates a virtual object from a picture of this one. Lots of approaches were proposed to achieve this reconstruction. Authors in [3] have used a moving depth camera (Kinect) to achieve a high detailed 3D reconstructions of an indoor scene in real-time. Another approach was proposed in [6], where they used a set of images of the scene combined with the help of the user to assign labels to polygonal regions. To build the 3D model, they only used the intersection of the polygons. There are lots of other techniques dedicated to 3D reconstruction. For further information, see [2], [4].

To build a 3D model of an object from its picture, we reverse the process that gave us the picture in first place. In other words, project the object's pixels from 2D picture coordinates to 3D space coordinates. The main problem here is the source of the picture (camera). The projection matrix used to take the picture is not public information, and most camera constructors don't provide it.

Zhang in [5] proposed a calibration technique, where he deduces the camera's intrinsic and extrinsic parameters. The approach requires that the camera observes a planar pattern shown at a few (at least two) different orientations. Then, a closed-form solution is calculated, followed by a non-linear optimization to refine the results.

This paper introduces another technique to solve camera calibration that will use Particle Swarm Optimization to find Homography between images of a calibration pattern (chessboard). Once the Homography is calculated, camera's intrinsic parameters can be easily deduced.

Problem Statement

A. Homography

The homography is a non-singular linear relationship between points on planes. The homography between two views plays an important role in the geometry of multiple views. Images of points on a plane in one view are related to corresponding image points in another view by a planar homography using a homogeneous representation. The homography depends on the intrinsic and extrinsic parameters of the cameras used for the two views and the parameters of the 3D plane [1].

B. Mathematical definition

In the complex plane, a Mobius transformation is frequently called a Homography. These linear-fractional transformations are expressions of projective transformations on the complex projective line, an extension of the complex plane. In higher dimensions Homogeneous coordinates are used to represent projective transformations by means

of matrix multiplications. With Cartesian coordinates matrix multiplication cannot perform the division required for perspective projection. In other words, with Cartesian coordinates a perspective projection is a non-linear transformation.

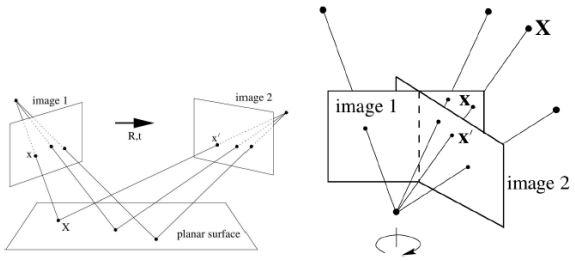


Fig. 1. Homography

Given:

$$m = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} M' = \begin{bmatrix} w \times X \\ w \times Y \\ w \end{bmatrix} H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$$

Then:

$$m = H \times M'$$

Where:

$$M = \frac{M'}{w} = \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

With H: Homography matrix. M: The model points in space. m: The model points projection on the picture.

C. Homography estimation

Points are the simplest and fundamental features that can be used for estimating homography. They have also been the main focus of the researchers for homography estimation. These methods recover the homography as a linear relationship on point and line features. These are the main ones:

- Direct Linear Transformation (DTL) using point correspondences.
- DLT using line correspondences.
- Robust estimation using RANSAC.
- Statistical optimization.

Particle Swarm Optimization

The basic idea of the PSO algorithm works by having a population (swarm) of candidate solutions (particles). These particles wander around in the search-space according to a few simple formulas. The movements of the particles are guided by their own best known position in the search space as well as the entire swarm's best known position.
When improved positions are discovered these will guide the movements of the swarm. The process is repeated and a suitable solution will eventually be discovered.

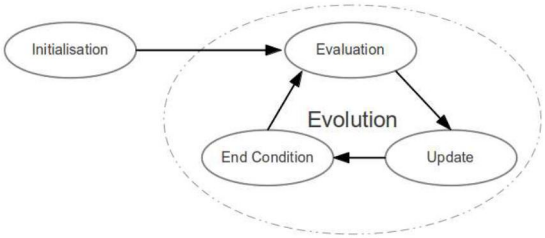


Fig.2. General cycle of particle swarm optimization process

Resolution of the Homography Matrix Based on PSO

A. First estimation of H

Having the model points coordinates in space (the axes $(\overrightarrow{X}, \overrightarrow{Y})$ and the origin chosen in the picture) and their correspondent ones in the image pixels, we follow the inverse operation for matrices multiplication.

$$H \times M = m \qquad (1)$$

Each particle will represent a vector of nine real positions which represent the solution of the Homography matrix. In the next step every particle will try to find a solution for:

$$m - H \times M = 0$$

B. Algorithm description

The resolution of the Homography matrix PSO-based undergoes several steps to find the best or the optimal values of the matrix.

a) Step 1. Initialization: For each particle, we initialize a random position and a random velocity value; this one must be within a specified interval.

b) Step 2. Evaluation: In PSO, the success of the method is in the good choice of the evaluation function which is called fitness function. The main purpose for using it is to minimize/maximize its result. For our problem we used the function:

$$m - H \times M = 0$$

With m, M are 3×n (n points number chosen in the picture) matrices and H is a 3 × 3 matrix described as follow:

$$H = \begin{bmatrix} x1 & x4 & x7 \\ x2 & x5 & x8 \\ x3 & x6 & x9 \end{bmatrix}$$

With Xi (i=1...9) are the position vector of the particle. So, to calculate the fitness, the particle must have a position that satisfies Eq.(1). To have a good constraint on the position vector we modified Eq.1 to:

$$m - y = m - (M^T \times H^T) = 0 \qquad (2)$$

With:

$$y_{i,1} = x1 \times M_{i,1} + x2 \times M_{i,2} + x3 \times M_{i,3} \qquad (3)$$

$$y_{i,2} = x4 \times M_{i,1} + x5 \times M_{i,2} + x6 \times M_{i,3} \qquad (4)$$

$$y_{i,3} = x7 \times M_{i,1} + x8 \times M_{i,2} + x9 \times M_{i,3} \qquad (5)$$

After the calculation of y, we did choose the Euclidean distance function to minimize:

$$fitness = \sqrt{\sum_{i=1,j=1}^{n,3} (m_{i,j} - y_{i,j})^2} \qquad (6)$$

Last thing to do in this part is to check if the fitness of a given particle is its best one to save or else keep the old one, in addition we look for the best fitness of neighbors and also save it for the update step.

c) Step 3. The update: In this step, we update every particle's velocity and position using the standard equations.

Experimentation

A. Application on synthesized images

For the experimentation, we tested the algorithm using synthesized images and real ones. First, we select the same feature points in two pictures. Next we launch the swarm with pixel's positions as data. The final result is the Homography matrix i.e the best position of the particle that has the smallest fitness.

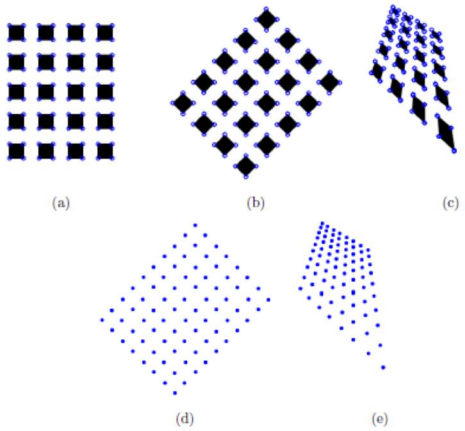


Fig.3: synthesized chessboard with 80 points

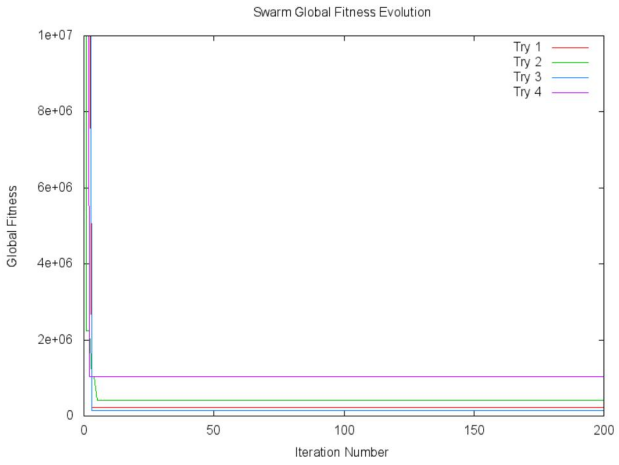


Fig.4: The evolution of the swarm using 4 points dataset (A zoom in region)

B. Application on INRIA dataset

We used Krystian Mikolajczyk's applet for Stereo Vision, and INRIA dataset to test our approach,

the applet is available here. The figures 5, 6 show the superiority of PSO Homography over the used approach in the applet (SVD).

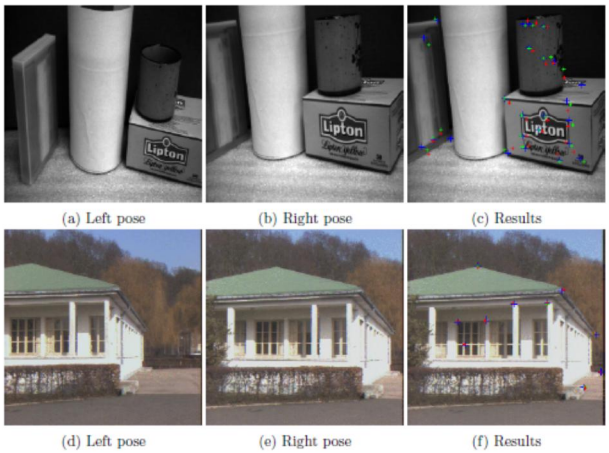


Fig.5: Error Projection comparison: Blue crosses are the chosen points, Green are the projection using PSO Homography, Red are the Mikolajczyk's applet results

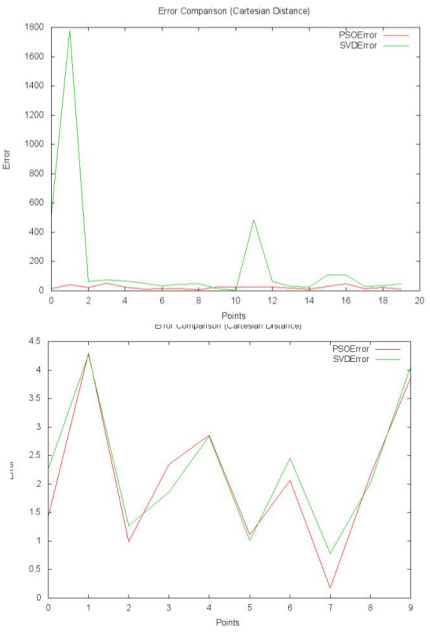


Fig.6: Error Projection comparison (Cartesian Distance) Red: PSO Error, Green: SVD Error

Discussion

We proposed PSO for Homography calculation in camera calibration process because lots of raisons, we may explains the important ones.

- First of all, we know that the computational approach for Camera Calibration Homography is an approximation that is found in the Least Square Sense with the calculation of Singular Value Decomposition (SVD), which is very complicated technique.
- One may go with a computational method to gain precision and accuracy, but we proved that PSO gives better accuracy comparing to Mikolajczyk's applet which use SVD to solve Homography.
- Also we know that the classic approach is composed of too many stages; each one is more complicated than the other (Bidiagonalisation, Householder transformation, Diagonalisation. . .) not as PSO approach which is very simple once the fitness function is found.
- Last positive point about PSO is the flexibility, one could change neighboring architecture and other sophisticated techniques to speed up the convergence.

Conclusion

In this paper, we presented an alternative approach for Homography calculation using Particle Swarm Optimization. We tried to simplify a complex mathematical technique by using an evolutionary approach. The operation needs 4 points (minimum) coordinates to get good results. The experimentation showed easiness of implementation, and large gain of time and accuracy. Thus: the superiority of our approach.

References

1. Anubhav Agarwal, C. V. Jawahar, and P. J. Narayanan. A survey of planar homography estimation techniques. CiteSeerX – Scientific Literature Digital Library and Search Engine, 2005.
2. Riccardo Gherardi. Advances in 3D Reconstruction. PhD thesis, Universit`a degli Studi di Verona, March 2010.
3. Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Pushmeet Kohli Richard Newcombe, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, and Andrew Fitzgibbon. KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera, October 2011.
4. Branislav Miucusık. Two-View Geometry of Omnidirectional Cameras. PhD thesis, Czech Technical University, June 2004.
5. Zhengyou Zhang. A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11):1330–1334, 2000.
6. Remo Ziegler, Wojciech Matusik, Hanspeter Pfister, and Leonard McMillan. 3d reconstruction using labeled image regions. Eurographics Symposium on Geometry Processing, pages 1–12, 2003.

TRACK 2 "REPRESENTATION, PROCESSING, ANALYSIS AND UNDERSTANDING OF IMAGES"

TWO-PASS ADAPTIVE HISTOGRAM BASED METHOD FOR RESTORATION OF FOGGY IMAGES

V. Agarwal[1], S. Khandelwal[1], D. Goyal[1], J. Sharma[2] and A. Tiwari[3]

[1]The LNM Institute of Information Technology, Jaipur
[2]Microsoft India R&D Pvt Ltd, Hyderabad
[3]Indian Institute of Technology, Jodhpur
vinit.agarwal@lnmiit.ac.in, shubham.khandelwal@lnmiit.ac.in, goyalldushyant@gmail.com, jatinsha@microsoft.com, akt@iitj.ac.in

In this paper, we present a Local Histogram Equalization (LHE) based method to improve the visual quality of foggy images. From the fog-degraded image we separate all three color components and apply the enhancement algorithm to each of these components. To mitigate the issue of blocking artifacts generated from Local Histogram Equalization with non-overlapping blocks, we propose to process blocks by overlapping them. Processing smaller block size yields better visual quality at the expense of relatively higher processing time, therefore, we also propose to use statistically optimal block size to obtain a good visual quality in reasonable time duration.

Introduction

In practice, it is often observed that visual quality of images gets deteriorated due to various natural phenomena. Fog or haze is one such phenomenon, caused by presence of suspended water droplets in the air. Fog removal or defogging of an image is, therefore, an important requirement that applications ranging from navigation, outdoor surveillance etc.

The visual quality degradation of an image due to fog, is a function of fog density and the distance of the scene. Quality of visibility depends on the extent of scattering caused by water droplets present in the fog. All the methods available in literature can be broadly classified into two categories - Physical model based methods [1]-[6] and Image Processing based methods [7]-[10]. In the first category, Huang et al. [2] proposed a bilateral filtering based method for fast recovery of foggy images. Shwartz et al. [3] suggested a method to remove spatially varying contrast by stray radiance (airlight) method. The key claim in their method is that a hazy scene recovery is only possible by subtraction of airlight. Kim et al. [4] proposed to estimate airlight using a cost function based on human visual model. They subtracted the estimated airlight map from degraded image to enhance its quality. The luminance component of image is employed for airlight estimation. Oakley et al. [1], [5] suggested a simple correction for contrast loss in foggy scenes. They restored scene contrast by approximating distribution of radiance in the scene. This is done by a Gaussian function with known variance and mean. Their method does not use any estimate of weather information. Robby T. Tan et al. [6] estimated the color of skylight and the values of airlight.

In the second category i.e. Image processing based methods, Z. Xu et al. proposed Contrast Limited Adaptive Histogram Equalization (CLAHE) [7]. Major drawback of this method is that, often, noisy pixels also get enhanced along with enhancement of the foggy pixels. However, the same method when applied on medical images yielded good results in improving visibility of such images. They also proposed Bilinear Interpolation Dynamic Histogram Equalization [9] method. The given image is divided into sub-images of some smaller size and then partitioned into corresponding sub-histograms without domination in Jia et al. [10] who proposed using Content Adaptive Local Contrast Enhancement algorithm (CLAHE) and non-overlapping block processing only on luma component of YCbCr part of image. We propose a similar algorithm in this paper.

They applied their method in YCbCr color space with non-overlapping blocks whereas we apply the CLAHE algorithm on each of the RGB color spaces using overlapping blocks. We call this as pre-processing stage. Due to the overlapping, some of the pixels get equalized by its neighborhood pixels and haze is significantly removed. This causes much better nearer vision than obtained by methods reported in literature. Moreover, in the second stage we applied the CLAHE algorithm [7] for post-processing. The post-processing is applied on intensity component of the image processed in the first stage. This is done because intensity component also gets enhanced after reducing the noise level in the pixels of image.

Proposed Algorithm

In this paper, we present an overlapping block-by-block, contrast limited adaptive histogram equalization (CLAHE)-based method. This is intended to overcome limitations of CLAHE algorithm. The proposed method is based on the general observation that human vision system is highly sensitive to noise. The proposed method is described as follows.

The foggy image captured by camera is divided into blocks and these blocks are taken in such a way that they overlap over a small number of columns. Each block of the image is split into R, G, B channels respectively. Now each channel of the block is processed by CLAHE. Finally, the block processed in each component is converted back to RGB block image. The above procedure is followed for processing of all the overlapped blocks.

Now the processed image is converted from RGB color space to HIS color space. The reason of conversion in HIS representation is the fact that the human eye color sensation matches with HIS representation. After this conversion, the intensity component of the image is processed by CLAHE while Hue and Saturation components are left unchanged. Finally, the image processed in HIS color space is converted back to RGB color space. By using the block segmentation of an image for each region, as in the case of an image with various depths, the contribution of airlight vary according to region. The small region can be considered to have approximately same fog density in that region. Thus, processing a block according to its surrounding pixels.

Overview of CLAHE

In general, local histogram of a pixel, x, is the same as the histogram of pixels in a rectangular window with the pixel x into its center. Only the pixels within the local area are considered. But according to the characteristic of human vision, the visual systems change with the region and these systems are affected by the surrounding environment. To solve these problems, S.M. Pizer [8] proposed a method which is called contrast limit Adaptive histogram equalization (CLAHE). The CLAHE method applies histogram equalization to a contextual region. Each pixel of original image is in the center of the contextual region. The original histogram is clipped and the clipped pixels are redistributed to each gray level. The new histogram is different from the original histogram, because each pixel intensity is limited to a user-defined maximum. So CLAHE limits the noise enhancement. The algorithm proposed in this paper is intended to overcome the limitation of CLAHE algorithm.

Improved CLAHE

CLAHE establishes a maximum value to clip the histogram and redistributes the clipped pixels equally to each gray level. It limits the noise while enhancing the contrast. By increasing ClipLimit, the noise in the dark background is enhanced to an extent that the image starts looking messy. The parameter ClipLimit sets a limit for the contrast enhancement. The higher value of ClipLimit gives higher contrast and a more flat histogram. To brighten the foreground and to prevent a highly noisy background, the parameter value of ClipLimit is statistically obtained as 0.05 by experimenting on large number of images. CLAHE operates on small regions in the image, called tiles, rather than the entire image. If the number of tiles is increased by powers of 2, we will get higher

value of number of tiles, dividing the image into smaller regions for local histogram equalization. On increasing number of tiles, it gives a better contrast in the foreground. However, when the number of tiles is too large, the data points in each tile would be too few for equalization to do a good adjustment.

The method proposed in this paper has the following advantages over other local adaptive histogram equalization methods.

1. Processing image by overlapping blocks increases level of enhancement. Histogram is equalized on a pixel by pixel basis as each pixel characteristics is dependent on its surrounding i.e. overlapping block.
2. The level of enhancement for each region of the image is determined by local image statistics. These statistics are estimated on a block-by-block basis.

Detailed Procedure

The detailed steps of the proposed method are as follows:

1 The image is split into smaller blocks and the blocks are divided into R, G and B parts respectively. Apply CLAHE method on them. Then by overlapping the blocks, we take the smaller size of block and then divide each block image component into different tiles for local enhancement.

2 In addition, we perform post-processing to further improve the image quality by adjusting the illumination of the image.

3 In HIS color space, keep the H and S part unchanged and process Intensity part with the same clipping limit of CLAHE method.

4 After processing Intensity part of image, the H, S and processed I part are combined and converted back to RGB color space.

Experimental Results and Discussions

In order to evaluate performance of our method, we tested it on different types of images degraded by fog. Some results are shown in Fig 1. Fig. 1.d shows the original foggy image. Fig. 1.e shows the result by CLAHE [7]. The result is enhanced but the distant visibility still needs improvement. Fig. 1.f shows the result by Haze removal [11], but in this algorithm, the artifacts are also added which were not initially present in the given image. For instance, Fig. 1.f.ii shows that the output intensities are lowered giving black artifacts. Similarly, Fig.1.g is the result of Visibresto [12] which enhanced the image but the distant vision is still not clear. But Fig. 1.h shows the results of the proposed algorithm which are better than various methods compared. As we decrease size of the blocks, we found that haze removal capacity of our algorithm improves. In this algorithm, we have used 64x64 size of the block that results into a total of 81(9x9) number of tiles for the images of dimension.The proposed block size is obtained statistically.

Conclusion

In this paper, we present a contrast limited Content Adaptive Histogram Analysis that sets the visibility restoration from a single image without using any extra information and enhances the contrast of images. Our aim in this work is to improve the visual quality of distant objects. From the figures, we can see that most of the fog has been removed from the images. This gives better contrast enhancement and brightness compared to other techniques. We believe that our method can be very useful in various fields/systems such as outdoor surveillance systems, intelligent vehicle systems, remote sensing systems, graphics editors etc. Encouraged by good results of the system, the proposed method can be used for real time applications related to fog removal.

References

1. J.P. Oakley and B.L. Satherley, “Improving image quality in poor visibility conditions using a physical model for contrast degradation," Image Processing, IEEE Transactions on, vol. 7, no. 2, pp. 167-179, 1998.
2. Y.M. Huang, M.K. Ng, and Y.W. Wen, “Fast image restoration methods for impulse and gaussian noises removal," Signal Processing Letters, IEEE, vol. 16, no. 6, pp. 457-460, 2009.
3. S. Shwartz, E. Namer, and Y.Y. Schechner, “Blind haze separation” in Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. IEEE, 2006, vol. 2, pp. 1984-1991.

4. D. Kim, C. Jeon, B. Kang, and H. Ko, “Enhancement of image degraded by fog using cost function based on human visual model," in Multisensor Fusion and Integration for Intelligent Systems, 2008. MFI 2008. IEEE International Conference on. IEEE, 2008, pp. 64-67.
5. K.K. Tan and J.P. Oakley, “Enhancement of color images in poor visibility conditions," in Image Processing, 2000. Proceedings. 2000 International Conference on. IEEE, 2000, vol. 2, pp. 788-791.
6. R.T. Tan, “Visibility in bad weather from a single image," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. Ieee, 2008, pp. 1-8.
7. Z. Xu, X. Liu, and N.J.F.R. from Color, “Images using contrast limited adaptive histogram equalization image and signal processing, 2009. cisp'09," in 2nd International Congress, 2009, pp. 17-19.
8. S.M. Pizer, E.P. Amburn, J.D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J.B. Zimmerman, and K. Zuiderveld, “Adaptive histogram equalization and its variations," Computer vision, graphics, and image processing, vol. 39, no. 3, pp. 355-368, 1987.
9. Z. Xu and X. Liu, “Bilinear interpolation dynamic histogram equalization for fog-degraded image enhancement," 2010.
10. Z. Jia, H. Wang, R. Caballero, Z. Xiong, J. Zhao, and A. Finn, “Real-time content adaptive contrast enhancement for see-through fog and rain," in Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on. IEEE, 2010, pp. 1378-1381.
11. K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior," in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. Ieee, 2009, pp. 1956-1963.
12. J.P. Tarel and N. Hautiere, “Fast visibility restoration from a single color or gray level image," in Computer Vision, 2009 IEEE 12th International Conference on. IEEE, 2009, pp. 2201-2208.

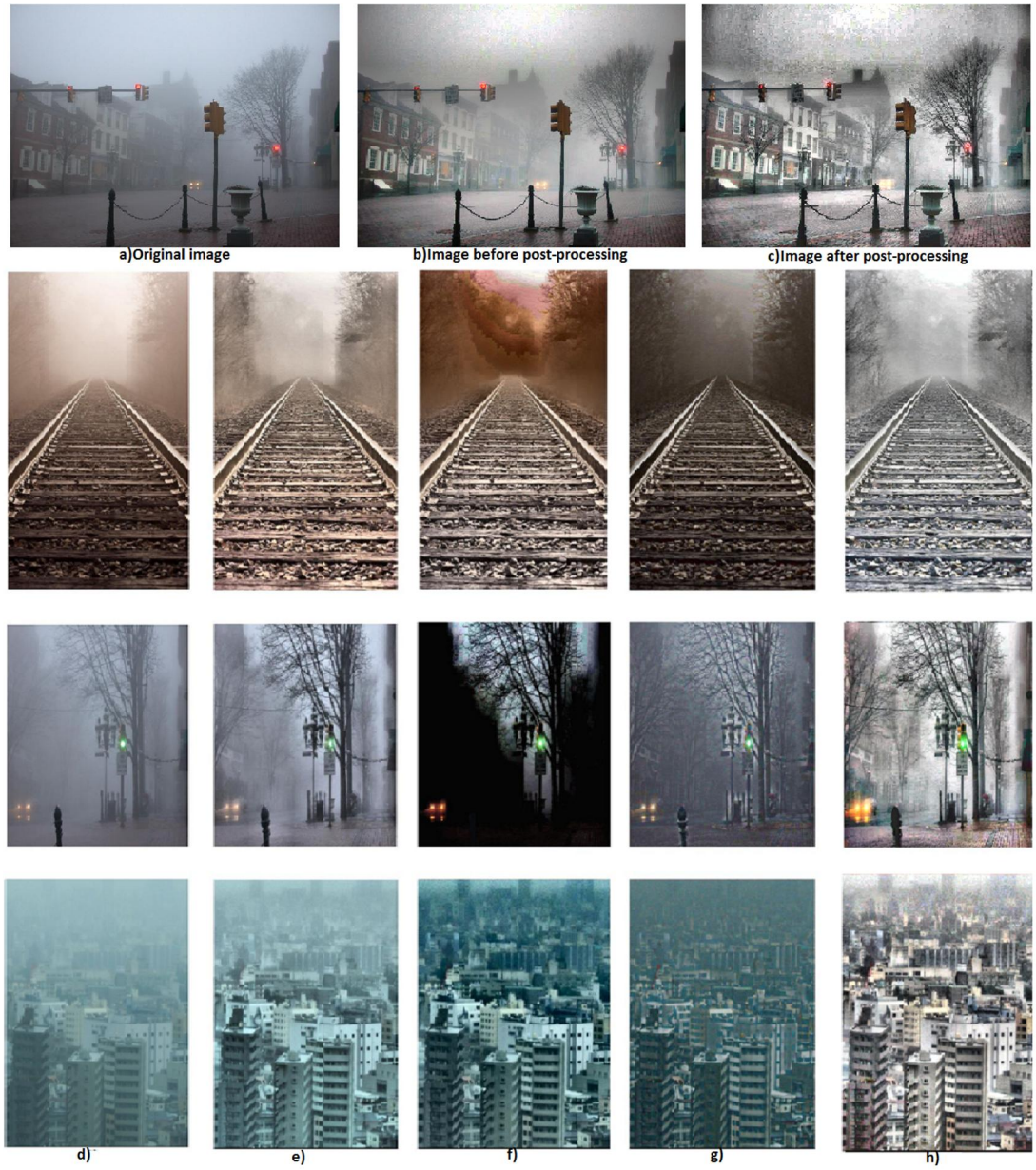


Fig.1. d) Input Image e) CLAHE [7] f) Haze Removal [11] g) Visibresto [12] h) Our Result

EDGE DETECTION ALGORITHM BASED ON NON LINEAR CELLULAR NEURAL NETWORKS

M.V. Akinin[1,2], M.B. Nikiforov[1,3]

[1] Moscow State University of Economics, Statistics and Informatics
Ryazan, Russian Federation [2]verzhak@gmail.com, [3]nikiforov.m.b@evm.rsreu.ru

In this research we give the theoretical description of an artificial nonlinear cellular neural network , different combinations of network parameters, that are used for edge detection in the image, results of the practical research of the possibility to use the network for the tasks of an operational edge detection.

Introduction

Modern applied image processing has some tasks that should be solved in the conditions of the current time. These tasks appear during the operational monitoring of the ecological situation, in the process of making automatic navigation systems that precise data on filming results and in some other tasks.

One of these tasks is the detection of contours that is enough to continue automatic matching of video data with topography maps [1]. The solution of this task, suitable to be used in mass-volume, must fulfill the following criteria:

- low time study;
- low computational costs(thus, low solution cost);
- accuracy that will be enough to make an automatic matching of the image with the video card.

The last demand supposes not detection of all edges, that can be seen, but detection of the most extensive ones and their additional (according to the contour's nature) approximation. So, for example, approximation may be necessary when we detect the edge of the water object, that is planted with vegetation, growing on the bank and not shown in the map.

Current edge detection algorithms don't fulfill all conditions that were specified previously:

- algorithms, based on differential operators and Sobel operator (including algorithm Canny) detect a plenty of contours, not important for the task of the automatic matching of the image with a video card, that leads to the increasing of computing and time difficulties of the matching algorithm;
- algorithms, based on wavelet conversions (Gabor wavelet and others), Hough method, require substantial computing expenses, connected with more operations with floating point.

The purpose of this research is to develop the edge detection algorithm that can fulfill all conditions described above.

In the present work we offer to use an algorithm based on nonlinear cellular neural networks for edge detection.

Theoretical description

Nonlinear cellular neural networks (NLCNN) is two-dimensional artificial neural system with the size $H \times W$ neurons. In this research we presume that NLCNN's size U .

Every neuron calculates its output y_{ij}^t at the moment of time t ccording to the collection of input data, taking into account the feedback inside neuron. In this work we consider NLCNN, which work is quantized in time- the feedback influences neuron's work with the retard of one network iteration.

The scheme of a neuron's structure is given in the picture 1.

The algorithm of neuron (i, j) on a iteration t consists [2]of the following steps:

1. the data of the original image U_{ij}^t come to the neuron's input, network neuron outputs Y_{ij}^t of the previous iteration and feedback's weight x_{ij}^{t-1}, besides $U_{ij}^t \subset U$ and $Y_{ij}^t \subset Y$, so each neuron has its own two dimensional local sensor field F_{ij}^t with the size $h^t \times w^t$.

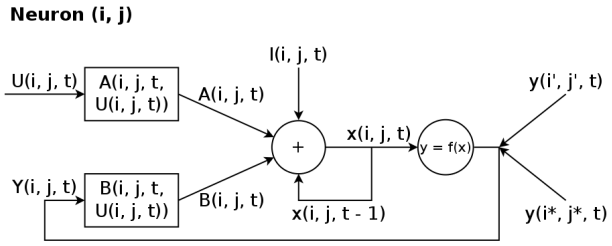


Fig. 1. Neuron

As U_{ij}^t we take the window, sized $\left(h^t, w^t\right)$ pixels in the image U with the center in the pixel $\left(i, j\right)$;

2. computing of (1) and (2):

$$\dot{A}_{ij}^t = \sum_{(i',j') \in F_{ij}^t} A_{ij}^t \left(u_{i'j'}^t\right), \quad (1)$$

$$\dot{B}_{ij}^t = \sum_{(i',j') \in F_{ij}^t} B_{ij}^t \left(y_{i'j'}^t\right), \quad (2)$$

where A_{ij}^t and B_{ij}^t functions are nonlinear functions of computing the nature of a neuron's connection with matrix elements U_{ij}^t and Y_{ij}^t;

3. computing of (3):

$$x_{ij}^t = -x_{ij}^{t-1} + \dot{A}_{ij}^t + \dot{B}_{ij}^t + I_{ij}^t, \quad (3)$$

where I_{ij}^t is threshold value;

4. computing of (4):

$$y_{ij}^{t+1} = f\left(x_{ij}^t\right), \quad (4)$$

where f is the function of the neuron activation.

It is shown in [2] that $Y^t \to Y_{const}$ and $t \to \infty$, thus, the network can be stopped if one of the following conditions is carried out:

- the network performed a rather great quantity of iterations;
- $\sum\sum \left|Y^{t+1} - Y^t\right| < \varepsilon$, where ε is chosen according to the task.

After network standstill, the matrix Y^{t+1} is accepted as output network matrix.

The time and the result of the network is adjusted to the following parameters:

- the sizes h^t, w^t of the local sensor neural field;
- functions A_{ij}^t and B_{ij}^t;
- thresholds I_{ij}^t;
- neuron activation function f.

The process of NLCNN work may be threaded. Within the scope of this work was made hardware-software system of edge detection, including multicore graphic processor, supporting common computing video card technology. During the experiment each neuron worked in a separate current, which were synchronized at the end of the network iteration.

Experimental part

In the experimental part we studied the time study of this algorithm and suitability of different combinations of the network value parameters for edge detection.

For thinning and approximation were used algorithms, described in [3].

In picture 2 the frame of text video consequence is given. Qualitative edge detection in the frame in the context of the task of semiautomatic matching with the video card is river edge detection, lakes edges detection and horizon detection.

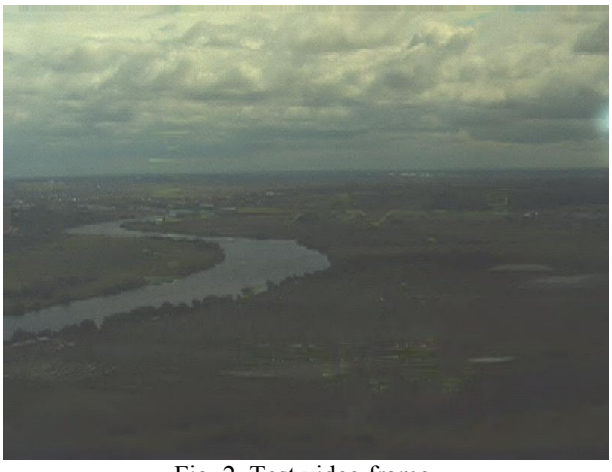


Fig. 2. Test video frame

In [2] are given the following values of the network, used for edge detection:

- $w^t = h^t = 3$, $I_{ij}^t = -1$;
- $A_{ij}^t\left(u_{i'j'}^t\right) = \begin{Bmatrix} -u_{i'j'}^t, & i' \neq i \cup j' \neq j \\ 8u_{i'j'}^t, & i' = i \cup j' = j \end{Bmatrix}$;
- $B_{ij}^t\left(u_{i'j'}^t\right) = \begin{Bmatrix} 0, & i' \neq i \cup j' \neq j \\ 2y_{i'j'}^t, & i' = i \cup j' = j \end{Bmatrix}$;
- $f(x) = \dfrac{|x+1| - |x-1|}{2}$.

The results of algorithm's running with these parameters are shown in picture 3 and 4.

Fig. 3. Edge detection results [2]

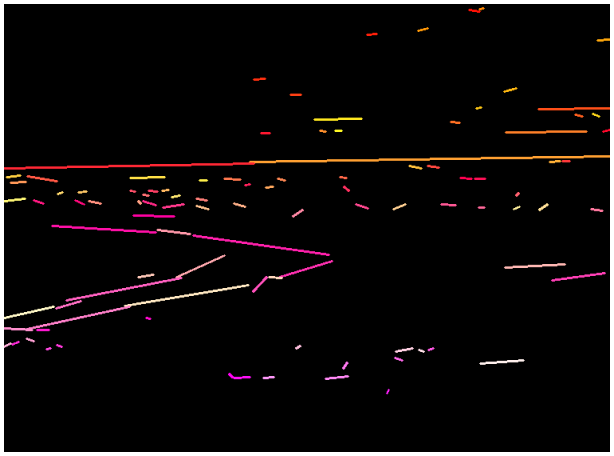


Fig. 4. Edge detection results - thin, approximated lines [2]

During this research, several modifications of the given parameters were analyzed. So, different ways of function f representation were researched. In the pictures 5 and 6 are given results of NLCNN work in the case of using hyperbolic tangent as neuron activation function.

Hyperbolic tangent (and any sigmoid function) allows obtaining gradual threshold modification and improves the result of edge detection of the objects.

Except hyperbolic tangent in NLCNN, the results are shown in pictures 5 and 6, were used some other improvements- changing size of sensor field and nonlinear function $A_{ij}^t(\ )$ and $B_{ij}^t(\ )$.

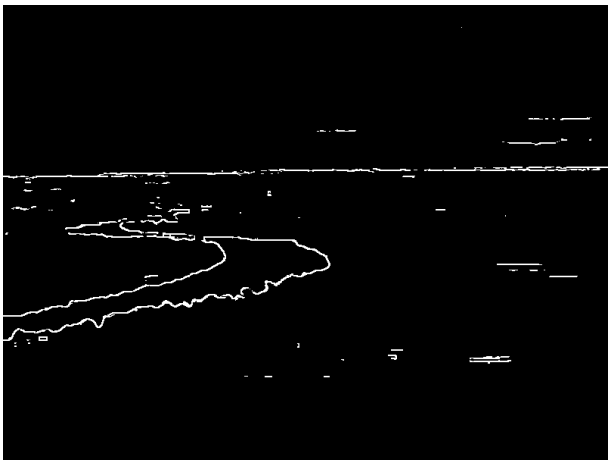


Fig. 5. The results of edge detention (hyperbolic tangent as a function)

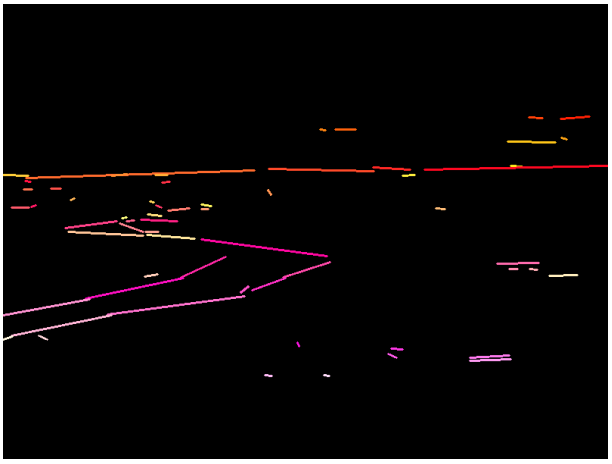


Fig. 6. The results of edge detention-thin approximated lines (hyperbolic tangent as a function)

In table 1 are given time studies for edge detection by different algorithms. Time study assessment was made in the computing system of the following components:

- hardware component:
- CPU - Intel Core i5 3337U (CPU speed : 2,7 kilo megahertz, 4 computational cores, the cache L3-3 megabytes);
- GPU - AMD Radeon HD 8750M (CPU speed-775 kilo megahertz; RAM is 2048 gigabytes (DDR3 technology); 384 computational cores);

- RAM is 4 gigabits DDR3;
- software:
- OS GNU/Linux (the version of Linux core 3.8.6);
- OpenCL standard realization in catalyst drivers of version 13.1.

Table 1. Comparison of temporal efficacy of edge detention algorithms

Algorithm	The time of text frame processing (sec., average, according to 1000 running of the algorithm)
NLCNN	0.01
Sobel	0.015
Canny	0.019
Prewitt	0.011
Gabor	0.04

Conclusions

According to the results of this research, we can make a conclusion that edge detection, based on NLCNN, fulfills conditions, required by the task of automatic image and video card matching in the algorithm of edge detection.
The parallel version of the edge detention algorithm based on NLCNN, has low time study, able to work in a relatively cheap (comparing with other means of threading computation) multicore processors, supporting common computing technology in the video card and allows to get enough accuracy of the most important edge detection (extensive and distinct) for automatic matching of the contours in the image.

References

1. S.I. Elesina, A.A. Loginov, M.B. Nikiforov. Correlation - extremal navigation systems. (in Russian) // Bulletin of the Ryazan State Radio Engineering University. Appendix. - Ryazan: Ryazan State Radio Engineering University. - 2012. - pp. 44 – 47.
2. L. O. Chua, T. Roska. Cellular neural networks and visual computing. Foundations and applications. - UK, Cambridge: Cambridge university press. - 2004.
3. R.C. Gonzalez, R.E. Woods. Digital Image Processing. - Prentice Hall. - 2002.

SEARCHING OF ZONES OF INTEREST ON VECTOR LOCALLY HOMOGENEOUS SCENES

R.M. Aleev[1,2], V.B. Fofanov[1,3]

[1] Kazan Federal University, 18 Kremlevskaya str., Kazan 420008, Russia; ph.: 7 (843)2315453;
[2]aleevrm@yandex.ru; [3]Viatcheslav.Fofanov@ksu.ru

We propose a technique for determining the so-called zones of interest, i.e., scene fragments that contain given objects and their neighborhoods. We use a collection of spatially juxtaposed images as the initial data and do a vector locally homogeneous random field as a scene model.

Introduction

The shape of an object is an important unmasking feature. Optoelectronic systems project a scene onto a plane perpendicular to the observation direction. This allows one to classify objects not by their shape but by the shape of their projections.

The construction of projections which form a scene of objects is called the segmentation. Evidently, the segmentation of the whole scene is much more difficult than that of its fragment containing only a given object and some its neighborhood. We treat such fragments as zones of interest. Therefore, it makes sense to begin the determination of given objects on a scene with the determination of zones of interest.

In [1] this problem is solved in the case when the only initial information on a scene is its one image. In this paper we consider a general case when the scene interpretation is based on using a collection of several spatially juxtaposed images. Such collections are formed by multispectral optoelectronic systems. At each pixel of a scene they concurrently measure the electromagnetic radiation in several spectral zones. We treat the collection of images obtained in such a way as the vector image of the scene.

In the first section we propose a mathematical model of a scene. In the second section we determine zones of interest and study their properties. In the third section we describe the method used for determining zones of interest.

1. Vector locally homogeneous scenes

Let Z^2 - be an integer grid in a plane. We treat a ν-dimensional random variable $\xi_z = (\xi_z^j)_{1\le j\le \nu}$ as a pixel and do a countable family in the form $(\xi_z)_{z\in Z^2}$ as a vector scene. Each coordinate of the pixel ξ_z^j, $1 \le j \le \nu$, is a scalar random variable which takes on values in the set $Y = \{0,...,n\}$. It describes the j-th property of the pixel, and its value becomes known only after the measurement (photographing) of the scene. All random variables are defined on one and the same probability space (Ω, A, P).

Let $\omega \in \Omega$ and $x_z^j = \xi_z^j(\omega)$ for all $z \in Z^2$ and $1 \le j \le \nu$. For each j we call a map $x^j : Z^2 \to Y$ in the form $x^j(z) = x_z^j$ the j-th scalar image of the vector scene and write $x^j = (x_z^j)_{z\in Z^2}$. We understand the image of the scene as the collection $\mathbf{x} = (x^j)_{1\le j\le \nu}$.

We are mainly interested in objects rather than in separate pixels. Each object on a vector scene is defined by a finite set $A \subset Z^2$, which

is called its projection, and a family $\xi_A = (\xi_a)_{a \in A}$ of vector random variables with values in $\mathbf{Y} = Y^{\nu}$. Let $\xi_A^j = (\xi_a^j)_{a \in A}$, $1 \le j \le \nu$. Then we can equivalently represent the object as $\xi_A = (\xi_A^j)_{1 \le j \le \nu}$. Its image is the family $\mathbf{x}_A = (\mathbf{x}_a)_{a \in A}$, $\mathbf{x}_A \in \mathbf{Y}^A$ or $\mathbf{x}_A = (x_A^j)_{1 \le j \le \nu}$.

We assume that for the object ξ_A the family $\xi_A^j = (\xi_a)_{a \in A}$, $1 \le j \le \nu$, is a fragment of a homogeneous random field with the mean value m_A^j and the covariation function K_A^j. We understand the mean value $E\xi_A$ of the object as the family $\mathbf{m}_A = (m_A^j)_{1 \le j \le \nu}$. In what follows we assume that $K_A^j(z) \to 0$ as $|z| \to +\infty$. In accordance with the Slutsky theorem,

$$\bar{x}_A^j = \frac{1}{|A|} \sum_{z \in A} x_z^j \to m_A^j$$

in the sense of probability. That is, $\bar{x}_A^j$ is an unbiased and consistent estimate for m_A^j. We treat scenes which satisfy indicated conditions as vector locally homogeneous scenes.

2. Zones of interest on vector locally homogeneous scenes

Let us treat points z and t in Z^2 as neighbor points, if the Euclidean distance $d(z,t) = 1$. For a finite set $A \subset Z^2$ we denote by $Fr(A)$ the boundary consisting of points of this set which have neighbors in $Z^2 \setminus A$.

We treat an object ξ_A, whose projection is a simply connected domain, as a spot, if it has the following properties: there exists a square Q in Z^2 such that $A \subset (Q \setminus Fr(Q))$, the family $\xi_{Q \setminus A}$, which is called a neighborhood or a background of the object, is a fragment of a vector homogeneous scene, and, finally, $d(\mathbf{m}_A, \mathbf{m}_{Q \setminus A}) > 0$. If so then we call ξ_Q the zone of interest of the object ξ_A.

Let us describe properties of spots necessary for determining zones of interest. We understand a square neighborhood centered at a point *z* with radius *r* as the subset

$$B(z,r) = \{t \in Z^2 : |t_1 - z_1|, |t_2 - z_2| \le r\},$$

consisting of $n = |B(a,r)|$ points.

Theorem 1. Let ξ_Q be a zone of interest containing a spot ξ_A and its background $\xi_{Q \setminus A}$, and let $\mathbf{x}_Q$ be a vector image of ξ_Q. If $B(a,r) \subset A$ and $B(z,r) \subset (Q \setminus A)$ are disjoint neighborhoods such that each of them consist of *n* points, while

$$\bar{\mathbf{x}}_a = \frac{1}{n} \sum_{t \in B(a,r)} \mathbf{x}_t \quad \text{and} \quad \bar{\mathbf{x}}_z = \frac{1}{n} \sum_{t \in B(z,r)} \mathbf{x}_t,$$

then $d(\bar{\mathbf{x}}_a, \bar{\mathbf{x}}_z) \to d(\mathbf{m}_A, \mathbf{m}_{Q \setminus A})$ as $n \to +\infty$ in the sense of probability.

Theorem 2. Let ξ_Q be a square on a vector homogeneous scene allowing the application of the Slutsky theorem, and let $\mathbf{x}_Q$ be its image. If $B(z,r)$ and $B(t,r)$ are disjoint neighborhoods in Q consisting of *n* points, and

$$\bar{\mathbf{x}}_z = \frac{1}{n} \sum_{t \in B(z,r)} \mathbf{x}_t \quad \text{and} \quad \bar{\mathbf{x}}_t = \frac{1}{n} \sum_{z \in B(t,r)} \mathbf{x}_z,$$

then $d(\bar{\mathbf{x}}_t, \bar{\mathbf{x}}_z) \to 0$ as $n \to \infty$ in the sense of probability.

Theorem 3. Let ξ_Q be a zone of interest containing an object ξ_A and its background

$\xi_{Q\setminus A}$. If $B(z_j,r)\subset(Q\setminus A)$, $1\le j\le s$, and $B(a,r)\subset A$ are pairwise disjoint square neighborhoods consisting of *n* points, while

$$\overline{\mathbf{x}}_a=\frac{1}{n}\sum_{t\in B(a,r)}\mathbf{x}_t \quad \text{and} \quad \overline{\mathbf{x}}_j=\frac{1}{n}\sum_{z\in B(z_j,r)}\mathbf{x}_z,$$

then the probability

$$P(d(\overline{\mathbf{x}}_a,\mathbf{m}_{Q\setminus A})>d(\overline{\mathbf{x}}_j,\mathbf{m}_{Q\setminus A}),1\le j\le s)$$

tends to 1 as $n\to+\infty$ with each positive integer *s*.

Theorem 4. Let ξ_Q be a square on a vector scene obtained by moving summation over a square neighborhood with radius $\hat{r}$, and let $\mathbf{x}_Q$ be its image. If $B(z_j,r+\hat{r})$, $1\le j\le s$, and $B(z,r+\hat{r})$ are pairwise disjoint square neighborhoods in Q, each of which consists of *n* points, while $\overline{\mathbf{x}}_z$ and $\overline{\mathbf{x}}_j$ obey equalities

$$\overline{\mathbf{x}}_z=\frac{1}{n}\sum_{t\in B(z,r)}\mathbf{x}_t \text{ and } \overline{\mathbf{x}}_j=\frac{1}{n}\sum_{t\in B(z_j,r)}\mathbf{x}_t,\quad 1\le j\le s,$$

then

$$P(d(\overline{\mathbf{x}}_z,\mathbf{m}_Q)>d(\overline{\mathbf{x}}_j,\mathbf{m}_Q),1\le j\le s)\to_0$$

as $s\to+\infty$ with any fixed *r*.

3. Determination of zones of interest

Let ξ_A be some spot; denote its diameter by $d(A)$. Assume that for any square such that its side length $L\ge d(A)+2$ and $A\subset Q\setminus Fr(Q)$ the family ξ_Q is the zone of interest for ξ_A. Let an integer number l be such that $d(A)+2\le l\le L$, and let z_0 be a node in Z^2. Denote orts $e_1=(1,0)$ and $e_2=(0,1)$. Choose a number Δ (we call it the step length) and construct on Z^2 the family $Q(l,\Delta)$ of squares with the side length of l, where the left upper vertex z obeys the equality

$$\mathrm{z}=\mathrm{z}_0+i\,\Delta e_1+j\,\Delta e_2,\ i\in Z,\ j\in Z.$$

The following assertion is valid.

Theorem 5. If Δ satisfies the inequality $1\le\Delta\le l-d(A)-1$, then there exists a square Q' in $Q(l,\Delta)$ such that $\xi_{Q'}$ is the zone of interest for ξ_A.

Theorem 5 implies that the family $Q(l,\Delta)$ for each object contains at least one zone of interest. In order to exclude from $Q(l,\Delta)$ squares which are not zones of interest, we refer all squares in $Q(l,\Delta)$ to two classes. Let the first class consist only of zones of interest and let the second one contain all the rest squares. Let us divide the boundary $Fr(Q)$ to s disjoint parts Fr_j, $1\le j\le s$, each of which contains n points, and calculate

$$\overline{\mathbf{x}}_j=\frac{1}{n}\sum_{t\in Fr_j}\mathbf{x}_t \text{ and } \overline{\mathbf{x}}_F=\frac{1}{|F(Q)|}\sum_{t\in F(Q)}\mathbf{x}_t.$$

As an indicator of a square Q centered at $a\in Z^2$ with a side length of l we consider the value of the map

$$\hat{f}_{\{a\}}(\mathbf{x})=$$

$$\sum_{j=1}^{s}I_{]0,+\infty[}(d(\overline{\mathbf{x}}_a,\overline{\mathbf{x}}_{Fr})-d(\overline{\mathbf{x}}_j,\overline{\mathbf{x}}_{Fr})).$$

By definition, $\hat{f}_{\{a\}}(\mathbf{x})$ equals the number of estimates for $\overline{\mathbf{x}}_j$ satisfying the inequality

$d(\overline{\mathbf{x}}_a, \overline{\mathbf{x}}_{Fr}) > d(\overline{\mathbf{x}}_j, \overline{\mathbf{x}}_{Fr})$, that is, $0 \le \hat{f}_{\{a\}}(\mathbf{x}) \le s$.

In order to prove the efficiency of $\hat{f}_{\{a\}}$ we consider two classes of equal squares. Denote by Θ_1 the class consisting of zones of interest whose center coincides with the center of the neighborhood $B(a,r)$ that belongs to the corresponding projection of the spot. Denote by Θ_2 the class consisting of fragments of homogeneous random fields (that is, containing no zone of interest). Vector locally homogeneous scenes obtained by moving summation over a neighborhood with radius $\hat{r}$ have the following property.

Theorem 6. Let $P(\Theta_1)$ and $P(\Theta_2)$ be a priori probabilities, let Fr_j, $1 \le j \le s$, be connected fragments consisting of *n* pixels of the boundary $Fr(Q)$ such that $d(Fr_i, Fr_j) > 2\hat{r}$, and let $\hat{f}_{\{a\}}$ be the indicator of the square ξ_Q corresponding to r and s. Then for any $\varepsilon > 0$ there exist $s(\varepsilon)$ and $n(\varepsilon)$ such that the error probability $e(h^{s(\varepsilon)}_{n(\varepsilon)})$ of the decision rule

$$\hat{h}^{s(\varepsilon)}_{n(\varepsilon)}(\hat{f}_{\{a\}}(\mathbf{x})) = \begin{cases} 1, \hat{f}_{\{a\}}(\mathbf{x}) = s(\varepsilon) \\ 2, \hat{f}_{\{a\}}(\mathbf{x}) < s(\varepsilon) \end{cases},$$

classifying squares in Θ, does not exceed ε.

Let us illustrate the proposed method with results of determining zones of interest with the help of infrared images (Figs. 1--3) obtained in spectral zones of 0.7-1.1, 3.0-5.0, and 8.0-12.0 μm, respectively. See Fig. 4 for the result of zones determination.

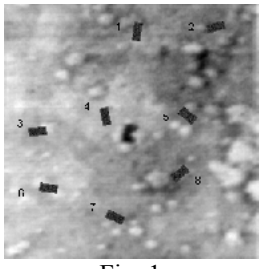


Fig. 1

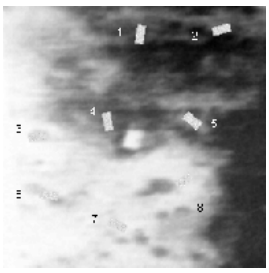


Fig. 2

Zone boundaries are colored green. For some objects we have obtained two zones, while six zones were classified erroneously.

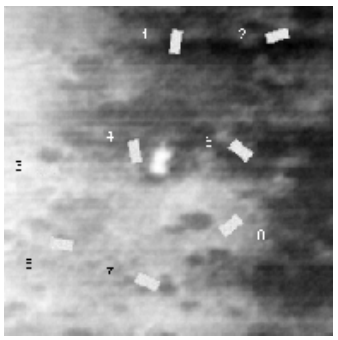


Fig. 3

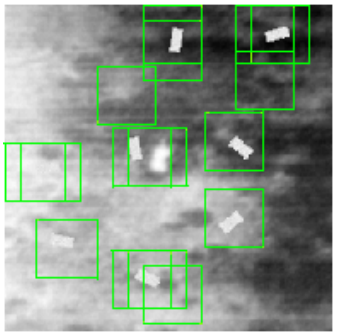


Fig. 4

Conclusion

We formally define the notion of a zone of interest on a vector locally homogeneous scene and propose a method for determining such zones with the help of spatially juxtaposed images.

References

1. R.M. Aleev, S.A. Martynov, V.B. Fofanov. Remarks on Searching Zones of Interest in Locally Uniform Scene // Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications. – 2012. - Vol. 22, No. 1. - PP. 150-156.

AUTOMATIZATION COLORIZE GRAYSCALE IMAGES BASED INTELLIGENT SCENE ANALYSIS[1]

A. Alekseev [2,3], V. Rozaliev[2,4], Y. Orlova[2,5]

[2] Volgograd State Technical University, Volgograd, The Russian Federation
[3]alekseev.yeskela@gmail.com, [4]vladimir.rozaliev@gmail.com, [5]yulia.orlova@gmail.com

This paper introduces a methodology for adding color to grayscale images based intelligent scene analysis. Towards this goal, we build on a technique that was recently developed to transfer colors from a user-selected source image to a target grayscale image with some modification. Search method will be shown a color source image. We propose improvements to existing solutions through intelligent scene analysis image. At the end we show results of colorize.

Introduction

Coloring grayscale images in a way that seems realistic to most human observers is a problem that has recently attracted renewed interest within the computer graphics community. The coloring problem amounts to replacing a scalar value stored at each pixel of a grayscale image by a vector in a multi-dimensional color space. This problem doesn't have a unique solution, because in the transition to grayscale image unsalvageable lost color information.

Today, there is some of software that allows coloring grayscale images. But in these systems humans must meticulously hand-color each of the individual [1] image regions. For example, Recolored system [2]. In addition, there are some articles that describe a fully automatic coloring [3][4], but they aren't used intelligent scene analysis.

Colorize technique

Fig. 1 illustrates the proposed technique to color a grayscale image using information from database of color images and database of objects. The technique consist of two steps: colorize overall part of image and colorize objects, if they are found. User can add and delete images of base of color images, but can't add or delete objects of base of objects, because to objects are put forward more stringent requirements.

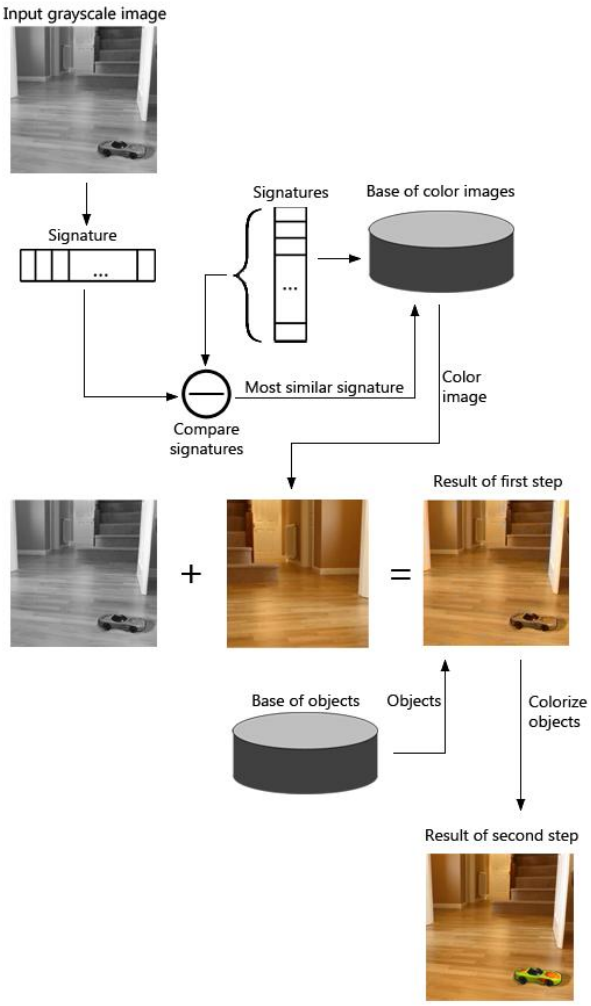


Fig. 1. Colorize technique.

Today, most of the existing systems stop at the first step. The second step is a key feature of the proposed technique.

[1] This work is supported in part by the Russian Foundation for Basic Research (the projects 12-07-00266-a, 12-07-00270-a, 13-07-00351-a, 13-07-00459-a, 13-07-00461-a).

Signature of image

Signature – 128 float values, obtained from normalized histogram of l channel lab color space. In [4] has been shown that 128 – optimum number of elements signature. Fig. 2 demonstrates a process for preparing signature.

Fig. 2. Signature of image

Two signatures are compared [5] on the basic of correlation (1) between them:

$$d_{correl}(H_1, H_2) = \frac{\sum_i^N H_1'(i) * H_2'(i)}{\sqrt{\sum_i^N H_1'^2(i) * H_2'^2(i)}}, \quad (1)$$

where H_1 and H_2 – signatures, N equals number of elements signature (N=128) and $H_k'(i)$ equals (2):

$$H_k'(i) = H_k(i) - \left(\frac{1}{N}\right)\left(\sum_j^N H_k(j)\right) \quad (2)$$

Base of color image

Fig. 3 illustrate adding image to base. The base contains color images and pairs of signature and name of the image. Name of image is unique. These pairs provide a fast search required image according its signature. User can add and delete image.

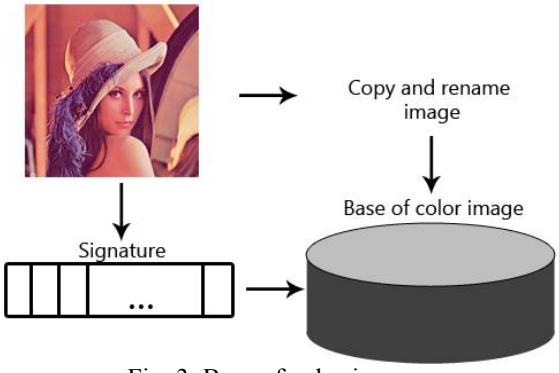


Fig. 3. Base of color image

Verification quality of the result

Method of verification based on structural similarity (SSIM) index [6]. The SSIM index is a full reference metric; in other words, the measuring of image quality based on an initial uncompressed or distortion-free image as reference, therefore we must do the following: for test color image get his grayscale variant and colorize it, then we get SSIM (3) for source color image and the result of colorize and multiply by 100%.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, (3)$$

where

μ_x – the average of x;

μ_y – the average of y;

σ_x^2 – the variance of x;

σ_y^2 – the variance of y;

σ_{xy} – the covariance of x and y;

$C_1 = (k_1 L)^2, C_2 = (k_2 L)^2$ two variables to stabilize the division with weak denominator;

L – the dynamic range of the pixel-values;

$k_1 = 0.01$, $k_2 = 0.03$ – constants.

Transfer color from source color image to target grayscale Colorize technique

Once an appropriate source color image has been selected for a given target grayscale image, we use the method of Welsh et al. [7] to transfer the chromatic information from the source to the target image, while keeping the luminance of the latter unchanged.

Welsh's method based compares small pixel neighborhoods on the luminance channels color and grayscale images. In small pixel neighborhood calculate the average and

variance. The modification is to change the size of neighborhood from 5x5 to 25x25.
Fig. 4. demonstrates a change SSIM index and time of coloring depending on the size of pixel neighborhood.

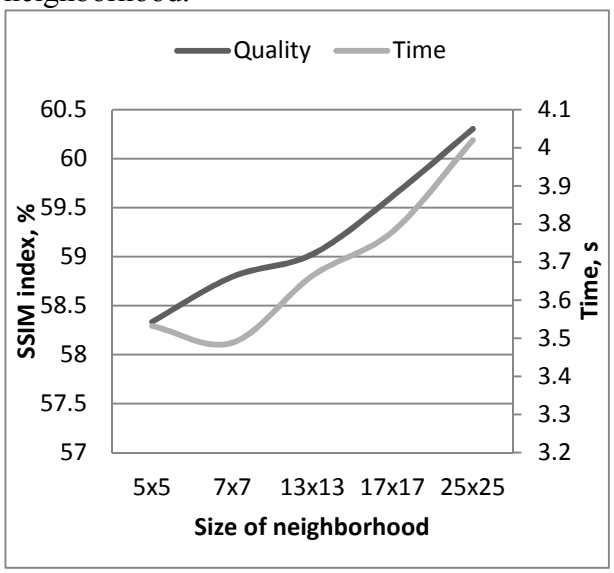


Fig. 4. The dependence the time of coloring and SSIM index of the pixel size of the neighborhood

Find objects on image

Object is a triple (4):

$$O = \{S, C, G\}, \qquad (4)$$

where
S – speeded-up robust features (SURF) interest points and descriptors;
C – Hu moment of main contour of object;
G – object color, it is values of l, a and b channels [8].
SURF is a robust local feature detector, first presented by Herbert Bay et al. in 2006. It uses an integer approximation to the determinant (6) of Hessian matrix (5). For features, it uses the sum of the Haar wavelet response around the point of interest.

$$H\big(f(x,y)\big) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}, \qquad (5)$$

$$\det(H) = \frac{\partial^2 f}{\partial x^2}\frac{\partial^2 f}{\partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y}\right)^2, \qquad (6)$$

where
H – The Hessian matrix for the two-dimensional function;
det(H) – determinant of Hessian matrix;
f(x,y) – arbitrary two-dimensional function.
With the SURF object is localized in the image. Then in this location we find all contours and find most similar Hu-moment. Moment (7) of contour is a gross characteristic of the contour computed by integrating over all of the pixels of the contour. There are the central (8), normalized (9) and Hu moments taken in the sum just displayed. The Hu invariant moments are linear combinations of the central moments. The idea here is that, by combining the different normalized central moments, it is possible to create invariant functions representing different aspects of the image in a way that is invariant to scale, rotation, and reflection [5].

$$m_{p,q} = \sum_{i=1}^{n} I(x,y) x^p y^q, \qquad (7)$$

where p is x-order and q is y-order, whereby order means the power to which the corresponding component is taken in the sum just displayed; I(x,y) – contour.

$$\mu_{p,q} = \sum_{i=0}^{n} I(x,y)\big(x - x_{avg}\big)^p \big(y - y_{avg}\big)^q \qquad (8)$$

where $x_{avg} = m_{10}/m_{00}$ and $y_{avg} = m_{01}/m_{00}$.

$$\eta_{p,q} = \frac{\mu_{p,q}}{m_{00}^{(p+q)/(2+1)}} \qquad (9)$$

As a result, objects on image localized based SURF method, then find contour based Hu invariant moments and colorize Welsh's method inside this contour.
Coloring objects on image allows improving the quality. On fig. 5 shows colorize image without colorize objects, index SSIM=77.58%.

Fig. 5. colorize image without colorize objects

On fig. 6 shows colorize image with colorize a car, index SSIM=79.62%. We received improving quality to 2% from this car.

Fig. 6. Colorize image with colorize objects

Results of colorize

On fig. 7 presents some results of colorize image, which were obtained by using software developed [9][10].

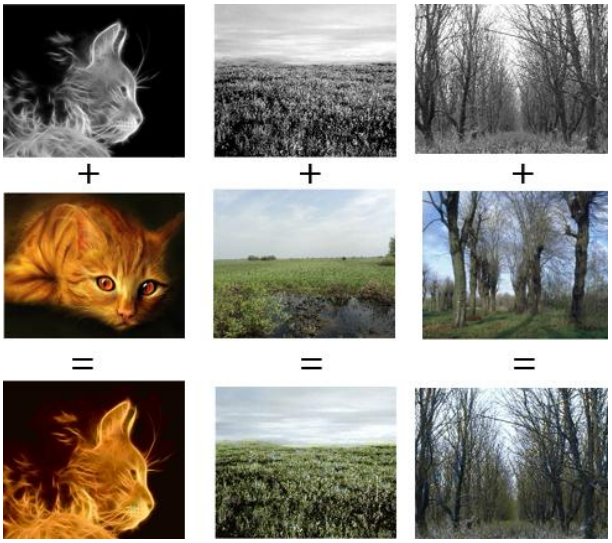


Fig. 7. Result of colorize

Conclusion

In this article we propose a methodology for adding color to grayscale images based intelligent scene analysis. Two steps of coloring allow improving quality as compared with only first step. The first step is based on Welsh's method transferring color between color and grayscale images. It has a major drawback: when a large number of different color tones of the image (e. g. photo of sport car) results can be severely incorrect. The second step must to fix this problem. It is based on the method of SURF and contour analysis. But coloring objects has a problem: a large number of objects are very time-consuming. To test the methodology developed a program [11], the article presented the results of work the developed software.

References

1. Rozaliev, V. L. at al. Application of neural networks and granulation in the construction of automated systems determine the emotional reactions of human / V. L. Rozaliev, A. S. Bobkov, O. S. Fedorov / Journal of VSTU. Series "Actual problems of management, computer science and informatics in technical systems" / VSTU. - Volgograd, 2010. Vol. 9. № 11. - P. 63-68.
2. Colorize Black and White Photos using Recolored. Retrieved from http://www.recolored.com/
3. Rathore, Y. at al. Colorization of Gray Scale Images using Fully Automated Approach // 2010. Vol. 7109. P. 16–19.
4. Vieira, L. F. M. at al. Fully automatic coloring of grayscale images // Image and Vision Computing. 2007. Vol. 25. № 1. P. 50–60.
5. Bradski G., Kaehler A. Learning OpenCV. 1st ed. / ed. Loukides M. Sebastopol: O'Reilly Media, 2008. P. 556.
6. Wang, Z. et al. Image quality assessment: from error visibility to structural similarity. // IEEE transactions on image processing: a publication of the IEEE Signal Processing Society. 2004. Vol. 13, № 4. P. 600–612.
7. Welsh, T., Transferring color to greyscale images / Welsh T., Ashikhmin M., Mueller K. // ACM Transactions on Graphics. 2002. Vol. 21. № 3. P. 277–280.
8. Bay H. et al. Speeded-up robust features (SURF) // Computer Vision and Image Understanding (CVIU). 2008. Vol. 110, № 3. P. 346 – 359.
9. Orlova, Y. A. Analysis of models and methods for increasing the efficiency of software design / Y. A. Orlova // Journal of VSTU. Series "Actual problems of management, computer science and informatics in technical systems" / VSTU. - Volgograd, 2010. Vol. 9. № 11. - P. 137-141. (in Russian)
10. Zaboleeva-Zotova, A.V. Automation of the initial stages of software design / A. V. Zaboleeva-Zotova, Y. A. Orlova // Journal of VSTU. Series "Actual problems of management, computer science and informatics in technical systems" / VSTU. - Volgograd, 2010. Vol. 8. № 6. - P. 121-124. (in Russian)
11. Orlova, Y. A. Algorithmic support of specification text analysis and modeling software / Orlova, Y. A. // Journal of VSTU. Series "Actual problems of management, computer science and informatics in technical systems" / VSTU. - Volgograd, 2010. – Vol. 8. № 6. – P. 68-72 (in Russian)

AUTOMATIC IMAGE MOSAICING AND SUPER RESOLUTION WITH SIFT AND RANSAC

L. L. Almeida[1,2], F. A. Silva[1], M. S. V. Paiva[2], L. A. C. Jorge[3]

[1] University of Western São Paulo (Unoeste), Pres. Prudente-SP, Brazil
llalmeida@unoeste.br, chico@unoeste.br
[2] São Carlos Engineering School (USP), São Carlos-SP, Brazil
mstela@sc.usp.br
[3] Brazilian Agric. Research Corporation (Embrapa), São Carlos-SP, Brazil
lucio.jorge@embrapa.br

The objective of this work is to develop efficient algorithms, robust and automated fusion image frames for mosaics and super resolution from the selection of a region in the mosaic. Image registration is a fundamental step in combining several images that make up the scene. Our research is based on the determination and extraction of characteristics defined by the SIFT algorithm (detector / descriptor) and RANSAC. Validations real data show the effectiveness of our techniques proposed. The difference of this work is the way to get the matching and merging of images because it occurs dynamically between elements common images that are stored in a sparse matrix.

Introduction

The development of techniques that enable the generation of high resolution images from a image sequence has been useful in several applications, such as for vehicle license plate identification [1], facial identification of possible suspects (criminals, for example) in images from secury systems [2][3], the generation of high-resolution aerial images from video pictures or digital camera, to determine diagnoses through medical images [4] or even the detection of diseases in their early stages. In astronomy, it can reveal new stars or better identify them by algorithms applied to these images, and also reconstruct old films [5][6], minimizing the amount of noise and blurring that appear in each frame designed. In all applications mentioned, it is essential to have visualization with a higher level of detail from one or more specific areas of interest in an image; this is essential to obtain high resolution. However, high-resolution images are not always available; this motivates the investigation of super-resolution techniques. Super-resolution is the process of obtaining a high resolution image from one or more low resolution images [3].

Digital images can not present a radiometric resolution (sharpness) and or geometric (size) satisfactory for analysis, whether due to natural factors such as excessive light, or even by the characteristics of the mechanism being used to obtain these images. It is possible that these images of poor quality have been reused, i.e. from a sequence of images of poor quality. It is possible to obtain a quality superior to those that are being analyzed. The brightness level of each image point of highest resolution is determined by the fusion of the corresponding points in other images, all transformed (rectified).

In this paper, we present an algorithm for the automatic construction of mosaics of high-resolution digital images. This is an area of current research in the fields of photogrammetry, computer vision, digital image processing and computer graphics. For the generation of a high resolution image from aerial images, which compose a bigger scene, evidenced in a mosaic, the correlation between the various images contained in the mosaic was used in this study, where one of the images was defined as the reference image for each process generation.

Mosaics of images can be used in several applications, even in biological images (mainly for microscopic images) [5], weather forecasts

and the construction of large collections of aerial and satellite images to control large areas of crops/pastures/forests [6]. The generation of high-resolution images based on a sequence requires first that the images belonging to the scene be rectified, i.e., the image registration must have been executed. The image registration is the process of overlapping two or more images of the same scene taken at different times, from different viewpoints, and or by different sensors. The process geometrically aligns two images. From this, you can create a new image with greater precision, that may be used for more detailed analyses, as evidenced in several studies related to this line of research [1].

The remaining sections of this paper are organized in the following way: in next section, we presented the techniques used for automated image registration used in this work. After, we present the methodology used in the work for the generation of super-resolution images. Finally, we present the experiments, results and discussion.

Techniques used for automated image registration

The match is the process adopted to accomplish the registration of two images, and for this, it is necessary to identify the region of overlap between them. This is done by finding common points. The initial difficulty is to find the keypoints in the first image, which can be located in the second image. One possible approach, also used in this study, applies the algorithm SIFT [4] to find points of interest in the two images, and then uses the RANSAC algorithm [4] to remove incompatible points.

SIFT Algorithm

This algorithm consists in a very efficient method to identify and to describe image keypoints, which is done by performing a mapping with different views of an object or scene, resulting in a vector with 128 values that describes each image keypoint. The algorithm consists of the following steps:

Scale-space extrema detection: The keypoints are detected applying a cascade filtering that identifies candidates that are invariant to scale. The scale-space is defined as a function $L(x,y,\sigma)$ in Equation 1, with an input image $I(x,y)$ [4].

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \quad (1)$$

where * is a convolution in x and y with the Gaussian $G(x,y,\sigma)$ in Equation 2.

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (2)$$

In order to detect stable keypoint locations in space-scale, Lowe [4] proposed the use of space-scale extrema in the difference-of-Gaussian (DoG) function convolved with the image $I(x,y)$, resulting in $D(x,y,\sigma)$, which can be computed from the difference of two nearby scales separated by a constant multiplicative factor k, as in Equation 3.

$$D(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) \quad (3)$$

The DoG is an approximation of the scale-normalized Laplacian of Gaussian $\sigma^2\nabla^2 G$ [4]. The maxima and minima of $\sigma^2\nabla^2 G$ produce the most stable image features.

Local extrema detection: from $D(x,y,\sigma)$, in [4], it is suggested that the local maxima and minima must be detected by comparing each pixel with its eight neighbors in the current image and nine neighbors in the scale above and below (26 neighbors). SIFT guarantees that the keypoints are located at regions and scales of high variations, which make these locations stable for characterizing the image.

Orientation assignment: The scale of the keypoint is used to select the Gaussian smoothed image L, with the closest scale, so that all computations are performed in a scale-invariant manner. The gradient magnitude $m(x,y)$ is computed with Equation 4.

$$m(x,y) = \sqrt{\Delta x^2 + \Delta y^2} \quad (4)$$

where $\Delta x = L(x+1, y) - L(x-1, y)$ and $\Delta y = L(x, y+1) - L(x, y-1)$. The orientation $\theta(x,y)$ is calculated by Equation 5.

$$\theta(x,y) = \arctan(\Delta y / \Delta x) \quad (5)$$

Keypoint description: The next step is to compute a descriptor for the local image region that is distinctive and invariant to additional variations, such as change in illumination or 3D viewpoint. In [4] is suggested that the best approach is to determine the magnitudes and directions of the gradients around the keypoint location. In this

approach the Gaussian image on the keypoint scale is used.

In order to find the match between two images it is possible to use the keypoints detected with the SIFT algorithm. In [5][6] is proved that the best match for each keypoint is found by identifying its nearest neighbor, which is defined by minimizing the Euclidean distance to the features vectors. To avoid an exhaustive search in the use of a data structure k-d tree that supports a balanced binary search is suggested to find the closest neighbor of the features and the heuristic algorithm Best-Bin-First (BBF) is used for the search.

RANSAC (RANdom SAmple Consensus) algorithm proposed by Xing and Huang [3] is a robust estimation method designed to identify the inliers[1] and outliers[2] from the set of keypoints detected by the SIFT algorithm. RANSAC is widely used for object recognition [3][6].

Methodology

This section describes the methodology used in the study to accomplish the generation of super-resolution images from sequences of images that form a mosaic of images (Fig. 1). The implementation of this methodology favored the validation process of the elements used to obtain the image in question.

We describe four important steps for the process application:

a. Collection of Images: The images were taken by digital cameras.[1]

b. Registration and Rectification: SIFT was used to determine the matching points between the images. After the parameters were determined, there was simultaneous rectification and resampling of the images that belong to a sequence relative to reference image. Based on the parameters that allowed the recording of images (detected and extracted by the SIFT and selected by RANSAC), we determined the matching between the images of interest. As the coordinates of the images are projected on the pixel fractions, it was necessary to perform an interpolation between the neighbors in order to determine the brightness level relative to what is desired to determine the super-resolution image on the basis of a real and not just calculated increase in resolution.

c. Fusion: the mosaic generation step is accomplished by merging the corresponding pixels in the rectified images belonging to the scene, with respect to the reference image. In parallel with this process, it was necessary to build data structures based on dynamic lists that merge to form a similar model for a sparse matrix. This process resulted in the labeling of each point of the original image mosaic to which it belongs, in order to generate a higher resolution image based on the original sequence. The implementation done allows these data structures to be expanded in any direction, displacing the original data and its indexes according to the insertion request.

d. Super-resolution: the last step is obtained by selecting the area of interest in the mosaic generated. From this, a super-resolution image is generated based on the data linked to the images that belong to the desired area.

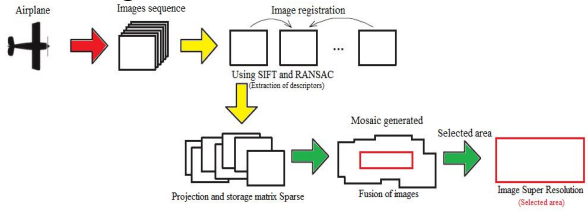


Fig. 1: Methodology for generating image super resolution from the selection of an area of a mosaic.

Steps to generate the images from mosaics

In this step it is necessary to label each point of the original image to which it belongs, to make it possible to generate a higher resolution image based on the original sequence, leading to formation of the tiles.

The structures used are described below, starting with the data structures to carry out the rectification in memory of each pixel, which was a sparse matrix (Fig. 2).

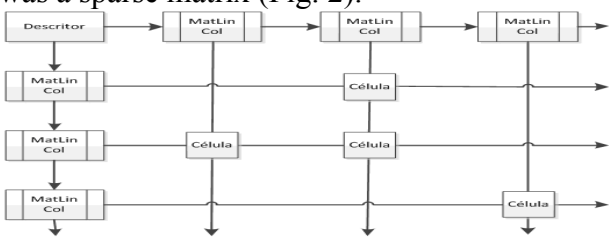


Fig. 2. Sparse matrix for storing the pixels of the images that belong to the mosaic.

Important fields that store elements are capable of accessing any element that is part of this structure.

[1] *inliers*: data points that fit a particular model within an error tolerance.

[2] *outliers*: data points that do not fit a particular model within an error tolerance.

The construction of the sparse matrix is made during the process of generating the panoramic image.

Results and Discussion

This section presents the experiments performed and validated by the tool based on the methodology adopted. The implemented system was tested with simulated data and real data. Several experiments were done with color images during the work, to determine the best sequence for the application of methods to generate an image with better resolution from image sequences that form a mosaic of images acquired with the purpose of extracting detailed information from the designed region. In the following subsections, we present some experiments that demonstrate the efficiency of the process.

Experiment - Sequence of urban aerial images

The set of aerial images used (300x300 pixels) is related to urban images with overlap, thus making possible the construction of the mosaic from those images for subsequent selection of a part of the same, in order to generate higher resolution images from the selected region. Fig. 3 shows the set of images analyzed.

Fig. 3. Sequence of urban images.

Fig. 4 shows the mosaic generated that also contains a storage structure of the points of overlap between the images that compose the mosaic.

Fig. 4. Image mosaic generated from the set of images of Fig. 3.

Fig. 5 shows the image super-resolution with twice the size of the selected part from the mosaic and generated based on real data from images that compose this area of the mosaic.

Fig. 5. Selected and extracted image after selecting the chosen region.

Fig. 6 (a) shows a extrapolation of the radiometric resolution of the original images used for generating the mosaic, while (b) shows the extrapolated resolution of the super resolution image obtained. It is evident that the serration existing in (b) was attenuated due to the use of real data of existing overlap, proving the efficiency of the proposed method.

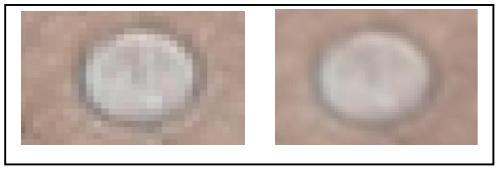


Fig. 6. Portions extrapolated from Fig. 2 and 3, respectively.

References

1. Y. Tian, K.-H. Yap, and Y. He, “Vehicle license plate super-resolution using soft learning prior”, Multimedia Tools and Applic, pp. 1-17, 2011.
2. Y. Tian, K.-H. Yap, and Y. He, “Vehicle license plate super-resolution using soft learning prior”, Multimedia Tools and Applic, pp. 1-17, 2011.
3. C. Xing, J. Huang, “An improved mosaic method based on SIFT algorithm for UAV sequence images”, IEEE Int Conf on Computer Design and Applications (ICCDA), p. 414-417, 2010.
4. D.G. Lowe, “Distinctive image features from scale-invariant keypoints”. Int Journal of Computer Vision, Vol. 60, No. 2, pp. 91–110, 2004.
5. H. Nasir, V. Stankovic, S. Marshall, “Singular value decomposi-tion based fusion for super-resolution image reconstruction”, Int Conf on Signal and Image Proc Applications (ICSIPA), 2011 IEEE, pp. 393-398.
6. Z. Dakun, L. Zhaoxin, J. Guiyuan, “A feature-based al-gorithm for image mosaics with moving objects, IEEE Int Conf on Intelligent Computing and Int Systems (ICIS), p. 26-29, 2010.

OPTIMAL FEATURES FUSION FOR SEMANTIC SEGMENTATION[1]

S. Anishchenko[2], M. Petrushan[2]

[2] A.B. Kogan Research Institute for Neurocybernetics, Southern Federal University, 194/1, Stachki Ave, 344090, Rostov-on-Don, Russia, sergey.anishenko@gmail.com

The new bottom-up semantic segmentation method with task-specified adaptation is proposed. Segmentation is performed on fused feature map, creating as the weighted sum of hue, saturation, and value components from HSV colour space. Similarity criterion for pixels grouping and maps fusion weights were optimized by particle swarm optimization (artificial bee colony). Method performance was evaluated on the facial image segmentation task. Segmentation performed on the maps mixture demonstrated better accuracy than single map-based segmentation, i.e. it contains the segment that fit the face area rectangle, detected by Viola-Jones method.

Introduction

Image and video segmentation is a computer vision branch aims at image separation into areas that have to fulfill certain conditions. In the case of semantic segmentation these conditions require the correspondence between segments and objects, i.e. the automatic segmentation have to be the same as manual one could be.

Within bottom-up approach [4] the segmentation is based on grouping of similar pixels in terms of low-level features, such as colour or brightness. The lack of enough information at local scales to detect semantically important regions is mentioned as the main disadvantage of bottom-up segmentation [2]. Such additional information about objects comes from top-down adaptation [3] in a form of constraints or after learning procedure. Several segmentation algorithms are based on combination of bottom-up segmentation and top-down presetting [5].

One of the fundamental problems within bottom-up segmentation is the formalization of pixels similarity criterion that can vary depending on the segmentation type: background subtraction, cosegmentation, objects separation, etc.

Background subtraction aims to separate constant background and moving foreground object in video sequence [7]. Homogeneity (pixel similarity) criterion for this segmentation type is based on features constancy during a period of time. Cosegmentation is a joint segmentation of images sequence in order to detect similar objects in them. Cosegmentation methods are based on background modeling and object classification algorithms [10].

Another fundamental segmentation problem is concerned with a selection of primary features for pixels grouping, for example grouping can be based on gray-scale intensity similarity, colour similarity, saturation similarity, etc., i.e. segmentation can be performed on different feature maps, like intensity map, hue map or gradient map.

We propose the new bottom-up semantic segmentation method, where homogeneity criterion and feature map are selected in top-down model construction procedure. Within our approach, instead of colour or intensity map processing, the segmentation is performed on the maps mixture, creating as the weighted sum of different feature maps. Hue, saturation, and value maps from HSV colour space are used in the mixture with different weights, optimized in order to provide the best object-related segmentation. This work extends segmentation approach suggested earlier in [9] where efficiency of simultaneous usage of components of various colour spaces has been demonstrated.

[1] Support of the Russian Foundation for Basic Research, grants N 12-01-31226 mol_a, N 12-01-31266 mol_a

The method was assessed on the facial image segmentation task. Similarity criterion parameters and maps fusion weights were optimized by swarm optimization method [1] to provide such separation where one of the segments fits to the face area rectangle, detected by Viola-Jones method [8].

Segmentation algorithm

The developed segmentation algorithm is based on the region growing technique. According to the taxonomy suggested in [6] the proposed algorithm relates to the image-domain based technique because it tries to satisfy both feature-space homogeneity and spatial compactness at the same time.

The main idea of the algorithm is to choose seed point and select all connected pixels satisfied the homogeneity criteria. These two steps iteratively repeat until the whole image is separated into homogeneous segments.

Seed point selection procedure starts from the origin of image coordinate system. The algorithm searches for the pixel which satisfies the following criteria. It must not belong to both overthresholded gradient map and previously segmented areas. This condition is summarized in (1).

$$S_n \notin (\Delta I > T) \bigcup_{i=1}^{n-1} A_i \qquad (1)$$

where S is the seed point, n – number of algorithm iteration, ΔI is image gradient, T – gradient magnitude threshold, A – segmented areas.

When seed point is found the flood fill algorithm is used to select all connected pixels which satisfy the homogeneity criteria (2) in the feature space.

$$P' - l < P < P' + u \qquad (2)$$

where P is the value of the currently observed pixel, P' is the value of one of the pixel neghbours, l – lower difference, u – upper difference.

In current implementation the feature space is one dimensional. In this work the colour components from HSV space and mix of these components have been used for experiments.

Feature spaces for segmentation

The developed segmentation algorithm takes into account pixels spatial compactness as well as homogeneity of pixels value in feature space. Any one-dimensional feature space can be used in the current implementation of algorithm. In this work four spaces have been examined to choose the most optimal one for semantic image segmentation:

1) Hue (H) from HSV,
2) Saturation (S) from HSV,
3) Value (V) from HCV,
4) Fused colour components (M) of HSV space (3).

$$M = k_1 H + k_2 S + k_3 V \qquad (3)$$

where k_1, k_2, k_3 – fusion weights, H, S, V – components of HSV colour space.

Video database

In this work two video sequences have been used. The first (n=181 frames) is for the optimization and second (n=40 frames) – for testing. Both video clips have been captured by off-the-shelf web-camera and contain one person sitting near computer (Fig. 1).

Fig. 1. Example of the frame from the video used for the parametric optimization of the segmentation algorithm.

Video was taken in average office condition with daylight source (window) located on left side of person. During capturing the second video the extra light source (fluorescent lamp) moving around the person has been used (Fig. 2).

Fig. 2. Example of the frame from the video used for testing the optimal parameters. There is the extra light source (fluorescent lamp) on the left side of person. Position of lamp varies within the video.

Parameters optimization

The aim of the conducted experiments is to optimize segmentation algorithm parameters to partition image into semantically meaningful areas, i.e. some of the segments have to fit a real world object. A face has been chosen as a target object. Four experiments differing by used feature space have been conducted.

The first video sequence was used for optimization of segmentation algorithm and colour components fusion weights. In particular the following parameter of segmentation algorithms varied during optimization:

1) lower difference l (2),
2) upper difference u (2),
3) fusion weight k_1, k_3, k_3 (3).

The particle swarm algorithm [1] has been used for optimization. Parameter l varied in the range [1; 20], u - [1; 20], k_1, k_3, k_3 - [0;1].

The cost function D was computed during optimization as sum of frames cost function D_f. D_f computation utilized information about face area detected by Viola-Jones algorithm. The frame was divided by segments using the developed segmentation algorithm. D_f was computed for each segment. The only one segment was selected which is the best fitted to the face area. It is done by searching for the minimum D_f over the frame segments (4).

$$D_f = \min_n[(S_n^{out} - kS_n^{in})/S_f] \qquad (4)$$

where n – a number of image segments, S^{out} – a number of pixels outside of the facial rectangle, S^{in} – a number of pixels inside of the facial rectangle, S_f – an area of the facial rectangle, k – an extra penalty coefficient equal to 5. This coefficient controls a balance between false positive (segment is outside of target object) and false negative (unconnected pixels inside the target object) segmentation.

Results

As a result of experiments the set of algorithm parameters (2) and colour components fusion weights (3) have been computed within optimization procedure based on the first video sequence (Fig. 1). The parameters are summarized in Table 1.

Table 1. The segmentation algorithm and colour component fusion parameters achieved by optimization procedure

Parameter	H	S	V	M
L	1.12	15.18	5.56	11.79
U	20	9.29	20	2.07
K_1				0.78
K_2				0.02
K_3				0.35

The testing of algorithm with optimal parameters has been done utilizing the second video sequence (Fig. 2, 3.a). To quantitatively estimate segmentation performance the cost function D_f for each frame was computed by the same manner as in the case of optimization (4). The average and variance of D_f over the video sequence are shown in Table 2 (variance is in brackets) along with the same values for the first video used in optimization.

Table 2. Result of optimization and testing

Experiment	*H*	*S*	*V*	*M*
Optimization	-3,01 (0,011)	-0,01 (0,699)	-2,23 (2,932)	**-3,14** (0,333)
Testing	-3,02 (0,019)	0,87 (0,179)	-1,824 (3,957)	**-3,34** (0,009)

It can be seen from the Table 2 that the minimum of cost function (bolded) was achieved for the feature space constructed by the fusion of colour components (*M*). Qualitatively this result can be seen in Fig. 3. It is obvious that segmentation result is better in the case of fused colour components (Fig 3.c).

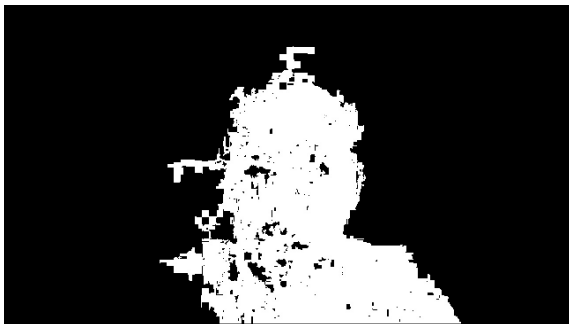
Conclusion

The seed growing semantic segmentation algorithm is proposed. Four feature spaces including fused ones have been investigated to achieve efficient semantic bottom-up segmentation. The optimal parameters of the segmentation and features fusion have been computed in the optimization experiments. Optimization was done utilizing video captured with constant light source. The testing has been performed using the video captured with moving light source.

It has been shown that the face area can be efficiently segmented using fused features map. In particular the mix of components of

HSV colour space has demonstrated the best efficiency.

a) 

b) 

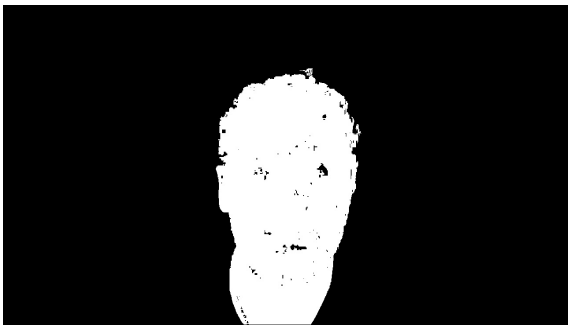
c) 

Fig. 3. Example of the segmentation of the video captured with extra lighting; a) – frame, b) – segmentation based on hue (*H*), c) – segmentation based on mixture of colour components (*M*).

References

1. B. Basturk and D. Karaboga. An Artificial Bee Colony (ABC) algorithm for numeric function optimization // In Proc. IEEE Swarm Intelligence Symp. - 2006.
2. David Marr. Vision: A Computational Investigation into the Human Representation and Pro-cessing of Visual Information // NY, USA: Henry Holt and Co., Inc. - 1982.
3. E. Borenstein and S. Ullman. Combined Top-Down/Bottom-Up Segmentation // IEEE Transactions on Pattern Analysis and Machine Intelligence. - 2008. - P. 2109 –2125.
4. H.G. Barrow and R.J. Popplestone. Relational Descriptions in Picture Processing // Machine Intelligence. - 1971. - Vol. 1 - P. 377–396.
5. J. Carreira and C. Sminchisescu. Constrained Parametric Min-Cuts for Automatic Object Segmentation // IEEE Conference on Computer Vision and Pattern Recognition. - 2010.
6. Luccheseyz, L., and S. K. Mitray. Color image segmentation: A state-of-the-art survey // Proceedings of the Indian National Science Academy (INSA-A). Delhi, Indian: Natl. Sci. Acad. – 2001. – P. 207-221.
7. M. Piccardi. Background substraction techniques: a review // Computer Vision Research Group (CVRG), University of Technology, Sydney (UTS). - 2004.
8. Paul A. Viola, Michael J. Jones. Robust Real-Time Face Detection // International journal of computer vision. - 2004. - Vol. 57, No. 2. - P. 137-154.
9. S. Anishenko, D. Shaposhnikov, R. Comley, X. Gao. A colour based approach for face segmentation from video images under low luminance levels. // In Proc. of the 11th IASTED International Conference on Computer Graphics and Imaging (CGIM 2010). – 2010. – P. 184-189.
10. Vicente, Sara, Carsten Rother, and Vladimir Kolmogorov. Object cosegmentation // Computer Vision and Pattern Recognition (CVPR). - 2011.

A NEW EFFECTIVE METHOD FOR SPATIAL DENSITY OF DISCRETE POINTS CLOUD RECONSTRUCTION AND ITS IMPLEMENTATION IN COMPUTER 3D GRAPHICS[1]

W. Antciperov[2,3], O. Evseev[2,4], Yu. Obukhov[2,5]

[2]Kotel'nikov Institute of Radioengineering and Electronics, Russian Academy of Sciences, Moscow, Russian Federation,
[3]antciperov@cplire.ru , [4]ev.mipt@gmail.com , [5]obukhov@cplire.ru

Report presents results of a new method of parametric approximation for the purposes of computer reconstruction of three-dimensional distributions of discrete micro-objects that are distributed along a series of two-dimensional cross-sections of the cloud of points. In the center of report is the discussion of the role of modeling and the actual models in problems of 3D neuron distribution visualization, especially in the case of large arrays of primary microscopy data. The article compares conceptual backgrounds and parts of the technical implementation of the proposed method and well-known three-dimensional reconstruction lofting method.

Introduction

The current state of computer science and technology let us significantly change traditional approaches to scientific research and engineering. Computer models and virtual 3D-graphics objects systematically substitute now real physical models and prototypes. And it is inevitable, since working with digital models provides us with high level of flexibility, a large number of options and quick critical characteristics calculation. Not the least is the significant resource saving (time, material and labor costs) that also can be achieved.

Models, used in the research and development, should adequately approximate those real or proposed objects, which they are intended for. There is a lot of model constructing methods and they vary in a wide range from analytical methods for creating theoretical models up to experimental ones for creating simulators that completely numerically repeat measurements of real prototypes. The most promising are synthetic methods, which optimally, in accordance with some theoretical "frame", interpolate relatively small amount of available measurements. Multitude of these methods includes three-dimensional reconstruction techniques – the subject of study.

These methods are used in many fields of applications. The most advanced of them are: creating CAD/CAM reengineering concepts, for instance in the automotive industry; planning of surgical prosthetics; making models of buildings, urban fragments and landscape simulation structures. It is worth noting also the use of three-dimensional reconstruction techniques in the model industry (custom fitting), virtual museums (virtual museum), etc.

Currently, three-dimensional computer reconstruction methods are quickly implemented in biomedical applications. The most popular these methods appeared in the problems of visualization and interpretation of the low resolution data gained from virtual cross sections, for example, in confocal microscopy, X-ray or CT and magnetic resonance imaging [1]. In such areas the three-dimensional reconstruction from rigidly positioned two-dimensional cross-section images become now a common tool of diagnostics and research [2].

[1] Support of the grant RFBR № 12-07-00104

Suggested method backstory

Recently we have developed a prototype of software and algorithmic complex for the three-dimensional neurons distribution visualization. This development was based on the concepts of reengineering and computer modeling and was the main content of the project "Research and development of software and algorithmic methods for neurons automatic recognition and reconstruction of their three-dimensional distribution" funded by the RF Ministry department of science and education. The purpose of this project was to automate the process of counting surviving neurons and the analysis of their spatial distribution in different schemes of Parkinson's disease modeling. The rough experimental data were presented by coordinates of neurons, obtained in the process of digital microscopic images of mouse brain slices analysis [3].

This software complex was designed as a modular system, where each module contains a set of algorithms intended for the tasks of a certain reconstruction process step (pipeline). The main steps of interactive reconstruction were: alignment of adjacent slices; construction of a general global three-dimensional coordinate system that is internally connected to the sample material; recognition of neurons in the images of initial slices, the formation of planar neuron distributions based on the EM-clustering, and 3-D visualization based on parametric 3D-surfaces graphics DirectX-oriented approximation algorithms [4].

Let us here briefly describe only the basic ideas that formed the basis of our 3-D visualization method, without lingering on the others mostly technical steps. The method is focused on the synthesis of a priori, delivered by the generalized model of Parkinson's disease model, and a posteriori information representing experimental data in the form of specified parametric distributions. The result of the method was to be a three-dimensional density neurons' distribution that approximates recognized neurons in the microscopic images. The developed system presents this distribution as 3D-surfaces of constant neurons' density level interpolating the family of two-dimensional contours to the space between sections. Interpolation is realized on the base of approach that is conceptually close to the well-known lofting method [5]. As loft-objects (L-objects), 3D-surface is constructed by forming a shell on the series of cross-sections curves arranged along a certain path. However, in contrast to the classical lofting method, profile curves are defined by parametric closed contours obtained from section of exponential family mixture (EM-clustering result) at specified density level. Instead of interpolation on space between sections of these contour lines, we, in method proposed, utilize the interpolation of distribution parameters. For this reason, we called it as parametric interpolation based reconstruction.

Known approaches - conventional lofting methods

One of the most widely used methods of three-dimensional reconstruction is a classical lofting (or skinning) method [5]. A set of cross-sections curves and path (along which they are located) are given in the form of splines. In most applications, splines are B-splines that are expressed as linear combinations of basis B-splines - finite functions represented by polynomials of p degree on the support:

$$C(u) = \sum_{j=0}^{n} B_{j,p}(u) P_j \qquad (1)$$

where $C(u)$ – u-parameterized spline, $\{ B_{jp}(u) \}$ – B-spline basis.

A set of $n+1$ control points $\{P_j\}$ in (1) explicitly defines the shape and position in space spline $C(u)$, which approximates this sequence and needs not pass through any of these points. In other words, $C(u)$ in (1) is a such curve with prescribed smoothness that most closely approximates the spatial jogged line through points $\{P_j\}$. The main feature of B-splines is their locality: a change in one of the control points P_i leads to $C(u)$ deformations only in the immediate vicinity due to the finiteness of the basic B-spline $B_{ip}(u)$.

Three-dimensional reconstruction spline-based lofting method is realized by following steps:

1. First build B-spline curves in each cross-section of the L-object and compute parameterization knot vector.
2. Then, bring B-spline curves to a common degree and knot vector.
3. And, finally, carry out interpolation of the B-spline curves towards the path of loft (called longitudinal direction), i.e. construct a two-parametric B-spline surface $S(u,v)$:

$$S(u,v)=\sum_{i=0}^{m}\sum_{j=0}^{n}B_{j,p}(u)B_{i,q}(v)P_{i,j}\,. \qquad (2)$$

Where p degree B-splines in the cross-sections (u parameterization) and q degree B-splines along the longitudinal direction (v parameterization) may be different, though, in practice, p and q are equal. Matrix of control points $\{P_{ij}\}$ in (2) uniquely specifies the location of a surface in space.

On the basis of above-mentioned, we can imagine surface lofting as a result of contour B-spline varying curve shifting upon the other B-spline curves along the L-object axis. This circumstance causes B-spline surfaces to inherit a lot of one-dimensional spline features.

Formation of B-spline surfaces by lofting methods (2) is widely used in modern CAD-software. This method is very flexible and is able to create complex surfaces, meeting requirements of high smoothness as well. Nowadays most professional graphic software suites use non-uniform rational B-spline (NURBS) – the special case of B-spline curves.

Suggested approach - parametric approximation method

Being remarkable, classical lofting method, however, may turn out inefficient in practice due to huge amount of calculations. Inefficiency is more evident in case when contours in each section can be analytically expressed as relatively simple curves, defined by few parameters. Just the case is 3D-reconstruction based on modeling.

In our project of neuron density distribution 3D-reconstruction, the function of neuron distribution density $\rho(\vec{x})$ was modeling as mixture of exponential distributions:

$$\rho(\vec{x})=\sum_j n_j\rho(\vec{x}\,|\,\vec{\Theta}_j), \qquad (3)$$

where $\vec{x}$ – 2D-vector in section, n_j – number of neurons and $\vec{\Theta}_j$ – parameters of normalized neuron distribution density $\rho(\vec{x}\,|\,\vec{\Theta}_j)$ in j-th cluster. Density affiliation with parametric exponential family (EF) assumes that $\rho(\vec{x}\,|\,\vec{\Theta}_j)$ looks as:

$$\rho(\vec{x}\,|\,\vec{\Theta})=\frac{1}{Z(\vec{\Theta})}\exp\left(\sum_i\Theta_i u_i(\vec{x})\right)h(\vec{x}), \qquad (4)$$

.

where $h(\vec{x})$ – some measure density, common for entire EF distributions, $\vec{\Theta}=\{\Theta_1,...,\Theta_k\}$ – set of parameters, $\vec{u}(\vec{x})=\{u_1(\vec{x}),...,u_k(\vec{x})\}$ – set of EF sufficient statistics, $Z(\vec{\Theta})$ – normalizer.

The structure of neuron distribution density (3) as exponential mixture was chosen for simplicity of parameters $\vec{\Theta}_j$ estimation on the base of j-th cluster data observed set $\{\vec{x}_{j,1},...,\vec{x}_{j,n_j}\}$. In fact, due to certain feature of parametric exponential families [6], maximum likelihood parameter estimations $\hat{\vec{\Theta}}_j$ looks as:

$$\hat{\vec{\Theta}}_j=\vec{\nabla}G(\hat{\vec{\eta}}_j),\quad \hat{\vec{\eta}}_j=\frac{\sum_{i=1}^{n_j}\vec{u}(\vec{x}_j)}{n_j}, \qquad (5)$$

Where $\vec{\eta}$ are the moment parameters, since for EF $\vec{\eta}=<\vec{u}(\vec{x})>$ – sufficient statistics moments (4), and $G(\vec{\eta})=\max_{\vec{\Theta}}(\sum_i\Theta_i\eta_i(\vec{x})-\ln Z(\vec{\Theta}))$ is Legendre transform of normalizer logarithm.

Iterative EM-algorithm, adapted to poorly intersecting clusters, was used to work out parametric density distribution as (3), where, along with evaluation of parameters $\vec{\Theta}_j$ (5), there is need to evaluate n_j as well (i.e. clustering coordinates of full neuron array within the slice). Detailed approach of point cloud clusterization by means of 2D-gaussian mixture model is stated in article [4].

This model of distributed surface data representation as parametric distribution (3) avails to optimize 3D-reconstruction process by interpolation of parameters $\{n_j\}$ and $\{\vec{\Theta}_j\}$ to intersection space, and then recover density distribution using these parameters. Article [8] presents detailed accuracy reference in resolving neuron 3D-distribution according to set of layer wise 2D-distributions. Approximation accuracy criterion is worked out, and, using it, one can easily estimate correctness of approximation in any location between sections. Presented method of approximation is suitable for areas, where this criterion does not exceed certain level, otherwise one should use other methods, for instance B-spline lofting.

Conclusion

Computer simulation of spatial distribution of point cloud becomes more effective using proposed concept of distribution parameters interpolation. Advantage of proposed algorithms is the significant reduction of required computational resources due to turn multitudinous interpolative operations in spatial dimension of flat sections distribution to the task of interpolation of much less parameters modeling distributions.

Aforesaid method applied to reconstruction 3D-distribution of neurons from black substance digital images [9] approved proposed algorithms advantages – simplicity, reliability and weak resource consumption. Looking ahead, methods seem to be expanding and more effective in future research.

References

1. B.Geige. Three-dimensional modeling of human organs and its application to diagnosis and surgical planning. // Institut National de Recherche en Informatique et Automatique, Report 2105. -1993.
2. Allen Human Brain Atlas in 3-D. // Allen Institute for Brain Science. http://human.brain-map.org/static/brainexplorer
3. I.B. Gurevich, E.A. Kozina, A.A. Myagkov, M.V. Ugryumov, V.V. Yashina. Automating Extraction and Analysis of Dopaminergic Axon Terminals in Images of Frontal Slices of the Striatum. // Pattern Recognition and Image Analysis.- 2010. Vol. 20, P. 349-359.
4. Y.V. Obukhov, V.E. Antsiperov, O.V. Evseev. Neuronal 3D distribution reconstruction system based on a priori morphological model and series of real 2D-images of brain sections // Journal of Radio Electronics.- 2012. No.7, http://jre.cplire.ru/jre/jul12/5/text.pdf.
5. L. Piegl, W. Tiller. The NURBS book. // Springer, Heidelberg, New York.-1997.
6. F. Nielsen. K-MLE: A fast algorithm for learning statistical mixture models. // IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).- 2012. P.), 869-872.
7. R.M. Neal, G.E. Hinton. A view of the EM algorithm that justifies incremental, sparse, and other variants. // in “Learning in graphical models”, MIT Press Cambridge, USA.- 1999. P. 355 – 368.
8. V.E. Antsiperov, O.V. Evseev, Y.V. Obukhov. Neuronal 3D distribution reconstruction system based on a priori morphological model and series of real 2D-images of brain sections // Journal of Radio Electronics.- 2012. No.8, http://jre.cplire.ru/jre/aug12/8/text.pdf
9. V.E. Antsiperov, Yu.V. Obukhov, O.V. Evseev. Computer neurons 3D-distribution reconstruction on the base of neurological Parkinson’s disease model and 2D-microscopic section image series. // International conference "Intelligent Information Processing" IIP-9, Montenegro, Budva.- 2012. P. 540-543.

DEVELOPMENT OF THE SYNTHESIS ALGORITHM OF THREE-DIMENSIONAL OBJECTS FROM VIDEO IMAGES[1]

A.E. Barinov[2,3], A.A. Zakharov[2,4]

[2]Murom Institute of the Vladimir State University, Murom, The Russian Federation
[3]barinov-a-e@ya.ru, [4]aa-zaharov@ya.ru

The article describes the development of automatic algorithm for reconstruction of three-dimensional scenes from the video sequence. The algorithm is based on finding the point correspondences between frames. Presents the experimental research of the algorithm.

Introduction

One of the main tasks of computer vision is the task of reshaping the images. Distinguish active and passive optical methods of three-dimensional reconstruction. Active reconstruction methods based on the use of range finders, which have high costs. Passive optical three-dimensional reconstruction techniques include the manipulation of two-dimensional images. The passive methods include [1-4]:

- shape from stereo;
- shape from shading;
- shape from texture;
- shape from focus;
- shape from contour;
- shape from motion.

Recently, the actively developing shape from motion. In this case the reshaping of the scene from the sequence of images from different positions taken at certain times. Most of the camera is also considered to be unknown and restored in the course of solving the problem. To create a three-dimensional reconstruction of the system is suggested to use a method based on the movement of the camera in space. This method is called by the video sequence reconstruction (shape from video). Restoration of the geometry objects in the scene in such systems is difficult and requires significant user interaction. It is therefore necessary to develop new methods and software for the automatic reconstruction of the geometry of three-dimensional scenes of varying complexity. The proposed reconstruction algorithm for video sequence is as follows. Of the total number of images selected by the two adjacent frames. The aim is to find correspondences with the adjacent frame sequence. Personnel are calculated characteristic reference points three-dimensional scene. Other images are processed relative to the first frame. At each stage of the reconstruction of the original refined and expanded.

Synthesis algorithm of three-dimensional objects from video images

The algorithm is based on the implementation of the following steps in constructing a three-dimensional model from a video sequence.

Step 1. Receiving a set of raster images using a digital video camera. Shooting a subject comes from different points of view. The camera moves in a straight line. The direction of the optical axis of the camera exactly perpendicular to the direction of motion (Fig.1).

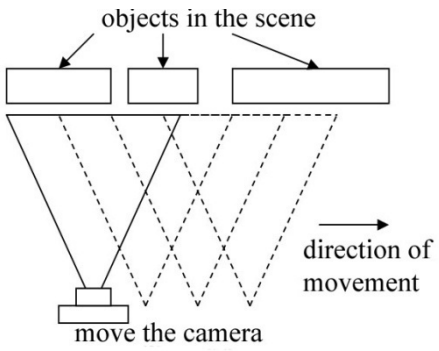


Fig. 1. The scheme for obtaining a video sequence

The error in determining the coordinates of the camera will not seriously affect the accuracy

[1] Support by the grant RFBR № 13-07-97523

of the reconstruction, as the movement going in a straight line.

Step 2. Finding point correspondences in the images. At this stage it is necessary to establish a correspondence between the point features of different images of the same scene. To find the point correspondences are used to handle them. A handle - an identifier of the key points that distinguishes it from a group of points. Coincidentally descriptors are allocated corresponding to each other key points. Descriptors should ensure the invariance of finding the correspondence between the singular points of the transformations of images. The paper was used descriptor SIFT (Scale Invariant Feature Transform).

Step 3. The calculation of the coordinates of control points. Reference points will be called a point whose projections were identified in the image detector. Two pictures with images of the same site, is derived from two points in space, called a stereoscopic pair of images (stereo pair). The calculations use the concept of epipolar configuration. In the simplest case of two identical parallel to the image plane parallel to the base of the stereo system (Fig.2).

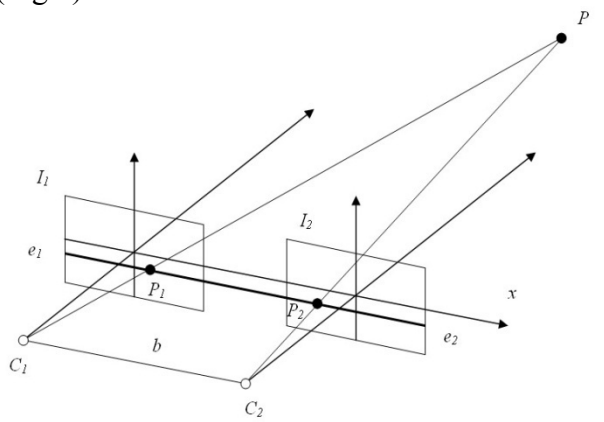


Fig.2. Epipolar configuration of a normal pair of images

Epipolar plane is the plane which the three-dimensional point P, the optical centers of two cameras C_1 and C_2, two projection P_1 and P_2 of point P on the image plane.

Epipolar lines are called two straight e_1 and e_2, which are direct intersection of epipolar plane with two planes I_1 and I_2 images.

The distance between the optical center of the camera is called the base b. Knowing epipolar configuration parameters, one can calculate the spatial coordinates of the point P.

The camera is calibrated in advance. Calibration task is to determine how the pixels located on the image plane of the camera relative to the three-dimensional space of points, which are required to obtain images from the camera. Allocate internal and external parameters of the camera. The internal parameters are:

- point of intersection of the optical axis and the image plane;
- scale factors for the horizontal and vertical pixel dimensions;
- focal length f : the distance from the optical center to the image plane;
- lens distortion factor: the scaling factor used to model the radial distortion of the lens.

External parameters describe the location and orientation of the camera coordinate system in three-dimensional space. These include the transfer parameters $t = \begin{bmatrix} t_x & t_y & t_z \end{bmatrix}^T$ and rotation settings $R = \begin{bmatrix} r_{11} & r_{12} & r_{13} & 0 \\ r_{21} & r_{22} & r_{23} & 0 \\ r_{31} & r_{32} & r_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

The algorithm is designed for the special case: there are internal and external parameters of the camera. Since the optical axis of the camera is always directed perpendicular to the motion, the rotation options are ignored.

In our system, the coordinates of the projections of a point P on the image plane can be calculated as follows

$$x_P = b\frac{(x_{P_1} + x_{P_2})}{2(x_{P_1} - x_{P_2})},$$
$$y_P = b\frac{y_{P_1} + y_{P_2}}{2(x_{P1} - x_{P_2})}, \qquad (1)$$
$$z_P = fb/(x_{P_1} - x_{P_2}).$$

Step 4. Building a three-dimensional model from the computed coordinates. According to the calculated coordinates is the construction of three-dimensional primitives. The geometric model is constructed using the polygonal geometry. This stage is one of the easiest to implement, as there are efficient algorithms for constructing the surface of the known coordinates of the vertices.

Conclusion

At realization of algorithm, you need to find a compromise between the size of the field of view, accuracy and computation of three-dimensional coordinates of the found point correspondences. At small base small initial errors in definition of coordinates can lead to considerable errors of calculation of depth. If adjacent positions of the camera considerably are remote from each other it can lead to difficulties in comparison of dot features because of mutual blanking of objects on maps and changes of the map of the scene.

Algorithm experimental researches on the object having the form of a parallelepiped have been spent. Maps of angular markers have been put on the parallelepiped sides. Shooting of the scene from several positions has been made. By the received results the schedule of dependence of quantity of the found dotted pairs from a corner of offset concerning a camera home position (fig. 3) has been constructed.

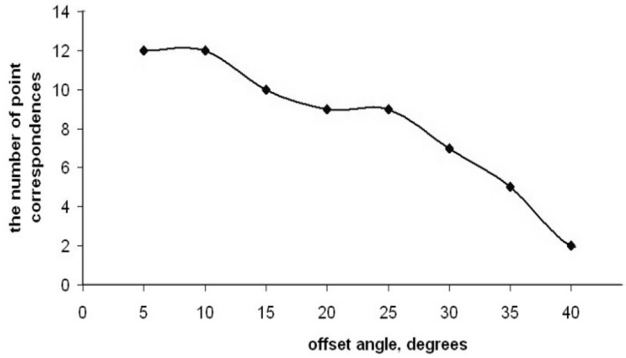


Fig. 3. Graph of the the number of point pairs found on the angle of move the camera

Analysis of the graph shows that with increasing displacement angle is significantly reduced number of point correspondences found automatically. Thus, the system parameters have been selected, at which the displacement angle of the current position is equal to 25 degrees. There have been studies of the developed algorithm on real objects (Fig. 4).

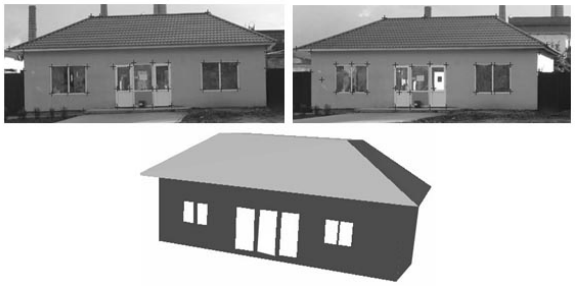


Fig. 4. Building three-dimensional model from the image sequence

It should be noted that the quality of the synthesized scenes depends on the image complexity. Algorithm does not always work correctly with images, which show objects with complex geometric structure, or if there are overlapping.

To increase the accuracy of reconstruction in the future we plan to use a pre-pre-defined information about the elements of the scene: the perpendicularity, parallelism, verticality, horizontality. In this case, the three-dimensional reconstruction is based on probabilistic methods of recognition.

The algorithm is implemented in software. Reconstruction of the scene is carried out automatically.

The proposed approach for restoration of three-dimensional data can be used alone or in conjunction with other methods of reconstruction to refine results.

References

1. El-Hakim S.F. Semi-automatic 3d reconstruction of occluded and unmarked surfaces from widely separated views/ S.F. El-Hakim// Proc. ISPSRS Symp. – Corfu. – 2002. – P. 143–148.
2. Hartley R. Multiple View Geometry in Computer Vision/ R. Hartley, A. Zisserman. – Cambridge University Press, 2000. – 655 p.
3. Pollefeys M. Detailed realtime urban 3D reconstruction from video/ M. Pollefeys, D. Nister, J.-M. Frahm// International Journal of Computer Vision. – 2008. – V.78 (2-3). – P. 143–167.
4. Szeliski R. Computer Vision: Algorithms and Applications/ R.Szeliski. – Springer-Verlag New York Inc, 2010. – 979 p.

OPTIMIZING THE COMPUTATIONAL COMPLEXITY OF THE ALGORITHM FOR ADAPTIVE COMPRESSION OF BINARY RASTER IMAGES [1, 2, 3]

A.V. Borusyak[4,5], Yu.G.Vasin[4]

[4] Research Institute of Applied Mathematics and Cybernetics
Lobachevsky Nizhni Novgorod State University
National Research University
603005, Russia, Nizhni Novgorod, Ulyanov St.10, UNN RIAMC.
Tel. (831)4362361, [5]sw-bor@yandex.ru

In this paper, we consider the problem of optimizing the computational complexity of the algorithm for adaptive compression of binary raster images (BRI) based on statistical coding using context modeling.

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. To date, many different universal compression algorithms have been proposed. However, at the same time, in many areas the need to develop specialized algorithms focused on a specific data type remains relevant. 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 monochrome black-and-white binary images (MBI). Among the most effective methods of data compression with a definite structure and characteristics are contextual data compression methods, where PPM is of primary importance.

Statement of the problem

A compression algorithm has been developed and implemented for MBI based on the PPM method of context modeling [12]. This implementation of the algorithm has demonstrated high efficiency in compression ratio, but the time consumption was too high. Besides, in larger images with low redundancy, memory overflow may occur due to the growth of the tree of context models. This paper discusses some ways to optimize the algorithm with respect to runtime and RAM consumption. We also tried to solve the task of expanding support for various image formats and of making the application suitable for use on various platforms.

Methods of solution

Initially, the order of pixels in the context was organized in such a way that one could get the context of any lower order with one operation of bit shift. However, the operation of calculating a new context of the maximum order is performed much more frequently than the operation of calculating the context of a lower order. Due to this, it was decided to optimize the function of calculating the new maximum order context. To solve this problem, the sequence of context pixels was changed from the user-specified fixed one to a serial line by line order, from the leftmost context pixel located on the same line with the pixel being coded to the rightmost one, located as far as possible vertically from the pixel being coded. This structure of context formation allows the calculation of a new context in two steps, with the use of information from the previous context. In the

[1]The reported study was partially supported by RFBR, research project № 13-07-00521_a
[2]The reported study was partially supported by RFBR, research project № 12-07-33107
[3]The reported study was partially supported by RFBR, research project № 13-07-12211

first step, the basis for the current context is calculated from the previous context using a single bit shift operation that includes the values of all the pixels belonging to the previous and current contexts. At the second step, the rightmost pixels of each line of the current context are added to the variable of the current context. Thus, the number of operations of addressing the image was reduced from the whole context value to the value of the context height. When moving to a new line of the image, it is sufficient to take only such pixels from the image that are at the same level or to the right of the pixel being coded, the remaining pixels will be outside the image boundaries and can be considered zero pixels. To calculate the context when moving to a new line, a fixed list of serial numbers is used of those pixels that do not fall outside the screen and are either to the right or at the same level as the pixel being processed. Calculation of lower-order contexts is made from the highest order context using a predetermined mask. The mask is a predetermined fixed list of serial numbers of maximum-order context pixels.

In order to limit the consumption of RAM, an upper limit was imposed on the number of context models stored in the tree of context models. When the specified threshold was exceeded, the AVL tree of maximum-level context models was deleted from the memory and replaced with an empty AVL tree. Due to its simplicity, this approach has minimized the time required for this operation. At the same time, due to the fact that the lower-level context models remained untouched after the operation, and due to the use of the method of information inheritance, the compression ratio has decreased only slightly.

Since many black-and-white documents have rectangular white margins around the main image, the detection of such margins was implemented, and their subsequent compression was performed separately from the content of the image. By means of serial search of all the pixels and finding the minimum and maximum values of the coordinates of black pixels in the image that are of the black colour, the boundaries of a white rectangular area framing the image are found. These boundaries, in terms of the coordinates of the upper left point and the lower right point of the meaningful part of the image, are recorded in the header of the output file. Next, the meaningful part of the image is separated and the resulting image is compressed as normal. This approach allowed us to significantly increase the speed of compression of images with white margins, and to increase the compression ratio of images.

To each context model, a reference to the ancestor context model was added, in order to accelerate the operation of finding the ancestor context model, which is used when inheriting information and moving to lower-order contexts. Context model trees are bounded from the top with respect to the number of context models contained. When a preset threshold is reached, the tree is completely cleared. This approach allows one to limit the amount of memory consumed when encoding noisy images with minimal loss in the coding rate and insignificant loss in compression efficiency.

To solve the problems of extending support to different formats and applicability for different platforms, the project was transferred from the Borland C++ Builder 6 programming environment to the Qt 4.8.2 programming environment. This transition has made it possible to improve performance, to make the program structure more flexible, to add support for such image formats as png and tiff, and also to provide the opportunity for compiling applications to run on different operating systems.

To improve the performance, the updating of the encoding (decoding) process percentage indicator is performed not every time when the transition occurs to a new line of the image, but only once in h/100 of such transitions, where h is the height of the image being processed.

Results

To test the effectiveness of the algorithm, experiments were performed on a test set of images and compared with an earlier version of the algorithm.

Table 1. Performance comparison of the program before and after the above-mentioned improvements

No.	Time (s)	
	Before improvements	After improvements
1	2,156	0,984
2	7,859	5,468
3	10,25	4,297
4	932,32	712,634
5	433,996	180,706

The results of experimental testing are given in the table.

Conclusion

The above approaches have reduced the coding time on average by the factor of 1.9. Memory consumption was also reduced, thus avoiding memory overflow when encoding images with low redundancy. Support for the most common image formats (png, tiff) has been added, along with the possibility to compile programs for different operating systems.

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. Salomon D. Compression of data, images and sound. - Moscow: Tekhnosfera, 2006.
3. Markov A.A. "Introduction to Coding Theory". M: Nauka. 1982.
4. Lidovsky V.V. "Information Theory": Manual. - M.: Sputnik Company, 2004. – 111 p. ISBN 5-93406-661-7.
5. Savitch W. Programming in C++. - St. Petersburg.: Piter, 2004.
6. Internet resource www.compression.ru
7. Specification for tiff http://cs.mipt.ru/docs/comp/eng/develop/formats/tiff/main.pdf
8. Specification for PNG http://www.w3.org/TR/PNG/
9. Specification for DJVU http://www.celartem.com/product/doc/specs/sdjvuspec.djvu
10. 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.
11. Zherzdev S.V. Image compression in Notices to Mariners // Microsoft Technologies in the theory and practice of programming. Proceedings. Nizhni Novgorod: University of Nizhni Novgorod Press, 2007. p.328.
12. Borusyak A.V., Vasin Yu.G., Zherzdev S.V. Compression of Binary Graphics Using Context Simulation // Pattern Recognition and Image Analysis, Vol. 23, No. 2, 2013 pp. 207-210

SEGMENTATION USING NON LOCAL ACTIVE CONTOURS AND BELIEF FUNCTIONS

Foued Derraz [1], Antonio Pinti [3,4] , Laurent Peyrodie [2] , Miloud Boussahla[5], Hechmi Toumi[4]

[1] Facult Libre de Medicine, 59046,France, foued.derraz@icl-lille.fr
[2] Hautes Etudes d'Ingénieur Lille, France, Laurent.peyrodie@hei.fr
[3] ENSIAME – UVHC, Université de Valenciennes, 59300 Valenciennes, France antonio.pinti@univvalenciennes.fr
[4] A5 EA 4708, I3MTO, CHRO 1, rue Porte Madeleine, 45032,Orléans, France, hechmi.toumi@univ-orleans.fr
[5] Telecommunication laboratory, Abou bekr Belkaid university, tlemcen, Algeria, m_boussahla@mail.univ-tlemcen.dz

In this paper, we propose a novel and rigorous framework for globally segmentation based on Non-Local Active contours that use the Dempster-Shafer framework. Instead of the traditional formulation of Non Local Active Contours, we investigated the evidential framework to find an alternative to Bayesian framework. The proposed method used all features issued from vector-valued image, including texture and color as descriptor to drive the Non Local Active Contours curve evolution. This formulation allows us to combine the foreground/background issued from the multiple channels in the same framework and to incorporate the heterogeneous knowledge in order to reduce the imprecision due to noise poor contrast, weak or missing boundaries of objects, inhomogeneities, etc. The statistical relation between the image channels is ensured by Dempster Shafer rule. We illustrate the performance of our segmentation algorithm using some challenging medical images.

Introduction

Image segmentation is one of the most important tasks in medical imaging applications, where segmentation has been becoming a powerful computer-aided tool for cancer or pathological detection, diagnosis and treatment as well as for surgical planning [1]. Segmentation can provide measurements for the location, area, volume of desired object to detect, and information allowing a dynamical analysis of anatomical structures. One of numerous contributions gained the attention of scientists in the world, which concerns the original work of Caselles *et. al.* [3], introducing firstly the geodesic Active Contours (AC) models.

Since then, AC models has been developed and also proved to be one of the most robust methods for segmentation of medical images [1,2,3,4]. The key idea of these models relies on the use of curve evolution to detect objects in a given image. More precisely, this consists in deforming an initial curve towards object boundaries, under some constraints issued from that image. There are two main approaches for AC models: the edge-based models [1,3] and the region-based models [1,2]. The edge-based models also referred as geometric AC, utilize the image gradient to guide the evolving curve toward object boundaries. These models take benefice of the advantages of the level set method and allows automatic change of topology[1,2,3].

The region-based models make use of statistical information of region instead of using image gradient, offering hence better performance in the case of noise and weak boundaries or discontinuous boundaries and handling homogeneous regions. However, by using global statistics, such models are not effective for segmenting objects with noise poor contrast, weak or missing boundaries between imaged objects, inhomogeneities, as in the case of magnetic resonance , positron emission tomography or computed tomography images. One way to overcome these difficulties is to exploit the high level knowledge about usual objects in an image. This will ease the interpretation of low-level cues extracted from images which may be highly beneficial for the segmentation methods based on AC models. Statistical knowledge [1,6] and

additional information such as texture and shape can improve the AC based segmentation methods [1,4]. To overcome these difficulties, the statistical knowledge are incorporated in the AC formulation using Dempster-shafer framework [5, 6,7].

The Dempster-Shafer (DS) framework [8] has been combined with either a simple thresholding [5], a clustering algorithm [9], a region merging algorithm [6] or with an AC algorithms [7].

In this paper we propose to use the evidential framework [8] to fuse several statistical information's in a new descriptor and incorporates it the AC models. The fusion of informations issued from different feature channels, e.g., color channels and texture offers an alternative to the Bayesian framework [7,8,9]. Instead of fusing separated probability densities, the evidential framework allows both inaccuracy and uncertainty. This concept is represented using Belief Functions (BFs) [7, 9, 10, 11] which are particularly well suited to represent information from partial and unreliable knowledges. The use of BFs as an alternative to the probability in segmentation process can be very helpful in reducing uncertainties and imprecisions using conjunctive combination of neighboring pixels. First, it allows us to reduce the noise and secondly, to highlight conflicting areas mainly present around weaks boundaries . In addition, BFs have the advantage to manipulate not only singletons but also disjunctions. This gives the ability to represent both uncertainties and imprecisions explicitly. The disjunctive combination allows the transferring of both uncertain and imprecise information on disjunctions [7, 11]. Finally, the conjunctive combination is applied to reduce uncertainties due to noise while maintaining representation of imprecise information at the boundaries between areas on disjunctions. In this paper, we highlight the advantage of evidential framework, to define a new descriptor based on the BFs to incorporate it in the formulation of the Non Local AC models [14, 15, 16]. In Section 2, we review the the segmentation based globally Non Local active contours model . In section3 we proposed a fast algorithm based split Bregman of our segmentation algorithm. In Section 4, we demonstrate the advantages of the proposed method by applying it to some challenging to some challenging images.

Globally Non Local Active Contours

Our Non-Local AC model is inspired from that in [14,16] and formualates the segmentation problem as a stationary point of the energy:

$$\min\left\{E_b(\chi)+\lambda E_{data}(\chi)\right\} \tag{1}$$

where λ is weighting parameter that should be adapted to the expected regularity of the boundary of the region. The characteristic function framework represent the segmented region Ω as:

$$\chi(\mathbf{x})=\begin{cases}1 & \mathbf{x}\, is\, inside\, \Omega\\ 0 & \mathbf{x}\, is\, outiside\, \Omega\end{cases} \tag{2}$$

Note that within the missing region only the term $E_b(\chi)$ regularizes the contours, which is thus expected to be a straight line segment in each connected component. Penalization using higher order curve derivatives could be used to allow curvilinear contours completion, such as those proposed in [6, 7].

We introduce an energy functional $E_{data}(\chi)$ enforcing the similarity of features located either inside or outside the contoured region Ω :

$$E_{data}(\chi)= \iint \varphi(\mathbf{x},\mathbf{y})K_\sigma(\mathbf{x}-\mathbf{y})P(\mathbf{x},\mathbf{y})T(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y} \tag{3}$$

The function φ restricts the comparison of pairs of patches that are in the same region Ω using Dempster-Shafer framwork. The function $T(\mathbf{x},\mathbf{y})$ excludes patches whose center is located in the damaged region from the comparison, this function can be formulated as:

$$T(\mathbf{x},\mathbf{y})=\begin{cases}0 & \mathbf{x}\, or\, \mathbf{y}\, are\, inside\, local\, region\\ 1 & otherwise\end{cases} \tag{4}$$

Note that the parameter σ is important since it controls the scale of the local homogeneity one requires for the object to be segmented.

The non-local interaction between two patches is measured using a weighted L2 distance:

$$P(\mathbf{x},\mathbf{y})=\int_t K_\sigma(z)\left\|p_{\mathbf{x}}^I(z)-p_{\mathbf{y}}^I(z)\right\|^2 dz \tag{5}$$

Where $K_\sigma(t)$ is the gaussien kernel and σ is weighting parameter used to give more influence to the central pixel. Note that missing pixels in patches are not accounted for the computation of the distance.

The patch around a pixel $\mathbf{x}$ is defined as:

$$p_{\mathbf{x}}^I(z)=I(\mathbf{x}+z) \tag{6}$$

For pixel neighbors $z\in[-\tau/2,\tau/2]$.

In order to produce the foreground regionand the background, the energy $E_{NL-AC}(\chi)$ is maximized, w.r.t the evolving domain, is done with the shape derivative tool [9]. Thus, the Eulerian derivative of E_{data} in the direction ξ is as follows:

$$\left\langle \frac{\partial E_{data}(\chi)}{\partial t}, \xi \right\rangle = \int_{\partial\Omega} V_{data} \left\langle \xi(s), N(s) \right\rangle ds \quad (7)$$

Where $\vec{N}$ is an exterior unit normal vector to the boundary $\partial\Omega$, $\left\langle \varepsilon, \vec{N} \right\rangle$ is the Euclidean scalar product and s is the arc length parametrization.

Dempster-shafer rules

The evidential framework [8, 9, 10, 11] is defined by the frame of discernment $\Omega = \{\Omega_1, \Omega_2, ..., \Omega_n\}$, composed of n single mutually exclusive subsets Ω_i. Each element of discernment can be associated with an elementary mass function noted $m(\Omega)$. The representation of imprecision and uncertainty is provided through the definition of the plausibility (Pl) and belief (Bel) function, which are both derived from a mass function (m). This function m is defined by: $m : 2^{\Omega} \to [0,1]$:

$$\begin{aligned} & m(\varnothing) = 0 \\ & \sum_{\Omega_i \subseteq \Omega} m(\Omega_i) = 1 \\ & Bel(\Omega) = \sum_{\Omega_i \subseteq \Omega} m(\Omega_i) = 1 \\ & Pl(\Omega) = \sum_{\Omega_i \cap \Omega \neq \varnothing} m(\Omega_i) \end{aligned} \quad (8)$$

When $m(\Omega) > 0$, Ω is a so called focal element [6,8,10,11].

In this framework, independent masses m_m are defined within the same frame of discernment as:

$$\begin{aligned} & m\left(\Omega_{i=\{1,...,n\}}\right) = m_1\left(\Omega_{i=\{1,...,n\}}\right) \otimes \\ & m_2\left(\Omega_{i=\{1,...,n\}}\right) ... \otimes m_m\left(\Omega_{i=\{1,...,n\}}\right) \end{aligned} \quad (9)$$

The total belief assigned to a focal element Ω_i is equal to the belief strictly placed on the region Ω_i. Then belief function can expressed as:

$$Bel(\Omega_i) = m(\Omega_i) \quad (10)$$

In this paper, we proposed to incorporate the comparison function as a belief function:

$$\varphi(\mathbf{x}, \mathbf{y}) = Bel(\Omega_{roi}) = m(\Omega_{roi}) \quad (11)$$

Fast Algorithm based on Split Bregman

A fast and accurate minimization algorithm for TV problem is introduced in [4,12]. We propose to solve our segmentation in this new framework and we formulate the variational problem as:

$$\min_{\chi,d}\left(E(\chi,d)\right) = \int_{\Omega} \left|d(\mathbf{x})\right| d\mathbf{x} + \lambda \int_{\Omega} V_{data}\, \chi(\mathbf{x})\, d\mathbf{x} \quad (12)$$

The vectorial function d enforces $d = \nabla\chi$ using the efficient Bregman iteration approach [7, 14] defined as:

$$\begin{cases} \left(\chi^{k+1}, d^{k+1}\right) = \operatorname{argmin} \left\{ \begin{array}{l} \lambda \int_{\Omega} V_{data}\, \chi \\ + \frac{\mu}{2} \int_{\Omega} \left| d - \nabla\chi - b^k \right|^2 \end{array} \right\} \\ b^{k+1} = b^k + \nabla\chi^k - d^{k+1} \end{cases} \quad (13)$$

The minimizing solution χ^{k+1} is characterized by the optimality condition:

$$\mu\Delta\chi = \lambda V_{data} + \mu div\left(b^k - d^k\right), \chi \in [0,1] \quad (14)$$

A fast approximated solution is provided by a Gauss-Seidel iterative scheme [7]. Finally, the minimizing solution is given by soft-thresholding:

$$d^{k+1} = shrink(\nabla\chi^{k+1} + b^k, \mu^{-1}) \quad (15)$$

Then, the final active contour is given by the boundary of the set $\left\{ \mathbf{x} \in \Omega \middle| \chi(\mathbf{x})^{final} \geq \frac{1}{2} \right\}$. The two iteration schemes are straightforward to implement. The two iteration schemes are straightforward to implement. Finally, V_{data} is updated at each iteration using the belief function given in (11) and (9).

Results

We introduced AC model that integrates belief function as statistical region knowledge. To illustrate and demonstrates the accuracy of our segmentation method, we present some results of our method and compare them to segmentation done by the traditional AC model based vector value image and the model proposed in [15]. The three methods are evaluated on 20 color images taken from the Berkeley segmentation dataset [15] using F-measure criterion. Traditional segmentation and method in [16,18] boundaries for the initialization around the object to be segmented, our method is free initialization and the segmentation done by the three method are presented in Figure.1. The proposed method gives the best

segmentation and the F-measure is better than the other methods (see Table.1)

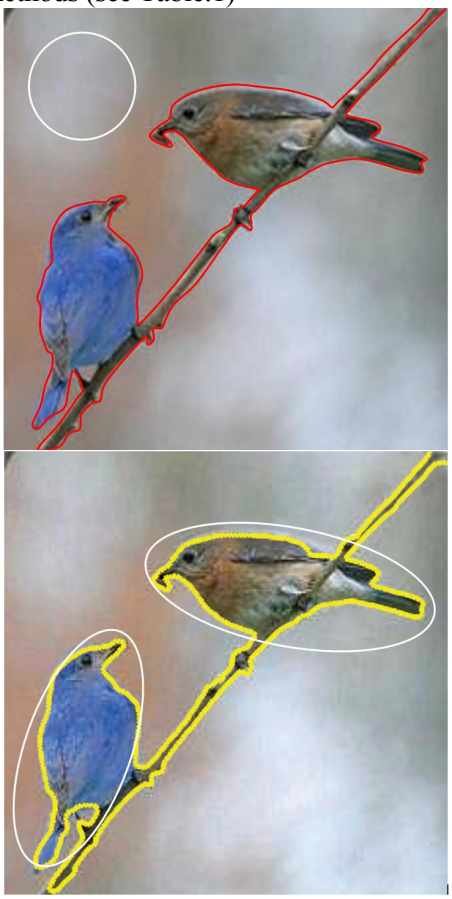


Fig. 1: Images taken from the Segbench Berkeley dataset [15]. In yellow color, the segmentation done by [16], in red color the segmentation done by our method. In the white color the initial curve of each model.

Table 1. Quantitative evaluation of the segmentation(F-measure)

Image	Our method	Method in [16]
Image a	0.94	0.91

Conclusion

In this paper, we have investigated the use of the evidential framework for AC model using DS theory. In this framework, we have investigated in particular the determining of mass function which represents a difficult task. The results have shown that proposed approach give the best segmentation for color and textured images. The experimental results show that the segmentation performance is improved by using the three information sources to represent the same image with respect to the use of on information. Indeed, there are some drawbacks of our proposed method. The proposed is very high time consuming for calculating the mass functions. Furthermore, the research of other optimal models to estimate the mass functions in the DS theory and the imprecision coming from different images channels are an important perspective issue of our work.

References

13. Daniel Cremers, Mikael Rousson, and Rachid Deriche, “A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape,” Int. J. Comput. Vision, vol. 72, no. 2, pp. 195–215, Apr. 2007.
14. Tony F. Chan, B. Yezrielev Sandberg, and Luminita A. Vese, “Active contours without edges for vector-valued images,” Journal of Vis. Communi. and Image Repres., vol. 11, pp. 130–141, 2000.
15. V. Caselles, R. Kimmel, G. Sapiro, "Geodesic Active Contours", Inter. Journal Computer Vision, vol. 22, no. 1, pp. 61-79, 1997.
16. Xavier Bresson, Selim Esedoglu, Pierre Vandergheynst, Jean-Philippe Thiran, and Stanley Osher, “Fast global minimization of the active contour/snake model,”
17. Michele Rombaut and Yue Min Zhu, “Study of dempster-shafer theory for image segmentation applications,” Image and Vision Computing, vol. 20, no. 1, pp. 15–23, 2002.
18. Benoït Lelandais, Isabelle Gardin, Laurent Mouchard, Pierre Vera, and Su Ruan, “Using belief function theory to deal with uncertainties and imprecisions in image processing,” in Belief Functions, 2012, pp. 197–204.
19. Björn Scheuermann and Bodo Rosenhahn, “Feature quarrels: the dempster-shafer evidence theory for image segmentation using a variational framework,” in Proceedings of the ACCV’10 - Volume Part II, Berlin, Heidelberg, 2011, pp. 426–439, Springer-Verlag.
20. Arthur P. Dempster and Wai Fung Chiu, “Dempster shafer models for object recognition and classification,” Int. J. Intell. Syst., vol. 21, no. 3, pp. 283–297, 2006.
21. Marie-Hélèene Masson and Thierry Denoeux, “Ecm: An evidential version of the fuzzy c,” Pattern Recognition, vol. 41, no. 4, pp. 1384–1397, 2008.
22. Fabio Cuzzolin, “A geometric approach to the theory of evidence,” IEEE Trans. on Syst., Man, and Cyber., Part C, vol. 38, no. 4, pp. 522–534, 2008.
23. Thierry Denoeux, “Maximum likelihood estimation from uncertain data in the belief function framework,” IEEE Trans. Knowl. Data Eng., vol. 25(1), pp. 119–130, 2013.
24. Tom Goldstein, Xavier Bresson, and Stanley Osher, “Geometric applications of the split bregman method: Segmentation and surface reconstruction,” J. Sci. Comput., vol. 45, no. 1-3, pp. 272–293, 2010.
25. David R. Martin, Charless C. Fowlkes, and Jitendra Malik, “Learning to detect natural image boundaries using local brightness, color, and texture cues,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 5, pp. 530–549.
26. X. Bresson and T. Chan, “Non-local unsupervised variational image segmentation models,” UCLA CAM Report 08-67, 2008.
27. G. Gilboa and S. Osher, “Nonlocal linear image regularization and supervised segmentation,” SIAM Mul. Model. and Simul., vol. 6, no. 2, pp. 595–630, 2007.
28. M. Jung, G. Peyré, L.D. Cohen: Non-local Active Contours. International Conference on Scale Space and Variational Methods in Computer Vision (SSVM), 255-266, Ein Gedi, Israel, 2011.

CROSS-VALIDATION OF PARAMETRIC ACYCLIC MODELS OF INTERRELATED OBJECTS[1]

S. Dvoenko[2,3], D. Sang[2,4]

[2] State University of Tula, 300600, Tula, Lenin Ave., 92, Russia
[3]dsd@tsu.tula.ru, [4]dvietsang@gmail.com

The tree-like model of a Markov random field is under investigation. The proposed approach is based on replacing an arbitrary graph of a MRF by a linear combination of acyclic parametric Markov models. The acyclic Markov models were previously suboptimal without tuning their parameters. Here we propose a simplified cross-validation procedure to evaluate the quality of a decision rule in the recognition problem and tune the Markov parameters as hyperparameters.

Introduction

It is often necessary to make coordinated decisions on objects collected in a single array of an interrelated data. Such interrelations between array elements are presented by adjacency graphs. The adjacency graph of the linearly ordered arrays is a chain.

Hidden Markov models appear to be very efficient for chain neighboring elements [1]. But for arbitrary adjacency graphs with cycles the problem of the Markov random field recognition becomes a *NP*-problem [2, 3]. Previously [4–6] the tree-like model of a Markov random field was proposed and the algorithm [6] for combining given tree-like Markov models was developed. Based on the Gauss-Seidel scheme, the procedure for estimating of optimal weights of graphs in linear combination was also proposed [6].

The transition matrix is a parameter of the one-sided Markov model. This matrix can be specified by only a unique diagonal element [4–6]. Methods of searching [7] for the optimal transition matrix and weights of acyclic adjacency graphs were proposed to tune the set of the acyclic Markov models.

On the other hand, Markov parameters can be considered as hyperparameters. But the tuning of hyperparameters based on the full cross-validation scheme requires an extremely high computational complexity.

In this work we propose a simplified cross-validation scheme to reduce the computational complexity and evaluate the Markov parameters as hyperparameters.

Acyclic Markov models in data array recognition problem

Let $t \in T$ be an element of a data array T. A symmetrical antireflective binary relation is determined for the set of array elements $t \in T$ by an undirected graph G without loops with edges $(s,t) \in G$ to connect a pair of adjacent elements $s \in T$ and $t \in T$.

The array T is presented as a two-component field (X, Y). The observed part $Y = (\mathbf{y}_t, t \in T)$ takes values from a certain set $\mathbf{y}_t \in \Theta$. The hidden part $X = (x_t, t \in T)$ needs to be restored and takes values from a finite set of class numbers $\Omega = \{1, 2, ..., m\}$.

Let observations $\mathbf{y}_t, t \in T$ be conditionally independent $\psi_t(\mathbf{y}_t \mid X) = \psi_t(\mathbf{y}_t \mid x_t)$ relative to the hidden field X.

Let the hidden field of classes X be restored by the Bayesian estimation

$$\hat{X}(Y) = (\hat{x}_t, t \in T), \hat{x}_t(Y) = \arg\max_{x_t \in \Omega} p_t(x_t \mid Y).$$

Let the hidden field X be a Markov field relative to the acyclic graph G. It was shown [4] a prior and a posteriori fields of hidden classes X are one-sided Markov ones relative to the same graph G.

[1] Supported by RFBR grants №13-07-00529, №11-07-00634

The supervised learning is the usual technique to evaluate a posteriori distributions $p_t(x_t \mid \mathbf{y}_t)$ in the pattern recognition problem. The hidden field X can be recognized by two passes along the graph G [5, 6]. That procedure was previously named as the basic algorithm.

In [6] the combining algorithm was proposed. In this algorithm the unknown a posteriori marginal probability distributions for hidden classes relative to the initial adjacency graph are replaced by a linear combination of distributions relative to a certain set of acyclic graphs $G_k, k = 1,...,K$ approximating the initial adjacency graph of the data array elements. Graph weights in combination are determined based on the Gauss-Seidel scheme [6].

In [4–6] the particular model of a Markov random field was proposed. Let one-sided MRF X be a homogeneous ergodic reversible finite Markov chain that has the stationary distribution p_i, $i \in \Omega$ in the root t^*.

The root t^* is naturally associated with the start of processing. Let us make the stationary distribution as the root one. Then the marginal a prior distributions in all elements are the same as the stationary distribution. Such a hidden field X is totally defined by a matrix $Q(m \times m) = (q_{ij}; i, j = 1,...,m)$, where q_{ij} is the transition probability from i-th to j-th states.

The stationary distribution $p_i, i \in \Omega$ of an ergodic reversible Markov chain satisfies to conditions $p_i q_{ij} = p_j q_{ji}; i, j \in \Omega$ relative to transition matrix Q. If the stationary distribution is uniform, then $q_{ij} = q_{ji}; i, j \in \Omega$, and Q becomes symmetric and twice stochastic.

It is proposed for the particular model that the transition matrix Q consists of the identical diagonal and the identical non-diagonal elements. That matrix Q can be specified by only single value of its diagonal element q as the unique Markov parameter to be tuned.

Algorithms for parametric tuning of an acyclic graph combination

Three algorithms [7] for simultaneously adjusting diagonal elements and weights of the linear combination of a given acyclic graph set were proposed. The general idea of them is to include the searching of the diagonal element in the process of graph weight adjusting based on the Gauss-Seidel scheme [6]. Variations of the diagonal element q in the range $1/m \le q < 1$ and the graph weights w in the range $0 \le w \le 1$ are considered like the coordinate-wise descent. In the algorithm A1 all Markov models for graphs G_k, $k = 1,...,K$ are defined by a single diagonal element q. In two other algorithms A2 and A3 each graph G_k, $k = 1,...,K$ corresponds to a separate acyclic Markov model with diagonal element q_k, $k = 1,...,K$. It is shown [7] these algorithms demonstrate a high level of recognition quality. It is compared, for instance, with the algorithm TRWS [8, 9], which is considered as the most efficient one nowadays.

Hyperparameter tuning problem

In general case, the model parameters can be divided into natural parameters and hyperparameters. Natural parameters reflect the essential properties of a model. But in many cases, it is required to determine some extra-parameters, for instance, to control the configuration of the parametric model or a prior distribution of natural parameters given. These extra parameters are called hyperparameters.

Sometimes the difference between natural and hyperparameters is quite unclear. The interpretation of parameters as natural or hyperparameters can be determined by investigator's point of view. In particular, such a situation arises in the problem of selecting parameters for the combination of acyclic adjacency graphs. Markov parameters and graph weights can be viewed as either natural or hyperparameters.

It is a complicated problem to determine the correct hyperparameters, because they have to restrict the permissible decision rules in order to avoid overfitting. Thus they can't be directly used in the error minimization process based on the training set to estimate the natural parameters. In model selection problem hyperparameters are usually tuned by the cross validation scheme. In practice the k-folds cross-validation is often used with $k = 5$ or $k = 10$. In this technique

the data is split into k roughly equal-sized parts.

Some overfitting for the validation data can occur in the cross-validation-based hyperparameter optimization. Therefore, it may be necessary to use another so-called test set that doesn't participate in the cross-validation process. The statistical efficiency of the tuned model is finally evaluated based on the test set [10]. The standard way of hyperparameter tuning is the grid search, which is simply an exhaustive searching through a specified set of hyperparameter values. That tuning scheme requires considerable time for multiple training runs that are applicable only to "fast" learning algorithms. There are no more than 2-3 hyperparameters can be tuned together by cross-validation.

In this paper, in order to avoid the excessive computational complexity we consider graph weights as natural parameters, and Markov parameters as hyperparameters.

In addition, we propose a "hybrid approach" consists in adjusting Markov parameters as natural in the same procedure with adjusting graph weights based on the Gauss-Seidel scheme [7]. But the quality of the decision rule is evaluated by a simplified cross-validation scheme that is similar to hyperparameter tuning scheme.

Simplified cross-validation scheme

In this paper the segmentation problem of raster texture images is investigated. Segmentation is performed on 100 simulated images (Fig. 1) with the size of 201×201. The textures of three classes are realizations of normally distributed variables with slightly different means in the feature space of RG-color components. The error rate of independent recognition for such images is not less than 30%.

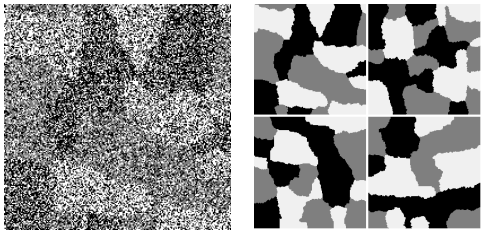


Fig. 1 – The first simulated image and the exact segmentations of first four simulated images

In Fig. 2 the acyclic adjacency graphs convenient to be used for texture images are shown.

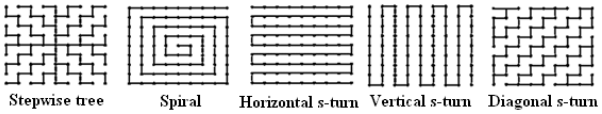


Fig. 2 – Acyclic adjacency graphs

Generally the initial set of images is divided into training and test sets, where the test set usually contains 25% of the total volume of the initial set. Such a partition of the initial set is often repeated 100 times. By repeating partitions some images may never participate in test, and some others may never participate in training.

In the simplified scheme of cross-validation every time we leave only one image as a test and the remaining images are used to perform cross-validation. It ensures that each image is used as a test only once, and all images participate in training.

In order to reduce the computational complexity we apply the non-classical cross-validation scheme using only one image for training.

It is known that a relatively small amount of training set leads to overfitting. Then, using only one image for training may not be sufficient. On the other hand, an image with the size of 201×201 consists of about 40 thousands of pixels. If each pixel is treated as an object, then using one image will mean training on about 40 thousands of objects. This number of training objects is large enough.

Another reason allowing us to use the non-classical cross-validation scheme instead of the classical one is that the results of Markov parameter selection on one simulated image and on most of them are not very different from each other. It is confirmed as follows. For each image we build the error line gained by combining acyclic adjacency graphs with equal weights for changing value of diagonal element. An analogous error line is built on the remaining images without itself. Experiments show that the values of diagonal element, giving the minimums of errors in both cases, are almost identical (Fig. 3).

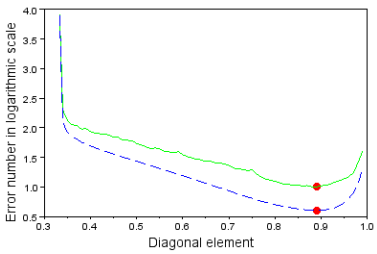


Fig. 3 – Lines of recognition errors

Furthermore, performing validation on most of images allows us to judge about the

generalization ability of the gained decision rule just right on the validation set. Therefore, among all decision rules we can choose such that potentially has the best generalization ability by defining it as the one that gives the lowest validation error. Validation error is the average number of errors on the validation set.

That foundation allows us to create the simplified scheme of cross-validation for estimating the general error. The procedure is as follows.

1. Exclude sequential image from a given image set and treat it as a test.
2. The other images form the training set on which the non-classical scheme of cross-validation is performed. Another image is selected for training. The remaining images are used for validation. Define the set of parameters that gives the lowest validation error.
3. Perform recognition on the test image based on the gained decision rule.
4. Average test error. This error is considered as the recognition error on the general set.

It should be emphasized that in such a scheme the cross-validation error is defined by averaging all validation errors gained by different sets of parameters. However, the test error is evaluated by the best set of parameters among them. Therefore, the general error should be less than the cross-validation error. On the contrary, in the standard schemes the cross-validation error often gets lower.

In practice, if it is required only to give the optimal tuning of parameters, we can build a simpler procedure without test images. The non-classical scheme of cross-validation is performed throughout a given image set and it is necessary just to select the parameter tuning, providing the least validation error.

Experimental results

Average cross-validation errors and general errors are shown in Fig. 4. The algorithm for weight fitting by Gauss-Seidel method is the worst. The quality of the algorithm for selecting unique diagonal element and graph weights is worse than two schemes with multiple diagonal elements.

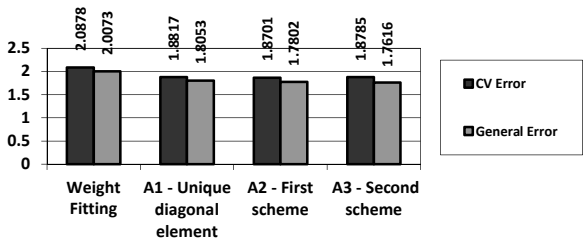


Fig. 4 – General and average cross-validation errors

If the non-classical scheme of cross-validation is performed on all 100 images without test ones, then the set of parameters which gives the least validation error is formed by the following values: graph weights (0,465; 0,13; 0,141; 0,141; 0.123) and diagonal elements (0,37; 0,61; 0,9; 0,92; 0,82), corresponding to different acyclic graphs (Fig. 2).

Conclusion

In this paper we propose a procedure of cross-validation, which allows reducing the computational complexity of hyperparameter tuning. This cross-validation procedure permits to evaluate the statistical properties of the decision rule, and determine the optimal combination of acyclic adjacency graphs with Markov parameters considered as hyperparameters.

References

1. L.R. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition // Proc. IEEE, 77. 1977. V. 2. P. 257–286.
2. S.Z. Li. Markov Random Field Modeling in Image Analysis. L: Springer–Verlag, 2009. 371 p.
3. M.J. Wainwright, M.I. Jordan. Graphical Models, Exponential Families, and Variational Inference // Foundations and Trends® in Machine Learning. 2008. V. 1. P. 1–305.
4. S.D. Dvoenko, A.V. Kopylov, V.V. Mottl. Pattern Recognition in Interrelated Data: The Problem, Fundamental Assumptions, Recognition Algorithms // Automat. Remote Control 65 (1), 127-141 (2004).
5. S.D. Dvoenko, A.V. Kopylov, V.V. Mottl. Pattern Recognition in Interrelated Data: The Problem, Fundamental Assumptions, Recognition Algorithms // Automat. Remote Control 66 (12), 2019-2032 (2005).
6. S.D. Dvoenko. Recognition of dependent objects based on acyclic Markov models // Pattern Recognition and Image Analysis. – 2012. – Vol. 22, No.1. – P. 28-38.
7. S.D. Dvoenko, D.V. Sang. Recognition of raster textured images based on parametric acyclic Markov

models // Proc. 22-th GraphiCon (Moscow, 2012). – M.: Maks Press, 2012. – P. 139-143.
8. V. Kolmogorov. Convergent Tree-Reweighted Message Passing for Energy Minimization // IEEE Trans. PAMI – 2006. – Vol.28, No.10. – P.1568-1583.
9. R. Szeliski, R. Zabih, D. Scharstein and others. Comparative Study of Energy Minimization Methods for Markov Random Fileds with Smoothness-Based Priors // IEEE Trans. PAMI – 2007. – Vol.6, No.6. – P.1068-1080.
10. M. Bishop. Pattern Recognition and Machine Learning. New York: Springer, 2006, 738 p.

ALGORITHM FOR ESTIMATING THE SPECTRAL CHARACTERISTICS OF REGULAR TEXTURES[1]

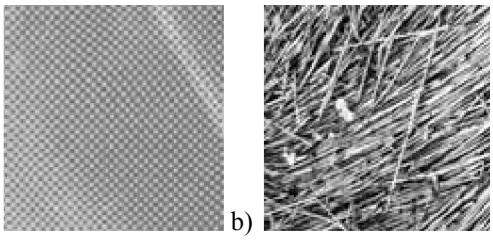
V.A. Fedoseev[2]

[2]Image Processing Systems Institute of the RAS
443001, Molodogvardeiskaya st., 151, Samara, Russia; +7(846)3378084, vicanfed@gmail.com

In this paper we propose an algorithm for estimating the spectral characteristics of the regular textures on digital images. Such characteristics represent grid structure of local maxima on image power spectrum. Search and selection of the local maxima are implemented using the procedure of non-linear filtering of the image spectrum in polar coordinates. It is shown that the results of the proposed algorithm can be helpful in solving of different texture analysis problems, such as texture segmentation, defect detection and periodic noise filtering.

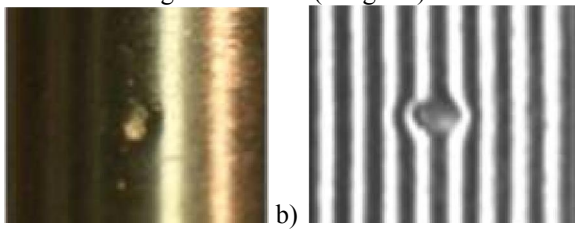
Introduction

Term "regular textures" is usually used to determine textures which are successfully described by structured approach [1], according to which the texture is described by a set of primitive *texels* arranged in a kind of regular or near-regular order [2]. Examples of regular textures can be found on large-scale images of printed documents (see Fig. 1a) or textile products, periodically noised images, striped snapshots of industrial items (see Fig. 2), etc.



a) b)

Fig. 1. Examples of regular (a – halftoned image) and irregular textures (b – grass)



a) b)

Fig. 2. A fragment of a cylindrical item (a) and its snapshot made by a periodical light pattern (b) [4]

Let us select the basic problems arising in analysis of such images.

1. *Texture image segmentation.* This problem occurs, for example, when we analyze printing documents to extract *texture watermarks* – a kind of protection information embedded into image to be printed by using different halftoning parameters in different image areas [3].
2. *Texture defect detection.* This problem arises in the quality analysis of textile [4, 5], industrial products [7], as well as printing products (so-called Heliotest inspection [6]).
3. *Periodic noise filtering.* This problem arises, for example, in the processing of remote sensing data [8] or in inverse halftoning (recovery of halftone image from binary, see Fig. 5) [9].

To solve all these problems it is possible to use methods which use as input data some measured spectral characteristics of the analyzed images. In this paper we propose a set of such characteristics of regular textures, estimated from image power spectrum, which then can be used in solving the above-mentioned problems, and an algorithm for finding these characteristics. Also in the paper some examples of the algorithm results, as well as examples of solved texture analysis problems using found spectral characteristics are provided.

A set of spectral characteristics of a regular texture

Power spectrum of non-sinusoidal regular texture besides the main harmonic contains multiple harmonics, which are repeated with step equal to the texture frequency (in the case of asymmetric texture), or the doubled texture frequency (in symmetric case). Figure 3 shows

[1] This work was financially supported by the Russian Foundation for Basic Research (Projects 12-01-00822, 12-07-31056, 13-01-97007, 13-01-12080) by Russian President Grant MK-3863.2013.9

examples of asymmetric regular textures and corresponding power spectra containing peaks at the multiple harmonics.

We propose to use main frequency and step of multiple harmonics as distinguishing information about regular texture. Thus, for a one-dimensional texture sought-for characteristics are main frequency vector $\mathbf{M}_1 = (\mu_{11}, \mu_{12})$ in 2D frequency space and grid step vector $\mathbf{\Delta}_1 = (\delta_{11}, \delta_{12})$ in the same space. Power spectrum of such a texture can have local maxima in coordinates $\mathbf{M}_1 + k\mathbf{\Delta}_1$, $k \in \mathbb{Z}$. Two-dimensional periodic texture in addition characterized by vectors $\mathbf{M}_2 = (\mu_{21}, \mu_{22})$ and $\mathbf{\Delta}_2 = (\delta_{21}, \delta_{22})$, and its power spectrum can contain peaks in the parallelogram grid nodes with coordinates $\mathbf{M}_1 + k\mathbf{\Delta}_1 + \mathbf{M}_2 + \kappa\mathbf{\Delta}_2$, $k, \kappa \in \mathbb{Z}$.

Some regular textures can have more simplified model. Thus, images obtained by a periodical light pattern (see Fig. 2) have a periodicity in only one direction, consequently, for these textures $\mathbf{M}_2$ and $\mathbf{\Delta}_2$ are zero-vectors. Printed halftoned images are characterized by a square grid cell, therefore for them

$$\mu_{21} = -\mu_{11}, \mu_{22} = \mu_{12}, \delta_{21} = -\delta_{11}, \delta_{22} = \delta_{12}.$$

Description of the algorithm

The proposed algorithm for finding the set of spectral characteristics of the regular texture consists of two stages:

1) construction of a primary (excessive) list of power spectrum local maxima;
2) determination of grid nodes from this list corresponding to regular textures.

As follows from introduction, the range of images considered in this paper is quite broad, and the origin of some of them suggests the possibility of having multiple textures in the image. The projected algorithm should define the characteristics set for all textures presented in the image.

Local maxima search is performed on the power spectrum image, represented in polar coordinates, using well-known peak filter [10]. In polar coordinates, the use of peak filter with fixed parameters at different distances from origin is more correct than in Cartesian coordinates, because in such case significant peaks at low frequencies can be omitted, and false detection of minor peaks at high frequencies can happen. After peak filtering further selection of only those peaks whose value T times greater than the average of their local neighborhood can be carried out. Then the coordinates of the selected local maxima are translated back into Cartesian and fill a list P in descending order of their corresponding power spectrum values. The list item $P(i)$ will be further denoted as (m_1^i, m_2^i).

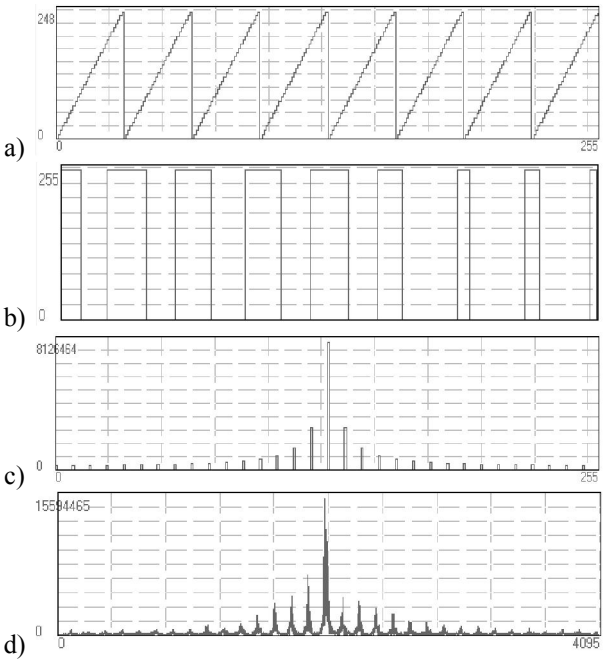


Fig. 3. A periodical signal (a), a halftoned image line profile (b) and amplitude spectrums of them (c-d)

A method of finding grid nodes from the list of found peaks used in the second stage, varies depending on a priori information about the shape of a regular texture grid. In total there are 5 variants: one-dimensional texture, two-dimensional texture with square, diamond, rectangular or parallelogram grid.

For simplification of the presentment and for notation cumbersome reduction we consider in details a method of finding nodes for square grid, typical for halftoned images, and then briefly describe modifications for other cases.

While the list P is not empty, repeat the following steps:

Step 1. $\mathbf{M}_1 = (\mu_{11}, \mu_{12}) := P(0)$ is the first item in P. It is extracted from P and added to the resulting list P_L which finally will contain all detected nodes of the moving lattice. The list P_L is previously cleaned.

Step 2. Calculation of the lattice step. If $\exists i : P(\mathrm{i}) = \left(m_1^i, m_2^i\right) \in u_{2\cdot\varepsilon}\left(2\mu_1, 2\mu_2\right)$, where $u_{2\cdot\varepsilon}\left(2\mu_1, 2\mu_2\right)$ is circular neighborhood with radius 2ε of the point $\left(2\mu_1, 2\mu_2\right)$, then $\boldsymbol{\Delta}_1 := \mathbf{M}_1$, move to step 3.

Otherwise, similarly check for peak in $u_{3\cdot\varepsilon}\left(3\mu_1, 3\mu_2\right)$. If it is found, then $\boldsymbol{\Delta}_1 := 2\mathbf{M}_1 = \left(2\mu_{11}, 2\mu_{12}\right)$; move to step 3.

Otherwise, we decide that the peak $\left(\mu_{11}, \mu_{12}\right)$ does not form a grid, which means that it does not indicate a regular texture. Move to step 1.

Step 3. Checking for existence of an orthogonal texture.

If $\exists i : P(\mathrm{i}) = \left(m_1^i, m_2^i\right) \in u_{\varepsilon}\left(-\mu_{11}, \mu_{12}\right)$, then $\mathbf{M}_2 = \left(m_1^i, m_2^i\right)$, move to step 4; else $\mathbf{M}_2 = \mathbf{0}$, $\boldsymbol{\Delta}_2 := \mathbf{0}$, move to step 6.

Step 4. Calculation of the lattice step for the orthogonal texture. If $\boldsymbol{\Delta}_1 = \mathbf{M}_1$, then $\boldsymbol{\Delta}_2 := \mathbf{M}_2$, else $\boldsymbol{\Delta}_2 := 2\mathbf{M}_2$.

Step 5. Search for all the peaks corresponding to grid nodes formed by the vectors $\mathbf{M}_1$, $\mathbf{M}_2$, $\boldsymbol{\Delta}_1$, $\boldsymbol{\Delta}_2$. For each $k, \kappa \in \mathbb{Z}$; $b_1, b_2 \in [0,1] \cap \mathbb{Z}$ check existence in P of local maxima

$$\left(m_1^i, m_2^i\right) \in u_{\max(k,\kappa)\cdot\varepsilon}(b_1(\mathbf{M}_1 + k\boldsymbol{\Delta}_1) + b_2(\mathbf{M}_2 + \kappa\boldsymbol{\Delta}_2)). \quad (2)$$

Found peaks are extracted from P and added to P_L.

Step 6. We have found the regular texture with parameters $\mathbf{M}_1$, $\mathbf{M}_2$, $\boldsymbol{\Delta}_1$, $\boldsymbol{\Delta}_2$ and the list of detected nodes P_L. For search of other textures move to step 1. ■

The method of finding grid nodes for one-dimensional texture is obviously a simplification of the considered method without steps 3 and 4. In the case of texture with diamond grid step 3 is complicated, because the check for the second direction of the texture should be performed not in the neighborhood of one point, but within a predetermined angle $0 < \varphi \le \pi/2$ from $\mathbf{M}_1$ at the same distance from the origin. In this case, a sign that it is found a two-dimensional texture instead of two one-dimensional, is the presence of peaks $\left(m_1^i, m_2^i\right)$ of the form (2) for which $b_1 = b_2 = 1$. For generic grids search range of $\mathbf{M}_2$ is determined not only by the angle φ, but also by the maximum aspect ratio $\vartheta \in \mathbb{R} \cap (0;\infty)$: $1/\vartheta \le |\mathbf{M}_2| \le \vartheta$. For rectangular grids search range is defined only by ϑ because the angle is fixed.

Examples and Discussion

Figure 4 shows an example of the algorithm's work for the image form Figure 1a. As a result three peak grids are found which corresponds to three different regular textures.

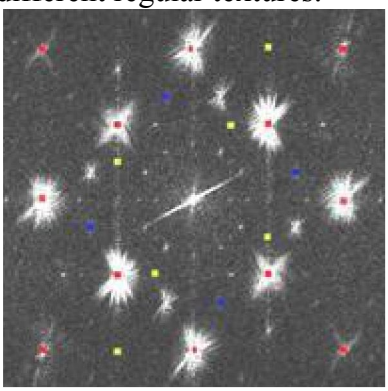
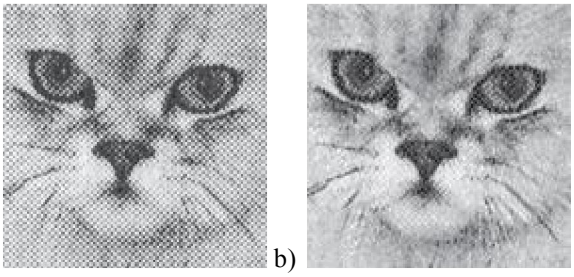
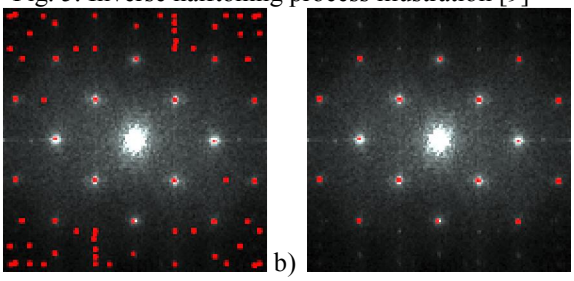


Fig. 4. Peak lattices on power spectrum of Fig. 1a, found by the proposed algorithm. Peaks of different lattices are marked with different colors



a) b)

Fig. 5. Inverse halftoning process illustration [9]



a) b)

Fig. 6. Illustration of first (a) and second (b) stages of the algorithm for Fig. 5a analysis

The feasibility of using the proposed algorithm for inverse halftoning problem (as well as for more general problem of periodic noise filtering) is illustrated in Figures 5-6. Fig. 5a shows a screened image and the result of its processing by Gaussian notch filter

constructed from the peaks of its power spectrum (example from [9]).

Figure 6 shows the result of analysis of image from Fig. 5a by the proposed algorithm: a – the first stage of peak search and selection, and b – the second stage of grid finding. As a result, one texture grid was found, which did not include some of the peaks selected at the first stage. Therefore, when we recover a halftone image there is no need to apply notch filter in neighborhood of these frequencies.

In Figures 7-8 we see examples of texture defect detection based on results of the proposed algorithm's work: a – original images, b – results of Gabor filtering at the frequency $\mathbf{M}_1$ of the found texture. Fig. 8c shows for comparison the results obtained by the well-known approach of Dmitry Chetverikov [4]. Visually apparent that the localization of the defect in Fig. 8b is performed better.

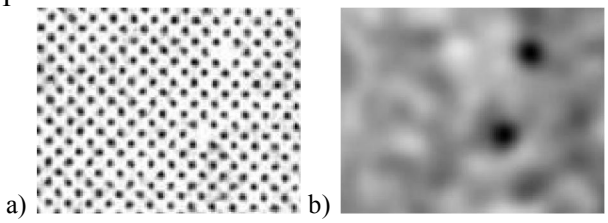


Fig. 7. Defect detection on the print sample

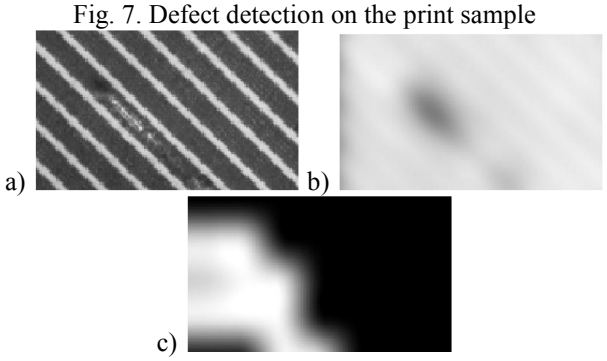


Fig. 8. Texture defect detection results: a – source image, b – by the proposed method, c) by Chetverikov's method [4]

Fig. 9 shows an example of extracting texture watermark with the use of our algorithm results. More details about benefits of its use in this problem can be found in [11].

Conclusion

In this paper, the algorithm for estimating the spectral characteristics of regular textures in digital image (frequency of the main harmonic, step of multiple harmonics in frequency space) is proposed. A number of examples show that the algorithm output can be successfully used to solve the problems of texture segmentation, defect detection and periodic noise filtering.

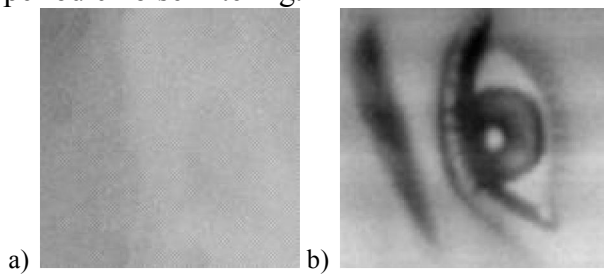


Fig. 9. Example of texture watermark extraction: a) source image, b) processing result

References

1. Haralick R. M. Statistical and structural approaches to texture //Proceedings of the IEEE. – 1979. – Vol. 67. – No. 5. – P. 786-804.
2. Lin H. C., Wang L. L., Yang S. N. Regular-texture image retrieval based on texture-primitive extraction //Image and Vision Computing. – 1999. – Vol. 17. – No. 1. – P. 51-63.
3. Fedoseev V. A., Mitekin V. A. Extraction method for textural watermarks of various linear raster patterns orientation type // Automation, Control, and Information Technology (ACIT), 2010 IASTED International Conference on. – Acta Press, 2010. – P. 15-17.
4. Chetverikov D. Fundamental structural properties of textures. – PhD. Th., MTA SZTAKI, Budapest, 2002.
5. Chan C., Pang G. K. H. Fabric defect detection by Fourier analysis //Industry Applications, IEEE Transactions on. – 2000. – Vol. 36. – No. 5. – P. 1267-1276.
6. Vartiainen J. et al. Detection of irregularities in regular patterns //Machine Vision and Applications. – 2008. – Vol. 19. – No. 4. – P. 249-259.
7. Caulier Y. et al. Segmentation and classification of anomalies in periodic structures //Journal of Electronic Imaging. – 2008. – Vol. 17. – No. 3. – P. 03301401-03301413.
8. Schowengerdt R. A. Remote sensing: models and methods for image processing. – Academic press, 2006.
9. Russ J. C. Processing images in frequency space //The Image Processing Handbook, ed. – 1995. – Vol. 2. – P. 283-346.
10. Soifer V. A. Computer Image Processing, Part I: Basic concepts and theory. – VDM Verlag. – 2009.
11. Sergeyev V., Fedoseev V., Mitekin V. Gabor Filter Based Attack on Printed Documents Protection Methods via Digital Watermarks //Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on. – IEEE, 2012. – P. 265-268.

TRACE TRANSFORM OF SPATIAL IMAGES[1]

N.G. Fedotov[2,3], S.V. Ryndina[2,4], A.A. Syemov[2,5]

[2] Penza State University (Penza, Russia), Volodarskogo, Penza, Russia
[3]440000, phone: +7(8412)524123, fedotov@pnzgu.ru
[4]440039, phone: +7(8412)552909, svetlanar2004@yandex.ru
[5]Prospekt Stroitelej, Penza, Russia, 440060,
mob. phone: +79374157515, t_pmd3c@mail.ru

The method considered in the article is a follow-up of the 2D Trace-transform method introduced and developed by N.G. Fedotov. A 3D Trace-transform both possesses the advantages of the 2D Trace-transform and uses the current practical experience in grappling with problems related to the invariance to a group of motion when analyzing and recognizing images.

Introduction

Research in pattern recognition has been carried out since the middle of the past century. Quite a number of various algorithms have been developed by now to solve pattern recognition tasks.

Technological developments of the late 20th century enhanced significantly the application of computer vision techniques and helped create 3D models of objects to solve a great variety of applied problems. The one of a predefined database automatic search for 3D models proves particularly topical.

Various techniques of spatial image analysis are generally highly tailored to analyze and recognize a specified type of objects. Moreover, there are very few techniques totally invariant to a group of motions of 3D objects, which would allow to obtain features more resistant to coordinate noise.

The present paper considers a mathematical model of applying Trace-transform to spatial images recognition, as well as the invariance of the features obtained to a group of motions.

Mathematical model of a Trace-transform

Suppose F is an original 3D model. Let us define a plane as tangent to the reference unit sphere with the center in the origin, passing through a base point x and at a distance r from the origin with the predetermined angles ω and φ, where η = [cosφ sinω, sinφ sinω, cosω] – is a unit vector in R^3, ω – angle between planes Π(η, r) and OXY, φ – angle between planes Π(η, r) and OXZ.

Let us scan the model with a grid of parallel planes with distance Δ*h* between the planes at angles *ω* and *φ*. The relative position of 3D object *F* and each scanning plane Ï $(\boldsymbol{\eta}(\omega,\varphi),r)$ are characterized by the number G, computed according to a certain rule *HyperT*: *G*=HyperT(F∩Π(η(ω,φ),r)). The characteristics indicated could be provided by a sectional area, character of the neighborhood of such an area, *etc*. Then, scanning should be performed for the new value of angle *ω*, which has got the discrete increment Δ*ω,* with a lattice of planes with the same distance Δ*h* between the scanning planes. Again, we apply the same selected *HyperT* rule to the meet of the new plane Π(η(*ω*+Δ*ω*,*φ*),r) with the 3D object *F*, and go on like that until we make a halfway equal to π radian.

Scanning is going on for the new value of angle *φ*, with the discrete increment Δ*φ*. In a like manner, we "go round the object" from Π(η(*ω*,*φ*),r) to Π(η(*ω*,*φ*+π–Δ*φ*),r) for each angle *φ*+Δ*φ*.

For the density of planes to stay uniform, it is significant that angles are not to be changed arbitrarily, but according to a support grid to be considered in the *Features invariant to rotation* Section of the article.

The result of computing *HyperT*-functional depends upon three parameters (r,ω,φ) of the

[1] Support of the grant project №12-07-00501

plane, hence the set o numbers $G = HyperT(F \cap \Pi(\eta(\omega,\varphi),r))$ obtained through scanning, is a *hypertrace*-transform. Therefore, if every image, obtained through cutting the reference 3D model with a plane, is matched with a certain informational feature $\Pi(F)$ by the rule *HyperT*, then with the numerical analysis, the result of 3D trace-transformation could be conveniently represented as a three-dimensional *hypertrace*-matrix ***3GM***, its axis 0ω being directed horizontally, axis 0φ – vertically, and axis 0r – depthward.

So, every row of the three-dimensional matrix contains element features to be computed from the images which result from the meets of scanning planes with the reference object, all angles ω (from ω to $\omega+\pi-\Delta\omega$) having been detoured, with the fixed values of angle φ and distance r. Accordingly, every column of the matrix contains element features to be computed from the images which result from the meets of scanning planes with the reference object, all angles φ (from φ to $\varphi+\pi-\Delta\varphi$), with the fixed values of angle φ and distance r. And, finally, every depth row of the matrix contains element features to be computed from the images which result from the meets of scanning planes with the reference object, for all values of distance r, values of angles ω and φ , being fixed.

Hence, couple $(\omega_i, \varphi_j, r_k)$ matches matrix element with the number (i, j, k) and the meaning $\Pi(F)$.

According to the theory considered, three-dimensional *hypertrace*-matrix having been filled, *hyper*quadruple feature forming goes on. With the *hyper*diametric functional *HyperR*, we process a depth row of matrix ***3GM***. It can be assigned, for example, as $\int HyperT(\omega, \varphi, r)dr$. The processing having been performed, the given three-dimensional *hypertrace*-matrix ***3GM*** becomes a two-dimensional matrix ***2GM***, its each column and row representing a 2π-periodic curve. Processing column after a column with the functional *HyperΩ*, we obtain a system of numbers – an ordinate vector, its indiscrete analog being a 2π- periodic curve. Then the system of numbers obtained is treated with the 3D-cyclic functional *HyperΦ*, which results in a certain feature of the projection image Res(F). It can be assigned, for instance, by the second harmonic amplitude, in a Fast-Fourier transformation in its indiscrete variant.

Hence, the structure of the studied image feature is composed of four *hyper*functionals: $\mathrm{Re}\,s(\mathrm{F}) = Hyper\Phi \circ Hyper\Omega \circ Hyper\mathrm{R} \circ Hyper\mathrm{T}(\mathrm{F}_{\mathrm{sect}})$.

Let us describe now the creation of informational feature $\Pi(F)$ of the in-plane figure, and the *hypertrace*-functional *HyperT* .

Each image on the projection of the reference 3D model on each of the built planes should be scanned to extract certain valuable features of distinctness (*e.g.*: figure area, contour perimeter, specific points, *etc.*). To find a feature of a 2D image of the projection image, the 2D trace-transform introduced and described by N.G. Fedotov [1], is used.

The image of a figure on the section F_{sect} is scanned with a lattice of parallel lines $l(p, \theta)$ with the distance Δp between the lines. Parameters p and θ are the section plane polar coordinates of line l. The feature of the image in question is structured as a composition of three functionals: $\Pi(F) = HyperT(F_{сеч}) = \Theta \circ P \circ T(F \cap l(\rho,\theta))$.

Please note that the coordinate system position on the section plane and its orientation as to the section image is absolutely irrelevant, feature obtained by 2D trace-transform is totally invariant to translation and rotation of the 2D-image on the projection plane [2].

Hence, the feature of the 3D-model in question is composed of four *hyper*functionals, one of them being applied to the joint impact of three 2D-functionals.

To obtain a greater discriminating power, a feature vector of the image is formed by selecting different combinations of 2D and 3D analogies of circular, diametric and trace-functionals.

Now then image scanning itself involves obtaining distribution of sections (or projections with a certain choice of a trace-functional, *etc.*) from all field-of-vision angles (with the required accuracy). The type of scanning in question correlates with the natural physical type of scanning, when a physical object to be digitized it is placed onto a plane and scanned from all directions. It helps form 3D model features at the stage of scanning, for scanning and processing are performed in like manner.

Features invariant to translation

The property of translation invariance for every functional ζ will be of the form: $\zeta(f(x+a)) = \zeta(f(x))$ for all permissible *f(x)* and $\forall a$.

It is common knowledge that the form of a 3D object projection onto a fixed plane stays the same when transferring the object to any vector *V*. Hence, transferring a reference 3D object only increases the dimension of the corresponding trace-matrices by a number of zero-elements, transferring them along corresponding axes. Thereafter, if functional HyperR only processes all zero elements in a Trace-matrix 3GM, all its values will be alike.

Features invariant to rotation

The standard enumeration of all the angles ω and φ of the planes' grid, for an indiscrete case, topologically, yields a model of a unit sphere. Let us pair each plane for r=1 with its point of tangency with a given sphere.

For a discrete case, a common global map shows a denser cluster of points in the neighborhood of poles then at the equator. That is why we need to build a support net in a different manner to provide for a uniform distribution (density) of points on the sphere.

Such a net can be obtained in a variety of ways, *e.g.:* using regular polyhedrons inscribed in spheres. Such type of polyhedrons, though, is bounded from above, and maximum accuracy can be obtained when n=20.

Another way is to form a net as a set of points which are intersections of lines passing through the sphere center, with the sphere itself. (Fig.1).

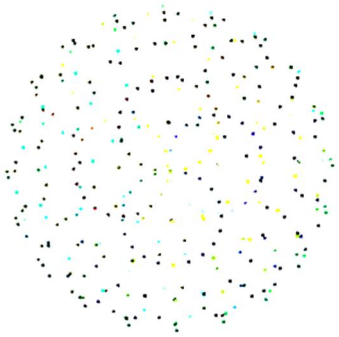


Fig. 1. Distribution of points at the intermediate stage: two poles are emerging.

Formation is performed according to the formulae (rcosφ sinω, rsinφ sinω, rcosω), where $\varphi \in [0, 2\pi - \Delta\varphi]$ *and* $\omega \in [0, 2\pi - \Delta\omega]$, $\Delta\varphi = \Delta\omega$.

Such a sphere lacks poles, or, in other words, every point based on the sphere can be considered a pole. Such designs result in forming triangles on the sphere (Fig.2).

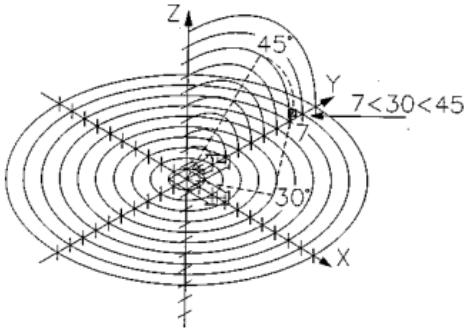


Fig. 2. Pattern of point distribution on a sphere.

Looking for a stronger evidence to the fact of that sort of distribution uniformity is beyond the scopes of the present article.

In the functional view, invariance of features looks as follows: $\zeta(f(Z \cdot x)) = \zeta(f(x))$ for all permissible *f(x)* and $\forall a$.

In terms of trace-matrices, rows and columns of matrix 2GM will be shifted to the *ω-th* and the *φ-th* number frontwards or backwards, correspondingly, which depends on the angle sign, their sequence order remaining flat.

Accordingly, its *hypertrace*-image (*i.e. hypertrace*-matrix graphic representation) will be shifted along the horizontal axis 0ω and the vertical axis 0φ to a certain distance equal to the corresponding angles of rotation.

Sensitivity to scaling

The property of sensitivity to scaling with the degree of sensitivity k for every functional ζ will be of the form: $\zeta(\mu \cdot f(x)) = \mu^k \cdot \zeta(f(x))$ for all permissible *f(x)* and $\forall a$.

For homothety, a depth row range (axis 0r) of matrix 3GM expands ($\mu > 1$) or contracts ($\mu < 1$), for the scaled object *F'* would intersect μ times more or fewer planes than the reference object *F*. Hence the number of non-zero values in the *i*-th row of matrices 3GM' and 3GM for objects F' and F, correspondingly, will be μ times different.

Results

The authors of the present article are now carrying out experiments to verify the validity of the method suggested. The first results obtained from the simplest 3D images indicate the efficiency of the method suggested.

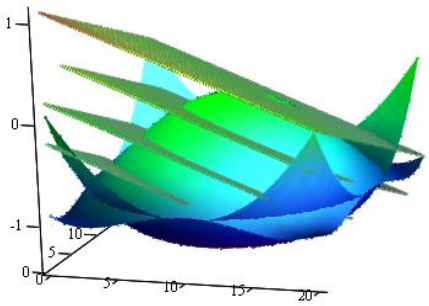


Fig. 3. Object cut with parallel planes.

Fig. 3 illustrates an example of trace-transform application to a spatial image.

Conclusion

It should be noted that, the scanning process being discrete, in practice, we can only state an approximate invariance of a feature.
Nevertheless, a scan step decrease results in a higher scanning accuracy. That is why the research in quadruple features stability with reference to the scan step and coordinate noise, proves topical for image recognition.
An additional point to emphasize is that applying the indicated features of functionals in a *hyper*quadruple composition at each stage, we can obtain the resulting feature invariant to translation and rotation, but sensitivity to scaling.

References

1. Fedotov N.G. The Theory of Image-Recognition Features Based on Stochastic Geometry // Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications. – 1998. – V. 8, № 2. – P. 264-266.
2. Fedotov N.G. Metodi stohasticheskoi geometrii v raspoznovanii obrazov. – M.: Radio I Svaz, 1990. – 144 p (in Russian).

3D SCENE RECONSTRUCTION FROM UNRECTIFIED STEREO IMAGES[1]

V. Fursov[2,3], S. Bibikov[2,3], Ye. Goshin[2,3]

[2] Image Processing Systems Institute of Russian Academy of Science
151, Molodogvardeyskaya st., 443110 Samara, Russia,
fursov@smr.ru
[3] Samara State Aerospace University,
34, Moskovskoye shosse, 443086 Samara, Russia,
fursov@ssau.ru

Our work provides a novel way to the 3D-scene reconstruction from stereo pairs of images. A traditional approach supposes a rectification of stereo images before searching for corresponding points and disparity map generation. In this paper we propose an approach in which the corresponding points are determined directly on the epipolar lines on the unrectified stereo images. A complete three-dimensional model of an object is constructed by scanning all epipolar lines. A set of experiments of scene reconstruction is presented.

Introduction

Image matching is an important part of stereo computer vision algorithms. When stereo images are taken from two distinct viewpoints, some difficulties are presented. In this case projective distortions on stereo pairs are usually significantly different from one another. To overcome these difficulties, a rectification technique is used. Rectification of stereo images is a transformation, in which corresponding points on stereo images are arranged in the same rows. Mainly, two types of rectification are used: the projective (or planar) one [1][2] the and polar one [3][4]. They have various scopes of application, but there is a common disadvantage of the rectification approach. It is the necessity for conversions of the images. The images and objects are distorted considerably, owing to projective or polar transformation. E.g., feature points detection performed on the new interpolated image will cause increase in errors.

Application of the technology of a 3D scene reconstruction proposed in this paper to a considerable degree allows to avert the above mentioned problem. This method has much in common with the polar rectification approach, except that the rectified images are not generated. Matching points are searched directly on the image on the epipolar lines. Then, for each pair of corresponding points, spatial coordinates of the scene are computed.

Problem definition

In this paper the obscura camera model is used. [5] Let us consider the case when the intrinsic parameters of the cameras, as well as their position and orientation are known. To characterize them, camera parameters matrices are defined:

$$\mathbf{K}_1 = \mathbf{K}_2 = \mathbf{K} = \begin{bmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad (1)$$

where f is focal length of the cameras, (u_0, v_0) are cameras' principal points location in the coordinate system associated with these cameras. [6]

Let us introduce a global coordinate system and the coordinate systems of first and second cameras with their centers at points $\mathbf{c}_1$, $\mathbf{c}_2$ in the global coordinate system.

Suppose $\mathbf{M}$ is a coordinate vector of some point in the global coordinate system. Coordinate vectors of this point in the

[1]This work was financially supported by the Samara Region Administration and the RFBR (grants № 12-07-00581-a, № 12-07-31208).

coordinate systems of the first and second cameras $\mathbf{m}_1$ and $\mathbf{m}_2$ are defined as

$$\mathbf{m}_1 = \mathbf{P}_1\mathbf{M}, \quad (2)$$

$$\mathbf{m}_2 = \mathbf{P}_2\mathbf{M}, \quad (3)$$

where the projection matrices are defined as

$$\mathbf{P}_1 = \mathbf{K}_1\left[\mathbf{R}_1 \vdots \mathbf{t}_1\right], \quad (4)$$

$$\mathbf{P}_2 = \mathbf{K}_2\left[\mathbf{R}_2 \vdots \mathbf{t}_2\right]. \quad (5)$$

$\mathbf{R}_1$, $\mathbf{R}_2$ are the matrices of 3×3-dimension, describing the rotation of the coordinate systems of the first and second cameras on the global coordinate system, and $\mathbf{t}_1 = \left[t_{1,x}, t_{1,y,} t_{1,z}\right]^T$, $\mathbf{t}_2 = \left[t_{2,x}, t_{2,y,} t_{2,z}\right]^T$ are coordinates of the origin of the global coordinate system in the coordinate systems of the first and second cameras respectively, defined as

$$\mathbf{t}_1 = -\mathbf{R}_1\mathbf{c}_1, \quad (6)$$

$$\mathbf{t}_2 = -\mathbf{R}_2\mathbf{c}_2. \quad (7)$$

If appropriate projections $\mathbf{m}_1, \mathbf{m}_2$ (2), (3) on the first and second camera images are known, point $\mathbf{M}$ coordinates in three-dimensional space can be calculated as the intersection of rays $(\mathbf{c}_1, \mathbf{m}_1)$ and $(\mathbf{c}_2, \mathbf{m}_2)$.

Correspondence points matching

The corresponding points are searched on the sets of points, located on epipolar lines $\mathbf{l}_1$ and $\mathbf{l}_2$, which are defined using a given (or computed) fundamental matrix [7][8].

Elementary fragments of a specific pair of points in the images of these sets are defined on the neighborhood of these points, limited by the squares whose vertices are the four nearest pixels on the image, i.e. sides of the square are limited to two consecutive (by rows and columns) pixels.

Let us assume that the distance between the nearest pixels is equal to 1, and the coordinates of the vertices of the unit square are defined in the associated reference system and are, respectively, (0,0), (0,1), (1,0), (1,1).

Let (x_1, y_1) and (x_2, y_2) are points coordinates in the reference system related to the unit squares. These points belong to these unit squares and lie on epipolar lines $\mathbf{l}_1, \mathbf{l}_2$, and $f(x_1, y_1)$, $f(x_2, y_2)$ are the intensity values at these points obtained by bilinear interpolation of intensity values in vertices $f_1(0,0)$, $f_1(0,1)$, $f_1(1,0)$, $f_1(1,1)$ and $f_2(0,0)$, $f_2(0,1)$, $f_2(1,0)$, $f_2(1,1)$:

$$f(x_1, y_1) \approx a_1 + b_1x_1 + c_1y_1 + d_1x_1y_1, \quad (8)$$

$$f(x_2, y_2) \approx a_2 + b_2x_2 + c_2y_2 + d_2x_2y_2, \quad (9)$$

where

$$a_1 = f_1(0,0), \quad a_2 = f_2(0,0)$$

$$b_1 = f_1(1,0) - f_1(0,0), \quad b_2 = f_2(1,0) - f_2(0,0)$$

$$c_1 = f_1(0,1) - f_1(0,0), \quad c_2 = f_2(0,1) - f_2(0,0)$$

$$d_1 = f_1(0,0) - f_1(1,0) - f_1(0,1) + f_1(1,1),$$

$$d_2 = f_2(0,0) - f_2(1,0) - f_2(0,1) + f_2(1,1).$$

Let us refer to the fragments consisting of several unit squares as composite, and the interpolated intensity values at given points on the epipolar lines belonging to these squares are designated as $f^{(k)}(x_1, y_1)$, $f^{(k)}(x_2, y_2)$, $k = \overline{1, N}$, where N is the number of unit squares (including the central one), and $(x_1, y_1), (x_2, y_2)$ are the coordinates of given points on epipolar lines $\mathbf{l}_1, \mathbf{l}_2$ in the reference systems associated with the unit squares. Figure 1 shows the examples of composite fragments formation on stereo images for nine points along epipolar lines $\mathbf{l}_1, \mathbf{l}_2$. It should be pointed out that the comparison of fragments of the first and second images should be performed in the same order. Thus, depending on a specified step, several points may appear on the same elementary fragment.

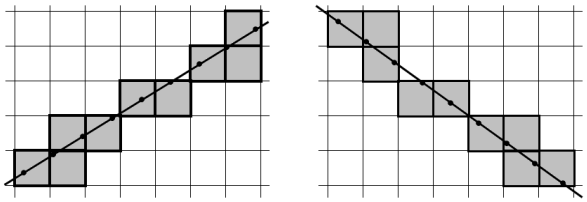


Fig. 1 – Fragment construction example

Descriptors, based of intensity values $f^{(k)}(x_1, y_1), f^{(k)}(x_2, y_2)$, $k = \overline{1, N}$, are calculated at nine points on these fragments by bilinear interpolation of the values in the vertices of the respective unit squares.

3D-scene points reconstruction

Once points $\mathbf{m}_1, \hat{\mathbf{m}}_2$ have been matched, point M coordinates can be calculated. To solve this problem, ratio (2), (3) with (4), (5), (6), (7) for homogeneous coordinates can be rewritten as follows:

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = \frac{1}{g_1}\mathbf{K}_1\mathbf{R}_1\left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} - \mathbf{c}_1\right), \qquad (10)$$

$$\begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \frac{1}{g_2}\mathbf{K}_2\mathbf{R}_2\left(\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} - \mathbf{c}_2\right), \qquad (11)$$

where

$[x_1, y_1, 1]^T = \mathbf{m}_1$, $[x_2, y_2, 1]^T = \hat{\mathbf{m}}_2$.

Equating the coordinate vectors, on basis of equations (10), (11), the following equation is obtained:

$$g_1\mathbf{R}_1^{-1}\mathbf{K}_1^{-1}\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} - g_2\mathbf{R}_2^{-1}\mathbf{K}_2^{-1}\begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \mathbf{c}_2 - \mathbf{c}_1 . \qquad (12)$$

Expression (12) is a system of three equations in two unknowns g_1 and g_2.

Since points $\mathbf{m}_1$ and $\hat{\mathbf{m}}_2$ were found with the epipolar constraints, only one point in 3-dimensional space corresponds to these points; therefore g_1 and g_2 are also uniquely determined. Then the coordinates of point $\mathbf{M}$ for predetermined point $\mathbf{m}_1$ and its corresponding point estimation $\hat{\mathbf{m}}_2$ are calculated using relations (10) or (11). These calculations are repeated for all pairs of corresponding points belonging to the epipolar lines of the same plane. A complete three-dimensional scene is formed by combining all of the points obtained on the sheaf of planes.

Outcome of experiment

For experimental verification of the technology operability, a three-dimensional model of the scene is used. This model represents a plane with some subjects situated on it. A 3D-model of the scene was constructed with the use of a raytracing program *POV-Ray*.

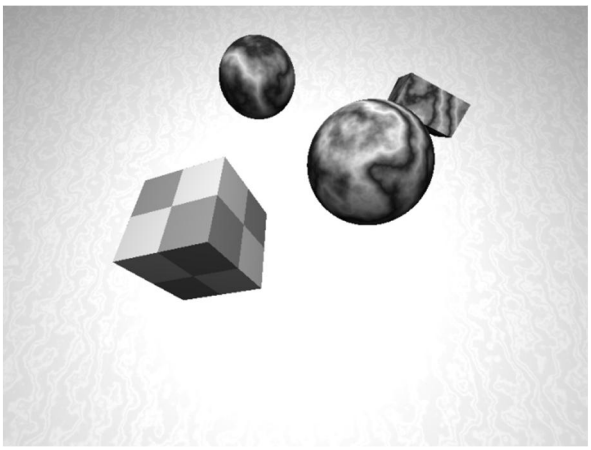


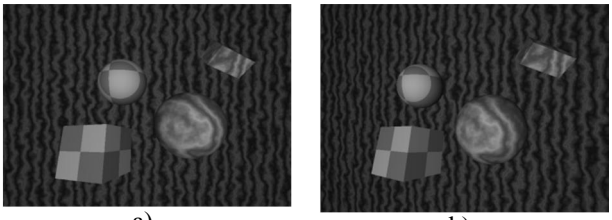
Fig. 2 – Scene overview

Figure 2 shows an overview of the scene model used in the experiment. Figures 3a and 3b show two stereo images of the scene, which were obtained with the following camera parameters (hereinafter, indices 1 and 2 are used to designate the left and right views, respectively):

$$\mathbf{K}_1 = \mathbf{K}_2 = \mathbf{K} = \begin{bmatrix} 400 & 0 & 400 \\ 0 & 400 & 300 \\ 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{R}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{C}_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

$$\mathbf{R}_2 = \begin{bmatrix} 0.9848 & 0 & 0.1736 \\ 0 & 1 & 0 \\ -0.1736 & 0 & 0.9848 \end{bmatrix}, \mathbf{C}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$



a) b)

Fig. 3 – Stereo images: a) left view; б) right view.

For these cameras parameters a fundamental matrix is computed as follows:

$$\mathbf{F} = \begin{bmatrix} 0 & 7.66292 \cdot 10^{-6} & -0.00229887 \\ 0 & 0 & -0.0176516 \\ 0 & 0.0143183 & 1 \end{bmatrix}$$

This matrix is used to generate epipolar lines on the left and right images.

For the matching on the epipolar lines, vectors of 9×1 dimension are used. These vectors are

composed only of intensity values of nine elementary fragments along the epipolar line.
Figure 4 shows the results of the 3D-scene reconstruction mentioned above. These results are obtained using the four above-mentioned point descriptions. It is obvious that as the dimension of description vectors increases, the quality of the scene reconstruction increases as well, while the noise intensity connected with matching errors decreases.
However, in all the experiments the size of the composite fragment containing only 9 unit squares along the epipolar lines remains low. Increased information content of the composite fragments is achieved by increasing the number of calculated characteristics. It is clear that this is accompanied by an increase in computing, in exchange for quality.

Fig. 4 – Example of scene reconstruction

Conclusion

The absence of an explicit rectification of images and usage of small fragments for matching make the proposed technology advantageous in terms of reducing the computational complexity. At the same time, the results of experiments show that the use of epipolar constraints and composite fragments allows to obtain a sufficiently high quality scene reconstruction.

References

1. Hartley, R. I. (1999). Theory and practice of projective rectification. International Journal of Computer Vision, 35(2), 115-127.
2. Monasse, P., Morel, J. M., & Tang, Z. (2010). Three-step image rectification. In Proceedings of the British Machine Vision Conference.
3. Pollefeys, M., Koch, R., & Van Gool, L. (1999). A simple and efficient rectification method for general motion. In Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on (Vol. 1, pp. 496-501). IEEE.
4. Häming, K., & Peters, G. (2007, May). Extension of the generalized image rectification–Catching the infinity cases. In Proc. 4th International Conference on Informatics in Control, Automation, and Robotics (ICINCO 2007)(J. Zaytoon, J.-L. Ferrier, JA Cetto, and J. Filipe, eds.) (Vol. 2, pp. 275-279).
5. Forsyth, D. & Ponce J. (2004). Computer Vision: A Modern Approach. Moscow: “Williams” Publisher.
6. Hartley, R., & Zisserman, A. (2000). Multiple view geometry in computer vision (Vol. 2). Cambridge.
7. Fursov V. Conformed Identification of the Fundamental Matrix in the Problem of a Scene Reconstruction, using Stereo Images / V. Fursov, Ye. Goshin // Image Mining. Theory and Applications. Proceedings of IMTA-4 2013. – 2013. – P. 29-37.
8. Luong, Q. T., & Faugeras, O. D. (1996). The fundamental matrix: Theory, algorithms, and stability analysis. International Journal of Computer Vision, 17(1), 43-75.

IMAGE PREPROCESSING FOR STOCHASTIC GRADIENT ESTIMATION OF INTERFRAME GEOMETRICAL DEFORMATIONS USING FEATURE POINTS[1]

P.V. Jakshankin[2,3], A.G. Tashlinskii[2,4]

[2] Ulyanovsk State Technical University, ul. Severnyi Venets 32, Ul'yanovsk, 432027 Russia, telephone (88422)778102, e-mail: [3]yakspavel@yandex.ru, [4]tag@ulstu.ru

Image preprocessing algorithm increasing the parameters' domain for based on feature points stochastic gradient estimation of interframe geometrical deformations is considered. The preprocessing includes the Gaussian filter's applying.

Introduction

Stochastic gradient algorithms (SGA) are good choice for estimation of interframe geometrical deformations of images (IGDI) [3]:

$$\hat{\bar{\alpha}}_{t+1} = \hat{\bar{\alpha}}_t - \Lambda_{t+1}\bar{\beta}_{t+1}(Q(Z_{t+1}, \hat{\bar{\alpha}}_t)), \qquad (1)$$

where $\bar{\alpha}$ - vector of estimates of IGDI's parameters; $\bar{\beta}$ - estimate of the gradient of the objective function measuring the quality of obtained estimates, Λ_t - the learning rate matrix defining the increase of parameters' estimates on t-th iteration; Z_{t+1} - samples drawn from $Z^{(1)}$ and $Z^{(2)}$ used to find $\bar{\beta}$ on iteration $(t+1)$.

In this work the vector of parameters $\bar{\alpha} = (h_x, h_y, \varphi, k)$ was examined, where h_x and h_y - translations among axes, φ - rotation angle and k - scale factor.

Assuming the initial guess for translation $h_0 = 0$, rotation angle $\varphi_0 = 0^0$ and scale factor $k_0 = 1$, SGA allows estimating IGDI parameters in a relatively small range. The suggested preprocessing procedure objects to expand the effective range of SGA.

Preprocessing

The preprocessing procedure consists of two consecutive steps – feature detection on the reference and target images and formation of 2D Gaussian waves around each feature point.

In the first stage on the reference image (fig. 1,a) and the target image (fig. 1,b) feature points are allocated using FAST detector [2] (fig. 1,c and d). This detector is chosen due to its stable results obtained by relatively short period of time. The idea of the detector is simple – the Bresenham [1] circle of radius R is considered around each pixel (let us denote its intensity as L_0). A point is recognized as a feature if there is an arc of N pixels on the circle N with the intensities $L_1 \dots L_N$ so that $\forall i \in (1..N)\ L_i \geq L_0 + \lambda$ or $\forall i \in (1..N)\ L_i \leq L_0 - \lambda$, where λ - a threshold. Let us modify the algorithm: associate each feature point with the periodicity p using information from the detector:

$$p = \frac{N}{l}, \qquad (2)$$

where l - the length of Bresenham circle. Radius $R = 3$ was examined. Consider this radius the length of the circle l is equal to 16 pixels and possible periodicities p depending on the arc length N are shown in table 1. Let us notice that for this radius the minimal value of N recognizing a point as a feature usually taken equal to 9.

Table 1. The probabilities of feature points

N	< 9	9	10	…	16
p	not a feature point	9/16	10/16	…	1

Some points may not be features. Such points are not of interest for the problem. Due to this

[1] Support of the RFBR grant 12-01-97014

fact, they are changed to black pixels on the reference and target images.

In the second stage (fig. 1,e and f) the 2D Gaussian wave is formed around each feature point with coordinates (x_h, y_h) on the reference and target images. The parameters of these waves are the width σ and magnitude A_h equal to probability p_h evaluated in the first stage. The greater the arc length N, the more likely it is that a point will stay feature in case of geometrical and intensities' deformations. Therefore, the influence of such points should be greater. Thus, the intensity L of not feature point with coordinates (x, y) is defined using equation:

$$L = \sum_{h \in H} A_h L_h e^{-\frac{(x-x_h)^2+(y-y_h)^2}{2\sigma^2}}, \quad (3)$$

where H - the set of all feature points.

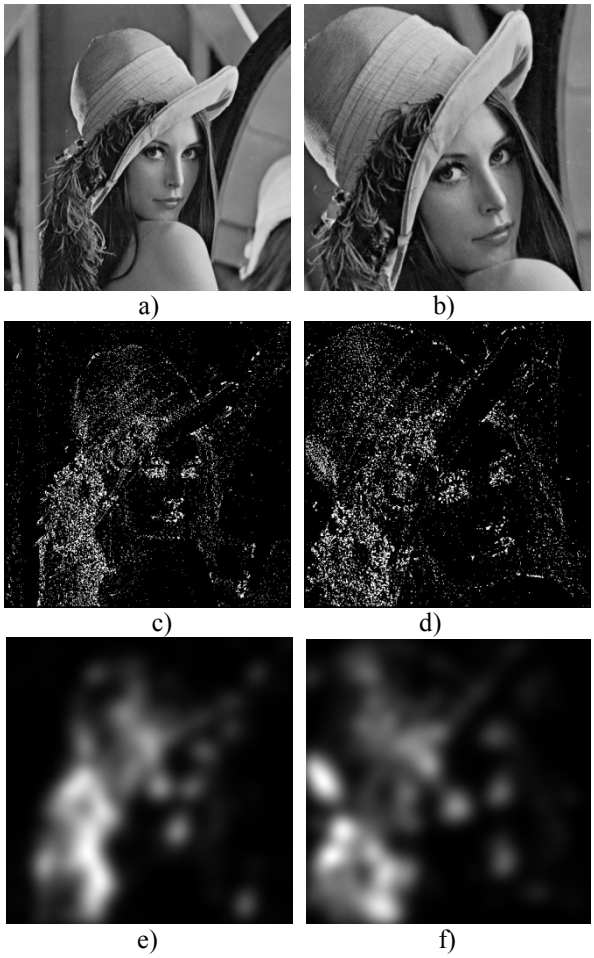


Fig. 1. Preprocessing of reference and target images for estimation of IGDI parameters.

Let us note that the value of width parameter σ is a tradeoff between required accuracy of estimates and the width of the SGA effective range – the greater σ, the broader SGA effective range would be but with lower accuracy of estimates.

After preprocessing, SGA is used to estimate IGDI. Due to inefficiency, points not influenced by features are excluded from the samples.

Experiments

To show the work of the algorithm let us consider some experimental results. The image Lena (fig. 1,a) from the library of South California University was used. Table 2 shows the parameters of the experiment.

Table 2. The experiment parameters

Experiment No		1	2	3	10
IGDI	h_x	5	10	…	50
	h_y	5	10	…	50
	φ	5	10	…	50
	k	1.05	1.1	…	1.5
σ^2 additive white noise		5	5	…	5
The iteration number		50	50	…	50
Confidence interval of the IGDI parameters		$0{,}9 * h_x \le \hat{h}_x \le 1{,}1 * h_x$ $0{,}9 * h_y \le \hat{h}_y \le 1{,}1 * h_y$ $0{,}9 * \varphi \le \hat{\varphi} \le 1{,}1 * \varphi$ $0{,}9 * k \le \hat{k} \le 1{,}1 * k$			

Even if only one parameter falls out of the confidence interval, the estimation procedure was considered as a failure. Fig. 2 shows the dynamics of failures' number for each experiment. One can see that the number of failures of SGA without preprocessing in the third experiment is already equal to 41, while SGA with preprocessing performed well up to tenth experiment, i.e. for this image and this parameters the effective range of SGA was expanded by three times at least.

At the same time, according to the table 3 there was a loss of accuracy due to the use of the Gaussian filter and incomplete information about the images (only feature points remain).

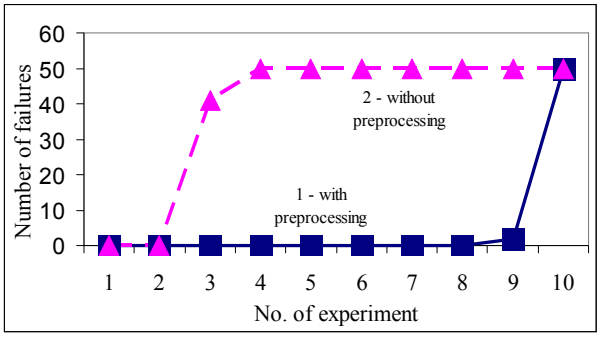


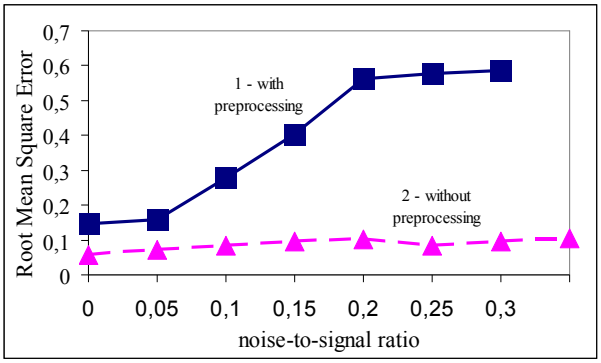
Fig. 2. The dynamics of failures' number of SGA with preprocessing (1) and without (2).

Table 3. The mean of failures of IGDI parameters estimation

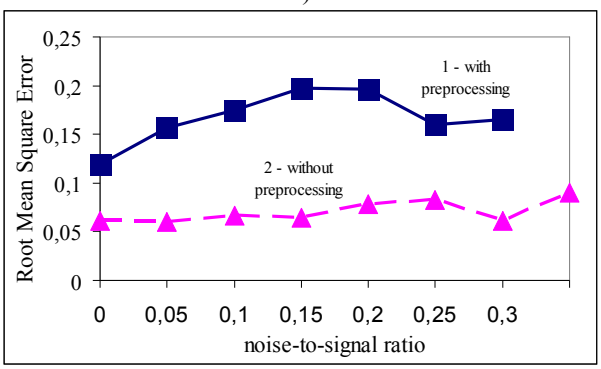
Experiment No	The mean errors (above – for SGA without preprocessing, below – for SGA with preprocessing)			
	h_x	h_y	φ	k
1	0,11 0,05	0,13 0,07	0,09 0,02	0 0
2	0,19 0,05	0,16 0,06	0,09 0,04	0 0
3	0,22 0,13	0,19 0,18	0,14 0,12	0 0
4	0,44 -	0,24 -	0,13 -	0 -
5	0,49 -	0,86 -	0,20 -	0,01 -
6	0,63 -	0,38 -	0,27 -	0 -
7	1,03 -	0,3 -	0,34 -	0,01 -
8	1,73 -	0,48 -	0,61 -	0,02 -
9	2,83 -	3,27 -	1,14 -	0,04 -
10	-	-	-	-

The second experiment was objected to examine the influence of intensities' distortions on the estimation quality. Fig. 3,a and fig. 3,b show the dependence of mean square error of estimates of parameters h_x and φ on noise-to signal ratio. In this case, the failure was considered if the absolute error of translation estimation was higher than 2 or the absolute error of rotation angle estimation was higher than 2°.

Let us notice that when noise-to-signal ratio is greater than 0.3 there was a failure in each experiment. False feature points and undetected real feature points due to noise cause this.



a)



b)

Fig. 3. The dependence of mean square error of estimation of translation (a) and rotation angle (b) on the noise-to-signal ratio.

Conclusion

Examined preprocessing of images can considerably expand the effective range of SGA for the problem of IGDI parameters' estimation. SGA with preprocessing is more sensitive to noise due to FAST detector using raw intensities. The use of only feature points with applying Gaussian filter influences the accuracy of estimation. However, SGA with preprocessing can be used as a good initial guess for further estimation with higher accuracy.

References

1. Rogers D. Algorithmic principles of machine graphics. — M.: Mir, 1989. (in Russian).
2. Rosten E. Fusing points and lines for high performance tracking / E.Rosten, T. Drummond // IEEE International Conference on Computer Vision, vol.2, 2005 – pp. 1508–1511.
3. Tashlinskii A.G. Estimation of parameters of spatial deformations of images' sequences. – Ulyanovsk: izdatelstvo UlGTU. 2000. 132 p. (in Russian).

NIGHT IMAGE ENHANCEMENT VIA BRIGHT CHANNEL PRIOR

Hoyong Jang[1,2], Jungyu Kang[1,3]

[1] Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea
[2]hoyong.jang@kaist.ac.kr, [3]cmiller2air@kaist.ac.kr

In this paper, we introduce a uncomplicated but efficient algorigthm to enhance night image. We define features of night images and extract the illumination of an image by choosing brightest channel of region and we obtain enhanced images with compensation of dark region based on Retinex theory. The results show that our methods have high performance despite its simplicity.

Introduction

There are many algorithms for object detection and recognition and they have high performance[6][7] in daytime. They, however, show poor detection or recognition results in night because of low and variable illumination.

Many algorithms have been developed to enhance night images and they can be divided into two ways, 'Contrast Enhancement' and 'Retinex Theory'. The contrast enhancement is to make better distribution of contrasts. Histogram Equalization is one of these algorithms; it makes an image have uniform contrast distribution. Though histogram equalization showed good result with short computation time, it has some limitations. First, it increases contrast of an object that has same color under different illuminations. Second, it also increases contrast of noises in an image, so histogram equalization of noise images show poor results.

Retinex theory, another algorithm, describe a intensity of some point (x,y) in an image represented as product of illumination and reflection value of the point. It shows better performance than histogram equalization, but there is a question that how illumination can be estimated from an image exactly.

We propose a novel algorithm to estimate illumination by using Bright Channel Prior and perform night image enhancement. We will describe details at next sections.

Related Works

Yamasaki et al.[1] did night image enhancement on surveillance camera. It used a daytime scene of same place and made background bright so the camera could detect moving object easily. It, however, needs daytime scene of same place and only enhances background.

There are some algorithms[2][3] that use retinex theory for night image enhancement. Rao et al.[2] converted an RGB channel of image to HSI channel and consider intensity values filtered by low pass filter as illumination and saturation values as reflection.

Wong et al.[3] took probability theory to estimate illumination. The algorithm calculates illumination values if there are given values of product of illumination and reflection.

These algorithms didn't fully decompose illumination from an image. In [3], picture was taken under white light but estimated illumination had color(ex. hair, eyes).

Panagopoulos et al.[4] proposed an algorithm which takes bright channel prior for estimating and removing shadows.

In this paper, we choose bright channel prior for estimating illumination because it is less affected by reflection of an image and compensate dark region of the image based on retinex theory.

Approach

First, features of night images should be considered. We will treat only images with artificial light, for example, streetlight, light from building, car light and etc. In other cases, only moonlight exists, histogram equalization would show best results because of uniform distribution of illumination.

In the night image defined above, we can find two features.

- At least 1 light source exists in the image.
- There is large contrast between light source regions and other regions.

Histogram equalization cannot work well because of above features. However the image can be enhanced with making dark region bright if we know exact illumination of a night image.

A feature, at least one pixel has high intensity in bright region of images, informs brightness of that region. For above feature, we denote bright channel prior which takes an idea from Dark Channel Prior[5] to estimate exact illumination.

$$J^{bright}(\mathbf{x}) = \max_{y \in \Omega(\mathbf{x})} \left(\max_{c \in \Omega(r,g,b)} J^{c}(y) \right) \qquad (1)$$

A patch is set with size n x n and a pixel value of bright channel prior is the maximum value of the patch. A pixel of bright channel prior can have high value in two cases; 1) a pixel in bright region and 2) a pixel has white color. We, however, don't need to consider the case 2) because a pixel in white and dark region has low value due to low light.

Fig. 1. Original image (left)
and Bright Channel Prior (right)

The result of bright channel prior is shown in Fig. 1 and it reveals illumination independent of colors of objects.

Image Enhancement

We should infer which region has little or a lot of illumination. Since bright channel prior was derived from patches, shape of patch affects estimated illumination. For this reason, the estimated illumination is filtered with Gaussian filter.

To obtain amount of illumination needed, we call that Illumination Shortage, at a dark pixel, we subtract the estimated one from saturation values which is the maximum value of intensity. Since the difference of actual brightness is non-linear to that of intensity values of the image,

$$y' = \frac{1}{1 + \exp(-\lambda(y - m))} \qquad (2)$$

We employ sigmoid function of intensity like (2) to fit the brightness. y are intensity values of Bright Channel Prior filtered and λ, m are parameters which are set only once.

Fig. 2. Illumination Shortage

Illumination shortage is visualized in Fig. 2. In the figure, dark regions are shown bright because the regions need more light and bright regions are represented in dark for the same reason.

Compensation of illumination is performed based on the retinex theory. The retinex theory adapts idea from the truth that human eyes recognize same color dissimilar under different illumination, such as shade, color of light etc. This can be represented as an equation (3).

$$I(x,y) = L(x,y) \times R(x,y) \qquad (3)$$

$I(x,y)$ is intensity of arbitrary point of image. $L(x,y)$, $R(x,y)$ are luminance and reflection, respectively. Luminance means brightness also reflection means object color without effect of light at the pixel.

If we take logarithm scale of equation (3), product is converted to summation. Since luminance compensation is represented as adding illumination shortage like equation (4).

$$\log I' = \log(L + L') + \log R \qquad (4)$$

Then, we use rule of logarithm to above equation.

$$\log I' = \log L + \log R + \log(1 + \frac{L'}{L}) \qquad (5)$$

The sum of logarithm of luminance and reflection is that of intensity.

$$\log I' = \log I + \log(1 + \frac{L'}{L}) \qquad (6)$$

The result is shown in Fig. 3

Fig. 3. The result of proposed algorithm

Dataset are night images downloaded from the internet. We performed proposed algorithm on the dataset and compared result with histogram equalization.

Conclusion

We proposed a novel algorithm. The algorithm comprised of illumination estimation and shortage compensation. It performs efficiently, but it has limitations. If there are black objects in an image, it is considered as dark region and color of the object will be deformed. It, however, shows good results if there are no black objects larger than patch size.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No.NRF-2011-0017202 and NRF-2010-0028680)

References

1. Yamasaki, Akito, et al. "Denighting: Enhancement of nighttime images for a surveillance camera." Pattern Recognition, 2008. ICPR 2008. 19th International Conference on. IEEE, 2008.
2. Rao, Yunbo, et al. "Illumination-based nighttime video contrast enhancement using genetic algorithm." Multimedia Tools and Applications (2012): 1-20.
3. Wong, Alexander, David A. Clausi, and Paul Fieguth. "Adaptive Monte Carlo Retinex method for illumination and reflectance separation and color image enhancement." Computer and Robot Vision, 2009. CRV'09. Canadian Conference on. IEEE, 2009.
4. Panagopoulos, Alexandros, et al. "Estimating shadows with the bright channel cue." Trends and Topics in Computer Vision. Springer Berlin Heidelberg, 2012. 1-12.
5. He, Kaiming, Jian Sun, and Xiaoou Tang. "Single image haze removal using dark channel prior." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009.
6. Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. Vol. 1. IEEE, 2001.
7. Schneiderman, Henry, and Takeo Kanade. "A statistical method for 3D object detection applied to faces and cars." Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on. Vol. 1. IEEE, 2000.
8. X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, “Face Recognition Using Laplacianfaces”, IEEE Trans-actions on Pattern Analysis and Machine Intelligence,vol. 27, no. 3, pp. 328-340, 2005.
9. Salvador, E., Cavallaro, A., Ebrahimi, T.: Cast shadow segmentation using invari-ant color features. Computer Vision and Image Understanding 95 (2004) 238{259
10. Fu, Huiyuan, Huadong Ma, and Shixin Wu. "Night Removal by Color Estimation and sparse representation." Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, 2012.

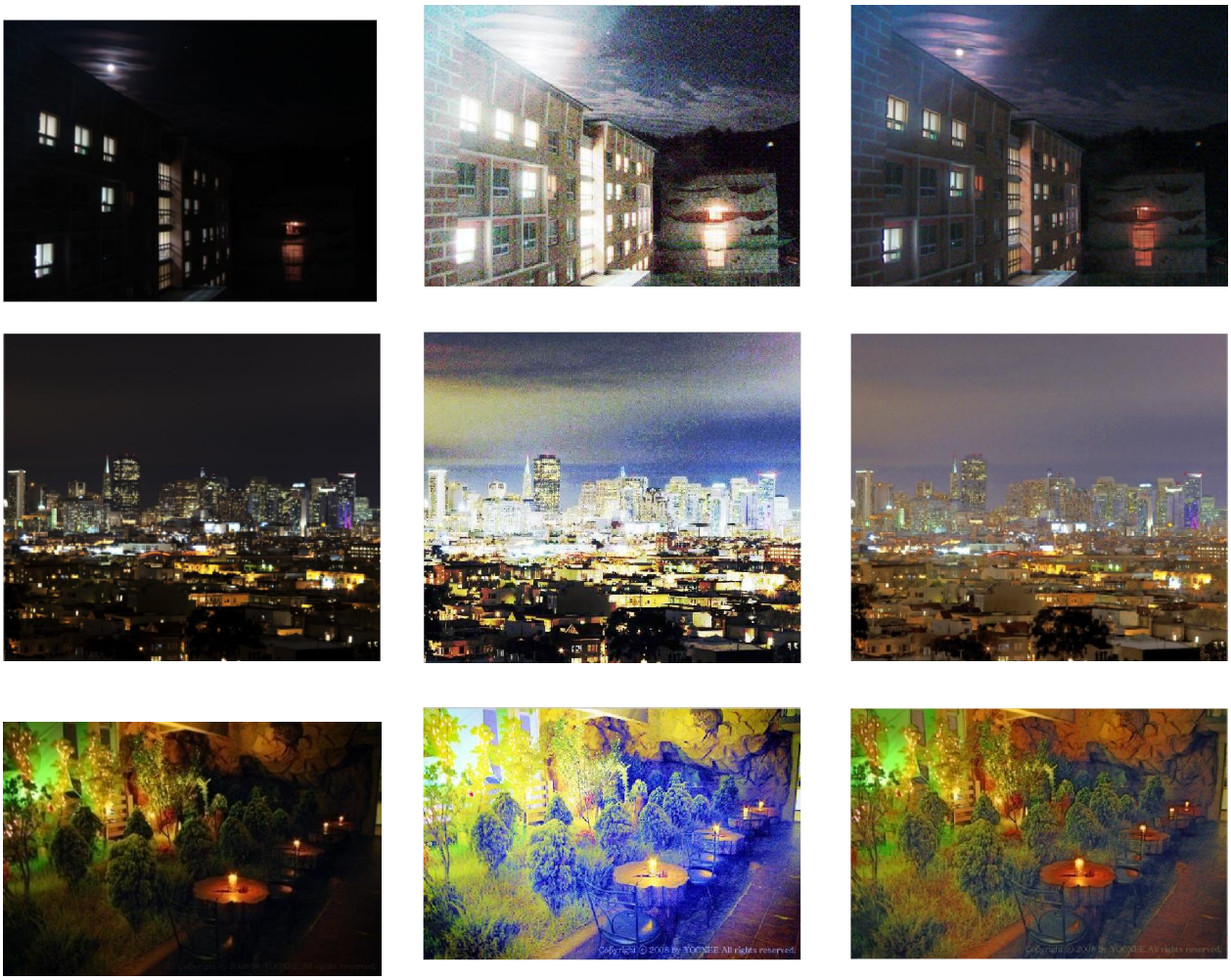


Fig. 4. Compare results of enhancement Original image(left), Histogram Equalization(middle), and proposed algorithm(right)

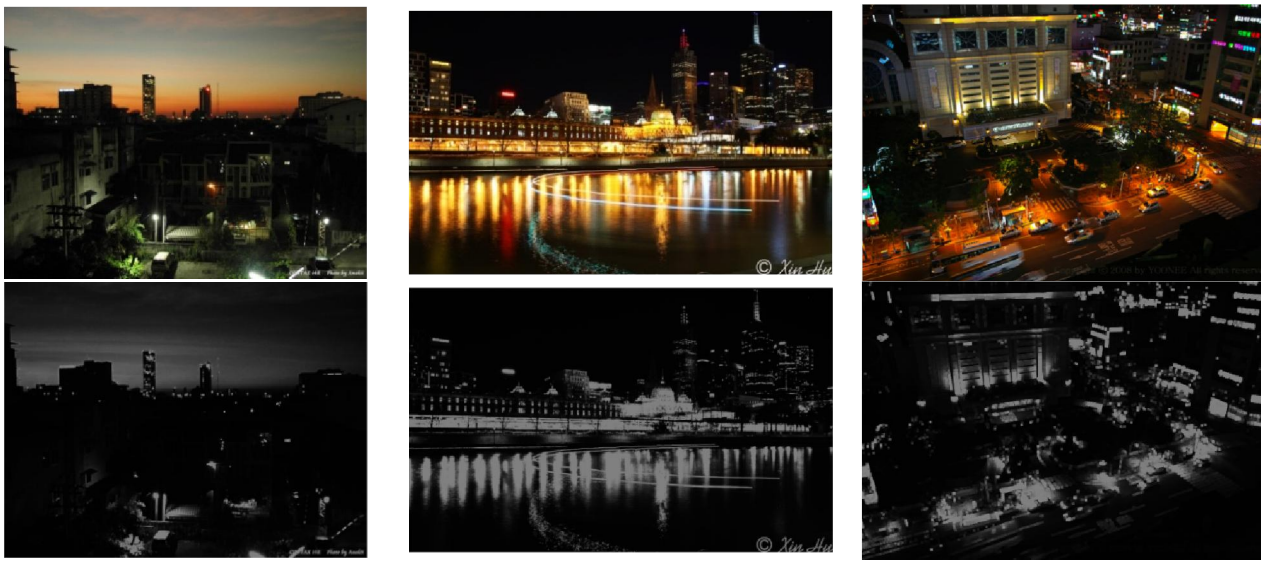


Fig 5. Estimated illumination(bottom) of images(top).

IMAGE SEGMENTATION IN MEDICAL IMAGING VIA GRAPH-CUTS[1]

M. Jirik[2,3], M. Zelezny[2,4]

[2] Department of Cybernetics at University of West Bohemia, Pilsen, Czech Republic
[3]mjirik@kky.zcu.cz, [4]zelezny@kky.zcu.cz

An organ segmentation is usually first step of liver treatment. We introduce an semi-automatic method for liver segmentation based on Graph-Cuts. Our experiments compare expert segmentation with our algorithm. We compare two different sets of parameters. Our software implementation is freely available

Introduction

Living donor transplantation and other modern methods of liver treatment are usually based on computed tomography (CT). Our work is motivated by two following clinical application. First is living-related liver transplantation (LRLT). It is the case when healthy voluntary donor gives a part of his liver to treated person. Second are oncologic resection. It is treatment for patient with liver cancer.
Manual extraction of individual anatomical information of liver and its vascular system is complex mental and very time consuming work. It is difficult to mentally construct 3D vessels anatomy from planar data of CT. Machine learning techniques provide wide range of methods which can facilitate human operator work. Computer assisted planning enables individual anatomy visualization and gives a support for operability decisions.
Our application is written in Python and software can be downloaded from http://github.com/mjirik/liver-surgery

Methods

There are some steps in our application which need to be done, so we can get satisfactory result. First step is data acquisition. Medical data are stored in DICOM format. Each slice is usually stored in single file. We used pydicom library with some improvements to read data. There are some limitations of this library. We have added ability to read DICOM overlay. It was important because this is way how we obtain an expert annotation of our data.

There are two groups of liver segmentation methods, semi-automatic and automatic. Automatic methods works without any sort of operator interactivity (for example [9]). Semi-automatic algorithms require some user intervention and the result is operator dependent (for example [12]). For clinical applications user control over the result is great advantage of these methods as long as lower error rate as shows [6]. Survey on liver segementation methods is presented in [8].
Segmentation used in this paper is based on Graph-Cut (GC). First use of of max-flow/mim-cut algorithms to minimize certain energy functions in computer vision problems is described in [11]. Segmentation problem is converted into graph issue. Boykov et al. showed in [3, 4, 1] max-flow/min-cut algorithm with some important improvements. For our application we have used implementation which is described in [7, 5].
By using Graph-Cuts, we minimize cost function E(A):

$$E(A) = \lambda R(A) + B(A) \qquad (1)$$

Here A is labeling. Each pixel A_p can represent 'object' or 'background'. Term $R(A)$ is related to region properties and $B(A)$ is related to boundary properties of image. The λ coefficient weights region term versus boundary term.

[1] The work has been supported by the grant of The University of West Bohemia, project number SGS-2013-032 and Experimental surgery – new trends in Biomechanics (registration number CZ1.07/2.2.00/15.0049)

$$R(A) = \sum_{p \in P} R_p(A_p) \qquad (2)$$

Region term is used for setting penalty to each pixel p which describes how similar is its intensity to the model of background or object. As show image 1 the graph is constructed based on input data and selected cost function. By its partition into two disjoint subsets image segmentation is performed.

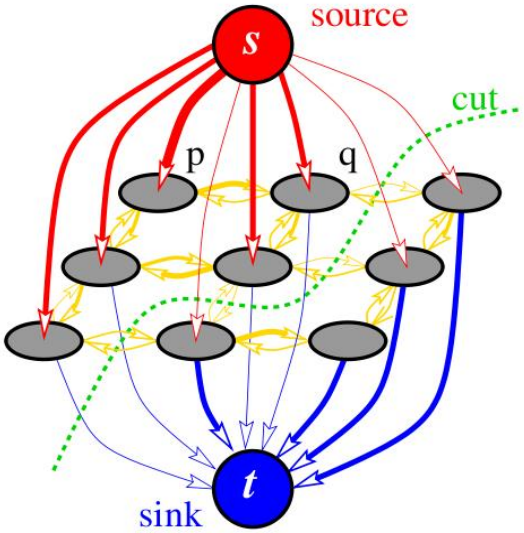


Fig 1: Graph-Cut [Boykov2004]

Main problem is edge weights setting. There are two types of edges in the graph. N-links (neighboor-links) are associatad with edeg properties and it connects (usualy) two neighboring nodes. T-links (terminal-links) are linked to region image properties and are connected to two terminals in graph.

Weights of T-links R_p to object and background vertex are given by model of object and background.

$$R_p(obj) = -\ln\left(Pr(I_p|O)\right) \qquad (1)$$
$$R_p(bkg) = -\ln\left(Pr(I_p|B)\right)$$

In our case likelihood $Pr(I_p|O)$ and $Pr(I_p|B)$ for object and background are given by gaussian mixture model with three components. It is based on image density (intensity) of data from user interaction. Image Model parameters are estimated by expectation maximization (EM) algorithm [10].

$$B(A) = \sum_{p \in P} B_{\{p,q\}} \cdot \delta(A_p, A_q) \qquad (3)$$

$$\delta(A_p, A_q) = \begin{cases} 1 & if A_p \neq A_q \\ 0 & otherwise \end{cases} \qquad (4)$$

Term $B(A)$ reflects boundary penalties of segmentation. Boykov et al. [2] suggests using function that penalizes discontinuities between pixels. We used constant penalty a for segmentation boundary. It means that objects with large surface area are more penalized then objects with same volume and smaller surface area.

Well known weakness of graph-cut algorithm is memory usage. It quickly increase with image size because of large number of edges in the constructed graph. For three dimensional data has every pixel 8 connections (6 N-links and 2 T-links) which makes this problem even more acute. We face it with two processes. First approach is setting of region of interest (ROI). Memory usage is much lower if when we work with the certain data subset.

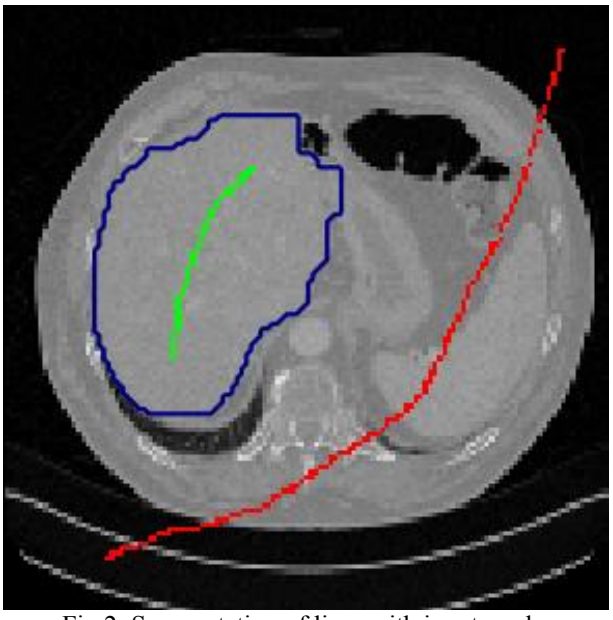


Fig 2: Segmentation of liver with input seeds

Second preprocessing step is data resampling. Data are resized to defined voxel size. It decreases computational complexity, and in addition to that, we get voxel with equal three dimensions. It allows easier setup of $B(A)$ energy term. N-links in all directions are set on same constant value.

In our experiments we used set of 8 patients. For each case we have contrast enhanced computer

tomography scans. Venous and arterial phase CT images were obtained. We have used venous data because of better contrast between liver and other tissues. Data are anotated by experts. They manually segmented the liver.

Results

Experiment in this part is based on comparison of expert manual segmentation and semiautomatic methods based on Graph-Cut segmentation. Fist step in our experiment is setting of region of interest. It is done with manual sekection. Then we used resampling to voxels with all dimensions equal to 2 mm.
The method is tested with two different setups. In first experiment si parameter a set to 30 while in second experiment is this value equal to 15.
All resulst are shown in table 1. First column is expert segmentation. Second column is Graph-Cut segmentation with $a = 30$. Third column is GC with $a = 15$. In last two columns is shown difference between manual segmentation and semi-automatic method. Last row of the table shows average values. For differences is constructed with absolute value.

Table 1

Data	Expert [ml]	GC a=30 [ml]	GC a=15 [m]	Diff a=30 [%]	Diff a=15 [%]
D1	597	723	694	-21,1	-16,2
D2	1151	1175	1073	-2,1	6,8
D3	1006	1122	1043	-11,5	-3,7
D4	1757	1641	1680	6,6	4,4
D5	1425	1314	1346	7,8	5,5
D6	1391	1331	1340	4,3	3,7
D7	3013	2758	2887	8,5	4,2
D8	1653	1730	1743	-4,7	-5,4
avg	1499	1474	1475	8.33	6.24

Discussion

As you can see from table 1 semi-automatic segmentation brings alternative to manual segmentation of liver. Manual segmentation take about 30 minutes from expert time. Semiautomatic method is performed in 10 minutes. Difference between both setting is almost one percent. Lower constant a bring better results but it is more time consuming.

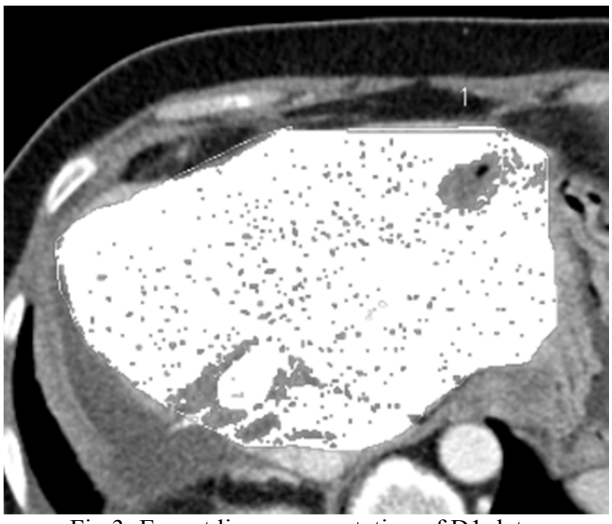


Fig 3: Expert liver segmentation of D1 data

In the experiment we compare expert with semiautomatic method. Results of expert are operator dependent. Image 3 show problematic liver segmentation of data D1. Machine based semi-automatic segmentation gives more consistent outputs.

Conclusion

Our work introduces possibilities of computer assisted diagnostic for liver treatment. Semi-automatic methods can save time of operator and brings consistent performance. Our algorithm can measure volume with error 6.24%. The experiments shows that careful parameter setup of methods can give us some improvement. Our algorithm is freely available.

References

1. Boykov, Y.; Jolly, M.-P.: Interactive Organ Segmentation Using Graph Cuts . 2000: s. 276-286.
2. Boykov, Y.; Kolmogorov, V.: An Experimental Comparison of Min-cut/Max-flow Algorithms for Energy Minimization in Vision. 2001: s. 359-374.
3. Boykov, Y.; Veksler, O.; Zabih, R.: Markov random felds with efcient appro-ximations. Computer Vision and Pattern Recognition (CVPR), 1998, 1998: s. 648-655
4. Boykov, Y. Y.; Jolly, M. P.: Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on, ro£ník 1, 2001: s. 105-112 vol.1

5. Andrew Delong, Anton Osokin, Hossam N. Isack, and Yuri Boykov. Fast Approximate Energy Minimization with Label Costs. International Journal of Computer Vision, 96(1):1-27, July 2011
6. Heimann et al. Comparison and evaluation of methods for liver segmentation from CT datasets. IEEE transactions on medical imaging, 28(8):1251-65, August 2009.
7. Vladimir Kolmogorov and Ramin Zabih. What Energy Functions Can Be Minimized via Graph Cuts? IEEE Trans. Pattern Anal. Mach. Intell., 26(2):147-159, 2004.
8. Ahmed M. Mharib, Abdul Rahman Ramli, Syamsiah Mashohor, and Rozi Binti Mahmood. Survey on liver CT image segmentation methods. Artificial Intelligence Review, 37(2):83-95, April 2011.
9. L Rusko and G Bekes. Fully automatic liver segmentation for contrast-enhanced CT images. MICCAI07: Segmentation in the Clinic, pages 143-150, 2007.
10. 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, B, vol. 39, no. 1, pp 1-38, 1977
11. D. Greig, B. Porteous and A. Seheult. Exact Maximum a posteriory estimation for binary images. Journal of the Royal Statistical Society, series B, , pp 271-279, 1989
12. D. Selle, B. Preim, A. Schenk and H. O. Peitgen. Analysis of vasculature for liver surrgical planning. Journal of Medical Imaging. Vol. 21, pp 1344-1357, 2002

IMAGE DENOISING VIA ENERGY ORIENTED SPARSE REPRESENTATION[1]

Jungyu Kang[2,3], Hoyong Jang[2,4], Chang D. Yoo[2,5]

[2]Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea
[3]cmiller2air@kaist.ac.kr, [4]hoyong.jang@kaist.ac.kr, [5]cdyoo@kaist.ac.kr

This paper considers a denoising of zero-mean white additive Gaussian noise from an image. Sparse representation technic has shown good performance in denoising noises while keeping the image from being blurred. To optimize the conventional sparse representation algorithm, a concept of energy function is combined. An energy function to specify the noise level, estimation technic to find the optimal level of representation, result of the experiments, and further possible improvements are introduced.

Introduction

This paper considers a traditional image denoising problem. Here the noise is considered to be additive noise. Here our goal is to find estimation of original image x from given noised image y. Although usually the noise n varies through conditions such as camera, environment, etc., here we only consider zero-mean white Gaussian with standard variance σ. The Gaussian distribution becomes a general model for various noises since the sum of random variable tend to a Gaussian distribution.

Since this is a very traditional problem and studied for a long time, various approaches were conducted. Although the results varies with the texture or complexity of the input image, algorithms such as BM3D [4], SKR[10], and K-LLD [3] showed best performances in denoising zero-mean white Gaussian noises. Yet, it is considered that it can still be improved. Among many competitive denoising approaches, our method focuses on using sparse representation[6].

More specifically, our method aim to improve existing denoising algorithm using sparse representation [6]. This paper aims to optimize the reconstruction rate of the sparse representation using a specifically designed energy function instead of simple sum of square error.

Fig 1 – Denoising result

The remainder of the paper is structured as follows: Section 2 briefly provides preliminary knowledge about dictionary learning, sparse representation algorithm, and the conventional sparse representation based denoising algorithm by Elad et al.. Section 3 introduces the proposed algorithm about designing the energy function and estimating the original image. Section 4 provides example results and evaluations of the proposed algorithm. Finally, section 5 concludes the paper with descriptions of future works.

Related Works

Dictionary learning is a data representing algorithm by solving a matrix factorization problem. Let us consider a set $Y = \{\mathbf{y^1}, \dots, y^n\} \in R^{m \times n}$ which is a set of m dimensional n signals. In our problem, the **Y** will be the input noised image and **y** will be

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No.NRF-2011-0017202 and NRF-2010-0028680)

the patches of the noised image. The goal of dictionary learning is to reconstruct the input signal **Y** as linear combination of two matrices, **D** and **A**. Here $\mathbf{D} \in R^{m \times k}$ denotes a m-dimensional trained dictionary containing k basis vectors. $\mathbf{A} \in R^{k \times n}$ denotes a decomposition coefficient for n signals. In detail, the dictionary learning aims to build an appropriate over-completed dictionary that can represent the input signals well with given constraints. This is achieved by optimizing the following equation.

$$\min_{D \in D, A \in A} \left|\left|Y - DA\right|\right|_2^2 \quad s.t \ \sum_{i=1}^{n} \Omega(\alpha_i) \quad (1)$$

Here, the constraint omega is a constraint to prevent overfitting and exploit certain property such as nonnegativity, sparsity, or structure to the coefficients. To optimize the equation, various optimization technics has been proposed [5, 7, 8]. In this paper, K-SVD algorithm [1] is used to learn **D** and **A** in a same time.

After learning the appropriate dictionary using the KSVD algorithm, we aim to reconstruct the input signals. However, since we want to enforce sparsity to our solution, l1 norm is used as the constraint term. Therefore the equation above can be reformulated as below:

$$\min_{A \in A} \sum_{i=1}^{n} \left|\left|y_i - D\alpha_i\right|\right|_2^2 + |\alpha_i| \quad (2)$$

Here, since the above equation is a convex function, the optimal solution can be obtained with convex optimization [2]. However, the convex optimization usually takes long time to get a solution, so we use OMP [9] instead.

The Orthogonal Matching Pursuit (OMP) algorithms is a greedy algorithm which seeks the best basis vector from the given basis vectors D to reconstruct the input signal in each iteration. It projects the whole basis vectors to the signal and finds the basis vector which has the maximum value. The corresponding coefficient value to the basis is obtained by simply performing a pseudo-inverse, which provides a same solution with the least square problem. The residual is updated by subtracting the projection of the chosen basis. These procedures are iteratively done until the result reaches the stopping rule.

The major advantages of this algorithm are that it works very fast, it is easy to implement, and the stopping rule can be designed as the user wants. Usually the number non-zero terms or reconstruction error are used as the stopping rule. However, in this paper, we introduce the energy function to control the sparsity and reconstruction error of the solution.

Proposed Algorithm

To start, we first introduce the conventional image denoising algorithm using sparse representation [6]. In the algorithm, a dictionary for the sparse representation is prepared from either natural images or the noised image itself using the K-SVD algorithm. Using these dictionaries, the algorithm defines an optimization problem below

$$arg\min_{X, \alpha_{ij}} \sum_{ij} \left|\left|R_{ij}X - D\alpha_{ij}\right|\right|_2^2 + \lambda \left|\left|Y - X\right|\right|_2^2 + \sum_{ij} \mu_{ij} \left|\left|\alpha_{ij}\right|\right|_0 \quad (3)$$

Here, $\mathbf{R}_{ij}$ is a matrix which indicates the ij-th patch of X. Therefore this first term controls the reconstruction error between the input noisy patch and the reconstructed patch. The second term ensures the relationship between the reconstructed image and the input image and the third term enforces sparsity to the solution. The parameter and μ_{ij} are weight coefficients that balance the sparsity and reconstruction rate. The goal of the parameter is to have the reconstruction error $\left|\left|Y - X\right|\right|$ below $C\sigma^2$. Where a is the standard variation of the noise and C is a constant.

The solution for the problem above is derived using OMP. Here $C\sigma^2$ is used as the parameter for the stopping rule of the OMP algorithm. Each target patches are reconstructed iteratively with the given dictionary **D** while the squared sum of the residual goes below the stopping parameter which is a function of the $C\sigma^2$. The patches derived from these routines are then aligned and overlapped parts are averaged linearly. Then the reconstructed image **X** will be the estimation of the original image **X** and this showed a great performances. Especially, the results using the dictionary from the noised image itself showed

the better performance than the dictionary from natural images.

From the algorithm explained above, a simple question arises. Will it really give us the optimal solution? How the solution will change if the OMP algorithm had more or less iterations? Of course, less iteration will give less reconstructed image which may results in large difference between the reconstructed image the original image. In the same way, more iteration will give better reconstruction of the input image, which may lead to reconstruction of the noise.

Then what would be the optimal number of iterations? From the heuristic experiment, we found that the conventional algorithm which sum of square of the residual as the parameter for stopping rule does not give us the optimal number of the iterations. Now, the goal of our algorithm will be defining the better measure and parameters for the OMP algorithm's stopping rule.

To achieve this, we define the energy function to determine the noise level of the image. Therefore the noised image will have higher energy value than the original image. Here we assume that the energy function is linear so the condition stated below holds:

$$E(\mathbf{X}+\mathbf{N}) = E(\mathbf{X}) + E(\mathbf{N}) \quad (4)$$

If the condition above holds, the E(**N**) can be easily obtained from the training data by simply subtracting the energy value of the original image from the noised image. In this paper, the absolute sums of the gradients in x and y directions are used as the energy function. By comparing the energy values of the patches from training original images and the noised images, we can get the difference between the two sets. Using this value we estimate the energy value of the original image from the noised image. Of course, the energy function stated above, absolute sum of gradients, does not exactly holds the linearity condition above, but the condition fairly holds. In this way, the energy value of the original image can be estimated from the energy value of the noised image.

In our method, to denoise a noised image, we find the average energy value difference caused by the noise. Using the randomly generated patches, we add the same noise to the patches and compare the average noise differences between original patches and noised patches. Using the average energy value difference between noised image and original image, we reformulate (4) as follows:

$$\underset{X,\alpha_{ij}}{arg\min} \sum_{ij} \left|\left|R_{ij}X - D\boldsymbol{\alpha}_{ij}\right|\right|_2^2 + \boldsymbol{\lambda}\left|\left|E(Y) - E(X)\right|\right|_2^2 + \sum_{ij} \boldsymbol{\mu}_{ij} \left|\left|\alpha_{ij}\right|\right|_0 \quad (6)$$

Here only the second term has been changed into the difference of the energy value from the squared sum of errors.

To solve the above equation, we again use the OMP. However, in the OMP algorithm, we use the difference of the energy values as the stopping rule instead of the sum of square error. In other words, we keep the iteration going until the energy function of the residual reaches to the average energy value difference. This will allow us to control the level of reconstruction so we avoid reconstruction the noise while estimating the original image.

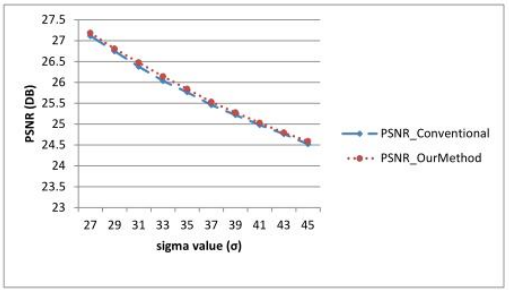


Figure 2. Comparison between the conventional method(blue circle, dashed line), and our method(red square, dotted line) using the global dictionary

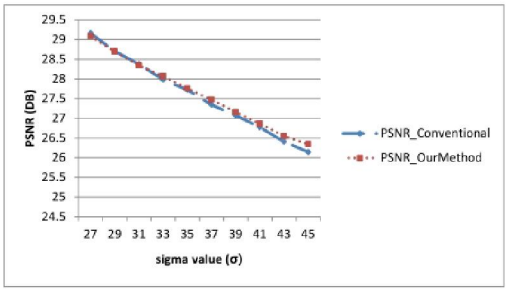


Figure 3. Comparison between the conventional method(blue circle, dashed line), and our method(red square, dotted line) using the adaptive dictionary

Experimental Result

In this section we evaluate and compare the results from our algorithm with the conventional algorithm. The tests have been conducted on the Barbara image and the Boat with various noise levels. The exact same test has been conducted with [6], with the same dictionaries and noises for the comparison. Here the dictionary used was the dictionary from [6] which the size is 64 256. We used 8 8 image patches, which is the same size with the conventional algorithm. Table 1 and 2 compares the denoising results of our algorithm and the conventional algorithm. As the result shows, our method showed better performance in noise levels area for results using the global dictionary. For the results using the adaptive dictionary, our method show competitive result and gets better as the noise level and the complexity of the image increases.

Table 1. Summary of the denoising PSNR result for Barbara image.

$\sigma/PSNR$	Global Dictionary		Adaptive Dictionary	
	Elad et al.	Our	Elad et al.	Our
27/19.5	27.11	**27.18**	**29.16**	29.08
29/18.9	26.74	**26.80**	**28.70**	28.69
31/18.3	26.37	**26.47**	28.31	**28.34**
33/17.7	26.03	**26.14**	27.98	**28.06**
35/17.2	25.76	**25.84**	27.72	**27.76**
37/16.8	25.45	**25.53**	27.34	**27.47**
39/16.3	25.21	**25.27**	27.07	**27.15**
41/15.9	24.98	**25.03**	26.77	**26.86**
43/15.4	24.76	**24.79**	26.40	**26.55**
45/15.0	24.52	**27.52**	26.14	**26.34**

Table 1. Summary of the denoising PSNR result for Barbara image. Left column is the result using Elad et al. [6] and the right column is the result using our algorithm for various noise levels.

Table 2. Summary of the denoising PSNR result for Boat image.

$\sigma/PSNR$	Boat	
	Elad et al.	Our
27/19.5	29.24	**29.41**
29/18.9	**29.07**	28.95
31/18.3	28.53	**28.70**
33/17.7	**28.35**	28.29
35/17.2	**28.06**	27.98
37/16.8	27.78	**27.87**
39/16.3	27.41	**27.63**
41/15.9	27.17	**27.23**
43/15.4	26.90	**27.00**
45/15.0	26.74	**26.78**

Table 2. Summary of the denoising PSNR result for Boat image. Left column is the result using Elad et al. [6] and the right column is the result using our algorithm for various noise levels.

Conclusion

This paper has presented the algorithm developed from the conventional denoising method using sparse representation. It used the absolute of gradients as the energy function and used it as the stopping rule for the OMP algorithm. It showed slight but clearly better performances for the denoising using the global dictionary and competitive performances for the adaptive dictionary.

Yet, this algorithm has high potential to be improved by simply defining the better energy function for modeling the input noise. Also estimation could be developed by doing the regression about the energy values from the training data. This will clearly give us better estimation about the original image's energy value. Finally, this regression will allow us to model more complex noises such as noises from multimodal Gaussians.

References

1. M. Aharon, M. Elad, and A. Bruckstein. K-svd: Design of dictionaries for sparse representation. Proceedings of SPARS, 5:9–12, 2005..
2. S. Boyd and L. Vandenberghe. Convex optimization. Cambridge university press, 2004.
3. P. Chatterjee and P. Milanfar. Clustering-based denoising with locally learned dictionaries. Image Processing, IEEE Transactions on, 18(7):1438–1451, 2009.
4. K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. Image Processing, IEEE Transactions on, 16(8):2080–2095, 2007.
5. B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. The Annals of statistics, 32(2):407–499, 2004.
6. M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. Image Processing, IEEE Transactions on, 15(12):3736–3745, 2006.
7. J. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent. Journal of statistical software, 33(1):1, 2010.
8. H. Lee, A. Battle, R. Raina, and A. Y. Ng. Efficient sparse coding algorithms. Advances in neural information processing systems, 19:801, 2007.
9. Y. C. Pati, R. Rezaiifar, and P. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on, pages 40–44. IEEE, 1993.
10. H. Takeda, S. Farsiu, and P. Milanfar. Kernel regression for image processing and reconstruction. Image Processing, IEEE Transaction on, 16(2):349-366, 2007

REFLECTIVE SYMMETRIZATION OF FEATURE POINTS IN IMAGES [1]

A. Karkishchenko[2], V. Mnukhin[3]

[2] North Caucasian Federal University,
2 Kulakov Avenue, 355029 Stavropol, Russia; karkishalex@gmail.com
[3] Southern Federal University, 44 Nekrasovskiy Lane, 347928 Taganrog, Russia;
mnukhin.valeriy@mail.ru

In this work, we consider the symmetrization problem, that is the problem to obtain more accurate information about location of points based on a priori knowledge of their symmetries. Methods to solve the symmetrization problem with respect to vertical and inclined axes of reflectional symmetry are considered jointly with the more general symmetrization with respect to an indefinite reflection axis. The methods produce the minimal deformation that enhances approximate symmetries present in a given arrangement of points.

Introduction

Symmetry is a central concept in many natural and man-made objects and plays a crucial role in visual perception, design and engineering. Several recent efforts in shape analysis have focused on detecting symmetries in 2D and 3D shapes [1–3]. Numerous applications have successfully utilized this type of information, e.g., for model reduction [3], scan completion [4], segmentation [5], shape matching [1], etc. In many cases low-level symmetry analysis is based on investigations of so-called *feature points*, whose exact meaning depends on the resolving problem.

One of the most common problems, where methods of "refinement by symmetry" can be efficiently used, is the biometrical identification, when the correct location of feature points is crucial. In particular, the accuracy of human face detection and recognition strongly depends on the measurement precision of pupils of eyes location [6]. Then the fact of near, but imperfect, reflective symmetry of human full-faces can be used to improve the accuracy [7]. In fact, usually methods of human face detection are based on the position analysis of several dozens feature points, which are either coupled in pairs, symmetric with respect to a vertical axis, or situated in the axis.

In this work, we present several methods to obtain more accurate information about location of feature points, based on *a priori* knowledge of their symmetry. Note that positions of points in images are always known with some drift that depends on such factors as the image quality, noise levels in a vicinity of the points, the processing algorithm, and so on. As a result, evaluated coordinates could fail the symmetry conditions even for those points, which are in fact symmetric. So it is reasonable to use the information about symmetry to specify positions of feature points. Besides, the symmetrization itself should be done with minimal deformation of points positions that enhances approximate symmetries present in a given arrangement of points.

Symmetrization with respect to a vertical axis

Let $P=\{p_1,...,p_n\}$ be the set of all feature points given by their coordinates $p_k=(x_k,y_k)$. Assume that the reflection symmetry axis coincides with the coordinate axis Oy and that the feature points are ordered in such a way that the points $P_R=\{p_1,...,p_m\}$ are in the right half-plane, the corresponding points $P_L=\{p_{m+1},...,p_{2m}\}$ are in the left half-plane,

[1] The work was supported by RFBR grants 11-07-00591 and 13-07-00327.

and the rest of points $P_O = \{p_{2m+1}, ..., p_n\}$ are situated in the axis Oy, as Fig. 1 shows.

We associate with the ordered set P the following $2n$-dimensional vector:

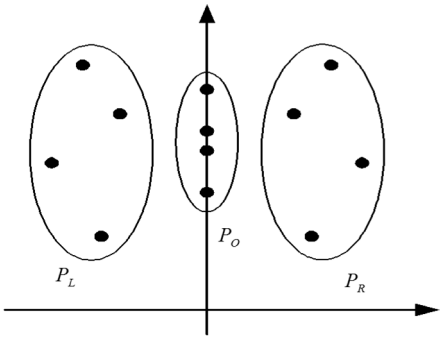


Fig. 1. Symmetry with respect to the vertical axis Oy.

$$(p_1, p_2, ..., p_n) \leftrightarrow (x_1, ..., x_n, y_1, ..., y_n)^T = X.$$

Evidently, all such vectors form the vector space R^{2n}. Taking into account the partition into the classes P_R, P_L, P_O, note that if the positions of all the feature points are precisely known, the following conditions must be satisfied:

$$\begin{aligned} x_i &= -x_{m+i}, \quad i = 1, ..., m; \\ y_i &= y_{m+i}, \quad i = 1, ..., m; \\ x_i &= 0, \quad i = 2m+1, ..., n. \end{aligned}$$

It is easy to check that the set of all vectors in R^{2n} under these conditions form an n-dimensional subspace $R_{Sym} \subset R^{2n}$. Note that the "symmetrization" of a given arrangement P of feature points means finding the vector $X_s \in R_{Sym}$ that is the best approximation of X by the Euclidean norm:

$$X_s = \arg \min_{Z \in R_{Sym}} \|Z - X\|.$$

Thus, X_s is the orthogonal projection of X onto the subspace R_{Sym}.

To find the projection X_s, let Q be the projection matrix onto the subspace R_{Sym}, and let A be a $(2n \times n)$-matrix with R_{Sym} as its column space, so that columns of A form a basis for the subspace R_{Sym}. Then, since A is a matrix with linearly independent columns, we know [8, p. 164], that

$$Q = AA^+ = A\left(A^T A\right)^{-1} A^T,$$

where A^+ is the pseudoinverse of A, and so

$$X_s = QX.$$

To complete the solution, note that

$$A^T = \begin{pmatrix} I_m & -I_m & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_m & I_m & 0 \\ 0 & 0 & 0 & 0 & 0 & I_{n-2m} \end{pmatrix},$$

and so

$$Q = \frac{1}{2} I_{2n} + \frac{1}{2} \begin{pmatrix} -S & 0 \\ 0 & S \end{pmatrix},$$

where S is the following $n \times n$-matrix:

$$S = \begin{pmatrix} 0 & I_m & 0 \\ I_m & 0 & 0 \\ 0 & 0 & I_{n-2m} \end{pmatrix}.$$

Hence, for the vertical reflextion axis the symmetrization can be easily performed by producing simple operations over coordinates of feature points. Namely, to symmetrize a pair of points $p_i \in P_R$ and $p_{m+i} \in P_L$, we need to take a pair of points with coordinates

$$\left(\pm \frac{|x_i| + |x_{m+i}|}{2}, \ \frac{y_i + y_{m+i}}{2} \right).$$

To symmetrize a point from P_O, we need to zeroize its *x*-coordinate, and leave its *y*-coordinate without changes.

Symmetrization with respect to an arbitrary axis

Now, let the reflection axis be defined by the equation $y = ax + b$, (where $a \neq 0$), as it is shown in Fig. 2. Then the symmetrization problem can be solved by switching to a new coordinate system $(O'x'y')$. Namely, assume that the y'-axis of the new system coincides with the reflection axis, the origin O' is in the point, where y'-axis crosses Oy, and $O'x'$ is

orthogonal with $O'y'$ in such a way that $(O'x'y')$ is a "right-handed" system.

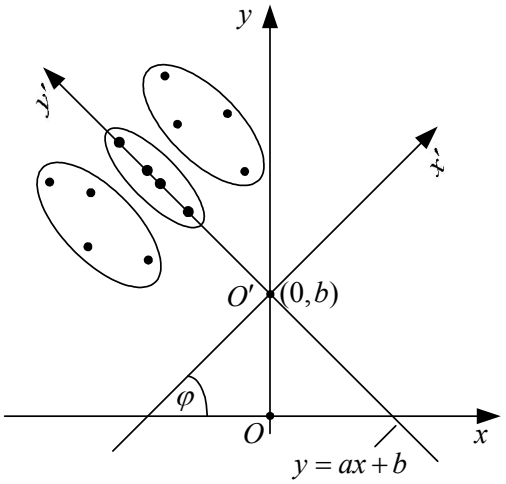


Fig. 2. Symmetrization with respect to an arbitrary axis.

Let (x,y) be coordinates of an arbitrary point in the "old" (Oxy)-system, and let (x',y') be coordinates of the same point in the "new" $(O'x'y')$-system. Then

$$\begin{aligned} x' &= x\cos\varphi + (y-b)\sin\varphi \\ y' &= -x\sin\varphi + (y-b)\cos\varphi \end{aligned},$$

where the values of $\cos\varphi$ and $\sin\varphi$ follow from the condition $\cot\varphi = -a \neq 0$, so that

$$\sin\varphi = \frac{1}{\sqrt{a^2+1}}, \qquad \cos\varphi = -\frac{a}{\sqrt{a^2+1}}.$$

Just as above, we consider the vector

$$X = \left(x_1,...,x_n,y_1,...,y_n\right)^T \in R^{2n},$$

formed by coordinates of the feature points with respect to the "old" (Oxy)-system, and the vector X' of coordinates of the same points with respect to $(O'x'y')$. These vectors are related by

$$X' = R\left(X - bK\right),$$

where $K = (0,...,0,1,...,1)^T$, and

$$R^T = \begin{pmatrix} \cos\varphi \cdot I_n & \sin\varphi \cdot I_n \\ -\sin\varphi \cdot I_n & \cos\varphi \cdot I_n \end{pmatrix}.$$

It converts the symmetrization with respect to an arbitrary axis into the previously solved problem. Indeed, we may symmetrize X' in the same way, as it has been done in the previous section, and then return the symmetrized vector X'_s back into the (Oxy)-system:

$$X_s = RX'_s + bK = R^T\left(QR^T\left(X - bK\right)\right) + bK,$$

or

$$X_s = GX + bHK,$$

where

$$G = RQR^T \quad \text{and} \quad H = I - RQR^T.$$

Now, it as an easy matter to work out that

$$G = \frac{1}{2}I_{2n} + \frac{1}{2}\begin{pmatrix} -\cos 2\varphi \cdot S & \sin 2\varphi \cdot S \\ \sin 2\varphi \cdot S & \cos 2\varphi \cdot S \end{pmatrix};$$

the expression for H is similar.

Symmetrization with respect to an unknown axis

Assume now, that the parameters a and b of a reflection symmetry axis are unknown, though the existence of such axis follows from the nature of the problem under solution. In this section we use the previously developed methods to determine the parameters in such a way that symmetrization could be achieved by the minimal deformation. For this we need the following notations:

$$\sigma = (\underbrace{1\ ...\ 1}_{n}\underbrace{1\ ...\ 1}_{n})^T,\ \sigma_0^1 = (\underbrace{1\ ...\ 1}_{n}\underbrace{0\ ...\ 0}_{n})^T,$$
$$\sigma_1^0 = (\underbrace{0\ ...\ 0}_{n}\underbrace{1\ ...\ 1}_{n})^T.$$

So let the reflection axis be defined by $y = ax + b$, where $a = -\cot\varphi \neq 0$ and b are undefined parameters. The corresponding optimization problem is

$$\left\| AY - R^T(\varphi)\left(X - b\sigma_1^0\right)\right\|^2 \xrightarrow{Y,\varphi,b} \min,$$

where $AY = Z$ is the resulting vector of coordinates of symmetrized points. In other words, to achieve the optimal symmetrization we need to find the parameters φ, b, and the minimizing vector Y.

Let $F\left(Y,\varphi,b\right)$ be the expression under the minimization. Writing it in a matrix form and using the minimization criteria, we would get the following equations:

$$\frac{\partial F(Y,\varphi,b)}{\partial Y}=0,$$

$$\frac{\partial F(Y,\varphi,b)}{\partial \varphi}=0,$$

$$\frac{\partial F(Y,\varphi,b)}{\partial b}=0.$$

A series of appropriate calculations and transformations produce the system

$$\begin{cases} AY = QR^T(\varphi)\left(X - b\sigma_1^0\right), \\ \left(X - b\sigma_1^0\right)^T \dfrac{dR(\varphi)}{d\varphi} AY = 0, \\ \left(\sigma_1^0\right)^T R(\varphi) AY = \left(\sigma_1^0\right)^T \left(X - b\sigma_1^0\right). \end{cases}$$

The first (matrix) equation gives us the symmetrized vector in relation with φ and b. The last two scalar equations can be used to find the parameters φ and b of the reflection axis.

To write down the final solution, assume that the vector X is partitioned into the "centralized" blocks

$$X = \left(\dot{x}_1^T \dot{x}_2^T \dot{x}_3^T \dot{y}_1^T \dot{y}_2^T \dot{y}_3^T\right),$$

where $\dot{x}_i = x_i - x_{av}e$, $\dot{y}_i = y_i - y_{av}e$, $i = 1,2,3$, e is the vector of 1's of an appropriate dimension, and

$$x_{av} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad y_{av} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

are averages of x- and y-coordinates of the evaluated feature points. Then, after some tedious transformations of the last two equations of the system, the next result follows:

$$\operatorname{tg} 2\varphi = \frac{(\dot{x}_1,\dot{y}_2)+(\dot{x}_2,\dot{y}_1)+(\dot{x}_3,\dot{y}_3)}{(\dot{x}_1,\dot{x}_2)-(\dot{y}_1,\dot{y}_2)+\frac{1}{2}(\dot{x}_3,\dot{x}_3)-\frac{1}{2}(\dot{y}_3,\dot{y}_3)},$$

$$b = y_{av} + x_{av}\cot\varphi.$$

Substitute the evaluated parameters φ and b into the first equation of the system, we can find the symmetrization $AY = Z$ of the original vector X. As the last step of the solution, the symmetrized vector should be transformed into the original "old" coordinate system.

Conclusion

We have considered the problem of feature points symmetrization with respect to reflectional symmetry. It has many different applications, since reflection is the most simple and common kind of 2D symmetry. At the same time, rotational, dihedral and translational symmetries are also quite common for 2D image processing problems, and the skew affine symmetries are under the special interest because of real-world applications. Considera-tion of 3D symmetries seems to be even more challenging. Rigorous mathematical solutions of such symmetrization problems could be quite involved but would lead to improved and extended methods of image processing.

References

1. J. Podolak, P. Shilane, J. Giesen, M. Gross, L. Guibas. Example-based 3D scan completion // Symposium on Geometry Processing. – 2005.– P. 23-32.
2. A. Martinet, C. Soler, N. Holzschuch, F. Sillion. Accurate detection of symmetries in 3D shapes // ACM Trans. Graph. – 2006. – Vol. 25, No. 2. – P. 439-464.
3. N.J. Mitra, L.J. Guibas, Pauly. Partial and approximate symmetry detection for 3D geometry // ACM Trans. Graph. – 2006. – Vol. 25, No. 3. – P. 560-568.
4. S. Thrun, B. Wegbreit. Shape from symmetry // Int. Conference on Computer Vision. – 2005.
5. P. Simari, E. Kalogerakis, K. Singh. Folding meshes: Hierarchical mesh segmentation based on planar symmetry // Proc. Symposium on Geometry Processing. – 2006.
6. A.N. Karkishchenko, I.N. Grechukhin. Statistical face recognition based on the geometry of feature points (in Russian). // Large-Scale Systems Control. – 2012. – Vol. 38. – Moscow. – P. 78-90.
7. A.N. Karkishchenko, I.N. Grechukhin. Localization of feature points based on the natural symmetries of images (in Russian). // Proc. of the Conference "Intellectual Control Systems for Rail Transport".– 2012.–Moscow.–P.262-265.
8. G. Strang. Linear Algebra and Its Applications. – Thomson Brooks/Cole. – 2006. – 487 p.

IMAGE SEGMENTATION BY OPTIMAL AND HIERARCHICAL PIECEWISE CONSTANT APPROXIMATIONS

M. Kharinov[1]

[1] **St. Petersburg Institute for Informatics and Automation of RAS,**
14–th line, 39, 199178, St. Petersburg, Russia
khar@iias.spb.su

Piecewise constant image approximations of sequential number of segments or clusters of disconnected pixels are treated. The method of majorizing of optimal approximation sequence by hierarchical sequence of image approximations is proposed. A generalization for multidimensional case of color and multispectral images is foreseen.

Introduction

In terms of deductive Descriptive approach [3, 7] the preliminary stage of image transformation from the original form into a recognizable form is considered in the report. For this purpose, the piecewise constant image approximations are studied. According to the formal quality assessment the optimal approximations are the best, since they minimally differs from the image in the values of total squared error E or standard deviation σ of the image pixels from the averaged pixels of approximation. On the other hand, hierarchical approximations, compared with nonhierarchical ones, are far preferable for image recognition tasks. Whether these requirements are compatible - that is the question to be discussed.

Analytical description

Let's consider the general case of multi-dimensional pixel clustering for color or multispectral images. Let I_1 and I_2 be two central data points of averages intensities for clusters 1 and 2, respectively. Let n_1 be the number of pixels in the cluster 1 and n_2 be the number of pixels in the cluster 2. Then the increment ΔE_{merge} of the total squared error E caused by the merging of specified clusters and the reduction of the number of clusters per unit is given by the following formula:

$$\Delta E_{merge} = \frac{\|I_1 - I_2\|^2}{\frac{1}{n_1} + \frac{1}{n_2}} \geq 0 \ , \qquad (1)$$

where the symbol $\| \ \|$ denotes an Euclidean distance.

Just this quantity is minimized in the version [1] of Mumford-Shah model [10, 11]. In the version [6], the formula differs by an additive term, and in FLSA version [2] by a multiplicative factor to take into account the total length of the boundaries between the segments (clusters of connected pixels).

Let's write down the formula for splitting of the cluster 1, when its $k < n_1$ pixels with the central data point I, initiate a new cluster. In this case, cluster 1 is split into two clusters of k and complementary $n_1 - k$ pixels, and cluster splitting is accompanied by increase of the cluster number per unit along with a non-positive increment ΔE_{split} of the total squared error:

$$\Delta E_{split} = -\frac{\|I - I_1\|^2}{\frac{1}{k} - \frac{1}{n_1}} \leq 0 \, . \qquad (2)$$

The composition of splitting and merging of clusters induces a correction operation without changing the number of clusters, which is accompanied by an increment $\Delta E_{correct}$ of the total squared error:

$$\Delta E_{correct} = \frac{\|I - I_2\|^2}{\frac{1}{k} + \frac{1}{n_2}} - \frac{\|I - I_1\|^2}{\frac{1}{k} - \frac{1}{n_1}}, \qquad (3)$$

where the negative term in (3) describes the increment of the total squared error E, caused by converting of k pixels from cluster 1 into a separate cluster, and the first term in (3) describes the increment of E caused by merging of the initiated cluster with the cluster 2, in accordance with (1) and (2).

Noteworthy that by simplifying of the formula (3) *K*-means method [8, 9, 14] is derived. Applying (3) precisely, we have proposed for the clustering of pixel sets a more accurate method [5], which in one-dimensional case provides a calculation of a complete sequence of optimal image approximations that are treated in multi-threshold Otsu method [12, 13]. Then, using (1) and (2) we have developed top-down/bottom-up segmentation algorithm $F_u v$ that for given image u and any masking image v produces a hierarchical sequence $\{v_i\}$ of approximations $v_1, v_2, ..., v_g$. The sequence of approximations $F_u v = \{v_i\}$ depending on successive cluster number i is described by a *convex* monotone sequence of corresponding values $E_1 \geq E_2 \geq , ..., \geq E_g = 0$ of total squared error E:

$$E_i \leq \frac{E_{i-1} + E_{i+1}}{2}, \quad i = 2, 3, ..., g-1, \qquad (4)$$

where E_1 corresponds to a trivial image approximation v_1, containing a single cluster, and $E_g = 0$ describes the last approximation $v_g = u$, containing the clusters of equal pixels.

In the case of $v = u$ the algorithm $F_u u$ produces a sequence $\{u_i\}$ of approximations $u_1, u_2, ..., u_g$, which is reproduced when replacing the masking image by any of specified approximations:

$$F_u u_1 = F_u u_2 = ... = F_u u_g \equiv F_u u. \qquad (5)$$

The property (5) is characteristic for the sequence of approximations $u_1, u_2, ..., u_g$, since in a different choice of image-mask, particularly, in the choice of an optimal approximation as masking image, this property is lost.

Thus, in one-dimensional case of gray-scale image, a sequence of optimal approximations is majorized by the self-consistent sequence of hierarchical approximations, which eventually should be generalized to the multidimensional case of color and multispectral images.

Experimental results

Optimal and hierarchical approximations for standard Lenna image in 2, 4, 8 and 16 gradations are illustrated top-down in Fig. 1.

Fig. 1. Optimal and hierarchical image approximations.

Fig. 1 shows the optimal approximations on the left. The hierarchical approximations are shown on the right. Two bottom

approximations visually coincide with each other and with the original image.

The sequential values E_i of the total squared error for approximations containing up to 10 clusters of pixels are given in the table 1 and graphically illustrated in Fig. 2.

Table 1. Total squared error

g	Optimal	Hierarchical
1	204664605,4	204664605,4
2	61548497,96	70364664,51
3	29502852,42	31708474,68
4	14675887,34	15629646,62
5	8967579,334	10149880,78
6	6605810,961	7647674,675
7	4691315,544	5568946,159
8	3697423,421	4054979,449
9	3042513,759	3447078,552
10	2473873,467	2853095,235

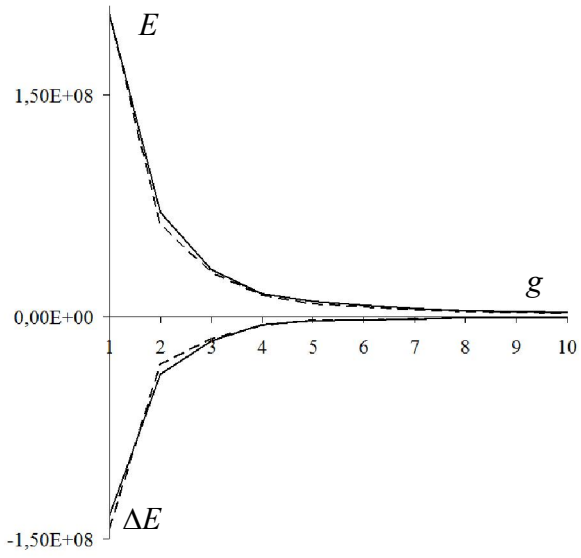


Fig. 2. Convex dependencies of E on cluster number g.

Fig. 2 shows the behavior of the total squared error for the optimal (dashed line) and hierarchical (solid line) approximations of the image depending on the number of clusters. The upper graph shows a monotonic decrease of the total squared error E_i itself and the lower graph shows the monotonic increase of its increment: $\Delta E_i = E_{i+1} - E_i, i = 1, 2, ..., g-1$. So, both dependencies are convex.

The overall results on segmentation are briefly presented graphically in Fig. 3, which demonstrates the dependencies of the standard deviation σ on the cluster number in the range from 1 to 1000 (σ is related with the total squared error by the formula: $E = N\sigma^2$).

In Fig. 2 the central gray curve separates the upper two curves describing partitioning of the image into connected segments from the lower curves that describe the image segmentation by clustering of disconnected pixels.

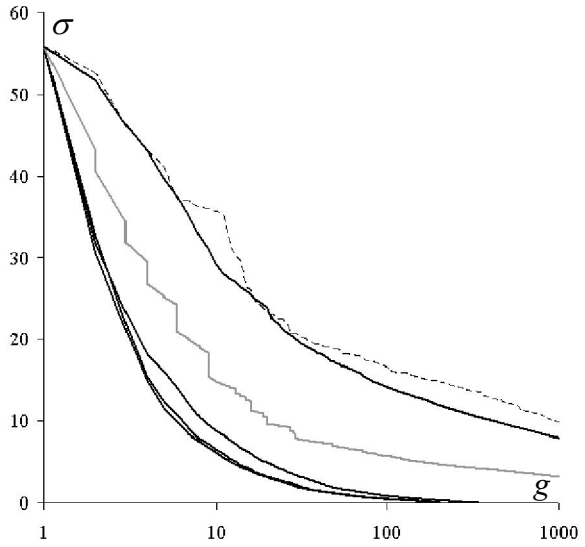


Fig. 3. Dependencies of σ on cluster number g (logarithmic scale along horizontal axes).

Upper dashed curve describes image segmentation according FLSA version [2] of Mumford-Shah model. The next curve just bellow dashed one describes segmentation according [1].

The gray curve demonstrates the calculations in our extended version of Mumford-Shah model, wherein the number of operations with the pixel sets is tripled according to (1)-(3). The algorithm is characterized as an algorithm of region growing, wherein just only the minimization of a total squared error determines how segment appears, how it is subdivided into the parts and how one segment captures the part of another with possible splitting of segment into disconnected pieces. The smooth curve sections correspond to hierarchical subsequences of image approximations and stepwise drops describe the conversion of one hierarchy into another.

Two of the lowest coalescing curves in Fig.3 describe the same optimal and majorizing approximations that illustrated by Fig. 2. The next overlying curve corresponds to a sequence of approximations generated by the algorithm $F_u v$ for the mask v obtained by resizing of the image u in sixteen times for

each dimension. At first, the algorithm $F_u v$, starting from given approximation v generates a complete sequence of approximations by top-down and bottom-up segmentation, and then edits them, converting into resulting approximation sequence $\{v_i\}$ that corresponds to a convex sequence $\{E_i\}$ (4). To apply the algorithm in multidimensional case of color and multispectral images, additional masking images should be introduced.

The image partitions for an increasing number of clusters do not depend from the linear transformation of intensities, and the algorithms of generating of image approximations commute with an image enlargement by doubling of pixels.

Conclusion

Thus, in the report we perform a comparative analysis of segmentation techniques based on advanced *K*-means method, a complete multi-thresholding method according N. Otsu and also extended Mumford-Shah model with the triple number of operations with clusters or segments [4, 5].

We draw attention that a sequence of optimal approximations of an image in general case is not hierarchical. However, it turns out that a sequence of optimal approximations is majorized by a sequence of hierarchical approximations. Furthermore, in both cases the distances between the image and its approximations depending on a number of clusters are described by a convex sequence in values of total squared error that fall down with increasing number of clusters. It is equally important that our MOAS method of majorizing of optimal approximation sequence is accessible for generalization to the case of multidimensional case of colors and multispectral images.

In contrast to the nonhierarchical approximations, the hierarchical approximations provide fast generation, storage in a fixed lesser volume of RAM, as well as effective image processing, skipping the repetitions of clusters or segments. We therefore look forward to the implementation of our approach in image processing domain, especially in the practice of automatic object detection.

References

1. A.S. Bugaev, A.V. Khelvas Pilot studies and development of methods and tools for analysis and automatic recognition of media flows in global information systems // Scientific and Technical Report, Moscow: Moscow Institute of Physics and Technology, - 2001, -Vol. 1, - 140 p. (in Russian).
2. D.J. Crisp, T.C. Tao Fast Region Merging Algorithms for Image Segmentation // The 5th Asian Conf. on Computer Vision (ACCV2002), Melbourne, Australia, - 2002. - P 1–6.
3. I.B. Gurevich and V.V. Yashina Descriptive Approach to Image Analysis: Image formalization space // Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications, - 2012. -Vol. 22, No.4, - P. 495–518.
4. M. V. Kharinov Stable segmentation of digital image // Computer Graphics and Vision (Graphi-con'2012) / Proceedings of the 22th International Conference on Computer Graphics and Vision, - Moscow: MSU, 01–05 October 2012, - P. 208-213. (see also arXiv:208.2655, in Russian).
5. M.V. Kharinov Reclassification formula that provides to surpass K–means method // arXiv preprint, arXiv: 1209.6204, 2012. - 6 p.
6. G. Koepfler, C. Lopez, J.M. Morel A Multiscale Algorithm for Image Segmentation by Variational Method // SIAM Journal on Numerical Analysis. - 1994. - Vol.31, No.1, - P.282–299.
7. I.V. Koryabkina Effective ways and means of image descriptions in pattern recognition, PhD thesis, Institution of Russian Academy of Sciences Dorodnicyn Computing Centre of RAS, Moscow, - 2006. - 20 p. (in Russian).
8. S.P. Lloyd Least squares quantization in PCM // IEEE Transactions on Information Theory - 1982. No 28 (2), - P. 129-137.
9. J.B. MacQueen Some Methods for classification and Analysis of Multivariate Observations // Proc. Fifth Berkeley Symp. Math. Stat. and Probab., Berkeley: University of California Press, -1967. Vol. 1, - P. 281–297.
10. D. Mumford, J. Shah Boundary detection by minimizing functionals // Proc. IEEE Comput. Vision Patt. Recogn. Conf. - San Francisco, -1985. - P. 22-26
11. D. Mumford, J. Shah Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems // Communications on Pure and Applied Mathematics, -1989. Vol. XLII, No. 4, - P. 577–685.
12. N. Otsu A Threshold Selection Method from Gray–Level Histograms // IEEE Transactions on systems, MAN, and CYBERNETICS, Vol. SMC–9, No. 1, - 1979. - P. 62-66.
13. Ping-Sung Liao, Tse-Sheng Chen, Pau-Choo Chung A Fast Algorithm for Multilevel Thresholding // J. Inf. Sci. Eng. - 2001. Vol. 17 (5), - P. 713-727.
14. H. Steinhaus Sur la division des corps materiels en parties // Bull. Acad. Polon. Sci. - 1956. C1. III Vol. IV, - P. 801-804.

DISCRIMINATION OF SIMILAR OBJECTS USING PARTIAL PHASE INFORMAION[1]

V. Kober[2,3], V. Karnaukhov[2,4], O. Milukova[2,5]

[2] Institute for Information Transmission Problems Bolshoi Karetnii 19, 127994 Moscow, Russia, [4]vnk@iitp.ru, [5]milukova@iitp.ru

[3] Department of Computer Science, Division of Applied Physics CICESE, Ensenada, B.C. 22860, Mexico, vkober@cicese.mx

Pattern recognition based on phase-only filtering designed with partial phase information is presented. The modified phase-only filter is able to discriminate similar objects is proposed. The recognition performance of the suggested filter is compared with that of common correlation filters in terms of discrimination capability. Computer simulation results are provided and discussed.

Introduction

Pattern recognition is one of the most important topics in image and signal processing. One of the ways to take into account features of a signal such as edges, contours, abrupt signal change etc. is to use the phase information of the signal Fourier transform. Moreover, the phase information often is invariant to a linear signal distortion [1]. The classical matched correlation filter [2] is based on the phase information and optimized with respect to mean squared error (MSE) for detection of objects in additive Gaussian noise. Many different types of correlation filters for pattern recognition based on correlation have been proposed [3]. A traditional way to design correlation filters is to make filters that optimize different criteria. Some of these measures can be essentially improved using an adaptive approach to the filter design [4-7]. According to this concept, we are interested in a filter with good performance characteristics for a given observed scene, i.e., with a fixed set of patterns or a fixed background to be rejected, rather than in a filter with average performance parameters over an ensemble of images.

One of the most important performance criteria in tasks of pattern recognition is the discrimination capability (DC), or how well a filter detects and discriminates between classes of objects. A theoretical analysis of correlation methods was made by Yaroslavsky [4]. He suggested a filter with minimum probability of anomalous localization errors (false alarms) and called the optimal filter (OF). An important feature of the OF is its adaptivity in application to pattern recognition or target detection because its frequency response takes into account the power spectrum of wrong objects or an observed scene background to be rejected. The disadvantage of the OF in optical implementation is its extremely low light efficiency.

Another fruitful approach to synthesis of adaptive correlation filters with improved performance parameters is based on idea of zero-blocking the phase-only filter [8] (POF) transmission at some frequencies [9-12].

In this paper, we suggest another method derived from the definition of discrimination capability to design a filter for real-valued images with a given value of discrimination capability using partial phase information of POF spectrum components of similar objects. With the help of computer simulation we compared the performance of the proposed method with that of common correlation filters in terms of discrimination capability.

[1] Support of the grant RFBR 11-07-00361 Development of particularized software system for automatic diagnostic and restoration of distorted images.

Design of phase-only filter with improved discrimination performance

For simplicity of the analysis, we assume that an input scene contains the target and a similar pattern to be rejected. These two objects are nonoverlapping. The discrimination capability for pattern recognition may be expressed as [4]

$$DC = 1 - \frac{\left|C^B\right|^2}{\left|C^T\right|^2} \qquad (1)$$

where C^T and C^B are the target and the non-target correlation peaks, respectively; that is, C^B is the maximum in the correlation plane over the background area to be rejected; C^T is the maximum in the correlation plane to be recognized. In the case of two objects (the target and the non-target), C^B is the cross-correlation peak, and C^T is the auto-correlation peak. For the filter synthesis, let us assume that $t(x,y)$ and $b(x,y)$ represent respectively the target centered at the origin, and the non-target placed in the scene at coordinates that the cross-correlation intensity maximum appears at the origin. The POF is a basic filter for filter synthesis. The transfer function of a conventional POF is given by [8]

$$H_{POF}(u,v) = \frac{T^*(u,v)}{\left|T(u,v)\right|} = \begin{cases} \exp\left(-i\Phi_t(u,v)\right), & if \quad \left|T(u,v)\right| \neq 0 \\ 0 & \text{otherwise} \end{cases} . \qquad (2)$$

Let us apply the filter to the target image. The auto-correlation peak for the centered target image coincides with the origin. For the case of the non-target image, we repeat a similar procedure, however, instead the auto-correlation intensity peak we use the cross-correlation intensity peak (defined as the cross-correlation intensity maximum). Let

$$T(u,v) = \left|T(u,v)\right| \exp\left[i\,\Phi_t(u,v)\right]$$
$$B(u,v) = \left|B(u,v)\right| \exp\left[i\,\Phi_b(u,v)\right] \qquad (3)$$

be the Fourier transforms of the target $t(x,y)$ and the non-target $b(x,y)$, respectively. Here $\Phi_t(u,v)$ and $\Phi_b(u,v)$ are their phase distributions. When the POF is used, and the input scene is real-valued, and it contains the reference image $t(x,y)$ and the object to be rejected $b(x,y)$, the discrimination capability (1) can be written as

$$DC = 1 - \frac{\left|2\int_{-\infty}^{+\infty}\int_{0}^{+\infty}\left|B(u,v)\right|\cos\left[\,\Delta\Phi(u,v)\right]dudv\right|^2}{\left|2\int_{-\infty}^{+\infty}\int_{0}^{+\infty}\left|\,T(u,v)\,\right|dudv\right|^2} . \qquad (4)$$

By substituting a predefined value of the DC, say $\left(1-\varepsilon^2\right)$, into (4) after some manipulations, we arrive at a relation consisting of the product of two expressions:

$$\left(\int_{-\infty}^{+\infty}\int_{0}^{+\infty}\left(\left|B(u,v)\right|\cos\left[\,\Delta\Phi(u,v)\right]-\varepsilon\left|T(u,v)\right|\right)dudv\right)$$
$$\left(\int_{-\infty}^{+\infty}\int_{0}^{+\infty}\left(\left|B(u,v)\right|\cos\left[\,\Delta\Phi(u,v)\right]+\varepsilon\left|T(u,v)\right|\right)dudv\right)=0 . \qquad (5)$$

Our objective is to force the product to be zero, or, for the case of numerical filter synthesis, to minimize the absolute value of the product in (5). Note that the product is equal to zero if at least one of the product parentheses is equal to zero. This can be achieved using only partial phase information at some frequencies. To enable the use of a computer, we replace all the integrals by sums and discretize all continuous signals according to a sampling rate. The total algorithm consists of two independent steps. First, samples of the first expression in (5) are sorted into variational row based on their magnitudes. Then we choose the minimum quantity of samples in the variational row to be chosen until the equality condition in (5) is satisfied. If, for instance, the calculation result of the first parenthesis in (5) is greater than zero we assign the amplitudes in variational row equal to zero starting with the highest value sample.

In this way we choose the minimum number of elements of the first expression in (5). Second, the above-described procedure is repeated for the samples of the second parenthesis in (5). By comparing the two results, we shall obtain a support region of the POF with the minimum quantity of selected elements. In general, for real and positive images it is only necessary to take into account the first expression in (5).

Computer simulation

In this section, we apply the conventional POF, the OF, and the modified phase-only filter described in section 2 on the same test image. The scene shown in Fig. 1 consists of two similar character objects placed in 256x256 array: E and F (the letter F is contained in the letter E).

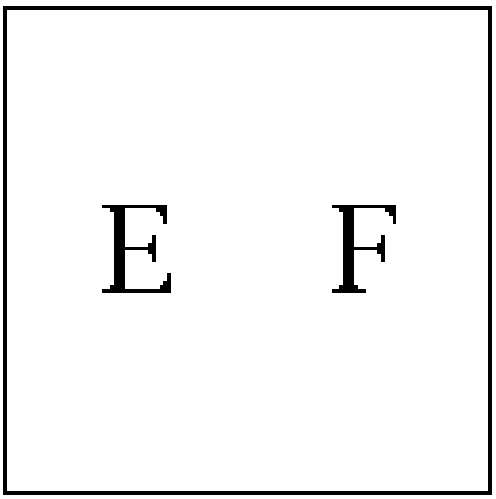


Fig. 1. Test scene.

The target is the letter F. The correlation performance in terms of the number of blocked (unused) elements (NBE) normalized to the number of image pixels versus the discrimination capability (DC), obtained with the proposed method is shown in Fig. 2. The performances with respect to the DC of the POF and the OF are 0.1 and 0.7, respectively. Note that the proposed algorithm essentially improves the DC compared with the POF and may achieve better discrimination than that of the OF. This means we can design filters which are more adaptive to a given scene than the OF. The price of these improvements is the decrease of the light efficiency compared with that of the POF. The set of frequencies for which the filter magnitude is not equal to zero (selected phase distribution) is called the region of support.

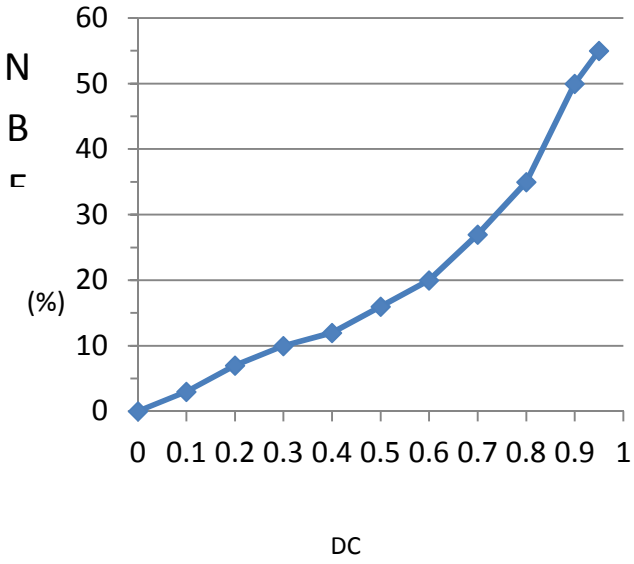


Fig. 2. Performance of the proposed method in terms of NBE versus DC.

The binary image shown in Fig. 3 is the region of support computed for the test image by the proposed method.

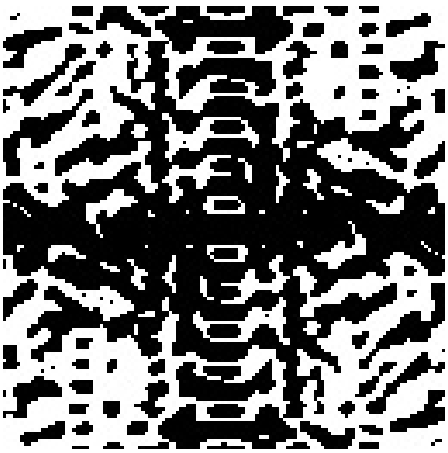


Fig. 3. Selected phase components (white areas) yielding DC of 0.85.

The image center represents the origin (zero frequency), the black regions are where the POF is blocked, and the white represents places where the POF is transparent. The number of used phase components of 45% helps us to yield the value of the DC of 0.85.
The correlation intensity distributions obtained with the conventional POF and with the

proposed method are presented in Figs. 4a and 4b, respectively.

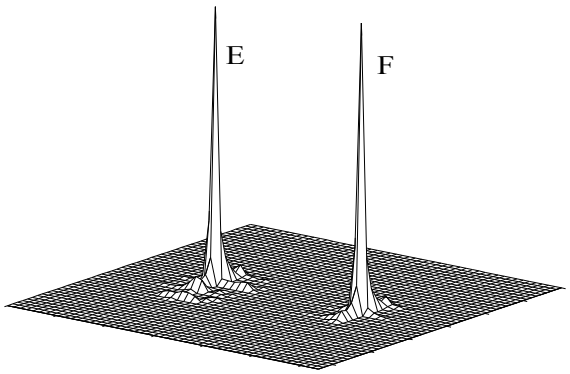


Fig. 4a. Correlation intensity distribution for the test image obtained with the conventional POF.

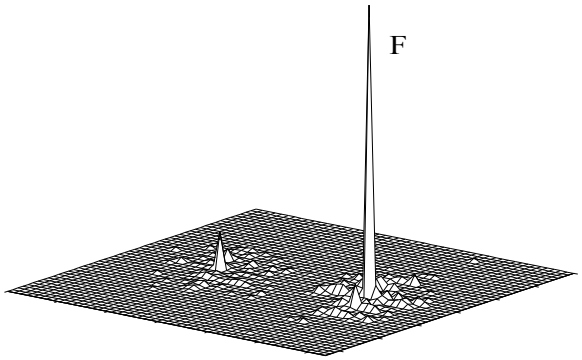


Fig. 4b. Correlation intensity distribution for the test image obtained with the proposed POF.

Note that in Fig. 4a the output correlation plane of the conventional POF has two almost equally high correlation peaks. In contrast, the output correlation plane in Fig. 4b obtained by the proposed algorithms consists of a high correlation peak at the position of the target. These results clearly demonstrate the advantages of the proposed method.

Conclusion

We used the definition of discrimination capability to design filters for real-valued images with a given value of discrimination capability using partial phase information of POF spectrum components of similar objects. With the help of computer simulation we compared the performance of the proposed method with that of common correlation filters in terms of discrimination capability.

The design method was derived under assumption that the maximum cross-correlation is centered at the origin and is not moved during the blocking operation. For tested images that we have used in computer simulation this assumption is satisfied. A generalization of the method to an input scene with several objects to be rejected is straightforward.

In future it is of interest to investigate the use of phase information for recognition of linear distorted similar objects imbedded into a noisy cluttered background [13]. The proposed method can be adapted for these circumstances.

References

1. O. Milukova, V. Kober, V. Karnaukhov, I.A. Ovseevich. Spectral analysis of distorted images in restoration problems // Pattern Recognition and Image Analysis. - 2010. -Vol. 20, No. 3. – P. 335-340.
2. A.B. VanderLugt, Signal detection by complex filtering // IEEE Trans. Inf. Theory. -1964. -Vol. 10. –P. 135-139.
3. V.B.V.K. Kumar and L. Hassebrook. Performance measures for correlation filters // Applied Optics. – 1990. - Vol. 29. –P. 2997-3006.
4. L.P. Yaroslavsky, The theory of optimal methods for localization of objects in pictures // in progress in Optics XXXII, E. Wolf, Ed., Elsevier. -1993. -P. 145-201.
5. B. Javidi, J. Wang, Design of filters to detect a noisy target in nonoverlapping background noise // Journal OSA (A), -1994, Vol. 11. –P. 2604-2612.
6. V. Kober and J. Campos, Accuracy of location measurement of a noisy target in a nonoverlapping background // Journal OSA (A). -1996, Vol. 13. –P. 1653-1666.
7. V. Kober, M. Mozerov, and I. A. Ovseevich. Adaptive Correlation Filters for Pattern Recognition // Pattern Recognition and Image Analysis. -2006. -Vol. 16, N. 3. - P. 425-431.
8. J. Horner, P. Gianino. Phase-only matched filtering // Applied Optics. -1984. Vol. 23. –P. 812-816.
9. V.B.V.K. Kumar and Z. Bahri. Phase-only filters with improved signal to noise ratio // Applied Optics. – 1989. - Vol. 28. – P. 250-257.
10. V.B.V.K Kumar, W. Shi, C. Hendrix. Phase-only filter with maximally sharp correlation filter // Optics Letters. – 1990. - Vol. 15. - P. 807-809.
11. E. Ahouzi, J. Campos and M.J. Yzuel. Phase-only filter with improved discrimination // Optics Letters. – 1994. - Vol. 19. –P. 1340-1342.
12. V. Kober and I.A. Ovseyevich. Phase-only filter with improved filter efficiency and correlation discrimination. // Pattern Recognition and Image Analysis. – 2000. -Vol. 10, No. 4. - P. 514-519.
13. E.M. Ramos Michel, V. Kober. Design of correlation filters for recognition of linearly distorted objects in linearly degraded scenes // Journal OSA A. – 2007. - Vol. 24. No. 11. - P. 3403-3417.

ABOUT STOPPING OF THE BUILDING PROCESS OF HIERARCHICAL REGRESSION IN THE CASE OF CONSTRUCTION OF COMPUTATIONAL PROCEDURES OF LOCAL IMAGE PROCESSING[1]

V.N. Kopenkov[2]

**[2] Image Processing Systems Institute of RAS,
443001, Molodogvardeiskaja st., 151, Samara, Russia; +7 (846) 3378084,
vkop@smr.ru.**

The paper deals with the usage of the functional of full cross-validation when solving the problem of building a local signal/images processing procedures, based on empirical data (hierarchical regression). It is proposed a methodology of stopping of the process of formation of various combinations of training and control samples, and the process of constructing of processing procedure, based on the interval estimation of the functional of sliding quality control.

Introduction

The tasks of local processing and analysis of digital images are widely used in various fields of human activity. At the same time, the distinction of processing algorithms and handling large amounts of data (satellite images, remote sensing data, hyperspectral data, multi-dimensional signals and so on), taking into account the requirements of the high performance leads to the need of usage non-linear local transformation routines [3,5,7]. One of the most common approaches currently in use is the realization of cybernetic principle of "black box" (the terms of other authors is the processing via recognition, processing basing on precedents). In this case transformation itself and its parameters are determined based on analyzing of the input and output signals, or images [4,6,7].
The classic approach to constructing of approximately universal procedures of local adaptive digital signal and images processing, which implements the principle of "black box" is based on usage of technique of artificial neural networks[7]. As alternative, but substantially less researched version of described solution is the use of hierarchical computational structure, such as the decision trees and regression trees [4,6]. In both cases, one of the main problems is the solution of stopping the learning process, as well as the veracity of the obtained result.
Usually, for estimation of the generalization capability and selecting a stopping rule of the learning procedure is used statistical theory of Vapnik–Chervonenkis [1], which relates the three parameters of training: training error, veracity (reliability), and the length of training dataset. But estimations of statistic theory is highly inflated and not take into account the possible rearrangement of the elements of training and testing datasets. A more efficient way to estimate of the generalization capability is the use of combinatorial theory of reliability of Vorontsov [2], which is based on evaluation of the functional of full cross-validation, assuming verification all possible combinations of division dataset into training and testing part. The correct solution of the task of constructing of image processing procedures which takes into account all combination of training and control datasets is unrealizable in practice because of the giant enumeration of variants of different combinations. Therefore, this article proposes decision procedure, which allows to determine the need to stopping or continuation of formation of training and testing datasets is necessary for constructing

[1]This research was partly financially supported by:the Ministry of Education and Science of the Russian Federation (under the decree of the Government of the Russian Federation of 09.04.2010 № 218),RFBR projects: 12-07-3117512_мол_а, 12-07-31175\13, 12-07-00021-а, 12-07-00751-а, 13-07-97006-р_поволжье_а,

processing algorithm according to their concrete number.

Task specification

The model of local image processing, which implement the principle of "black box" (processing through the recognition or processing based on precedents), in most cases, suggests decomposition of the transformation into two stages: the formation of the description of the fragments of image (formation of the features) and calculation of transformation results [4]. The construction of the transformation is realized on the basis of processing precedents – a set of matched pairs of images (the "entry" X and the desired "outcome of the processing" Z) with the minimization of transformation errors.

Then have to perform the analysis of the generalization capability of constructed procedure on the basis of combinatorial theory and full sliding control functional:

$$Q^{st}(\mu(\Omega),\Omega)=\frac{1}{N}\sum_{n=1}^{N}\nu(\mu(\Omega_n^s),\Omega_n^t),$$

where Ω – finite set of objects for training (precedents), which in this case corresponds to the number of the image pixels,

$\mu(\Omega)$ – algorithm (method) of teaching on a set of training dataset Ω,

$(\Omega_n^s,\Omega_n^t),\ n=1,2,...,N$ – all possible partitioning variants of the dataset Ω on a training and control datasets.

$\nu(\mu(\Omega_n^s),\Omega_n^t)$ – frequency of errors of algorithm $\mu(\Omega_n^s)$, constructed based on a dataset Ω_n^s and tested on dataset Ω_n^t.

The total number of all possible N decompositions of dataset is equal C_T^s. General scheme of construction procedure is shown in Figure 1.

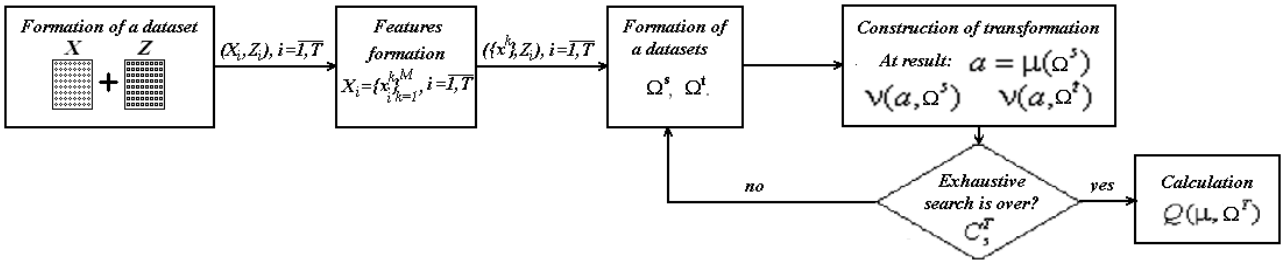


Fig. 1. Scheme of construction procedure with exhaustive search on datasets

Is quite logical is the fact that when working with images solution of problem of constructing the processing procedures taking into account the all combination of training and testing dataset is unrealizable in practice because of the giant of busting on various combinations of datasets. Therefore, it is necessary to develop a method to determine the possibility of stopping a forming, or the need to continue training and testing datasets generation on the base of a finite number of them.

The proposed solution

With rather large volumes of the sampling can be assumed that the error rate of the algorithm has a binomial distribution with t degrees of freedom (the length of the testing dataset) and the probability of "success" is p (quality of the algorithm on a control set). In this way:

$$\nu(\mu(\Omega_n^s),\Omega_n^t)\sim Bin(t,p).$$

Function of the probability is specified as:

$$p_v(r)=C_t^r p^r(1-p)^{t-r},\ \ r=\overline{0,t}.$$

Then the distribution of the functional full cross-validation is evaluated:

$$Q^{st}(\mu(\Omega),\Omega)=\frac{1}{N}\sum_{n=0}^{N-1}\nu\left(\mu(\Omega_n^s),\Omega_n^t\right)\sim Bin(N\cdot t,p).$$

Based on the analysis of functional $Q_1^{st}\sim Bin(N_1\cdot t,p_1),\ \ Q_2^{st}\sim Bin(N_2\cdot t,p_2)$ for the different subsets of features (in the case when we decide whether to recalculate the feature space, or to stop the process of building a processing procedure) under the assumption of $p_2<p_1$, with veracity γ (and correspondingly $p_1<p_2$, with the veracity $(1-\gamma)$) the decision about continuing or stopping of process of generating of different combinations of training and control datasets

(and transition to the next subset of features) can be taken.

In such case the quality of the algorithm on dataset Ω, accordingly can be estimated as:

$$\nu(\mu(\Omega),\Omega^T) \sim \frac{1}{|\Omega|}\sum_{\omega_i\in\Omega} I(\omega_i,\mu(\omega_i)),$$

where $I(\omega_i,\mu(\omega_i)) = \begin{cases} 1, & p \\ 0, & 1-p \end{cases}$.

Moreover, if $n >> 1$ (the number of objects – and that is justified, because corresponds to the image size) and the λ is fixed, we obtain the Poisson distribution with parameter λ:

$$Bin(n, \lambda/n) \approx P(\lambda).$$

In this case to make a decision of stopping generation of various combinations of training and control datasets, and about the transition to the next subset of features it is necessary to calculate confidence intervals for the expectation of a Poisson distribution for the functional full cross-validation on a datasets N_1 and N_2 in form:

$$\left[\lambda_1 - \frac{\tau_{1-\alpha/2}\sqrt{\lambda_1}}{\sqrt{N_1}}, \lambda_1 + \frac{\tau_{1-\alpha/2}\sqrt{\lambda_1}}{\sqrt{N_1}}\right]\left[\lambda_2 - \frac{\tau_{1-\alpha/2}\sqrt{\lambda_2}}{\sqrt{N_2}}, \lambda_2 + \frac{\tau_{1-\alpha/2}\sqrt{\lambda_2}}{\sqrt{N_2}}\right],$$

where $\tau_{1-\alpha/2}$ – quantile of distribution $N_{0,1}$ for level $1-\alpha/2$ $(\alpha = 1-\gamma)$.

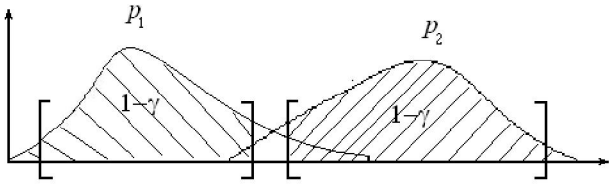


Fig.2. Constructing of confidence intervals

The decision of stopping of generation of various combinations of datasets and about the transition to the next subset of features is taken at a time when a separation of calculated confidence intervals on adjacent steps is achieved.

Illustration of the process of processing algorithm constructing (training).

As a family of features have been used characteristics obtained as a result of local discrete wavelet transforms of signal and image. Such features corresponds to the following specifications:

existence of the efficient calculation algorithm [8];

completeness of description of the input signal;

consistently obtaining and using of features that eliminates the problem of exhaustive search on the features space.

Issues related to the features formation on the basis of algorithms of local DWT, as well as their advantages by using in local image processing tasks, is considered in [8].

As the processing procedure (transformation function) were examined the regression trees and the artifical neural network.

Figure 3 shows an example of training of a regression tree, for different sets of features (K=1,2,3,...,12, with a gradual increase). The graphs show the noise reduction (ε^2/D_V) with the increasing of a regression tree depth (Hav). Figure 4 presents a statistics of process of regression tree construction on a various combinations of training and control datasets (group of points of each colors correspond to the optimum value of quality for a given set of features K=1,2,3,...,12, in the case of exhaustive search a some number of variants of partitioning (Ω_n^s, Ω_n^t), $n = 1,2,...,N$ of a dataset Ω on training and control part).

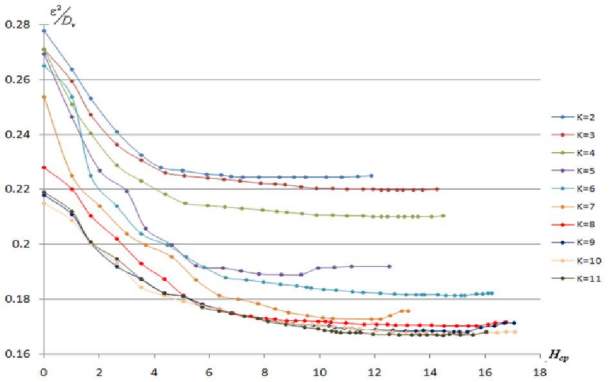


Fig. 3. Process of training of processing procedure at various features space

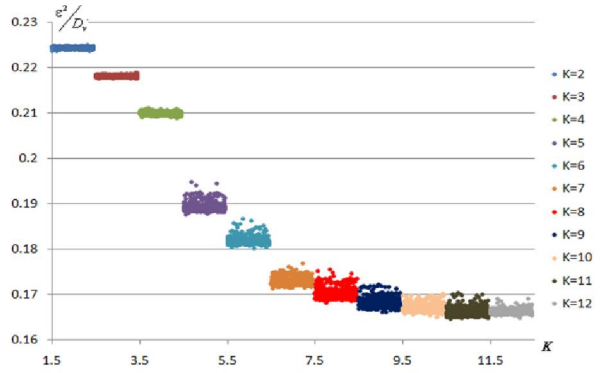


Fig. 4. Statistics of quality of procedures for different combinations of training and control datasets

Figure 5 is a graph of the construction process of local image processing procedures, with confidence intervals, for the optimal values of quality (the dependence of noise reduction on the number of features), and Figure 6 - calculation of the required number of combinations of training/testing datasets for the making a decision of switching to the next set of features (the number of combinations required for the separation of the confidence intervals on the adjacent steps).

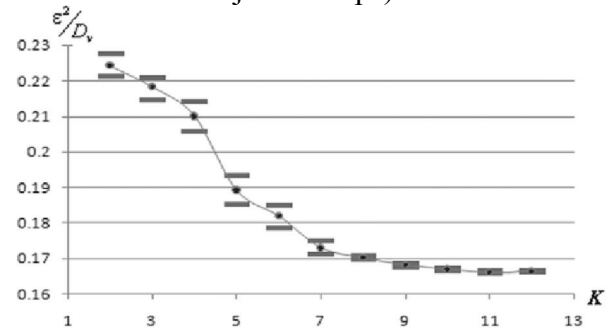


Fig. 5. Construction of the processing procedure with confidence intervals

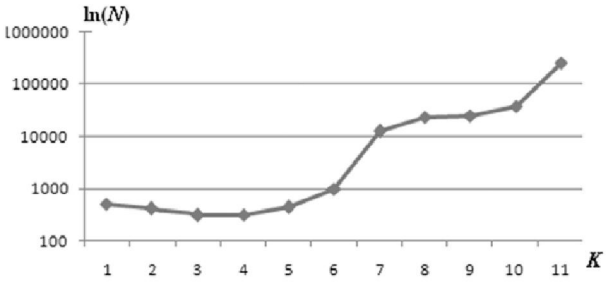


Fig. 6. Calculating of combinations number

Conclusion

The proposed method of estimation of the required number of training iterations of the algorithm processing and, as a consequence, the stopping rule of the formation of different combinations of training and testing datasets based on their particular number allows to use the full functionality of combinatorial theory and a functional of full cross-validation control during the constructing (training) processing procedures, which are tuning on the bases of a training dataset. As a result, it is possible to avoid the problems of retraining / poorly trained processing algorithms and, at the same time, construct the local processing procedure with predetermined computational complexity and veracity, and with the best quality (for an existing training dataset).

References

1. Vapnik, V.N. The theory of pattern recognition / VN Vapnik, A. Chervonenkis / - Moscow: Nauka, 1974. (in russian)
2. Vorontsov K. A combinatorial approach to assessing the quality of training algorithms / / Mathematical problems of cybernetics / Ed O.B. Lupanov. Moskow. Fizmatlid. 2004. Vol. 13. p. 5–36. (in russian)
3. Woods, R. Digital Image Processing / R. Woods, R.Gonsales / - M: Technosphere, 2005. - 1072 p. (in russian)
4. Kopenkov V.N., Myasnikov V.V. An algorithm for automatic construction of a local non-linear processing procedures images based on hierarchical regression. Computer optics, 36(2):257-266, 2012. (in russian).
5. Methods of computer image processing. Part II: Methods and algorithms, Ed. V.A. Soifer, Moskow. Fizmatlid. 2009.
6. Breiman L. Classification and regression trees / Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone.// Monterey, Calif., U.S.A.: Wadsworth, Inc. – 1984.
7. Haikin, S. Neural Networks: A Comprehensive Foundation / M.: «Vilyams», 2006. 1104 p.
8. Kopenkov V. Effecient algorithms of local discret wavelet transform with HAAR-like bases. Pattern Recognition and Image Analysis. Vol 18 No 4 2008 pp. 654-661.

THE SYNTHESIS OF WAVE MODEL OF A MULTIDIMENSIONAL RANDOM FIELD WITH A GIVEN NON-MONOTONE CORRELATION FUNCTION[1]

V.R. Krasheninnikov[2]

[2]**Ulyanovsk State Technical University, Severny Venets St. 32, Ulyanovsk, 432027 Russia, tel: +7-8422-77-82-61, fax: 8-8422-43-02-37, e-mail: kvr@ulstu.ru**

To effectively solve various problems of multidimensional data (images) processing their mathematical formulation is needed. It includes a mathematical description of the image. Different models of random fields (RF) are typically used for this purpose. A desirable property of a model is the ability to solve the problems of correlation analysis and synthesis. For example, the problem of RF wave model synthesis is soled rather easily, if RF has a certain decreasing correlation function (CF). In this paper a version of the wave model with a non-monotonic CF is introduced.

Introduction

In a wave model RF is defined by an equation [1]

$$x(\bar{j},t) = \sum_{\{k:\tau_k \le t\}} f((\bar{j},t),(\bar{u}_k,\tau_k),\bar{\omega}_k), \quad (1)$$

where $x(\bar{j},t)$ is the value of RF at a point $\bar{j}$ at time t; $(n+1)$-dimensional domain of $S = \{(\bar{j},t)\}$ field can be grid or continuous; $\{(\bar{u}_k,\tau_k)\}$ is a discrete field of random points (FRP) in $(n+1)$-dimensional continuous space; $\bar{\omega}_k$ is a random vector of f function parameters. This field can be represented as a result of the impact of random disturbances or waves $f((\bar{j},t),(\bar{u}_k,\tau_k),\bar{\omega}_k)$ generated at random places $\bar{u}_k$ at random time τ_k and changing in accord with a certain law in time and space. A choice of function f, FRP and parameters $\bar{\omega}_k$ allows to obtain a wide range of RF, which includes the Poisson fields, multidimensional filtered Poisson processes, weighted sums, the random walk model, parametric models [2], etc.

Let's consider a particular case when

$$f((\bar{j},t),(\bar{u}_k,\tau_k),\bar{\omega}_k) = \\ = \exp(-\mu\,|\,t-\tau_k\,|)\,g(\rho_k/R_k)\xi_k, \quad (2)$$

where FRP is Poisson with a constant density λ; $\rho_k = |\,\bar{j}-\bar{u}_k\,|$ is the distance between $\bar{j}$ and $\bar{u}_k$; $\{R_k\}$ is a system of independent non-negative and identically distributed random variables (RV) with probability density function (DF) $\omega(\alpha)$; $\{\xi_k\}$ is a system of centered, independent and identically distributed RV. In this case, the waves are stationary, independent of each other, have spherical sections in space and exponentially decay with time; the system $\{\xi_k\}$ determines the intensity of the waves, and $\{R_k\}$ determines their spatial scale. Generated field X is stationary, homogeneous, has zero mean and isotropic in space CF. If

$$g(y) = c\exp(-2y^2) \quad (3)$$

CF takes the form of

$$V(\rho,t) = \frac{c^2\pi^{n/2}\lambda}{2^{n+1}\mu}e^{-\mu t}\int_0^\infty \alpha^n \exp\left(-\frac{\rho^2}{\alpha^2}\right)\omega(\alpha)d\alpha, \quad (4)$$

When $\rho = t = 0$ using (4) we find the variation of the field

$$\sigma_n^2 = \frac{c^2\pi^{n/2}\lambda}{2^{n+1}\mu}M[R^n], \quad (5)$$

which is proportional to the density of FRP λ, the effective interval $1/\mu$ of wave attenuation and the n^{th} degree mean observation of spatial scale R.

Simulation of a discrete RF on the n-dimensional grid $\{\bar{j}\}$ with the quantization step Δt using this particular model may be arranged using the following algorithm. At the initial instant $t_0 = 0$ field values in all nodes are equal to zero. At any succeeding moment $t_m = m\Delta t$ on a continuous space or a net,

[1] This study was supported by RFBR grant 13-01-00320

which overlaps $\{\bar{j}\}$, Poisson FRP with density $\lambda\Delta t$ is formed. In each generated FRP point $\overline{u}_k$ RV ξ_k and R_k are played and then the transformation of all field values on grid

$$x_{\bar{j}}^{t_m} = x_{\bar{j}}^{t_{m-1}} \exp(-\mu\Delta t) + \sum_k g(\rho_k / R_k)\xi_k \quad (6)$$

is performed . The advantage of this algorithm is its recurrence, making it easy to implement a programmed field simulation.

Correlation analysis and synthesis

Let's consider now the solution of correlation analysis and synthesis problems. Formula (4) means that the constructed field has an exponential normalized correlation functions (NCF) $e^{-\mu t}$ in time and NCF

$$r(\rho) = \frac{1}{M[R^n]}\int_0^\infty \alpha^n \exp(-\frac{\rho^2}{\alpha^2})\omega(\alpha)d\alpha \quad (7)$$

in space. Thus, solving the problem of analysis, when PDF $\omega(\alpha)$ is specified, the required NCF can be found analytically or by numerical integration. While solving the problem of synthesis when the NCF *r(ρ)* is set, it is necessary to solve integral equation (7) for the unknown PDF $\omega(\alpha)$.

The implementation of a monotone correlation function

Since the analytical solution of the synthesis problem is not always possible, let's consider its approximate solution. From (7) it follows that for degenerate distribution (R=α=const) we obtain

$$\exp(-\rho^2 / \alpha^2). \quad (9)$$

on space NKF.

Let an arbitrary non-increasing NKF $r(\rho)$ be set. Let's approximate it as the sum of gaussoids with positive coefficients:

$$r(\rho) \approx h(\rho) = \sum_i q_i \exp(-\rho^2 / \alpha_i^2), \quad (10)$$

where $\sum_i q_i = 1$. Then, for a discrete distribution $P(R = \alpha_i) = k^{-1} q_i / \alpha_i^n$, where $k = \sum q_i / \alpha_i^n$ the generated field will have NCF exactly equal to $h(\rho)$. Thus, the given model allows approximately to solve the problem of synthesis by varying only the probability distributions of R scale.

The implementation of a non-monotone correlation function

A significant limitation to the above decision of the synthesis problem is the assumption of the given CF $r(\rho)$ monotonicity, as otherwise approximation (10) of a non-monotonic function which decreases in gaussoids is impossible. To generated RF with non-monotonic CF we need non-monotonic space waves. For this let's take function

$$g(y) = \sum_{k=1}^{K} c_k \exp(-\beta_k y^2), \quad (11)$$

instead of (3). That is, each wave will be the sum of gaussoids. Then, for the degenerate distribution (R = α = const) in space we get NCF

$$B\sum_{k,l=1}^{K} \frac{c_k c_l}{(\beta_k + \beta_l)^{n/2}} \exp\left(-\frac{\beta_k \beta_l}{\beta_k + \beta_l}\frac{\rho^2}{\alpha^2}\right), \quad (12)$$

instead of (9) where the normalizing factor

$$B = \left(\sum_{k,l=1}^{K} \frac{c_k c_l}{(\beta_k + \beta_l)^{n/2}}\right)^{-1}. \quad (13)$$

In particular, if $n = 2, K = 2$, i.e. for two-dimensional RF with waves as sums of two gaussoids, we obtain NCF as a sum of three gaussoids:

$$B\left(\begin{array}{l} \frac{c_1^2}{2\beta_1}\exp\left(-\frac{\beta_1\rho^2}{2\alpha^2}\right) + \\ + \frac{2c_1c_2}{\beta_1+\beta_2}\exp\left(-\frac{\beta_k\beta_l\rho^2}{(\beta_k+\beta_l)\alpha^2}\right) + \\ + \frac{c_2^2}{2\beta_2}\exp\left(-\frac{\beta_2\rho^2}{2\alpha^2}\right) \end{array}\right) \quad (14)$$

with normalizing factor

$$B = \left(\frac{c_1^2}{2\beta_1} + \frac{2c_1c_2}{\beta_1+\beta_2} + \frac{c_2^2}{2\beta_2}\right)^{-1}. \quad (15)$$

For example, if $c_1 = 2, c_2 = -1, \beta_1 = 0.4, \beta_2 = 0.2$ we have NKF

$$\begin{array}{l} h(\rho) = 6\exp\left(-\frac{0.2}{\alpha^2}\rho^2\right) - \\ -8\exp\left(-\frac{2}{15\alpha^2}\rho^2\right) + 3\exp\left(-\frac{0.1}{\alpha^2}\rho^2\right). \end{array} \quad (16)$$

To find NCF at non-degenerate distribution of the scale factor it is necessary to substitute (11) in (7) instead of $\exp(-\rho^2 / \alpha^2)$.

The broader class of CF can be obtained using the shifted gaussoids:

$$g(y) = \sum_{k=1}^{K} c_k \exp(-\beta_k (y - a_k)^2), \qquad (17)$$

i.e. a simulated image will be composed of "ringed" spots and NCF will take the form of

$$B \sum_{k,l=1}^{K} \frac{c_k c_l}{(\beta_k + \beta_l)^{n/2}} \exp\left(- \frac{\beta_k \beta_l}{\beta_k + \beta_l} \frac{(\rho + a_l - a_k)^2}{\alpha^2} \right). \qquad (18)$$

Formula (12) is a particular case of (17) at zero shifts $a_k = 0$. If we use equally shifted gaussoids ($\beta_k = \beta$) in (18), we obtain

$$B \sum_{k,l=1}^{K} c_k c_l \exp\left(- \frac{\beta}{2\alpha^2} (\rho + a_l - a_k)^2 \right). \qquad (19)$$

Note that in (19) the dimension *n* of RF affects only the normalization factor *B*. If we get multiple shifts $a_k = (k-1)a$, then

$$g(y) = \sum_{k=0}^{K-1} c_k \exp(-\beta (y - ka)^2), \qquad (20)$$

$$B \sum_{k,l=0}^{K-1} c_k c_l \exp\left(- \frac{\beta}{2\alpha^2} (\rho - (k-l)a)^2 \right). \qquad (21)$$

In this case, wave (20) and NCF (21) have the same structure. In particular, if $c_k = (-1)^k, \alpha = 1$, we obtain:

$$g(y) = \sum_{k=0}^{K-1} (-1)^k \exp(-\beta (y - ka)^2), \qquad (22)$$

$$B \sum_{k,l=0}^{K-1} (-1)^{k+l} \exp\left(- \frac{\beta}{2} (\rho - (k-l)a)^2 \right). \qquad (23)$$

Conclusion

The article introduces a method to synthesize a wave model, which allows to represent and simulate multi-dimensional RF with an arbitrarily given CF. For this purpose it is necessary to approximate a given CF as a sum of shifted gaussoids. In addition it is possible to simulate both individual images and their sequences with NCF exponential in time.

References

1. K. K. Vasil'ev , V. R. Krasheninnikov. Statistic Analysis of Multi–Dimensional Images (UlGTU, Ulyanovsk, 2007) [in Russian].
2. A.S. Shaligin, U.I. Palagin. Applied Methods of Statistic Modeling (Leningrad, Mashinostroenie, 1986) [in Russian].

THE METHOD OF HIGHLIGHTING TEXTURE AREAS IN IMAGES

E.E. Kurbatova[1,2], E.V.Medvedeva[1,3], A.A. Shemyakina[1,4]

[1]Vyatka State University, Kirov, The Russian Federation,
[2]kurbatovae@gmail.com, [3]emedv@mail.ru, [4]anna-a-sh@mail.ru

The texture segmentation method, which is based on random Markov fields. The estimate probability of transition while switching from one image element to another represents a textural characteristic. This method efficiently locates texture areas with different statistical characteristics and reduces computational costs.

Introduction

Segmentation is a commonly used method of image analysis in the earth's surface monitoring systems. It makes possible to locate the specified areas in the images. There are areas on the earth's surface images that do not have clearly determined borders or significant details. In this case, it is essential to analyse the texture of these areas for their identification.

Papers [1] referred to show how random Markov fields could be used to characterize a texture. Although in processing digital halftone (or colourful) images with number of intensity levels equal 2^g there are some difficulties with storing and handling transition probability matrixes with dimension of $2^g \times 2^g$. Such way of image processing requires high computing facilities. The arrangement of halftone g-digit images as a collection of g binary images [2,3] allows reducing computing costs due to handling transition probability matrixes with dimension of 2x2.

This paper contains the description of digital halftone images texture segmentation method based on random Markov fields, which allows locating extensive areas with similar statistical characteristics efficiently and using less computing facilities.

Texture segmentation method

Consider g-digit halftone image to be a collection of g binary images. As there is a statistical concatenation of elements of each binary image, we regard elements of halftone image as forming a two-dimension Markov chain with several states, and elements of binary image as forming a two-dimension Markov chain with two equiprobable states $M_1^{(l)}$, $M_2^{(l)}$ and horizontal transition probability matrix ${}^1\Pi = \left\| {}^1\pi_{ij}^{(l)} \right\|_{2\times 2}$ and vertical transition probability matrix ${}^2\Pi = \left\| {}^2\pi_{ij}^{(l)} \right\|_{2\times 2}$.

The fragment of lth binary image represented in Fig. 1 conforms random Markov field.

Vicinity of element $\nu_3^{(l)}$ consists of elements $\nu_1^{(l)}$, $\nu_2^{(l)}$.

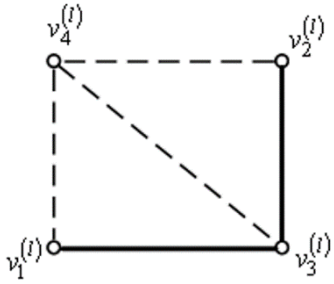


Fig. 1. Fragment of a binary image

As two-dimensional homogeneous Markov chain fully depends on transition probability matrix and initial probability vector, it is proposed to use the estimate of Markov chain transition probability as a textural characteristic. Assume that textural characteristic does not vary appreciably all over an area, but possesses considerably different values in different areas. Consider that texture areas are clear in high-order digits of elements of halftone image, it is essential to

process high-order binary images to locate those areas. Let us use several-dimensional mathematical image model described in [2] to evaluate transition probabilities $\pi_{iii}^{(l)}$ of two-dimensional Markov chain. Element $\nu_3^{(l)}$ state probability (Fig. 1) is defined by element $\nu_3^{(l)}$ entropy in regard to the state of vicinity $\Lambda_{ij}^{(l)} = \left\{\nu_1^{(l)}, \nu_2^{(l)}\right\}$ of element $\nu_3^{(l)}$:

$$H\left(\nu_3^{(l)}\middle|\nu_2^{(l)},\nu_1^{(l)}\right) = -\log\frac{w\left(\nu_3^{(l)}\middle|\nu_1^{(l)}\right)w\left(\nu_3^{(l)}\middle|\nu_2^{(l)}\right)}{w\left(\nu_3^{(l)}\middle|\nu_2^{(l)},\nu_1^{(l)}\right)}, \quad (1)$$

where $w\left(\nu_3^{(l)}\middle|\nu_1^{(l)}\right)$, $w\left(\nu_3^{(l)}\middle|\nu_2^{(l)}\right)$ is one-dimensional adjacent element transition probability density; $w\left(\nu_3^{(l)}\middle|\nu_2^{(l)},\nu_1^{(l)}\right)$ is transition probability density of two-dimensional Markov chain.

Transition probability density of two-dimensional two-state Markov chain can be represented as:

$$w\left(\nu_3^{(l)}\middle|\nu_2^{(l)},\nu_1^{(l)}\right) = \\ = \sum_{i,j,q=1}^{2}\pi\left(\nu_3^{(l)} = M_i^{(l)}\middle|\nu_1^{(l)} = M_j^{(l)};\nu_2^{(l)} = M_q^{(l)}\right)\times \quad (2)$$
$$\times\,\delta\left(\nu_1^{(l)} - M_j^{(l)}\right)\times\delta\left(\nu_2^{(l)} - M_q^{(l)}\right),$$

where $\delta(\cdot)$ is delta-function.

Using (1) for different combinations of vicinity $\Lambda_{ij}^{(l)}$ states we can calculate matrix $\mathbf{\Pi}^{(l)}$ as

$$\mathbf{\Pi}^{(l)} = \begin{Vmatrix} \pi_{iii}^{(l)} & \pi_{iij}^{(l)} \\ \pi_{iji}^{(l)} & \pi_{ijj}^{(l)} \\ \pi_{jii}^{(l)} & \pi_{jij}^{(l)} \\ \pi_{jji}^{(l)} & \pi_{jjj}^{(l)} \end{Vmatrix} = \begin{Vmatrix} \alpha_1^{(l)} & \alpha_1'^{(l)} \\ \alpha_2^{(l)} & \alpha_2'^{(l)} \\ \alpha_3^{(l)} & \alpha_3'^{(l)} \\ \alpha_4^{(l)} & \alpha_4'^{(l)} \end{Vmatrix} \begin{matrix} i,j = \overline{1,2}; \\ i \neq j; \\ l = \overline{1,g}; \end{matrix} \quad (3)$$

where matrix entries are symmetrical and meet the normality condition

$$\alpha_q^{(l)} + \alpha_q'^{(l)} = 1.$$

Values of matrix $\mathbf{\Pi}^{(l)}$ entries (3) depend on the values of matrixes ${}^1\mathbf{\Pi}^{(l)}$ and ${}^2\mathbf{\Pi}^{(l)}$ entries

$$\begin{aligned} &\alpha_1^{(l)} = \pi_{iii}^{(l)} = \pi(\nu_3^{(l)} = M_1^{(l)}|\nu_1^{(l)} = M_1^{(l)};\nu_2^{(l)} = M_1^{(l)}) = \\ &= {}^1\pi_{ii}^{(l)}\ {}^2\pi_{ii}^{(l)} \Big/ {}^3\pi_{ii}^{(l)}; \\ &\alpha_4^{(l)} = 1-\alpha_1^{(l)}; \\ &\alpha_2^{(l)} = \pi_{iji}^{(l)} = \pi(\nu_3^{(l)} = M_1^{(l)}|\nu_1^{(l)} = M_1^{(l)};\nu_2^{(l)} = M_2^{(l)}) = \\ &= {}^1\pi_{ii}^{(l)}\ {}^2\pi_{ij}^{(l)} \Big/ {}^3\pi_{ij}^{(l)}; \\ &\alpha_3^{(l)} = 1-\alpha_2^{(l)}, \end{aligned} \quad (4)$$

where ${}^3\pi_{ij}^{(l)}$ is a transition probability matrix ${}^3\mathbf{\Pi}^{(l)} = {}^1\mathbf{\Pi}^{(l)} \times {}^2\mathbf{\Pi}^{(l)}$ entry.

Estimated value of horizontal transition probability ${}^1\hat{\pi}_{ii}^{(l)}$ depends on average length of sequence of like-sing elements [3]:

$${}^1\hat{\pi}_{ii}^{(l)} = 1 - \frac{2p_1^{(l)}}{\hat{\chi}^{(l,r)}}, \quad (5)$$

where $\hat{\chi}^{(l,r)}$ is the estimated value of the average length of the sequence of equal elements in the l th binary image on the r th step of the estimate adjustment and $p_1^{(l)}$ is the initial probability ($p_1^{(l)} = 0{,}5$).

From the second line of the binary image using the set of elements $\psi = \left\{\nu_1^{(l)}, \nu_2^{(l)}, \nu_3^{(l)}, \nu_4^{(l)}\right\}$ and the previously calculated estimated value of horizontal transition probability ${}^1\hat{\pi}_{ii}^{(l)}$ we can calculate the value of vertical transition probability ${}^2\hat{\pi}_{ii}^{(l)}$ and estimate the value of transition probability $\hat{\pi}_{iii}^{(l)}$ in two-order Markov chain:

$$\hat{\pi}_{iii}^{(l)} = \frac{{}^1\hat{\pi}_{ii}^{(l)} \cdot {}^2\hat{\pi}_{ii}^{(l)}}{{}^3\hat{\pi}_{ii}^{(l)}}, \quad (6)$$

where ${}^3\hat{\pi}_{ii}^{(l)} = {}^1\hat{\pi}_{ii}^{(l)}\,{}^2\hat{\pi}_{ii}^{(l)} + {}^1\hat{\pi}_{ij}^{(l)}\,{}^2\hat{\pi}_{ij}^{(l)}$.

Method of the "scanning frame" allows for local changes of statistical characteristics [4]. Frame size depends on the specified accuracy and the requirement to minimize computational costs.

During the processing of the first row and columns frame size increases according to the processing element index until it reaches the adjusted value. On each step of processing the frame size is determined by expressions

$$\begin{aligned} m &= \overline{1,(2i-1)}, \quad i = \overline{2,(M-1)/2}; \\ n &= \overline{1,(2j-1)}, \quad j = \overline{2,(N-1)/2}, \end{aligned} \quad (7)$$

where i, j are the coordinates of processed element; M, N are adjusted width and height of the scanning frame.

The average transition probability $\tilde{\pi}_{iii}^{(l)}$ for the central element of the frame is estimated by counting the average transition probability within the frame.

$$\tilde{\pi}_{iii}^{(l,r,k)} = \frac{1}{m \times n} \sum_{r=1}^{m} \sum_{k=1}^{n} \hat{\pi}_{iii}^{(l,r,k)} . \qquad (8)$$

In the next step, the frame is shifted by one element to the right and downwards and the average transition probability is estimated for every element of the binary image.

To locate the areas with different texture it is necessary to compare the calculated estimate $\tilde{\pi}_{iii}^{(l)}$ with a threshold value. The threshold value h between different textures is estimated $\tilde{\pi}_{iii}^{(l)}$, the choice of which being based on the image histogram analysis.

If it is known a priory that the image contains two textures with different statistical characteristics, two labels (0 and 1) will be enough to enumerate them. All the elements, for which estimate $\tilde{\pi}_{iii}^{(l)}$ exceeds the threshold value, are labelled with 1, all the rest are labelled with 0.

If a digital grey-tone image contains several textures, for each texture there should be a unique label. In this case, several threshold values corresponding the textures of the image are used.

Modelling results

Modelling utilized real and artificial images formed by means of two-dimensional mathematical model and an algorithm defined in [2]. To get the quality assessment of the image segmentation an artificial image was formed accordingly to the set layout. The texture segmentation in every case was preformed with 21×21 frame.

The performed research proved the frame of this size to be the most efficient as for the ratio of segmentation quality to processing time. The modelling was preformed in a Matlab environment. Fig. 2b represents the segmentation example of the artificial image with the size of 1024×1024 with four texture areas (as it is shown in fig. 2a), and a histogram of the image texture characteristic (fig. 2c). The image contains areas of four textures with transition probability matrixes:

$${}^{1}\Pi_{1}^{(l)} = {}^{2}\Pi_{1}^{(l)} = \begin{Vmatrix} 0{,}9 & 0{,}1 \\ 0{,}1 & 0{,}9 \end{Vmatrix}, \quad {}^{1}\Pi_{2}^{(l)} = {}^{2}\Pi_{2}^{(l)} = \begin{Vmatrix} 0{,}7 & 0{,}3 \\ 0{,}3 & 0{,}7 \end{Vmatrix},$$

$${}^{1}\Pi_{3}^{(l)} = {}^{2}\Pi_{3}^{(l)} = \begin{Vmatrix} 0{,}6 & 0{,}4 \\ 0{,}4 & 0{,}6 \end{Vmatrix} \text{ and } {}^{1}\Pi_{4}^{(l)} = {}^{2}\Pi_{4}^{(l)} = \begin{Vmatrix} 0{,}5 & 0{,}5 \\ 0{,}5 & 0{,}5 \end{Vmatrix}.$$

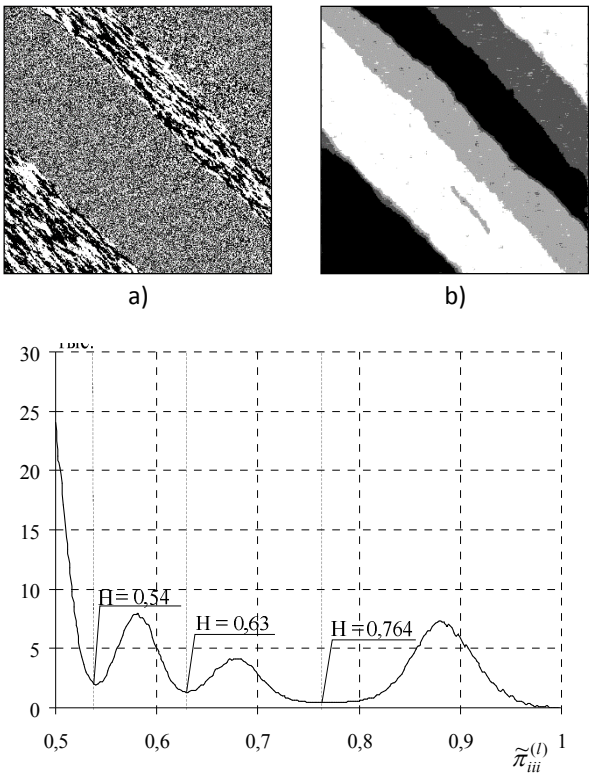


Fig. 2. Segmentation results of an artificial image: a) initial texture mapping; b) result of segmentation using the devised method; c) image texture characteristic histogram

The number of histogram peaks equals the number of textures in texture mapping. The threshold value is set equal to a minimum value between two adjacent peaks of the histogram. The thresholds set using the artificial images with different statistical characteristics can be used to perform the segmentation of other images that contain texture areas with the same statistical characteristics. Fig. 3 represents the segmentation results of the satellite image (with the size 693×709), on which the area of the forest is effectively highlighted. The

segmentation was performed with the use of a 6th binary image.

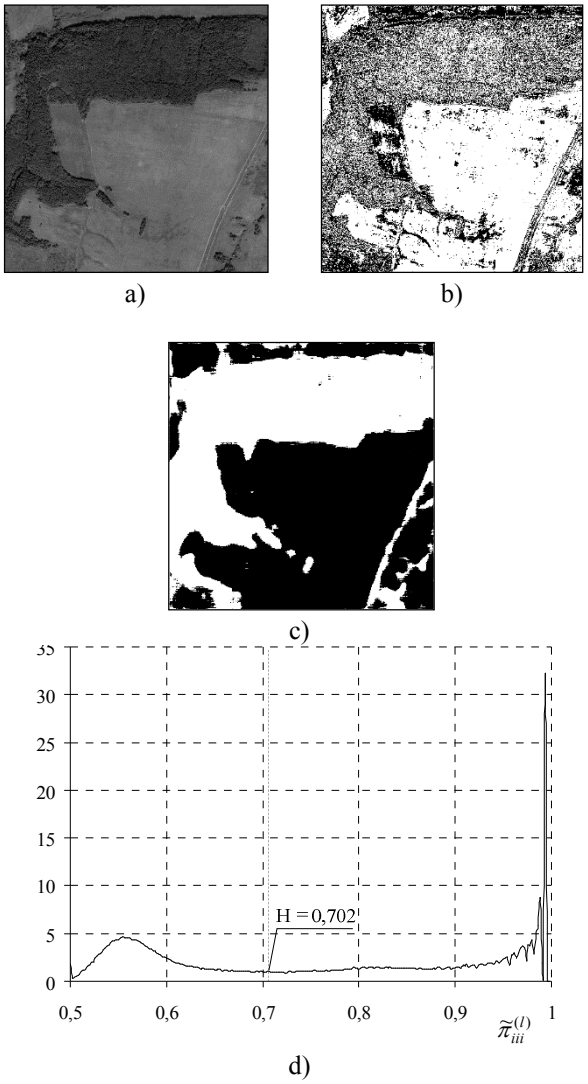


Fig. 3. The segmentation results of a satellite image: a) initial image; b) a binary image, used for segmentation; c) image texture characteristic histogram; d) the result of segmentation using the devised method

The results adduced in fig. 3 show that the devised method efficiently locates two segments; one for the forest area, the other for the field area (fig. 3c). The texture characteristic histogram (fig. 3d) contains two main peaks corresponding two areas in the image. For the forest area $\widetilde{\pi}_{iii}^{(l)} = 0{,}55$ and for the field area $\widetilde{\pi}_{iii}^{(l)} = 0{,}95$. The threshold value is set between two peaks of the histogram.

To assess the quality of performing the texture area locating method and to compare it with known contour methods the quantity of falsely labelled elements was calculated by comparing the segmentation results with the texture mapping. The calculating results are listed in the table.

Comparison results

grey-tone image type	devised method	Canny method	Sobel method	Laplacian of Gaussian method
artificial image	0,0079	0,7665	0,4811	0,5956
real image	0,0388	0,7268	0,4707	0,5969

The results show that the devised method segmentation fault is more than 10 times less in comparison with the known methods based on the calculation of a gradient. Hence, the devised texture area locating method based on random Markov fields efficiently locates extensive areas with similar statistical characteristics and less computational costs.

References

1. Kovtun I.V. Texture image segmentation based on random Markov fields (Control systems and machines, 2003, № 4. pp. 46-55.)
2. Petrov E.P. “Synthesis method of mathematical models of video based on several-dimensional Markov fields” Ed. by E.P. Petrov, E. V. Medvedeva, A. P. Metelev (Nonlinear World, № 4, 2011, pp.213-231.)
3. Medvedeva E. V. Digital grey-tone images informational characteristics computing method Ed. by E. V. Medvedeva, E.P. Petrov (Infocommunication Technologies, vol.6, №3, 2008. pp.104-109.)
4. Shapiro L. Computer Vision Ed. by Linda G. Shapiro, George C. Stockman; (Moscow, BINOMIAL. Knowledge laboratory, 2006, 752 p.) [in Russian].

VEHICLE VIDEO DETECTION METHOD[1]

V. Kustikova[2], I. Meyerov [2], N. Zolotykh [2]

**[2] Lobachevsky State University of Nizhni Novgorod,
Gagarin ave., 23, Nizhni Novgorod, 603950, itlab.ml@cs.vmk.unn.ru**

This paper considers the vehicle video detection and tracking problem. The mathematical model of the problem is formulated. A new method is proposed. The method consists of video partition into blocks of equal length and detection for the first and the last block frames with the following reconstruction of vehicle locations at the intermediate frames. Proposed method demonstrates above 70% true positive detection rate, and approximately 0.3 false positives per frame on some real traffic video sequences.

Introduction

The video-based vehicle detection is considered. This problem comes from traffic stream analysis, traffic violation detection, automatic traffic light control and other important applications. At present the problem still has no final solution so further research in this field make practical sense.

This paper is organized as follows. First an overview of existing methods of problem solution is given. Then a mathematical model of the problem is formulated. The next section provides a new approach for the problem solution. Then the experimental results are discussed.

Related work

Detailed overview of modern methods of the video-based vehicle detection problem solution is given in [25]. The problem involves a lot of researchers [4, 11, 12, 18]. Most of the works are based on the license plate numbers detection [2, 14, 24]. Advantage of this method is simplicity of implementation however its usage requires high definition video. The other methods detect vehicle location as self-contained object, i.e. these methods are not intended to detect single parts and primitives. Such methods are less exacting to the quality of input video. These methods are divided into two subgroups [13, 20, 23]:

- Motion detection and blob tracking methods with the following vehicle identification [21];
- Vehicle detection and tracking methods [20].

Image pre-processing methods are generally used before object detection and tracking. Particularly pre-processing includes shadow suppressions [5, 6, 10, 19] and regions of interest (traffic lanes or road area) highlighting [1].

Video-based vehicle detection problem

Method of video detection deals with input video stream, that is represented as a frames sequence $I_0, I_1, \ldots, I_{N-1}$, where N is a number of frames, w, h are frames width and height. Assume that the object *location* is defined by the bounding box placement [8, 9, 20].

The problem of vehicle video detection consists in mapping each frame I_k into the set of objects locations B_k. Thus the problem of video detection is reduced to construction of mapping φ:

$$\varphi : I_k \to B_k \qquad (1)$$

where $B_k = \{b_l^k, \quad l = 0, \ldots, |B_k| - 1\}$, $b_l^k = ((x_1^l, y_1^l), (x_2^l, y_2^l) [, s^l, c^l])$, (x_1^l, y_1^l), (x_2^l, y_2^l) are coordinates of left upper and right lower corners of the bounding box

1 This research was done at Intel-UNN Lab “Information Technologies” at CMC department and supported by the Russian Federation President Grant “Scientific School-1960.2012.9”, Federal Grant-in-Aid Program “Human Capital for Science and Education in Innovative Russia” (Governmental Contract No. 14.B37.21.0393).

$(0 \le x_1^l < x_2^l < w,\ 0 \le y_1^l < y_2^l < h)$, $s^l \in \mathbb{R}$ is a confidence, c^l is a vehicle class.

The model under consideration requires construction of objects tracks. Let us consider the map ψ from the vehicles locations of the frame I_k to the vehicles locations of the frame I_{k+1}. For the certain frame the object disappears from the camera's field of view. So let us expand a set of locations by the singular location. In this order we choose a box with illegal coordinates that degrades to the point $b = ((-1, -1), (-1, -1)[, s, c])$. The map ψ is:

$$\psi : B_k \to B_{k+1} \cup \{b\} \qquad (2)$$

Let us assume that I_k is the first frame where the object was detected and that the object is contained in the bounding box with index $r_0(k)$, and that q is the number of frames where the object is visible. Then we say that *a track* is a sequence of locations

$$T_{r_0(k)}^{\ k} = \left(b_{r_0}^{\ k}, b_{r_1}^{\ k+1}, \ldots, b_{r_{q-1}}^{\ k+q-1}\right), \qquad (3)$$
$$b_{r_i}^{\ k+i} = \psi\left(b_{r_{i-1}}^{\ k+i-1}\right), i = \overline{1, q-1}.$$

Vehicle video detection method

Let us divide the video into images *blocks* of the length *step*. The idea of the proposed method is to detect vehicles for the first and the last block frames and reconstruct the locations for the intermediate frames. We denote:

- B_i is a set of objects locations that are detected for the frame I_i;
- B_i' is a set of objects locations for the frame I_i. These objects have the track identifier, i.e. these objects were visible in the previous block;
- B_i'' is a set of objects locations. For these objects it is decided that they present in the frame I_i inside the matching bounding box;
- T^i is a state of a set of tracks at the moment of processing the frame I_i.

In the proposed method the blocks are handled sequentially. Decisions on the objects locations are made in the process. Computational scheme of each method's iteration consists of the following stages:

1. Vehicle *detection* in the frame I_{i+step} and construction of locations set B_{i+step}.
2. *Matching* the locations sets B_i and B_i' to the set B_{i+step}.
3. *Construction* and *correction of the sets* B_i'', B_{i+1}'', … , $B_{i+step-1}''$, B_{i+step}', B_{i+step}, T^{i+step}.
4. *Saving of the locations* that are contained in the sets B_i'', B_{i+1}'', … , $B_{i+step-1}''$.
5. *New images block loading*.
6. *Working sets update*: $B_i = B_{i+step}$, $B_i' = B_{i+step}'$, $B_i'' = B_{i+1}'' = \ldots = B_{i+step-1}'' = \varnothing$.

Detection stage supposes usage of vehicle detection algorithm for static images. The quality of a chosen detector has a significant influence on the final result.

Let us consider in detail the stage *matching*. The goal of this stage is to construct matching sets MBS_i and MB_i for locations from B_i' to B_{i+step} and from B_i to B_{i+step}. In order to achieve this goal an exhaustive search for all possible pairs of locations is executed. And then we match images of the objects restricted by bounding boxes. Matching supposes feature points detection and complete matching of the descriptors with the following cutoff of outliers by RANSAC [17]. Experiments showed that the most efficient descriptor for this task is SURF [3]. For locations with fixed index (from B_i' or B_i) the matching with the maximal number of inliers is chosen. The maximal number of inliers should be more than some fixed threshold value. In order to ensure that different locations didn't match into the same locations of the set B_{i+step} one more additional run is executed. As a result of this run the choice in accordance with the maximal number of matching is made.

Construction and correction of working sets is the most labor-consuming stage that consists of three steps:

1. *Working sets update based on matching results in* MBS_i. This step supposes extension of existing tracks in T^i. For this purpose the elements of MBS_i are viewed and then the pair of matching locations in B_i' and B_{i+step} is chosen. After that the locations in the intermediate frames are reconstructed. The borders of the bounding box for locations from B_i' are displaced on some confidence interval. As a result the region of feasible object location on the next frame is defined. Then the matching of regions pairs (procedure is the same as for the second scheme stage) is executed as well as correction of bounding box coordinates. Coordinates are corrected based

on positional relationship of feature points that belong to the set of inliers and to the bounding box of the previous frame. It is assumed that bounding boxes sizes are the same for all frames in the block. Then based on reconstructed locations, track identifier is defined equal to the identifier of the chosen location from B_i'. And they are placed into the sets $B_{i+1}'', \ldots, B_{i+step-1}''$. Examined locations from B_i' are relocated into B_i'', examined locations from B_{i+step} are relocated into B_{i+step}'.

2. *Working sets update based on matching results in* MB_i. As a result new tracks are constructed or the existing tracks are updated on the condition that object was outside of camera's field of view for some previous time period. Initially an attempt to reconstruct the locations of the intermediate frames similarly to the previous step is executed. Then we look for such track in T^i that constructed set of locations is feasible continuation for it (matching of the final and initial fragments). If such track doesn't exist then the new track is created otherwise the set is attached to the existing track. In accordance with this the location track identifier is defined and the constructed locations are placed into the sets $B_{i+1}'', \ldots, B_{i+step-1}''$. Examined location is relocated from B_i into B_i''. The location from B_{i+step} is placed into в B_{i+step}' on the condition that the track is reconstructed for all intermediate frames. Otherwise we consider the examined matching of locations as outlier.

3. *Working sets update in accordance with current state of the set* B_{i+step}. The aim of this step is revealing the situations when the object appears on one of the intermediate frames. For this purpose the tracks are reconstructed inversely in time similarly to the first step of correction. Then we look for correspondent track in T^i (overlapping situation). If such track exists then it is attached by reconstructed locations otherwise the new track is created. Reconstructed locations are assigned by the correct track identifier and these locations are placed into the working sets $B_{i+step-1-lt}'', \ldots, B_{i+step-1}''$, where *lt* is a number of the reconstructed locations. Then examined location from B_{i+step} is placed into B_{i+step}'.

The rest of the method stages are technical and support capability of results saving with the following transition to the next frames block consideration. Note that such scheme allows detecting objects at the moments of appearing in the frame and leaving the frame when the object becomes so feasible that detector is able to find its location. Obtaining of previous/future locations may be achieved after the whole video processing or in parallel with the main iterations execution after the tracks are constructed.

Software implementation

The software implementation uses OpenCV computer vision library [15]. We chose Latent SVM [7] as a detection algorithm. The vehicle classifier (class CAR) was trained using PASCAL Visual Object Challenge 2007 data set [16]. Additionally the *cut* of locations that belong outside the marked region of interest was executed. Source code is available for download [22].

Experimental results

We collected videos of traffic (25 FPS, 720x405):

- *track_10_5000-7000* (2000 frames = 80 s, ~3000 bounding boxes, 58 tracks, left figure) – video with vehicles of the only class CAR, that move in 4 lanes of the same direction.
- *track_10_7000-8000* (1000 frames = 40 s, ~1000 bounding boxes, 29 tracks, right figure). Here we have objects of two classes, CAR and BUS.

Note that partially visible objects are present at each set of bounding boxes.

In this paper the following factors are used to evaluate the detection quality: *average precision* (AP) [16]; *the true positive rate*, (TPR); *the false detection rate* (FDR); *the average false positives per frame* (FPF) [13, 20, 21]. It is assumed that the object is detected correctly if overlapping percentage of the detected and marked bounding boxes is more than some threshold value (for TPR, FDR and FPF threshold value was chosen equal to 50%). Method parameter *step* was chosen equal to 5. The results are shown in the table. Falling of the factors for the second video is the consequence of the BUS class objects presence. Small buses were detected as

cars (mini-vans) because of its similarity of appearance.

Video	AP	TPR (%)	FDR (%)	FPF
track_10_5000-7000	0.68	74.8	19.9	0.27
track_10_7000-8000	0.68	71.3	32.4	0.38

Mean processing time per one images block is 3.5 s (Windows 7, Intel® Core i5 CPU, 2 physical and 2 virtual cores, 2.40 GHz, 2 GB).

Conclusion

This paper presents the method of vehicle video detection. The method is based on video partition into the images blocks and the following reconstruction of locations for the intermediate frames. Our method achieves above 70% true positive detection rate (TPR), and gives only 0.3 false positives per frame (FPF). Experimental results demonstrate the proposed approach is promising. Note that the comparison with the other methods is not clear due to the absence of free implementations and public traffic video sequences. Future work includes experiments on the detector capability to find partially visible vehicles and the evaluation of tracking. Vehicle classification problem is still to be investigated.

References

1. J. Arrospide et al. Robust Vehicle detection through multi-dimensional classification for on broad video based systems // In Proc. of the IEEE Int. Conf. on Image Processing. – 2008. – P. 2008-2011.
2. C. Arth, F. Limberger, H. Bischof. Real-Time License Plate Recognition on an Embedded DSP-Platform // In Proc. of the CVPR. – 2007. – P. 1-8.
3. H. Bay et al. SURF: speed up robust features // CVIU. – 2008. – Vol.110, No.3. – P. 346-359.
4. N. Buch, J. Orwell, S.A. Velastin. A Review of Computer Vision Techniques for the Analysis of Urban Traffic // IEEE Trans. on intelligent transp. systems. – Vol. 12, No.3. – 2011. – P. 920-939.
5. R. Cucchiara et al. Statistical and knowledge based moving object detection in traffic scene // In Proc. of the IEEE Int. Conf. on Intel. Transp. Syst. – 2000. – P. 27-32.
6. R. Cucchiara et al. Improving Shadow Suppression in Moving Object Detection with HSV Color Information // In Proc. of the IEEE Int. Conf. on the Intel. Transp. Syst. – 2001. – P. 334-339.
7. P. Druzhkov et al. On some new object detection features in OpenCV library // PRIA: Advances in Mathematical Theory and Applications. – 2011. – Vol.21, No.3. – P. 384-386.
8. P.F. Felzenszwalb et al. Object Detection with Discriminatively Trained Part Based Models // IEEE Trans. on Pattern Analysis and Machine Intelligence. – 2010. – V.32, No.9. – P. 1627–1645.
9. P.F. Felzenszwalb et al. Cascade object detection with deformable path model // In Proc. of the CVPR. – 2010. – P. 2241-2248.
10. G.S.K. Fung et al. Towards Detection of Moving Cast Shadows for Visual Traffic Surveillance // Systems, Man, and Cybernetics. – 2001. – Vol. 4. – P. 2505-2510.
11. S. Gauglitz, T. Holerer, M. Turk. Evaluation of Interest Point Detectors and Feature Descriptors for Visual Tracking [http://cs.iupui.edu/~tuceryan/pdf-repository/Gauglitz2011.pdf].
12. M. Kafai, B. Bhanu. Dynamic Bayesian Networks for Vehicle Classification in Video // IEEE Trans. on Ind. Inf. – Vol. 8, No.1. – 2012. – P. 100-109.
13. Z.W. Kim, J. Malik. Fast Vehicle Detection with Probabilistic Feature Grouping and its Application to Vehicle Tracking // In Proc. of the CVPR. – 2003. – Vol.1. – P. 524-531.
14. Macroscop systems [http://macroscop.com].
15. OpenCV Library [http://opencv.org].
16. PASCAL Visual Object Challenge 2007 [http://pascallin.ecs.soton.ac.uk/challenges/voc].
17. J. Ponce, D.A. Forsyth. Computer vision. A modern approach. – 2004. – 465 p.
18. R. Rad, M. Jamzad. Real time classification and tracking of multiple vehicles in highways // J. Pattern Rec. Letters. – 2005. – Vol. 26, Issue 10. – P. 1597-1607.
19. A. Sanin, C. Sanderson, B.C. Lovell. Shadow Detection: A Survey and Comparative Evaluation of Recent Methods // Pattern Recognition. – 2012. – Vol. 45, No.4. – P. 1684–1695.
20. S. Sivaraman, M. Trivedi. A General Active-Learning Framework for On-Road Vehicle Recognition and Tracking //IEEE Trans. on Intel. Transp. Syst. – 2010. – Vol.11, No.2. – P. 267-276.
21. X. Song, R. Netavia. A Model-based Vehicle Segmentation Method for Tracking // In Proc. of the ICCV. – 2005. – Vol.2. – P. 1124-1131.
22. Source code [http://www.itlab.unn.ru/file.php?id=735].
23. Y.M. Tsai et al. An intelligent vision-based vehicle detection and tracking system for automotive applications // In Proc. of the IEEE Int. Conf. on Consumer Electronics. – 2011. – P. 113-114.
24. VOCORD Traffic [http://www.vocord.ru/218].
25. N. Zolotykh, V. Kustikova, I. Meyerov. A review of vehicle detection and tracking methods in video // N. Novgorod: Vestnik of UNN. – No.5(2). – 2012. – P. 347-357 (In Russian).

OBJECT CLASSIFICATION VIA ENSEMBLE OF MULTICHANNEL IMAGES USING TREE-STRUCTURED REPRESENTATIONS[1]

M.M. Lange[2,3], D.Y. Stepanov[2]

[2] Dorodnicyn Computing Centre, Russian Academy of Sciences, Moscow, Russian Federation, [3]lange_mm@ccas.ru

A metric classifier of composite objects given by collections of patterns is suggested. The classifier is constructed in a space of the tree-structured object representations using a kernel-based decision rule. An efficiency of the classifier is shown by experimental estimations of recognition error rates for the objects given by couples of color face and grayscale signature. A computational advantage factor of a hierarchical decision algorithm as against an exhaustive decision algorithm is evaluated.

Introduction

In many applications, a solution of object classification problem leads to estimating error rate and computational complexity of a decision algorithm. This approach is relevant especially in the case when the submitted objects are given by large data levels and a number of classes is sufficiently great. An efficient solution of the problem can be achieved using multiresolution object representations by trees or pyramids [1-3]. Multilevel structures of these representations are suitable for using hierarchical decision algorithms that provide a decrease of a computational complexity with respect to an exhaustive decision algorithm.

In [4,5], tree-structured pattern representations are used for efficient classification of objects given by one-channel images. A metric classifier based on a multilevel network of templates is developed and a guided search algorithm for decision templates is suggested. Given number of classes c, a computational complexity of this algorithm is $O(c \log c)$ **as** against $O(c^2)$ for an exhaustive search.

In this paper, we suggest a generalization of the above classifier for the objects given by multichannel images as well as by an ensemble of the image sources.

Statement of the Task

Let composite objects be generated by M independent sources of images and a sequence of sets

$$\mathbf{A}= (\mathbf{A}_1,..., \mathbf{A}_s,..., \mathbf{A}_M) \qquad (1)$$

is an ensemble in which $\mathbf{A}_s$ is a set of objects of the s-th source. In (1), any composite object $A \in \mathbf{A}$ is given by a sequence $A=(A_1,..., A_s,..., A_M)$ where $A_s \in \mathbf{A}_s$. Also, each set $\mathbf{A}_s$ contains the objects of c+1 classes so that $\mathbf{A}_s = \{\mathbf{A}_{si}\}_{i=0}^{c}$, where each class $\mathbf{A}_{si}$ of number $i > 0$ includes the objects of semantically the same type and the zero class $\mathbf{A}_{s0}$ combines all other objects. So, the composite objects in (1) belonging to c positive classes and to the zero class are called own and alien, respectively. There are given a priory probabilities P_{own} and $P_{alien} = 1 - P_{own}$.

Any object $A_s = \{A_{sq}\}_{q=1}^{N_q} \in \mathbf{A}_s$ is given by a collection of patterns extracted in N_s-channel image (one pattern per each channel) and the object A_s is described by a collection of N_s representations as follows

$$A_s \to A_s^L = \{A_{sq}^L = (a_{sq}^0,\ldots,a_{sq}^l,\ldots,a_{sq}^L)\}_{q=1}^{N_s}. \qquad (2)$$

In (2), the each representation A_{sq}^L is given by a perfect binary tree of order L in which a_{sq}^l is the l-th level of the tree and $A_{sq}^l = (a_{sq}^0,\ldots,a_{sq}^l)$ is a subtree of order l in A^L [4]. As the representation units are the nodes of the binary tree and a number of the nodes of the l-th level is $\| a_{sq}^l \| = 2^l$, the collection (2) yields the multiresolution representation of the object A_s. So, a representation of any composite object $A = (A_1,..., A_s,..., A_M)$ in the ensemble (1) is

[1] This work is supported by the RFBR, project 12-01-00920-a

given by the corresponding collection of M representations of the form (2).

Fig.1 shows an example of the collection of trees of order L=8 for a composite object that consists of one channel grayscale signature (N_{sign}=1) and three channel color face (N_{face}=3). The nodes of the trees are elliptic primitives taken in principle axes of the pattern segments.

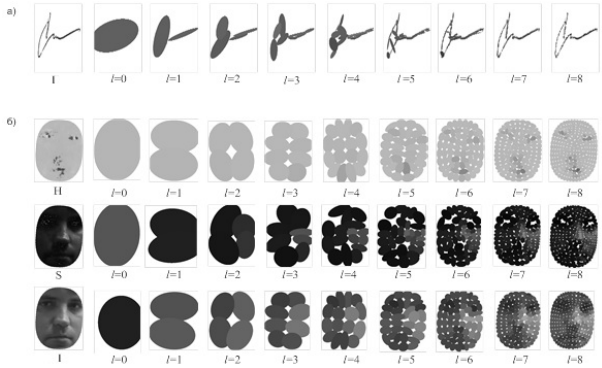


Fig. 1. Tree-structured representation of a composite object given by I-signature (a) and HSI-face (b).

In [5], there was defined a dissimilarity measure $d^l_{sq}(A,\hat{A}) \geq 0$ of order $l=1,...,L$ for any pair of patterns $(A_{sq},\hat{A}_{sq})$ represented by subtrees $(A^l_{sq},\hat{A}^l_{sq})$ in (2). This measure allows us to define dissimilarity measures

$$d^l_s(A,\hat{A}) = \sum_{q=1}^{N_s} \gamma_{sq} d^l_{sq}(A,\hat{A}), \quad l=1,...,L, \tag{3}$$

$$d^l(A,\hat{A}) = \sum_{s=1}^{M} \eta_s d^l_s(A,\hat{A}), \quad l=1,...,L, \tag{4}$$

for pairs of objects represented by $(A^l_s,\hat{A}^l_s)$ and $(A^l,\hat{A}^l)$, respectively, where the normalized weights $\gamma_{sq} \geq 0$ and $\eta_s \geq 0$ are estimated while learning.

Let a collection of clusters

$$\mathbf{B}_s = \{\mathbf{B}_{si} \subset \mathbf{A}_{si}\}_{i=1}^{c} \subset \mathbf{A}_s, \ s=1,...,M,$$

be a learning set of objects of the s-th source. Then a collection

$$\mathbf{B}=\{\mathbf{B}_i = (\mathbf{B}_{1i},...,\mathbf{B}_{si},...,\mathbf{B}_{Mi})\}_{i=1}^{c} \subset \mathbf{A} \tag{5}$$

yields a learning set of the composite objects taken from the ensemble (1). In each cluster $\mathbf{B}_{si} \in \mathbf{B}_s$, a collection of templates $\hat{\mathbf{B}}_{si}$ is selected so that $||\hat{\mathbf{B}}_{si}||=\hat{m}_{si} \leq ||\mathbf{B}_{si}||=m_{si}$. The templates of each source are combined into collections

$$\hat{\mathbf{B}}_s = \{\hat{\mathbf{B}}_{si} \subset \mathbf{B}_{si}\}_{i=1}^{c} \subset \mathbf{B}_s \ , s=1,...,M,$$

that yield a collection of composite templates

$$\hat{\mathbf{B}}=\{\hat{\mathbf{B}}_i = (\hat{\mathbf{B}}_{1i},...,\hat{\mathbf{B}}_{si},...,\hat{\mathbf{B}}_{Mi})\}_{i=1}^{c} \subset \mathbf{B}. \tag{6}$$

Using the measure (4) and the collection of templates (6), we define discriminant functions

$$g^l(A\,|\,\hat{\mathbf{B}}_i) = \sum_{\hat{B}\in\hat{\mathbf{B}}_i} q^l(\hat{B})\, K_{d^l}(A,\hat{B}), \ i=1,...,c. \tag{7}$$

of order $l=1,\ldots,L$. Here, $q^l(\hat{B}) \geq 0$, $\sum_{\hat{B}\in\hat{\mathbf{B}}_i} q^l(\hat{B}) = 1$, is a weight of the composite template $\hat{B}=(\hat{B}_1,...,\hat{B}_M)$ represented by $\hat{B}^l=(\hat{B}^l_1...,\hat{B}^l_M)$, and $K_{d^l}(A,\hat{B})$ is a kernel which values belong to [0,1] and decrease as $d^l(A,\hat{B})$ increases. Given composite object $A \in \mathbf{A}$, the functions (7) of maximum order L yield a decision rule as follows

$$i^*=[\max_{i=1}^{c} g^L(A\,|\,\hat{\mathbf{B}}_i) > 0] \arg\max_{i=1}^{c} g^L(A\,|\,\hat{\mathbf{B}}_i), \tag{8}$$

where $[f]$ is an indicator of f. If $g^L(A\,|\,\hat{\mathbf{B}}_i)=0$, i =1,..,c, the decision (8) yields the zero class of alien objects.

The decision (8) can be made either by an *Exhaustive Search* (ES) algorithm in the space of object representations of order L or by a *Guided Search* (GS) algorithm in all spaces of the object representations of order $l=1,\ldots,L$.

Learning the classifier consists in selecting the collection of the composite templates (6) as well as in estimating the weights in the measures (3) and (4). Testing the classifier consists in carrying out experiments on person recognition via the ensemble of color faces and grayscale signatures.

Learning the Classifier

Selection of the templates. The learning procedure is based on constructing Tree-Structured Covers (TSCs) of the clusters by spheres of a form

$$S^L(\hat{B},D^L(\hat{B}))=\{B\,|\,d^L(B,\hat{B}) \leq D^L(\hat{B})\} \tag{9}$$

where $\hat{B}$ and $D^L(\hat{B})$ are the sphere center and radius, respectively, and $d^L(B,\hat{B})$ is the dissimilaraty measure (4) of maximum order L. Given cluster $\mathbf{B}_i$, the cover is constructed by dichotomous partitioning $\mathbf{B}_i$ into nonintersecting

fragments $\{\mathbf{B}_{ij}\}_{j=1}^{\hat{m}_i}$ of cardinalities m_{ij} so that $m_i = \sum_{j=1}^{\hat{m}_i} m_{ij}$, and by calculating the spheres (9) for these fragments. For each $\mathbf{B}_{ij}$, the center and the radius of the sphere are defined as follows

$$\hat{B}_{ij} = \arg\min_{\hat{B}\in\mathbf{B}_{ij}} \max_{B\in\mathbf{B}_{ij}} d^L(B,\hat{B}) \qquad (10)$$

$$D^L(\hat{B}_{ij}) = (1-\alpha_i)D^L_{\min}(\hat{B}_{ij}) + \alpha_i D^L_{\max}(\hat{B}_{ij}). \quad (11)$$

Here, $\alpha_i \in [0,1]$ is a parameter; $D^L_{\min}(\hat{B}_{ij})$ is a value of $d^L(B,\hat{B})$ between the center $\hat{B}_{ij} \in \mathbf{B}_{ij}$ and the most dissimilar object $B \in \mathbf{B}_{ij}$; $D^L_{\max}(\hat{B}_{ij})$ is a value of $d^L(B,\hat{B})$ between $\hat{B}_{ij} \in \mathbf{B}_{ij}$ and the least dissimilar object $B \in \mathbf{B} \setminus \mathbf{B}_i$ subject to

$$D^L_{\max}(\hat{B}_{ij}) > D^L_{\min}(\hat{B}_{ij}).$$

Fig. 2 shows an example of a binary tree-like structure of the cover (a) and a scheme of setting the covering spheres for a pair of adjacent fragments (b).

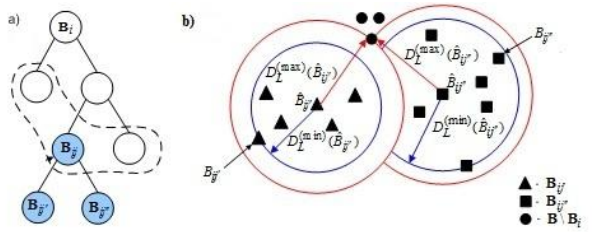


Fig. 2. TSC of the cluster $\mathbf{B}_i$ by spheres: a) binary tree of fragments; b) covering spheres of adjacent fragments.

At each dichotomous step, the partition $\mathbf{B}_{ij} \to (\mathbf{B}_{ij'}, \mathbf{B}_{ij''})$ is made for a leaf node $\mathbf{B}_{ij}$ of a maximum dispersion $\max_{B\in\mathbf{B}_{ij}} d_L(B,\hat{B}_{ij})$ with respect to the center (10). The fragments $(\mathbf{B}_{ij'}, \mathbf{B}_{ij''})$ include the objects that are the nearest by the measure $d^L(B,\hat{B})$ with respect to the most dissimilar supporting objects $(B_{ij'}, B_{ij''})$ in $\mathbf{B}_{ij}$, respectively.

Proposition 1. TSC of $\mathbf{B}_i$ yields collections $\hat{\mathbf{B}}_i^L = \{\hat{B}_{ij}^L\}_{j=1}^{\hat{m}_i}$ and $\mathbf{D}^L(\hat{\mathbf{B}}_i) = \{D^L(\hat{B}_{ij})\}_{j=1}^{\hat{m}_i}$ of order L. TSCs of c clusters generate a sequence of pairs $\{\hat{\mathbf{B}}_i^l, \mathbf{D}^l(\hat{\mathbf{B}}_i)\}_{i=1}^c$, $l = 1,\ldots,L$, giving TSC classifier by the decision rule (8).

Corollary 1. Each pair $(\hat{\mathbf{B}}_i^l, \mathbf{D}^l(\hat{\mathbf{B}}_i))$ yields the weights $q^l(\hat{B}_{ij}) = m_{ij}/m_i$ and the kernel in (7) as follows

$$K_{d^l}(A,\hat{B}_{ij}) = 2^{\frac{-d^l(A,\hat{B}_{ij})}{D^l(\hat{B}_{ij})}},\ l = 1,2,\ldots,L-1,$$

$$K_{d^L}(A,\hat{B}_{ij}) = 2^{\frac{-d^L(A,\hat{B}_{ij})}{D^L(\hat{B}_{ij})}}\ [d^L(A,\hat{B}_{ij}) \le D^L(\hat{B}_{ij})].$$

Estimations of parameters. For $\{(\hat{\mathbf{B}}_i^L, \mathbf{D}^L(\hat{\mathbf{B}}_i))\}_{i=1}^c$, estimations of the parameters $\{\hat{m}_i, \alpha_i\}_{i=1}^c$ are obtained using a leave-one-out cross validation procedure [6].

Proposition 2. Given TSC classifier, an estimation of error rate over the learning set (5) is defined by a mean value

$$\varepsilon_{\mathbf{B}} = \frac{1}{c}\sum_{i=1}^{c} \varepsilon_{\mathbf{B}}^{cv}(\hat{m}_i,\alpha_i),$$

where $\varepsilon_{\mathbf{B}}^{cv}(\hat{m}_i,\alpha_i) = \varepsilon_{\mathbf{B}\setminus\mathbf{B}_i}^{cv}(\hat{m}_i,\alpha_i)P_{alien} + \left(\frac{(c-1)}{c}\varepsilon_{\mathbf{B}\setminus\mathbf{B}_i}^{cv}(\hat{m}_i,\alpha_i) + \varepsilon_{\mathbf{B}_i}^{cv}(\hat{m}_i,\alpha_i)\right)P_{own}$

is a cross validation error rate function, and $\varepsilon_{\mathbf{B}_i}^{cv}(\hat{m}_i,\alpha_i)$ and $\varepsilon_{\mathbf{B}\setminus\mathbf{B}_i}^{cv}(\hat{m}_i,\alpha_i)$ correspond to FRR and FAR for the unit classifier given by the pair $(\hat{\mathbf{B}}_i^L, \mathbf{D}^L(\hat{\mathbf{B}}_i))$.

Corollary 2. Minimum of $\varepsilon_{\mathbf{B}}$ is achieved by $(\hat{m}_i,\alpha_i) = \arg\min_{1\le\hat{m}_i'\le m_i;\ 0\le\alpha_i'\le 1} \varepsilon_{\mathbf{B}}^{cv}(\hat{m}_i',\alpha_i'),\ i = 1,\ldots,c.$

Clearly, the TSCs that are formulated in the above propositions and corollaries can be constructed using the measures (3) and (4) of order L. We obtain cross validation error rates $\{\varepsilon_{\mathbf{B}_{sq}}\}_{q=1}^{N_s}, s = 1,\ldots,M$, and $\{\varepsilon_{\mathbf{B}_s}\}_{s=1}^{M}$ for learning collections $\mathbf{B}_{sq}$ and $\mathbf{B}_s$, respectively. These rates yield estimations $\{\{\gamma_{sq} = \mu_s \log\varepsilon_{\mathbf{B}_{sq}}\}_{q=1}^{N_s}\}_{s=1}^{M}$ and $\{\eta_s = \mu\log\varepsilon_{\mathbf{B}_s}\}_{s=1}^{M}$, where μ_s and μ are the normalizing multipliers.

The TSC classifier of the composite objects provides a fusion recognition scheme based on *General Dissimilarity Measure* (GDM) of the form (4) for a given ensemble of sources. We show an advantage of the GDM fusion scheme as against the known *Weighted Majority Vote* (WMV) fusion scheme [7].

Decision algorithms and complexity

We consider ES algorithm and GS algorithm for the decision (8) using a multilevel network of the templates given by $\{\hat{\mathbf{B}}_i^l, \mathbf{D}^l(\hat{\mathbf{B}}_i)\}_{i=1}^c$, $l = 1,\ldots,L$. A

computational complexity of these algorithms is measured by a number of pairs of compared nodes in the representations (2) while calculating the discriminant functions (7). The ES algorithm uses the last level and calculates $c_L = c$ functions of order L. The GS algorithm uses all levels and calculates $c_l = \left\lfloor c2^{-\beta(l-1)} \right\rfloor$ functions of order l at the l-th level. Here, $\beta > 0$ and c_{l+1} is a number of the largest functions taken from c_l.

Proposition 3. For $\beta = \frac{1}{L}\log_2 c \geq 1$, the computational complexities of ES and GS algorithms satisfy the estimations

$$C^{\mathrm{ES}} = 2c(2^L - 1)\sum_{s=1}^{M} N_s \hat{m}_s^{\mathrm{mean}} = O(c^2),$$

$$C^{\mathrm{GS}} \leq 2cL\sum_{s=1}^{M} N_s \hat{m}_s^{\max} = O(c\log_2 c),$$

where $\hat{m}_s^{\mathrm{mean}} = c^{-1}\sum_{i=1}^{c} \hat{m}_{si}$, $\hat{m}_s^{\max} = \max_{i=1}^{c} \hat{m}_{si}$.

Corollary 3. For $c \to \infty$, the estimations in proposition 3 provide a computational advantage factor $C^{\mathrm{ES}}/C^{\mathrm{GS}} \simeq c/\log_2 c$.

Experimental Results

The developed methodology were used for classification of the composite objects given by couples of a grayscale signature and a colour face in HSI model. The signatures were taken from a competition database [8], the images of faces were made by the authors.

Source	P_{own}=1		P_{own}=0.75	P_{own}=0.50	P_{own}=0.25
	SVM	TSC	TSC	TSC	TSC
I^{Sign}	2E-4	0.005	0.006	0.010	0.015
H^{Face}	0.002	0.008	0.010	0.012	0.015
S^{Face}	0.009	0.012	0.015	0.017	0.019
I^{Face}	0.017	0.012	0.014	0.020	0.025
HSI^{Face}_{GDM}	0.001	0.001	0.005	0.007	0.010
HSI^{Face}_{WMV}	0.006	0.005	0.007	0.009	0.012
$(I^{Sign}, HSI^{Face})_{GDM}$	7E-5	3E-4	0.001	0.003	0.005

Fig. 3. Mean error rates of signature and face classification by ES algorithm

Source		P_{own}=1	P_{own}=0.75	P_{own}=0.5	P_{own}=0.25
		TSC	TSC	TSC	TSC
I^{Sign}	ε	0.015	0.020	0.025	0.030
	θ	5.362	5.975	6.274	8.366
HSI^{Face}_{GDM}	ε	0.012	0.018	0.020	0.025
	θ	5.228	8.366	8.496	9.803
$(I^{Sign}, HSI^{Face})_{GDM}$	ε	0.010	0.014	0.018	0.021
	θ	5.400	8.184	8.366	9.653

Fig. 4. Mean error rate ε and factor $\theta = C^{\mathrm{ES}}/C^{\mathrm{GS}}$ for classification by GS algorithm

The experiments were carried out by a scheme of 30 times 2 fold cross validation [6] for P_{own}=1; 0,75; 0,50; 0,25. The experimental results were obtained for the suggested TSC-classifier and for the known SVM classifier [9]. The tables of the results are shown in Fig.3 and Fig.4 for ES algorithm and GS algorithm, respectively. Error rates of the SVM classifier are given for $P_{own} = 1$ only as there is no the reject function in the applied SVM version. In Fig.3, the error rates demonstrate an advantage of GDM fusion scheme with respect to WMV scheme for face classification. Fig.4 shows the factors of computational advantage of GS algorithm as against ES algorithm.

Conclusion

In this paper, we suggest the new metric classifier of 2D objects given by multichannel images. The classifier is constructed in the space of tree-structured object representations that allow us to apply the fast hierarchical decision algorithm instead of the exhaustive search algorithm. Given ensemble of sources, we suggest the new fusion classification scheme based on the General Weighted Measure in a set of the object representations. The suggested fusion scheme showed the less error rates in comparison with the known Weighted Majority Vote fusion scheme.

References

1. Rosenfeld A. Quadtrees and Pyramids for Pattern Recognition and Image Analysis. Proc. 5th Int. Conference on Pattern Recognition. Miami Beach, 1980, pp. 802–811.
2. Torsello A. Matching Hierarchical Structures for Shape Recognition // PhD thesis. The University of York, 2004.
3. Elfiky N. M., Khan F. S., Weijer J., Gonzalez J. Discriminative Compact Pyramids for Object and Scene Recognition // Pattern Recognition, 2012, vol. 45, no. 4, pp. 1627 - 1636.
4. Ganebnykh S.N., Lange M.M. Classification of 2D Grayscale Objects in a Space of Multiresolution Representations // Pattern Recognition and Image Analysis, 2009, 19 (4), 591-602.
5. Ganebnykh S.N., Lange M. M., Stepanov D. Y. Metric Classifier Using Multilevel Network of Templates // Pattern Recognition and Image Analysis, 2012, 22 (2), 265-277.
6. Theodoridis S., Koutroumbas K. Pattern Recognition // Elsevier, 2009.
7. Kuncheva L.I.. Combining Pattern Classifiers // Wiley and Sons, New Jersey, 2004.
8. www.cs.ust.hk/svc2004: International Signature Verification Competition, 2004.
9. Open Source Computer Vision Library. Available from opencv.willowgarage.com.

RIDGE AND TREE FEATURE DETECTION ON IMAGES[1]

A. Levashov [2,3], D. Yurin[2,4]

[2] Laboratory of Mathematical Methods of Image Processing
Faculty of Computational Mathematics and Cybernetics
Lomonosov Moscow State University
Leninskie Gory, Moscow 119991, Russia
[3]alexeylevashov89@gmail.com , [4]yurin@cs.msu.ru

In this article a new ridge detector with improved non-maxima suppression procedure in scale-space is proposed. Using the results of scale space ridge detection as primary markup the lines are prolonged up to branch points and restored in places where they are not detected. In our method the adaptive choice of threshold and flood fill are used. The algorithm is tested on synthetic, landscape and medical images.

Introduction

There are several major algorithms for ridge detection. Apparently the first of them was the approach [2] that does not use scale-space analysis and therefore its disadvantage is that it detects only thin lines. The approach [5] is based on the scale-space analysis, but it represents the theoretical basis for it and not practically effective algorithm. For the last 15 years there have been only a few cases where it was used in scientific researches. The detector [3] based on generalization of NMS for the case of 3d scale space (x, y, t) can been seen as more effective for ridge detection. However one of this approach`s disadvantages is that it does not detect the branch points on the lines and in the adjacent areas. It interrupts the lines connection and significantly complicates the tree structure analysis on images. The examples of such tasks are masking of tree branches and wires on the non-professional photos for retouching, road and river net finding on aerospace images for the cartographic information updating, selection of blood veins on medical images and processing of 3D images of computer tomography for medical diagnostic. The branch points are essential for such tasks, but often the lines are not detected in the adjacent areas for several reasons. First, there is an ambiguity in choosing the direction for NMS procedure, because the eigenvalues are almost equal. Second, in the adjacent areas two or more ridges lie closely and, as a result, Laplace operator can give a maximal response on bigger scale (two times or more which leads to lines breakdown). Third, the filter response can diminish due to inconsistency between actual image structure and ridge mathematical model, which implies: the solid color line in the range of central lobe of the second derivative of Gaussian function and background with another color on the sidelobes. This article is devoted to suppression of disadvantages described above.

The ridge detection

In this section we shortly recall the algorithm [3], let $I(x,y)$ is the intensity of input grayscale image. Scale space [6] is defined as convolution of the image $I(x,y)$ with Gaussian kernel:

$$L(x,y,t) = G(x,y,t) * I(x,y),$$
$$G(x,y,t) = \frac{1}{2\pi t} e^{-\frac{x^2+y^2}{2t}}, \quad t = \sigma^2, \tag{1}$$

We denote smoothed image $L(x,y,t)$ derivatives with subscripts as $L_x, L_y, L_{xx}, L_{xy}, L_{yy}, \ldots$. Such derivatives are equivalent to original image $I(x,y)$ convolution with appropriate derivatives of Gauss function $G(x,y,t)$. By direct differentiation it is easy to check [6] that scale-space representation of image $L(x,y,t)$ is satisfied to diffusion equation:

[1] The research was supported by the RFBR grant 13-07-00584.

$$L_t = \frac{1}{2}\left(L_{xx} + L_{yy}\right), \quad (2)$$

where derivatives are computed at scale t.

Consider the algorithm [3] for ridge detection with branch points missing. Let the thickness of lines is fixed and equal 2σ, then filter for ridge lines finding can be the Laplace operator or maximal eigenvalue of Hessian $\mathbf{H}_{22}$ in point (x, y)

$$\mathbf{H}_{22} = \begin{pmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{pmatrix}, \quad (3)$$

where the derivatives are computed on scale $t = \sigma^2$. In each point (x, y) we have a pair of eigenvectors: $(\bar{\mathbf{v}}_1, \bar{\mathbf{v}}_2)$, where $\bar{\mathbf{v}}_1$ corresponds to eigenvalue λ_1 such that $|\lambda_1| \geq |\lambda_2|$ and $\bar{\mathbf{v}}_1$ is normal to ridge. On this basis we obtain the algorithm of ridge lines detection as local maxima finding task (we find directional local maxima and minima of values λ_1 in direction $\bar{\mathbf{v}}_1$), i.e. we use the Non Maxima Supression (NMS) procedure with 3x3 window from Canny edge detector. If the detection scale t does not equal to thickness of line, the response of Laplacian (which is equal to $\Delta L = L_{xx} + L_{yy} = \lambda_1 + \lambda_2$) or eigen value $|\lambda_1|$ is less than that of for scale is equal to thickness. Then if we analyze the response on neighborhoods scales we can choose the optimal scale. In scale-space the 3D Hessian is:

$$\mathbf{H}_{33} = \begin{pmatrix} L_{xx} & L_{xy} & L_{xt} \\ L_{xy} & L_{yy} & L_{yt} \\ L_{xt} & L_{yt} & L_{tt} \end{pmatrix}, \text{ where using (2)}$$

$$L_{xt} = \frac{1}{2}\left(L_{xxx} + L_{xyy}\right), \; L_{yt} = \frac{1}{2}\left(L_{xxy} + L_{yyy}\right), \quad (4)$$

$$L_{tt} = \frac{1}{4}\left(L_{xxxx} + 2L_{xxyy} + L_{yyyy}\right).$$

In [3] the NMS procedure was modified as follows: the suppression is performed simultaneously in 2 directions of 3D scale space (x,y,t) using 3x3x3 window. The directions are selected as projections of 2 eigenvectors corresponding to 2 maximal eigenvalues of (4) on the plane normal to eigenvector $\bar{\mathbf{v}}_2$ of 2D Hessian (3). The discrete step on scales $\sigma_0, ..., \sigma_i, ..., \sigma_n$ in [3] are chosen as in [7]: $\sigma_0 = 1, \sigma_{i+1} = \sigma_i \sqrt[4]{2}$. For 3D NMS procedure implementation it is sufficient to hold in memory only 3 layers corresponding to 3 sequential scales. Processing such layers triplets sequentially under σ_i increasing, the local maximum points are collected.

False detection suppression

The method [3] outlined above detects also such features as blobs and some edges. In our approach the final results are presented in vector form instead of bitmap. After detection of ridges their spine pixel chains are collected in lists and the lists are united in a graph. For excluding blobs, the lists with small count of pixels are discarded. It is important that the minimal count of pixels in list is proportional to the ridge (or thick line) width. That is we impose a constraint for minimum elongation of the blobs, which are considered as ridges. This approach is differ from [3] where separation of ridges and blobs is based on ratio of eigen values of (3).

For suppression of edges in [3] in NMS procedure the points are discarded if next dot product is negative:

$$\left(\vec{\mathbf{g}}\left(\vec{\mathbf{x}} + \vec{\mathbf{v}}_1\right), \vec{\mathbf{g}}\left(\vec{\mathbf{x}} - \vec{\mathbf{v}}_1\right)\right) < 0. \quad (5)$$

Our experiments show that condition (5) can lead to erroneous discarding true ridge centers for example in cases when different side's backgrounds relatively to ridge have a different average intensity. We propose the modification of this condition:

$$\left(\vec{\mathbf{g}}\left(\vec{\mathbf{x}} + \sigma_i \vec{\mathbf{v}}_1\right), \vec{\mathbf{g}}\left(\vec{\mathbf{x}} - \sigma_i \vec{\mathbf{v}}_1\right)\right) < T_g \quad (6)$$

where σ_i corresponds to current scale, the threshold constant satisfied the inequality:

$$-1 < T_g < 0, \quad (7)$$

$\vec{\mathbf{g}} = (L_x, L_y)^T$ is the image intensity gradient.

Adaptive flood-fill

After the previous step we obtains the set of detected ridges represented as set of points $S = \{(\bar{\mathbf{x}}_j, \sigma_j, l_j)\}$, where $\bar{\mathbf{x}}_j$ – the coordinates of center points of ridge lines, σ_j – the scale, at which we detect ridge point with

coordinates $\bar{\mathbf{x}}_j$, l_j – the response value in this point (the response of Laplacian convolution computed on scale σ_j). Using this data, we build 3 images $M, I_{\min}, I_{\max}$. We set $M(\bar{\mathbf{x}}) = 1$ if pixel belongs to ridges, and $M(\bar{\mathbf{x}}) = 0$ elsewhere. $I_{\min}, I_{\max}$ – additional images showing the available ridges points intensity range. Let initially in all points $M(\bar{\mathbf{x}}) = 0$. Then for each points $\bar{\mathbf{x}}_j \in S$ we set

$$M(\bar{\mathbf{x}}_j) = 1, \quad I_{\max}(\bar{\mathbf{x}}_j) = I(\bar{\mathbf{x}}_j) + 0.5|l_j|, \quad I_{\min}(\bar{\mathbf{x}}_j) = I(\bar{\mathbf{x}}_j) - 0.5|l_j|. \tag{8}$$

As shown in [3], the Laplacian value on scale corresponding to ridge thickness is equal to difference of average intensities of background and of ridge.

Then we make flood fill of image M. The proposed algorithm is similar to [1], but we use additional conditions for pixel filling. Consider two neighborhood pixels $\bar{\mathbf{y}}_k$ and $\bar{\mathbf{y}}_j$ - $M(\bar{\mathbf{y}}_k) = 1, M(\bar{\mathbf{y}}_j) = 0$. If the condition

$$I_{\min}(\bar{\mathbf{y}}_k) < I(\bar{\mathbf{y}}_j) < I_{\max}(\bar{\mathbf{y}}_k), \tag{9}$$

is satisfied, we modify the pixel $\bar{\mathbf{y}}_j$ as:

$$M(\bar{\mathbf{y}}_j) = 1, \quad I_{\min}(\bar{\mathbf{y}}_j) = I_{\min}(\bar{\mathbf{y}}_k), \quad I_{\max}(\bar{\mathbf{y}}_j) = I_{\max}(\bar{\mathbf{y}}_k). \tag{10}$$

Now we can describe the algorithm of adaptive flood fill:

algorithm AdaptiveFloodFill($\bar{\mathbf{x}}, \bar{\mathbf{d}}$)

1. **if** $M(\bar{\mathbf{x}}) = 1$ return
2. **if** $NOT\left(I_{\min}(\bar{\mathbf{x}}) < I(\bar{\mathbf{x}} + \bar{\mathbf{d}}) < I_{\max}(\bar{\mathbf{x}})\right)$ return
3. $M(\bar{\mathbf{x}} + \bar{\mathbf{d}}) = 1$
   $I_{\min}(\bar{\mathbf{x}} + \bar{\mathbf{d}}) = I_{\min}(\bar{\mathbf{x}})$
   $I_{\max}(\bar{\mathbf{x}} + \bar{\mathbf{d}}) = I_{\max}(\bar{\mathbf{x}})$
4. AdaptiveFloodFill ($\bar{\mathbf{x}} + \bar{\mathbf{d}}, (0 \quad 1)^T$)
5. AdaptiveFloodFill ($\bar{\mathbf{x}} + \bar{\mathbf{d}}, (1 \quad 0)^T$)
6. AdaptiveFloodFill ($\bar{\mathbf{x}} + \bar{\mathbf{d}}, (0 \quad -1)^T$)
7. AdaptiveFloodFill ($\bar{\mathbf{x}} + \bar{\mathbf{d}}, (-1 \quad 0)^T$)

In the beginning of the algorithm for each points $\bar{\mathbf{x}}_i : M(\bar{\mathbf{x}}_i) \neq 0$ and for each neighborhoods $\bar{\mathbf{x}}_i + \bar{\mathbf{d}}$ we invoke the algorithm AdaptiveFloodFill($\bar{\mathbf{x}}_i, \bar{\mathbf{d}}$), which differs from [1] by 2 and 3 lines. For near border pixels we additionally check when coordinates $\bar{\mathbf{x}} + \bar{\mathbf{d}}$ are outside of image rectangle. This is the scheme of simple and not optimal realization of the flood filling algorithm. The faster algorithm is described, for example, in [1], and can easy be modified for our problem by including additional conditions for pixel filling (9), (10).

Ridge skeletization

The binary image M after the previous step will segment the image into the pixels belonging to the ridges and the background pixels. When at branch points the ridges change their color slightly in comparison with the background, the junction points will be added to the ridges, even if they were not found by the scale-space detector. Further the continuous skeletization of binary image can be applied and the skeleton can be used as the result. We use the algorithm [9], because in our approach it is desirable to retain the ridge points found by scale-space detector and only add junction and junction neighborhood ridge points as a result of thinning. The algorithm of skeletization via line thinning is based on step-by-step removing of the points which are not the centers in lines under additional condition that it is not ridge point obtained with scale-space detector. Each algorithm iteration includes two passes through all the pixels. On each pass the non-skeleton points are marked but removed only after the end of the pass. Around each point $\bar{\mathbf{x}}$ of the image M the window 3x3 is selected. We can enumerate pixels $P_1, P_2, ..., P_9$ is the window in order shown on the Fig.1. Descriptors $B(\bar{\mathbf{x}}) = \sum_{i=2}^{9} P_i$ and $A(\bar{\mathbf{x}})$ which is equal to the number of patterns 01 in the sequence $P_2, ..., P_9, P_2$ are calculated.

P_9 $(i-1, j-1)$	P_2 $(i-1, j)$	P_3 $(i-1, j+1)$
P_8 $(i, j-1)$	P_1 (i, j)	P_4 $(i, j+1)$
P_7 $(i+1, j-1)$	P_6 $(i+1, j)$	P_5 $(i+1, j+1)$

Fig. 1: pixel`s order in 3x3 window

On the first pass the pixel $\bar{\mathbf{x}}$ is marked to be removed from image M if all four next conditions are satisfied:

a) $2 \le B(\bar{\mathbf{x}}) \le 6$

b) $A(\bar{\mathbf{x}}) = 1$

c) $M(P_2) \cdot M(P_4) \cdot M(P_6) = 1$

d) $M(P_4) \cdot M(P_6) \cdot M(P_8) = 1$

On the second pass clauses c) and d) are changed to the following:

c') $M(P_2) \cdot M(P_4) \cdot M(P_8) = 1$

d') $M(P_2) \cdot M(P_6) \cdot M(P_8) = 1$

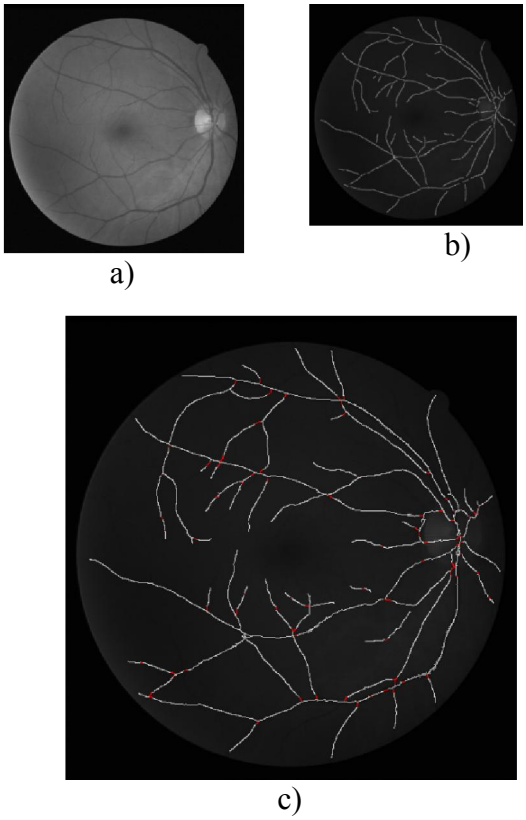
The iterations are being done until no pixels are removed during two passes. As we don't remove pixels detected by the scale-space algorithm, all the lines found initially retains and only the junction points between them are appended.

Results and conclusion

The ridge detection algorithm without brakes near branch points is proposed. We make some improvements in algorithm [3] and restore breaks in ridges near branch points with adaptive flood fill algorithm. We test our approach on synthetic and medical images. Some results for medical retina images are shown on Fig. 2. The ridges in the Fig. 2b have brakes near branch points while in the Fig. 2c most of these brakes are restored and shown in red. One of the disadvantages of the proposed algorithm is that our method does not detect the accurate ridge edges. But after we restore the ridges interconnections, the problem simplifies to finding edges between object (ridges united in tree) and the background. The task becomes bipolar labeling problem and can be solved via graph minimal cut between source (the set of connected ridge central points) and sink (background, or the points lying far from ridges longer than the ridge width).

References

1. S.V. Burtsev, Y.P. Kuzmin. "An efficient flood-filling algorithm" // Computers & graphics, Elsevier, -V. 17, -No 5, -P. 549–561, 1993
2. R. Haralick, "Ridges and Valleys on Digital Images" // Computer Vision, Graphics, and Image Proc., -V. 22, -No. 10, 1983
3. N. A. Khanina, E. V. Semeikina, D. V. Yurin. "Scale-space color blob and ridge detection" // Pattern Recognition and Image Analysis, -No. 1, -P. 221−227, 2012
4. M. Klaiber, L. Rockstroh, Z. Wang, Y. Baroud, S. Simon. "A Memory-Efficient Parallel Single Pass Architecture for Connected Component Labeling of Streamed Images" // Field-Programmable Technology, Int. Conf, -P. 159−165, 2012
5. T. Lindeberg, "Edge detection and ridge detection with automatic scale selection" // International Journal of Computer Vision, -V. 30, -No. 2, -P. 117–154, 1998
6. T. Lindeberg "Scale-Space Theory in Computer Vision" // Kluwer Academic Publishers/Springer, Dordrecht, Netherlands, 1994.
7. D.G. Lowe. "Distinctive image features from scale-invariant keypoints" // Int. Journal of Computer Vision, -V. 60, -No. 2, -P. 91−110, 2004
8. M.B. Dillencourt, H. Samet, M. Tamminen (1992). "A general approach to connected-component labeling for arbitrary image representations" // Journal of the ACM, -V. 39, -N. 2, -P. 253−280, 1992
9. T.Y. Zhang, C.Y. Suen "A fast parallel algorithm for thinning digital patterns" // Communications of the ACM, -V. 27, -No. 3, -P. 236−239, 1984.



a) b) c)

Fig. 2: a) Input image, b) scale-space ridge detection, c) the result of line connection

EFFECTIVENESS ANALYSIS OF SOME OBJECT IDENTIFICATION METHODS FOR BINARY IMAGES[1]

R.G. Magdeev[2], A.G. Tashlinskii[2]

[2] Ulyanovsk State Technical University, ul. Severnyi Venets 32, Ul'yanovsk, 432027 Russia, telephone (88422)778102, e-mail: tag@ulstu.ru

A comparative analysis of computational cost and incorrect identification probability of some based on comparison with a standard object identification methods is performed. Correlation-extreme algorithm, contour analysis method and stochastic gradient identification are considered as such methods. Investigated methods use directly images or object's features (contours) on these images.

Problem formulation

Video processing, remote sensing, processing of data from vehicle's cameras etc. are fields where object identification can be implemented. Although there are many publications the mentioned problem is still topical. One of the reasons is the requirement of real-time data processing.

In real situations objects' images are considerably different depending on the positions of light sources, the angle of observation, parameters of acquisition system. Change in observation angle leads to a change of an object's form. Obviously we can't store standard images for every acquisition conditions because a consideration of just the angle of observation, positions of light sources and focal length requires more than one hundred standard images of one object. If we consider k_i values for i-th parameter α_i then a use of L parameters $\overline{\alpha}$ requires $\prod_{i=1}^{l} k_i$, standards, which is unacceptably for real-time systems when k_i is large.

A problem of object identification is always reduced to seek for a spatial transformation minimizing distance between tested image Z and standard image Z_S in a given metric space. If a model of possible image deformations $T(\overline{\alpha})$ and objective function Q is given then parameters $\overline{a}*$ leading to objective function's extremum should be found [1 - 6]. Objective function Q is a similarity measure of Z and Z_S. It defines the metric space.

Object identification methods usually include some kind of image preprocessing, e.g. image segmentation allowing distinguishing an object and finding its location as the first approximation. Besides, image binarisation is usually used to simplify primitive objects identification [1, 3, 4]. Binary images are considered in this work.

In this work it is assumed that target object has similar geometrical shape on images. But they can be differ by scale factor k, rotation angle φ, translations h_x and h_y among the axis Ox and Oy. Moreover, images can have additive noise. In this situation, the object's image can be obtained up to noise from the standard image using similarity model [1].

A comparative analysis of popular object identification methods is described below. These popular methods are correlation-extreme algorithm (CEA), contour analysis method (CAM) and stochastic gradient identification (SGI). Investigated methods use directly images or object's features (contours) on these images. The comparison is performed by computational cost and a probability of incorrect object identification.

Computational cost of the methods

Let's estimate methods' computational cost.

The idea of CEA is to compute correlation function between tested image and standard image for every given combination of

[1] Support of the RFBR grant 12-01-97014

transformation parameters. If tasted image contains similar fragment then correlation function has a maximum at this position [2, 4].

The main stages of CEA are evaluation of correlation ratio for each possible object's location (with each standard), search for the maximum ratio and comparison with a threshold providing given correct identification probability.

Computational cost of CEA depends on a range of possible parameters. For $m \times n$ pixels standard image it is approximately

$$S_{CEA} \approx 4k_\kappa k_\varphi k_{hx} k_{hy}(mn+1),$$

where: $k_{hx} = (M-m)/\Delta h$, $k_{hy} = (N-n)/\Delta h$, $k_\kappa = (\kappa_{max} - \kappa_{\min})/\Delta\kappa$ and $k_\varphi = (\varphi_{max} - \varphi_{\min})/\Delta\varphi$ - the number of standards for parameters h_x, h_y, κ and φ respectively; $\varphi_{max(\min)}$ and $\kappa_{max(\min)}$ - the maximum (minimum) rotation angle and scale ratio; $\Delta\kappa$, $\Delta\varphi$ and Δh - steps of parameters. If object's orientation is not bounded, we obtain the following

$$S_{CEA} \approx \frac{8\pi(\kappa_{max} - \kappa_{\min})(M-m)(N-n)(nm+1)}{(\Delta h)^2 \Delta\kappa\Delta\varphi}. \quad (1)$$

Use of CEA allows recognizing objects by their contours. In order to extract information about object's form the contour is considered as closed contour vector [2, 3]. The length of this contour (the number w of its basic components) [3] is normalized. After that, normalized correlation ratio between obtained contour vector and vector obtained from standard by circular shift of its components is computed. If correlation ratio oversteps the threshold then the object is identified.

Assuming a use of the Canny edge detector [3] with noise reduction based on Gaussian filter [1] and Fast Fourier Transform (FFT), and Sobel operator [2] to estimate the intensity gradient, let's estimate the computational cost of CEA to detect object boundaries. The main stages of CEA [3] and their number of required basic operations are given in table 1.

The computational cost of CAM:

$$S_{\text{CAM}} \approx 2MN(\log(MN)+15) + 16(n+m) + 6w^2 + 4w \quad (2)$$

Table 1. Computational cost of CEA

CEA stages	The number of operations
Noise reduction	$2NM\log(NM)$
Gradient estimation	$12NM$
Non-maximum suppression	$8NM$
Double threshold filtering	$2NM$
Tracing edges through the image and hysteresis thresholding	$8NM$
Representing contours as vectors	$16(n+m)$
Normalization of the contour's length, where w – the contour's length (the number of basic vectors)	$4 \cdot w$
Normalized correlation function calculation	$6 \cdot w^2$

In SGI algorithm identification parameters $\overline{\alpha}$ are searched for recurrently with standard location keeping unchanged [5]:

$$\hat{\overline{\alpha}}_t = \hat{\overline{\alpha}}_{t-1} - \mathbf{\Lambda}_t \overline{\beta}_t,$$

where $\overline{\beta}_t$ - gradient estimation of Q depending on $\hat{\overline{\alpha}}_{t-1}$ and the number of iteration $t = \overline{0,T}$; $\mathbf{\Lambda}_t$ - the learning rate matrix [5]. For identification correlation ratio is a common choice for Q. Effective range of the method is limited and it defines the number of standards. Fig. 1 shows standards (rectangular, elliptical and triangular) and estimates of objective function $\hat{Q}$ obtained using SGI on the number of iterations for each standard in case of rectangular object on the tested image. One can see that for the case of rectangular standard Q estimates converge to minimum within 600 iterations. For the other standards this estimates are stochastic with rather similar variance.

In order to increase convergence of estimations $\overline{a}$ and to increase the effective range of SGI it is reasonable to implement low-pass filtering, e.g. using Gaussian filter. As mentioned above, it requires about $2NM\log(NM)$ basic operations. Computational cost of SGI by itself is considered in [6]. It depends on a way of gradient estimation. If correlation ratio is used as objective function, the computational cost is approximately $(51\mu+91)T$ up to $(69\mu+48)T$ basic operations, where μ - the size of local sample used at each iteration, T -

the number of iterations. Therefore, computational cost of SGI is in average equal to:

$$S_{\text{SGI}} \approx 2MN(\log(MN)+15)+(60\mu+70)T \quad (3)$$

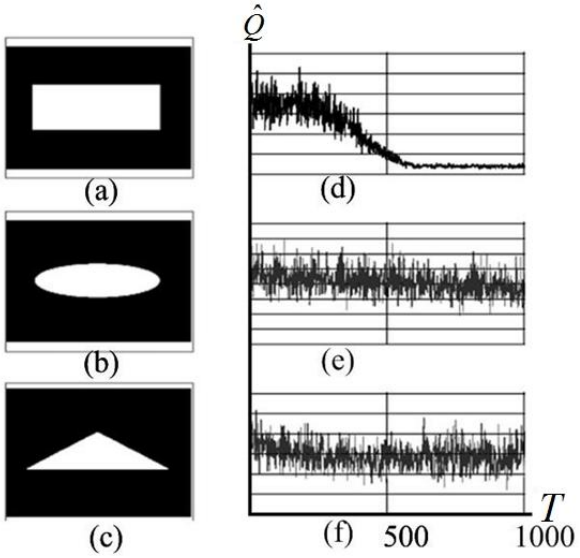


Fig. 1. Objective function estimation using SGI.

Fig. 2 shows the dependence of computational cost of tested algorithms on the size of image ($M = N$) with the same object size 128x128 pixels. Curve 1 represents CEA with $\kappa_{max} = 1.4$, $\kappa_{\min} = 0.6$, $\Delta h = 2$, $\Delta\kappa = 0.2$, $\Delta\varphi = 0.05$; curve 2 – CAM with $t = 32$; curve 3 – SGI with $\mu = 20$, $T = 2000$.

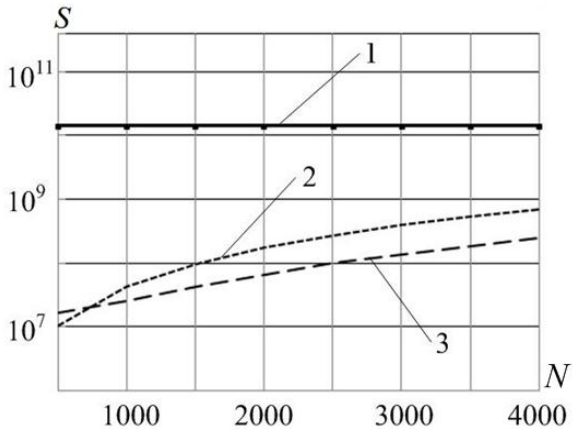


Fig. 2. The dependence of computational cost of the methods on the image size with the same object size.

One can see that when image has less than $5\cdot10^5$ pixels, CAM has the least cost. If it is more than $5\cdot10^5$ pixels, SGI has the least cost. Computational cost of CAE is 100 times higher and it depends quadratically on the image size (it is slightly reflected at the picture).

,Fig. 3 shows the dependence of computational cost on the object's size ($n = m$) when the image size is constant (1024x1024 pixels). The parameters of methods were equal to mentioned above. One can see that computational cost CAM and SGI are weakly dependent on the object's size. For CAE this dependence is approximately quadratic. SGI requires the lowest cost and CAE – the highest: when $n = 200$ they differ by 100 times, when $n = 700$ - by 1000 times.

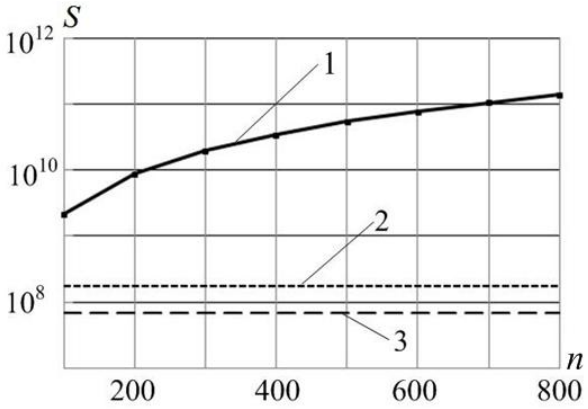


Fig. 3. The dependence of computational cost of the methods on the object's size with the same image size.

An experiment has shown that when $M = N = 512$, $m = n = 256$ after 200 implementations the average evaluation time of CAE is equal to 14 minutes, CAM – 0.6 seconds, SGI – 0.9 seconds. Let's notice that three initial approximations were used for SGI because the effective range of this method for used number of iteration is about $\pm 60^0$. A calculation of computational cost for the same conditions gives the following results: $S_{\text{CEA}} \approx 1{,}4\cdot10^{10}$, $S_{\text{CAM}} \approx 1{,}1\cdot10^{7}$, $S_{\text{SGI}} = 1{,}7\cdot10^{7}$. It is confirmed with experimental results.

The probability of incorrect identification

The probability of incorrect identification P_{er} was determined experimentally. And the influence of additive Gaussian noise with signal-to-noise ratio from 1 to 10 and misalignment between standard and tested objects (which is crucial for SGI) were examined.

Fig. 4 shows curves of dependence of P_{er}

on signal-to-noise ratio. CEA has the best noise stability due to the large sample size. Due to high computational cost relatively large steps of identification parameters were used ($\Delta h = 2$, $\Delta\kappa = 0.2$, $\Delta\varphi = 0.05$). This led to some incorrect identification. SGI has also shown good noise stability. With weak noise ($q < 8$) it has lower P_{er}. This can be explained by high accuracy of parameters' identification. Due to errors in contour detection CAM showed several times worse results for each q.

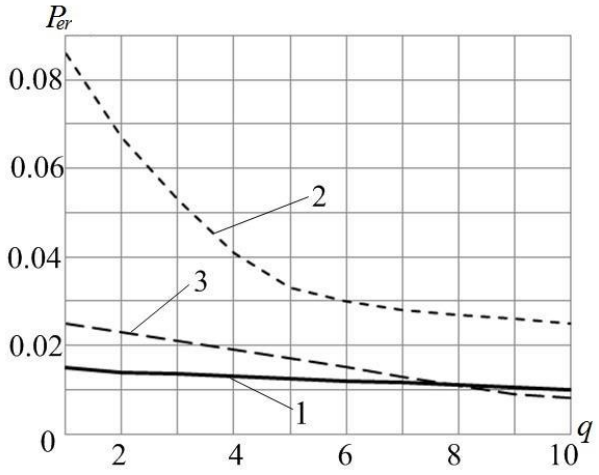


Fig. 4. The dependence of incorrect identification probability on signal-to-noise ratio.

Fig. 5 shows the incorrect identification probability dependence on misalignment of standard object and tested object when $q = 10$. One can see that this parameter is crucial only for SGI due to low effective range. For SGI in this case P_{er} increases by 4 times with misalignment distance increasing from 0 to 40.

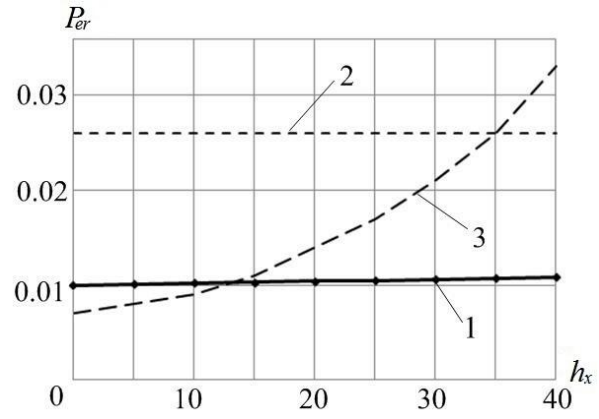


Fig. 5. The dependence of incorrect identification probability on misalignment of standard and tested objects

Conclusion

The comparative analysis showed that the computational cost of examined methods is dependent on the image size. CAM has the lowest cost with small images, with big – SGI. CEA has quadratic dependence of the cost on the image size and it is 100 times higher.

The computational cost of CAM and SGI is slightly dependent on the object size, but for CEA it also has a quadratic form. SGI requires the lowest computational cost, CEA – the highest.

Due to the largest sample size CEA has the highest noise stability. Due to relatively large steps used for identification parameters ($\Delta h = 2$, $\Delta\kappa = 0.2$, $\Delta\varphi = 0.05$) CEA has some incorrect identification. SGI has also shown good noise stability. But the probability of the correct identification for this method is dependent on misalignment between tested and standard objects. The probability of incorrect identification for CAM is several times higher is case of noise due to errors in contour detection.

Thus, SGI is better choice among examined methods for continuous data processing due to high performance and good noise stability.

References

1. Gonzalez, R.C., Woods, R.E. Digital Image Processing, Prentice Hall. - New Jersey, 2002.
2. Gruzman I.S., Kirichuk V.S. et al. Digital image processing in information systems. – Novosibirsk: NGTU publishing, 2002. (in Russian).
3. Introduction into contour analysis / Edited by. J.A. Furman – M.: FIZMALIT, 2003 – 592p. (in Russian).
4. Pratt U. Digital image processing. – M.: Mir, 1982, V.1. (in Russian).
5. Tashlinskii A. Computational Expenditure Reduction in Pseudo-Gradient Image Parameter Estimation / Computational Science – ICCS 2003, V. 2658. Proceeding, Part II. Berlin: Springer, 2003, p. 456-462.
6. Tashlinskii A.G., Smirnov P.V. and Zhukov S.S. Analysis of methods of estimating objective function gradient during recurrent measurements of image parameters / Pattern Recognition and Image Analysis, 2012, Vol. 22, No. 2, pp. 393–399.

ANALYSIS OF THE GRADIENT DESCENT METHOD IN PROBLEMS OF SIGNAL RESTORATION[1]

A. Makovetskii[2,4], V. Kober [2,3,5]

[2] Department of Mathematics, Chelyabinsk State University, Chelyabinsk, Russia
[4]artemmac@csu.ru
[3] Department of Computer Science, CICESE, Ensenada, B.C. 22860, Mexico
[5]vkober@cicese.mx

In the work, the problem of restoration of signals blurred by a linear operator is considered. An image as a function of two variables can be well described with two variations. One of them is a metrical characteristic of continuously differentiable functions. Another one is called a linear variation. The linear variation is a topological characteristic of a function of two variables. We carry out a detailed analysis of the gradient descent method for signal restoration. Restoration errors of the method can be overcome by using linear variations. Computer simulation illustrates restoration results by the gradient descent method.

Introduction

Numerous papers on image restoration consider the restoration of functions distorted by a linear operator using different variations of the gradient descent method [1-4]. In this paper, analysis of the gradient descent method for signal restoration is performed in order to find shortcomings of the method and to improve its performance using linear variations of a function.

Let $\boldsymbol{u}$ be an original image, $\mathbf{A}$ be a distorting compact operator, and $\boldsymbol{u}_0$ be a degraded image, that is, $\boldsymbol{u}_0 = \mathbf{A}\,\boldsymbol{u}$. Consider the restoration problem of the original image $\boldsymbol{u}$ using known $\mathbf{A}$ and $\boldsymbol{u}_0$. Denote by $\mathrm{J}(\boldsymbol{u})$ the functional:

$$J(\boldsymbol{u}) = \| \mathbf{A}\boldsymbol{u} - \boldsymbol{u}_0 \|_{L_2}^2 . \qquad (1)$$

To restore the function $\boldsymbol{u}$, the following variation problem is stated:

$$\boldsymbol{u} = \arg min \| \mathbf{A}\boldsymbol{u} - \boldsymbol{u}_0 \|_{L_2}^2 , \quad \boldsymbol{u} \in BV(\Omega) \qquad (2)$$

where Ω is the domain of the function $\boldsymbol{u}$ in $\mathbb{R}^2$. A common way of solving the problem in (2) is to use the gradient descent method [5]. The iterations of the gradient descent method are given as

$$\boldsymbol{u}_{k+1} = \boldsymbol{u}_k - \alpha_k \cdot J'(\boldsymbol{u}_k) . \qquad (3)$$

Here $J'(\boldsymbol{u})$ is the gradient of the functional $J(\boldsymbol{u})$. In the functional space L_2 the gradient is given by

$$J'(\boldsymbol{u}_k) = 2\mathbf{A}^*(\mathbf{A}\boldsymbol{u} - \boldsymbol{u}_0) , \qquad (4)$$

where $\mathbf{A}^*$ is an adjoin operator. α_k is a parameter of the gradient method step. This parameter is selected in such a way to satisfy the following condition:

$$J(\boldsymbol{u}_{k+1}) < J(\boldsymbol{u}_k) . \qquad (5)$$

The norm is a metrical characteristic of a function of two variables. Continuous functions of two variables also have a set of topological characteristics called linear variations. Kronrod [6] introduced the notion of a regular component of the level set of a continuous function of two variables. The simplest topological characteristic in the linear variation theory is a number of regular components for all level sets of a function. Full information about linear variations of a

[1] This work is supported by RFBR grant № 13-01-00735

function is contained in the one-dimensional tree of a function of two variables.

Linear variation for continuous and discrete functions

Let $\Phi_u(t)$ be the number of regular components of a level set t for a continuous function [6]. The first Kronrod's linear variation is defined as

$$V(u) = \int_{-\infty}^{+\infty} \Phi_u(t)\, dt \ . \qquad (6)$$

Let $\boldsymbol{w}$ be a binary discrete function $\boldsymbol{w} = (w_{i,j})$, where $w_{i,j} \in \{0,1\}$, for all pairs (i,j). A subset of such pairs (i,j) when $w_{i,j} = 1$ and all elements of the subset are connected by the 8-connectivity, is called the connected component of the binary function $\boldsymbol{w}$. For a number $k \in \mathbb{N}$ and a discrete function $\boldsymbol{u}$ we define the following indicator function :

$$\chi_k(u_{i\,j}) = \begin{cases} 1, u_{i\,j} \geq k \\ 0, u_{i\,j} < k \end{cases}. \qquad (7)$$

Definition. The number $\mathrm{V}_k(u_{i,j})$ of connected components for a level k, $k \in \mathbb{N}$ of the discrete function $\boldsymbol{u}$ is called the number of connected components of the binary discrete function $\chi_k(u_{i,j})$.

Definition. The linear variation $\mathrm{V}(u_{i,j})$ of a discrete function $\boldsymbol{u}$ is defined as follows:

$$V(u_{i\,j}) = \sum_{k=0}^{+\infty} V_k(u_{i\,j}) \ . \qquad (8)$$

Let us compute the discrete gradient $\nabla u_{i\,j}$ of $\boldsymbol{u}$ at (i,j) as

$$\nabla u_{i\,j} = (\, u_{i+1\,j} - u_{i\,j}, u_{i\,j+1} - u_{i\,j}\,) \ . \qquad (9)$$

Suppose that if the pair (i,j) is outside of the domain of the function $\boldsymbol{u}$, then $u_{i\,j} = 0$.

Application of gradient descent method for continuous functions

Let us consider a linear operator $\mathbf{A}$ of a centered horizontal blurring in the functional space $L_2(\Omega)$ that degrades a function $\boldsymbol{v} \in L_2(\Omega)$:

$$\mathbf{A}v(x) = \frac{1}{2\Delta}\int_{x-\Delta}^{x+\Delta} v(t)\, dt \ , \qquad (10)$$

where x is an arbitrary point of the set Ω.

Remark. The operator $\mathbf{A}$ is a self-adjoint operator, that is, $\mathbf{A}^* = \mathbf{A}$.

Suppose that the function $v(x)$ is a parabola $ax^2 + b$, where $a < 0$.

Remark. The function $v(x) = ax^2 + b$ does not belong to of the functional space $L_2(\Omega)$ when $\Omega = \mathbb{R}$. Nevertheless, assuming that $\Omega = [-R, R]$, where R is a sufficiently large number, the performance of the gradient decent method at the origin with a finite number of steps for the parabola defined on the whole axis $\mathbb{R}$ coincides with that for the parabola defined on the interval $\Omega = [-R, R]$.

Consider the behavior of a set of functions $\{A^1 v(x), A^2 v(x), \ldots, A^k v(x)\}$.

Proposition 1.

$$\mathbf{A}^k \boldsymbol{v} = ax^2 + \left(b + \frac{k}{3} a\Delta^2\right). \qquad (11)$$

Let us apply the gradient descent method to $\boldsymbol{u}_0$. Denote by $\boldsymbol{u}_1$, $\boldsymbol{u}_2$,…, $\boldsymbol{u}_k$ functions obtained from the function $\boldsymbol{u}_0$ using an appropriate number of steps of the gradient method. For the given operator $\mathbf{A}$, and the given function $\boldsymbol{v}$ we obtain

$$\begin{aligned} \boldsymbol{u}_{k+1} &= \boldsymbol{u}_k - 2\,\alpha_{k+1}\,\mathbf{A}^*(\,\mathbf{A}\,\boldsymbol{u}_k - \boldsymbol{u}_0\,) \\ &= \boldsymbol{u}_k - 2\,\alpha_{k+1}\,\mathbf{A}^2\boldsymbol{u}_k + 2\,\alpha_{k+1}\,\boldsymbol{A}\boldsymbol{u}_0 \\ &= \boldsymbol{u}_k - 2\,\alpha_{k+1}\,\mathbf{A}^2\boldsymbol{u}_k + 2\,\alpha_{k+1}\,\mathbf{A}^2\boldsymbol{v} \quad . \qquad (12) \end{aligned}$$

where α_k is a parameter of the gradient method at the *k*th step.

Proposition 2.

$$\begin{aligned} \boldsymbol{u}^k = ax^2 + (b + \frac{1}{3}a\Delta^2(\,1 - 2\alpha_1\,)\cdot \\ \cdot(\,1 - 2\alpha_2\,)\ldots\cdot(\,1 - 2\alpha_k\,))\,, \qquad (13) \end{aligned}$$

where α_k is a parameter of the gradient method at the *k*th step.

Corollary.

$$\mathbf{A}\boldsymbol{u}_k = ax^2 + (b + \frac{1}{3}a\Delta^2((\ 1-2\alpha_1\ ) \ldots (\ 1-2\alpha_k\ ) + 1\ ))\,. \qquad (14)$$

Note that the gradient method exactly restores the position of the maximum of the function. In the case of the optimal parameter value $\alpha_i = \frac{1}{2}$, the function maximum is exactly restored too. However, the probability of choosing the optimum value of α_i using (5) in the method equals zero. It can be shown that for small values of the parameters α_i,

$$\lim_{k\to\infty}\left(b + \tfrac{1}{3}\alpha\Delta^2(\ 1-2\alpha_1\ )(\ 1-2\alpha_2\ ) \ldots (\ 1-2\alpha_k\ )\right) = 0. \quad (15)$$

If the number of steps of the gradient method is a finite number, the difference between the values of the maxima of the original and restored functions is not zero. It follows from

$$\mathbf{A}\boldsymbol{u}_k - \boldsymbol{u}_0 = \mathbf{A}\boldsymbol{u}_k - \mathbf{A}^1\boldsymbol{v} = \\ = \tfrac{1}{3}a\Delta^2(\ 1-2\alpha_1\ )\cdot \ldots \cdot (\ 1-2\alpha_k\ )\,. \quad (16)$$

A regular stopping rule of the gradient method depends on the difference value,

$$J(\boldsymbol{u}_{k+1}) - J(\boldsymbol{u}_k). \qquad (17)$$

If the difference is less than a small predefined value ϵ, then the iterative process is stopped. Consequently,

$$\|\ \mathbf{A}\boldsymbol{u}_{k+1} - \boldsymbol{u}_0\ \|^2 - \|\ \mathbf{A}\boldsymbol{u}_k - \boldsymbol{u}_0\ \|^2 = \\ = \left(\frac{1}{3}a\Delta^2(\ 1-2\alpha_1\ )\cdot \ldots \cdot (\ 1-2\alpha_k\ )\right)^2 \\ \cdot (\ {\alpha_{k+1}}^2 - 4\alpha_{k+1})\,. \qquad (18)$$

The difference value between the maxima of the functions $\boldsymbol{u}_{k+1}$ and $\boldsymbol{v}$ is given by

$$\tfrac{1}{3}a\Delta^2(\ 1-2\alpha_1\ )\cdot \ldots \cdot (\ 1-2\alpha_{k+1}\ )\,. \quad (19)$$

Comparing (18) and (19), one can observe that for small values of α_i a saturation of the gradient method in accordance with the criterion (17) occurs when the value of expression (19) is substantially greater than ϵ.

Therefore, for the considered model the gradient method is able to localize local extrema, but often is unable to exactly restore the values of the extrema.

Denote by $w(x,y)$ a function of two variables,

$$w(x,y) = v(x) \times \boldsymbol{I}, \qquad (20)$$

where $\boldsymbol{I}$ is a binary function of region of support (for instance, for selecting an arbitrary segment of the function). Let the function $w(x,y)$ be a parabolic cylinder. For the function of two variables, now we apply the approach described above for the one-dimensional case.

Linear variation $V(\boldsymbol{w})$ for a parabola $v(x)$ defined on the segment $[-R,R]$ can be written as

$$V(\boldsymbol{w}) = v(0) - v(R)\,. \qquad (21)$$

If we know the height of the original parabola, the projection method [7] can be used to a set the linear variation values computed over a function restored by the gradient method. The projection method helps to restore pretty well the values of local function extrema at know positions.

Computer Simulation

Let us consider the following discrete function shown in Fig. 1:

$$v_i = \begin{cases} 0,\ 0 \le\ i \le 471 \\ -\frac{(x-5\ 00)^2}{4} + 210,472 \le\ i \le 528 \\ 0,\ 529 \le\ i \le 1000 \end{cases}. \quad (22)$$

The maximum value equals to $v_{5\ 00} = 210$. The When a linear blurring with $\Delta = 5$ is used, a saturation of the gradient method occurs at the maximum of the restored function with $v_{5\ 00} = 209.36$.

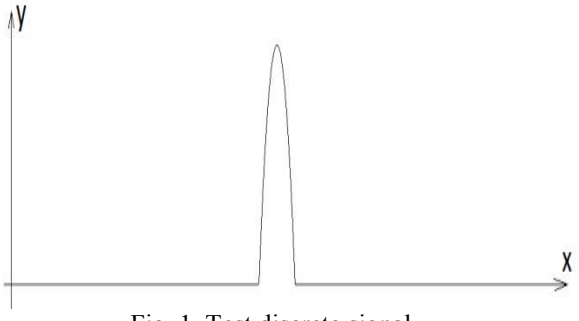


Fig. 1. Test discrete signal.

The restoration of a blurred function with $\Delta = 10$ yields the maximum value of the output function of $v_{5\ 00} = 205.44$. For the latter case, the restored function and difference between the original and restored functions are shown in Figs. 2 and 3, respectively.

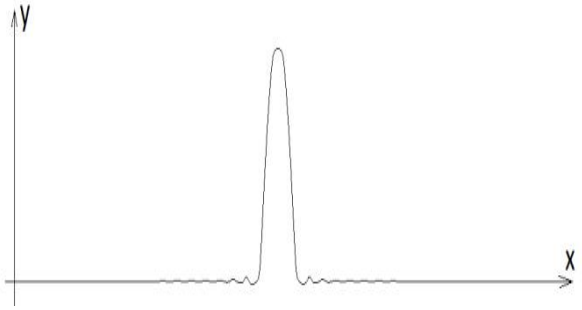


Fig. 2. Restored discrete signal for the case $\Delta = 10$.

Fig. 3. Difference between the original and restored signals for the case $\Delta = 10$.

Conclusion

In this paper we analyzed the performance of the gradient decent method for restoration of signals blurred by a linear operator. Computer simulation illustrated restoration results by the gradient descent method. Restoration errors of the method can be overcome by using linear variations and early proposed projection method.

References

1. L.I. Rudin, S. Osher. Total variation based image restoration with free local constraints // Proc. Int. Conf. Image Processing (ICIP-94). – 1994. – Vol. 1, – P. 31-35.
2. A. Chambolle, P.L. Lions. Image recovery via total variational minimization and related problems //
3. Numer. Math. -1997. – Vol. 76. _ P. 167-188.
4. S. Osher, M. Burger, D. Goldfarb, J. Xu, W. Yin. An iterative regularization method for total variation based image restoration // Multiscale Modelling and Simulation. -2005. – Vol. 4, -P. 460-489.
5. A. Chambolle. An Algorithm for Total Variation Minimization and Applications // Journal of Mathematical Imaging and Vision 20, 2004, P. 89-97.
6. J.A. Snyman. Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms (Springer Publishing, 2005).
7. A. Kronrod. On functions of two variables // Uspehi Mat. Nauk. – 1950. Vol. 1, No. 35, - P. 24-134.
8. A. Makovetskii, V. Kober. Image restoration based on topological properties of functions of two variables // Proc. SPIE Applications of Digital Image Processing XXXV. -2012. – Vol. 8499, -P. 84990A.

A NEW METHOD FOR HIGH-CAPACITY INFORMATION HIDING IN VIDEO ROBUST AGAINST FRAME LOSS[1]

V.A. Mitekin[2]

[2] Image Processing Systems Institute, Russian Academy of Sciences, Russia
443001, Samara, Molodogvardeyskaya st. 151
vmitekin@gmail.com

This paper presents a new method for high-capacity information hiding in digital video. Algorithm proposed in this paper does not require a temporal synchronization to provide a hidden data extraction robust against frame loss. At the same time, proposed randomized distribution of hidden information bits across video frames allows to increase the hiding capacity proportionally to number of frames used. Proposed algorithm is also robust against "watermark estimation" attack aimed at estimation of hidden information without knowing the embedding key. Experimental results are provided to demonstrate declared features of proposed information hiding method.

Introduction

Methods for information hiding in digital video or image sequences, especially methods purposed for copyright protection and traitor tracing applications became a widely discussed topic during last two decades. According to [6] and [8], video watermarking and information hiding algorithms are usually based on "frame-by-frame" embedding approach, which means every frame of the host video is processed independently from other frames. However, as it was shown in [2], there are two potential security issues caused by this approach. First, if the same embedding key K is used to embed watermark into every frame of chosen video, then a simple attack aimed at hidden information extraction is possible. Attacker can analyze (by means of temporal averaging or using methods like an ICA, [1]) a large amount of video frames to find a "static" noise-like image component $D'(n,m)$, which was constantly added into every frame during information hiding. By this means, attacker can recognize the hidden information presence even if embedding key is unknown. Also, attacker can apply the "watermark estimation" attack ([2]) using an estimated $D'(n,m)$ value to extract and remove hidden information from host video without loss of video quality .

Second, if independent keys K_t , $t \in [0, T-1]$ are used in such a manner what information is embedded into t-th frame of host video using key K_t. Usage of independent embedding keys makes it impossible to estimate hidden information using multiple video frames. But in this case, frame counter synchronization between information embedder and detector must be provided to achieve robustness against frame loss. Several algorithms for frame counter synchronization were proposed during last decade. Chen [3] and Sun [4] suggest generating key K_t for information hiding using current frame statistical features, such as frame luminance mean and variance. This algorithm does not use counter information for key generation, so loss of one or several frames after information embedding will not influence information extraction from non-corrupted frames. In other words, algorithms proposed in [3] and [4] use frame contents instead of frame counter for embedding key generation, which makes information extraction totally independent from host video frame order. However, these algorithms have a several disadvantages. First, a number of equal keys will be generated if video contains a series of identical frames. Second, frame-based video processing, such as frame blurring, cropping and filtering, can affects statistical frame features needed for key generation. If these processing operations are introduced after information embedding, then keys generated during information extraction will not match

[1] This work was partially supported by RFBR, grant № 12-07-31056 , grant of Russian Federation president for young scientists, grant № МК-3863.2013.9

original keys and no information will be extracted.

Another approach to key generation for information hiding in video is described in [2], [5],[6]. A small set of frame keys *Kj.* is generated for information embedding , where $j=[1..J]$, $J \ll T$ and T is a total number of video frames. Next, information hiding for the t-th frame of host video is performed using j-th frame key, where $j= (i \bmod J)$. If frame synchronization is lost during hidden information extraction stage, an exhaustive search is performed trying all possible embedding keys to extract information from current desynchronized frame. On the other hand, if attacker knows the rule for K_j selection during embedding, he can choose only the frames with the same embedding key for performing "watermark estimation" attack. In next section this type of attack will be noted as "watermark estimation" attack with frame selection.

Also, frame counter resynchronization method based on recognition of additional temporal or spatial "synchronization patterns" is proposed in [7], [8]. A synchronization pattern, usually an *M*-sequence or Gold sequence, is added to certain parts of video signal to provide reference information (temporal offset, for example) for extraction procedure. This solution provides high robustness against video processing operations (cropping, filtering etc) and non-malicious frame desynchronization due to hardware issues and transmission errors. But, as it was demonstrated in [9], synchronization marks are by himself are subject to a specialized type of attacks called "watermark template attack", so overall data hiding scheme based on synchronization marks embedding has low security against malicious attacks.

In the next section a new method for data hiding robust against malicious and non-malicious frame desynchronization attacks (including frame loss caused by data transmission errors) is proposed, requiring no temporal synchronization at all. At the same time, proposed algorithm uses bit index randomization to split hidden information bits between host video frames, so a higher bit capacity is achieved compared to known algorithms. Experimental results are provided to illustrate declared features of proposed information hiding method.

Proposed information hiding algorithm

Proposed method is aimed at high-capacity information hiding in digital video and multilayer digital images (hyperspectral satellite images for example). As it was declared in [6] and [10], data hiding algorithm is insecure against watermark estimation attacks if both watermark and embedding key are identical for every frame of host video. Existing algorithms use varying key to provide hidden information security against estimation attacks. But there is another way to achieve a variability of noise-like distortions caused by information embedding. A constant frame key can be used for all video frames, but information intended for hiding will be randomly changed for every frame of host video.

Let H be the multi-bit information sequence intended for hiding into digital video, where H_i is the i-th bit of H and $i \in [0, L-1]$. According to proposed method, at the first stage H is partitioned into L independent "slices": $S_0, S_1, S_2, S_3 ... S_L$. Each slice consists of $M=round(\log_2(L)+1)$ bits, which are composed using bit concatenation by the following rule:

$$S_i = n_1 n_2 n_3 ... n_{M-1} H_i, \quad (1)$$

where *round* is a rounding up operation, and $n_1 n_2 n_3 ... n_{M-1}$ is a binary representation of index i. For example, for $L=8$ and $H_7=0$, S_7 will be composed as binary sequence *"1110"*, where *"111"* is a binary representation of *"7"*. Hereafter, first *M-1* bits of "slice" will be referred as an "index section" and the *M*-th bit of slice will be referred as an "informative section"

Next, at the information hiding stage one of slices $S_0, S_1, S_2, S_3 ... S_L$ is selected **randomly** and then embedded into current frame using **any multi-bit image hiding algorithm suitable for multiple bit information hiding in digital images**. The main idea of proposed method is to avoid any pre-determined or key-based order of slices during information embedding. Using the "slice" structure proposed in (1), hidden information can be extracted with no synchronization information at all. The bit sequence extracted from the host

frame at the receiving end can be divided into "informative" and "index" sections, and then *"*index section*"* of extracted slice can be used to determine the original location of "informative" bit in the multi-bit hidden information sequence *H*.

A following implementation of information hiding and extraction method is investigated in this paper, based on 2D spread spectrum modulation.

First**, at the hidden information embedding** stage two different binary 2D orthogonal sequences M_1 and M_2 (which means M_1 and M_2 has near-zero circular cross correlation values) of length $N \times M$ (where $N \times M$ is a frame size) are generated using secret key *K*.

Next, a hidden binary image *W(n,m)* of size $N' \times M'$, where $N' < N, M < M'$ is embedded into the *t*-th frame of digital video using the following rule:

$$I_t'(m,n) = \begin{cases} I_t(m,n) - \alpha \cdot (\hat{S}_{x,y}(M_1) - \hat{S}_{x',y'}(M_2)), \\ \quad if\ W(x-x', y-y') = 1; \\ I_t(m,n) - \alpha \cdot (\hat{S}_{x,y}(M_2) - \hat{S}_{x',y'}(M_1)), \\ \quad if\ W(x-x', y-y') = 0; \end{cases} \quad (2)$$

where $I_t(m,n)$ is the *t*-th frame before information embedding, $\hat{S}_{x,y}$ is a 2D circular shift of M_1 by *x* columns and *y* rows, $I_t'(n,m)$ - the *t*-th video frame after information embedding. Values *x, y, (x-x'), (y-y')* are positive integers selected according to the following restrictions:

$$\begin{array}{l} x \in [0, M-1], \\ y \in [0, N-1], \\ (x-x') \in [0, M'-1], \\ (y-y') \in [0, N'-1] \end{array}, \quad (3)$$

According to (2), each processed frame contains exactly one bit of binary image *W(n,m)*. Bits of *W(n,m)* are embedded into host frame sequence in a random order defined by *x, y, x', y'* values. This means the "index section" is formed using random *(x-x')* and (*y-y')* values which are later embedded into host frame by modulating M_2 and M_2 cyclic shifts.

To extract a hidden image *W'(n,m)* from the host video, fast normalized circular cross correlation ([11],[12]) using all possible cyclic shifts of current frame $I_t'(n,m)$ is computed to detect the presence of M_1 and M_2 sequences embedded into the current frame. Let C_1 be the maximum value of circular cross correlation between $I_t'(n,m)$ and M_1; *p* and *q* are vertical and horizontal shift values of a host frame, needed to produce the maximum value of circular cross correlation). Equally C_2 is the maximum value of circular cross correlation between M_2 and $I_t'(n,m)$, *p'* and *q'* are vertical and horizontal shift of a host frame. Hidden information slice is detected in current frame if the following logical statement *F* is true

$$F = ((C_1 < -T)\ AND\ (C_2 > T))\ OR \\ OR\ ((C_1 > T)\ AND\ (C_2 < -T)), \quad (4)$$

where *T* is a detection threshold $0<T<1$.

Next, to compute the bit indexes *n* and *m* for a current frame, "index section" is extracted using the equation:

$$\begin{array}{l} n = p - p', \\ m = q - q', \end{array} \quad (5)$$

After computing indexes *n* and *m*, "informative" bit must be extracted and is stored as pixel of extracted binary image *W'(n,m):*

$$W'(n,m) = \begin{cases} 1, if\ (C_1 < -T)\ AND\ (C_2 > T) \\ 0, if\ (C_1 > T)\ AND\ (C_2 < -T) \end{cases}, \quad (6)$$

If a number of video frames is much larger than $N \times M$, then it is possible to extract the same "index section" from multiple frames. This means a same bit of hidden information was independently hidden into several frames, providing additional redundancy and robustness against video distortions. In the next section, experimental results are provided to illustrate declared features of proposed information hiding method.

Experimental research

First, we investigated a correlation between noise-like patterns introduced to different video frames during information embeding,. For a set of 1000 frames with 640×480 pixels size, image

$$D_t(n,m) = I_t(m,n) - I_t'(m,n), \quad (7)$$

was computed, representing noise-like patterns introduced during information hiding into *t-th* frame.

Next, maximum circular cross correlation([11],[12]) value $C_{max}(t)$ between D_t and $D_{t+1.}$ was computed using all possible cyclic shifts of D_t.

In an ideal case of counter-dependent information hiding $C_{max}(t)=0$ for any *t*, which means noise patterns for neighboring frames are completely uncorrelated. At the other hand, if time-invariant embedding key is used and

hidden information is repeated for every frame, then $C_{max}(t)=1$ and the whole information hiding method is subject to the "watermark estimation" attack.

During proposed experiment, an averaged value of $C_{\max}(t)$ was estimated using a whole frame set:

$$\overline{C} = \frac{1}{1000}\sum_{t=0}^{999} C_{\max}(t), \qquad (8)$$

resulting in an averaged value $\overline{C} = 0.00002$. At the same experiment, maximal and minimal values of $C_{\max}(t)$ computed at whole frame set were ***0.03*** and ***-0.034***, respectively.

Obtained results allows to claim that the proposed method for information hiding introduced mostly uncorrelated frame noise patterns during information embedding into neighbouring frames.

Next, an experiment for investigation of proposed method robustness against frame loss and pixelwise frame operations was implemented. A hidden binary image $W(n,m)$, $n \in [0,31], m \in [0,31]$, was embedded into video sequence consisting of 5000 frames with 640×480 pixels frame size. Next, after image hiding, simple image processing was applied to every frame of video, combining frame histogram equalization and cropping of 3% frame pixels. Next, a total number of frames was decreased to 4500 using a randomized frame decimation of the host video.

At the detection stage, detection threshold $T=0.03$ was used to detect hidden information. As a result, proposed detection algorithm failed to detect hidden information in *182* frames of *4500*, and additionally *12* frames caused incorrect extraction of an "index section". But none of this errors caused resulting extracted image $W'(n,m)$ distortions, because of a majority voting scheme used for post-processing of extracted bits. According to these results, proposed method allows to extract hidden information even if original frame sequence is corrupted by losing 10% of total frame number.

Conclusion

A new information hiding method for video is proposed in this paper, allowing information detection and extraction without frame counter synchronization. Also, as it was shown by experimental investigation, proposed method does not introduce a fixed or time-recurring noise pattern during information, which makes it robust against known watermark estimation attacks. At the same time, proposed randomized distribution of hidden information bits across video frames allows to increase the hiding capacity proportionally to number of video frames.

References

1. Pavel G. Embedding, Extraction and Detection of Digital Watermark in Spectral Images : diss. – LAPPEENRANTA UNIVERSITY OF TECHNOLOGY, 2005.
2. Doërr G., Dugelay J. L. Security pitfalls of frame-by-frame approaches to video watermarking //Signal Processing, IEEE Transactions on. – 2004. – T. 52. – №. 10. – C. 2955-2964.
3. Chen C., Ni J., Huang J. Temporal statistic based video watermarking scheme robust against geometric attacks and frame dropping //Digital Watermarking. – Springer Berlin Heidelberg, 2009. – C. 81-95.
4. Sun S. W., Chang P. C. Video watermarking synchronization based on profile statistics //Security Technology, 2003. Proceedings. IEEE 37th Annual 2003 International Carnahan Conference on. – IEEE, 2003. – C. 410-413.
5. Holliman M. J., Macy W. W., Yeung M. M. Robust frame-dependent video watermarking //Electronic Imaging. – International Society for Optics and Photonics, 2000. – C. 186-197.
6. Lin E. T., Delp E. J. Temporal synchronization in video watermarking //Signal Processing, IEEE Transactions on. – 2004. – T. 52. – №. 10. – C. 3007-3022.
7. Delannay D., Macq B. Classification of watermarking schemes robust against loss of synchronization //Electronic Imaging 2004. – International Society for Optics and Photonics, 2004. – C. 581-591.
8. Delannay D. Digital watermarking algorithms robust against loss of synchronization : дис. – Dissertation for degree of Ph. D, Universite catholique de Louvain, Belgium, 2004.
9. Herrigel A., Voloshynovskiy S. V., Rytsar Y. B. Watermark template attack //Photonics West 2001-Electronic Imaging. – International Society for Optics and Photonics, 2001. – C. 394-405.
10. Lin E. T., Delp E. J. Temporal synchronization in video watermarking //Signal Processing, IEEE Transactions on. – 2004. – T. 52. – №. 10. – C. 3007-3022.
11. Tsai D. M., Lin C. T. Fast normalized cross correlation for defect detection //Pattern Recognition Letters. – 2003. – T. 24. – №. 15. – C. 2625-2631.
12. O'Ruanaidh J. J. K., Pun T. Rotation, scale and translation invariant digital image watermarking //Image Processing, 1997. Proceedings., International Conference on. – IEEE, 1997. – T. 1. – C. 536-539.

MODEL-BASED GRADIENT FIELD DESCRIPTOR FOR SEARCHING AND RECOGNITION OF IMAGES[1]

V. Myasnikov[2]

[2] Image Processing Systems Institute, Russian Academy of Sciences, Samara, Russian Federation, vmyas@smr.ru

In this paper we propose a new descriptor which is used to describe the digital image – a model-based gradient field descriptor. The derived descriptor characteristics, considered as digital image features, can be used to solve the problems of image analysis, recognition and retrieval effectively. The examples of such tasks solutions using the proposed descriptor are also represented in this paper.

Introduction

Local descriptors computed for specific areas of interest of the analyzed image are a powerful modern tool for solving a wide class of problems in digital image processing [1-6]. Currently, there are many different descriptors, which can be divided into the following categories.

Descriptors based on probability distributions, including SIFT [3], HOG [4] and LESH [1-2] descriptors, which are widely used because of their versatility and quality indicators values.

Descriptors based on spectral-frequency representation. The main way to generate descriptors of this category is to use image decomposition coefficients based on wavelets, Gabor filters, Fourier basis, the Radon transform, etc.

Differential descriptors are calculated as a numeric transform of the partial derivatives values of the brightness function of the analyzed image.

A new image descriptor is proposed in this paper, which is based on differential and probability properties of the analyzed image area. The specificity of this descriptor is a priori given (or predetermined) probability distribution of the gradient field that characterizes the analyzed image model and/or solving problem. Descriptor component values for a specific image (fragments) can be calculated as the values of the probability density of the argument as a specific field gradient, i.e. both are dependent on the instance (a specific image) and the model (the probability distribution). Such descriptor specificity enables to classify it as a model-oriented.

Examples of solving search and image recognition problems using the proposed descriptor are given.

Model-Based Gradient Field Descriptor

Let us consider a differentiable function of brightness $f(t_1,t_2)$, defined on the analysis area of the form $\left\{(t_1,t_2): T_1^{\min} \le t_1 \le T_1^{\max}, T_2^{\min} \le t_2 \le T_2^{\max}\right\}$. The analysis area of a digital image is considered as:

$$D = \left\{ (n_1,n_2): 0 \le n_i \le \frac{T_i^{\max} - T_i^{\min}}{\Delta}, i = \overline{1,2} \right\},$$

where Δ is a sampling interval of the original continuous brightness function.

Under the *digital gradient field* we will mean a function/image of the form:

[1] This research was partly financially supported by:
- RFBR, grants № 13-01-12080-ofi-m, 13-07-12103-ofi-m, 12-07-00021-a;
- the Ministry of Education and Science of the Russian Federation (under the decree of the Government of the Russian Federation of 09.04.2010 № 218);
- Basic Research Program of the Presidium of the Russian Academy of Sciences "Fundamental Problems of Informatics and Information Technologies" project 2.12.

$$g(n_1,n_2)\equiv\nabla f\left(n_1\Delta+T_1^{\min},n_2\Delta+T_1^{\min}\right)=$$
$$=\left.\begin{pmatrix}\partial f/\partial t_1\\ \partial f/\partial t_2\end{pmatrix}\right|_{\substack{t_1=n_1\Delta+T_1^{\min},\\ t_2=n_2\Delta+T_2^{\min}}}\quad (n_1,n_2)\in D$$

Each unit of the gradient field carries information about the *gradient module* $|g(n_1,n_2)|$ and *gradient orientation* $\arg(g(n_1,n_2))$ in relation to the coordinate axes. Let the digital gradient field of the analysis area be $\bar{g}\equiv\{g(n_1,n_2)\}_{(n_1,n_2)\in D}$. And let $\bar{G}$ be a *random digital gradient field* with the probability density $p_{\bar{G}}(\cdot)$, characterizing *observation model* and/or solving problem.

Under *the model-based gradient field descriptor* (MGFD) or the *descriptor handle to the model* $\bar{G}$ we will mean a vector

$$\begin{pmatrix}
p_{\bar{G}}(\bar{g})\\
\min\limits_{|\bar{g}|} p_{\bar{G}}(\bar{g})\\
\max\limits_{|\bar{g}|} p_{\bar{G}}(\bar{g})\\
\min\limits_{\arg\bar{g}} p_{\bar{G}}(\bar{g})\\
\max\limits_{\arg\bar{g}} p_{\bar{G}}(\bar{g})\\
\min\limits_{\arg\bar{g}}\min\limits_{|\bar{g}|} p_{\bar{G}}(\bar{g})\\
\max\limits_{\arg\bar{g}}\min\limits_{|\bar{g}|} p_{\bar{G}}(\bar{g})\\
\max\limits_{|\bar{g}|}\min\limits_{\arg\bar{g}} p_{\bar{G}}(\bar{g})\\
\min\limits_{\arg\bar{g}}\max\limits_{|\bar{g}|} p_{\bar{G}}(\bar{g})\\
\min\limits_{|\bar{g}|}\max\limits_{\arg\bar{g}} p_{\bar{G}}(\bar{g})\\
\max\limits_{|\bar{g}|}\max\limits_{\arg\bar{g}} p_{\bar{G}}(\bar{g})
\end{pmatrix}.$$

The value $\min\limits_{\arg\bar{g}} p_{\bar{G}}(\bar{g})$ is defined as the density $p_{\bar{G}}(\cdot)$ of the gradient field instance where the modulus of the gradient coincides with the modulus of the gradient $|g(n_1,n_2)|$ for the field $\bar{g}$ for each sample, and the phase gradient values $\arg(g(n_1,n_2))$ are selected so as to minimize the density $p_{\bar{G}}(\cdot)$. Other MGFD components are determined similarly.

For the MGFD components the following relationships are valid:

$$\min_{|\bar{g}|} p_{\bar{G}}(\bar{g})\le p_{\bar{G}}(\bar{g})\le\max_{|\bar{g}|} p_{\bar{G}}(\bar{g}),$$
$$\min_{\arg\bar{g}} p_{\bar{G}}(\bar{g})\le p_{\bar{G}}(\bar{g})\le\max_{\arg\bar{g}} p_{\bar{G}}(\bar{g}),$$
$$\min_{|\bar{g}|}\min_{\arg\bar{g}} p_{\bar{G}}(\bar{g})\le\min_{\arg\bar{g}} p_{\bar{G}}(\bar{g})\le\max_{|\bar{g}|}\min_{\arg\bar{g}} p_{\bar{G}}(\bar{g}),$$

etc.

Features of the Model-Based Gradient Field Descriptor

The features of the gradient field are defined as the derived MGFD characteristics. Using inequalities presented above it is convenient to introduce them as the relative position of specific MGFD components between its minimum and maximum values. So you can get seven major features of the form (here $\phi:\mathbf{R}_+\to\mathbf{R}$ is an arbitrary monotonically increasing function):

$$\vartheta_0=\frac{\phi(p_{\bar{G}}(\bar{g}))-\phi\left(\min\limits_{|\bar{g}|} p_{\bar{G}}(\bar{g})\right)}{\phi\left(\max\limits_{|\bar{g}|} p_{\bar{G}}(\bar{g})\right)-\phi\left(\min\limits_{|\bar{g}|} p_{\bar{G}}(\bar{g})\right)},$$
$$\vartheta_1=\frac{\phi(p_{\bar{G}}(\bar{g}))-\phi\left(\min\limits_{\arg\bar{g}} p_{\bar{G}}(\bar{g})\right)}{\phi\left(\max\limits_{\arg\bar{g}} p_{\bar{G}}(\bar{g})\right)-\phi\left(\min\limits_{\arg\bar{g}} p_{\bar{G}}(\bar{g})\right)},$$
$$\vartheta_2=\frac{\phi(p_{\bar{G}}(\bar{g}))-\phi\left(\min\limits_{\arg\bar{g}}\min\limits_{|\bar{g}|} p_{\bar{G}}(\bar{g})\right)}{\phi\left(\max\limits_{|\bar{g}|}\max\limits_{\arg\bar{g}} p_{\bar{G}}(\bar{g})\right)-\phi\left(\min\limits_{\arg\bar{g}}\min\limits_{|\bar{g}|} p_{\bar{G}}(\bar{g})\right)},$$
$$\vartheta_3=\frac{\phi\left(\min\limits_{\arg\bar{g}} p_{\bar{G}}(\bar{g})\right)-\phi\left(\min\limits_{\arg\bar{g}}\min\limits_{|\bar{g}|} p_{\bar{G}}(\bar{g})\right)}{\phi\left(\max\limits_{|\bar{g}|}\min\limits_{\arg\bar{g}} p_{\bar{G}}(\bar{g})\right)-\phi\left(\min\limits_{\arg\bar{g}}\min\limits_{|\bar{g}|} p_{\bar{G}}(\bar{g})\right)},$$

$$\vartheta_4 = \frac{\phi\left(\max\limits_{\arg\bar{g}} p_{\bar{G}}(\bar{g})\right) - \phi\left(\min\limits_{\arg\bar{g}} \max\limits_{|\bar{g}|} p_{\bar{G}}(\bar{g})\right)}{\phi\left(\max\limits_{|\bar{g}|} \max\limits_{\arg\bar{g}} p_{\bar{G}}(\bar{g})\right) - \phi\left(\min\limits_{\arg\bar{g}} \max\limits_{|\bar{g}|} p_{\bar{G}}(\bar{g})\right)}$$

$$\vartheta_5 = \frac{\phi\left(\min\limits_{|\bar{g}|} p_{\bar{G}}(\bar{g})\right) - \phi\left(\min\limits_{\arg\bar{g}} \min\limits_{|\bar{g}|} p_{\bar{G}}(\bar{g})\right)}{\phi\left(\max\limits_{\arg\bar{g}} \min\limits_{|\bar{g}|} p_{\bar{G}}(\bar{g})\right) - \phi\left(\min\limits_{\arg\bar{g}} \min\limits_{|\bar{g}|} p_{\bar{G}}(\bar{g})\right)},$$

$$\vartheta_6 = \frac{\phi\left(\max\limits_{|\bar{g}|} p_{\bar{G}}(\bar{g})\right) - \phi\left(\min\limits_{\arg\bar{g}} \max\limits_{|\bar{g}|} p_{\bar{G}}(\bar{g})\right)}{\phi\left(\max\limits_{\arg\bar{g}} \max\limits_{|\bar{g}|} p_{\bar{G}}(\bar{g})\right) - \phi\left(\min\limits_{\arg\bar{g}} \max\limits_{|\bar{g}|} p_{\bar{G}}(\bar{g})\right)}$$

The values ϑ_i characterize the similarity degree of the analyzed gradient field to potentially possible realizations of the random field $\bar{G}$, which is the model. The following inequalities are faithful:

$$\forall i = \overline{0,6} \quad 0 \le \vartheta_i \le 1,$$

and the higher features values mean higher similarity. The whole set of these features will be referred to below as the vector $\bar{\vartheta} \equiv (\vartheta_0, \vartheta_1, \ldots, \vartheta_6)^T$.

Using Descriptors for Searching and Recognition of Images

The obvious way to build a decision rule using MGFD is to use a conventional discriminant function of the Bayes classifier [7]:

$$d_l(g) = P(\Omega_l) p_{\bar{G}(l)}(g),$$

where $P(\Omega_l), p_{\bar{G}(l)}(g)$ are a priori probability of the class Ω_l and the corresponding to this class value of the probability density for the analyzed gradient field g (the first descriptor component).

More complex decision rules can be designed based on the combinations of individual decisions of "simple" classifiers - *experts* - each of which uses the MGFD features from its analysis sub-region. Thus, the classified fragment is divided into subregions (possibly overlapping) $D_i \left(i = \overline{0, I-1}\right)$ -, and the MGFD description and/or its features $\bar{\vartheta}^i \left(i = \overline{0, I-1}\right)$ are formed for each subregion.

Selecting a classification algorithm of each expert and the method of its setting, the recognition/analysis/research developer may continue to use various combinations of existing strategies of their decisions: the majority vote or the weighted vote, the decisive list, the estimates calculation algorithm, algorithms algebra [8], boosting algorithms [9], and others.

Applications

The proposed MGFD algorithm and the features based on it have been used in the development of two image processing systems.

Example 1. The aim of the first system is detection of passenger vehicles on digital images of optical remote sensing of the Earth. The processed images were obtained by aerial photography with spatial resolution 0.2 m. During construction of this system there have been used different groups of features (based on geometry, texture, structure), but the majority of features (over 80%) were MGFD features. These MGFD features correspond to specific areas of detecting object - a combination of experts' solutions approach was used. The decisive list was performed as a decision rule (to verify the hypothesis of the object existence). The detection system was tested on 30 images of urban areas, not included in the training set. The average number of objects in the images is 35. An example of the proposed method is shown in the figure below. The values of detection quality indicators are:

- correct detection probability – 95 %,
- number of false detected objects (per image) – 2,8.

The detailed description of the detection method that underlies the developed system is presented in

Example 2. The purpose of the second system was the detection of face location in identity paper digital image. The task was complicated by the fact that there was no restrictions on document type, digital image position, rasterization type when printing the images. There was also allowed the presence of glare

and flare as a consequence of paper coating imposition and laminating of uneven lighting elements; commercial printings on human photos; a high difference in front and background contrast (for faces in photos) is permitted, etc. [10].

Fig. 1. Detection of passenger vehicles on digital images of optical remote sensing of the Earth.

The main element of this solution is one MGFD feature - ϑ_2. The correct detection of face position probability for the developed system (assuming a position and size error of 20% and 30%, respectively) was 98%. The experiment was run for 500 identity papers.

Conclusion

A model-based gradient field descriptor for describing digital images was proposed. There were proposed descriptors features, allowing to solve the problems of image analysis, recognition and retrieval efficiently. The examples of real problems solutions showed the effectiveness of the proposed approach to analyze images.

Further investigations of this work include developing automatic procedures for large-format images description synthesis (e.g., space-based remote sensing data of the Earth) using MGFD to solve the problem of searching images in databases.

References

1. C. Schmid, R. Mohr, C. Bauckhage. Evaluation of interest point detectors // Int. Journal of Computer Vision. – 2000. – Vol. 37(2). – P. 151-172.
2. K. Mikolajczyk, C. Schmid. A performance evaluation of local descriptors // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2005. – Vol. 10(27), P. 1615-1630.
3. D. Lowe. Distinctive image features from scale-invariant keypoints // International Journal of Computer Vision. – 2004. - Vol. 2(60). – 91-110.
4. N. Dalal, B. Triggs. Histograms of Oriented Gradients for Human Detection // Proceedings of IEEE Conference Computer Vision and Pattern Recognition, San Diego, USA. – 2005. – P. 886-893.
5. D.A. Forsyth, J. Ponce. Computer Vision: A Modern Approach // New Jersey: Prentice Hall, 2002. - 693 p.
6. L.G.Shapiro, G.C.Stockman. Computer Vision // New Jersey: Prentice Hall, 2001. - 608 p.
7. K. Fukunaga. Introduction to Statistical Pattern Recognition // 2nd ed., New York: Academic Press, Inc. - 1991.- 591 p.
8. J.I. Zhuravlev. An algebraic approach to recognition or classifications problems // Pattern Recognition and Image Analysis. — 1998. — Vol. 8(1) — P. 59–100.
9. P. Viola, M.J. Jones. Rapid Object Detection using a Boosted Cascade of Simple Features // Proceedings IEEE Conf. on Computer Vision and Pattern Recognition. – 2001. – P. 511-518.
10. V.V. Myasnikov Methods of vehicle detection on the aero and space images of Earth remote sensing // Computer Optics. – 2012. – ISSN 0134-2452. (in printing, in Russian)

MOMENT MATCHING ESTIMATION METHOD FOR AN ASYMMETRIC GENERALIZED GAUSSIAN MIXTURE MODEL

N. Nacereddine[1], D. Ziou[2]

[1] Centre de Recherche en Soudage et Contrôle, 16002 Algiers, Algeria, nafaa.nacereddine@enp.edu.dz
[2] DMI, Université de Sherbrooke, Qc., Canada J1K 2R1, djemel.ziou@usherbrooke.ca

In this paper, the r-order moments of the asymmetric generalized Gaussian (AGG) distribution is originally computed. Then, the moment matching method associated to the expectation-maximization (EM) algorithm is used to estimate the AGG mixture model parameters. The obtained results are comparable to those of the maximum likelihood method which, however, manipulates high nonlinear equations (piece-wise function, log, etc.), contrarily to the proposed method where the calculus is less difficult

Introduction

An accurate modeling of unknown probability density functions (*pdf*s) of data, encountered in practical applications, can play an important role in the direction of simulation and design of modern signal processing systems [1].
The AGG distribution has been shown to not only model a wide range of statistical distributions (e.g. impulsive, Laplacian, Gaussian and uniform distributions) but also include the asymmetry [2]. The reason for which an AGG mixture model (AGGMM) can portray successfully a large class of signals (e.g. image gray level histogram, speech data) thanks to the AGG distribution flexibility. In this work, the moment matching estimation (MME) method is combined with the EM algorithm to estimate the parameters of an AGGMM. The resulted are then compared with the maximum likelihood estimation (MLE) method [2]. In fact, the MME method can constitute an alternative to MLE method where, in the latter, the parameter equations are characterized by a high non-linearity due especially to discontinuity and special functions (see [2] for more details).

Asymmetric generalized Gaussian distribution and its mixture

The probability density function *pdf* of an uni-dimensional asymmetric generalized Gaussian distribution (AGGD) is defined as

$$f(x|\vec{\theta}) = \begin{cases} \dfrac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)} e^{-[(-x+\mu)/\alpha_1]^\beta} & \text{if } x < \mu \\ \dfrac{\beta}{(\alpha_1+\alpha_2)\Gamma(1/\beta)} e^{-[(x-\mu)/\alpha_2]^\beta} & \text{if } x \geq \mu \end{cases} \quad (1)$$

for x ($x \in \mathbb{R}$), and where $\vec{\theta} = (\mu, \alpha_1, \alpha_2, \beta)^T$ ($-\infty < \mu < \infty, \alpha_1 > 0, \alpha_2 > 0, \beta > 0$) is a vector whose the components are the pseudo-mean, the left and right scale parameters and the shape parameter, respectively. Γ is the gamma function defined by $\Gamma(\xi) = \int_0^\infty e^{-t} t^{\xi-1} \mathrm{d}t$.
The parameter β dictates the exponential rate of decay as shown in Fig. 1.

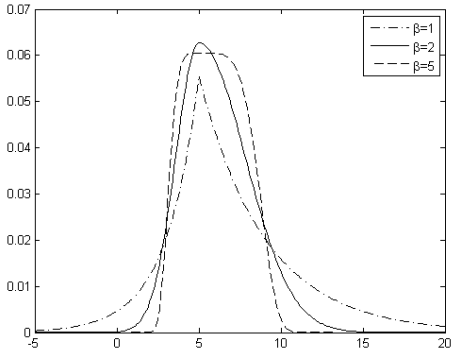


Fig. 1. The *pdf*s of the AGGD with $\mu = 5$, $\alpha_1 = 2$, $\alpha_2 = 4$, $\beta = \{1,2,5\}$

Special cases of AGGD can be obtained with some values of β as follows: asymmetric Laplacian distribution for $\beta = 1$, asymmetric Gaussian distribution for $\beta = 2$, impulse distribution for $\beta \to 0$ and uniform distribution for $\beta \to \infty$. Note that the scale parameters which express the width of the distribution are linked to standard deviations by

$$\alpha_i = \sigma_i \sqrt{\frac{\Gamma(1/\beta)}{\Gamma(3/\beta)}}, \quad i = 1,2 \qquad (2)$$

Let $X = \{x_i\}, i = 1,\ldots,n$ be a set of n realizations of a random 1D vector with an AGG density $f(x_i)$ given in (1). Thus, AGG mixture model (AGGMM) can be written as a weighted sum the M components $C_m (m = 1,\ldots,M)$ following AGG *pdf*

$$f(x_i \mid \vec{\Theta}) = \sum_{m=1}^{M} \pi_m f(x_i \mid \vec{\theta}_m) \qquad (3)$$

where $\vec{\Theta} = (\pi_m, \vec{\theta}_m)$ is a vector of parameters to estimate with π_m the prior probability of the m^{th} component which satisfies : $\pi_m > 0$ and $\sum_{m=1}^{M} \pi_m = 1$. The parameters $\vec{\Theta}$ can be estimated by using EM algorithm. In this work, the gray level histogram $h(g)$ ($g \in [0,255]$) of an image is estimated by a mixture of univariate AGGDs.

Moment Matching Estimation (MME)

Proposition 1. If X is a random variable with the AGG density and, for any $r \in$ IN, the r-order non central moment can be written as

$$EX^r = \frac{1}{(\alpha_1 + \alpha_2)\Gamma(1/\beta)} \times \sum_{k=0}^{r} \binom{r}{k} \left[(-1)^k \alpha_1^{k+1} + \alpha_2^{k+1} \right] \Gamma\left(\frac{k+1}{\beta}\right) \mu^{r-k} \qquad (4)$$

Proof.

$$EX^r = \frac{\beta}{(\alpha_1 + \alpha_2)\Gamma(1/\beta)} \left[\int_{-\infty}^{\mu} x^r e^{-\left(\frac{-x+\mu}{\alpha_1}\right)^{\beta}} \mathrm{d}x + \int_{\mu}^{\infty} x^r e^{-\left(\frac{x-\mu}{\alpha_2}\right)^{\beta}} \mathrm{d}x \right]$$

If we set the following change of variable $y = (-x+\mu)/\alpha_1$ if $x \le \mu$ and $y = (x-\mu)/\alpha_2$ if $x > \mu$, use the power series formula $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$ and the definite integral $\int_0^{\infty} x^{\nu-1} \exp(-\eta x^p) dx = \frac{1}{p} \eta^{-\nu/p} \Gamma(\nu/p)$ function and take the amount $\frac{\beta}{(\alpha_1 + \alpha_2)\Gamma(1/\beta)}$ equal to C, we will obtain

$$EX^r = C\left[(-1)^r \alpha_1^{r+1} \int_0^{\infty} \left(y - \frac{\mu}{\alpha_1} \right)^r e^{-y^\beta} \mathrm{d}y + \alpha_2^{r+1} \int_0^{\infty} \left(y + \frac{\mu}{\alpha_1} \right)^r e^{-y^\beta} \mathrm{d}y \right]$$

$$= C\left[(-1)^r \alpha_1^{r+1} \int_0^{\infty} \sum_{k=0}^{r} \binom{r}{k} y^k \left(-\frac{\mu}{\alpha_1} \right)^{r-k} e^{-y^\beta} \mathrm{d}y + \cdots \right.$$
$$\left. \cdots + \alpha_2^{r+1} \int_0^{\infty} \sum_{k=0}^{r} \binom{r}{k} y^k \left(\frac{\mu}{\alpha_2} \right)^{r-k} e^{-y^\beta} \mathrm{d}y \right]$$

$$= C\left[(-1)^r \alpha_1^{r+1} \sum_{k=0}^{r} \binom{r}{k} \left(-\frac{\mu}{\alpha_1} \right)^{r-k} \int_0^{\infty} y^k e^{-y^\beta} \mathrm{d}y + \cdots \right.$$
$$\left. \cdots \alpha_2^{r+1} \sum_{k=0}^{r} \binom{r}{k} \left(\frac{\mu}{\alpha_2} \right)^{r-k} \int_0^{\infty} y^k e^{-y^\beta} \mathrm{d}y \right]$$

$$= \frac{C}{\beta}\left[\sum_{k=0}^{r} \binom{r}{k} (-1)^{2r-k} \alpha_1^{r+1} \alpha_1^{k-r} \mu^{r-k} \Gamma\left(\frac{k+1}{\beta}\right) + \cdots \right.$$
$$\left. \cdots \sum_{k=0}^{r} \binom{r}{k} \alpha_2^{r+1} \alpha_2^{k-r} \mu^{r-k} \Gamma\left(\frac{k+1}{\beta}\right) \right]$$

$$= \frac{1}{(\alpha_1 + \alpha_2)\Gamma(1/\beta)} \sum_{k=0}^{r} \binom{r}{k} \left[(-1)^k \alpha_1^{k+1} + \alpha_2^{k+1} \right] \Gamma\left(\frac{k+1}{\beta}\right) \mu^{r-k} \quad \blacksquare$$

Proposition 2. Using the proposition 1, the first four moments (mean, variance, skewness, and kurtosis) of X are deduced as

Mean

$$me = EX = \mu - (\alpha_1 - \alpha_2) \frac{\Gamma(2/\beta)}{\Gamma(1/\beta)} \qquad (5)$$

Variance

$$\sigma^2 = E(X - EX)^2 = EX^2 - (EX)^2 = \frac{\alpha_1^3 + \alpha_2^3}{\alpha_1 + \alpha_2} \frac{\Gamma(3/\beta)}{\Gamma(1/\beta)} - (\alpha_1 - \alpha_2)^2 \frac{\Gamma(2/\beta)^2}{\Gamma(1/\beta)^2} \qquad (6)$$

Skewness

$$\gamma_1 = \frac{E(X - EX)^3}{\sigma^3} = \frac{EX^3 - 3EXEX^2 + 2(EX)^3}{\sigma^3} = \frac{\alpha_1 - \alpha_2}{\sigma^3} \left[-2(\alpha_1 - \alpha_2)^2 \frac{\Gamma(2/\beta)^3}{\Gamma(1/\beta)^3} + \cdots \right.$$
$$\left. 3\frac{\alpha_1^3 + \alpha_2^3}{\alpha_1 + \alpha_2} \frac{\Gamma(2/\beta)\Gamma(3/\beta)}{\Gamma(1/\beta)^2} - (\alpha_1^2 + \alpha_2^2) \frac{\Gamma(4/\beta)}{\Gamma(1/\beta)} \right] \qquad (7)$$

Kurtosis

$$\gamma_2 = \frac{E(X - EX)^4}{\sigma^4} - 3 = \frac{EX^4 - 3EXEX^3 + 6(EX)^2 EX^2 - 3(EX)^4}{\sigma^4} - 3$$
$$= \frac{1}{\sigma^4} \left[\frac{\alpha_1^5 + \alpha_2^5}{\alpha_1 + \alpha_2} \frac{\Gamma(5/\beta)}{\Gamma(1/\beta)} - (\alpha_1 - \alpha_2)^2 (4(\alpha_1^2 + \alpha_2^2) \frac{\Gamma(2/\beta)\Gamma(4/\beta)}{\Gamma(1/\beta)^2} \right.$$
$$\left. - 6\frac{\alpha_1^3 + \alpha_2^3}{\alpha_1 + \alpha_2} \frac{\Gamma(2/\beta)^2 \Gamma(3/\beta)}{\Gamma(1/\beta)^3} + 3(\alpha_1 - \alpha_2)^2 \frac{\Gamma(2/\beta)^4}{\Gamma(1/\beta)^4}) \right] - 3 \qquad (8)$$

Knowing that $\{x_1, x_2, \ldots, x_n\}$ is a set of n realizations of the random variable X, the empirical moments or sample moments denoted $me_s, \sigma_s^2, \gamma_{1s}$ and γ_{2s} are computed as

$$me_s = \bar{x} = \sum_{i=1}^{n} x_i / n \tag{9}$$

$$\sigma_s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 / n \tag{10}$$

$$\gamma_{1s} = \sum_{i=1}^{n} (x_i - \bar{x})^3 / (n\sigma_s^3) \tag{11}$$

$$\gamma_{2s} = \sum_{i=1}^{n} (x_i - \bar{x})^4 / (n\sigma_s^4) \tag{12}$$

In this subsection, the likelihood maximization and the moment matching methods are combined for a mixture of AGGDs. In fact, the E-step of the EM algorithm, as in MLE [2], will use the likelihood maximization iteratively to better fit the estimated mixture model to the real histogram while the moment matching will be used to estimate the AGGD parameters. The MME consists to equate the theoretical expressions of moments *me*, σ^2, γ_1 and γ_2 given by equations (5)–(8) with samples moment values given in (9)–(12) and then solving numerically the nonlinear system of equations to obtain the estimated parameter vector $\vec{\theta}_{mme} = (\mu_{mme}, \alpha_{1_{mme}}, \alpha_{2_{mme}}, \beta_{mme})^T$.

Experiments

To compare between the MME and MLE performances in fitting image intensity histograms assumed as AGGMM, two real images are used. The first image represents a magnetic resonance image (MRI) of a brain consisting, after background removal, of three modes : white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF) [3] (see Fig. 2). The second image is a radiographic image of a welded joint used in nondestructive testing (NDT) [4]. For EM initialization, all the initial prior probabilities $\pi_m(0)$ are taken equal to $1/M$; the pseudo-means $\mu_m(0)$ are chosen as $\mu_m(0) = H^{-1}(0.05) + m \times (H^{-1}(0.95) - H^{-1}(0.05))/(M+1)$ where H denotes the cumulated histogram; whilst $\sigma_{1,m}^2(0) = \sigma_{2,m}^2(0) = (\mu_2(0) - \mu_1(0))^2/10$ and all initial shape parameters $\beta_m(0)$ are taken equal to 1.8. The sum of squared error (SSE) is used to measure the histogram curve fitting

$$SSE = \sum_{g} \left(h(g) - f(g, \hat{\Theta}) \right)^2 \tag{13}$$

The segmented image is mapped by assigning the optimal realization $S(u,v)$ for each pixel (u,v) in the image according to the Bayes decision rule (see [2] for more details). Then, a region uniformity measure U is used to evaluate the segmentation performance method. It is given by

$$U = 1 - \sum_{m=1}^{M} \pi_m \sigma_m^2 / \sigma_T^2 \tag{14}$$

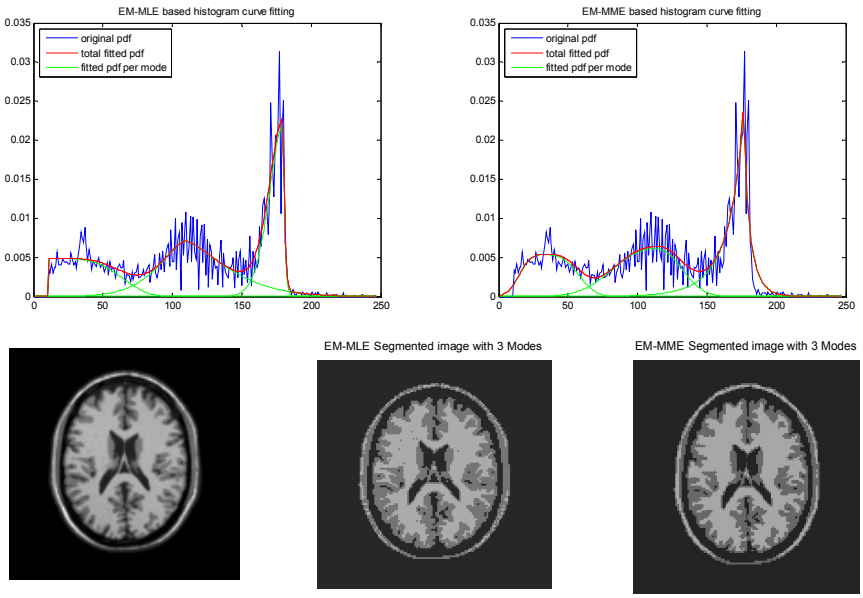
where π_m is the area ratio of the m^{th} segmented region and σ_m^2 its variance, whilst σ_T^2 is the total image variance. The highest is the value of U, the better is the segmentation. The estimated parameters, the histogram fitting errors and the segmentation evaluation measures for both images and methods (EM-MLE and EM-MME) are reported in Table 1. The *SSE* and U values obtained on the chosen images are comparable for the both methods.

Conclusion

In this paper, the moment matching method is applied on finite mixture of AGGDs to fit gray level histograms of a brain MRI and Radiographic NDT images. The resulted estimated parameters and fitting errors are comparable to those of the EM-MLE method in a reasonable agreement. Moreover, the solutions given by this method requires less computation effort thanks to its relative simpler nonlinear equations, comparatively to EM-MLE parameter equations where the nonlinearities are from many types (see [2] for more details).

References

1. K. Kokkinakis, A.K. Nandi. Exponent parameter estimation for generalized Gaussian probability density functions with application to speech modeling // Signal Processing. - 2005. – Vol. 85, No. 9. – P. 1852-1858.
2. N. Nacereddine, S. Tabbone, D. Ziou, L. Hamami. Asymmetric generalized Gaussian mixture models and EM algorithm for image segmentation // 20th International Conference on Pattern Recognition. - Istanbul, Turkey 23-26 Aug. 2010 - P. 4557-4560.
3. I.N. Bankman. Hanbook of Medical Imaging Processing and Analysis, Academic press.
4. N. Nacereddine, D. Ziou, L. Hamami. Fusion-based shape descriptor for weld defect radiographic image retrieval // The international Journal of Advanced Manufacturing Technology –2013.–Vol. 66, No 5-8.



5. Fig. 2. MRI of brain segmented in three regions (WM, GM and CSF) via MLE and MME on AGGMM

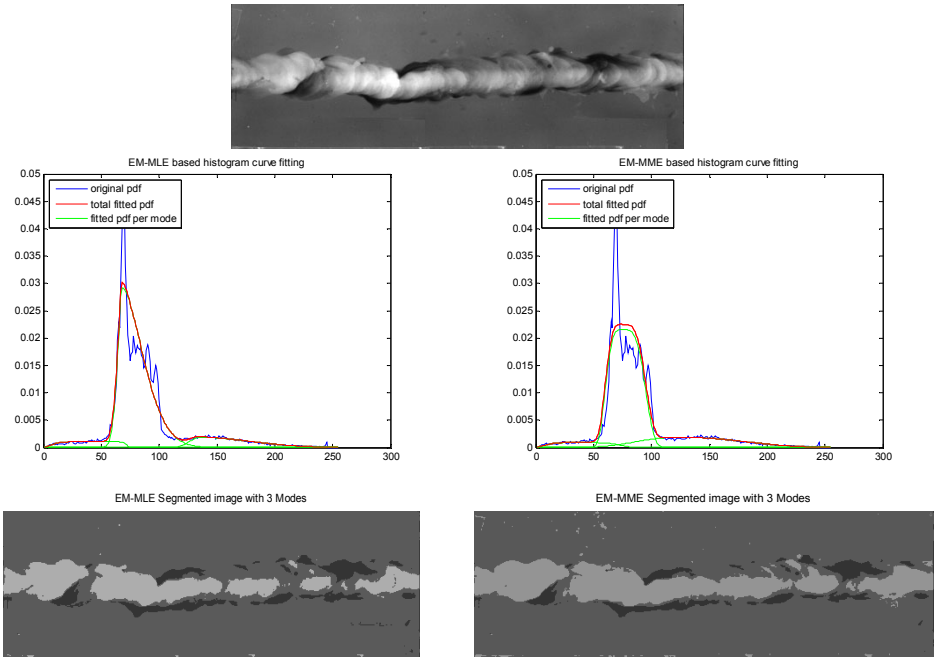


Fig. 3. Weld radiographic images segmented in three regions via MLE and MME on AGGMM

Table 1. Estimated parameters, histogram fitting errors and segmentation evaluation scores for AGGMM via EM-MLE and EM-MME

		π	μ	α_1	α_2	β	SSE ($\times 10^{-4}$)	U
Fig.2	MLE	{0.25 0.44 0.31}	{11 108.8 179}	{0.0 26.7 13.6}	{57.4 41.2 2.3}	{4.1 1.5 1.7}	**9.4**	0.960
	MME	{0.23 0.35 0.42}	{32.6 112.5 176.3}	{20.3 39.0 10.9}	{28.7 25.3 4.4}	{3.1 2.3 0.8}	11.7	**0.973**
Fig.3	MLE	{0.07 0.81 0.12}	{52 68.1 135}	{46 5.7 12}	{20.3 25.5 61.6}	{10.4 1.7 2}	**13.8**	**0.765**
	MME	{0.06 0.74 0.20}	{36.6 75.1 124.2}	{30.5 16.4 47.9}	{38.8 21.9 73.4}	{4.8 3.8 2.4}	23.5	0.753

A COMPARATIVE STUDY ABOUT OBJECT CLASSIFICATION BASED ON GLOBAL AND LOCAL FEATURES

H. Naeem[1], M. Minhas[2]

[1] HITEC University, Taxila, Pakistan hammadnaeem@gmail.com
[2] University of Minnesota, Duluth, USA maria.16988@gmail.com

Scene classification and object recognition is a hot area of research in the field of computer vision and has always fascinated researchers to explore strategies for optimization of results. Global and local features are manipulated to find a match in the images or scene categories. This paper mainly comprises of finding the scene labels based on the objects present in it. The image is transformed into a feature space and the classifier is trained to differentiate each class in the feature space. Various feature extraction techniques like SIFT and Dense Sift are explored in this paper to find an optimized result. Different classifiers were tested individually as well as their combinations to achieve better results. Combination of Sparse SIFT and Dense SIFT techniques was found to perform better compared to others.

Introduction

Object recognition in an image is an old yet inconclusive area of research in computer vision. The major problems encountered in object recognition are objects present in the scene having high clutter, illumination variations, high occlusion, high degree of geometric transformations and intra class variation.

Local features are more robust to clutter, occlusion, light changes and geometric transformations hence, strategies using them proved to be more successful [1]. Local features were successfully utilized for a long time but the problem with local features was that they normally lose all the information about the spatial layout of the image and hence lack the descriptive ability. In this paper, we have worked on the PASCAL dataset from 2006 in which there are ten different classes of object and we have to detect a particular object in a given image. The ten different object classes are: bicycle, bus, car, cat, cow, dog, horse, motorcycle, people and sheep. Classification is not easy as objects have high degree of geometrical transformation, illumination changes, clutter, occlusion and intra-class variations [1]. There are also cases in which an image contains objects from different classes. This only contributes in making classifier confused in deciding about the particular object in the image.

General Strategy

Fairly, large set of training images is provided for each of the ten classes. Classifier is trained with the help of these training images. First features are extracted from images of each class using any feature extraction technique and are then mapped according to the bag of words vocabulary.The feature spaces generated by extraction of features from the images of different classes are grouped into k clusters with the help of k-mean algorithm. The numbers of clusters generated define the size of vocabulary. This technique of grouping or clustering of features is called ‘Bag of words’technique.

Bag of words technique has demonstrated a fair deal of success and has surpassed many past techniques.Next histogram is generated for each of the bag of words feature and classifier is trained for each of the class. The classifier is trained to indicate the presence or absence of an object in the image with some level of confidence. Classifier is trained with positive and negative examples of the class to generate decision boundaries for a class in the feature space. Ground truths are provided for each image. They are compared with the labels

assigned by the classifier and ROC curves are generated showing the performance of technique for that particular class. It is necessary to mention that histograms are normalized in order to have same number of features from each image. For maximum computational efficiency, we normalize all histograms by the total weight of all features in the image, in effect forcing the total number of features in all images to be the same [3].

Experiments and Results

Results for the various strategies used are follows.

1) Sparse SIFT

SIFT features were tested with a number of parameters which are discussed in the following sections:

Vocabulary Parameters: As for SIFT, 'Bag of Features' technique was used. The parameters include number of images used for generating the vocabulary and size of the vocabulary (number of clusters or words in the vocabulary) for extracting the SIFT features from the images. These are explained in the following sections a and b.

a) Number of Images: For this, the number of images used to generate the vocabulary is varied, vocabulary size is kept constant at '500'. The numbers of images used are 5, 10 and 15. The results are shown in table 1.

Table 1. Results for varying number of images

	5	10	15
Bicycle	0.611	0.625	0.632
Bus	0.392	0.365	0.365
Car	0.665	0.685	0.673
Cat	0.628	0.649	0.655
Cow	0.638	0.673	0.679
Dog	0.486	0.493	0.471
Horse	0.401	0.394	0.426
Motorbike	0.568	0.515	0.508
Person	0.621	0.628	0.628
Sheep	0.717	0.725	0.711

It can be observed from table 1 that by increasing the number of images per class for vocabulary generation improved the results. However, no definitive conclusion can be drawn between the performance improvement and number of images used to generate vocabulary. It might be because of the quality of images used for that particular class.

b) Varying Vocabulary Size: In this, the vocabulary size was varied while keeping the number of images constant at '10' .The value of 'k' was varied in K-mean algorithm, which is used to cluster features to generate different sizes for vocabulary. The values of 'k' taken were 200, 500 and 1000. Thus we get 200, 500 and 1000 clusters or words in the vocabulary.The results are shown in table 2.

Table 2. Results for varying the vocabulary size

	200	500	1000
Bicycle	0.649	0.625	0.775
Bus	0.498	0.365	0.667
Car	0.728	0.685	0.745
Cat	0.495	0.649	0.614
Cow	0.701	0.673	0.815
Dog	0.496	0.493	0.549
Horse	0.452	0.394	0.400
Motorbike	0.542	0.515	0.573
Person	0.494	0.628	0.539
Sheep	0.607	0.725	0.814

A definitive trend can be seen from table 2 that by increasing the size of the bag of words, results were greatly improved hence a better result classification can be observed.

2) Dense SIFT

Table 3. Results for varying scale in dense SIFT

	1.5	5	10
Bicycle	0.785	0.812	0.726
Bus	0.718	0.599	0.645
Car	0.756	0.736	0.708
Cat	0.689	0.667	0.661
Cow	0.841	0.801	0.783
Dog	0.665	0.524	0.501
Horse	0.638	0.498	0.473
Motorbike	0.583	0.595	0.567
Person	0.617	0.559	0.583
Sheep	0.695	0.662	0.629

Next, the experiment was done with the dense SIFT by varying the size of frames used to extract features from the image. As in Dense SIFT, we scan the whole image and features are extracted uniformly from each part (Frame or window) of the

image, so by changing the scale, we effectively change the number of descriptors for an image. In this experiment, the vocabulary size was kept fixed at 500 and number of images used to build vocabulary had a constant valueof 10. The scale of the frame had values of 1.5, 5 and 10. The results are shown in table 3.

Results came out quite good for a scale value of 1.5. Generally, as we increase the value of the scale, most of the classes don`t have the good results as compared to low scale values. It is observed that Dense SIFT gives better results than Sparse SIFT but for some classes trend is opposite. DENSE SIFT has particularly given better result for the difficult class 'person' in which we get high value for AUC as compared to optimized sparse SIFT parameters. But for another difficult class, 'horse' results are not good for dense SIFT and sparse SIFT gives better results. So, it is difficult to conclude that dense SIFT is better than sparse SIFT, but still dense SIFT has an edge over sparse sift as results are quite good for some of the classes and average classification value is better for dense SIFT.

3) Classifiers

Performance of different classifiers was also tested in order to achieve better results.

a) Varying Classifier Type: Here, the main comparison is done between KNN classifier and Support Vector Machine SVM classifier.10 images were used to build a standard vocabulary of 500 words. The value of the threshold is kept constant at 0. Next, different classifiers are selected from the 'PRtool box' [2] and their performance is tested with the same dataset.

Table 4. Results for varying classifier types

	KNN-1	SVM-P1	SVM-R1
Bicycle	0.412	0.625	0.627
Bus	0.489	0.365	0.381
Car	0.312	0.685	0.697
Cat	0.624	0.649	0.632
Cow	0.739	0.673	0.675
Dog	0.592	0.493	0.484
Horse	0.457	0.394	0.491
Motorbike	0.513	0.515	0.521
Person	0.516	0.628	0.623
Sheep	0.565	0.725	0.729

The results of the experiment can be seen from table 4.It was observed that SVM classifier clearly performs better as compared to KNN classifier.

b) Combining Different Classifiers: In this, we combined different classifiers together. The classifiers combined were SVM, KNN and LDC.

Table 5. Results for combining different classifiers

Bicycle	0.322
Bus	0.522
Car	0.306
Cat	0.664
Cow	0.710
Dog	0.601
Horse	0.394
Motorbike	0.467
Person	0.532
Sheep	0.618

From table 5, it can be observed that combining different classifier`s does not produce good results. The performance of the individual classifiers was better compared to combined ones.

Combining Features

In sparse SIFT, we only take strong features and in Dense SIFT, we take features from a window scanned over the whole image. This thing brought an idea that the performance of both these features should be checked together. So for this purpose the sparse and dense SIFT features were combined.

Table 6. Results for combined Sparse and Dense SIFT

Bicycle	0.792
Bus	0.568
Car	0.765
Cat	0.753
Cow	0.849
Dog	0.652
Horse	0.514
Motorbike	0.634
Person	0.552
Sheep	0.714

From table 6, it can be seen that the performance is quite good as compared to the individual case. The main reason for this can be that by combining sparse and dense SIFT;

features from all over the image are taken along with the strong features. This might have helped the classifier to draw better decision boundaries, considering the image is now represented in a better way.

Optimization

Keeping track of all the experiments done in the previous sections and their optimized parameters, a final classifier is built. As both sparse and dense SIFT show competitive results and it was difficult to decide about the selection of one, final classification is done with both sparse and dense SIFT. For sparse SIFT, the value of the threshold is taken to be '0' and for dense SIFT, the value of the scale is taken to be '1.5' with the spacing step of '10'. This time the size of the vocabulary is taken as 500. Vocabulary size of 1000 and 1500 can also be used for further improvement of results. The results of the final classifier are shown in table 7.

Table 7. Results for final classification

	Sparse SIFT	Dense SIFT
Bicycle	0.792	0.817
Bus	0.667	0.725
Car	0.745	0.756
Cat	0.598	0.649
Cow	0.815	0.848
Dog	0.647	0.549
Horse	0.614	0.503
Motorbike	0.539	0.642
Person	0.683	0.540
Sheep	0.814	0.725

Conclusion

Object classification is a difficult and complex problem. There is no one technique which outclasses other techniques in all cases. Further, the performances of the classes are not uniform for any given technique. Different classes behave differently with the changed set of parameters. So, it is difficult to devise a single strategy which is equally good for all the classes.

In general, it is found that SIFT performs better than other techniques where dense SIFT has a slight edge over sparse SIFT. The number of images used to build vocabulary also affects the results. The greater the number of images better will be the results. But after a certain value, increase in number of images have a very small impact on result improvement but adds a huge computational cost. The size of the vocabulary is also a very important factor and results improve a lot with the increase in the vocabulary size. Bag of words is a smart technique and has proven quite good in object or scene classification techniques. Also it is observed that quality and quantity of the SIFT features also effects the performance of the classifier. The SVM classifier was found to be better than KNN classifier for this case. The quality (in terms of diversity and description for a particular class) of images used for the training and bag of words is also an important factor for classification performance.

References

1. M. Everingham, A. Zisserman, C. K. I. Williams, and L. Van Gool. The pascal visual object classes challenge 2006. PASCAL Network of Excellence on Pattern Analysis, statistical Modelling and Computational Learning.
2. R. Duin, P. Juszczak, P. Paclik, E. Pekalska, D. de Ridder, and D. Tax, PRTools4, A Matlab Toolbox for pattern Recognition, Delft University of Technology, February 2004.
3. A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms,"

INCREMENTAL GENERALIZED LOW RANK APPROXIMATION OF MATRICES FOR VISUAL LEARNING AND RECOGNITION

H. Nakouri[1], M. Limam[1,3]

[1] LARODEC Laboratory, University of Tunis, Tunisia,[1] nakouri.hayfa@gmail.com
[2] Dhofar University, Oman, Tunisia, [3] mohamed.limam@isg.rnu.tn

In this paper, we consider the problem of incrementally computing low-rank approximations of matrices. We propose an incremental low rank approximation algorithm of a sequence of matrices instead of one single matrix. Based on the generalized Low Rank Approximation of matrices approach, the proposed method uses an iterative approach to update the generated low rank structures. We tested the proposed incremental learning method on five face databases to evaluate its efficiency in terms of recognition rate and compared its performance with the Incremental Singular Value Decomposition method. We also showed that the overall reconstruction error is kept bounded during the incremental learning process.

Introduction

In recent years, increasingly more data items come naturally as $2D$ objects, such as $2D$ images. Currently, widely used methods for dimension reduction of these $2D$ data objects are based on the approach of vectorizing them on $1D$ vector and stacking them together in one large matrix. For example, each image A_i, $A_i \in \mathbb{R}^{\mathrm{r}\times\mathrm{c}}$, $i = 1, \dots, N$ is transformed to a vector a_i of length $l = \mathrm{r} \times \mathrm{c}$. The standard SVD is then applied to a matrix A containing all the vectors, $A = (a_i, \dots, a_N)$. This approach is largely used in image processing and is well known as Eigenfaces [1]. Nevertheless, this approach does not preserve the $2D$ of data images. The Generalized Low rank Approximation of Matrices (GLRAM) approach, proposed by Ye [2], captures explicitly the $2D$ nature and has three advantages over conventional SVD-based approaches: First, it deals with much smaller matrices, typically $r \times c$ matrices, instead of $n \times rc$ matrix in conventional approach. Second, the new approach requires substantially smaller memory storage at the same or even better than accuracy of reconstruction. Third, some of the operations on these rectangular objects can be done much more efficiently, e.g. object recognition, due to the well preservation of the $2D$ structure. While conventional low rank approximation approaches use a vectorized representation of the $2D$ image matrix, $GLRAM$ works with a representation that is closer to the $2D$ matrix representation and attempts to preserve spatial locality of the pixels. The vectorized representation of images may lead to loss of spatial locality information. On the other hand, GLRAM is a batch learning method. It means that all training data that have to be available to compute the low rank components during the training stage. The learning stops once the training data have been fully processed. If we want to incorporate additional training data into an existing projection matrix, the matrix has to be retrained with all training data. In turn, it is hard to scale up the developed systems and not easy to adapt a feature space for time varying and unseen data. To overcome this limitation, an incremental approach is a straightforward solution.

In this paper, we propose a new incremental feature extraction method called incremental GLRAM (IGLRAM) with less computational load and smaller memory waste than the GLRAM. Thus, the objective of this paper is to propose an incremental version of the GLRAM approach by updating the projection data matrices and thus we avoid computing the whole low rank structures from scratch whenever we have new data. Furthermore, we

apply this approach to incremental face recognition and classification.

This paper is organized as follows. The first section gives an overview of GLRAM. The second section introduces our incremental GLRAM (IGLRAM). Finally, we present our experimental results.

Generalized Low Rank Approximation of Matrices (GLRAM)

For a general form of matrix approximation, Ye [2] proposed a matrix space model for low rank approximation of matrices called GLRAM, which can be stated as follows. Let $A_i \in \mathbb{R}^{r\times c}$, for $i = 1, \dots, N$ be the N data points in the training set. We aim to compute two matrices $L \in \mathbb{R}^{r\times l1}$ and $R \in \mathbb{R}^{c\times l2}$ with orthonormal columns and N matrices $M_i \in \mathbb{R}^{l1\times l2}$ for $i = 1, \dots, N$ such that LM_iR^Tapproximates A_i well for all i. Mathematically, this can be formulated as an optimization problem:

$$\min_{\substack{L^TL=I_{l_1}\\ R^TR=I_{l_2}}} \sum_{i=1}^{N}\|A_i - LM_iR^T\|_F^2 \quad (1)$$

where F denotes the Frobenius norm. In order to find an optimal solution for the GLRAM, Ye has proved that the proposed GLRAM is equivalent to the following optimization problem

$$\max_{L,R} J(L,R) = \sum_{i=1}^{N}\|LA_iR^T\|_F^2 = \sum_{i=1}^{N} Trace\,[L^TA_iRR^TA_i^TL] \quad (2)$$

where L and R must satisfy $L^TL = I_{l1}$, $R^TR = I_{l2}$. Ye proposed Algorithm 1 to solve this optimization problem and defined the Root Mean Square Reconstruction Error (RMSRE) criteria to cope with the convergence issue of Algorithm 1. It is defined as:

$$RMSRE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\|A_i - LM_iR^T\|_F^2} \quad (3)$$

It has been shown that each iteration of the GLRAM algorithm will decrease the RMSRE value and that RMSRE is bounded. Hence, the GLRAM algorithm converges.

Algorithm 1 Generalized low Rank Approximation of Matrices (GLRAM)

INPUT: Matrices A_i, l_1 and l_2

OUTPUT: Matrices L, R and M_i

1) Obtain initial $L_0$for L and set $i = 1$
2) **WHILE** not convergent **DO**
3) Form the matrix M_R=$\sum_{j=1}^{n} A_j^T L_{i-1}L_{i-1}^T A_j$
4) Compute the $l2$ largest eigenvectors $\{\Phi_j^R\}_{j=1}^{l2}$ of M_R.
5) Let $R_i \leftarrow [\Phi_1^R, \dots, \Phi_{l2}^R]$
6) Form the matrix M_L=$\sum_{j=1}^{n} A_j R_{i-1}R_{i-1}^T A_j^T$
7) Compute the $l2$ largest eigenvectors $\{\Phi_j^L\}_{j=1}^{l1}$of M_L.
8) Let $L_i \leftarrow [\Phi_1^L, \dots, \Phi_{l1}^L]$
9) $i \leftarrow i + 1$
10) **END WHILE**
11) $L \leftarrow L_{i-1}$
12) $R \leftarrow R_{i-1}$
13) $M_j = L^TA_jR$

Incremental Generalized Low Rank Approximation of Matrices (IGLRAM)

In this section, we develop the incremental version of the generalized low rank approximation of matrices, called IGLRAM.

We assume that we have already built, using the GLRAM method as set of law-rank matrices L, R and M_i, $i = 1, \dots, N$ of N input images. Let $A_{N+1} \in \mathbb{R}^{r\times c}$, we propose an iterative procedure update the transformation matrices $R$and L considering the newly added image A_{N+1}. We aim to compute optimal $\tilde{L} \in \mathbb{R}^{r\times l1}$ and $\tilde{R} \in \mathbb{R}^{c\times l2}$ with orthonormal columns which are the updated L and R, respectively and the matrix M_{n+1}such that LM_iR^T is a good approximation for A_i, $i = 1, \dots, N+1$. This minimization problem can be mathematically formulated as:

$$\min_{\substack{L^TL=I_{l_1}\\ R^TR=I_{l_2}\\ M_i\in\mathbb{R}^{l1\times l2}, i=1,\dots,N+1}} \sum_{i=1}^{N+1}\left\|A_i - \tilde{L}M_i\tilde{R}^T\right\|_F^2 \quad (4)$$

Similar to GLRAM, $\tilde{L}$ and $\tilde{R}$ are considered as the two-sided linear transformations on the data in matrix form, with $\tilde{L}$ and $\tilde{R}$ as the transformations on the left and right sides, respectively. The novelty of our incremental approach is the use of double-sided transformations instead of one single

transformation in the case of traditional low-rank approximations [3]. The incremental process begins by updating the mean:

$$\bar{A}' = (n\bar{A} + A_{N+1})/(n+1) \qquad (5)$$

Then, we center the new input data:

$$A_{N+1} = A_{N+1} - \bar{A}' \qquad (6)$$

The iteration procedure is used to compute $\tilde{L}$ and $\tilde{R}$, the updated versions of L and R, respectively. We update the set of eigenvectors L and R by adding a new matrix and applying a rotational transformation. In order to do this, we first compute the residual matrix

$$h_R = A_{N+1} - A_{N+1}\tilde{R}\,\tilde{R}^T \qquad (7)$$

And normalize it in order to obtain $\hat{h}_R = \frac{h_R}{\|h_R\|_F}$ for $\|h_R\|_F > 0$ and $\|h_R\|_F = 0$ otherwise. The new matrix of the right transformation $\tilde{R} \in \mathbb{R}^{c\times l2}$ is computed by

$$\tilde{R} = [\tilde{R} \;\; Q_R]\, X_R \qquad (8)$$

Where $Q_R$is obtained by the orthogonal triangular decomposition $[Q_R, R_R] = QR(\hat{h}_R)$. $X_R$is a rotation matrix and it is the solution of the eigenproblem of the following form:

$$G_R X_R = X_R \Lambda'_{\mathrm{R}} \qquad (9)$$

We decompose $G_R$as:

$$G_R = \begin{bmatrix} \Lambda_{\mathrm{R}} & \tilde{R}_{i-1}^T A_{n+1}^T \\ (\tilde{R}_{i-1}^T \tilde{A}_{n+1}^T)^T & \mathbf{0} \end{bmatrix} \qquad (10)$$

Where $\Lambda_{\mathrm{R}}$are the computed eigenvalues of the M_R matrix (line 3 in Algorithm 1). We proceed in a similar way to update the left transformation $\tilde{L} \in \mathbb{R}^{r\times l1}$. Then, the updating process is repeated until we achieve convergence. Once we get optimal $\tilde{L}$ and $\tilde{R}$ transformation matrices, we compute the reducedmatrix $M_{N+1} = \tilde{L}^T A_{n+1} \tilde{R}$. The pseudo-code to update L and R following the above iterative procedure is given in Algorithm 2.

Algorithm 2 Incremental Generalized low Rank Approximation of Matrices (IGLRAM)

INPUT: $A_{N+1} \in \mathbb{R}^{r\times c}$, $L \in \mathbb{R}^{r\times l1}$, $R \in \mathbb{R}^{c\times l2}$, Λ_{L}, Λ_{R}, l_1, l_2 and $\bar{A}$.

OUTPUT: $\tilde{L} \in \mathbb{R}^{r\times l1}$, $M_{N+1} \in \mathbb{R}^{l1\times l2}$ and $\bar{A}'$

1) $\bar{A}' = (n\bar{A} + A_{N+1})/(n+1)$
2) Center new data $A_{N+1} = A_{N+1} - \bar{A}'$
3) $\tilde{R}_0 \leftarrow R$, $\tilde{L}_0 \leftarrow L$ and set $i = 0$
4) **WHILE** not convergent **DO**
5) $h_R = A_{N+1} - A_{N+1}\tilde{R}_{i-1}\tilde{R}_{i-1}^T$
6) $\hat{h}_R = \frac{h_R}{\|h_R\|_F}$
7) Obtain the QR decomposition $[Q_R, R_R] = QR(\hat{h}_R)$
8) Compute rotation matrix $X_R$as the eigenvectors of the G_R:
$$G_R = \begin{bmatrix} \Lambda_{\mathrm{R}} & \tilde{R}_{i-1}^T A_{n+1}^T \\ (\tilde{R}_{i-1}^T \tilde{A}_{n+1}^T)^T & \mathbf{0} \end{bmatrix}$$
9) $\tilde{R}_i \leftarrow [\tilde{R}_{i-1} \;\; Q_R]\, X_R$
10) $h_L = A_{N+1} - \tilde{L}_{i-1}\tilde{L}_{i-1}^T A_{N+1}$
11) $\hat{h}_L = \frac{h_L}{\|h_L\|_F}$
12) Obtain the QR decomposition $[Q_L, R_R] = QR(\hat{h}_L)$
13) Compute rotation matrix X_L as the eigenvectors of the G_L:
$$G_L = \begin{bmatrix} \Lambda_{\mathrm{L}} & \tilde{L}_{i-1}^T A_{n+1} \\ (\tilde{L}_{i-1}^T A_{n+1})^T & \mathbf{0} \end{bmatrix}$$
14) $\tilde{L}_i \leftarrow [\tilde{L}_{i-1} \;\; Q_L]\, X_L$
15) $i \leftarrow i + 1$
16) **END WHILE**
17) $\tilde{R} \leftarrow \tilde{R}_{i-1}$
18) $\tilde{L} \leftarrow \tilde{L}_{i-1}$
19) Compute $M_{N+1} = \tilde{L}^T A_{n+1} \tilde{R}$

Experiments

To test the behavior of the GLRAM algorithm for visual learning and recognition, we carried out a set of experiments. We used five different databases: YALE [4], AR [5], FERET [6], ORL [7] and PIE [8] databases. Table 1. Provides brief information about this database.

Table 1. Used Face databases

%	Size	Subjects	# images
AR	64 64	171	105
ORL	112 92	40	476
PIE	251 251	68	200
YALE	151 151	15	60
FERET	80 80	48	192

All grayscale images are normalized. We use the canonical size of the images rather than the

original one based on the eye corner location. We use the Viola Jones face detector algorithm [9] to capture canonical faces.

Reconstruction Error. We tested the set of images from the AR database, to test the quality of low rank representation when reconstructed incrementally using our method. During the process of the incremental learning, we monitored the quality of the momentary representation of a subset of training images. At each step, we compute the RMSRE value between the original image and its reconstruction.

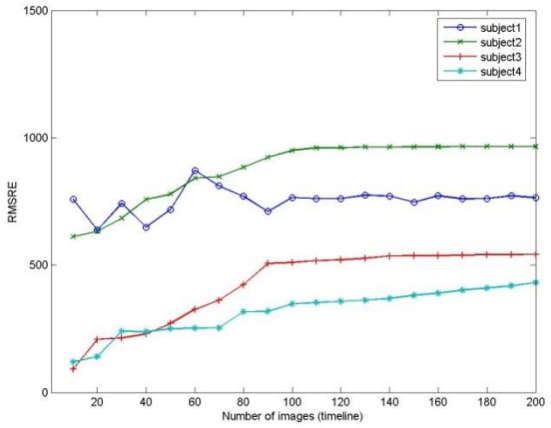


Fig. 1. Reconstruction Error for a subset of images during the incremental learning.

Fig. 1 shows that for all images the reconstruction error is bounded. This indicates that the incremental representation does not deteriorate dramatically during the learning process.

Recognition performance. To show the efficiency of our incremental approach on large datasets, we compare the IGLRAM method to batch GLRAM and to the popular Incremental Singular Value Decomposition ISVD [4] in terms of recognition rate. For each database, we use $2/4$ of images for batch learning using the GLRAM method, $1/4$ of images for incremental learning using the proposed IGLRAM method and the remaining $1/4$ of images are used for the recognition test. Table 2 shows that the recognition performance of our incremental method outperforms that of ISVD with the different used databases. This supports the importance of preserving the 2D structure of image data during the learning process. On the other hand, the proposed method gives less better results than GLRAM. This is due to the fact that our algorithm do not update the reduced representations $M_i$of the previously learned images in each new incremental learning.

Table 2. Recognition performance

%	ISVD	GLRAM	IGLRAM
AR	50.70	80.35	67.36
ORL	52.70	76.70	60.70
PIE	54.30	85.30	64.90
YALE	61.20	84.20	63.35
FERET	63.80	88.80	69.20

Conclusion

In this paper, a new incremental algorithm for low rank approximation of a sequence of matrices is presented. Experiments show that the proposed algorithm better performance than traditional ISVD in terms of recognition rate and reconstruction error. We also showed the convergence of the algorithm due to the stability of the RMSRE criteria.

References

1. M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, no. 1, pp. 103–108, Jan. 1990.
2. J. Ye, "Generalized low rank approximations of matrices," in Machine Learning, 2004, pp. 887–894.
3. P. Hall, D. Marshall, and R. Martin, "Adding and subtracting eigenspaces with eigenvalue decomposition and singular value decomposition," Image Vision Comput., vol. 20, no. 13-14, pp. 1009–1016, 2002.
4. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," 1997.
5. A. Martinez and R. Benavente, "The ar face database," Tech. Rep., 1998.
6. P. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET Database and Evaluation Procedure for Face Recognition Algorithms," Image and Vision Computing, vol. 16, no. 5, pp. 295–306, 1998.
7. F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," pp. 138–142, 1994.
8. T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression (pie) database," 2002.
9. P. Viola and M. Jones, "Robust real-time face detection," International Journal of Computer Vision, vol. 57, pp. 137–154, 2004.

CONTOUR DESCRIPTION METHOD BASED ON MULTIRESPONSE STATISTICAL MODELS

A. S. Naumov[1]

[1] Yaroslav-the-Wise Novgorod State University,
Russian Federation, 173003, Velikiy Novgorod, B.S.-Peterburgskaya st., 41
tel. +79212029083, e-mail: alex.naumov53@mail.ru

A method for image contour description proposed, which is based on contour approximation with multiresponse piecewise-linear segment models and further dominant point detection. Method provides combined usage of geometrical and color information to detect dominant points. Received descriptor could be applied in various tasks related with pattern recognition, image analysis and particularly for reassembling images from fragments of arbitrary shape.

Introduction

In applications related with pattern recognition and reconstruction images from fragments [1] one of the most informative feature of analyzed object is its border shape and color allocation along the border, presented by contour line with dominant (corner) points. In case of shape dominant points have high local curvature value, and in case of color they have noticeable change of color or/and brightness in the neighborhood.

Numerous methods for dominant point detection are known. The most widespread of them (for example IPAN99 [2]) detect corners by inscribing a triangle with defined parameters to the current curve segment and then use technique of non-maximum suppression to detect points with locally maximum curvature. Those methods are mostly empiric and require setting several independent parameters which should be chosen depending on image resolution, corner strength and other features of contour shape. Statistical characteristics of approximation and vector nature of contour line not considered. They are suitable only for geometric contour, so it is necessary to implement additional methods to identificate color dominant points.

Due to this a statistic method for contour description was developed, that uses unified model for geometric and color contour. It allows getting more reliable and factual description, provides universality in compare with existing methods.

Contour model

Let initial information about object contour presented by matrix, each row of which (a single point) includes four variables: x – horizontal pixel coordinate, y_1 – vertical pixel coordinate, y_2 – color coordinate a in the *Lab* color space, y_3 – color coordinate b.

Contour is approximated with piecewise-linear function, each single section of which begins and ends in the dominant points. Linear function chosen because we need distinctly locate noticeable corners which have been smoothed if used higher degree polynomials. Multiresponse linearly parameterized model of linear section of image contour can be presented by vector function

$$\mathbf{Y}_{ij} = \mathbf{P}_i^{\mathrm{T}}\left(x_{ij}\right)\mathbf{B}_i + \mathbf{E}_i \qquad (1)$$

where i – section index, $j = \overline{1, n_i}$ – point index in the *i*-th section, n_i – number of points in the *i*-th section, $N = \sum_{i=1}^{u} n_i$ – total number of points in the image fragment contour, $\mathbf{P}_i^{\mathrm{T}}\left(x_{ij}\right) = \left\{\mathrm{f}_1\left(x_{ij}\right), \mathrm{f}_2\left(x_{ij}\right), \mathrm{f}_3\left(x_{ij}\right)\right\}$ – function that describes dependency of response variables from independent variable x_{ij} (in this case x_{ij} – abscissa value of geometric variable in the j-th point of i-th section),

$\mathbf{Y}_{ij} = \{y_{1ij}, y_{2ij}, y_{3ij}\}^{\mathrm{T}}$ – vector of response values (in this case ordinate y_{ij} and two color variables), $\mathbf{B}_i = \{b_{1i}, b_{2i}, \ldots, b_{li}\}^{\mathrm{T}}$ – coefficients vector for model of *i*-th section, $\mathbf{E}_i = \{e_{1i}, e_{2i}, \ldots, e_{ki}\}^{\mathrm{T}}$ – vector of residuals for the *i*-th section with covariance matrix $\mathbf{V}_{\mathbf{E}i}$.

Matrix $\mathbf{P}(\mathbf{X})$ for 3-response model of the *i*-th section takes the form

$$\mathbf{P}_i^{\mathrm{T}}(x_{ij}) = \begin{Bmatrix} 1 & x_{ij} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & x_{ij} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_{ij} \end{Bmatrix} \quad (2)$$

With normally distributed observation errors according to maximum likelihood method, estimated coefficients for *i*-th section of contour can be calculated as [3]:

$$\hat{\mathbf{B}}_i = \mathbf{V}_{\mathbf{B}i} \sum_{j=1}^{n_i} \mathbf{P}(x_{ij}) \mathbf{V}_{\mathbf{E}i}^{-1} \mathbf{Y}_{ij}, \quad (3)$$

where $\mathbf{V}_{\mathbf{E}i}$ – covariance matrix of observation errors for *i*-th section, and $\mathbf{V}_{\mathbf{B}i}$ – covariance matrix of estimated coefficients for *i*-th section, which is

$$\mathbf{V}_{\mathbf{B}i} = \left[\sum_{j=1}^{n_i} \mathbf{P}(x_{ij}) \mathbf{V}_{\mathbf{E}i}^{-1} \mathbf{P}^{\mathrm{T}}(x_{ij}) \right]^{-1}. \quad (4)$$

To estimate covariance matrix of the observation errors we need preliminary choose contour section with homogeneous color and linear form of geometric contour. Estimation of covariance matrix of observation errors $\mathbf{V}_{\mathbf{E}}$ is calculated as residuals covariance matrix for this homogeneous section:

$$\hat{\mathbf{V}}_{\mathbf{E}i} = \mathbf{V}_{\mathbf{R}i} = \frac{k}{nk - l} \sum_{j=1}^{n_i} \mathbf{R}_{ij} \mathbf{R}_{ij}^{\mathbf{T}}, \quad (5)$$

where *n* – number of points in selected section, $\mathbf{R}_{ij} = \overline{\mathbf{Y}}_{ij} - \mathbf{P}_i^{\mathrm{T}}(x_{ij}) \mathbf{B}_i$ – residual vector for homogeneous segment model.

Contour description algorithm

1. Set initial number of section points for initial calculation of estimated coefficients of the model (1) with formula (3) and covariance matrix with formula (4), $n_i = n_{i\min} > k/m$, where *m* – number of independent variables (in this case *m*=1) [3]. For a model of type (2) minimum allowable number of points are 3. Set number of test points n_p, which will be used to test hypothesis of their attachment to the current section model.
2. Set start point of the image contour, *i*=0.
3. Set initial number of points $n_{i\min}$ for the next *i*=(*i*+1)-th section.
4. Build the linear model for the *i*-th section: calculation of estimated coefficients according to the formula (3), covariance matrix of estimated values (4), likelihood function logarithm (6) and it's variance (7).
5. For specified number of test points n_p perform hypothesis test about that this points (not included in the estimated coefficients calculation), belongs to the current section model calculated at the step 4.
6. If hypothesis is rejected, then admit point n_i as dominant point and start calculation of the next section (step 3). If hypothesis confirmed, increase total current number of section points $n_i = n_i + 1$ and update (recalculate) current section model with all points (step 4).
7. Repeat next point check procedure (step 5) until dominant point is found. This point considered as a first point of the next section.
8. Calculation process with steps 2-5 continues for next sections until the last point of the contour achieved.

Dominant point detection

Dominant point detection accomplished by testing hypothesis about attachment of the next test point of the contour to the model of current section. For the current section average logarithm of likelihood function (log-likelihood) $\bar{l}_c$ calculated as [5]

$$\bar{l}_c = \frac{1}{n} \sum_{j=1}^{n_i} \left[\overline{\mathbf{Y}}_j - \mathbf{P}^{\mathrm{T}}(x_j) \hat{\mathbf{B}} \right]^{\mathrm{T}} \hat{\mathbf{V}}_{\mathbf{E}}^{-1} \left[\overline{\mathbf{Y}}_j - \mathbf{P}^{\mathrm{T}}(x_j) \hat{\mathbf{B}} \right] \quad (6)$$

and it's variance s_c^2 as

$$s_c^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} \left(l_{cj} - \bar{l}_c \right)^2. \quad (7)$$

For the next test point, considering that it belongs to the current *i*-th section model, log-likelihood function value calculated as:

$$l_{\mathrm{p}} = \left[\mathbf{Y}_{\mathrm{p}} - \mathbf{P}^{\mathrm{T}}\left(x_{\mathrm{p}}\right)\hat{\mathbf{B}}_i\right]^{\mathrm{T}} \hat{\mathbf{V}}_{\mathbf{E}}^{-1}\left[\mathbf{Y}_{\mathrm{p}} - \mathbf{P}^{\mathrm{T}}\left(x_{\mathrm{p}}\right)\hat{\mathbf{B}}_i\right], \quad (8)$$

where $\mathbf{Y}_{\mathrm{p}}$ – response vector of test point, x_{p} – independent variable value for the test point, $\hat{\mathbf{B}}_i$ – vector of estimated coefficients of the current section model.

Test point belongs to the current section model if this inequality is true:

$$l_{\mathrm{p}} - \bar{l}_c \le u_{1-\alpha} s_c, \quad (9)$$

where $u_{1-\alpha}$ - normal distribution quantile with significance level of α. To perform robust dominant point detection inequality (9) should be tested for n_p consecutive points.

Figure 1 shows, how developed method works in case of geometric contour. A part of contour approximated with two linear sections. Point with x_{i+1}, y_{i+1} coordinates is the dominant point and the boundary of the 2 linear sections of approximation $\mathbf{Y}_i(x)$ and $\mathbf{Y}_{i+1}(x)$. As can be seen, y_{i+1} - original value of y coordinate of the contour point, y'_{i+1} - value estimated by the model, n_p – number of test points used for checking criterion (9). Difference $y_{i+1} - y'_{i+1}$ in practical applications is insignificant, so original contour point could be considered as a dominant point. Color contour function approximated in a similar way.

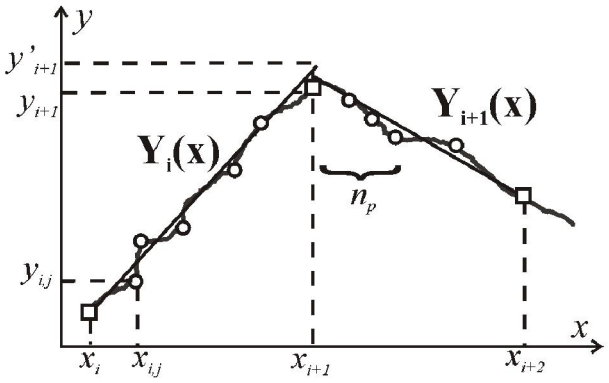


Fig.1. Linear sections and dominant point.

Implementation and experimental results

Proposed method was implemented in the program system and tested on several types of original images. Figure 2 shows an image of the military jet plane. Detected dominant points of the geometrical contour give visually sufficient accuracy and adequacy in shape approximation. If necessary, the additional filtering could be done to get more coarse representation and delete less valuable points. Figure 3 presents the fragment image of the ceramic painted object with marked dominant points of geometrical and color contour. For clarity they are shown at two separate pictures.

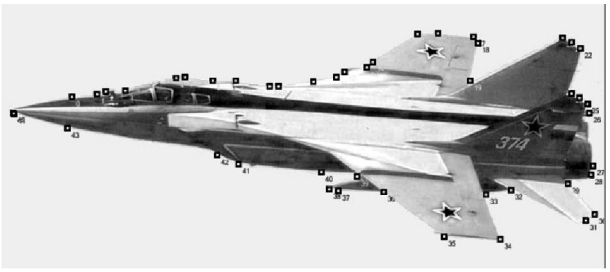
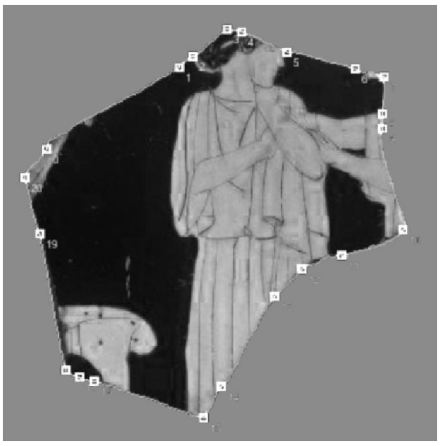
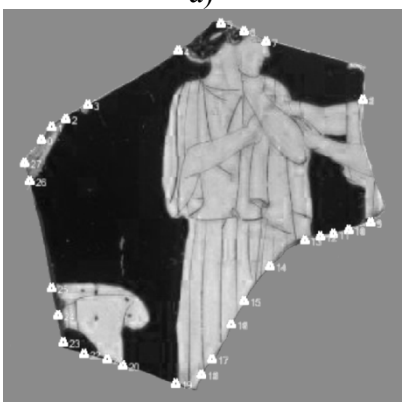


Fig.2. Object shape representation with dominant points.

Due to minimizing edge inaccuracies, color contour was extracted with offset from geometrical contour towards the fragment center. Geometrical contour was extracted by usual edge detection methods. Parameter values have been chosen so that $n_{i\min} \approx 2n_p$.



a)



b)

Fig.3. Ceramic fragment image with detected dominant points for: a) geometric contour; b) color contour.

As it's seen, object shape presented accurately, remarkable color transitions along the contour were also identified. This allows applying received descriptors for searching matching contour parts and locating junctions in a problem of reassembling images from fragments.

Experimental results demonstrated working ability and efficiency of the method. For achieving best results it's required to choose optimal parameter values $n_{i\min}$ and n_p according to the type of fragments and characteristics of the specific task. Method can be applied separately to the geometric and color contour or simultaneously for the both.

Using x coordinate as independent variable imposes a limit for vertical lines approximation, so it's appropriate to use path length along the curve u as independent variable, considering x and y as dependent variables and parameterized as $x(u)$, $y(u)$. Total number of response variables could also be varied according to specific task, for example, including an L value of the Lab color model and other alterations.

Conclusion

Proposed approach for contour description allows taking into consideration statistical characteristics of the image fragment contour. Multiresponse model gets possible to use additional (color) variables for contour approximation and apply the same model for geometrical and color contour. This increases method versatility and expand scope of possible applications in compare with existing methods. Developed algorithm and its modifications could be applied for description of contours and arbitrary curves in different applications related with pattern recognition, reassembling images from fragments and other practical tasks.

References

1. A.S. Naumov, S.A. Lutsiy Forming of the contour descriptor for the problem of image synthesis from fragments of a random shape // Vestnik NovSU: Natural and technical sciences, - 2012. - №.68. - P.68-74. (in Russian)
2. Chetverikov D. A Simple and Efficient Algorithm for Detection of High Curvature Points in Planar Curves // Computer Analysis of Images and Patterns. Lecture Notes in Computer Science, - 2003. - Vol.2756. - P.746-753.
3. S. A. Popov, G. M. Emelyanov. Color Correction of Digital Images by Means of Multiresponse Regression Models // Pattern Recognition and Image Analysis, - 2002. - Vol. 12, №. 2. - P. 145-149.
4. S. A. Popov, G. M. Emelyanov. Color Oservation of Digital Images: Error Analysis // Pattern Recognition and Image Analysis, - 2005 Vol. 15, № 2. - P. 108-112.
5. Rencher, Alvin C. Methods of Multivariate Analysis / A John Wiley & Sons, Inc. Publication. Brigham Young University, 2002.– 708 p.

ILLUMINANT COLOR CORRECTION, USING COLOR SHAPE UNITS METHOD[1]

A. Nikonorov[2,3]

[2] S. P. Korolyov Samara State Aerospace University, 34 Moskovskoye Shosse, 443086 Samara,
[3] Image Processing Systems Institute of RAS 151 Molodogvardeyskaya st., 443001 Samara,
artniko@gmail.com

This paper presents the color-shape units model for correction of non-isoplanatic deviation in the scene illumination. We propose an identification of the color correction function on the set of color shape units with the Hausdorff metric. Also a necessary condition which allows to obtain an adequate form of the color correction function is presented. The experiments performed on real images confirm the high quality of the proposed color correction technique.

Correction of the illuminant deviations in color image

This paper mainly considers a three channel image. The retina or a trichromatic (or sometimes more than trichromatic) sensor generally collects three or more measurements depending on the spectral reflectance of the observed scene surfaces as well as the spectral irradiance that is incident upon it.

Definition 1.

We define the three-vector of measurements obtained in each pixel as:

$$\mathbf{p}(\mathbf{x}) = \begin{bmatrix} p^1(\mathbf{x}) \\ ... \\ p^3(\mathbf{x}) \end{bmatrix} = \int R(\lambda,\mathbf{x}) I(\lambda,\mathbf{x}) \mathbf{F}(\lambda) \quad (1)$$

where R and S are $R \times Z_2 \to [0,1]$ functions of wavelength λ, and the first one is the spectral reflectance of the scene surfaces and the second is the spectral irradiance that is incident at each scene point. $\mathbf{F}(\lambda) = [F^{\{1\}}(\lambda),...,F^{\{K\}}(\lambda)]^T$ is the spectral transmittance distribution of color sensors. The color space, which contains points p(x) is normalized using L_2 norm or using unicontrast norm CIE Lab [1]. For the color correction as well as for color segmentation we can use a five dimensional space Υ which merges the color and spatial spaces [2]:

$$[\mathbf{x}, \mathbf{p}(\mathbf{x})] \in \Upsilon = Z_2 \oplus R_K \quad (2)$$

Metric of this space is defined as:

$$\|\mathbf{x}, \mathbf{p}(\mathbf{x})\| = \|\mathbf{x}\| + \|\mathbf{p}(\mathbf{x})\| \quad (3)$$

Most of the existent color correction techniques, such as "grey world" or retinex approaches [3], are based on the assumption that the incident irradiance is isoplanatic. Thus these techniques transform the image color to the value which corresponds to a certain ideal irradiance or illuminant – $I_I(\lambda)$:

$$\mathbf{p}(\mathbf{x}) \to \mathbf{p}_I(\mathbf{x}) = \int R(\lambda,\mathbf{x}) I_I(\lambda) \mathbf{F}(\lambda) d\lambda \quad (4)$$

This approach is not accurate enough for several tasks of high fidelity color correction. This type of problems of color correction is considered in this paper. The problem is to correct non-isoplanatic deviation in illumination and to restore an image with the given illumination $I_0(\lambda)$:

$$\mathbf{p}(\mathbf{x}) = \int R(\lambda,\mathbf{x}) I(\lambda,\mathbf{x}) \mathbf{F}(\lambda,\mathbf{x}) d\lambda \quad (5)$$

$$\to \mathbf{p}_0(\mathbf{x}) = \int R(\lambda,\mathbf{x}) I_0(\lambda) \mathbf{F}(\lambda) d\lambda$$

The problems (4) and (5) are ill-posed, and the only way to solve them is to use *a priori* knowledge of some kind. Thus color invariant methods use "grey world" or "grey edges" assumptions [4]. We propose to use the prior knowledge of the colors of small closed patches of the image, as a color correction specialist does.

[1] This work is partially supported by RFBR (No. 11-07-12051-ofi-m, 12-07-00581-a, 12-07-31208)

In a number of papers [5], it is shown that the dichromatic model provides the necessary accuracy for color correction. The modified dichromatic model includes a diffuse reflection $L_B(\lambda,\mathbf{x})$, a specular reflection $L_S(\lambda,\mathbf{x})$, and an ambient light $L_A(\lambda)$ as separate terms:

$$L(\lambda,\mathbf{x}) = L_B(\lambda,\mathbf{x}) + L_S(\lambda,\mathbf{x}) + L_A(\lambda) \qquad (6)$$

Known illumination distortions are described by a specular reflectance term $L_S(\mathbf{x},\lambda)$ and shadowing multiplier $H(\mathbf{x})$. Using (1) and (6) it can be shown that pixel of the image is defined as:

$$\mathbf{p}(\mathbf{x}) = \int [H(\mathbf{x})(I_B(\lambda)R(\lambda,\mathbf{x}) + \qquad (7)$$
$$+H(\mathbf{x})(I_C(\lambda,\mathbf{x}) + I_S(\lambda,\mathbf{x})R(\lambda,\mathbf{x})) +$$
$$+I_A(\lambda)R(\lambda,\mathbf{x})]\mathbf{F}(\lambda)d\lambda$$

As follows from the models of specularity and shadowing [5], $L_S(\mathbf{x},\lambda)$ and $H(\mathbf{x})$ in (7) are smooth functions of $\mathbf{x}$. Hence, there exist closed curves γ such that the values of the following functions are constant:

$$H(\mathbf{x}) = H_\gamma, I_s(\lambda,\mathbf{x}) = I_{s\gamma}(\lambda) \quad \forall \mathbf{x} \in \gamma \qquad (8)$$

For such curve γ, (7) gets the following form:

$$\mathbf{p}(\mathbf{x}) = \int [H_\gamma(\mathbf{x})(I_B(\lambda)R(\lambda,\mathbf{x}) + \qquad (9)$$
$$+H(\mathbf{x})(I_C(\lambda,\mathbf{x}) + I_{s\gamma}(\lambda,\mathbf{x})R(\lambda,\mathbf{x})) +$$
$$+I_A(\lambda)R(\lambda,\mathbf{x})\mathbf{F}(\lambda)]d\lambda$$

in the this expression only spectral reflectance $R(\lambda,\mathbf{x})$ depends on $\mathbf{x}$.

Let there exist such patches $\mathbf{u}_1 \subset \gamma_1$ and $\mathbf{u}_2 \subset \gamma_2$, that:

$$R(\lambda,\mathbf{x}_1) = R(\lambda,\mathbf{x}_2) = R(\lambda), \qquad (10)$$
$$\forall \mathbf{x}_1 \in \mathbf{u}_1, \quad \forall \mathbf{x}_2 \in \mathbf{u}_2$$

Taking (10) into account, the color inside the patch $\mathbf{u}$ depends only on this patch. Hence for two patches on different curves γ_1 and γ_2 it is possible to define a *color correction function F()*, which corrects illumination distortions on these curves:

$$\mathbf{p}_{\gamma_2}(\mathbf{u}_{\gamma_2}) = F(\mathbf{p}_{\gamma_1}(\mathbf{u}_{\gamma_1}), \mathbf{a}) \qquad (11)$$

where $\mathbf{a}$ is a parameter vector.

Let us specify the set of patches, for which the color correction function is defined (11), using the *color-shape units model*.

Color correction based on the color shape units model

Def. 2. The color shape unit.

A color shape unit consists of two units, which are shape and color:

$$C = \left\{ \begin{array}{l} \mathbf{p}(\mathbf{x}): \\ \mathbf{p}(\mathbf{x}) = \int R(\mathbf{x},\lambda)S(\mathbf{x},\lambda)\mathbf{F}(\mathbf{x},\lambda)d\lambda \end{array} \right\}, \qquad (12)$$
$$\mathbf{x} \in S \subset Z_2$$

Shape unit is a point set on a plane $\mathbf{x} \in S \subset R_2$ which forms a closed region.

Color unit is a point set C in color space, where each point corresponds to the shape unit S.

Color shape unit (CSU) is defined as implicative map of the shape unit's points into the color space:

$$u : \mathbf{x} \in S \to \mathbf{p}(\mathbf{x}) \in C \qquad (13)$$

Pixel of the CSU can be represented as a point in five dimensional space Υ with the following metric:

$$\left\| [\mathbf{x},\mathbf{p}(\mathbf{x})]^T, [\mathbf{y},\mathbf{p}(\mathbf{y})]^T \right\| = \|\mathbf{x},\mathbf{y}\| + \|\mathbf{p}(\mathbf{x}),\mathbf{p}(\mathbf{y})\| \qquad (14)$$

Metric (14) defines the distance between points of two CSU. Using this distance we can define the Hausdorff metric for CSU itself:

$$\|\mathbf{u}_i,\mathbf{u}_j\|_H = \max \left(\begin{array}{l} \max\limits_{\mathbf{x}\in\mathbf{u}_i} \min\limits_{\mathbf{y}\in\mathbf{u}_j} \left\| [\mathbf{x},\mathbf{p}(\mathbf{x})]^T, [\mathbf{y},\mathbf{p}(\mathbf{y})]^T \right. \\ \max\limits_{\mathbf{y}\in\mathbf{u}_j} \min\limits_{\mathbf{x}\in\mathbf{u}_i} \left\| [\mathbf{x},\mathbf{p}(\mathbf{x})]^T, [\mathbf{y},\mathbf{p}(\mathbf{y})]^T \right. \end{array} \right. \qquad (15)$$

There are two special cases of the Hausdorff metric: $\|\ ,\ \|_c$ is for color units and $\|\ ,\ \|_s$ for shape units:

$$\|\mathbf{u}_i,\mathbf{u}_j\|_c = \max \left(\begin{array}{l} \max\limits_{\mathbf{x}\in\mathbf{u}_i} \min\limits_{\mathbf{y}\in\mathbf{u}_j} \|\mathbf{p}(\mathbf{x}),\mathbf{p}(\mathbf{y})\|, \\ \max\limits_{\mathbf{y}\in\mathbf{u}_j} \min\limits_{\mathbf{x}\in\mathbf{u}_i} \|\mathbf{p}(\mathbf{x}),\mathbf{p}(\mathbf{y})\| \end{array} \right) \qquad (16)$$

$$\|\mathbf{u}_i,\mathbf{u}_j\|_s = \max \left(\max\limits_{\mathbf{x}\in\mathbf{u}_i} \min\limits_{\mathbf{y}\in\mathbf{u}_j} \|\mathbf{x},\mathbf{y}\|, \max\limits_{\mathbf{y}\in\mathbf{u}_j} \min\limits_{\mathbf{x}\in\mathbf{u}_i} \|\mathbf{x},\mathbf{y}\| \right) \qquad (17)$$

Using (16) and (17), the metric $\|\ ,\ \|_H$ (15) could be expressed by:

$$\|\mathbf{u}_i,\mathbf{u}_j\|_H = \|\mathbf{u}_i,\mathbf{u}_j\|_c + \|\mathbf{u}_i,\mathbf{u}_j\|_s \qquad (18)$$

Using (9), (10) and CSU definition, the color correction problem is formalized as:

$$\mathbf{a} = \arg\min \left\| \mathbf{p}(\mathbf{u}^i_{\gamma_2}), F(\mathbf{p}(\mathbf{u}^i_{\gamma_1}), \mathbf{a}) \right\| \qquad (19)$$

$$\{\mathbf{u}^i_{\gamma_1}\} \subset \gamma_1, \{\mathbf{u}^i_{\gamma_2}\} \subset \gamma_2 \qquad (20)$$

Metric in (19) is the Hausdorff metric (15).

The color correction function (19) transforms the set $U = \{u_i\}$ of CSU from the distorted

part of the image to another set $U^0 = \{u_i^0\}$ of CSU which are distortion free. Let us assume that this set is known *a priori*, details could be found in [6].

We define a value of color correction function for whole CSU as:

$$F(\mathbf{u},\mathbf{a}) = \{F(\mathbf{x}_i, p(\mathbf{x}_i),\mathbf{a})\},\quad \mathbf{x}_i \in \mathbf{u} \qquad (21)$$

Observe that the color correction function does not change the spatial position of CSU, and metric (15) takes the following form:

$$\|\mathbf{u}, F(\mathbf{u},\mathbf{a})\|_H = \|\mathbf{u}, F(\mathbf{u},\mathbf{a})\|_c = \qquad (22)$$

$$\max\left(\begin{array}{l}\max\limits_{\mathbf{x}\in\mathbf{u}}\min\limits_{\mathbf{y}\in\mathbf{u}}\|F(\mathbf{x},\mathbf{p}(\mathbf{x}),\mathbf{a}),\mathbf{p}(\mathbf{y})\|,\\ \max\limits_{\mathbf{y}\in\mathbf{u}}\min\limits_{\mathbf{x}\in\mathbf{u}}\|F(\mathbf{x},\mathbf{p}(\mathbf{x}),\mathbf{a}),\mathbf{p}(\mathbf{y})\|\end{array}\right)$$

Finally, the task of the color correction function identification takes the following form:

$$\mathbf{a} = \arg\min_{\{\mathbf{u}_i\},\{\mathbf{u}_i^0\}} \|F(\mathbf{u}_i,\mathbf{a}),\mathbf{u}_i^0\| \qquad (23)$$

where $\{\mathbf{u}_i\}$ is a set of distorted CSU, and $\{\mathbf{u}_i^0\}$ is a set of distortion free CSU.

The color correction curves, which are generally used by color correction specialists, are monotone functions [7]. For different color gradations in the original image corresponds to different gradations in corrected image. Mathematically, it is equal to the following rule:

$$p_k(x_1) < p_k(x_2) \Rightarrow F(p_k(x_1)) < F(p_k(x_2)) \qquad (24)$$
$$| F(p_k(x_1)) > F(p_k(x_2))$$

According to (9), the pixel's color in a CSU depends on $I(\lambda,\mathbf{x})$ and $R(\lambda,\mathbf{x})$. For each set of CSU $I(\lambda,\mathbf{x})$ is constant, accordingly to (10). Hence, for each CSU, color depends only on $R(\lambda,\mathbf{x})$. Therefore, to construct a correct set of CSU for solving (23), we need to construct matching between two sets of CSU. For this, it is necessary to choose pairs of CSU with close values of $R(\lambda,\mathbf{x})$.

A pair of two matched CSU (*i,j*) we define as:

$$(i,j): \|R(\lambda,\mathbf{x}_i), R(\lambda,\mathbf{x}_j)\| < t_R, \qquad (25)$$
$$\forall \mathbf{x}_i \in u_i, \forall \mathbf{x}_j \in u_j$$

It is possible to prove the following necessary condition of CSU matching for necessary form of color correction curve.

Theorem 1. The necessary condition for CSU matching

Necessary condition for (25) is the following

$$\forall p(\mathbf{x}_j^1) > p(\mathbf{x}_k^1) : F(p(\mathbf{x}_j^1)) > F(p(\mathbf{x}_k^1)) \qquad (26)$$

Correction function identification using CSU matching condition

In general case optimization with the use of the Hausdorff metric is difficult, non-convex and computationally complicated problem. This problem can be solved using an evolutionary optimization approach. For correction problem it can be solved for each color coordinate independently. Using (24) general task (23) takes the following form:

$$\|F(p(\mathbf{u}_1),\mathbf{a}), p(\mathbf{u}_0)\|_C = \qquad (27)$$
$$= max(|\max(F(p(\mathbf{u}_1))) - \max(p(\mathbf{u}_0))|,$$
$$|\min(F(p(\mathbf{u}_1))) - \min(p(\mathbf{u}_0))|)$$

taking into account the condition (26), (27) is equal to the following minimization problem:

$$\mathbf{a} = \arg\min_{\forall\{u_0,u_1\}}((\max(F(p(\mathbf{u}_1))) - \qquad (28)$$
$$-\max(p(\mathbf{u}_0)))^2,$$
$$(\min(F(p(\mathbf{u}_1))) - \min(p(\mathbf{u}_0)))^2)$$

So, we need only two points min(p(**u**)), max(p(**u**)) to describe each CSU. Matching (25) is performed as follows:

$$\{p_0,p_1\}_k, \{p_0,p_1\}_{k+1} \in P \longleftrightarrow \qquad (29)$$
$$\longleftrightarrow p_{0,k} = \min(\mathbf{u}_{0,c}^i),$$
$$p_{0,k+1} = \min(\mathbf{u}_{0c}^i),$$
$$p_{1,k} = \min(\mathbf{u}_{1,c}^i),\quad p_{1,k+1} = \min(\mathbf{u}_{1,c}^i)$$

Finally, identification problem of color correction function using condition of theorem 1 takes the following form:

$$\mathbf{a} = \arg\min_{\{p_0,p_1\}\in P}((F(p_1,\mathbf{a}) - p_0)^2) \qquad (30)$$
$$F'(p_1,\mathbf{a}) \ge 0$$

The second condition written for monotonically non-decreasing correction function, for non-increasing function the sign in inequality must be inverted.

For three-order polynomial fitting of function *F* expression (30) takes the form:

$$\mathbf{a} = \arg\min_{\{p_0,p_1\}\in P}(a_3p_1^3 + a_2p_1^2 + a_1p_1 + a_0 - p_0)^2 \qquad (31)$$

$$3a_3p_1^2 + 2a_2p_1 + a_1 \ge 0 \qquad (32)$$

This is the convex minimization problem described by Kuhn — Tucker theorem conditions, and it can be solved as non-negative LSM problem.

Outcome of the experiment

As an example of the color correction based on identification (31) – (32) we consider shadow removing from real image, shown in fig. 1.

Fig. 1. Figure caption.

Correction function identified by this CSU matching provides the color correction which is shown in Fig 3. The correction result is not accurate enough.

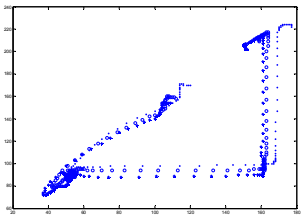


Fig. 2. Figure caption.

Correction function identified by this CSU matching provides the color correction which is shown in Fig 3. The correction result is not accurate enough.

Fig. 3. Figure caption.

Using expressions (31)-(32) as a model for RANSAC optimization and the by ordinate matching as initial set we achieve the reduced set of matched CSU. The resulting set is shown in Fig. 4.

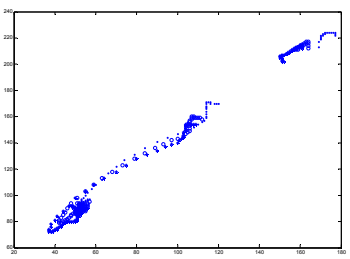


Fig. 4. Figure caption.

Correction result for this set of CSU matching is shown in fig 5. The quality is higher than in fig. 3.

Fig. 5. Figure caption.

Conclusion

In this paper a new basic approach to the distortion correction in color images based on the color shape units model is presented. Theoretical grounds and numerical methods are analyzed. The correction results on real images are shown.

References

1. Jadd D., Wyszecki G., Color in Business, Science, and Industry, 1978, 576p.
2. H.D.Cheng, X.H.Jiang, Y.Sun, J.Wang, Color image segmentation: advances and prospects, Pattern Recognition, Vol.34, Is.12, 2001, pp. 2259–2281
3. Brainard D. H., Wandell B. A., Analysis of the retinex theory of color vision, J. Opt. Soc. Am., Vol. 3, No. 10, pp 1651-161, 1986.
4. A. Gijsenij, T. Gevers, J. van de Weijer, Computational Color Constancy: Survey and Experiments, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. X, NO. X, 2010, pp. 1-15
5. Salvador E., Shadow segmentation and tracking in real-world conditions, PhD Thesis, EPFL, 2004, 194p.
6. S.A. Bibikov, V.A. Fursov, A.V. Nikonorov, Shadow Artifacts Correction on Fine Art Reproductions, IMTA 3, 2010, Angers, France, May 2010. – pp. 3-12.
7. Margulis D., Modern Photoshop Color Workflow, MCW Publishing, 2013, 480p.

TWO-DIMENSIONAL FAST FOURIER TRANSFORM ALGORITHM MODIFICATION BY ANALOGUE OF COOLEY-TUKEY FOR RECTANGULAR SIGNAL

M. Noskov[1,2], V. Tutatchikov[1,3]

[1]Institute of Space and Information Technology Siberian Federal University, 26 Kirenskogo Street, Krasnoyarsk, 660074, Russia, [2]mvnoskov@yandex.ru, [3]vtutatchikov@mail.ru

In the article the algorithm of two-dimensional fast Fourier transform for rectangular signal, which is the two-dimensional analogue of the Cooley-Tukey algorithm, is given. The results of numerical experiments are presented.

Introduction

The discrete Fourier transform (DFT) has several important applications owing to the existence of effective algorithms of the DFT calculation. For example, the DFT can be used for spectral analysis of many-dimensional signals (space imaging, image processing).

In the paper algorithms of the DFT calculation are examined. The algorithms presented are distinctive because of their computational complexity: the two-dimensional DFT calculation by a method of partitioning into rows and columns, which are calculated by means of the one-dimensional fast Fourier transform (FFT), and the two-dimensional FFT by analogue of the Cooley-Tukey algorithm.

The algorithm description

Let us have a look at the signal f, which is a two-dimensional periodic signal with a period of 2^s by the first coordinate and $2^{s+\vartheta}$ by the second coordinate. The counts are given as $f_{k,t}$, where $k = 0:2^s$, $t = 0:2^{s+\vartheta}$.

The discrete Fourier transform for the signal f is given in the formula:

$$F_{l,m} = \sum_{k=0}^{2^s-1}\sum_{t=0}^{2^{s+\vartheta}-1} f_{k,t} e^{\frac{2\pi imt}{2^{s+\vartheta}}} e^{\frac{2\pi ilk}{2^s}} \qquad (1)$$

The two-dimensional DFT can be calculated by means of one-dimensional DFT combinations. For this reason F is calculated as follows [1]:

$$F_{l,m} = \sum_{k=0}^{2^s-1}\left[\sum_{t=0}^{2^{s+\vartheta}-1} f_{k,t} e^{\frac{2\pi imt}{2^{s+\vartheta}}}\right] e^{\frac{2\pi ilk}{2^s}} \qquad (2)$$

The sum in the square brackets is the one-dimensional DFT calculation, for example, in rows, the outer sum, therefore, is the one-dimensional DFT calculation in columns. Let us transform the formula by partitioning the second coordinate into even and odd components [2]:

$$\begin{aligned} F_{l,m} &= \sum_{t=0}^{2^{s+\vartheta}-1}\left[\sum_{k=0}^{2^s-1} f_{k,t} e^{\frac{2\pi ilk}{2^s}}\right] e^{\frac{2\pi imt}{2^{s+\vartheta}}} = \\ &= \sum_{t_1=0}^{2^{s+\vartheta-1}-1}\left[\sum_{k=0}^{2^s-1} f_{k,2t_1} e^{\frac{2\pi ilk}{2^s}}\right] e^{\frac{2\pi im2t_1}{2^{s+\vartheta}}} + \\ &+ \sum_{t_1=0}^{2^{s+\vartheta-1}-1}\left[\sum_{k=0}^{2^s-1} f_{k,2t_1-1} e^{\frac{2\pi ilk}{2^s}}\right] e^{\frac{2\pi im(2t_1-1)}{2^{s+\vartheta}}} = \\ &= \sum_{t_1=0}^{2^{s+\vartheta-1}-1}\left[\sum_{k=0}^{2^s-1} f_{k,2t_1} e^{\frac{2\pi ilk}{2^s}}\right] e^{\frac{2\pi imt_1}{2^{s+\vartheta-1}}} - \\ &- \sum_{t_1=0}^{2^{s+\vartheta-1}-1}\left[\sum_{k=0}^{2^s-1} f_{k,2t_1-1} e^{\frac{2\pi ilk}{2^s}}\right] e^{\frac{2\pi imt_1}{2^{s+\vartheta-1}}} e^{\frac{\pi im}{2^{s+\vartheta-1}}} = \\ &= f^1_{k,2t_1} - e^{\frac{\pi im}{2^{s+\vartheta-1}}} f^1_{k,2t_1-1} \end{aligned} \qquad (3)$$

where

$$\begin{aligned} f^1_{k,2t_1} &= \sum_{t_1=0}^{2^{s+\vartheta-1}-1}\left[\sum_{k=0}^{2^s-1} f_{k,2t_1} e^{\frac{2\pi ilk}{2^s}}\right] e^{\frac{2\pi imt_1}{2^{s+\vartheta-1}}}, \\ f^1_{k,2t_1-1} &= \sum_{t_1=0}^{2^{s+\vartheta-1}-1}\left[\sum_{k=0}^{2^s-1} f_{k,2t_1-1} e^{\frac{2\pi ilk}{2^s}}\right] e^{\frac{2\pi imt_1}{2^{s+\vartheta-1}}} \end{aligned} \qquad (4)$$

$f^1_{k,2t_1}$ - two-dimensional even and $f^1_{k,2t_1-1}$ - two-dimensional odd subsignals of the

$2^s : 2^{s+\vartheta-1}$ dimension. Sequentially applying the procedure (3) of partitioning the second coordinate into even and odd components for subsignals (4),), we will find the one-dimensional FFT of the length 2^ϑ signal over the final subsignals (5) of the $2^s : 2^s$ dimension. For these the two-dimensional FFT algorithm by analogue of the Cooley-Tukey algorithm has already been described [3]:

$$f^{\vartheta}_{k,2t_\vartheta} = \sum_{t_\vartheta=0}^{2^s-1}\left[\sum_{k=0}^{2^s-1} f_{k,2t_\vartheta} e^{\frac{2\pi i l k}{2^s}}\right] e^{\frac{2\pi i m t_\vartheta}{2^s}},$$
$$f^{\vartheta}_{k,2t_\vartheta-1} = \sum_{t_\vartheta=0}^{2^s-1}\left[\sum_{k=0}^{2^s-1} f_{k,2t_\vartheta-1} e^{\frac{2\pi i l k}{2^s}}\right] e^{\frac{2\pi i m t_\vartheta}{2^s}} \quad (5)$$

Let us calculate the total amount of operations. The procedure (3) takes $2^{\vartheta-1}\vartheta$ operations of complex multiplications and $2^{\vartheta+1}\vartheta$ complex addition. The two-dimensional FFT by analogue of the Cooley-Tukey algorithm over the final signal (5) of the $2^s : 2^s$ dimension requires $3 \cdot 2^{2s-s} s$ operations of complex multiplications and $2^{2s+1} s$ operations of complex addition. As a result, the total number of operations needed for processing the initial signal $f_{k,t}$ where $k = 0 : 2^s$, $t = 0 : 2^{s+\vartheta}$ is: $3 \cdot 2^{2s+\vartheta-3}(s+\vartheta)$ operations of complex multiplications and $2^{2s+\vartheta+1}(s+\vartheta)$ operations of complex addition. If we calculate the FFT of the initial signal $f_{k,t}$ by partitioning into rows and columns, the number of operations required will be: $2^{2s+\vartheta-1}(2s+\vartheta)$ of complex multiplications and $2^{2s+\vartheta-3}(2s+\vartheta)$ operations of complex addition.

The obtained results

For the algorithm testing the program in the programming language C++ with using of the MPI library has been written. The testing was conducted on PC with following characteristics:

- Processor: Intel Core 2 Duo CPU T8100 2.1 GHz;
- RAM: 2 GB;
- Operating system: Windows XP Service Pack 3.

The presented algorithms working time is given in table 1. The initial signal example is given in picture 1.

Table 1 – The two-dimensional FFT working time for a rectangular signal, sec

Height	Width	FFT RC	FFT KT
1024	2048	1.105	0.666
1024	4096	2.414	1.413
1024	8192	6.057	3.848

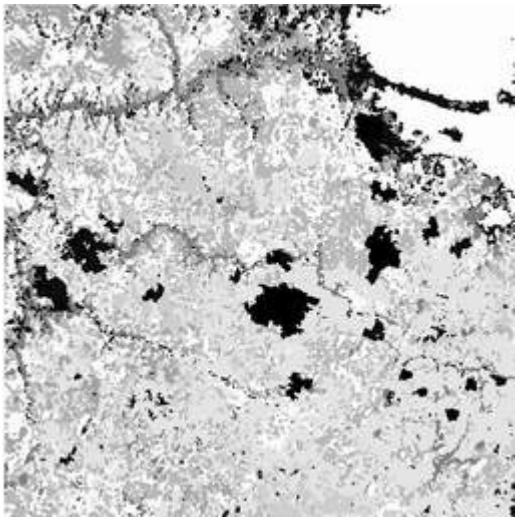


Fig 1 – Initial signal example

Conclusion

The modified algorithm of the two-dimensional FFT by analogue of the Cooley-Tukey algorithm requires fewer complex operations of addition and multiplication and is faster than its analogue of the two-dimensional FFT calculation by rows and columns.

References

1. Pissis S. Parallel Fourier Transformations using shared memory nodes. MSc in High Performance Computing. The University of Edinburgh Year of Presentation: 2008;
2. Starovoytov A.V., About multidimensional analog of algorithm of Cooley-Tukey, The bulletin of the Siberian state space university of a name of academician M.F.Reshetnev, issue 1 (27), 2010. Pages 69-73 (In Russian);
3. Tutatchikov V.S., About calculation of two-dimension fast Fourier transform, 8-th Open German-Russian Workshop "PATTERN RECOGNITION and IMAGE UNDERSTANDING" ORGV-8-2011. Pages 219-291.

CONTOUR ANALYSIS AND IMAGE SUPERIMPOSITION TASK IN COMPUTER VISION SYSTEMS

A. I. Novikov[1,2], V. A. Sablina[1,3], M. B. Nikiforov[1,4], A. A. Loginov[1,5]

[1] Ryazan State Radio Engineering University, 390005, Ryazan, Gagarina 59/1
[2]phone: 8-906-542-29-34, e-mail: novikovanatoly@yandex.ru
[3]phone: 8-920-638-95-79, e-mail: sablina.v.a@evm.rsreu.ru
[4]phone: 8-910-907-75-07, e-mail: nikiforov.m.b@evm.rsreu.ru
[5]phone: 8-910-507-44-32, e-mail: loginal@mail.ru

The contour analysis application possibility for image superimposition task solution in multispectral computer vision systems is investigated. This issue is considered for both images from sensors and images synthesized on the basis of the digital terrain map. To solve the designated main task and a number of the auxiliary tasks the original algorithms are suggested. The initial real images processing and their subsequent superimposition examples are represented.

Introduction and problem definition

The intensive scale increase of the freight and passenger carriage by air makes extremely important the flight accident prevention problem. This problem solution is especially urgent for airport areas where an air picture is the most complex. The computer technologies state-of-the-art, mathematical image processing methods elaboration degree for the images from multispectral sensors in computer vision systems make possible to solve multidisciplinary aircraft navigation tasks in the real-time mode [1, 2]. One of the key tasks consists in the aircraft precise space coordinates determination. This task is more urgent in the aircraft landing stage. It is known that existing technologies give aircraft coordinates with errors. These errors can be estimated and therefore aircraft coordinates can be corrected using the superimposition of the observed image and the corresponding image fragment synthesized from the digital map.

The known image superimposition methods pertain to correlation algorithms group [3]. Sufficiently high reliability of these methods is attained at the price of time-consuming operations. This problem practically excludes the possibility of the correlation algorithms application in real aircraft on-board systems. In the issue the new superimposition methods search task for the images of various natures appears. These methods have to provide considered task solution in real-time.

One of the possible solution methods search directions for assigned task comes to finding in the image pair some similar objects set, their parameters calculating, one image to another image plane geometric transformation construction on this basis. The known search methods of such objects on the basis of the whole image analysis also require time-consuming operations and at the same time can't provide high reliability.

It is suggested to construct objects comparison methods on the basis of their contours analysis [4]. The main objects contours remain invariable except for existing geometric deformation. Therefore contour analysis methods investigation with regard to contours comparison and superimposition in the aerial photographs is of a certain interest. Besides such approach to the image superimposition gives the real-time algorithms development possibility so long as vector contours processing algorithms have lower computational complexity than raster image processing algorithms.

The main image pair superimposition task solution stages on the basis of contour analysis are:

– initial image processing and edge detection;
– closed contours obtaining and not enough informative lines removing;
– closed contours vector description;
– initial superimposition by transformation in complex plane;
– corresponding points detection on the compared contour pair;
– image superimposition.

Algorithms approbation is carried out with visible image pair. Each image of a pair contains the same islands but there is time delay between images registrations.

Tasks solution algorithms

It is suggested to perform image smoothing and edge detection using the algorithms which real images processing examples are represented in [5]. The second task solution, viz. closed contours obtaining and not enough informative lines removing, is attained by standard morphological processing procedures. Therefore these two algorithmic stages are excluded from the consideration.

The next stage is transition from binary raster descriptions for extracted closed contours of main objects in images to vector descriptions. For this purpose authors developed approximating polygon finding algorithm. This algorithm is based on approximating polygon vertices selection using local maxima of angle cosine estimations for these vertices. The remaining closed contour points are removing. Angle cosine estimations are calculating using the certain number of neighbor points on both sides. And the number of such points is algorithm adjustable parameter. This parameter gives the control of the detailing degree of the contour approximation by polygon. The algorithm results for two different values of this parameter are shown in Fig. 1.

Further this algorithm is considered in brief. For each contour point the cosine estimation is calculated

$$\cos\alpha = \frac{p_1 p_2 + \hat{k}_1' \hat{k}_2'}{\sqrt{1+\hat{k}_1^2}\sqrt{1+\hat{k}_2^2}}$$

for the angle between the lines $L_1 : y = \hat{k}_1 x$ and $L_2 : y = \hat{k}_2 x$ passed through this point and m pixels on both sides of this point. Angle coefficient for each line can be found $\hat{k} = \sum_{i=1}^{m} y_i x_i \Big/ \sum_{i=1}^{m} x_i^2$. Here $(x_i ; y_i)$ is points $A_i, i = \overline{1,m}$ coordinates in moving coordinates Oxy for angle coefficient k_1 and $(x_i ; y_i)$ is points $B_i, i = \overline{1,m}$ for angle coefficient k_2. Coefficients p_1 and p_2 possess the value 1 or –1 depending on sum $\sum_{i=1}^{m} x_i$ sign. In concord with the changes of these coefficients signs the coefficients k_1 and k_2 signs also change.

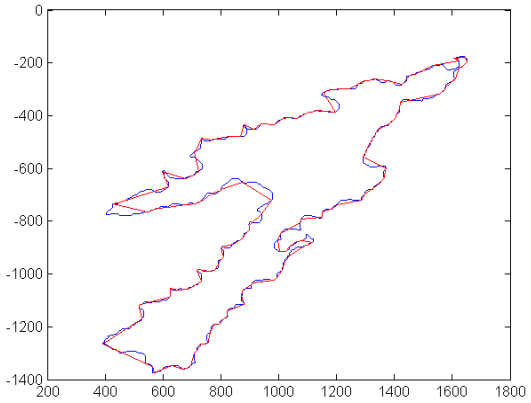


Fig. 1. Two closed contour vector description variants for the islands in the first test image

In the given case approximating polygon finding algorithm makes possible to reduce the number of vertices for the closed islands contours from approximately 5 thousands (all points without exception) to approximately 6 hundreds (blue contours in Fig. 1) or to approximately 2 hundreds (red contours in Fig. 1). The enlarged fragment for contours in Fig. 1 is shown in Fig. 2. The blue contour describes the shore line of the island in detail and is calculated using 3 neighbor points on the both sides of each vertex, and the red contour describes the shore line more schematic and is calculated using 350 neighbor points on the both sides of each vertex.

The fourth algorithmic stage consists in the initial superimposition of two images by the complex plane transformation of the second image points to the first image. This algorithm is based on the transformation $z_k^{(2)} = z_{np} \cdot z_k^{(1)}$, $z = x + iy$ where $z_k^{(1)}$ and

$z_k^{(2)}$ are the points on the original image and the second image superimposed with the first respectively, $z_{np} = x_{np} + iy_{np}$ is the desired complex plane transformation. To find complex number z_{np} it is necessary to search out a pair of corresponding points on each of the both contours. In the considered algorithm variant the points belonging to the area diameter ends is selected as such points, i.e.

$$\{M_1, M_2\} = \arg\max_{M_i, M_j \in \partial D} \rho(M_i, M_j).$$

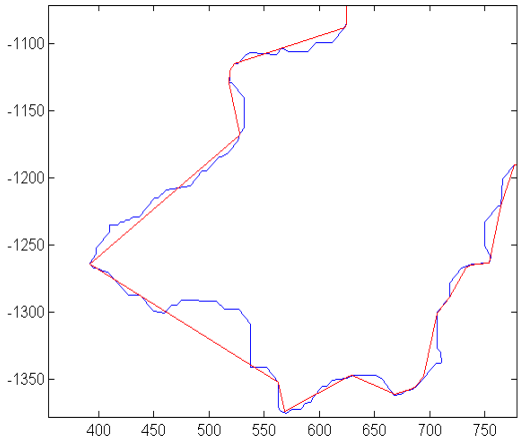


Fig. 2. Closed contour vector description of the island in detail and more schematic

The points M_1, M_2 form the vector $\bar{a} = (x_2 - x_1; y_2 - y_1)$ which coordinates are the real and the imaginary parts of the complex number

$$z_{np} = (x_2 - x_1) + i \cdot (y_2 - y_1)$$

respectively.

The considered superimposition algorithm result for the islands contours is shown in Fig. 3.

In Fig. 3 it is observed noticeable divergence between the superimposed contours. It can be explained by projective distortions which can't be taken into account within the considered algorithm. However this algorithm has high performance and so can be used for the initial contours superimposition.

For more qualitative contours superimposition problem solving the further schematization of the islands contours is done at the expense of removing the vertices with angles which are close to the straight angle. The vertices with angle cosines lower than –0.4 are thrown out. Finally it is succeeded to decrease the number of approximating polygon vertices to approximately 30.

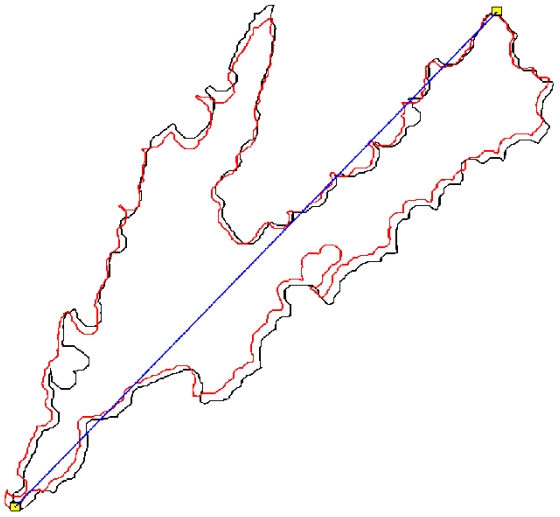


Fig. 3. The contours superimposition result in the complex plane

The next fifth algorithmic stage has to provide corresponding points detection on the compared contour pair. As far as the contours might have no global similarity (e.g. the part of an object in an image is hidden) it is required to find and to compare with each other the similar fragments of the contours. To do this it is necessary to estimate such local similarity with a certain measure. For this purpose it is suggested to generate the auxiliary subcontour in the neighborhood for each of the remaining vertices.

Then each subcontour of one contour is compared to all of the subcontours of the other contour for the purpose of finding the similar fragments of the existing contours. As the similarity measure for subcontours it is suggested to use autocorrelation function in combination with additional parameters of the vertex. It is suggested to calculate the autocorrelation function as follows. Let

$$\Gamma = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$

is coordinate description of the contour Γ Then autocorrelation function represents a vector $\bar{\tau}$ which components are

$$\tau_k = \frac{|(\Gamma_k, \Gamma_0)|}{(\Gamma_0, \Gamma_0)}, \; k = \overline{1, n-1}$$

where

$$(\Gamma_k, \Gamma_0) = \sum_{m=1}^{n} (x_m + iy_m)(x_{k+m(\mathrm{mod}\, n)} - iy_{k+m(\mathrm{mod}\, n)}).$$

Autocorrelation function $\bar{\tau}$ is calculated for

each of the constructed subcontours. Thereupon similarity measure is calculated between each subcontour γ_m belonging to the second contour and some fixed subcontour γ_0 belonging to the first contour

$$\Delta_m = \sum_{k=1}^{[n/2]} \left| \tau_k^{(m)} - \tau_k^{(0)} \right|.$$

Then values Δ_m are compared with a certain specified threshold ε. Subcontour γ_m is included in the set of contours which are close to subcontour γ_0 if $\Delta_m < \varepsilon$, and it is excluded from this set otherwise.

The final selection among the close subcontours is performed using main contours comparable vertices additional parameters such as vertex angle, mutual position in comparison with the other contour vertices.

In the last algorithmic stage the corresponding point pairs selected in the previous stage is used for projective transformation search. This transformation serves the contours and the entire images superimposition.

The contours superimposition using the complex contour analysis methods makes possible to take into account three kinds of affine image transformation (the shift, the rotation and the scaling of the image) [6]. However at the same time the projective deformations aren't taken into account. The influence of these deformations can be taken into account by the homography matrix

$$H = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{31} & 1 \end{pmatrix}$$

which elements h_{ij} are calculated from the corresponding points coordinates selected in the previous stage. The superimposed vector islands contours together with the selected four corresponding point pairs are shown in Fig. 4. Their superimposition result using the homography matrix is shown in Fig. 5.

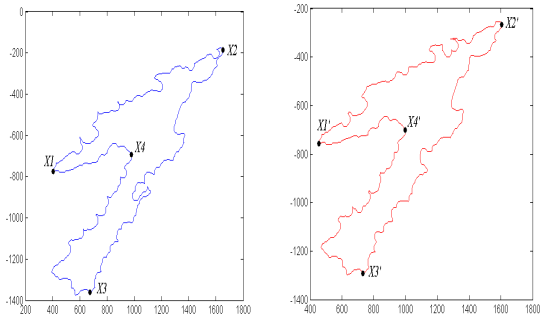


Fig. 4. The first and the second contours for the superimposition with the selected supporting points

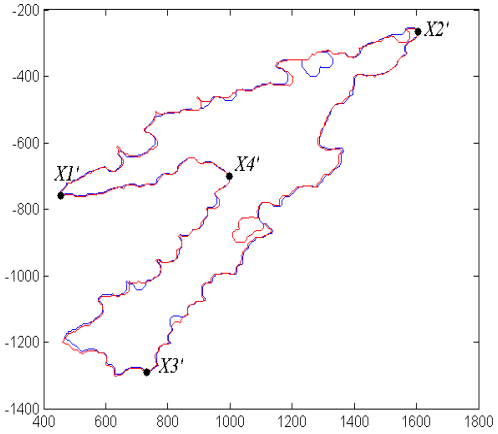


Fig. 5. The contours superimposition result using the homography matrix

References

1. 1. David A. Forsyth and Jean Ponce, Computers Vision: A Modern Approach. New York: Prentice-Hall, 2011. 792 p.
2. 2. Rafael C. Gonzales and Richard E. Woods, Digital Image Processing, 3nd ed. New Jork: Prentice-Hall, 2008. 954 p.
3. 3. Vjatcheslav K. Baklitski, Correlation Navigation and Aiming Methods. Tver, Knizhnyj klub, 2009. 360 p.
4. 4. Yakov A. Furman (ed.) Contour Analysis Introduction and its Image and Signal Processing Application. Moscow: Nauka, 2002. 592 p.
5. 5. Victoria A. Sablina, Anatoly I. Novikov, Michael B. Nikiforov, and Alexander A. Loginov "An Approach to the Superimposition Problem in Multispectral Computer Vision Systems", 2nd Mediterranean Conference on Embedded Computing (MECO) Proceedings. Budva, Montenegro, 2013. pp. 117-120.
6. 6. Alexander A. Loginov, Anatoly I. Novikov, Victoria A. Sablina, and Olga V. Shcherbakova "Complex Contour Analysis Feasibility Study of Contour Classifiation and Superimposition Problems", RSREU Bulletin № 1 (Issue 43). Ryazan, 2013. pp. 20-24.

CONNECTED ATTRIBUTE FILTERING BASED ON CONTOUR SMOOTHNESS

G. K. Ouzounis[1], E. R. Urbach[2], M. H. F. Wilkinson [3]

[1] Digital Globe Inc.,
1601 Dry Creek Drive, Suite 260, Longmont, CO 80503, USA.
georgios.ouzounis@digitalglobe.com
[2] University of Groningen PO Box 83, Croydon, NSW 2132, Australia,
erikurbach@yahoo.com
[3] Johann Bernoulli Institute for Mathematics and Computer Science,
University of Groningen, PO Box 407, 9700 AK Groningen, Netherlands.
m.h.f.wilkinson@rug.nl

A new attribute measuring the contour smoothness of 2-D objects is presented in the context of morphological attribute filtering. The attribute is based on the ratio of circularity to non-compactness and has a maximum of 1 for a perfect circle. It decreases as the object boundary becomes irregular. Computation on hierarchical image representation structures relies on five auxiliary data members and is rapid. Contour smoothness is a suitable descriptor for detecting and discriminating man-made structures from other image features. An example is demonstrated on a very-high-resolution satellite image using connected pattern spectra and the switchboard platform.

Introduction

In image analysis, the detection of man-made structures remains a challenge due to the limited descriptors available to distinguish them from other image features. Linearity is defined as the ratio between the principal axes of image components from its PCA analysis. It can distinguish artificial structures from others but can easily be misled for example by a row of adjacent trees, which has a high linearity measure too. Corners [1,2] are good indicators of artificial structures but do not comply with the framework of connected image analysis and moreover, are very sensitive to the image radiometry. Other shape-based descriptors like rectangularity or triangularity [3] are derived from moments of inertia and describe best fitted shapes rather than actual component attributes, i.e. suitable for post-processing.

In this paper a new component attribute is proposed measuring directly the smoothness of connected component contours. The proposed attribute is utilized by morphological connected attribute filters [4] that are discussed next. Following this, the definition and derivation of contour smoothness is given along with an experiment in detecting man-made structures form very high resolution satellite imagery. A brief discussion and conclusions are given at the end of the paper.

Image content organization

The organization of the image information content into connected regions or components is a well-established strategy in image analysis [5] allowing for efficient and edge preserving processing. Attribute filters [4] are morphological operators deciding on accepting or rejecting connected components based on some attribute measure. They find use in a wide range of applications [6] and can be implemented efficiently on hierarchical image representation structures like the Max-Tree [7] and the Alpha-Tree [8], or directly based on the Union-Find algorithm [9] or other similar component labeling schemes. An important tool based on attribute filters is the pattern spectrum [10] which is a multi-dimensional attribute histogram or feature space. The switchboard platform [11] is a wrapper of the pattern spectrum allowing for real-

time interaction with the image space via the tree data-structure used for image representation.

Contour smoothness

The proposed measure of contour smoothness is a ratio of two geometric descriptors; the circularity and non-compactness of a connected component. As the contour of a connected component becomes rougher the circularity reduces while the non-compactness increases thus the ratio decreases along. Contour smoothness has a maximum of 1 for a circle.

The circularity of a component C is a normalized measure of compactness:

$$cicularity = \frac{P^2}{4\pi A}, \qquad (1)$$

in which P is the perimeter and A is the area (pixel count) of C. Circularity has a maximum of 1 for a perfect circle and reduces as the shape deviates from a circle.

The non-compactness is a scale invariant measure given by:

$$non-compactness = \frac{2\pi I}{A^2}, \qquad (2)$$

in which I is the moment of inertia of a connected component given by:

$$I = \sum_C (x-\bar{x})^2 (y-\bar{y})^2 \qquad (3)$$

The terms x, y are pixel coordinates and $(\bar{x}, \bar{y})$ are the coordinates of the component's centroid:

$$\bar{x} = \frac{1}{A}\sum_C x \text{ and } \bar{y} = \frac{1}{A}\sum_C y. \qquad (4)$$

The term 2π in (2) is to normalize the attribute to a minimum value of 1 for a perfect circle. To improve scale invariance of (3) a further term is added that accounts for the pixel shape; pixels are assumed to be unit squares, each with a moment of inertia equal to:

$$\int_{-0.5}^{0.5}\int_{-0.5}^{0.5} (x^2 + y^2)\, dy\, dx = \frac{1}{6} \qquad (5)$$

as in [12]. Expanding (3) further and correcting for the square shape of pixels we obtain:

$$I = \sum_A x^2 + \sum_A y^2 - \frac{\left(\sum_\alpha x\right)^2}{A} - \frac{\left(\sum_\alpha y\right)^2}{A} + \frac{A}{6}. \qquad (6)$$

Definition. The contour smoothness cs of a connected component is given by:

$$cs = \frac{\frac{P^2}{4\pi A}}{\frac{2\pi I}{A^2}} = \frac{A P^2}{8\pi^2 I}. \qquad (7)$$

The contour smoothness measure is scale invariant but sensitive to the definition of the perimeter measure P. A simple definition for P is the count of all pixels of a connected component that border (4-way adjacency) a pixel of different intensity. Perimeter measures like the city-block or contour pixel counts based on 8-way adjacency require further corrections in (7). The computation of (7) can be done incrementally by the aid of 5 auxiliary data members; pixel count, perimeter pixel count, sum of x, sum of y, and sum of squares ($x^2 \square y^2$). The same pool of auxiliary data allows for a wider range of attributes to be computed and offers a robust connected-component characterization.

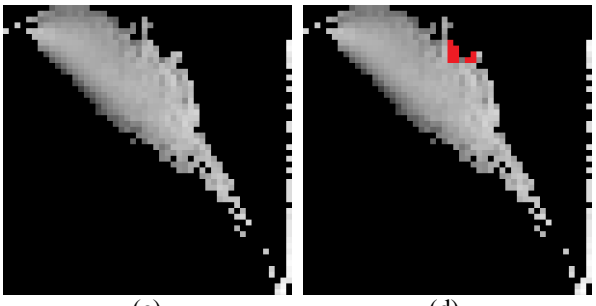
Experiments

The objective of the experiment presented in this section is to extract man-made structures from satellite imagery that are characterized by a varying degree of size and contour smoothness. To exploit the image information content efficiently a Max-Tree structure is computed from the input image shown in Fig.1.(a). All auxiliary data needed for the computation of component attributes are collected during the construction of the tree.

Once the tree is constructed, a connected-pattern spectrum [10,14] is computed by a simple pass through the structure. This is a

(a)

(b)

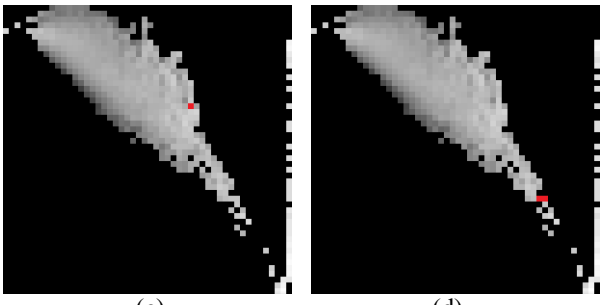


(c) (d)

Fig.1. (a) A sample panchromatic image tile of an urban area by the WorldView 2 sensor (©DigitalGlobe Inc.). (b) Image segmentation based on contour smoothness using the switchboard platform on the pattern spectrum in (c); selected switches are shown in (d).

(a)

(b)



(c) (d)

Fig.2. (a-b) Image segmentations of the image in.(a) based on contour smoothness using the switchboard platform on the pattern spectrum in.(c); selected switches are shown in (c) and (d) respectively.

normalized attribute histogram obtained by measuring the contribution of all connected components represented by the tree as a function of their attribute value, to the respective bins. Fig.1.(c) shows the connected pattern spectrum of (a), in which the x-axis corresponds to the size attribute (logarithmic mapping) and the y-axis to the contour smoothness attribute (logarithmic mapping). Each bin intensity is a normalized value [0-255] of the area of all components satisfying the bin extrema conditions multiplied by each

component's lifetime. The origin is at the top left of each plot; the first bin corresponds to all components having the smallest area and smallest contour roughness (inverse of smoothness) in the decomposition. The pattern spectrum is interfaced by the switchboard platform [11] that allows for interactive and real time image segmentation. Fig.1.(b) and Fig.2.(a),(b) show segmentation instances based on the respective switch selections marked with red on each pattern spectrum.

Discussion

The previous experiment demonstrated that contour smoothness along with size, offers a good approximation of man-made structures in very high resolution satellite imagery. That is due to the fact that most built-up appears with smooth contours by contrast to non-man-made objects. Thresholds on contour smoothness are subject to the image information content, and as shown in Fig.2.(b) beyond a certain value range, segmented objects become meaningless.

Contour smoothness is a non-increasing attribute [4] and so is the respective attribute filter. For the latter, when computed on a hierarchical image representation structure, special filtering rules apply. Contour smoothness filters implemented on the Max-Tree [7] require the subtractive rule [10]; if implemented on the Alpha-Tree [8] they require the max-rule [8].

The formalization of contour smoothness as a function of five auxiliary data members, allows it to be computed incrementally during the tree construction or in a single pass through the image following tree construction. Typical timing for a panchromatic image with a raster size of 10,000×20,000 pixels is 2.83 s. on a 64bit Intel Core i7-2760QM CPU@2.40GHz machine with 4GB of RAM. Using the parallel algorithm in [13,14] this can be reduced to less than one second on four cores.

Conclusion

A new attribute measuring the contour smoothness of connected components was presented. Contour smoothness can be used to discriminate man-made structures from other image features and an example was demonstrated. Future work includes methods for automatic threshold selection based on connected component statistics.

References

1. M. Pesaresi, A. Gerhardinger, F. Kayitakire. A robust built-up area presence index by anisotropic rotation-invariant textural measure. // IEEE JSTARS - 2008. - Vol. 1, No. 3. - P. 180-192.
2. C. Harris. M. Stephens. A combined corner and edge detector. // Proc. 4th Alvey Vision Conference. - 1988. - P. 147-151.
3. P. L. Rosin. Measuring Shape: Ellipticity, Rectangularity, and Triangularity. // Machine Vision and Applications. - 2003. - Vol.14, No. 3. - P. 172-184.
4. E. J. Breen. R. Jones. Attribute openings, thinnings, and granulometries. // Computer Vision and Image Understanding. - 1996, - Vol.64, No. 3. - P. 377-389.
5. P. Salembier. J. Serra. Flat zone filtering, connected operators and filters by reconstruction. // IEEE TIP. - 1995. - Vol. 4, No. 8. – P. 1153-1160.
6. P. Salembier, M. H. F. Wilkinson. Connected operators: a review of region-based morphological image processing techniques. // IEEE Signal Processing Magazine. - 2009. - Vol. 6. - P. 136-157.
7. P. Salembier. A. Oliveras. L. Garrido. Anti-extensive connected operators for image and sequence processing. // IEEE TIP. - 1998. - Vol.7, No. 4. - P. 555-570.
8. G.K. Ouzounis, P. Soille. The Alpha-Tree algorithm. //Publications Office of the European Union. - 2012.
9. A. Meijster. M. H. F. Wilkinson. A comparison of algorithms for connected set openings and closings. // IEEE TPAMI. - 2002. - Vol. 24, No. 4. - P. 484-494.
10. E. R. Urbach, J. T. B. Roerdink. M. H. F. Wilkinson. Connected shape-size pattern spectra for rotation and scale invariant classification of gray-scale images. // IEEE TPAMI. - 2007, Vol.29, No. 2. - P. 272-285.
11. G. K. Ouzounis. V. Syrris. L. Gueguen. P. Soille. The switchboard platform for interactive image information mining. // ESA-EUSC-JRC Proc. 8th Int. Conf. Image Information Mining. - 2012. P. 26-30.
12. M. A. Westenberg, J. T. B. Roerdink. M. H. F. Wilkinson. Volumetric attribute filtering and interactive visualization using the max-Tree representation. // IEEE TIP. - 2007. - Vol.16, No.12. - P.2943-2952.
13. M. H. F. Wilkinson, H. Gao. W. H. Hesselink, J. E. Jonker. A. Meijster. Concurrent computation of attribute filters on shared memory parallel machines. //IEEE TPAMI. - 2008, - Vol. 30, No. 10, - P. 1800-1813.
14. M. H. F. Wilkinson, U. Moschini. G. K. Ouzounis, M. Pesaresi. Concurrent computation of connected pattern spectra for very large image information mining. // ESA-EUSC-JRC Proc. 8th Int. Conf. Image Information Mining. - 2012. P. 21-25.

USING CUDA TECHNOLOGY FOR HIGH-DIMENSIONAL IMAGE FILTERING[1]

V. Panishchev[2,3], V. Sharkovskiy[2,4]

[2]South-Western State University, Russian Federation
[3]gskunk@yandex.ru, [4]sharkoon13@yandex.ru

Considered the software implementation of digital filters for images using the technology NVIDIA CUDA ®.

Introduction

Image processing is the most interesting and important direction. With the increase in size of the image is increased and the processing time. One solution is to parallelization, in practice the calculations are carried out up to 16 threads.

We used technology NVIDIA CUDA (Compute Unified Device Architecture) with which the calculations are made on the graphic accelerators company NVIDIA. This technology allows the use of 512 to 1,024 streams, depending on the model of the graphics accelerator of the company [1].

Subheading

To evaluate the performance were considered: Sobel operator, the Fourier transform, the inverse Fourier transform, filtering images by minimizing the smoothing functional with communication (Tikhonov regularization method).

For the Fourier transforms (forward and reverse) technology was used CUDA [1], allowing the use of the processing power of NVIDIA graphics accelerator. Calculations are based graphics accelerator NvidiaGeForceGTX 650. This graphics accelerator supports up to 1,024 threads running at one time. The amount of memory used for calculations - about 870 megabytes. Since the calculation is on the graphics accelerator, it is worth noting the principle of work and writing code. The function that is executed on GPU is called a kernel. The kernel does not have a return value. Common core prototype is:

__global__ void gpuKernel (/ * arg1 * /, / * arg2 * /, / * ... * /).

Before you call the kernel, you must perform the following steps:

1. Allocate memory for the array on GPU.
2. Copy an array of computer memory to memory graphics accelerator.
3. Calculate the number of threads that will be used in the calculations.
4. Call the core.
5. Synchronize the device to make sure there are no errors in the calculation of.
6. Copy the result from the memory of the graphics accelerator in the computer's memory.
7. Free memory on the graphics accelerator.

For greater clarity, the following results are presented. Calculation of the masking operations Sobel operator is performed on GPU in 1024 threads. Time given in Table 1.

Table 1. The computation of convolution operator Sobel

image Size	Run-time, using the technology of CUDA, seconds	Execution time on the CPU 16 threads, seconds
1902 x 908	0,15	1,37
2457 x 2091	0,24	4,15
3726 x 3108	0,38	9,08
5029 x 4288	0,75	16,95

[1] Support of the grant MK-2932.2013.8

Since used both direct and inverse Fourier transform without the use of FFT algorithms, the size of the image were taken not more than 160x160 pixels. The larger the image size, the greater the run time for the CPU. The direct and inverse Fourier transforms are presented in Tables 2.

Table 2. Calculation time direct Fourier transform

image Size	Run-time, using the technology of CUDA, seconds	Execution time on the CPU 16 threads, seconds
64 x 64	0,21	8,07
96 x 96	0,3	40,57
128 x 128	0,51	125,9
160 x 160	0,97	306,3

Table 3. Time computation of the direct Fourier transform

image Size	Run-time, using the technology of CUDA, seconds	Execution time on the CPU 16 threads, seconds
64 x 64	0,25	8,47
96 x 96	0,32	41,57
128 x 128	0,53	132,9
160 x 160	0,98	310,3

CUDA technology provides all the computing power of the graphics accelerator. This technology provides the ability to parallelize the computation and produce them without the participation of the CPU. This allows you to process images of large size with a small waste of time.

References

1. Nvidia Cuda C Programming Guide. NVIDIA Corporation. 2012.

COMPLEX SCENE ANALYSIS USING HIERARCHICAL SPARSE CONCEPT REPRESENTATION

Sanghyuk Park[1], Jaesik Yoon[2], Chang D. Yoo[3] , and Jaecheol Kwon[4]

[1,2,3] Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea
[4] Future Research Lab., Korea Telecom R&D Center, Seoul, Republic of Korea
[1]shine0624@kaist.ac.kr, [2]jaesik817@kaist.ac.kr, [3]cdyoo@ee.kaist.ac.kr, [4]jckwon@kt.com

This paper considers a hierarchical sparse concept representation for complex scene analysis. Low-level visual features, foreground pixels and optical flows, are commonly used for the scene analysis. However, these features are sensitive to noise and have high-dimension to represent complex behavior of various objects. The considered hierarchical sparse concept representation aims to cope with noise of low-level features and reducing uncertainty in determining behavior patterns. In this paper, the task of complex scene analysis is formalized as discovering high-level behavior patterns which represent video context. The experimental results show that the considered algorithm yields good performance using complex video datasets.

Introduction

Automatic detecting representative behavior pattern has become a challenging task for complex scene analysis (CSA). In the past few years, various algorithms have been proposed to extract semantic behavior patterns from video scenes. They focused on analyzing complex and crowded video scene based on discovered specific patterns from low-level visual features. Previous algorithms usually consider video data as a set of visual features which are represented using a histogram of occurrences of feature, and the extracted features are directly used for CSA. However, these feature based algorithms have mainly two problems: (1) low-level visual features have limitation to represent high-level semantic information of behavior patterns. Hence, high dimensional feature space is required and computational complexity is increased; and (2) low-level visual features are sensitive to noise and some occlusion from movements of various objects in the complex video scene.

Recently, statistical machine learning based algorithms have shown good performance in the task of CSA based on: multi-scale methods [1, 2], hierarchical methods [3, 4, 5], Bayesian methods [6, 7], etc. These algorithms try to implicitly handle varying noise and uncertainty in discovering semantic behavior patterns from high dimensional visual feature spaces.

This paper considers a hierarchical sparse concept representation (HSCR) algorithm for the complex video scene analysis. In this paper, the task of CSA is considered as the problem of discovering a set of high-level semantic behavior patterns using hierarchical dictionary learning and sparse concept representations. The considered HSCR algorithm aims to cope with noise of high dimensional low-level visu-al features and reducing uncertainty in discov-ering semantic behavior patterns while pre-serving discriminative information of primary patterns and dimensionality reduction.

Non-object-based bag-of-words representation is used as low-level visual features similar to previous literature. However, the considered HSCR algorithm tries to discover multi-level behavior patterns instead of using noisy low-level visual features directly. Complex video scenes usually contain multiple movements of various objects over time and space, this paper assumes that these movements can be categorized into low-level and high-level behavior patterns. Low-level behaviors (e.g. moving of vehicle and pedestrians) can be characterized using representative movements which occur in a small range of spatial and temporal regions. High-level behaviors (e.g. traffic flows by signal) can be described as periodically repeated patterns using a set of low-level behavior patterns. Considering these hierarchical structures of behavior patterns can be helpful to analyze video scene more accurately than using single behavior patterns. In considered HSCR, behavior patterns of each level can be described using small number of basis which

related to most representative behavior pattern in each level. For this, hierarchical basis learning and sparsity constraint are considered not only reducing dimensionality of features but also preserving discriminative information.

Review of Sparse Representation

Let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N] \in \Re^{M \times N}$ is a set of N data with high-dimensional feature space. Matrix factorization (MF) algorithms [8, 9, 10] are common approaches to condense data by discovering a set of new basis vector and the new representation with respect to the new basis for each data. The aim of MF is finding two matrices, dictionary $\mathbf{D}$ and sparse representation $\mathbf{A}$, whose product can well approximate $\mathbf{X}$. Given a dictionary $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_K] \in \Re^{M \times K}$, the sparse representation $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_N] \in \Re^{K \times N}$ for $\mathbf{X}$ can be obtained by solving:

$$\mathbf{A} = \text{argmin}_{\mathbf{A}} \| \mathbf{X} - \mathbf{D}\mathbf{A} \|_{\text{F}}^2 \quad \text{s.t.} \forall i, \| \mathbf{a}_i \|_0 \leq \varepsilon \qquad (1)$$

Where, $\| \mathbf{X} - \mathbf{D}\mathbf{A} \|_{\text{F}}^2$ denotes the reconstruction error, and $\| \mathbf{a}_i \|_0 \leq \varepsilon$ is the sparsity constraint. Each column of $\mathbf{D}$ is a basis vector and each column of $\mathbf{A}$ is the K dimension representation of the original input data with respect to the new basis $\mathbf{D}$. In this sense, MF can be regarded as a dimensionality reduction algorithm since it tries to reduce the dimension from *M* to *K*. The performance of sparse representation $\mathbf{A}$ depends critically on the constructed $\mathbf{D}$.

Hierarchical Sparse Concept Representation

Given a data set of high dimensional feature $\mathbf{X}$, the purpose of HSCR algorithm is finding low dimensional high-level behavior patterns while reducing noise and dimensionality of features with reducing information loss. It can be obtained by solving the minimization problem as follows:

$$\min_{\mathbf{D}_{\text{L}}, \mathbf{D}_{\text{H}}, \mathbf{S}} \sum_{i=1}^{N} [\| \mathbf{x}_i - \mathbf{D}_{\text{L}} \mathbf{D}_{\text{H}} \mathbf{s}_i \|_2^2 + \phi \| \mathbf{D}_{\text{H}} \mathbf{s}_i \|_1] \qquad (2)$$

where the dictionary $\mathbf{D}_{\text{L}} \in \Re^{M \times Z}$ is the basis of low-level behavior patterns and $\mathbf{D}_{\text{H}} \in \Re^{Z \times K}$ is the basis of high-level behavior patterns. The matrix $\mathbf{S} \in \Re^{K \times N}$ indicates HSCR of input data $\mathbf{X}$ and ϕ is the regularization parameter. Eq. (2) is not jointly convex to $\mathbf{D}_{\text{L}}, \mathbf{D}_{\text{H}}, \mathbf{S}$. However, it is convex with respect to each of them if others are fixed. Hence, the HSCR can be designed using an iterative algorithm to alter-natively optimize each matrix base.

The HSCR algorithm has three-step in terms of dimensionality reduction and MF based on sparse concept coding algorithm [12]. The first step of HSCR is low-level behavior basis learning by exploring the low-dimensional intrinsic geometric structure of the input data instead of using visual feature space directly. Spectral regression [11, 12] is used for reducing dimension of data and finding manifold of the ambient space to model the local geometric structure of data with weight $\mathbf{W}$. Similarly, in the previous MF algorithm, considered HSCR tries to learn low-level behavior basis matrix $\mathbf{D}_{\text{L}}$ which can well fit to $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^{Z} \in \Re^{N \times Z}$. It can be obtained by solving the optimization problem as follows:

$$\min_{\mathbf{D}_{\text{L}}} \| \mathbf{Y} - \mathbf{X}^T \mathbf{D}_{\text{L}} \|_{\text{F}}^2 + \alpha \| \mathbf{D}_{\text{L}} \|_2^2 \qquad (3)$$

where $\| \mathbf{D}_{\text{L}} \|_2^2$ is a regularization term to avoid over-fitting and α is the regularization parameter. Where, the optimal $\mathbf{Y}$ can be obtained by solving Eq.(4) using the minimum eigenvalue eigen-problem of $\mathbf{L}\mathbf{Y} = \lambda \mathbf{D}_{\text{L}} \mathbf{Y}$.

$$\mathbf{Y}^* = \text{argmin}(\mathbf{Y}^T \mathbf{L} \mathbf{Y}) / (\mathbf{Y}^T \mathbf{D}_{\text{L}} \mathbf{Y}) \qquad (4)$$

Here, $\mathbf{L} = \mathbf{D}_{\text{L}} - \mathbf{W}$ and $\mathbf{D}_{ii} = \sum_j \mathbf{W}_{ij}$. The weight matrix $\mathbf{W}$ is constructed using EMD distance. After obtaining low-level behavior basis $\mathbf{D}_{\text{L}}$, the second step is computing the low-level sparse concept representation $\mathbf{S}_{\text{L}}$ by solving the optimization problem as follows:

$$\min_{\mathbf{S}_{\text{L}}} \sum_{i=1}^{N} [\| \mathbf{x}_i - \mathbf{D}_{\text{L}} \mathbf{s}_{\text{L,i}} \|_2^2 + \beta \| \mathbf{s}_{\text{L,i}} \|_1] \qquad (5)$$

where $\mathbf{S}_{\text{L}} \approx \mathbf{D}_{\text{H}} \mathbf{S}$ and $\| \mathbf{s}_{\text{L,i}} \|_1$ is the L1-norm regularization to enforce the sparsity. The Least Angel Regression [13] is used for solving the optimization problem in Eq. (5). In this step, $\mathbf{S}_{\text{L}}$ can be regarded as representative low-level behavior patterns using small number of basis of $\mathbf{D}_{\text{L}}$. The last step of HSCR algorithms is learning of high-level behavior basis and finding HSCR by solving the optimization problem as follows:

$$\min_{\mathbf{D}_H,\mathbf{S}_H} \sum_{i=1}^{N} [\| \mathbf{s}_{L,i} - \mathbf{D}_H \mathbf{s}_{H,i} \|_2^2 + \gamma \| \mathbf{s}_{H,i} \|_1] \quad (6)$$

where $\mathbf{D}_H$ indicates the basis matrix of high-level behaviors and $\mathbf{S}_H$ is the HSCR using $\mathbf{D}_H$. For this, $\mathbf{D}_H$ can be obtained using Eq.(3) with respect to $\mathbf{S}_L$, and $\mathbf{S}_H$is calculated using Eq.(5) with respect to $\mathbf{S}_L$ and $\mathbf{D}_H$, iteratively. Finally, the considered HSCR algorithm can be convert original high-dimensional low-level features $\mathbf{X}$ to low-dimensional HSCR $\mathbf{S}_H$, while removing noise using dimensionality reduction and preserving discriminative information.

Experiments and Results

In this section, we investigate the effectiveness of the considered HSCR algorithm for CSA. The task of CSA is considered as a multi-class clustering problem similar to previous literatures [2, 3, 7, 14, 15]. For this, we convert each high-dimensional test data into low-dimensional HSCR by using the learned low-level and high-level behavior basis from training data. Then, each converted test data is clustered into the *K* number of primary high-level behavior patterns by the nearest centroid criterion. To verify the effectiveness of HSCR, various experiments were conducted on the publicly available QMUL [3] and CVBASE '06 [16] dataset. All dataset contains complex behavior patterns of a large number of objects such as vehicles and human movements. A ***detailed*** description of experimental parameter is summarized in Table 1.

Table 1. Parameter setup (train)

Dataset	M	N	Z	K	α	β	γ
Junction-1	4176	73	8	2	0.1	0.02	0.4
Roundabout	4176	146	8	2	0.2	0.02	0.56
Basketball	4292	50	4	2	0.2	0.16	0.08
Handball	4524	100	8	2	0.2	0.32	0.1

The Traffic datasets: As shown in Figure 1, The QMUL dataset contains complex traffic scenes and have been extensively used in previous CSA literatures [2, 3, 7, 14, 15]. Each dataset was recorded by a stationary camera with a resolution of 360×288 (25 FPS).

(a) Junction-1 (b) Roundabout

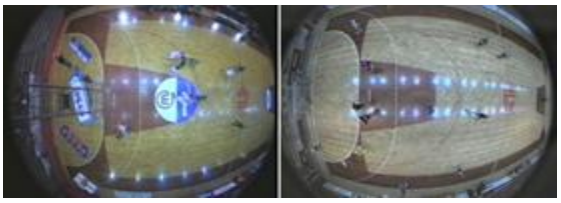
Fig. 1. QMUL datasets

Each video clip was spatially quantized into 36×29 cells and 4 directions of quantized optical flow were extracted in each cell. Hence, the size of low-level feature dimension M is 4176. The test dataset consist of 39, 59 video clips for the Junction-1 and Roundabout, res-pectively. For fair comparison between the considered HSCR algorithm and the state-of-the-art algorithms, the same datasets and ground truth labels [3, 14, 15] are used. For the database in hand, there are two main temporal phases of the primary behavior patterns: vertical and horizontal traffic flows. The quantitative results of the clustering accuracy based on primary high-level behavior patterns are represented in Table 2.

Table 2. Clustering accuracy (QMUL)

Algorithms	Junction-1 (%)	Roundabout (%)
K-means	53.75	63.79
pLSA [3]	89.74	84.46
HpLSA [3]	76.92	72.30
Cas-pLSA [2]	89.70	76.20
DDP-HMM [7]	87.18	85.14
EMD-L1 [15]	92.30	86.40
SparseEMD [16]	89.74	90.00
SRC [10]	92.30	64.41
D-KSVD [9]	92.30	62.70
SCC [12]	89.74	89.83
HSCR	**94.87**	**91.53**

As shown the experimental results in Table 2, it confirmed that the considered HSCR algorithm outperforms the state-of-the-art algorithms using QMUL dataset.



(a) Basketball (b) Handball

Fig. 2. CVBASE '06 dataset

The Sports dataset: As shown in Figure 2, CVBASE '06 dataset depicts complex sports

environments both basketball and handball. The data was recorded by a wide view camera in 5 minutes with 366×288 pixels (25 FPS) and 10 minutes with 384×288 pixels (25 FPS) for basketball and handball dataset, respectively. This sports video includes complex movement from various players in the stadium without moving rules. Hence, it is more difficult to distinguish behavior patterns. To discover high -level behavior patterns, we use two t***ypes*** of group activity as high-level behavior patterns: team offense and defense. To obtain the low-level visual features, video data were divided into several video clips every 3 sec. The test dataset consists of 50, 100 video clips for the basketball and handball data, respectively. The quantitative results of the clustering accuracy based on primary high-level behavior patterns are represented in Table 3. The experimental results show that considered HSCR algorithm outperforms all the others.

Table 3. Clustering accuracy (CVBASE '06)

Algorithms	Basketball (%)	Handball (%)
K-means	84.0	65.0
pLSA [3]	64.0	67.0
HpLSA [3]	70.0	72.0
SRC [10]	94.0	87.0
D-KSVD [9]	94.0	87.0
SCC [12]	76.0	70.0
HSCR	**96.0**	**88.0**

Conclusion

This paper considers a HSCR algorithm for complex scene analysis. The HSCR algorithm determines hierarchical behavior pattern using low-level and high-level behavior basis. The discovered HSCR can be capture most primary pattern to analyze behaviors from the complex video scene, while HSCR algorithm tries to removing noise using dimensionality reduction and sparsity in each level and preserving discriminative behavior patterns using hierarchy.
In the experiments, the HSCR has been used for discovering recurrent primary high-level behaviors in various complex video datasets and has been compared with state-of-the-art algorithms. The extensive experimental results show that the considered algorithm achieved good performance for complex scene analysis. In the future, we will extend considered HSCR algorithm for irregular behavior detection.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No.NRF-2011-0017202 and NRF- 2010-0028680)

References

1. Y.Yang, J.Liu, and M.Shah, "Video scene understanding using multi-scale analysis", ICCV, pp.1669-1676, 2009.
2. J.Li, S.Gong, and T.Xiang. "Learning behavioural context", IJCV, Vol:97(3), pp.276-304, 2012.
3. J.Li, S.Gong, and T.Xiang, "Global behaviour inference using probabilistic latent semantic analysis ", BMVC, pp.193-202, 2008.
4. X.Wang, X.Ma, and W.Grimson, "Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models", IEEE Trans. on PAMI, Vol:31(3), pp.539-555, 2009.
5. R.Emonet, J.Varadarajan, and J.Odobez, "Extracting and locating temporal motifs in video scenes using a hierarchical non parametric bayesian model", CVPR, pp.3233–3240, 2011.
6. T.Hospedales, S.Gong, and T.Xiang. "A markov clustering topic model for mining behaviour in video", ICCV, pp.1165–1172, 2009.
7. D.Kuettel, M.Breitenstein, L.V.Gool, and V.Ferrari, "What's going on? discovering spatio-temporal dependencies in dynamic scenes", CVPR, pp.1951-1958. 2010.
8. M.Aharon, M.Elad, and A.Bruckstein, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation", IEEE Trans. on Signal processing, Vol:54(11), pp.4311-4322, 2006.
9. Q.Zhang, B.Li, "Discriminative k-svd for dictionary learning in face recognition", CVPR, pp.2691-2698, 2010.
10. J.Wright, AY.Yang, A.Ganesh, S.Sastry, and Y.Ma,"Robust face recognition via sparse representation", IEEE Trans. on PAMI, Vol:31(2), pp.210-227, 2009.
11. D.Cai, X.He, and J.Han, "Spectral regression for efficient regularized subspace learning", ICCV, pp.1-8, 2007.
12. D.Cai, H.Bao, and X.He,"Sparse concept coding for visual analysis", CVPR, pp.2905-2910, 2011.
13. B.Efron, T.Hastie, I.Johnstone, and R.Tibshirani, "Least Angle Regression", The Annals of Statistics, Vol:32(2), pp.407–499, 2004.
14. G.Zen and E.Ricci, "Earth mover's prototypes: A convex learning approach for discovering activity patterns in dynamic scenes", CVPR, pp.3225-3232, 2011.
15. G.Zen, E.Ricci, and N.Sebe, "Exploiting sparse representations for robust analysis of noisy complex video scenes", ECCV, pp.193-213, 2012.
16. http://vision.fe.uni-lj.si/cvbase06/index.html.

IMAGE RECONSTRUCTION FROM PHASE USING HERMITE PROJECTION METHOD[1]

E. A. Pavelyeva[2,3], A. S. Krylov[2,4]

[2] Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University E-mail: [3]paveljeva@yandex.ru, [4]kryl@cs.msu.ru

Approximation of Fourier transform of a function using Hermite functions (AFTH) has been considered. It was shown that synthesis results of phase and magnitude for different images using this approximation enhanced synthesis results of discrete Fourier transform. The conditions of uniqueness of approximated function restoration from its AFTH phase were presented. New image boundaries treatment for AFTH method was introduced.

Introduction

Phase of Fourier transform contains more information than its magnitude [1]. In this article the approximation of Fourier transform using Hermite functions (AFTH) and AFTH phase are proposed. The comparison of synthesis results of phase and magnitude from different images using AFTH and discrete Fourier transform shows the effectiveness of AFTH use. New algorithm of image boundaries treatment is applied to the images before the transform of images into the frequency domain. Use of this algorithm reduces the errors of Gibbs effect. In the proposed algorithm the solution of Dirichlet problem for the Laplace equation is found. For AFTH method the conditions of uniqueness of approximated function restoration to within a scale factor from its AFTH phase only information are found in the last section of this article.

Use of Hermite functions for Fourier transform approximation

The Hermite functions [2] are defined as:

$$\psi_n(x) = \frac{(-1)^n e^{\frac{x^2}{2}}}{\sqrt{2^n n!\sqrt{\pi}}} \cdot \frac{d^n(e^{-x^2})}{dx^n}.$$

The Hermite functions are eigenfunctions of the Fourier transform (FT) $F[f] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-i\lambda x} dx$ with eigenvalues $\pm 1, \pm i$: $F[\psi_n] = (-i)^n \psi_n$, form a full orthonormal in $L_2(-\infty, \infty)$ system of functions and are computationally localized in spatial and frequency domains. Two-dimensional Hermite functions are defined as: $\psi_{m,n}(x, y) = \psi_m(x) \cdot \psi_n(y)$, so they have similar properties as one-dimensional Hermite functions in $L_2(R^2)$.

Let's introduce some definitions that will be used later in the article.

Def. 1: Let $f_n(x)$ be an approximation of expansion of $f(x) \in L_2(-\infty, \infty)$ in a Fourier series of Hermite functions (Hermite projection method [3]):

$$f(x) \approx f_n(x) = \sum_{k=0}^{n} c_k \ \psi_k(x), \qquad (1)$$

where $c_k = \int_{-\infty}^{\infty} f(x) \psi_k(x) dx$ are the Hermite coefficients. We call the expression

$$HF_{f,n}(x) = \sum_{k=0}^{n} c_k \ (-i)^k \ \psi_k(x) \qquad (2)$$

as approximation of Fourier transform using Hermite functions (AFTH).

Remark 1: Since $F[\psi_n] = (-i)^n \psi_n$ then $HF_{f,n} = F[f_n]$.

In every point x complex value $HF_{f,n}(x)$ can be written in exponential form:

$$HF_{f,n}(x) = A_{f,n}(x) \cdot e^{i\varphi_{f,n}(x)},$$

where $A_{f,n}(x)$ is the magnitude, $\varphi_{f,n}(x) = \arg HF_{f,n}(x)$ is the phase of $HF_{f,n}(x)$ (AFTH phase). If $\mathrm{Re}\, HF_{f,n}(x) = 0$ and $\mathrm{Im}\, HF_{f,n}(x) = 0$ we suppose that

[1] The work was supported by RFBR grant 13-07-00438

$\varphi_{f,n}(x) = 0$. Then $\varphi_{f,n}(x)$ is defined for all $x \in (-\infty, \infty)$ and takes values $(-\pi, \pi]$.

Remark 2: $tg\varphi_{f,n}(x) = \dfrac{\mathrm{Im}\, HF_{f,n}(x)}{\mathrm{Re}\, HF_{f,n}(x)}$ if $\mathrm{Re}\, HF_{f,n}(x) \neq 0$. If $\mathrm{Re}\, HF_{f,n}(x) = 0$ we suppose that $tg\,\varphi_{f,n}(x) = 0$.

Def. 2: Let $F_n(x)$ be an approximation of complex function $F(x) \in L_2(-\infty, \infty)$ using Hermite functions $\psi_0(x), ..., \psi_n(x)$:

$$F(x) \approx F_n(x) = \sum_{k=0}^{n} C_k\, \psi_k(x), \qquad (3)$$

where C_k are complex coefficients. We call the expression

$$HF^{-1}{}_{F,n}(x) = \sum_{k=0}^{n} C_k\, (-i)^k\, \psi_k(-x) \qquad (4)$$

as approximation of inverse Fourier transform using Hermite functions (AIFTH).

Remark 3: If $f_n(x)$ is an approximation of $f(x) \in L_2(-\infty, \infty)$ using Hermite functions $\psi_0(x), ..., \psi_n(x)$, $HF_{f,n}$ is AFTH of $f(x)$, $HF^{-1}[HF_{f,n}(x)]$ is AIFTH of $HF_{f,n}$, then $HF^{-1}[HF_{f,n}(x)]$ is real function and

$$f_n(x) = HF^{-1}[HF_{f,n}(x)] \ \forall x.$$

Def. 3: Let $f_{m,n}(x, y)$ be an approximation of $f(x, y) \in L_2(R^2)$ using Hermite functions $\psi_{i,j}(x, y)$, $i = 0,1,..., m$, $j = 0,1,..., n$:

$$f(x, y) \approx f_{m,n}(x, y) = \sum_{k=0}^{m} \sum_{l=0}^{n} c_{k,l}\, \psi_{k,l}(x, y) \qquad (5)$$

We call the expression

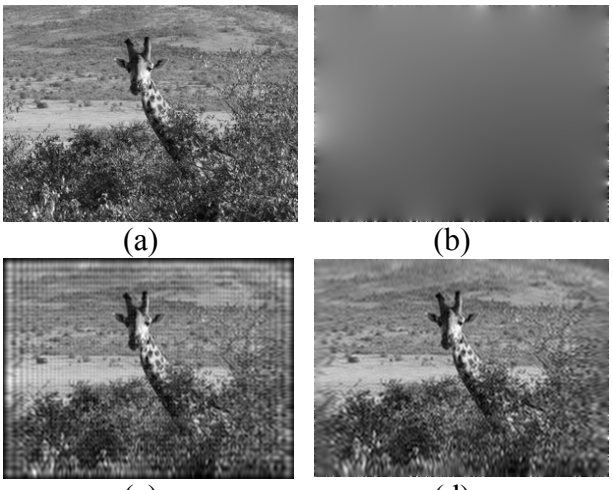
$$HF_{f,\mathrm{m},n}(x, y) = \sum_{k=0}^{m} \sum_{l=0}^{n} c_{k,l}\, (-i)^{k+l} \psi_{k,l}(x, y) \qquad (6)$$

as approximation of Fourier transform using Hermite functions (AFTH) in two-dimensional case. The definition of AIFTH in two-dimensional case is introduced similarly.

Boundary Gibbs effect suppression

Let we have an image with the intensity $I(x, y)$. We can interpret it as two-dimensional function $f(x, y) \in L_2(R^2)$ which is equal to $I(x, y)$ inside the rectangle $D = [-A, A] \times [-B, B]$ and is zero outside D. We rescale Hermite functions in $A/\sqrt{2m+1}$ times along the *x*-axis and $B/\sqrt{2n+1}$ times along the *y*-axis so that $\psi_{m,n}(x, y)$ is roughly localized in *D* [3]. After the rescaling all used Hermite functions are close to zero near the boundaries of D, so the approximation $f_{m,n}(x, y)$ in (5) can have Gibbs effect near the boundaries. To overcome this problem we find continuous function $L(x, y)$ in D with continuous second derivatives in $(-A, A) \times (-B, B)$ which is equal to $f(x, y)$ on the boundary of D. We find it as the solution of the Dirichlet problem for the Laplace equation:

$$\begin{cases} \Delta L(x, y) = 0, \ -A < x < A, -B < y < B, \\ L(-A, y) = f(-A, y) \\ L(A, y) = f(A, y) \end{cases}, \ -B \le y \le B, \qquad (7)$$
$$\begin{cases} L(x, -B) = f(x, -B) \\ L(x, B) = f(x, B) \end{cases}, \ -A \le x \le A.$$



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

Fig. 1. (a) the original image 300×225; (b) base function $L(x, y)$; (c) $f_{m,n}(x, y)$; (d) $\tilde{f}_{m,n}(x, y) + L(x, y)$.

We call the solution $L(x, y)$ of problem (7) as base function of $f(x, y)$ and find it using the finite element method.

Then we approximate the function $\tilde{f}(x, y) = f(x, y) - L(x, y)$ which is zero on the boundary of D and add to derived approximation $\tilde{f}_{m,n}(x, y)$ the base function $L(x, y)$. It can be

seen (fig. 1) that the Gibbs effect is suppressed for $\widetilde{f}_{m,n}(x,y)+L(x,y)$ ($m=150$, $n=100$).

Synthesis of phase and magnitude of images

It is well-known that the phase of Fourier transform contains much more information than its magnitude [1]. The classic approval of this statement is the following: two images $f(x,y)$ and $g(x,y)$ are taken. Let $F[f]=A_f\cdot e^{i\varphi_f}$ and $F[g]=A_g\cdot e^{i\varphi_g}$ are their discrete Fourier transforms. Then a synthesis is applied: two images are generated, where first image has the magnitude of FT of $f(x,y)$ and phase of FT of $g(x,y)$, and vice versa for the second image. We analyze the same synthesis using phases of AFTH $\varphi_{f,n}(x)$, $\varphi_{g,n}(x)$ and magnitudes of AFTH $A_{f,n}(x)$, $A_{g,n}(x)$, and we also use AIFTH instead of inverse FT. For both FT and AFTH methods of synthesis the base function $L(x,y)$ is subtracted before the transform of image into the frequency domain.

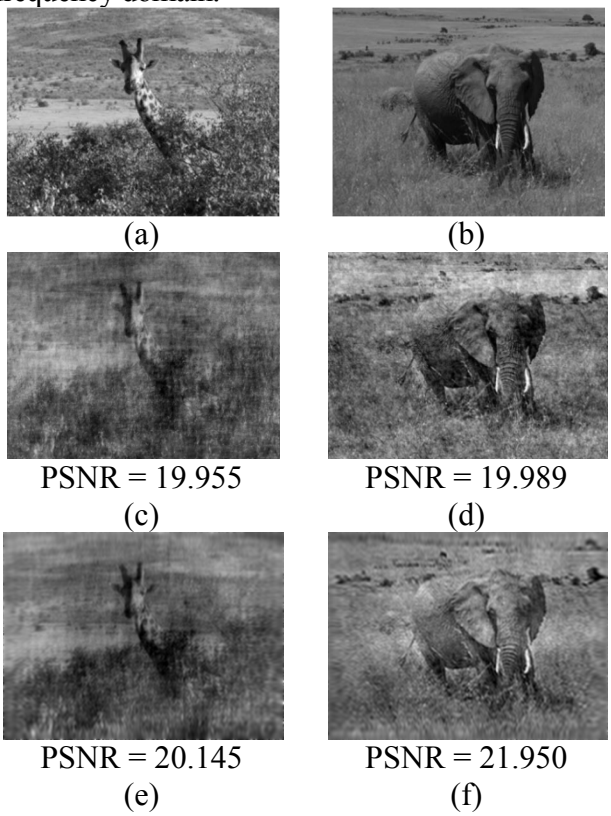


Fig. 2. (a), (b) the original images; synthesis results for: (c) phase of giraffe and magnitude of elephant; (d) phase of elephant and magnitude of giraffe. (e) phase of AFTH of giraffe and magnitude of AFTH of elephant; (f) phase of AFTH of elephant and magnitude of AFTH of giraffe.

The comparisons of images $f(x,y)$, $F^{-1}\left[A_g\cdot e^{i\varphi_f}\right]$ and $HF^{-1}\left[A_{g,n}\cdot e^{i\varphi_{f,n}}\right]$ are illustrated in fig. 2. PSNR values of the original and obtained images with the same phase information are calculated.

Table 1 shows the comparisons between synthesis using FT and AFTH of some 256×256 images from SIPI database [4]. We use $m=n=120$ for AFTH calculation (5-6). In all cases the magnitude of "lenna" image is taken. All images have better PSNR value in case of synthesis using AFTH. The synthesis results of phase of image "barbara" and magnitude of image "lenna" are shown in fig. 3.

Fig. 3. (a) "barbara", (b) "lenna"; (c) syntesis using FT; (d) synthesis using AFTH.

Table 1. PSNR values for FT and AFTH synthesis.

image	FT synthesis	AFTH synthesis
airplane	16.663	18.347
baboon	17.008	17.367
barbara	18.823	20.264
cameraman	12.949	13.435
couple	19.564	21.902
girl	18.364	20.096
girl2	12.729	13.361
girl3	11.881	12.697
house	15.289	16.576
jelly_beans1	17.885	19.009
jelly_beans2	17.362	18.303
peppers	17.712	18.569
sailboat	13.976	14.683
splash	15.065	16.046
tiffany	15.941	17.467
tree	12.477	12.820

The obtained results show that use of AFTH phase enhances the results of use of FT phase.

Signal restoration from phase of AFTH only

There are uniqueness theorems [1] of function restoration to within a scale factor from phase only information of its discrete Fourier transform. In this article we present the conditions (Theorem 2) under which the approximation $f_n(x)=\sum_{i=0}^{n} c_i\psi_i(x)$ of $f(x)\in L_2(-\infty,\infty)$ is uniquely specified to within a scale factor by phase of its AFTH.

Fix a number of Hermite functions n. Let $f_n(x)=\sum_{i=0}^{n} c_i\psi_i(x)$, $g_n(x)=\sum_{i=0}^{n} d_i\psi_i(x)$ are approximations of $f(x), g(x)\in L_2(-\infty,\infty)$ and are not even and not odd; $\varphi_{f,n}(x)$, $\varphi_{g,n}(x)$ denote phases of AFTH of $f(x)$ and $g(x)$. Let $\varphi_{f,n}(x)=\varphi_{g,n}(x)$ for $x\in(-\infty,\infty)$ where $\operatorname{Re} HF_{f,n}(x)\neq 0$ and $\operatorname{Re} HF_{g,n}(x)\neq 0$. We will determine coefficients $d_0,d_1,...,d_n$ from known $c_0,c_1,...,c_n$.

Let $D_n=\begin{bmatrix}d_0 & d_1 & ... & d_n\end{bmatrix}^T$ be a column matrix of unknown coefficients. The condition that the phases are equal when $\operatorname{Re} HF_{f,n}(x)\neq 0$ and $\operatorname{Re} HF_{g,n}(x)\neq 0$ means that $tg\varphi_{f,n}(x)=tg\varphi_{g,n}(x)$ or

$$\frac{-\sum_{k=0}^{\left[\frac{n+1}{2}\right]-1} c_{2k+1}(-1)^k\psi_{2k+1}(x)}{\sum_{k=0}^{\left[\frac{n}{2}\right]} c_{2k}(-1)^k\psi_{2k}(x)}=\frac{-\sum_{k=0}^{\left[\frac{n+1}{2}\right]-1} d_{2k+1}(-1)^k\psi_{2k+1}(x)}{\sum_{k=0}^{\left[\frac{n}{2}\right]} d_{2k}(-1)^k\psi_{2k}(x)}.$$

This condition can be rewritten in a matrix form:

$$A_nD_n=0,$$

where matrix A_n has n rows and $n+1$ columns, rows and columns are numbered from zero.

We denote $p_{m,l}^k=2^k k!C_m^kC_l^k$. $\left[\frac{n}{2}\right]$ is integer part of a number $\frac{n}{2}$.

Theorem 1: $A_1=\begin{bmatrix}-c_1 & c_0\end{bmatrix}$,

$$A_2=\begin{bmatrix}0 & -c_2 & c_1\\ -c_1 & c_0-p_{1,2}^1\cdot c_2 & p_{1,2}^1\cdot c_1\end{bmatrix},$$

$A_{n+1}=A_n^0+B_{n+1}$, where $A_n^0=\begin{bmatrix}0 & 0\\ A_n & 0\end{bmatrix}$,

$B_{n+1}=\begin{bmatrix}\bar{b}_0 & \bar{b}_1 & ... & \bar{b}_n & \bar{b}_{n+1}\end{bmatrix}=$

$$=\begin{bmatrix}
0 & ... & 0 & 0 & 0 & 0 & -c_{n+1} & c_n\\
0 & ... & 0 & 0 & c_{n+1} & 0 & -p_{n,n+1}^1\cdot c_{n+1} & b_{1,n+1}\\
... & ... & -c_{n+1} & 0 & p_{n-2,n+1}^1\cdot c_{n+1} & 0 & -p_{n,n+1}^2\cdot c_{n+1} & b_{2,n+1}\\
0 & ... & -p_{n-4,n+1}^1\cdot c_{n+1} & 0 & p_{n-2,n+1}^2\cdot c_{n+1} & 0 & -p_{n,n+1}^3\cdot c_{n+1} & b_{3,n+1}\\
b_{\left[\frac{n}{2}\right],0} & ... & ... & ... & ... & ... & ... & ...\\
0 & ... & -p_{n-4,n+1}^{n-4}\cdot c_{n+1} & 0 & p_{n-2,n+1}^{n-3}\cdot c_{n+1} & 0 & -p_{n,n+1}^{n-2}\cdot c_{n+1} & b_{n-2,n+1}\\
... & ... & 0 & 0 & p_{n-2,n+1}^{n-2}\cdot c_{n+1} & 0 & -p_{n,n+1}^{n-1}\cdot c_{n+1} & b_{n-1,n+1}\\
0 & ... & 0 & 0 & 0 & 0 & -pp_{n,n+1}^{n}\cdot c_{n+1} & b_{n,n+1}
\end{bmatrix},$$

where $\bar{b}_j$ denotes j^{th} column of matrix B_{n+1}, $b_{i,j}$, $i=0,1,...,n$, $j=0,1,...,n+1$ are elements of B_{n+1}. Element $b_{\left[\frac{n}{2}\right],0}$ of matrix B_{n+1} is zero if n is odd and is $(-1)^{\frac{n}{2}+1}c_{n+1}$ if n is even. The column

$$\bar{b}_{n+1}=-\frac{1}{c_{n+1}}\left(\sum_{k=0}^{\left[\frac{n}{2}\right]} c_{n-2k}\cdot\bar{b}_{n-2k}\right).$$

Theorem 2: Let F_n be the minor of $(n-1)$ order of matrix A_n obtained from the matrix A_n by deleting the zero row and the last two columns. Let $\varphi_{f,n}(x)=\varphi_{g,n}(x)$ for $x\in(-\infty,\infty)$ where $\operatorname{Re} HF_{f,n}(x)\neq 0$ and $\operatorname{Re} HF_{g,n}(x)\neq 0$. Then $g_n(x)=\beta\cdot f_n(x)$ for $\beta=const$ if and only if $F_n\neq 0$.

If we know the coefficients $c_0,c_1,...,c_n$ of approximation (1) of $f(x)\in L_2(-\infty,\infty)$ it is easy to calculate the minor F_n and thus to verify the uniqueness conditions for restoration of approximation $f_n(x)$ by AFTH phase $\varphi_{f,n}(x)$.

Conclusion

In this article an approximation of Fourier transform using Hermite functions was considered. The algorithm of image boundaries region treatment for AFTH and FT was introduced. The synthesis results of phase and magnitude from different images shows the effectiveness of AFTH use. The found conditions of uniqueness of approximated function restoration from its AFTH phase enable us to control AFTH phase based image processing.

References

1. A. V. Oppenheim, J. S. Lim. The importance of phase in signals //Proc. of the IEEE, 69(5), pp. 529-541, 1981.
2. E.C. Titchmarsh. Introduction to the theory of Fourier integrals // Oxford: Clarendon Press, vol. 337, 1937.
3. A. S. Krylov, D. N. Korchagin. Fast Hermite Projection Method // LNCS, vol. 4141, pp. 329-338, 2006.
4. SIPI images database http://sipi.usc.edu/database/

THE FILTERING OF PERIODICAL SIGNAL WITH ALIQUANT OWN PERIOD TO ITS DISCRETIZATION INTERVAL[1]

A. Reznik [2], V. Efimov[2,3],Y. Bondarenko[2,4]

[2] Institute of Automation and Electrometry, Siberian Branch, Russian Academy of Sciences, pr. Akademika Koptyuga 1, Novosibirsk, 630090 Russia, [2]reznik@iae.nsk.su
[3] efimov_vit@mail.ru
[4] bjuv@iae.nsk.su

Various approaches to solving the problem of filtration of a continuous periodic signal distorted by stationary additive non-correlated noise are considered. Particular attention is paid to the case where the known signal period is non-divisible by the signal sampling period.

Introduction

The problem of identification of a continuous periodic signal *S(t)* [1] from a mixture of the signal and additive amplitude non-correlated noise *ξ(t)* is simple if the signal period *T* is known:

$$S(t) = S(t + mT),\ m - \text{integer}. \qquad (1)$$

In this case, it is sufficient to unite *M* signal periods to be processed, for instance, in the zero period and to perform usual averaging:

$$S^0(t) = \frac{1}{M}\sum_{m=0}^{M-1}(S(t) + \xi(t + mT)). \qquad (2)$$

Here $\xi(t + mT)$ is the additive noise in the *m*-th period.

If, in sampling a mixture of the signal and noise, the signal period *T* contains an integer number *2L+1* of sampling intervals Δ, Eq. (2) transforms to

$$S^0(0.5\Delta + r\Delta) = \frac{1}{M}\sum_{m=0}^{M-1}(S(0.5\Delta + r\Delta) + \xi(0.5\Delta + r\Delta + mT)) \qquad (3)$$
$$r = \overline{0,2L}.$$

Such processing in the case of stationary non-correlated noise with a dispersion σ_ξ^2 and zero mean leads to reduction of the signal/noise ratio by a factor of *M*, i.e., the noise dispersion acquires the value σ_ξ^2 / M.

Processing of a periodic signal whose period is non-divisible by the sampling interval

The remainder of T/Δ is a simple fraction**.** Let us first consider the case where the signal period can be presented as

$$T = (N + q / s)\Delta \qquad (4)$$

where *q/s* is a rational fraction ($0 < q/s < 1$).

An analysis of a sequence of the adjacent signal periods shows that the number of different variants of the distribution of samples over the period *T* coincides with the fraction denominator *s*.

If the number of processed periods is *sM*, then periods with identical arrangement of the samples inside the period are first grouped, and then they are processed in accordance with Eq. (3). As a result, the signal/noise ratio decreases by a factor of *M*.

[1] This work was supported by the Russian Foundation for Basic Research (Grant No. 13-01-00361), by the Presidium of the Russian Academy of Sciences (Grant No. 15.11/2012), and by the Siberian Branch of the Russian Academy of Sciences (Integration Project No. 16/2012 between the Siberian Branch of the Russian Academy of Sciences and the National Academy of Sciences of Belarus).

Uniting the resultant *s* periods into one, we obtain the period

$$T = (sN + q)\Delta / s, \qquad (5)$$

inside which the samples are located with the sampling interval Δ/s. As the number of samples in the unitedperiod is greater approximately by a factor of *s* than the number of samples in an individual period, this fact can be used to further reduce the signal/noise ratio by a factor of *s*. The total decrease in this ratio is *sM* without distortion of the periodic signal with a limited number of harmonics.

Another method of filtration of the periodic signal is possible. Multiplying the left and right sides of equality (4) by the denominator of the rational fraction *q/s*, we obtain another period (*T1*), which is divisible by the sampling interval Δ:

$$T_1 = sT = (sN + q)\Delta. \qquad (5a)$$

This period contains harmonics with the index divisible by *s*, and dispersion is decreasing at S times.

Application of analytical transformations

Let us consider a periodic signal with an odd number of harmonics (the value of $N + 1$ is odd):

$$f(t) = a_0 + \sum_{k=1}^{N/2} a_k \cos(\frac{2\pi}{T} kt) + \sum_{k=1}^{N/2} b_k Sin(\frac{2\pi}{T} kt). \qquad (6)$$

If we know the values of *f(tr)* on the set of abscissas *tr* ($r = 0,N$) on the time interval T ($-0.5T < t0 < t1 < t2 < \ldots < tr < \ldots < tN < 0.5T$), then this signal can be correctly described by the identity [2]

$$f(t) = \sum_{r=0}^{N} f(t_r) W_r(t - t_r), \qquad (6a)$$

For an odd number of samples, the sampling function is [2]

$$W_r(t - t_r) = \prod_{\substack{u=0 \\ u \neq r}}^{N} \frac{\sin(\frac{\pi}{T}(t - t_u))}{\sin(\frac{\pi}{T}(t_r - t_u))}, r = \overline{0, N}. \qquad (7)$$

If the number $N + 1$ (number of samples) is even in the formula for the sampling function, there appears the factor

$$\cos(\frac{\pi}{T}(t + \sum_{\substack{k=1 \\ k \neq r}}^{N} t_k)) / \cos(\frac{\pi}{T} \sum_{k=0}^{N} t_k)$$

After preliminary processing of the entire set of periods and constructing the mathematical description for each period, the final filtration reduces to summation of the mathematical descriptions of all *M* periods and subsequent division of this sum by the number of periods.

The signal reconstruction error dispersion caused by amplitude non-correlated noise of the samples becomes smaller by a factor of *M* than the dispersion in the case of signal filtration over one period.

Filters with a finite pulse characteristic

Let us consider the use of filters of lengths $L = 2n + 1$ [3] and $L = 2n$ with the sampling functions

$$W_p = \frac{\cos(\omega \Delta p)}{\sum_{r=-n}^{n} \cos^2(\omega \Delta r)}, \quad p = \overline{-n, n} \qquad (8a)$$

$$W_p = \frac{\cos(\omega \Delta (2/p/ - 1)/2)}{2\sum_{r=1}^{n} \cos^2(\omega \Delta (2r - 1)/2)} \; p = \overline{-n, n}, p \neq 0 \qquad (8b)$$

These filters do not distort the signal of frequency ω and reduce the signal/noise ratio because the noise dispersion after filtration is

$$< \varepsilon_\xi^2 >= \sigma_\xi^2 \Big/ \sum_{r=-n}^{n} \cos^2(\omega \Delta r) \qquad (9a)$$

and

$$< \varepsilon_\xi^2 >= \sigma_\xi^2 \Big/ (2\sum_{r=1}^{n} \cos^2(\omega \Delta (2r - 1/2))). \qquad (9b)$$

If the signal contains only one harmonic, the non-divisibility of the period T by the sampling interval Δ does not affect the process of non-correlated noise suppression; as the filter length is increased, the residual noise dispersion decreases approximately by a factor of n in accordance with Eqs. (12a) and (12b). If the signal contains several harmonics, the non-divisibility of the period T by the interval Δ leads to mutual penetration of harmonics into the results of filtration of an individual harmonic.

The coefficient of transmission of the frequency λ by the filter tuned to the frequency ω is determined by the obvious relation for the filter of length $L = 2n + 1$:

$$W_{\omega}(\lambda) = C(\omega,\lambda)_{2n+1} \Big/ \sum_{r=-n}^{n} \cos^2(\omega\Delta r),$$

(10a)

the corresponding relation for the filter of length $L = 2n$ is

$$W_{\omega}(\lambda) = C(\omega,\lambda)_{2n} \Big/ 2\sum_{r=1}^{n} \cos^2(\omega\Delta(2r-1)/2). \quad (10b)$$

Using Eqs. (8a), (8b), (9a), (9b), (10a), and (10b), we can construct a system of linear equations preventing the mutual penetration of harmonics.

The analysis performed in this paper allows us to estimate the minimum possible error of filtration of a continuous periodic signal distorted by stationary additive non-correlated noise; with minor complications of processing algorithms, it is also possible to take into account the non-divisibility of the signal period T by the signal sampling interval Δ.

References

1. D. Jackson, Fourier Series and Orthogonal Polynomials, Dover Books on Mathematics, 1935.
2. V. M. Efimov, A. L. Reznik, and A. V. Torgov, “Sampling Functions in Nonuniform Sampling of a PeriodicSignal,” Optoelectron., Instrum., Data Proc. 46 (6), 521–532, 2010.
3. V. M. Efimov, “Limiting Capabilities of One Method of Statistical Processing of Measurement Results for a Stationary Random Process,” Avtometriya, No. 6, 112–115, 1965 (in russian).

COMPUTING OPTIMIZATION ALGORITHM ESTIMATION PARAMETERS OF ROTATION THREE-DIMENSIONAL OBJECTS BASED ON HOUGH TRANSFORM

A.A. Rozhentsov[1,2], K.V. Morozovskiy[1,3]

[1]424000 Volga State University of Technology, Yoshkar-Ola, Lenin Sq. 3, Department of radio engineering and biomedical systems.
[2]krtmbs@volgatech.net, [3]MorozovskiyKV@volgatech.net

This paper presents an approach to the definition of the rotation parameters 3D images based on Hough transform. Approach to computing optimization presented algorithm on the basis of the CUDA technology is shown. The algorithm of finding parameters of rotation on the basis of Hough transform on the CPU and GPU is described.

Introduction

One of the tasks which are often solved at processing of three-dimensional images is estimate the parameters of rotation three-dimensional objects. With the known numbering of points[1], for example in astroorientation systems where parameter estimation step is preceded by identification of stars in the frame, parameter estimation can be performed by matrix method, and the solution of this task doesn't cause difficulties. In situations where the observed image is affected by various noising factors, unknown numbering of points original object and observed object, does not match their number of points, the observed object is presented by a fragment, determination parameters of rotation by a matrix method becomes impossible. The solution to this problem is possible on the basis of the correlation algorithms, but it has highly complexity in conditions priori uncertainty about the parameters of rotation, scaling and transfer and it difficult to implement in real or near to the time scale. To increase the performance of the algorithm was proposed to use the technology NVIDEO CUDA.

The main method is developed by Hough transform. The main advantages of this method is its invariance to the numbering of points. Introduced in [2] modified generalized Hough transform allows to determine the coordinates centers of the shape independently of the rotation, scaling and shift parameters. However, this transformation does not allow to determine the rotation parameters.

Defining the parameters of rotation

Rotation points of three-dimensional object can be described by a rotation tensor [3]

$$P = m \otimes m + \cos(\alpha)(E - m \otimes m) + \\ + \sin(\alpha)(m \times E) \quad (1)$$

where P - the tensor of rotation, $\otimes$ - a sign of the tensor product, $m \otimes m$ - second-order tensor, m - axis of rotation, α - the angle of rotation around the axis and E - the unit tensor

Rotation is described by the following expression

$$v' = Pv \quad (2)$$

where v' - the rotating vector, v - the source vector.

Then the rotating vector is defined as:

$$v' = v\cos(\alpha) + (m \times v)\sin(\alpha) + \\ + m(m \cdot v)(1 - \cos(\alpha)) \quad (3)$$

On the basis of (3) to create expressions for determining the coordinates of the rotated vector:

$$\begin{aligned} &(1-\cos\alpha)(m_1v_1 + m_2v_2 + m_3v_3)m_1 + \cos\alpha v_1 + \\ &+ \sin\alpha(m_2v_3 - m_3v_2) = v_1' \\ &(1-\cos\alpha)(m_1v_1 + m_2v_2 + m_3v_3)m_2 + \cos\alpha v_2 + \\ &+ \sin\alpha(m_3v_1 - m_1v_3) = v_2' \\ &(1-\cos\alpha)(m_1v_1 + m_2v_2 + m_3v_3)m_3 + \cos\alpha v_3 + \\ &+ \sin\alpha(m_1v_2 - m_2v_1) = v_3' \end{aligned} \quad (4)$$

Solving (4) with respect to m1, m2, m3 get two solutions. The first solution:

$$m_1 = f_1(v_1, v_2, v_3, v_1', v_2', v_3', \alpha)$$
$$m_2 = f_2(v_1, v_2, v_3, v_1', v_2', v_3', \alpha) \quad (5)$$
$$m_3 = f_3(m_1, m_2, v_1, v_2, v_3, v_1', v_2', v_3', \alpha)$$

The second solution:

$$m_1' = f_1'(v_1, v_2, v_3, v_1', v_2', v_3', \alpha)$$
$$m_2' = f_2'(v_1, v_2, v_3, v_1', v_2', v_3', \alpha) \quad (6)$$
$$m_3' = f_3'(m_1, m_2, v_1, v_2, v_3, v_1', v_2', v_3', \alpha)$$

View of function $f_1(v_1, v_2, v_3, v_1', v_2', v_3', \alpha)$, $f_2(v_1, v_2, v_3, v_1', v_2', v_3', \alpha)$, $f_3(m_1, m_2, v_1, v_2, v_3, v_1', v_2', v_3', \alpha)$, $f_1'(v_1, v_2, v_3, v_1', v_2', v_3', \alpha)$, $f_2'(v_1, v_2, v_3, v_1', v_2', v_3', \alpha)$ and $f_3'(m_1, m_2, v_1, v_2, v_3, v_1', v_2', v_3', \alpha)$ is presented in the appendix.

At determining parameters of rotation original data is coordinates points [4,5] of original and processed 3D images. On the basis of coordinates points can be calculated the normal vectors to the surface of the object, that rotate similarly object and insensitive to the transformation of the offset. At determining parameters of rotation it is necessary to estimate four unknown parameters: three parameters – coordinates of an axis of rotation and the fourth - angle of rotation. According to expressions (5-6) free parameter Hough transform is the angle of rotation. Coordinates vectors of normal to a surface of the original and rotated objects will be basic data for determination parameters of rotation. For free parameter it is necessary to set limits of its change. For an unambiguous choice of the direction axis of rotation, the angle α is set in the range of from 0^0 to 180^0.Since each normal vector of the original object is compared to the each normal vector of rotating object, the present version Hough transform is not sensitive to the numbering of points.

At calculation coordinates of axes rotation for the next vector of a normal of original object and a vector of a normal of rotated object find coordinates of axis of rotation at some angle α. Since the angle α will take all possible values, the result is determined the coordinates of the axes of rotation possible in the given conditions, the intersection it in the accumulate array, which will peak. When calculating the rotation parameters are situations where solutions of (5) and (6) do not exist or do not satisfy the normalization rotation axis m. In this case, the vote is not carried out and it is believed that the rotation with these parameters is not possible.

Since the unknowns are the four parameters of rotation, then the accumulate array should be a four-dimensional array, which makes the implementation of this algorithm. However, as the vector of the rotation axis is normalized, and for its representation two of its component there is enough knowledge, and the third can be calculated to within a sign. In this regard, proposed to use an array of three-dimensional accumulate array $A[m_y, m_z, \alpha]$ comprising the coordinates rotary axis y, z and the rotation angle. As the knowledge of coordinates of peak in the accumulate array doesn't allow to define directly a sign the third components of a vector of an axis of rotation, It is necessary to re-calculate the coordinates of the axes of rotation and if indexes of accumulate array received thus coincide with found earlier, the calculated values of the coordinates of the rotation axis and the rotation angle is taken as the evaluation parameters of rotation.

In view of the algorithm for finding the rotation parameters based on Hough transform will include the following steps:

1. The calculation of normal vectors of the original and the rotated object.
2. Given the normal vector of point of the original object - v.
3. Given the normal vector of point rotating object - v'.
4. For a certain angle of rotation α of the normal vectors of points of the original object and the rotating object
5. Calculation in accordance with the (5,6) coordinates axis of rotation
6. If the axis of rotation exists and its length is equal to unit, the corresponding cell of the accumulate array is incremented $A[m_y, m_z, \alpha]$.
7. Finding the index of the peak in the accumulate array.
8. Repeat steps 1 - 6 and search matched cell to index of peak accumulate array.
9. End.

The algorithm for finding the parameters of rotation on the GPU

The algorithm, finding parameters of rotation (fig. 1), has three enclosed cycles, a choice a vector of a normal original object, a vector of a normal rotated object and value an angle of rotation. Thus, the complexity of the algorithm

for determining the parameters of rotation is $O(n^3)$.

Implementing this algorithm on the CPU involves the use of three enclosed cycles, as already shown, is the choice vector of normal original object (Fig. rectangle 1), vector of normal rotating object (Fig. Rectangle 2) and the value of the angle of rotation (Fig. Rectangle 3).

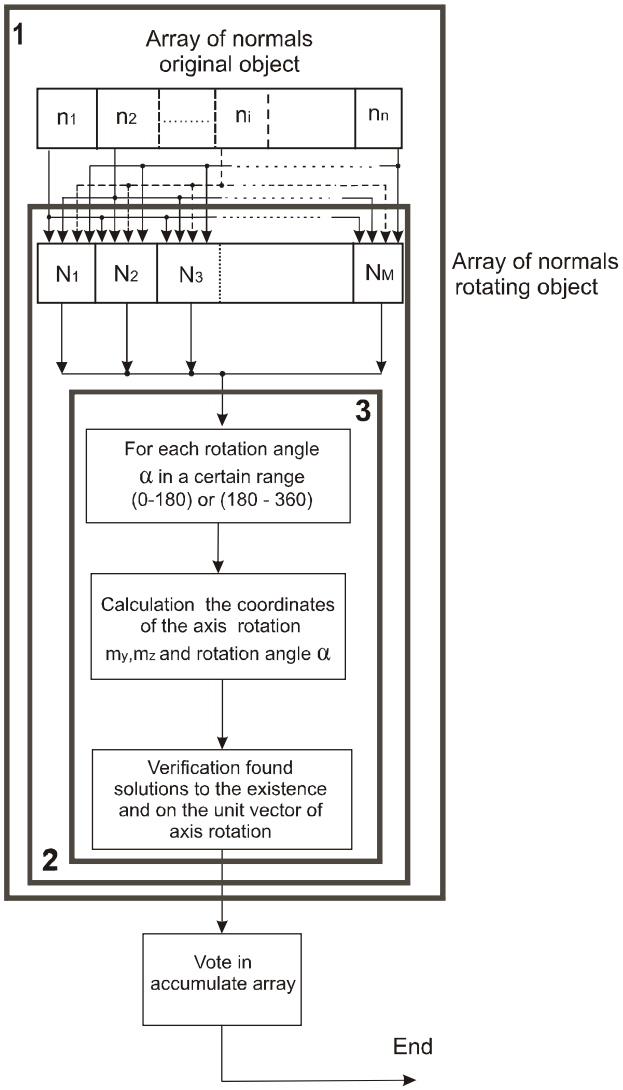


Fig. 1 The algorithm for finding the parameters of rotation on the Hough transform implemented on the GPU

In this triple loop conducted the check of existence vector axis of rotation with the given parameters of rotation and unit-length vector axis of rotation. After verification carried the vote in a three-dimensional accumulate array. The implementation of the algorithm for finding the parameters of rotation on the GPU suggests that the first two cycles, select vector of normal for original object and select vector of normal for rotating object (Fig. rectangle1 and 2) are defined on the GPU as a two-dimensional Grid. Its horizontal dimension defined by the number of counts the normal vector for original object and the vertical number of counts the normal vector for rotating object.

The GPU is made the third cycle - the enumeration of the possible values of the angle of rotation α (Fig. rectangle 3), checking the existence vector axis of rotation for given rotation parameters, check the unit vector length axis of rotation. If the vector axis of rotation exists and its length a unit, that determined the index of cell the three-dimensional accumulate array. Similarly, for this function are provided with the appropriate specified __ device__.

All combinations of the rectangular coordinate grid, comprising numbering threads, defines all possible combinations of the normal vectors original object and normal vectors rotating object. It provides parallel execution on the GPU enumeration algorithm, implemented on the CPU in the first two cycles of the algorithm for finding the parameters of rotation. Each thread of two-dimensional grid is loops in a cycle values the angles of rotation (Fig. rectangle 3) and finds the coordinates of the rotation axis. Since all the threads run in parallel on GPU, for a guaranteed vote in the accumulate array by all threads, the operation addition in cells is implemented using atomic operations (GPU).

Analysis of the performance showed that the use of NVIDIA CUDA technology to implement an algorithm for determining the parameters of the rotation based on Hough transform provides a reduction of the computation time of up to 140 times compared to the CPU

References

1. Lurie A.I. Analytical Mechanics -Moscow:“Fizmatlit” Publisher-1961-824p. “Russian”.
2. Rozhentsov A.A., K.V. Morozovskiy, A.A. Baev The modified generalized Hough transformation of processing three-dimensional images with unknown parameters of rotation and scaling / A.A. Rozhentsov, K.V. Morozovskiy, A.A. Baev // Optoelectronics. - V.49. - Novosibirsk, 2013. -№2. -p. 30-41. “Russian”.
3. Zhilin P.A. Vectors and tensors of the second rank in three-dimensional space. St. Petersburg: “Nestor” Publisher, 2001.p. 276. “Russian”.
4. Rozhentsov, A.A. Estimation of parameters and pattern recognition of three-dimensional images of objects with

disordered points / A.A. Rozhentsov, A.A. Baev, A.S. Naumov // Optoelectronics. - V.46. - Novosibirsk, 2010. -№1.-p.57-69. “Russian”.

5. Furman, J.A. Matching angular parameters and vector descriptions of 3D objects group point / Y.A. Furman, I.L. Egoshina, R.V. Eruslanov // Optoelectronics. - Novosibirsk, 2012. - V.48, № 6. - p. 3-17. “ Russian”.

Appendix

Solution 1:

$$m_2 = f_2 = \frac{\sqrt{(1-\cos(\alpha))(v_2'+v_2)^2((-v_2^2-v_3^2-v_1^2)\cos(\alpha)+v_2'v_2+v_3'v_3+v_1'v_1)}+(v_1'v_3-v_3'v_1)\cos(\alpha)+v_1'v_3-v_3'v_1}{(v_1^2+v_1'v_1+v_3^2+v_3'v_3+v_2(v_2'+v_2))\sin(\alpha)}$$

$$m_3 = f_3 = \frac{(v_3'+v_3)\sqrt{((-v_2^2-v_3^2-v_1^2)\cos(\alpha)+v_2'v_2+v_1'v_1+v_3'v_3)(\cos(\alpha)+1)(v_2'+v_2)^2}-(\cos(\alpha)+1)(v_2'+v_2)(-v_2'v_1+v_1'v_2)}{(v_2^2+v_2'v_2+v_1^2+v_1'v_1+v_3(v_3'+v_3))(v_2'+v_2)\sin(\alpha)}$$

$$m_1 = f_1 = \frac{(-v_2m_2^2-m_2v_3m_3+v_2)\cos(\alpha)+v_2m_2^2+m_2v_3m_3-v_2'+\sin(\alpha)m_3v_1}{-m_2v_1+m_2\cos(\alpha)v_1+\sin(\alpha)v_3}$$

Solution 2:

$$m_2' = f_2' = \frac{-\sqrt{(1-\cos(\alpha))(v_2'+v_2)^2((-v_2^2-v_3^2-v_1^2)\cos(\alpha)+v_2'v_2+v_3'v_3+v_1'v_1)}+(v_1'v_3-v_3'v_1)\cos(\alpha)+v_1'v_3-v_3'v_1}{(v_1^2+v_1'v_1+v_3^2+v_3'v_3+v_2(v_2'+v_2))\sin(\alpha)}$$

$$m_3' = f_3' = \frac{-(v_3'+v_3)\sqrt{((-v_2^2-v_3^2-v_1^2)\cos(\alpha)+v_2'v_2+v_1'v_1+v_3'v_3)(\cos(\alpha)+1)(v_2'+v_2)^2}-(\cos(\alpha)+1)(v_2'+v_2)(-v_2'v_1+v_1'v_2)}{(v_2^2+v_2'v_2+v_1^2+v_1'v_1+v_3(v_3'+v_3))(v_2'+v_2)\sin(\alpha)}$$

$$m_1' = f_1' = \frac{(-v_2m_2^2-m_2v_3m_3+v_2)\cos(\alpha)+v_2m_2^2+m_2v_3m_3-v_2'+\sin(\alpha)m_3v_1}{-m_2v_1+m_2\cos(\alpha)v_1+\sin(\alpha)v_3}$$

AN AUTOMATIC IMAGE SEGMENTATION ALGORITHM INVOLVING SHORTEST PATH BASINS[1]

T. Ryba[2]

[2] University of West Bohemia, Pilsen, Czech Republic, tryba@kky.zcu.cz

Image segmentation is a challenging task that is involved in almost every image processing system. Nowadays, methods based on graph theory are more and more popular. Nevertheless fully automatic methods that are both precise and robust are still hard to find. In this paper a new method based on shortest path in a graph is presented. This method automatically places seed points that are further used for image segmentation in the sense of path basins.

Introduction

Image segmentation is one of many fundamental problems in computer vision. Currently there are many different approaches for solving this task that are more or less successful. The gestalt theory described by Wertheimer in [6] shows that both boundary and region information should be used. In that case a segmentation method should lead to results that are precise and meaningful at the same time. Methods based on graph theory are capable of use both of these information and still producing results in an efficient way.

Related work

Using graph theory for image analysis is getting more and more popular nowadays. Image segmentation methods based on graph theory could be divided into five classes: methods based on shortest paths, minimal spanning trees, graph cut based methods with cost function, graph cut based methods on Markov random fields and the rest [1].

Further categorization could be done in sense of interactivity that needs to be done before running the algorithm. Probably the best known methods for image analysis that is based on graph theory is graph cut [8] and random walker [5]. This method could be used not only for image segmentation but basically for any kind of energy minimizing tasks as well.

While these methods are both efficient and precise they're also interactive. Therefore a set of user-defined or prelabeled pixels (so called seed points) must be given. These seed points could be obtained by user interaction or by some kind of preprocessing where an apriori information is needed. But not always such a preprocessing step or user interaction is possible and/or demanded.

Felzenszwalb and Huttenlocher presented in [2] an automatic segmentation method based on minimal spanning trees that incorporates differences between two sub-graphs and differences inside a sub-graph. Another well-known automatic method that is often used for method efficiency comparison was published by Shi and Malik in [3]. Here the minimization problem is formulated into a generalized eigenvalue problem. Another automatic method can be found in [4].

Representing an image as a graph

The algorithm presented here is based on graph theory thus it is necessary to represent an input image as an undirected weighted graph $G(V,E)$. Each node $v_i \in V$ corresponds to an image pixel and edges from the set E connect nodes with respect to a neighborhood system, Von Neumanns neighborhood was used in this work.

For each edge $(v_i, v_j) \in E$ a positive weight $w_{i,j}$ is calculated that corresponds to a similarity measure between two nodes. There's plenty of possibilities how to calculate these edge weights. Probably the most simple take use of absolute intensity difference:

[1] The work has been supported by the grant of The University of West Bohemia, project number SGS-2013-032.

$$w_{i,j} = \left|\mathrm{Im}(v_i) - \mathrm{Im}(v_j)\right|, \qquad (1)$$

where $\mathrm{Im}(v_i)$ corresponds to image intensity of the node (v_i). The formula (1) is simple and very fast to compute, however, it doesn't reflect any higher information of the image and in some cases its growth could be to slow. Thus we used in our work different well known formula [8], [1]:

$$w_{i,j} = \frac{1}{\exp\left(-\frac{\left(\mathrm{Im}(v_i) - \mathrm{Im}(v_j)\right)^2}{2\sigma^2}\right)}, \qquad (2)$$

where σ could be interpreted as a camera noise or expected intensity difference inside the segmented object [8].

Algorithm overview

The algorithm presented here is based on shortest path in a graph representing a segmented image. The shortest path is calculated in the sense of the single source variant of Dijkstra's algorithm [7]. That means that for a given source point the shortest paths to all other nodes in a graph are calculated. To achieve better efficiency of the algorithm it is possible to stop the calculation if there's no point that is closer to the source than a prespecified distance (threshold).

All points that are closer than this threshold form a basin and are considered to be in the same class. After the basin is determined another seed point needs to be chosen. Picking new seed point isn't straightforward and it's described in the next section.

Once all points have been labeled the algorithm ends. Few postprocessing steps could be done. For example it is possible to merge classes where the merging procedure could be driven by mean intensity difference of appropriate seed points or whole path basins.

Choosing seed points

To be the method automatic some heuristic for choosing seed points needs to be used.

A naive approach is to choose such a point that most differs from previous seed point. The difference could be represented by intensity difference or geometric distance, for example. The main drawback of this approach is that the two most different areas will switch among themselves until all their points will be considered as the seed points.

To overcome this problem we used a penalizing energy. This energy will increase in the neighborhood of current seed point thus making less possible that a point from this neighborhood will be chosen. The easiest way how to define the proper neighborhood is to use the Euclidean distance. It is reasonable to decrease the penalizing energy with increasing distance so that the farther points get lower penalty.

Nevertheless, using the Euclidean distance doesn't take into account information from any local image features. Therefore points that are close to the seed point but from different class would be penalized as well. Hence it is reasonable to involve an image feature that would tell us that we are crossing boundary between different classes. Using intensity gradient as this image feature is a good start. Points with high gradient represent the boundary points but crossing them doesn't necessary mean that we are crossing to a different class.

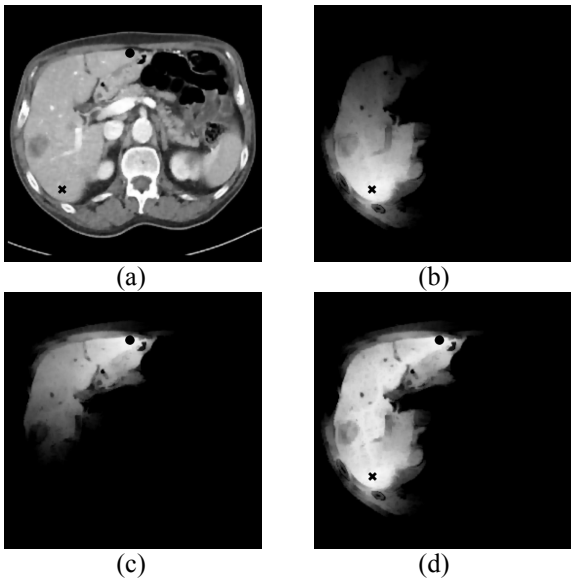


Fig. 1: In the input image (a) two seeds (cross and circle) are chosen. In figures (b) and (c) local energy of given seeds are shown, fig. (d) shows accumulative penalizing energy.

Better solution could be obtained again using the single source shortest path algorithm. Points with high gradient (edges) will stop the algorithm from further propagating but still the algorithm can reach the points from the same class that are

behind these edges. Another advantage is that we already had computed these distances from previous step. The energy at a point is therefore set to an opposite value of the shortest path as shown in equation (3):

$$E(p) = t - dist(p,s), \qquad (3)$$

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

Example of local energies for different seed points as well as overall energy for choosing a new seed point is shown in fig. 1.

To overcome the preference of choosing a new seed point near an edge in the image the final energy $E_f(p)$ at a point p is calculated as a weighted mean of calculated energy $E(p)$ and image gradient $\nabla \mathrm{Im}(p)$ at the same point:

$$E_f(p) = \alpha \cdot E(p) + (1-\alpha) \cdot \nabla \mathrm{Im}(p) \qquad (4)$$

Once a seed point is chosen this local energy is added to the overall (accumulative) energy. Than the next seed point will be chosen as a point with the smallest value of this energy.

Experiments and results

Here presented method was applied on different kind of images to show its usability. In all experiments we use the same values of parameters, specifically: distance threshold $t = 150$, weighting coefficient $\alpha = 0.5$ and number of iterations was set to 30. The first experiment was the segmentation of Lena image (fig. 2). The segmentation leads to meaningful results as shown in fig. 2 (d). Due to shading the face on the image is heterogeneous in image intensity and thus it was split to more regions.

(a)

(b)

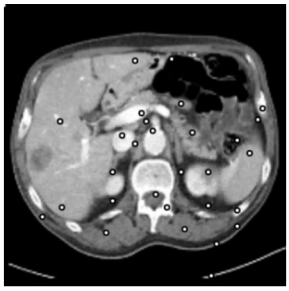
(c)

(d)

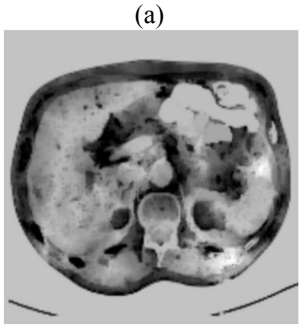
Fig. 2: Input image (a) is sequentially analysed and seed points are placed in it (b). Accumulative energy is shown in (c) and final segmentation in (d).

Another experiment was made on a medical image (fig. 3). It is a slice of CT of abdomen where the main part of the image represents the liver. As shown in fig 3 (d) the liver was segmented well but no seed was placed inside the hypodense tumor which is localized left in the liver. This is due to low contrast in the image in which case the spread of energy of one seed isn't affected with high image gradient.

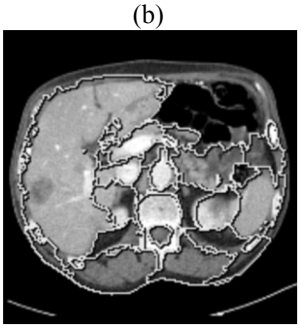
(a)



(b)



(c)

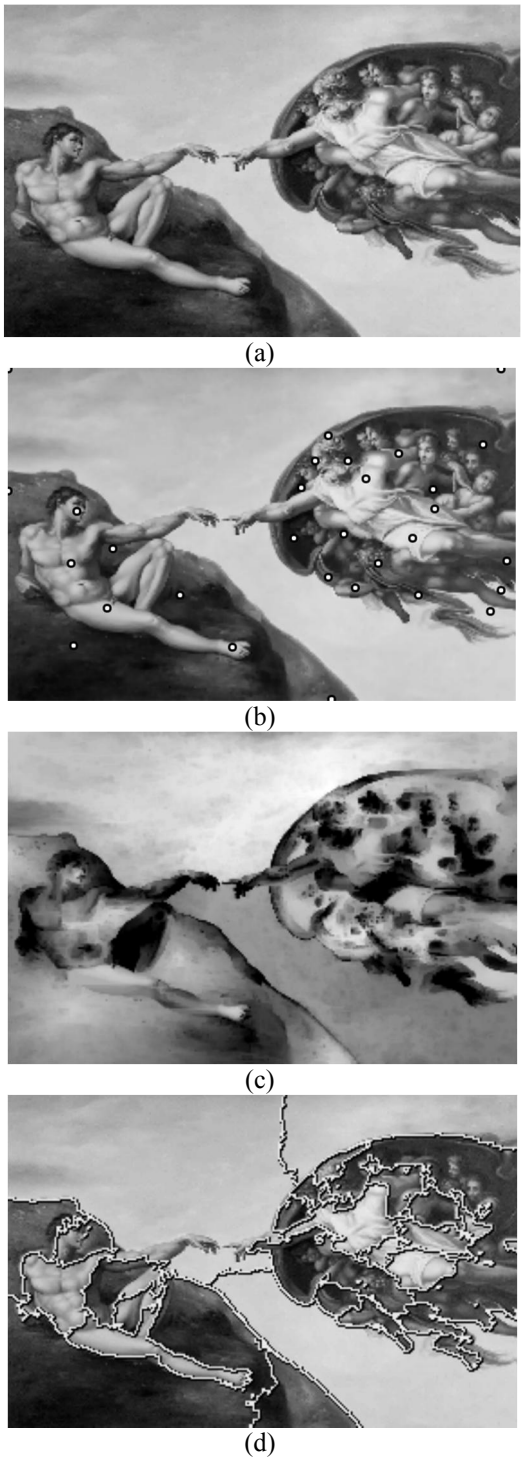


(d)

Fig. 3: Input image (a) shows a tomographic image of liver. Placed seed are shown in (b), accumulative energy is shown in (c) and final segmentation in (d).

Another possible drawback of this method is its iterative character. In the case where the input image consists of higher number of smaller objects or of non-homogenous objects the method needs higher number of iterations to segment the whole image. An example of this situation is shown on the well-known painting The Creation of Adam shown in fig. 4. The black regions in fig. 4(c) correspond to points that aren't in any path basin. Increasing number of iterations could deal with this problem

but it could also yield to an oversegmentation of the input image.



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

Fig. 4: Segmentation of painting The Creation of Adam (a). Automatically placed seeds are shown in (b), corresponding accumulative energy is shown in (c) and final segmentation in (d).

Conclusion

A new algorithm based on shortest distance in a graph was presented in this paper. This algorithm automatically places seed points in the image. A basin for each such placed seed point is then calculated and represents an image region. To place the seed points in a meaningful way an accumulative energy is calculated after a seed placement. This energy is based on single-source Dijkstra's algorithm for shortest path and image gradient is incorporated as well. New seed is then chosen as a point with smallest accumulative energy.

The performed experiments show the usability of this method on wide range of image types. A possible improvement of this method could lay for example in the merging of obtained regions or in improvements of placing the seed points. The method could be also seen only as a method for automatic placement of the seed points and the segmentation could be done with another method, e.g. graph cut [8] or random walker [5].

References

1. B. Peng, L. Zhang, D. Zhang. A survey of graph theoretical approaches to image segmentation. Pattern Recognition, 46 (3), pp. 1020-1038, 2013.
2. P.F. Felzenszwalb and D.P. Huttenlocher. Efficient graph based image segmentation. International Journal of Computer Vision, 59(2), pp.167-181, 2004.
3. J. Shi and J. Malik. Normalized Cuts and Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 22(8), pp. 888-905, 2000.
4. I.J. Cox, S.B. Rao, Y. Zhong. Ratio Regions: A Technique for Image Segmentation. In Proceedings of the International Conference on Pattern Recognition, pp. 557-564, 1996.
5. L. Grady. Random walks for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(11), pp.1768-1783, 2006.
6. M. Wertheimer. Laws of Organization in Perceptual Forms (partial translation). A Sourcebook of Gestalt Psycychology, pp. 71-88, 1938.
7. E.W. Dijkstra. A note on two problems in connexion with graphs. Numerische Mathematik, 1(1), pp. 269-27, 1959.
8. Y. Boykov, G. Funka-Lea. Graph Cuts and Efficient N-D Image Segmentation. International Journal of Computer Vision, 70(2), pp. 109-131, 2006.

MODEL SECRET SHARING SCHEMES IN SYSTEMS TRANSMIT VIDEO

D. Sagaydak[1,2], R. Faizullin[1,3]

[1] Omsk State Technical University, Department «Complex Protection of Information, Omsk, Russian Federation,
[2]sagaydak.dmitriy@gmail.com, [3]frt@omgtu.ru

This paper proposes two embodiments of secret sharing schemes on non-equal shares with an essentially small part of the secret for secure storage and video information processing in video data centers, which are based on RGB-representation (RBG: Red, Green, Blue) of each video frame. For each of the proposed scheme software implementation in the programming environment Borland C + + Builder 6 is made.

Statement of the problem

Now offered enough suitable secret sharing schemes, such as for example [1] and [2], but these works are not carried out a fundamental emphasis on the size of the secret. And just as no studies on the application secret sharing schemes to ensure the confidentiality of video data when they are stored on servers in the data center. Thus, in this paper we propose two variants of secret sharing schemes for unequal shares, where part of the secret will only be a small part of the transmitted video.

Description of the first secret sharing schemes

Suppose there is a stream of uncompressed video information expansion MxN. Initially this is the division of the video stream into frames, then for each frame pixels line by line reading is performed (from 1 to M). For each pixel of each line read there is its representation in the RGB format, that is three decimal numbers from 0 to 255, the values of which are translated into a binary number system and recorded one after the other. The result is a bit sequence consisting of 24 bits. $T=[r_1\ r_2\ r_3\ r_4\ r_5\ r_6\ r_7\ r_8\ g_1\ g_2\ g_3\ g_4\ g_5\ g_6\ g_7\ g_8\ b_1\ b_2\ b_3\ b_4\ b_5\ b_6\ b_7\ b_8]$, wherein r_1 - r_8, g_1 - g_8, b_1 - b_8 – serial number 24 bit sequence. Above each sequence the following permutation is performed: $\overline{T}=[r_1\ g_1\ b_1\ r_2\ g_2\ b_2\ r_3\ g_3\ b_3\ r_4\ g_4\ b_4\ r_5\ g_5\ b_5\ r_6\ g_6\ b_6\ r_7\ g_7\ b_7\ r_8\ g_8\ b_8]$.

Next, the sequence $\overline{T}$ is divided into blocks of four bits, which are the fractional part of decimal numbers a_i:

$a_1=0, r_1\ g_1\ b_1\ r_2$; $a_2=0, g_2\ b_2\ r_3\ g_3$;

$a_3=0, b_3\ r_4\ g_4\ b_4$; $a_4=0, r_5\ g_5\ b_5\ r_6$;

$a_5=0, g_6\ b_6\ r_7\ g_7$; $a_6=0, b_7\ r_8\ g_8\ b_8$.

Thus, for each pixel line of the image, an array A dimension $6\cdot M$ is made consisting of the received pixel for each a_i pixel of the image line. Next the average value of an array is found A:

$$sr_A=\frac{\sum_{i=1}^{6\cdot M} a_i}{6\cdot M} \qquad (1)$$

and formed a new array $\overline{A}$ dimension $6\cdot M$ consisting of elements $\overline{a_i}=sr_A-a_i$, wherein $i=1..6\cdot M$, sr_A - the value obtained in (1).

Next is the new array F, dimension $6\cdot N$ and consisting of elements:

$f_i=-a_{i-1}+2a_i-a_{i+1}$, if $i\neq 1$ and $i\neq 6\cdot M$;

$f_i=-a_{i-1}+2a_i-a_1$, if $i=1$ and $i\neq 6\cdot M$;

$f_i=-a_{6\cdot M}+2a_i-a_{i+1}$, if $i=6\cdot M$ and $i=1$.

From the right-hand side of the above equation we can compile a square and symmetrical with the respect to the diagonal matrix of the form:

$$H = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 & 0 & -1 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 & 0 \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 \\ -1 & 0 & 0 & 0 & 0 & -1 & 2 \end{pmatrix} \quad (2)$$

All the elements of the array F is less than 1, that is $f_i < 1$. From the experiments it was found that for later recovery of secretions it is necessary to make the calculation of the elements of the array F up to six to nine decimal places.

In various cases, the following can serve as a key:

a) The sequence of the signs of all elements in the array F, in such case an open information is an array $\overline{F}$ dimension $6 \cdot M \cdot N$ and is composed of elements $|f_i|$. The user will receive a key sequence of the signs of the elements of the array F, which is compressed without loss by the principle: $-- \rightarrow -$; $-+ \rightarrow +-$; $+- \rightarrow ++-$; $++ \rightarrow +++$.

b) The sequence of the signs of some of the elements of the array F, whereas in such case an open information is an array $\overline{F}$ dimensionality $6 \cdot M \cdot N$. The user will receive a key sequence consisting of retired characters and places of their withdrawal.

c) A sequence consisting of the average values sr_A obtained from (1).

Figure *1a* shows the original image frame, and Figure *1b* - shows a picture frame, for example, when an attacker attempts to recover the original image frame on the elements $|f_i|$ of the array $\overline{F}$. Recovery is made by direct translation of the array $\overline{F}$ elements $|f_i|$ in the pixel values in the format of RGB.

a) *b)*

Fig. 1 The original frame image (a), the frame image when you try to restore it by the elements of the array (b)

From the figure1b it is clear that when you try to restore the original frame image, some hard edges of image are observed. To avoid the appearance of these boundaries it is possible to implement the transformation of the original frame in a few rounds, two rounds are often enough. The resulting image after the first round is rotated 90 degrees (clockwise or counterclockwise) and is the second round. It is also possible, for example, instead of the second round perform a permutation of rows and columns as described in [3].

Figure 2 shows the image obtained when trying to restore the original frame image in the presence of an attacker array $\overline{F}$, consisting of the elements $|f_i|$, after two rounds of transformation of the original frame without performing any permutation of the columns and rows of the image after the first round of transformation.

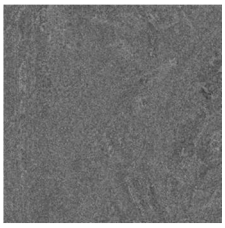


Fig. 2 Frame image when you try to restore it by the array elements after two rounds of the transformation of the initial frame image

Restoring the original image.

The user has the key S and data F. Since the rapid reconstitution of the secret is required it is supposed to use the Jacobi method, the matrix H shown in (2). For this matrix are the eigenvalues λ_i and eigenvectors υ_i, then the secret can be recovered in the following way: $F\overline{A} = S$, where $\overline{A}$ - is the sought-for secret.

1. $S = \sum_{i=1}^{6 \cdot M} \beta_i \cdot \upsilon_i \Rightarrow \beta_i = (S, \upsilon_i)$, i.e., the scalar multiplication of vectors S and υ_i, where S - the key (numeric vector), β_i - the scalar multiplication of the vector S by the eigenvectors υ_i of the matrix H.

2. We find the elements of the vector $\overline{A} = \sum_{i=1}^{6 \cdot M} \frac{\beta_i}{\lambda_i} \cdot \upsilon_i$. Source array elements A are equal $a_i = sr_A - \overline{a_i}$. After reverse permutation ($\overline{T} \Rightarrow T$), and performing the conversion value of the pixels from binary to decimal, will be able to restore the original image.

Description of the second secret sharing scheme

Let us consider the problem of storing a large number of data sets, the record length of which vary considerably. Suppose n bit vectors $A_1,...,A_n$ which dimensions are equal to $M_1,...,M_n$, respectively, and the variance M_i is evenly distributed in a fairly large range.

In this case, there arises a problem of economical data records which is currently solved in different ways, e.g., sharding, the introduction of various types of data such as CHAR and VARCHAR, separation of data markers , the creation of disk arrays such as RAID 0.

Consider the algorithm, which is a delta-primitivism code Elias (universal code for the encoding of integers, developed by Peter Elias) [4]. The first l bits are zeroed, where l - is the length of the record number n in the binary system, then the record itself is number n in the binary system, for example, suppose we are given $n = 3$ bit vectors, then write the number n in binary notation is 11, then $l = 00$. Therefore, the first step we obtain the following sequence of numbers - 0011. Next m_i bits are filled with zeros where m_i - the number of bits required to record the length of the vector A_i in the binary system.

Hence, in the above example $n = 3$, there are three bit vectors A_1, A_2, A_3, let $A_1 = 111011$, $A_2 = 10111$, $A_3 = 101$, then the dimension of these bit sequences are equal $M_1 = 6$, $M_2 = 5$, $M_3 = 3$, respectively $m_1 = 000$, $m_2 = 000$, $m_3 = 00$. Thus, the second phase would have the following sequence - 0001100001010011. The third step is to generate a sequence of consecutive record vectors themselves A_1, A_2, A_3. So, the above example would have the following sequence:0011000110000101001111101110111101.

It is also possible to use the method proposed above as the basis of secret sharing schemes for the video stream, and we try to construct an algorithm that does not require significant computational resources.

Suppose there is a certain flow ("TV"), the format of 720x576 pixels at 25 frames per second in the format RGB.

Dividing the given stream into frames is made. By-line reading of pixels frame is produced, and then for each pixel of the line its values in RGB format in the binary system (dimension of 24 bits, ie 8 bits for each color)is found and written one after the other in a single line, creating a sequence consisting of zeros and ones. Each sequence of quite long lines of read pixel is divided into n bit vectors $A_1, A_2, ..., A_n$, of different dimensions $M_1,...,M_n$. In its turn, the following operations are performed over the received bit vectors $A_1, A_2, ..., A_n$: $X_2 = A_3 \oplus A_4$,…, $X_{n-2} = A_{n-1} \oplus A_n$, $X_{n-1} = A_n \oplus A_1$ (where «$\oplus$» - bitwise addition modulo 2), received bit-vectors $X_1, X_2, ..., X_{n-1}$ are written one after the other. It should be noted that the bit vectors $A_1, A_2, ..., A_n$ have different dimensions and when executing «$\oplus$» if $A_{n-1} > A_n$ or $A_{n-1} < A_n$, smaller bit sequence is complemented to a greater bit sequence own repetition symbols, right, starting with the first. Performing the operation «$\oplus$» must retain the record length of the received bit vectors X_i, that is zeros received previously are not trimmed.

The resulting prefix bit vectors $A_1, A_2, ..., A_n$ for direct entry padded to the right a bit vector A_1 and, for example, maintained by the user. Recorded one after the other bit-vectors $X_1, X_2, ..., X_{n-1}$ are transferred to the data center.

Also to get bit vectors $X_1, X_2, ..., X_{n-1}$ instead of the operation «$\oplus$», we may use one of the block cipher encryption modes such as CBC (Cipher Block Chaining), CFB (XOR mode with feedback, Cipher Feedback). This

is necessary to remove a possible correlation between the bit vectors $X_1, X_2, ..., X_{n-1}$.

Thus, if the attacker gets to know the sequence consisting of a series of vectors X_i stored, he will not be able to restore the original sequence without knowledge of the generated prefix. If the bit sequence is stored at the data center server and the prefix is kept by the client, the proper level of confidentiality of stored information will be provided. If the attacker attempts to recover the original frame image, knowing the original image size, and only a bit sequence without a prefix, the result will be, for example, the following image (Figure 3).

a) *b)*

Fig. 3 The original frame image (a), the frame image when you try to restore it without the knowledge of the prefix

But even if the attacker will know the part of the prefix and the data will be divided into equal parts, but not A_1, he will still not be able to restore the original image as the problem is reduced to solving the uncertain system of n equations with $n+1$ unknowns:

$$\begin{cases} A_2 \otimes A_3 = X_1 \\ ..... \\ A_n \otimes A_1 = X_{n-1} \end{cases}$$

You can only restore the image selection bits, but when the record length A_1 is greater than 80 bits, the task becomes fundamentally unsolvable, because it is impossible to fulfill the computing using the equipment available.

Restoring the original image.

Assume that the user has full prefix formed on stage secret sharing, with a recorded track bit vector A_1. The server stores the data center portion of the recorded sequentially one after the other bit vectors X_i. Image size standard and it is known to all.

1) First, the analysis of the user's prefix is realised. The number of first l bit prefix member, equal to zero is analyzed. After recording a sequence of zeros we choose the sequence consisting of ones and zeros, which is equal to the length of a previous sequence of zeros. The selected sequence is translated from binary to decimal, and thus determine the number of bit vectors A_i.

2) The next step calculates the number of zeros written after the binary representation of bit vectors. And after these zeros the bit sequence is counted by the number of zeros. The resulting bit sequence is equal to the size of the binary M_1 vector A_1.

3) Next, a similar operation is carried out step 2) for the total number of vectors A_i defined in 1).

4) Thereby resulting in the steps 1) -3) is determined by the number of bit vectors A_i, and their length. Next, the vector A_1 is withdrawn from the prefix.

5) Then, taking into account the lengths of bit vectors A_i the system of equations is solved:

$$\begin{cases} A_2 \otimes A_3 = X_1 \\ ..... \\ A_n \otimes A_1 = X_{n-1} \end{cases}$$

Thus we find all bit-vectors A_i.

References

1. Chin-Pan Huang. Sharing the secret-dimensional images using conventional wavelets and multiveyvletov. / / Technology and Electronics, Vol 56b, № 1, P. 53-62.
2. A.A. Svench, R.T. Faizullin. Secret sharing scheme based on the metric characteristics of data for secure transmission of video conferencing. // Computer Optics V.31, № 1, 2007, Samara IPSI RAS - P. 105-111.
3. R.T. Faizullin, G.S. Rzhanitsyn. Building security system with video verification tasks graph isomorphism. / / Computer Optics, № 29, 2006, IPSI RAS, Samara, IPSI RAS - P. 84-93.
4. Elias P. Universal codeword sets and representations of the integers // IEEE Transactions on Information Theory, 1975, vol. IT-21, № 2. P. 194-203.

ENERGY SPECTRUM METHOD FOR IDENTIFICATION OF A LINEAR OBSERVATION MODEL WITHOUT CORRELATION MODEL OF INPUT SIGNAL[1]

V.V. Sergeyev[2,3], A.Yu. Denisova[2,4]

[2] IPSI RAS, Russia, 443001, Samara, Molodogvardeyskaya st. 151, [3]vserg@smr.ru, [4]anna3138@yandex.ru

This paper describes a modification of energy spectrum method for identification of linear observation model in the case when the model of correlation function of an input signal is unknown. As a distortion factors smooth and additive gauss noise are considered.

Introduction

Linear observation model describes signal transformation in video tract of remote sensing systems of the Earth [1]. In practice distortion identification seems to be very interesting problem, in particular for blind signal recovering. In the article [2] energy spectrum method was developed for images with isotropic exponential undivided autocovariance function. An estimation of an impulse response was dealt with estimation of input signal parameters such as correlation coefficient and variance in constraint that impulse response was narrow. In case of a priori absence of analytical model of correlation function (ACF) or difficulties in acquiring accurate experimental estimates of input signal parameters it is interesting to modify energy spectrum method in non parametric way.

Problem statement

Linear observation model [3] consists of input signal $x(n_1,n_2)$ *convolution with impulse response of distortion system* $h(k_1,k_2)$ *with addition independent from signal white noise* $v(n_1,n_2)$*:*

$$y(n_1,n_2)=h(n_1,n_2)**x(n_1,n_2)+v(n_1,n_2),\quad 0\le n_1,n_2\le N-1 \qquad (1)$$

where n_1,n_2 - integer arguments. Let input signal in the model (1) to be stationary random process with null average. Problem is to find unknown impulse response $h(k_1,k_2)$ using only output signal $y(n_1,n_2)$.

Further we will assume that:

1. *Impulse response is finite:*

$$\begin{cases} h(k_1,k_2)\neq 0, & |k_1|\le K,\ |k_2|\le K \\ h(k_1,k_2)=0, & \text{иначе} \end{cases}, \qquad (2)$$

and it holds normalization condition:

$$\sum_{k_1=-K}^{K}\sum_{k_2=-K}^{K} h(k_1,k_2)=1. \qquad (3)$$

Also frequency response (Fourier transformation of impulse response) is:

$$H\left(e^{i\omega_1},e^{i\omega_2}\right)=\sum_{k_1=-\infty}^{\infty}\sum_{k_2=-\infty}^{\infty} h(k_1,k_2)e^{-i(\omega_1k_1+\omega_2k_2)} \qquad (4)$$

real and non negative function.

2. *Impulse response has smooth effect, i.e. affects signal as a low pass filter, that is true for all distortion systems in practice in remote sensing.*

Energy spectrum method

Energy spectra (ES) of input and output signal are related by expression [3]:

$$\Phi_Y\left(e^{i\omega_1},e^{i\omega_2}\right)=\left|H\left(e^{i\omega_1},e^{i\omega_2}\right)\right|^2\Phi_X\left(e^{i\omega_1},e^{i\omega_2}\right)+D_V, \qquad (5)$$

where $\Phi_X\left(e^{i\omega_1},e^{i\omega_2}\right)$, $\Phi_Y\left(e^{i\omega_1},e^{i\omega_2}\right)$ – ES of input and output signals respectively. A module of frequency response can be derived from

[1]This research was financially supported by:
- Russian Foundation for Basic Research, grants: № 13-01-12080-ofi, 13-07-97006-r_povolgie_a;
- the Ministry of Education and Science of the Russian Federation (under the decree of the RF Government № 218, date 09.04.2010: contract № 02.G36.31.0001, date 12.02.2013)
- Basic Research Program of the Presidium of the Russian Academy of Sciences "Fundamental Problems of Informatics and Information Technologies" project 2.12.

formula (5), but if the first constrained is valid module can be ignored, i.e. frequency response is reconstructed unambiguously:

$$H\left(e^{i\omega_1}, e^{i\omega_2}\right) = \sqrt{\frac{\Phi_Y\left(e^{i\omega_1}, e^{i\omega_2}\right) - D_V}{\Phi_X\left(e^{i\omega_1}, e^{i\omega_2}\right)}}. \quad (6)$$

Than impulse response can be calculated by means of inverse Fourier transform from frequency response [3]:

$$h(k_1, k_2) = \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} H\left(e^{i\omega_1}, e^{i\omega_2}\right) e^{i(\omega_1 k_1 + \omega_2 k_2)} d\omega_1 d\omega_2. \quad (7)$$

As follows from formulas (6) and (7), it is needed to have estimations of input and output signal ES and noise variance to evaluate impulse response.

Estimation of noise variance can be produced by averaging ES in high frequency domain. [2].

The most difficult problem is to estimate input signal energy spectrum because there is no given analytical expression for it. It seems to be possible to estimate input ES from output signal ACF.

The idea of nonparametric estimation of input signal energy spectrum

Expression (6) makes relation only between ES of input and output signals, i.e., it is enough to have signal $\tilde{x}(n_1, n_2)$ with equivalent to origin source signal energy representation and to use ES of equivalent signal in formula (7).

Impulse response eliminates high frequency component of input signal according to second constraint. Using sharpen filter and noise reducing filter must have effect of signal restoration and filtered image can be used as energy equivalent estimation of input signal $\tilde{x}(n_1, n_2)$. Sharpen filter must approximate high frequency domain of input signal and it is needed noise removal filter, because there is no noise in input signal.

Implementation and experimental research

Median filter was implemented as noise removal part of composite filter because of its good noise reduction properties [3]. As for sharpen part of composite filter can be used extreme filter or combination of rang filters [3].

Algorithm of input signal ES estimation implemented here as an example is iterative algorithm and uses median and rang filters. A condition to stop iterations is that root mean square error of signal on neighbor iterations must be less than given threshold Δ or when iterative process accedes given maximum number of iterations.

On each iteration:

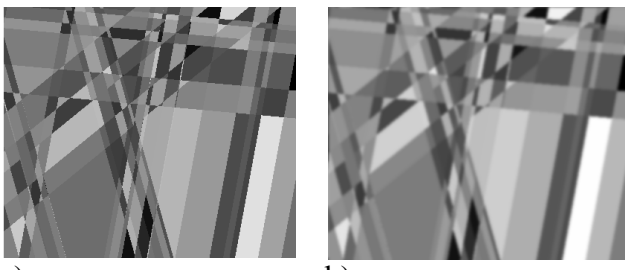
1. Median filtration with window 3×3 is done.
$$x_1(n_1, n_2) = \underset{m_1, m_2 \in W(n_1, n_2)}{med}\left\{y(m_1, m_2)\right\}, \quad (8)$$
2. Than maxmin filtration with window 3×3 is produced:
$$x_2(n_1, n_2) = \begin{cases} \max\limits_{m_1, m_2 \in W(n_1, n_2)}\left\{x_1(m_1, m_2)\right\}, & x_1(n_1, n_2) > \bar{x}_1 \\ \min\limits_{m_1, m_2 \in W(n_1, n_2)}\left\{x_1(m_1, m_2)\right\}, & x_1(n_1, n_2) \le \bar{x}_1 \end{cases}, \quad (9)$$

where $W(n_1, n_2)$ is set of indexes belonging to window at the position (n_1, n_2), $\bar{x}_1$ – average of window elements.

Filtered image is used as "equivalent" energy representation of input signal $\tilde{x}(n_1, n_2)$. ES $\Phi_Y\left(e^{i\omega_1}, e^{i\omega_2}\right)$ and $\Phi_X\left(e^{i\omega_1}, e^{i\omega_2}\right)$ calculated as periodogramms with Gaussian masking and then formulas (7) and (8) are used to estimate impulse response.

It is necessary to filter image iteratively because of rang and extreme filters give much contour distortions on angles and median filtration can compensates some of them. It was chosen smallest window size because angle distortions is also smaller with small windows.

It is preferred to use maxmin filter than extreme filter also because of smaller angle distortions of first.



a) b)

Fig. 1 – a) original image, b) smoothed by Gaussian window 11×11 image

Experiments were done for set of mosaic fields [4] with random lines directions (Fig.1)/ Intensity of lines regulates correlation

coefficients. In experiment correlation coefficients were from 0,7 to 0,99.
In experiment Gaussian impulse responses with widows sizes from 3×3 to 21×21 where used.
Experiment shows that without noise new energy spectrum method modification allows reconstruct 90% of most significant part of impulse response with relative error less than 25%. The example of restoration is on the Fig.2.

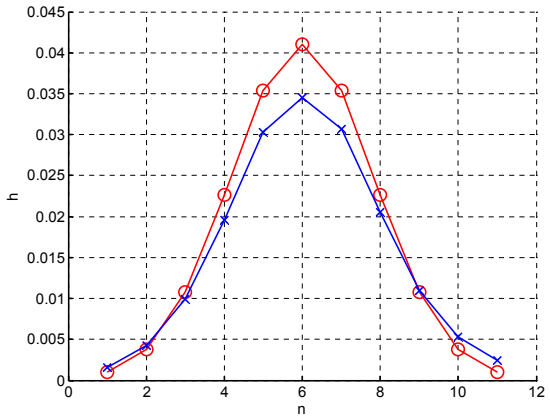


Fig.2. Experiment with Gaussian window with size 11×11: red line – original impulse response, blue line – restored impulse response.

But noise addition makes results not applicable.

Conclusion

Proposed in this paper modification of energy spectrum method allows to estimate impulse response from output signal with unknown correlation model.
Method of construction energy equivalent input signal approximation requires effective noise and sharpen filters with good performance on corners.

References

1. Shovengerdt, R. A., Remote sensing. Models and methods of image processing / Shovengerdt, R. A. - Moscow: "Tekhnosfera" Publishing. – 2010. [In Russian]
2. Sergeyev, V.V., Spectral-Energy Identification Method of the Linear Observation Model for Remote Sensing of the Earth/ V.V. Sergeyev, A.Yu. Denisova //Pattern Recognition and Image Analysis. – 2011. – Vol.21, No.2. – P. 311-313.
3. Methods of computer processing of images /Edited by V.A. Soifer – 2 ed. Ispra. – Moscow: Fizmatlit, 2003. – 784 Pp. [In Russian].
4. Modestino, J. W., Stochastic image models generated by random tessellations of the plane / J.W. Modestino, R.W. Fries, A.L. Vickers //Computer graphics and image processing.– 1980. – Vol. 12(1). – P. 74-98.

PRESENTATION AND COMPARISON METHODS FOR SEMANTICALLY DIFFERENT IMAGES

N.L. Shchegoleva[1], G.A. Kukharev[1], E.I. Kamenskaya[1]

[1]Saint Petersburg Electrotechnical University "LETI", ul. Professora Popova 5, St. Petersburg, 197376, Russia Tel.: +7(812) 234-26-82. E-mail: stilhope2009@gmail.com

This paper discusses the methods of presentation and comparison for semantically unrelated images with assessment of their similarity in original feature space, and in Space of Canonical Variables (SCV). Comparison in original feature space is based on color histograms, phase correlation and structural similarity index. The projection of the source images in SCV is implemented using two-dimensional canonical correlation analysis algorithm presented in this paper, and the measure of their similarity in SCV is based on the phase correlation.

Introduction

In recent decades, biometric technologies of human recognition and analysis rapidly develop [1]. Very complex problems are solved by means of biometric technologies, for example, problem human-computer interaction or problem of brain-computer interface. These tasks include observation of a person from different, seemingly unrelated, angles, and use of information from these observations for the interpretation or understanding of other observations. That is why recently interests of specialists in computer science and pattern recognition were devoted to mathematical methods allowing to transform two unrelated sets of data from the original feature space into a new common features subspace in which the original data sets are highly correlated. These methods are based on projections Partial Least Squares (PLS) and Canonical Correlation Analysis (CCA). CCA method was presented by Hotteling in the 1930s and initially described relationship between two one-dimensional data sets [2] with forecast of its application to multidimensional data. Two problems were found when using CCA for processing of digital images. The first problem is related to multi-dimensionality of initial data - images as 2D objects, with the third dimension determining brightness and the fourth dimension defining color components. The second is related to the problem of Small Sample Size, when dimension of the original feature space significantly exceeds the number of images in the original sample [3].

2D CCA/2D KLT algorithms

Suppose that we are given two sets of input data consisting of K images of size $M \times N$ each:

$$X = [X^{(1)} X^{(2)} \dots X^{(K)}],\ Y = [Y^{(1)} Y^{(2)} \dots Y^{(K)}] \quad (1)$$

where $X^{(k)}, Y^{(k)}$ – pairs of images, wherein $MN >> K$. We define the average images in each set of raw data:

$$\overline{X} = \frac{1}{K}\sum_{k=1}^{K} X^{(k)},\ \overline{Y} = \frac{1}{K}\sum_{k=1}^{K} Y^{(k)},\ k=1, 2, \dots, K.$$

If we assume that $\overline{X}^{(k)} = (X^{(k)} - \overline{X})$ and $\overline{Y}^{(k)} = (Y^{(k)} - \overline{Y})$, then instead of (1) we obtain:

$$\begin{aligned} \widetilde{X} &= [\overline{X}^{(1)} \overline{X}^{(2)} \dots \overline{X}^{(K)}] \\ \widetilde{Y} &= [\overline{Y}^{(1)} \overline{Y}^{(2)} \dots \overline{Y}^{(K)}] \end{aligned}. \quad (2)$$

Then we calculate eight covariance matrices $C_{xx}, C_{xy}, C_{yy}, C_{yx}$ for rows (r) and columns (c) of source images respectively [4].

To further implement CCA we calculate four total scattering matrices:

$$S^{(total\,1,r)} = [C_{xx}^{(r)}]^{-1} C_{xy}^{(r)} [C_{yy}^{(r)}]^{-1} C_{yx}^{(r)}; \quad (3a)$$

$$S^{(total\,2,r)} = [C_{yy}^{(r)}]^{-1} C_{yx}^{(r)} [C_{xx}^{(r)}]^{-1} C_{xy}^{(r)}.$$

$$S^{(total\,1,c)} = [C_{xx}^{(c)}]^{-1} C_{xy}^{(c)} [C_{yy}^{(c)}]^{-1} C_{yx}^{(c)}; \quad (3b)$$

$$S^{(total\,2,c)} = [C_{yy}^{(c)}]^{-1} C_{yx}^{(c)} [C_{xx}^{(c)}]^{-1} C_{xy}^{(c)}.$$

The goal of 2D CCA is to determine the four projection matrices W_{x_1}, W_{x_2} and W_{y_1}, W_{y_2},

which transform raw data into the space of variables $X^{(k)} \to U^{(k)}$ and $Y^{(k)} \to V^{(k)}$. This is achieved by solving the four eigenvalue problems:

$$S^{(total\ 1,r)} W_{x_1}^{(r)} = \Lambda_{x_1}^{(r)} W_{x_1}^{(r)}; \quad S^{(total\ 2,r)} W_{y_1}^{(r)} = \Lambda_{y_1}^{(r)} W_{y_1}^{(r)} \qquad (4a)$$

$$S^{(total\ 1,c)} W_{x_2}^{(c)} = \Lambda_{x_2}^{(c)} W_{x_2}^{(c)}; \quad S^{(total\ 2,c)} W_{y_2}^{(c)} = \Lambda_{y_2}^{(c)} W_{y_2}^{(c)}, \qquad (4b)$$

where: $\Lambda_{x_1}, \Lambda_{y_1}, \Lambda_{x_2}, \Lambda_{y_2}$ - the diagonal matrices of eigenvalues; W_{x_1}, W_{x_2} and W_{y_1}, W_{y_2} - the projection matrices. The transformation of raw data into a new space is implemented as two-dimensional Karhunen-Loeve transform (KLT) in the form of:

$$U^{(k)} = W_{x_1}^T \overline{X}^{(k)} W_{y_1}, \quad V^{(k)} = W_{x_2}^T \overline{Y}^{(k)} W_{y_2} \qquad (5)$$

$\forall k$, where $U^{(k)}, V^{(k)}$ - matrices representing the original images in the new feature space.

The main components are the ones that have the highest values. Basic variability of covariance matrices can be described by $d<<K$ largest eigenvalues. The number of principal components for rows $\lambda_i^{(r)}$, i= 1, 2, …, d_1 and columns $\lambda_i^{(c)}$, i= 1, 2, …, d_2 may be estimated by any known method.

It should be noted that $d_1 << M$; $d_2 << N$ and $d_1 \neq d_2$ in general case. The lower limit of parameter d can be chosen experimentally at the stage of verification of the analysis.

For dimensionality reduction procedure (5) should be modified. To do this we choose d rows of matrices $W_{x_1}^T, W_{y_1}^T$ corresponding to d largest eigenvalues, and based on those we form the matrices of reduction F_{x_1} and F_{y_1}. We also choose d columns of matrices W_{x_2}, W_{y_2} corresponding to d largest eigenvalues, and based on those we form the matrices of reduction F_{x_2} and F_{y_2}. Then we perform the "truncated" two-dimensional Karhunen-Loeve transformation, which can be represented in the following form [3, 4] (the symbol "^" determines the difference of result from (5):

$$\hat{U}^{(k)} = F_{x_1} \overline{X}^{(k)} F_{x_2}, \quad \hat{V}^{(k)} = F_{y_1} \overline{Y}^{(k)} F_{y_2}, \quad \forall k. \qquad (6)$$

The size of resulting matrices in (6) is equal to $d \times d$ or $d_1 \times d_2$. The resulting matrices of variables are defined as

$$\hat{U} = [\hat{U}^{(1)} \hat{U}^{(2)} \ldots \hat{U}^{(K)}], \quad \hat{V} = [\hat{V}^{(1)} \hat{V}^{(2)} \ldots \hat{V}^{(K)}] \ \forall k. \qquad (7)$$

Their dimensions will be $d \times d \times K$ or $d_1 \times d_2 \times K$.

Image database for experiments

Consider the case when the original images do not belong to the same global class. An example of this is the images from the set «People and Dogs», containing portraits of dogs and their owners. Several sample images are shown in Fig. 1(*a*) [5]. Here each pair of images contains a portrait of a person, presumably dog owner, and a front view of his dog. The image base allows to experimentally test how our visual assessment of similarity of these portraits in their original form is comparable to the formal assessments of similarity in the original feature space and in the subspaces.

Comparison of images in the original feature space

If we consider pairs of images in Fig. 1(*a*), we can see a resemblance between an owner and his dog. This resemblance is determined by the same camera angle, similar form and color of person's and dog's hair, appropriate color and shades of person's clothing.

Let's calculate color brightness histogram for each image. Example of this histograms and mutual phase correlation between histograms for pair 1 shown on Fig. 1(*b*).

The peak of this correlation is greater than 0.75, with a relatively small values of its side, which indicates a high degree of similarity between the images in each pair. Maximum correlation (≈0.9) is reached for a pair number 1, whose images are the most expressive similarities. This means that the subjective and formal evaluation color factor indicate the existence of a similarity in the two images are semantically unrelated.

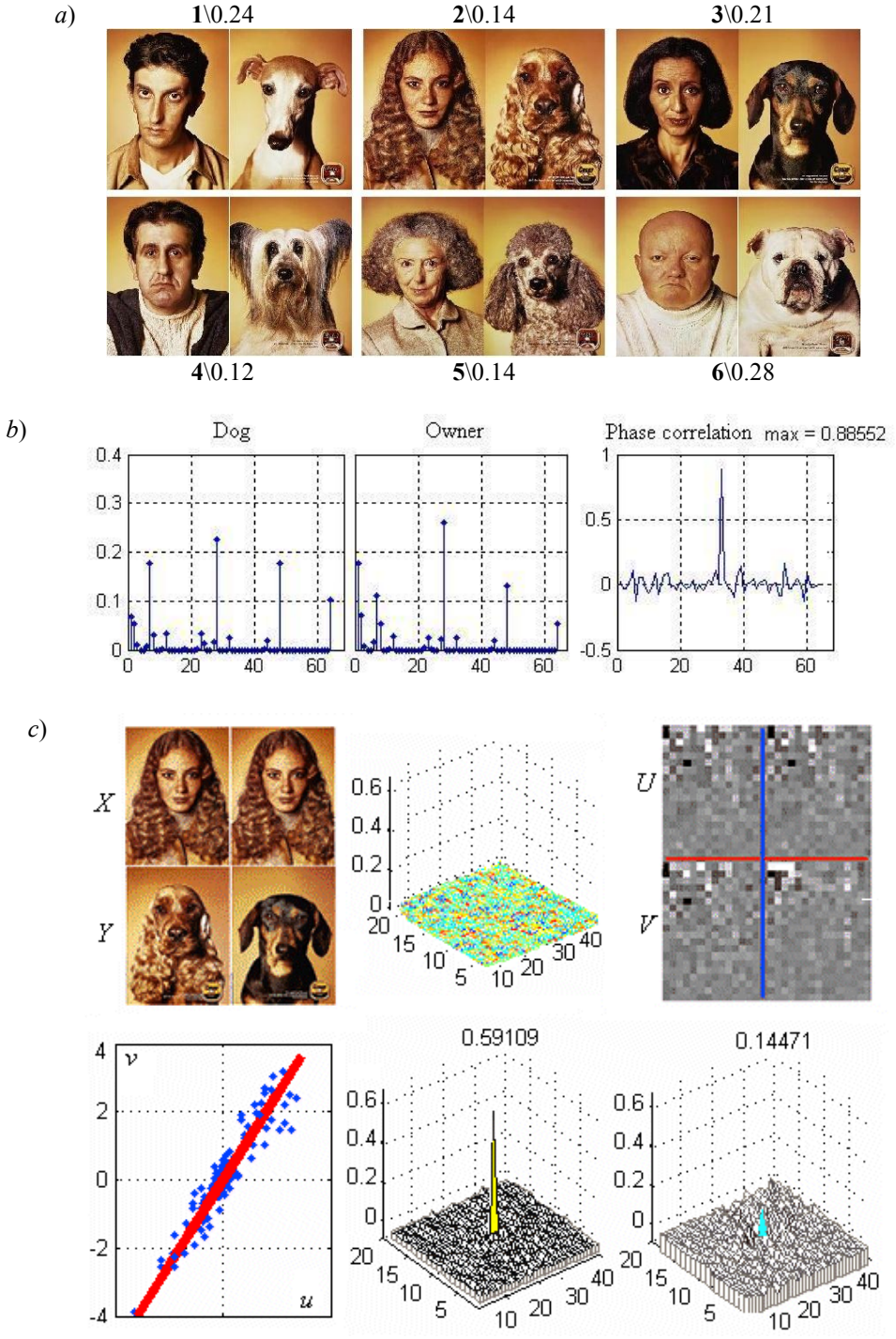


Fig. 1. Input data and processing result

Let's calculate mutual phase correlation between the image of an “owner” and a dog for a pair number 1. The top middle graph Fig. 1(*c*) show that the mutual phase correlation is almost zero. While it seems to us that the portraits are similar to each other, there is no actual correlation!

Let's compare their by means Structural SIMilarity Index (SSIM) [6]. The number of the pairs “owner-dog” and their SSIM value are shown in Fig. 1(*a*). Note that the values are quit low.

Representation of semantically different images in the space of canonical variables

Now let's check correlation between images "owner-dog" in the space of canonical variables. Two pairs of source images are shown at left top Fig. 1(c), the canonical variables U and V, corresponding to the original images are shown at top right Fig. 1(c). It is easy to see that three of the four canonical variables contain the same values

(luminance components) and, therefore, even look alike.

A graphical representation of the relationship between the variables in the new feature space for CCA, obtained by (7), is shown at left bottom Fig. 1(c). Note that canonical variables, shown as a function of U from V, have virtually linear relationship.

The phase correlation between the variables corresponding to the images from the same class is shown at bottom middle Fig. 1(c).

The phase correlation between the variables corresponding to the images from different classes is shown at bottom right Fig. 1(c).

From the results it is clear that in SCV between the variables U and V there was a correlation. This phase correlation between the variables from the same class is much higher value of 0.5 and significantly higher than the corresponding SSIM values. Correlation in the other case can be considered negligible, and the ratio of these correlations is 4 times.

High correlation of variables U and V in the space of canonical variables, their linear relationship to each other, "matching behavior" create prerequisites for mutual recognition and modelling of some variables from others. These tasks are easily implemented in the space of canonical variables.

Similar results were obtained for other pairs of source images, although we assumed that pairs of original images (owner-dog) belonged to semantically different global classes, and in fact their similarity has been proven only at the color level.

The main result of the analysis is as follows:

1) the fact of the correlation between the images of the dog and its owner in the space of canonical variables is confirmed, while any other ways could not confirm it;

2) indexing of images can be done using 2D CCA methods (search, recognition, model mapping of one image to another, reconstruction of images).

In general, this using example of developed algorithms 2D CCA/2D KLT has shown that they can be widely applied in recognition and classification of images, and also to decrease redundancy of image representation. The latter is due to the fact that selected value of parameter d << MN (MN is dimension of the original feature space).

Conclusion

The paper discussed the methods of presentation and comparison of semantically different images with assessment of their similarity in the original feature space, and presented the 2D CCA/2D KLT algorithm to implement the projection of these images and assessment of their similarity in the space of canonical variables.

To compare images in the original feature space we used their color histograms and phase correlation, two-dimensional phase correlation, and the structural similarity index. However, we could only partially prove similarity corresponding to subjective comparison of selected images.

It is shown that the «dissimilar» in the original feature space may be similar in the space of canonical variables. The projection in the space of canonical variables is implemented using 2D CCA/2D KLT presented in the paper and specifically designed to handle two sets of images. The results prove that 2D CCA/2D KLT methods can be widely used in search, pattern recognition and image classification tasks, and to decrease redundancy of images representation regardless of their semantic relationships.

References

1. Encyclopedia of Biometrics (Li Stan Z. (editor). – Springer Science+Business Media, 2009.
2. Hotelling H. Relations between two sets of variates // Biometryka. – 1936. –№28. – pp. 321 – 377.
3. Kukharev G., Forczmanski P. „Facial Images Dimensionality Reduction and Recognition by Means of 2DKLT". Journal Machine GRAPHICS & VISION, Vol.16, No. 3/4, 2007, pp. 401-425.
4. Kukharev Georgy, Tujaka Andrzej, Forczmański Paweł. Face Recognition using Two-dimensional CCA and PLS // International Journal of Biometrics. – 2011. – №3. – pp. 300 – 321.
5. www.popular-pics.com/ Funny_People_And_Dog_Similarity_Pictures__1
6. Dosselmann R., Yang X. D. A comprehensive assessment of the structural similarity index // SIViP. – 2011. –№5. – pp.81 – 91.

LOCALLY ADAPTIVE IMAGE DERINGING[1]

I. Sitdikov [2,3], A. Krylov [2,4]

[2] Faculty of Computational Mathematics and Cybernetics,
Lomonosov Moscow State University
3 thoughteer@gmail.com, [4] kryl@cs.msu.ru

In this paper we propose new locally adaptive image deringing methods. These methods are based on total variation minimization with an adaptively varying regularization parameter. The proposed approach improves visual quality of the resulting images by preserving more structural information comparing to existing methods. Attendant parallel algorithms were developed and implemented on GPU.

Introduction

We consider the problem of ringing artifact suppression, or deringing, for digital grayscale images. The ringing artifact can be seen as ripples or oscillations from high-contrast edges of the image. It is generally caused by partial or complete loss of high-frequency information. In practice, it usually appears as a result of excessive upsampling, compression, sharpening, or denoising of the image. It can also appear due to features of the measurement process, e.g. on MRI images [1]. Fig. 1 illustrates ringing induced by truncation of the image spectrum.

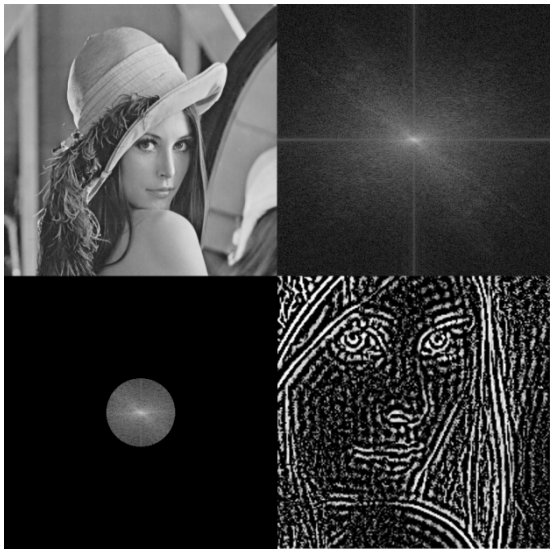


Fig. 1. Ringing induced by truncation of the image spectrum. Left-to-right, top-to-bottom: original image, spectrum of the original image, truncated spectrum, cropped and amplified absolute difference between the original and corrupted image.

This effect is known as Gibbs phenomenon.

Total Variation Minimization

One of the most widely used approaches to image deringing is total variation (TV) minimization [2]. The resulting image is obtained by solving the following strongly convex minimization problem:

$$J_1[u] = \| u - f \|_2^2 + \alpha \, \mathrm{TV}[u] \to \inf . \qquad (1)$$

Here f is a given distorted image, and TV is the total variation functional:

$$\mathrm{TV}[u] = \sum_p |\nabla u(p)|, \qquad (2)$$

where ∇u denotes the gradient of u, and the summation is taken over all the image pixels. Regularization parameter α stands for strength of ringing suppression: higher values correspond to smoother resulting images. An example is shown on Fig. 2.

As well as all image deringing methods this method deals with the trade-off between ringing suppression and keeping fine details of the image. In order to preserve more structural information (like textures) in the image, we propose locally adaptive image deringing methods, which select the regularization parameter at each pixel depending on local characteristics of the image.

[1] The work was partially supported by RFBR grant No. 13-07-00438

Fig. 2. Deringing via TV minimization. Distorted (*top*) and restored (*bottom*) image. The regularization parameter was set to the minimal value that visually suppresses the artifact. Distorted image: PSNR=30.78dB, SSIM=0.763. Restored image: PSNR=30.82dB, SSIM=0.778.

Locally Adaptive Methods

We start with determining optimal values of the regularization parameter at each pixel of a test image:

$$\alpha^*(p) = \inf \operatorname{Arg}\inf_{\alpha \geq 0} |u_\alpha(p) - u^*(p)|, \qquad (3)$$

where

$$u_\alpha = \arg\inf_u \left\{ \| u - f \|_2^2 + \alpha \, \mathrm{TV}[u] \right\}, \qquad (4)$$

and f is a distorted version of test image u^*. Optimal values of the regularization parameter are shown in Fig. 3. One can easily notice that the values should increase with the distance to the nearest edge. This fact encourages us to include an additional weighting function into the target functional:

$$\begin{aligned} &J_2[u] = \| \beta(u - f) \|_2^2 + \alpha \, \mathrm{TV}[u], \\ &J_2[u] \to \inf. \end{aligned} \qquad (5)$$

Higher values of β correspond to lower strength of artifact suppression, and vice versa. We select β in the following form:

$$\beta(p) = \exp\left\{ -\frac{\rho(p)}{3d} \right\}, \qquad (6)$$

where $\rho(p)$ is the distance from pixel p to the nearest edge, and d is the half-width of the oscillations, which is supposed to be known. Notice, that in this case functional J_2 is strongly convex, and there exists a unique solution of minimization problem (5).

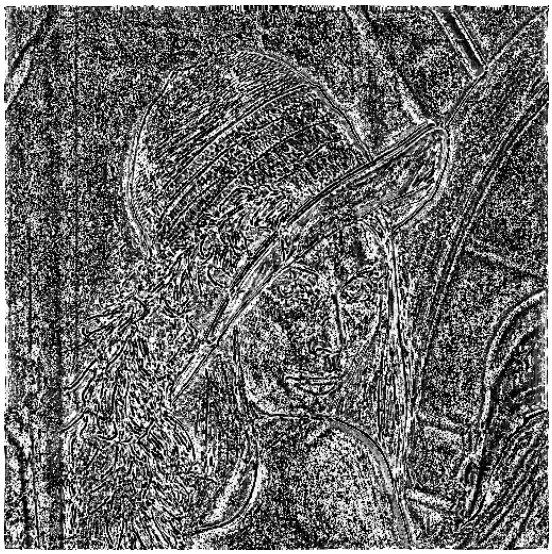


Fig. 3. Optimal values of the regularization parameter in the total variation minimization method (black pixels correspond to 0, white pixels correspond to 0.1).

Algorithm

The proposed locally adaptive image deringing methods involve the following multistage algorithm:

1. Apply the Canny edge detector [3] to given distorted image f and obtain set of edge points E. We use $\sigma = 1$ for the edge detection in our experiments. We apply neither threshold nor hysteresis.
2. Remove the edge points related to the oscillations, obtaining $E_M \subset E$.
3. For each pixel p calculate the distance to the nearest edge point: $\rho(p) = \inf_{q \in E_M} |p - q|$.

4. Calculate weighting function β, e.g. $\beta(p) = \exp\left\{-\frac{\rho(p)}{3d}\right\}$.
5. Minimize J_2 for fixed α. Return the solution as a result.

The described algorithm has only two parameters: regularization parameter α and half-width of the oscillations d.

On the second step of the algorithm we remove oscillations from the edge map obtained on the first step by checking the following condition for each pixel of the image [4]:

$$g(p) > G(p) = \max_q \frac{1}{2} g(q) \mathrm{e}^{-\frac{|p-q|^2}{2R_M^2}}, \qquad (7)$$

where g denotes modulus of the gradient of f, and R_M is the radius of mask G. In our experiments we use $R_M = 3d$. Condition (7) utilizes the fact that real edges have higher values of g than their oscillations. Efficient computation of the mask could be done as follows:

$$\begin{aligned} G(p) &= \exp\left\{-\frac{Y(p)}{2R_M^2}\right\}, \\ Y(p) &= \min_q \left\{X(q) + |p-q|^2\right\}, \\ X(q) &= -2R_M^2 \ln\frac{g(q)}{2}. \end{aligned} \qquad (8)$$

Here calculation of G is reduced to calculation of Y, which appears to be the generalized Euclidean Distance Transform (EDT) of X. There is a number of algorithms for performing an exact EDT in linear time [5].

On the third step we calculate $\rho(p) = \inf_{q \in E_M} |p-q|$. The calculation becomes trivial after reduction to EDT:

$$\begin{aligned} \rho^2(p) &= \min_q D(q) + |p-q|^2, \\ D(q) &= \begin{cases} 0, & q \in E_M, \\ \infty, & q \notin E_M. \end{cases} \end{aligned} \qquad (9)$$

Finally, we minimize functional J_2. Since J_2 is strongly convex, it is possible to minimize it using well-known subgradient descent [2]:

$$\begin{aligned} &u^0 = f, \\ &u^{n+1} = u^n - \tau_n J_2'[u^n], \\ &\tau_n = \frac{C}{n+1}, \ C = const, \\ &J_2'[u^n] = 2\beta^2(u^n - f) - \alpha \operatorname{div}\frac{\nabla u^n}{|\nabla u^n|}, \\ &u^* \approx u^N. \end{aligned} \qquad (10)$$

However, this method has an ultimately low rate of convergence. Recently, Yu. Nesterov proposed a more efficient algorithm for solving such non-smooth functional minimization problems [6]. Later it was successfully applied to the total variation minimization problem [7]. In our case that algorithm takes the following form:

$$\begin{aligned} &u^0 = f, \\ &G^{-1} = 0, \\ &L = 2 + 2\sqrt{2}\alpha\mu^{-1}, \\ &\eta^n = 2\beta^2(u^n - f) + \alpha \mathrm{TV}_\mu'[u^n], \\ &\mathrm{TV}_\mu'[u^n] = -\operatorname{div}\begin{cases} \frac{\nabla u^n}{|\nabla u^n|}, & |\nabla u^n| \ge \mu \\ \frac{\nabla u^n}{\mu}, & |\nabla u^n| < \mu \end{cases} \\ &v^n = u^n - \frac{\eta^n}{L}, \\ &G^n = G^{n-1} + \frac{n+1}{2}\eta^n, \\ &w^n = f - \frac{G^n}{L}, \\ &u^{n+1} = \frac{n+1}{n+2}v^n + \frac{2}{n+3}w^n, \\ &u^* \approx u^N. \end{aligned} \qquad (11)$$

Here μ is a smoothing parameter. Appropriate values of μ lie within the range $[0.001, 0.005]$. This algorithm provides an optimal rate of convergence, while preserving locality of the computations. It normally converges 10-20 times faster than the subgradient method.

In order to make our methods suitable for real-time and interactive applications we implemented all the attendant algorithms on GPU using the CUDA platform.

Results

A result obtained by the proposed locally adaptive method is shown on Fig. 4. Our approach leads to better PSNR and SSIM indices comparing to the original TV minimization method. These results also agree with the visual perception.

Fig. 4. Deringing via the proposed method. Distorted (*top*) and restored (*bottom*) image. The regularization parameter was set to the minimal value that visually suppresses the artifact. Distorted image: PSNR=30.78, SSIM=0.763. Restored image: PSNR=30.87, SSIM=0.781.

Our GPU version of the proposed algorithm is at least 10 times faster than the one implemented on CPU. It uses some basic image processing operations implemented in [8]. It takes less than 100 ms to process a 1MP grayscale image on the GeForce GTX 670 card.

Conclusion

We considered the problem of ringing artifact suppression for digital grayscale images. We proposed a family of new locally adaptive image deringing methods based on the TV minimization approach. The proposed methods provide better results in terms of PSNR and SSIM indices, as well as visual perception, comparing to the existing methods. Our GPU implementation of the proposed algorithm makes it suitable for real-time and interactive applications.

References

1. L.J. Erasmus, D. Hurter, M. Naude, H.G. Kritzinger, S. Acho. A short overview of MRI artefacts // South African Journal of Radiology. – 2004. – Vol. 8, No. 2. – P. 13-17.
2. A.V. Nasonov, A.S. Krylov. Adaptive Image Deringing // 19th International Conference on Computer Graphics GraphiCon'2009. – 2009. – P. 151-154.
3. J. Canny. A Computational Approach to Edge Detection // IEEE Transactions on PAMI. – 1986. – Vol. 8, No. 6. – P. 679-698.
4. A.S. Krylov, A.V. Nasonov. Adaptive Total Variation Deringing Method for Image Interpolation // Proceedings of International Conference on Image Processing. – 2008. – P. 2608-2611.
5. R. Fabbri, L. Costa, J. Torrelli, O. Bruno. 2D euclidean distance transform algorithms: a comparative survey // ACM Computing Surveys. – 2008. – Vol. 40.
6. Yu. Nesterov. Smooth minimization of non-smooth functions // Mathematical Programming. – 2005. – Vol. 103, No. 1. – P. 127-152.
7. P. Weiss, L. Blanc-Feraud, G. Aubert. Efficient Schemes for Total Variation Minimization Under Constraints in Image Processing // SIAM Journal on Scientific Computing. – 2009. - Vol. 31, No. 3. – P. 2047-2080.
8. I.T. Sitdikov, A.V. Nasonov, A.S. Krylov, Ding Yong. Parallel Implementation of Area Detection Algorithms for Image Ringing Artifact Analysis // Proceedings of 15th international conference "Digital Signal Processing and its Applications" DSPA'2013. – 2013. – Vol. 2. – P. 55-58. [in Russian]

SPECIFICS OF OBJECTIVE FUNCTIONS FOR RECURRENT ESTIMATION OF INTERFRAME GEOMETRIC DEFORMATIONS[1]

A.G. Tashlinskii, S.V. Voronov[2]

[2] Ulyanovsk State Technical University, ul. Severnyi Venets 32, Ul'yanovsk, 432027 Russia, telephone (88422)778102,
[3]tag@ulstu.ru, [4]s.voronov@ulstu.ru

Effectiveness of using mean squared difference, correlation ratio and mutual information as objective functions for estimation of interframe geometric deformations is analyzed. Additive Gaussian noise, linear and non-linear brightness distortions are considered disturbing factors.

Problem formulation

Estimation of interframe geometric deformations is one of the most widespread problems of digital image processing. There are many methods for solving this problem in Frequency domain [1] and in spatial domain [3]. Spatial domain methods operate directly on pixels and the problem of estimation of deformation's parameters $\boldsymbol{\alpha}$ becomes the problem of searching for the extreme point of multi-dimensional objective function $\mathrm{J}(\mathbf{Z},\boldsymbol{\alpha})$. 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 Ω on which images are defined. There is a wide variety of similarity measures, which can be used as objective functions [4]. The decision of which objective function to choose is usually based on specifics of images, deformation properties and conditions. The most common objective function for recursive estimation of interframe geometric deformations (e.g. stochastic gradient estimation [5]) are mean squared difference (MSD) and correlation ratio (CR). Recently, objective functions from the theory of information [2] are becoming more popular. Among these functions, the most interesting is mutual information (MI). Because of big computational cost, a use of MI was limited. This work is devoted to experimental comparison of the three objective functions mentioned above for different classes of interframe brightness distortions.

Formulas used for calculating

Following expression was used to estimate MSD:

$$\hat{\mathrm{J}}(\mathbf{Z},\boldsymbol{\alpha})=\frac{1}{\mu}\sum_{\bar{j}\in\Omega_Z}\left(\tilde{z}_{\bar{j}}^{(1)}-z_{\bar{j}}^{(2)}\right)^2, \qquad (1)$$

where $\bar{j}\in\Omega_Z$ - coordinates of samples $z_{\bar{j}}^{(2)}$ in set $Z=\left\{z_{\bar{j}}^{(2)},\tilde{z}_{\bar{j}}^{(1)}\right\}$ using to estimate $\mathrm{J}(\mathbf{Z},\boldsymbol{\alpha})$; $\tilde{z}_{\bar{j}}^{(1)}=\tilde{z}^{(1)}(\bar{j},\bar{\alpha})$ — values of brightness of continuous image $\tilde{Z}^{(1)}$ obtained from image $\mathbf{Z}^{(1)}$ using one of interpolation methods [6]; μ — size of a subset Z which is equal to cardinality of the set Ω_Z.

Correlation ratio can be defined as:

$$\hat{\mathrm{J}}(\mathbf{Z},\boldsymbol{\alpha})=r=\frac{\sum_{\bar{j}\in\Omega_Z}(\tilde{z}_{\bar{j}}^{(1)}-\hat{m}_{\tilde{z}1})(z_{\bar{j}}^{(2)}-\hat{m}_{z2})}{\mu\hat{\sigma}_{z1}\hat{\sigma}_{z2}}, \qquad (2)$$

where $\hat{m}_z$ and $\hat{\sigma}_z$ — estimations of mean and variance of images $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$. If r is not equal to ±1 than relationship between intensities $z^{(2)}$ and $\tilde{z}^{(1)}$ can be considered a

[1] Support of the RFBR grant 13-01-00555-a

measure of linearity using equation: $z_{\bar{j}}^{(2)} = \frac{\hat{\sigma}_{z1}}{\hat{\sigma}_{z2}}\left(\tilde{z}_{\bar{j}}^{(1)} - \hat{m}_{z1}\right) + \hat{m}_{z2}$.

Entropy estimations of images $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(2)}$ are needed to compute MI. They can be calculated using samples of images from subset Z

$$\hat{H}(\mathbf{Z}^{(1)}) = -\sum_i p_{z1}(z_i)\log p_{z1}(z_i),$$
$$\hat{H}(\mathbf{Z}^{(2)}) = -\sum_i p_{z2}(z_i)\log p_{z2}(z_i),$$

where p_{z1} and p_{z2} — estimations of marginal probability distributions of intensities of subsets $\left\{\tilde{z}_{\bar{j}}^{(1)}\right\}$ and $\left\{z_{\bar{j}}^{(2)}\right\}$ correspondingly. As these estimations we can use histograms of images; z_i - i-th level of intensity. Estimation of joint entropy:

$$\hat{H}(\mathbf{Z}^{(1)},\mathbf{Z}^{(2)}) = -\sum_i\sum_k p_{z1,z2}(z_i,z_k)\log p_{z1,z2}(z_i,z_k),$$

where $p_{z1,z2}$ — estimation of joint probability distribution. Because $\tilde{z}_{\bar{j}}^{(1)}$ and $z_{\bar{j}}^{(2)}$ are dependent, MI can be estimated in following way

$$\hat{\mathrm{J}}(\mathbf{Z},\bar{\alpha}) = \hat{H}(\mathbf{Z}^{(1)}) + \hat{H}(\mathbf{Z}^{(2)}) - \hat{H}(\mathbf{Z}^{(1)},\mathbf{Z}^{(2)}).$$

(3)

Notice that contrary to MI and CR, MSD shows the measure of dissimilarity. Therefore, when using MSD as objective function the goal is to minimize it.

Experimental results

Experiments were performed using simulated and real satellite images. The size of each image was 400x400 pixels. For simplicity, translation h among one of the axis within the range from -65 to 65 pixels was used as the parameter of interframe geometrical deformations. To estimate values of objective functions the following expressions were used: (1) – for MSD, (2) – for CR and (3) – for MI. Calculated values of MSD and MI have different dimensions. Their values were normalized (divided to the maximum values in examined range of translations) to make results easier to compare visually.

Let us consider some situations.

Examined images differ by unbiased additive Gaussian noise. To investigate the influence of noise on objective functions it is reasonable to use simulated images because we can priory define probability density function of intensities and correlation function. Simulated images with close to Gaussian probability density function of intensities and with correlation function close to Gaussian were synthesized based on a wave model [8]. Experiments were performed with noise-to-signal ratio q from 0 to 2 (q = noise's variance / signal's variance). Fig. 1 shows dependence of objective functions from translation h and noise with different variance (q=0.02, q=0.2, q=2). Hereafter curve 1 is MSD, 2 – CR, 3 – MI. One can see that in case of weak noise (q=0.02) MI has maximum slope. When using recurrent algorithms [7], this fact can lead to better convergence of estimates of interframe geometric deformations. Moreover, due to the biggest sharpness a use of MI can provide with more accurate estimates (the problem of computational cost is not considered in this work). However, when the level of noise increases the slope and sharpness of MI's characteristic decrease. When q changes from 0 to 2 the maximum of MI's characteristic decreased by 15 times. The MSD-curve behaves according to the theory (this function can be found analytically if correlation function and noise-to-signal ratio are given). The slope of the MSD-curve gradually decreases and the minimum increases from 0 to 0.52 (when q=2) with increasing noise level. CR-curve behaves approximately the same except the fact that it has maximum (not minimum) decreasing from 1 to 0.55.

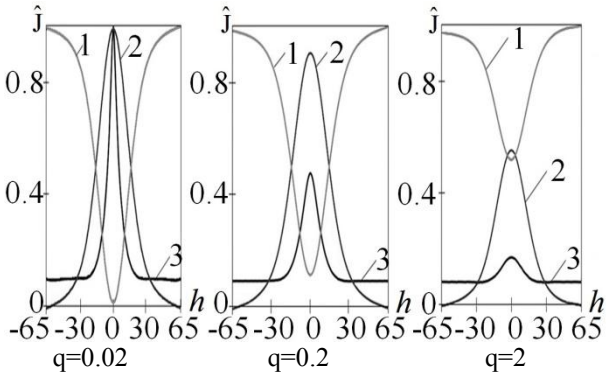


Fig.1. Estimation of objective functions for different noise-to-signal ration.

Fig. 2 shows the dependence of maximum slope (K) of MSD, CR and MI curves on q. One can see that when the noise level is low MI-curve has significantly higher slope in comparison with MSD and CR. However, when noise becomes stronger (approximately when q=0.1) MI starts to show worse results than MSD and CR. The maximum slope of MSD-curve smoothly decreases by 1.6 times and CR-curve – by 1.4 times with q increasing from 0 to 2. Thus, when images differ by additive noise and q >0.1 a use of MI potentially may not provide with a better convergence of recurrent algorithms used to estimate interframe geometric deformations.

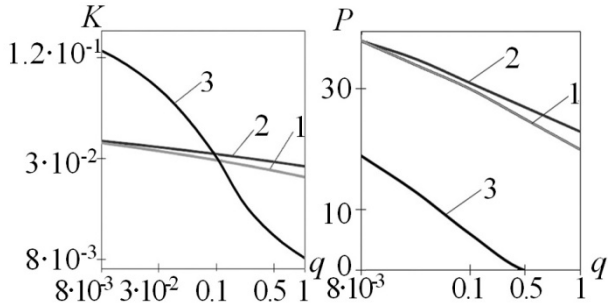


Fig.2. Maximum slope.

Fig.3. Potential effective range.

Effective range is another important measure of estimation procedures. It is a subdomain of parameters where estimates of interframe geometrical deformations have required accuracy with given constraints (on computational cost, the number of iterations etc.). One of the criteria of including a point in this subdomain is the slope of objective function's curve in this point. The slope need to exceed a threshold – a value which can't provide with the required convergence of estimates' vector $\boldsymbol{\alpha}$. To measure this parameter for investigated objective functions a conventional threshold was assumed. Potential effective range P of recurrent estimation procedures was estimated using this threshold (actual effective range depends on procedure's performance, a type of deformations etc.). Fig. 3 shows the dependence of potential effective range on q in case the threshold is equal to 0.035. One can see that MI shows two times worse result than MSD and CR. Without noise MSD and CR have equal P. Effective range of CR decreases by 1.7 times with increasing q to 1, for MSD – by 1.9 times. MI has non-zero effective range only when q ≤0.5.

Let us note that choosing parallel translation as $\boldsymbol{\alpha}$ doesn't affect much on generality of results because for any point the reasoning above is correct for Euclidian mismatch distance [5] as well. Euclidian mismatch distance integrally describes an arbitrary set of model's parameters of interframe geometric deformations.

Examined images differ by linear distortions of intensities. Fig. 4,a shows results when negative transformation was considered a linear distorion's function. In addition Gaussian noise was added to images. Due to inverse linear relationship between intensities of images, the value of CR has minimum when images are fully aligned. MSD also can't be used in this situation.

Images with non-linear distortions of intensities. Multimodal images (aquired by different sensors) is an example of images with non-linear distortions of intensities. Sattelite images acquired at different seasons and in different spectral range (fig. 5, where a half of one image is shown above the diagonale and a half of another image - below the diagonale) is an example of multimodal images. Fig. 4,b shows estimates of objective functions for this images. One can see that only CR and MI have sharp extremums. MI has higher slope in proximity of exremum. MSD does not have sharp extremum and can't be used as objective function, especially with a small subset.

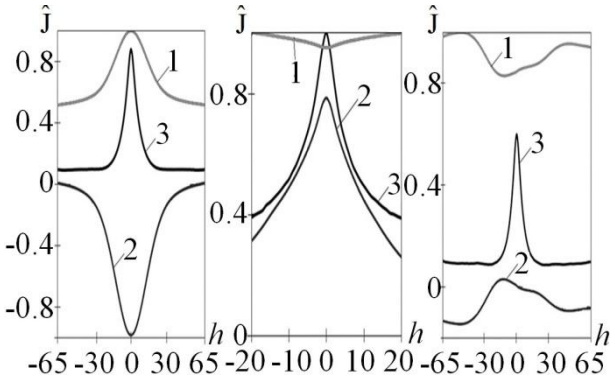


Fig. 4. Estimates of objective functions in case of linear and non-linear intensities' distortions.

In addition, when one of the images has just a region with negative transformation of pixels' intensities would also be an example of non-linear distortions of intensities. Fig. 5,c shows results of an experiment with

such images. Results show that in this situation MI is the only reasonable choice of objective function. Due to non-linear distortions, MSD and CR can't estimate magnitude of relationship between pixels' intensities.

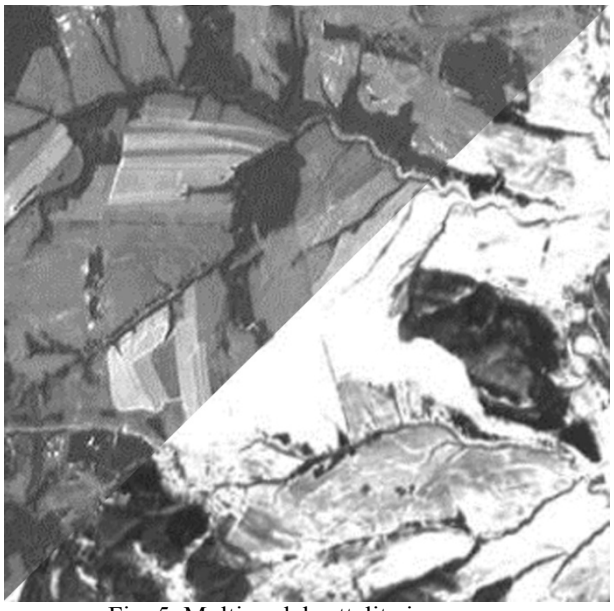


Fig. 5. Multimodal sattelite images.

Conclusion

Experiments show that MSD is the most reasonable choice of objective function for recurrent estimation of interframe geometrical deformations of images without multiplicative distortions of intensities. In case of low-level additive noise, MI has the highest slope. Therefore, it may have better convergence of estimation algorithms. However, when the level of noise raises the maximum of MI-curve dramatically decreases. CR and MSD provide with broader effective range. Consider this criteria MI loses by two times. MI shows the best results for multimodal images and images with linear distortion of intensities, CR is a bit worse. However, CR is much less computationally complex than MI. Finally, MI is the only effective measure (among investigated) for images with non-linear distortions of intensities.

References

1. Brown L G. A survey of image registration techniques / ACM C.S. 1992. V. 24. № 4. P. 325-376.
2. D'Agostino E, Maes F, Vandermeulen D, Suetens P. An information theoretic approach for non-rigid image registration using voxel class probabilities / Med Image Anal. 2006. V 6(3). P. 413-431.
3. De Castro E., Morandi C. Registration of translated and rotated images using finite Fourier transform / IEEE Transactions on Pattern Analysis and Machine Intelligence. 1987. Vol. 9. № 5. P. 700–703.
4. Goshtasby A.A. Image registration. Principles, tools and methods / Advances in Computer Vision and Pattern Recognition. Springer 2012. 441 p.
5. Tashlinskii A. Computational Expenditure Reduction in Pseudo-Gradient Image Parameter Estimation / Computational Science – ICCS 2003, V. 2658. Proceeding, Part II. Berlin: Springer. 2003. P. 456-462.
6. Tashlinskii A.G. Estimation of parameters of spatial deformations of images' sequences. – Ulyanovsk: izdatelstvo UlGTU. 2000. 132 p. (in Russian).
7. Tashlinskii A.G., Safina G.L., Voronov S.V. Pseudogradient optimization of objective function in estimation of geometric interframe image deformations / Pattern recognition and image analysis. 2012. V. 22. №. 2. P. 386-392.
8. Vasilev K.K., Krasheninnikov V.R. Statistical analysis of multidimensional images. – Ulyanovsk: izdatelstvo UlGTU, 2007. 172 p. (in Russian).
9. Viola P., Wells III W.M. Alignment by maximization of mutual information / International Journal of Computer Vision. 1997. Vol. 24. P. 137–154.

AREA AND CHAIN GROUP OBJECTS DISTINCTION BY ANALYSING OF MINIMAL SPANNING TREE ADGES SEQUENCES[1]

D. V. Urzhumov[2,3], A. V. Krevetsky[2,4]

[2]Volga state university of technology, Russia Mari-El 424000 Yoshkar-Ola Karl Liebknecht street 74 61
[3]danurzhumov@gmail.com, [4]krevetsk@mail.ru

The problem of distinguishing group objects is considered. The algorithm of discrimination, based on the analysis of minimal tree edges sequences, is developed. The resistance to the form of an associated continuous image is proved.

Introduction

The actual problems of the remote sensing data analysis are to recognize, to classify and to estimate parameter for groups of objects with size comparable to the resolution of the scanner.

As objects in the group are similar in size to the resolution of the scanner, images of them have degenerated form. So, the elements of group can be considered as point objects (PO), and the groups itself as a group point object (GrPO).

An efficient method of GrPO distinction is spectral analysis of contours of associated continuous images (ACI) [1]. Under this method contour is assigned a normalized autocorrelation function (NAcF) (1):

$$\nu(d) = \frac{(\Gamma, \Gamma(d))}{\left\|\Gamma\right\|^2} = \frac{\sum_{n=0}^{L-1} \gamma(n)\gamma^*(n+d)}{\sum_{n=0}^{L-1} \left|\gamma(n)\right|^2}, \quad (1)$$

where $\left\|\Gamma\right\|^2$ is a square of norm for ACI contour Γ consists of L elementary vectors (EV) $\gamma(n)$; d is offset of the start point; $\Gamma(d)=\{\gamma(n+d)\}0,L-1$; * – the complex conjugate.

Number chain GrPO may have ACI, with the change of EV direction (Fig. 1).

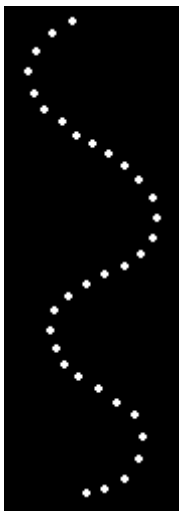


Fig. 1. Example of GrPO with change of EV direction.

It is clear that, GrPO with change of EV direction will have a narrower correlation interval. So, increasing frequency changes of direction may cause GrPO misinterpretation.

If ACI is represented as minimum spanning tree (MST) with the PO of GrPO as vertices, a chain GrPO contains the only sequence includes all vertices of MST [2]. Area GrPO contains several sequences (branches) of vertices.

Problem statement

The purpose of the study is to develop area and chain GrPO distinction algorithm, based on a structure of MST with the PO o GrPO as vertices.

[1] This work is supported by Russian Foundation for Basic Research (grant № 13-01-00427)

Approach and techniques

Define a branch as a sequence of adjacent MST edges, bounded by vertices with degree not equal 2, as shown (Fig. 2).

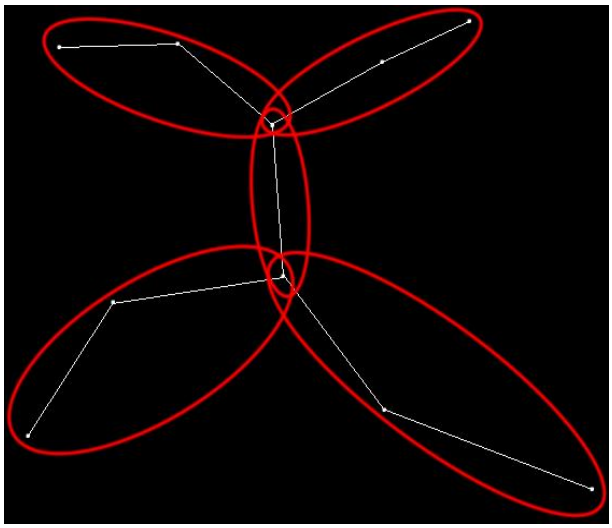


Fig. 2. Segmentation MST into branches.

The vertex is called *border vertex of branch* if its degree not equal 2 and *inner vertex of branch* otherwise.

Each inner vertex store:

- number of branch, it belongs;
- its number in branch;
- total length of edge sequences in the direction of increasing and decreasing numbers of vertices in a branch.

Border vertices may be associated with more than one branch; they store a set of triples associated with branches. Each triple contains:

- number of branch with border vertex at current;
- number for vertex in branch;
- total length of branch.

Branch can also be defined as a sequence of vertices, so it stores numbers of vertices ordered by increasing number in branch.

Segmentation in branches is performed during the construction of MST. Algorithm analyzes both the added vertex and the nearest to it (adjacent) already included to MST. There are three cases:

1. The adjacent vertex is an inner vertex for a branch of MST. In this case a branch is divided into two. The adjacent vertex becomes border vertex for each of them. The edge between added and adjacent vertices forms a new branch. The added and the adjacent vertices become border vertices for this branch.

2. The adjacent vertex is a border vertex for a single branch of MST. In this case added vertex becomes border and adjacent vertex becomes inner for this branch.

3. The adjacent vertex is a border vertex for several branches of MST. In this case the edge between added and adjacent vertices forms a new branch. The added and the adjacent vertices become border vertices for this branch.

Block diagram (Fig. 3) illustrates the algorithm is performed when a new vertex is added to MST.

It is clear that, the ideal chain GrPO may contain a single branch that contains all vertices and edges of GrPO. However real chain GrPO may be distorted by near located false or side PO. The following algorithm may be used to account this probability:

1. for each border vertex with degree greater than 1 associated branches are sorted by decreasing of branch total length.

2. two branches with the greatest length are considered as elements of the chain and are excluded;

3. other associated with the vertex branches are considered side branches, total edge length of them is summed to calculate general total length of side edges;

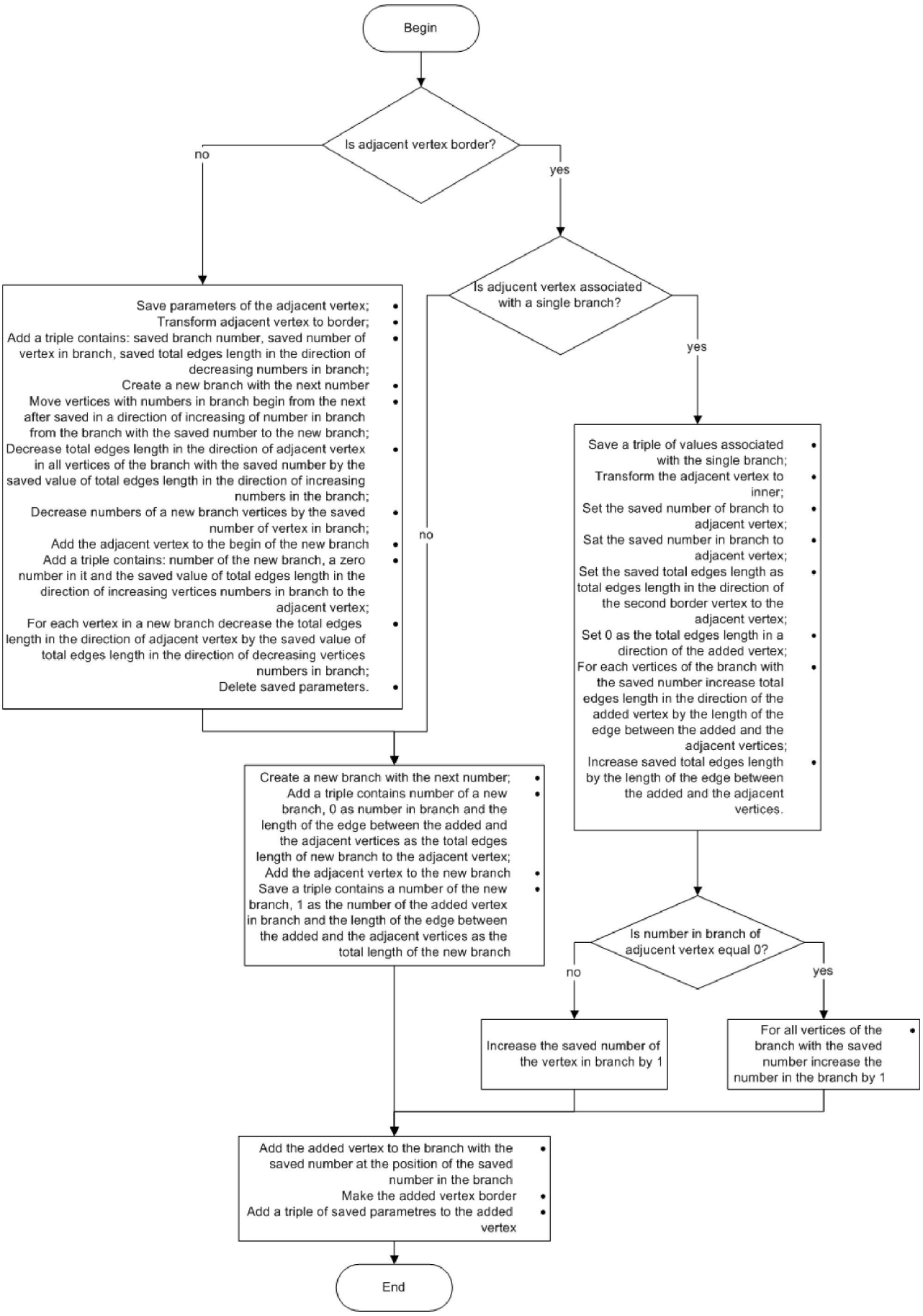


Fig. 3. Block diagram of algorithm for segmentation MST into branches.

4. The ration of general total length for side edges to general total length for all MST edges is compared with the threshold, in excess of which the hypothesis about the chain type of GrPO is rejected and the GrPO is defined as area.

Results

The proposed algorithm analyzes the ration of general total length for side edges to general total length for all MST edges. So it remains invariant to scaling and rotation of ACI. Total length evaluation allows excluding the influence of ACI shape. So, the proposed algorithm is resistant to contour EV direction changes.

The proposed algorithm is not required the construction of ACI contour. It is also integrated in the procedure of MST construction, which allows increasing performance of algorithm by decreasing the number of total enumerations for the set of PO at GrPO.

Conclusion

Therefore the proposed algorithm may be applied for GrPO distinction. Although algorithm works only with ACI in the form of MST, it is more resistant to ACI shape. Decreasing the number of total enumerations for the set of PO at GrPO increases the performance of algorithm and allows it to be executed in real time mode.

References

1. Krevetsky A.V. Mitrofanov V.I., Plekin V.J. Distinction of group point objects by the form of associated continues image // Proceedings of Higher Education. Radionics. – 1997, Vol. 40, №3.- P. 44-52 (in Russian).
2. Introduction to the contour analysis; application to image and signal processing / J.A. Furman, A.V. Krevetsky, A.K. Peredreev and others; edited by J.A. Furman. – 2-nd edition. – M.: PHYSMATHLITE, 2003. – 592 p.

TWICE STOCHASTIC MODELS OF IMAGES

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

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

This paper is devoted to the synthesis of algorithms of double stochastic AR models that can be used to form images that are similar in their properties to real signals. In addition to solving the problem of synthesis the filtering of multidimensional images generated by double stochastic models is considered.

Introduction

At present time the problems of multi-dimensional data processing have become particularly important. It occurs due to widespread use of the methods of various objects images registration, including a multi-spectral (up to 10 spectral bands) and hyperspectral (up to 300 bands) registration of the earth's surface. This survey results in obtaining of multidimensional arrays of information, determined by the spatial coordinates, spectral range and the pickup time. Significant volumes of the information to be processed can potentially improve the quality of processing of satellite images, but require new methods of qualitative and quantitative analysis of aerospace observation in the form of a single multi-dimensional aggregate.

For effective solution of various image processing problems their correct mathematical formulation is needed, which first of all includes the mathematical description, that is the image model as an object of study.

Despite the variety, the known mathematical models of multi-dimensional signals, usually used for solving the problem of estimation of multidimensional images [2, 5, 6, 8-11], have a number of drawbacks. Among these drawbacks considerable difficulties of description of spatially inhomogeneous and time non-stationary real stuff are of particular interest. Therefore, for achievement of a satisfactory solution to the problems of image processing it is very important to choose the appropriate mechanisms to model these observations.

Synthesis of double-stochastic model

It should be noted that double stochastic or mixed models of random fields are known in the literature for a long time. In [6] for image simulation it is proposed to use combinations of different methods of random fields generation. One of such combinations, for example, can be represented as follows:

$$B(x) = \eta(x)B_1(x) + (1-\eta(x))B_2(x), \quad (1)$$

where $B_1(x)$ - a realization of one of the microstructures (e.g. an object on the image), $B_2(x)$ - a realization of another microstructure (e.g. underlying surface), $\eta(x)$ - a realization of the field determining the interaction. In this case the multipliers $\eta(x)$ and $(1-\eta(x))$ perform switching from one realization $B_1(x)$ to another one $B_2(x)$.

In a number of works, such as [3, 4] it is proposed to use the double stochastic models to describe various processes associated with the description of the financial indices of the turbulent flow and network traffic. However, several problems arise when using these results to describe the multi-dimensional random fields. The resulting solutions are either unable to fully describe the characteristics of the variety of real images, or involve significant difficulties in the subsequent analysis and estimation. To overcome these difficulties, we assume that the random field is defined on a multi-dimensional rectangular grid Ω so that its values $x_{\bar{i}} = F(x_{\bar{j}}, \alpha_{\bar{i}}, \xi_{\bar{i}})$, where $\bar{i} \in \Omega$, $\bar{j} \in D$, $F()$ - a certain transformation; $\alpha_{\bar{i}}$ - model parameters, which represent a random field

independent of $\xi_{\bar{i}}$. Then the density of the random variable $x_{\bar{i}}$ can be defined as follows $\omega(x_{\bar{i}}) = \int_D \omega(x_{\bar{i}}|\alpha_{\bar{i}})\omega(\alpha_{\bar{i}})d\alpha_{\bar{i}}$.

The coefficients $\alpha_{\bar{i}}$ characterize the size and shape of the objects on the simulated image, and increase of the average values of these parameters will lead to increase of the average size of the objects on the simulated image. The choice of a method for converting the brightness values into a set of correlation parameters enables to monitor the values of images correlation functions (CF), which makes it possible to simulate random fields with correlation properties similar to real satellite images.

As an example let us consider a realization of the two-dimensional random field $X = \{x, \bar{i} \in \Omega\}$, obtained by using a model with multiple roots of the characteristic equation [9]. In this case, the simulated images are derived from one-dimensional autoregressive equations [9]. For example, for the second order AR process $x_i = 2\rho x_{i-1} + \rho^2 x_{i-2} + \xi_i$ the corresponding two-dimensional eight-point random field model can be obtained:

$$x_{ij} = 2\rho_x x_{i-1j} + 2\rho_y x_{ij-1} - 4\rho_x\rho_y x_{i-1j-1} - \\ - \rho_x^2 x_{i-2j} - \rho_y^2 x_{ij-2} + 2\rho_x^2\rho_y x_{i-2j-1} + \\ + 2\rho_x\rho_y^2 x_{i-1j-2} - \rho_x^2\rho_y^2 x_{i-2j-2} + b\xi_{ij} \quad (2)$$

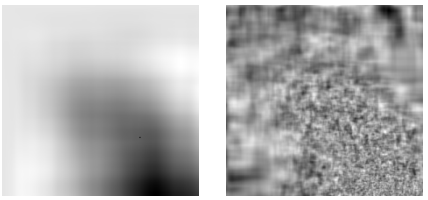
Assume that parameters ρ_x and ρ_y, used in the model, are not constant values, but are realizations of the two basic random fields, the brightness values of one of which will be converted into the set of correlation parameters $\{\rho_{xij}, i = 1,..,M1, j = 1,..,M2\}$, and brightness of the other will be, correspondingly, converted into the set of correlation parameters $\{\rho_{yij}, i = 1,..,M1, j = 1,..,M2\}$ in accordance with the following formulas:

$$\rho_{xij} = r_{1x}\rho_{x(i-1)j} + r_{2x}\rho_{xi(j-1)} - r_{1x}r_{2x}\rho_{x(i-1)(j-1)} + \varsigma_{xij}, \quad (3)$$

$$\rho_{yij} = r_{1y}\rho_{y(i-1)j} + r_{2y}\rho_{yi(j-1)} - r_{1y}r_{2y}\rho_{y(i-1)(j-1)} + \varsigma_{yij}, \quad (4)$$

where $\{\varsigma_{xij}\}$ and $\{\varsigma_{yij}\}$ – two-dimensional random fields of independent Gaussian random variables with zero mean and variances $M\{\varsigma^2_{xij}\} = (1-r_{1x}^2)(1-r_{2x}^2)\sigma^2_{\rho_x}$ and $M\{\varsigma^2_{yij}\} = (1-r_{1y}^2)(1-r_{2y}^2)\sigma^2_{\rho_y}$; $\sigma^2_{\rho_x} = M\{\rho^2_{xij}\}$, $\sigma^2_{\rho_y} = M\{\rho^2_{yij}\}$.

An example of the model image realization with varying correlation properties which characterize relief of the assumed snapshot of the Earth's surface is shown in Fig. 1, where a - basic random field; b – image frame with varying correlation properties.



a) (0.99;0.99) b) (0.9;0.9)

Fig. 1. Simulation of images with varying correlation properties

Filtering of double-stochastic model

Double-stochastic image models enable to give an adequate description of signals in a variety of information retrieval systems. In such systems, as a rule, signals are transmitted via the communication channels containing noise. At that a number of problems of processing of signals on noisy background arises, the most complicated of which is the problem of multidimensional RF filtering [2, 5-11]. To solve the problem of double stochastic RF filtering we first consider the following one-dimensional version of the field. Suppose we have a random sequence $\{x_i\}$, formed according to the formula $x_i = \rho_{i-1}x_{i-1} + \xi_i$, where $\rho_i = r\rho_{i-1} + \zeta_i$, r - correlation coefficient of the base process, ξ_i and ζ_i - normal random variable with zero mean and variances $V_{\bar{\xi}_i}$ and $V_{\bar{\zeta}_i}$ respectively. Suppose that observation $\{z_i\}$ is an additive mixture of sequence $\{x_i\}$ and white noise $\{n_i\}$ with variance V_n: $z_i = x_i + n_i$. Then the observation model can be written as follows: $z_i = F(\bar{x}_i) + n_i$; $\bar{x}_i = \varphi(\bar{x}_{i-1}) + \bar{\xi}_i$. For filtering of such a sequence we use non-linear Kalman filter [5]. In this case, we obtain the following filtering equation:

$$\begin{pmatrix} \hat{x}_i \\ \hat{\rho}_i \end{pmatrix} = \hat{\bar{x}}_{\text{э}i} + P_i \begin{pmatrix} 1 \\ 0 \end{pmatrix} V_n^{-1}\left(z_i - \hat{\rho}_{i-1}\hat{x}_{i-1}\right), \quad (5)$$

where $\hat{\bar{x}}_{эi} = \begin{pmatrix} \hat{\rho}_{i-1}\hat{x}_{i-1} \\ r\hat{\rho}_{i-1} \end{pmatrix}$,

$$P_i = P_{эi}\left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} V_n^{-1} \begin{pmatrix} 1 & 0 \end{pmatrix} P_{эi}\right)^{-1},$$

$$P_{эi} = \begin{pmatrix} \hat{\rho}_{i-1} & 0 \\ \hat{x}_{i-1} & r \end{pmatrix} P_{i-1} \begin{pmatrix} \hat{\rho}_{i-1} & \hat{x}_{i-1} \\ 0 & r \end{pmatrix} + \begin{pmatrix} \sigma_\xi^2 & 0 \\ 0 & \sigma_\zeta^2 \end{pmatrix}.$$

To filter a flat image, we can make use of the vector (row-wise) non-linear Kalman filter [9]. For this purpose we combine the image row elements into the vector $\bar{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iN})^T$. Then the image model (1) can be written as follows:

$$\bar{x}_i = \rho_{x(i-1)}\bar{x}_{i-1} + \upsilon(\bar{x}_{\rho(i-1)})\bar{\xi}_i, \qquad (6)$$

where $\bar{x}^T{}_{\rho i} = (\bar{x}^T{}_i \bar{\rho}^T{}_{xi} \bar{\rho}^T{}_{yi})$.

And the process of row-wise estimation can be written as follows:

$$\hat{\bar{x}}_{pi} = \hat{\bar{x}}_{эpi} + P_i \frac{\partial F^T}{\partial \bar{x}_{pi}} V_n^{-1}(\bar{z}_i - \hat{\bar{x}}_{эpi}) \qquad (7)$$

Let us compare the resulting algorithm with the Wiener filter [5, 8]. Despite the fact that the use of the latter generally requires a stationarity condition for the field $\{x_{ij}\}$ with very slow change in the basic fields $\{\rho_{xij}\}$ and $\{\rho_{yij}\}$ the CF of the mixed model can be very close to the CF of the Habibi model [1]:

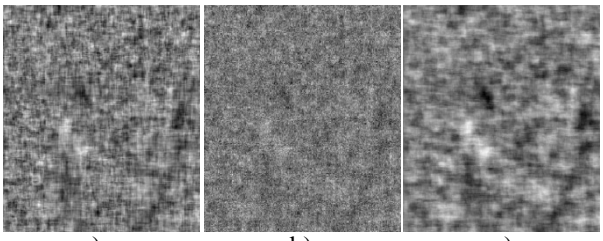
$$B_{ij} = \sigma_x^2 \rho_x^{|i|} \rho_y^{|j|} = \sigma_x^2 e^{\ln \rho_x |i|} e^{\ln \rho_y |j|} = \sigma_x^2 e^{-a|i|} e^{-b|j|} \quad (8)$$

Under the specified conditions this makes it possible to estimate $\{x_{ij}\}$ on the basis of the Wiener filter:

$$\hat{x}_{ij} = \sum_{q=1}^{M1} \sum_{u=1}^{M2} h_{qu} z_{i-q,j-u}, \qquad (8)$$

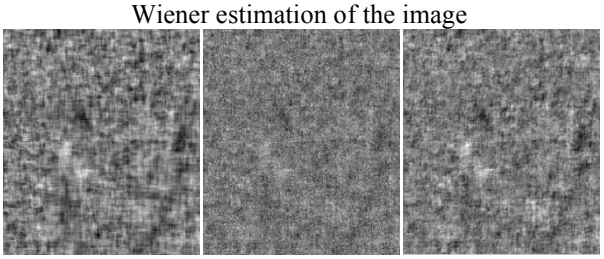
where h_{qu} - weight function which can be expressed through known relationships between CF, spectrum density and frequency response [8].

The results of the two filtering algorithms operation are shown in Fig. 2 and Fig. 3: a - the original image; b - noisy image; c – estimate of the image by means of the filter. The parameters of the image to be simulated: image size 240×240; $\sigma_x^2 = 1$; $m_{\rho_x} = 0.78$; $m_{\rho_y} = 0.78$; $\sigma_{\rho_x}^2 = 0.003$; $\sigma_{\rho_y}^2 = 0.003$; $r_{1x} = r_{2x} = r_{1y} = r_{2y} = 0.99$; $\sigma_n^2 = 3$.



a) b) c)

Fig. 2. a - image without noise; b - image with noise; c- Wiener estimation of the image



a) b) c)

Fig. 3. a - image without noise; b - image with noise; c- Kalman estimation of the image

The results of filtering error dependence versus noise variance are shown on the graphs in Fig. 4.

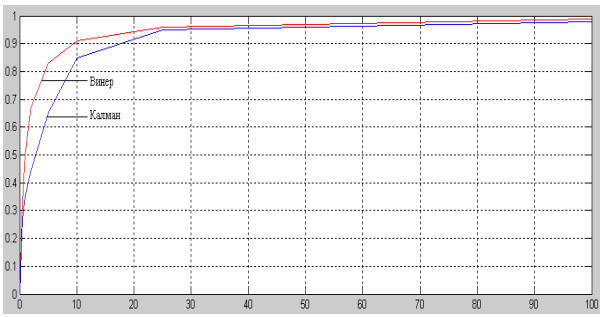


Fig. 4. The dependence of the filtering error variance versus noise variance

Filtering errors dependencies analysis at various parameters of the original images shows that for non-uniform images the Kalman filter has an advantage due to varying properties of the random field. This gain for special cases can attain 70% in the sense of the estimation variance.

Conclusion

Thus, in this paper a technique of imaging with varying properties is given and also synthesis of image recursive filtering algorithm is performed. The obtained results will enable in the future to develop mechanisms of forming of quasi inhomogeneous multidimensional images similar to the actual in the sense of their correlation properties.

References

1. Andriyanov N., Vasiliev K., Dementiev V. Performance analysis of two-dimensional double-stochastic field / / Proceedings of the International Conference "Radio-electronic devices and systems for information and communication technologies" (RES 2013), 2013. - P. 294-297 (in Russian)
2. Dudgeon D., Mersero R. Digital processing of multidimensional signals. – M.: World, 1988, 488 p. (in Russian)
3. Gorcev A., Zuevich V. Optimal estimation of the conditions of asynchronous double stochastic flow of events with an arbitrary number of conditions / / Bulletin of the Tomsk State University. - № 3 (12), 2010. - P. 41-53 (in Russian)
4. Korolev V. Probability and statistical analysis of random processes using mixed Gaussian models, ISBN :978-5-211-05863-7 Moscow: Moscow State Technical University, 2011, 510 p. (in Russian)
5. Sage A. Estimation theory and its application in communication and control / A. Sage, J. Melsa. — M. : Svyaz, 1976, 495 p. (in Russian)
6. Shalygin A., Palagin Y. Applied statistical modeling techniques, L.: Mechanical Engineering. Leningrad. Dep., 1986. – 320 p. (in Russian)
7. Vasiliev K. Presentation and fast processing of multi-dimensional images. / K. Vasiliev, V. Krasheninnikov, I. Sinitsyn V. Sinitsyn // Naukoemkie Technologii. - № 3. 2002. p. 4-24. (in Russian)
8. Vasiliev K., Krasheninnikov V. Filtering methods of multidimensional random fields. - Saratov: Saratov State University, 1990. - 128 p. (in Russian)
9. Vasiliev K., Krasheninnikov V. Statistical analysis of multidimensional images / K. Vasiliev, V. Krasheninnikov. - Ulyanovsk Ulyanovsk State Technical University, 2007. - 170 p. (in Russian)
10. Vasiliev K., Luchkov N. The filtering efficiency of random fields on multidimensional grids: Proceedings of the 63 scientific session devoted to Radio Day. - 2008. - P. 378-380. (in Russian)
11. Woods J.W. Two-dimensional Kalman filtering //Topics in Applied Physics, Berlin, 1981, v.42, pp.155-208.

STRUCTURAL ANALYSIS OF RASTER IMAGES[1, 2]

Yu.G. Vasin[3], V.P. Gromov[4]

[3] Research Institute of Applied Mathematics and Cybernetics
Lobachevsky Nizhni Novgorod State University
National Research University
603005, Russia, Nizhni Novgorod, Ulyanov St.10, UNN RIAMC.
Tel. (831)4362361, [4]pmk@unn.ac.ru

In this paper, we present the results of research into and practical use of methods and algorithms for purposeful splitting of raster data in relation to the tasks of semantic interpretation of document-type graphical data (schemes, plans, drawings, topographic maps, nautical charts, survey sheets, etc.), as well as images of Earth remote sensing (ERS).

Introduction

Significant amounts of graphic information (GI) used in communication systems determine the relevance of the rapid and reliable recovery of useful information, which is necessary to make pragmatic decisions, from this stream. Given the fact that the base non-derivative (primitive) information object of GI is a raster image without regard to its substantive content, it is important to design and develop methods and algorithms for fast and reliable assessment of its substantive content in terms of symbols. The solution of these problems is directly related to the effective scheme of splitting the original raster image into structural elements that are suitable for further analysis and synthesis of the document's symbol content using methods and algorithms of pattern recognition. Our paper offers an approach to the solution of this problem.

Statement of the problem

For graphic document data, the raster image will be considered to be an electronic copy of the graphic document with a standardized (in terms of color and geometry) symbol content. We will also assume that for the raster image, as an integral data object, essentially a spatially organized set of primitive pixel elements, the following operations have already been performed:

a) referencing to the real-world coordinate system;
b) elimination of pixel distortions;
c) restoring the color palette of the original image document.

Algorithms for solving these problems are well known, therefore they are not considered in this paper.

Thus, the problem of data structural analysis can be formulated as follows.

It is necessary to split the original raster image (bitmap) into such a hierarchical sequence of structural elements for which a high-performance scheme of their algorithmic design will be possible and, subsequently, a simple and reliable scheme for object interpretation. Below, we give the schemes for raster image splitting and the algorithmic content for their implementation, both for graphic images [1-3] and for Earth remote sensing images [4].

Models and algorithms for structural analysis of raster graphic images

The original set of pixels of the graphic image raster are broken down successively into the following subsets: the foreground pixels (symbols - "dark"), the background pixels (background - "light"). Then, the resulting subset is additionally split into 3 classes: discrete (those associated with the symbols of

1 The reported study was partially supported by RFBR, research project № 13-07-00521_a
2 The reported study was partially supported by RFBR, research project № 13-07-12211

small dimensions); linear (those belonging to line-type symbols); area (background).
Next, for each selected subset of pixels, the following hierarchical structural model is proposed:
- a pixel - an initial primitive element;
- a bar - a linear cluster of pixels;
- a simple raster object (SRO): a cluster of bars of the same colour with a single connection;
- a blob – a colour spot: a connected set of SROs of the same colour (strokes, pixels);
- symbol - a connected set of (in the general case) coloured SROs, strokes, and pixels;
- raster image – an image of a disjoint sets of plotted blobs (symbols).

Algorithm 1. For splitting pixels into background and symbol pixels, we used a simple algorithm based on a threshold scheme. A pixel was classified as a background pixel if its brightness was less than a predetermined threshold, otherwise the pixel should be assigned to the class of symbol pixels. The threshold value of brightness was chosen by the parameters of the normative color palette of the document.
Algorithm 2. To classify the pixels as belonging to a discrete, line or area raster objects, the following calculation scheme was used. A cluster of uniformly coloured pixels was selected from the unclassified pixels. If the dimensions of the cluster were less than those specified, then all the pixels of this spot were classified as belonging to discrete objects. Otherwise, for each pixel in the selected colour spot the minimum length of the segment of the bar section was calculated and if the length of this segment was less than some specified value, this pixel was classified as belonging to line objects, otherwise it was referred to the class of pixels of area objects.
The process of selection of the structural elements of the proposed raster model of the BAR, SRO or BLOB type involved the standard procedures of cluster analysis based on color similarity and spatial coherence of pixels. Application of the bar model of multicolour raster image can significantly (10 to 20 times) improve the performance of algorithms, not only during construction of the structural model of the raster image, but also at the stages of forming the geometrical (vector) model of the image and recognition of the symbol content of the graphic document.
At the geometric (vector) level of raster description, the following sequence of raster vector models is proposed. At the lowest level, the whole aggregate of blobs built at the pixel level is uniquely described by the contour model of the raster image as a set of disjoint closed non-intersecting planar curves, which are essentially a structured vector description of the boundaries of colored spots. At the second (higher) level, a segment-contour model was used for the vector description of the raster image. In this model, the aggregates of pixel structural elements belonging to discrete and area symbols are presented as a contour model, while pixel formations belonging to line symbols are represented as an aggregate of segments, i.e. as an aggregate of curves that describe the behavior of axial lines of line raster objects at the vector level. This model is not unequivocal and admits the reverse transition (rastering) of line raster objects of the graphic document being processed with the accuracy up to one pixel. Besides, it should be noted that the structural elements of these models: contours and segments are "coloured", i.e. they contain a colour characteristic in their description. This is a significant factor facilitating the automatic recognition of graphic symbols in the document.

Models, methods and algorithms for the structural analysis of raster images of Earth remote sensing

The data flows generated by the methods of Earth remote sensing (ERS) are formatted as a multispectral set of images for a given terrain area. Numerical zoning and mosaic coverage methods are used to provide positional alignment of images of different spectral components.
For the data thus formatted, in this paper ERS images will be treated as a matrix of vector elements

$$P[Pij]\ (i=1,2,3,..\ N;\ j=1,2,3,..,M), \qquad (1)$$

where:
Pij - vector object with components Pij={I1,I2,I3,..Ik};
M,N - pixel dimensions of the image field;

Ik - brightness intensity of the k-th spectral component;
k - dimension of the vector pixel element (actually it is the number of simultaneously registered spectral signals in different ranges of scanning radiation). Besides, it is assumed that for the terrain area of interest its object content has been formed using the methods of GIS applications.

$$O\{Op\},\ p=1,2,3,\ldots,L \qquad (2)$$

For the chosen model of remote sensing data, that are positionally related to a particular area of the Earth's surface, we studied the problem of focused structural analysis of the vector matrix (1).

The central task of the structural analysis is to split the original set of filling elements of the matrix (1) into such a subset of pixel clusters

$$K=\{Ks\},\ s=1,2.3,..L, \qquad (3)$$

that:
1) each pixel of the matrix (1) is always included in one and only one of the clusters (the condition of absence of intersection of clusters Ki * Kj =0, i=j, i,j=1,2,3,…,M);
2) this requires that each cluster Ks of (3) should be a positionally related component of the pixels having similar spectral composition.
To solve the problems of structural analysis of multi-channel Earth remote sensing data, the mathematical model was chosen of the source data in the form of N objects (pixels) set in the k+2 dimensional space of the combined aggregate of spectral and spatial features. In this case, the source data for solving the problems of structural analysis are in the form of a set of vector objects:

$$P=\{Pm\}\ m=1,2,..\ N\ ;\ N=\text{count X*count Y} \qquad (4)$$

Pm - original vector element of the data;
N - number of elements (pixels);
count X, count Y - linear pixel dimensions of the image (raster).
We have considered the following scheme for splitting the raster image by means of cluster analysis methods. Initially, clusters of pixels were selected successively only in the space of spectral features. The spectral similarity of 2 pixels Pi, Pj was measured by the metric of proximity:

a) with the account of absolute values of brightness of the spectral components

$$R(P_{i,}\ P_j) = \sum_{k=1}^{n}(P_k^i - P_k^J) \qquad (5)$$

b) in the invariant form with respect to brightness properties of spectral channels as the spectral correlation coefficient of 2 spatially distinct objects (pixels):

$$Kij= Cov(Pi,Pj)/(Cov(Pi\ Pi)*Cov\ (Pj\ Pj)) \qquad (6)$$

The procedure of clustering in the space of spectral features consisted in the following: a k-dimensional distribution histogram of vector pixel of matrix (1) in the space of k spectral components was constructed. For each local peak of the histogram, another cluster is formed of pixels that have not yet been clustered. Parameters of the constructed cluster are calculated as the average values of the spectral components of the pixels that fill it. At the final stage, spatial mapping is performed of the constructed spectral clusters - "spots." At this point, the process of structural description of the vector raster image (1) in the form of a model of spectral spots (areas of interest)

$$G= \{Gt\}\ ,\ t=1,2,3,..,S$$

is concluded and the data are prepared for solving the task of classification.

Classification of elements of an ERS raster image is a essentially a preliminary decoding of the semantic content of the image and consists in assigning each pixel of the image to one of the specified classes. In order to optimize the computational complexity of this task, its solution only involved the above-constructed aggregate of spectral spots G (areas of interest).
The procedure of classification was as follows:
1) a raster pattern was chosen for each class of objects;
2) for each sample, a recursively nested hierarchical system of spectral features Q,V was constructed:
$Q,V= \{Q_r, V_r\}$, r=1,2,..,u ;
where u is the number of levels of the feature tree;

Q_r, - is the average brightness in the raster window of the size $2^{(r-1)}$ x $2^{(r-1)}$;

V_r - is the average brightness variability in the raster window of the size $2^{(r-1)}$ x $2^{(r-1)}$.

Q_r and V_r are calculated recursively by the formula:

$$Q_0 = q; \quad V_0 = 0;$$

$$Q_{r+1} = \frac{1}{4}\left(\sum_k Q_{r,k}\right); \qquad (7)$$

$$V_{r+1} = \max_k Q_{r.k} - \min_k Q_{r,k}; \quad k = (1,2,3,4)$$

The pixel matrix of the areas of interest was scanned with the raster window of the standard pattern for a given class of objects. In the process of scanning, the current values of the spectral features were first calculated at each point by the formula (7), and then with the use of these values the similarity was assessed of the standard raster window and the current pixel window. If the measure of similarity was sufficient, then the pixels of the current window were assigned to the class being recognized. After applying this procedure for all of the standard patterns, the task of classification of image pixels was considered accomplished.

Conclusion

The proposed models and the structural analysis algorithms are implemented in the form of a dedicated software. Their experimental testing as components of the technologies for the processing of raster topographic maps and nautical charts, raster images of Earth remote sensing, as well as photographic images in the tasks of 3D-modeling of buildings and structures has confirmed their effectiveness.

References

1. Vasin Yu.G. Gromov V.P. Colour updating of graphic documents. –Pattern recognition and image analisis: new informationtechnologies -9 International Conference 2008
2. Vasin Yu.G. Bashkirov O.A., Rudometova S.B. Mathematical model for structured description of graphic images. - Automation of processing of complex graphic information: Inter-university collection. Gorky: GSU, 1984.
3. Vasin Yu.G, Vasin D.Yu., Gromov V.P..Structural description of raster data. // The 6th All-Russian Conference "Methods and tools for processing complex graphic information" / Nizhni Novgorod, UNN Press, 2001. p. 23-29.
4. Vasin Yu.G. Constructive approach in the problems of processing and analysis of signals and images-8th Open German-Russian Workshop “Pattern recognition and image understanding” (OGRW-8-2011:Workshop Proceedings. 2011.-pp.315-319.

AUTOMATION METHODS FOR TECHNOLOGIES TO PRODUCE DIGITAL GRAPHIC DOCUMENTS WITH WEAKLY FORMALIZED DESCRIPTION OF OBJECTS [1, 2]

Yu.G. Vasin[3,4], L.I. Lebedev [3,5]

[3] Research Institute of Applied Mathematics and Cybernetics
Lobachevsky Nizhni Novgorod State University
National Research University
603005, Russia, Nizhni Novgorod, Ulyanov St.10, UNN RIAMC.
[4]vasin@focus.ac.su, [5]lebedev@pmk.unn.ru

The paper describes the solution of the problem of object identification in graphic images based on recognition with self-learning to improve the efficiency of technologies for creating digital graphic documents. The results are implemented in the experimental technology of digital databases of graphic documents.

Introduction

In most cases, graphic documents contain objects whose description generally complies with the requirements. Some examples of such graphic documents are books, maps and plans, nautical charts and other types of documents. However, in many cases the rules of nomenclature description of objects in graphic documents are not observed. Among such documents are some nautical navigation plans, and in most cases, hydrographic survey sheets. As a rule, the resulting contour model for describing objects in the image does not match the standard format for such graphic documents and has specific features both in terms of the composition of the objects, and ways for representing them. Thus, the task of providing automated entry of such documents becomes difficult due to the low level of recognition of objects in the image. Problems of this kind are also encountered in the recognition of objects in the images of Earth remote sensing.

Statement of the problem

A peculiar feature of the above-mentioned graphic documents is their stylized representation form, despite all possible deviations from regulatory requirements to object representation. This means that all the objects that are similar in shape can be attributed to the same equivalence class. Therefore, to solve the problem of accurate recognition it is necessary to replace the basic set of standard patterns with a set of standard patterns generated by the objects of this particular document. The task of recognition will be solved in full, if each equivalence class is further assigned an appropriate identifier.

Methods for solving the problem

It is proposed to carry out object recognition based on the calculation of similarity estimates using the correlation and extreme contour method (CECM) [1,2]. CECM is invariant with respect to orthogonal transformations and scaling, which ensures the most complete composition of objects in the equivalence class. CECM is used in a self-learning mode with automatic replenishment of standard patterns based on descriptions of unidentified objects [3]. Thus, as a result of using CECM in the self-learning mode, two files will be generated at the output: the file of standard patterns and the file for the parametric descriptions of recognized objects, this latter file being a set of fixed-length entries. Each entry includes the data necessary to recover the object, including a link to the standard pattern whose assessment of similarity has the

1 The reported study was partially supported by RFBR, research project № 13-07-00521_a
2 The reported study was partially supported by RFBR, research project № 13-07-12211

greatest value. At this point, the first part of the recognition procedure, consisting of breaking down all the objects into equivalence classes, is completed. Next, each equivalence class should be assigned a corresponding identifier.

According to the proposed technology, the assignment of identifiers is performed upon visualization and examination of the standard patterns. Here, three possible outcomes of identification of the standard patterns are possible. In the first case, the standard pattern is some class of equivalence and it is necessary to assign it a name appropriate to the image. In Figure 1a, the standard pattern with the image of the number "3" is assigned the code of this character. In Figure 1b, a noise object corresponds to the standard pattern image, and there is no sense to interpret it in terms of symbol- sign coding.

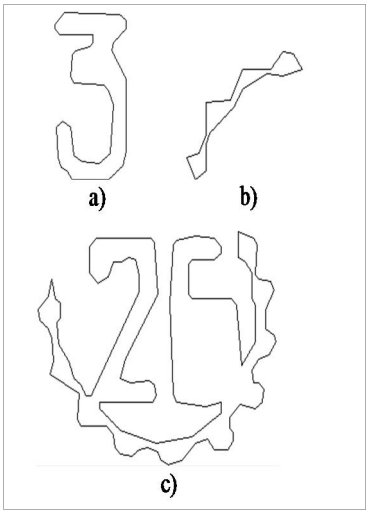


Fig.1. Identification of standard patterns

As a rule, noise objects are parts of contours that have fallen apart. The standard pattern in Figure 1c clearly bears some information load, the loss of which is undesirable when entering the document. Therefore, the input technology in such situations provides for two options for further actions of the user. In the first case, with the use of this standard pattern, all the objects are found of the corresponding equivalence class, and they are entered in the interactive mode. One disadvantage of this solution is the need to enter in the interactive mode not only the attribute data, but also the metric description of the elements of the document content. In the case when an element of the contents has a metric description, it is preferable to use the option of correcting the original image of the sign. Then, after implementing all the necessary corrections of the signs, a new contour representation of the image is formed with a different set of objects. This will enable the automatic assembly mode of such elements of the document content.

To implement the automatic assembly mode, certain requirements should be met with respect to the specification of the metric description of standard patterns. When forming inscriptions, the orientation of the standard patterns for their identification must be horizontal. Also, for reliable detection and identification of subscript and superscript characters, one must specify the anchor point and the form of standard patterns.

Results

The potential of the proposed technology to automate the input of the elements of the content of graphic documents can be demonstrated by using the example of the digital equivalent of the hydrographic survey sheet PGS2 [2]. Vector description of the image of this sheet consisted of 35905 contours. Figure 1 shows that the font style of the digits on the image does not match the font D-431, which must be used in the survey sheet. Moreover, such a style of characters does not match any one of the fonts used in the production of nautical charts. In the course of recognition of PGS2 image objects in the self-learning mode, 335 standard patterns have been obtained, which is more than an order of magnitude less than the number of objects. The examination of the standard patterns showed that the first group shown in Figure 1a should include 33 patterns, the second group in 1b -189 patterns, and the third group in 1c - 113 standard patterns.

In this document, labeling the standard patterns for the first group is a fast procedure, but the main part of time consumption is associated with the analysis of the third group 1c. For 64 standard patterns of this group, only the attributive characteristics need to be entered, while for 49 standard patterns it is desirable to implement corrections in the original description of the corresponding fragments of the image in order to enable the subsequent use of the metric of the

newly obtained contours to automatically generate the content elements of the document PGS2. It should be also noted that the style of the characters in the document will be fully reflected in the description of the standard patterns, which will naturally increase the quality of recognition. This, in turn, will reduce the time required for interactive input after the operation of automatic procedures when producing digital maps, survey sheets and other graphic documents.

Another problem that can be solved using CECM recognition in the self-learning mode is to detect certain kinds of objects or to identify areas of interest. In this case, one will also have to look through the standard patterns file and to specify the objects of interest to the user. Figure 2 shows a fragment of the satellite image of a sea area.

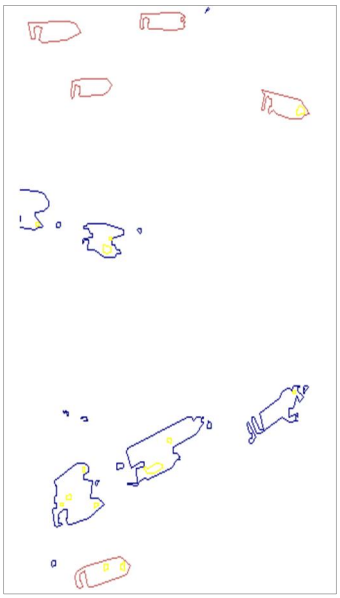


Fig.2. Detection of objects

When reviewing the resulting file of standard patterns, the contour corresponding to the vessel of a particular shape was specified. In this fragment of the image, 5 vessels are marked that have the shape similar to that of the standard pattern.

Conclusion

Improving the efficiency of automated procedures for entering graphic documents is largely determined by the recognition quality of the image objects. As follows from the above, the recognition quality depends on the standard patterns that are formed automatically in the self-learning mode. As a result, a significant reduction in time consumption has been achieved when digitizing nautical charts and survey sheets.

References

1. Vasin Yu.G., Lebedev L.I., Puchkova O.V.Contour correlation - extremal methods of detection and alignment of video information objects /Automated processing of complex graphic information: Inter-university collection of papers / ed. Yu.G. Vasin. – Gorky State University, Gorky, 1987. p.97-112.
2. Vasin Yu.G., Lebedev L.I. "The problem of finding harmonized descriptions in correlation - extremal contour methods of pattern recognition," Reports of the 15th All-Russian Conference "Mathematical Methods of Pattern Recognition (MMPR-15)", Moscow, 2011
3. Vasin Yu.G., Lebedev L.I. An effective format for representing graphic information // Pattern Recognition and Image Analysis, 2012, Vol. 22, No. 2, pp. 393–398. © Pleiades Publishing, Ltd., 2012.

COMPARISON OF 2D IMAGE SHAPE SIMILARITY MEASURES[1]

Yu. V. Vizilter[2,3], A. Yu. Rubis[2,4]

[2] State Research Institute of Aviation Systems (GosNIIAS),
125319, 7, Viktorenko str., Moscow, Russia
[3] viz@gosniias.ru,[4] arcelt@mail.ru

The problem of image correspondence measure selection for image comparison and matching is addressed. Most popular technique for image shape comparison utilizes the mutual information measure based on probabilistic reasoning and information theory background. Another well-known approach for "intensity-to-geometry" shape matching was proposed by Pytiev. In our previous work this Pytiev morphological approach was generalized for obtaining the pure "geometry-to-geometry" shape matching technique. In our other previous work the transform distance for geometrical difference evaluation of shapes named Geometrical Difference Index (GDI) was proposed and normalized similarity measure of image shapes based on GDI was defined. This paper contains some new experimental results and represents comparison of aforementioned similarity measures in multispectral matching application.

Introduction

This paper addresses the problem of image matching "just by shape" with no dependence on the concrete intensity or radiometric pixel values. For example, one can compare images of one scene captured at different seasons, different time of day, in different weather and lighting conditions, in different spectral ranges and so on.

The most popular technique for such image shape comparison utilizes the mutual information measure based on probabilistic reasoning and information theory background [1]. The other approach was proposed by Pytiev (so called "Pytiev morphology") based on geometrical and algebraic reasoning [2].

Mutual information approach is based on probabilistic reasoning. It provides the robust tool for matching of images with different intensities based on their joint 2D histograms. But these histograms cannot explain the geometrical idea of image "shape" in some evident form. Such mathematical "shape" formalism is given in evident form in the so-called "morphological" approach for image comparison proposed by Pytiev.

In the framework of this approach images are considered as piecewise-constant 2D functions

$$f(x,y)=\sum_{i=1}^{n} f_i \chi_{F_i}(x,y),$$

where n – number of non-intersected connected regions of *tessellation* $\mathbf{F}$ of the frame Ω, $\mathbf{F}=\{F_1,\ldots,F_n\}$; $\mathbf{f}=(f_1,\ldots,f_n)$ – corresponding vector of real-valued region intensities; $\chi_{Fi}(x,y)\in\{0,1\}$ – characteristic (support) function of i-th region:

$$\chi_{F_i}(x,y)=\begin{cases}1, & if\ (x,y)\in F_i;\\ 0, & otherwise.\end{cases}$$

Set of images with the same tessellation $\mathbf{F}$ is a convex and close subspace $F\subseteq L^2(\Omega)$ called *shape-tessellation* or simply *shape*:

$$\mathbf{F}=\left\{f(x,y)=\sum_{i=1}^{n} f_i \chi_{F_i}(x,y),\ \mathbf{f}=\{f_1,\ldots,f_n\},\mathbf{f}\in R^n\right\}$$

For any image $g(x,y)\in L^2(\Omega)$ the *projection onto the shape F* is determined as

$$g_F(x,y)=P_F g(x,y)=\sum_{i=1}^{n} g_{F_i}\chi_{F_i}(x,y),$$

$$g_{F_i}=\left(\chi_{F_i},g\right)\Big/\left\|\chi_{F_i}\right\|^2,\quad i=1,\ldots,n.$$

[1] Supported by RFBR grants 12-07-31186-mol_a, 13-08-01071-a.

Pytiev morphological comparison of images f(x,y) and g(x,y) is performed using the normalized morphological correlation coefficients of the following form

$$K_M(g,F)=\frac{\| P_F g \|}{\| g \|},\ K_M(f,G)=\frac{\| P_G f \|}{\| f \|}.$$

The first formula estimates the closeness of image *g* to the "shape" of image *f*. Second formula measures the closeness of image *f* to the "shape" of image *f*. In general $K_M(g,F)$ is not equal to $K_M(f,G)$. So, this morphological image matching score is *asymmetric* in contrary to *symmetric* linear correlation coefficient $K_N(f,g)$.

In our previous work [3] this Pytiev morphological approach was generalized for obtaining the pure "geometry-to-geometry" shape matching technique. In [4] the transform distance for geometrical difference evaluation of shapes named Geometrical Difference Index (GDI) was proposed and normalized similarity measure of image shapes based on GDI was defined. This paper contains some new experimental results and comparison of aforementioned similarity measures in multispectral image matching application.

Geometrical shape comparison

In [3] the geometrical shape comparison approach was developed based on Pytiev's morphological image analysis.

Let $f(x,y)$ form *F* is a piecewise-constant 2D function described above (in section 2). Image $g(x,y)$ from *G* is an analogous 2D function with *m* as a number of tessellation regions $\mathbf{G}=\{G_1,\ldots,G_m\}$; $\mathbf{g}=(g_1,\ldots,g_m)$ – vector of intensity values; $\chi_{Gj}(x,y)\in\{0,1\}$ – support function of *j*-th region. Let's introduce following additional set of "*S*-variables":

S – area of the whole frame Ω;

$S_i=\left\|\chi_{Fi}(x,y)\right\|^2$ – area of tessellation region F_i;

$S_j=\left\|\chi_{Gj}(x,y)\right\|^2$ – area of tessellation region G_j;

$S_{ij}=\left(\chi_{Fi}(x,y),\ \chi_{Gj}(x,y)\right)$ – area of intersection $F_i\cap G_j$.

Mean square effective morphological correlation coefficient (MSEMCC) for shapes F and G was determined as

$$K_M^2(F,G)=\sum_{j=1}^{m}\sum_{i=1}^{m}\frac{S_{ij}}{S}\frac{S_{ij}}{S_j}= \\ =\sum_{j=1}^{m}\sum_{i=1}^{m}K_\Omega(F_i,G_j)\,K_M^2(G_j,F_i), \qquad (1)$$

where $K\Omega(F_i,G_j)=S_{ij}/S$ – normalized *influence coefficient* for pair of regions F_i and G_j;

$K_M^2(G_j,F_i)=S_{ij}/S_j$ – square of *normalized morphological correlation* for pair of regions.

In [4] the special transform distance for geometrical difference evaluation of shapes named *Geometrical Difference Index* (*GDI*) was proposed. In addition, some normalized similarity measures of image shapes based on GDI were represented.

GDI metrics for comparison of two image shapes *F* and *G* was defined as:

$$d_H(F,G)=\sum_{j=1}^{m}\sum_{i=1}^{n}p_{ij}d_H(G_j,F_i)$$
$$d_H(G_j,F_i)=p_i+p_j-2p_{ij}$$
$$p_{ij}=S_{ij}/S \qquad p_i=S_i/S \qquad p_j=S_j/S$$

Metrics properties of GDI were proved in [4].

Most interesting GDI-based similarity measure proposed in [4] is a *Centered Metrical Similarity Coefficient* (*CMSC*). It is based on comparison with assumption about *F* and *G* shapes independence. In this case:

$$d_{HInd}(F,G)=\sum_{j=1}^{m}\sum_{i=1}^{n}p_ip_j(p_i+p_j-2p_ip_j).$$

So, CMSC was defined as:

$$K_{CHS}(F,G)=\frac{\left|d_{HInd}(F,G)-d_H(F,G)\right|}{d_{HInd}(F,G)} \qquad (2)$$

It has the following properties:

1. $K_{CHS}(F,G)\in[0,1]$;
2. $K_{CHS}(F,G)=1 \Leftrightarrow F=G$;
3. $K_{CHS}(F,G)=K_{CHS}(G,F)$;
4. $\forall i,j: p_{ij}=p_i\,p_j \Rightarrow K_{CHS}(F,G)=0$.

Experimental results

For comparison of geometrical correlation techniques and corresponding similarity

measures, mutual information criterion and Pytiev's morphological coefficient a number of experiments is performed over a set of real images including remote sensing and multispectral images. The scheme of experimental comparison was slightly modified relative to [3]. Some clear fragments of TV ("etalon" fragments) image were compared with all equal-sized fragments of corresponding IR image without segmentation. TV fragments were segmented with n =4 levels using least square optimal segmentation procedure. For segmented image fragments f (TV) and non-segmented g (IR) at each fragment g position (x,y) following similarity measures were calculated:

- mutual information $MI(F,G)$;
- square of centered Pytiev morphological coefficient $K_P = K_M^2(f - f_0, G)$;
- square of MSEMCC $Km = K_M^2(F,G)$ (1);
- CMSC $K_{CHS}(F,G)$ (4);

The quality of these measures was estimated by following statistics of 2D correlation function $C(x,y)$ – SNR (*signal-to-noise ratio*) and E (*exceeding* of first maximum to second one):

$$SNR = \frac{|C_1 - \mu|}{\sigma}; E = \frac{|C_1 - \mu|}{|C_2 - \mu|},$$

where C_1 – global maximum of correlation value; C_2 – second global maximum of correlation value out of some small neighborhood of first global maximum; μ – mean value of correlation function; σ – dispersion of correlation function.

Figure 1 demonstrates the example (Example 1) of experimental clear TV data and IR data corrupted by Gaussian noise. Figure 2 demonstrates 2D-graphs (correlation fields) of different correlation measures for matching of TV etalon and IR image (for Example 1). Corresponding numerical data are listed in Table 1. Example 2 (Figures 3, 4; Table 2) presents the other TV-IR image pair. Smaller etalon TV fragment is clear and IR fragment corrupted by Gaussian noise. According with these and other TV-IR matching examples the following conclusions can be made:

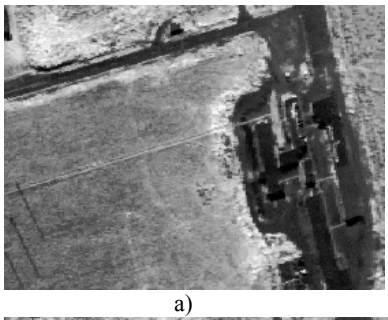
1. In case of relatively large fragment matching for clear and noisy IR images both Pytiev morphological coefficient and all geometrical correlation measures (1), (2) provide SNR and E values close to mutual information characteristics (a little bit worse or better).
2. In case of relatively small fragment matching for clear images as Pytiev's morphological coefficient as geometrical correlation measures (1), (2) provide quality values close to mutual information characteristics too. For noisy IR images matching values of CMSC (2) better than Pytiev's morphological coefficient and mutual information characteristics.

Table 1 Numeric data for TV-IR matching (Example 1).

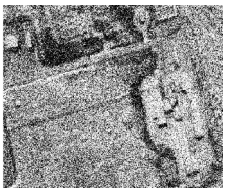
Measure	Max value	SNR	E
$MI(F,G)$	0.17562	5.6410	4.2552
$K_M^2(f - f_0, G)$;	0.13628	4.5855	2.7565
$K_M^2(F,G)$	0.05178	5.3068	3.5311
$K_{CHS}(F,G)$	0,01013	4.4291	3.6752

Table 2 Numeric data for TV-IR matching (Example 2).

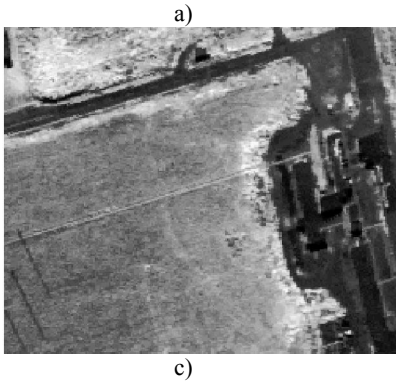
Measure	Max value	SNR	E
$MI(F,G)$	0.92975	9.2211	1.9517
$K_M^2(f - f_0, G)$;	0.13523	8.5589	1.5447
$K_M^2(F,G)$	0.01969	8.9991	1.7216
$K_{CHS}(F,G)$	0,00639	11.3575	1.9142



a)



b)



c)

Figure 1. Example 1: a) etalon TV fragment; b) test IR image; c) segmented TV fragment.

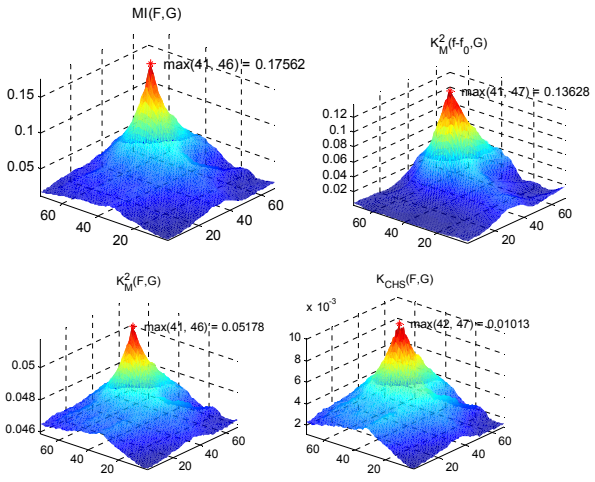


Figure 2. Correlation fields for TV-IR matching (Example 1).

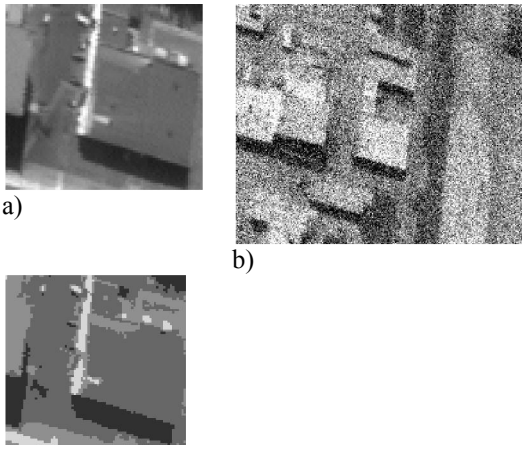


Figure 3. Example 2: a) etalon TV; b) test IR; c) segmented TV fragment.

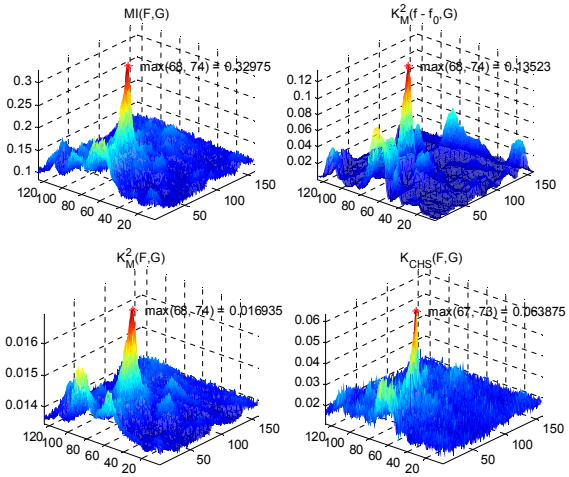


Figure 4. Correlation fields for TV-IR matching (Example 2).

Conclusion

The paper represents the comparative experimental research of previously proposed most robust geometrical correlation coefficients (1), (2) using 2D image shapes in multispectral image matching [3], [4]. These measures compared with Pytiev's morphological correlation coefficient [2] and mutual information criterion [1].

In the matching experiments quality characteristics of geometrical correlation coefficients (1), (2) was very close to mutual information. In some cases (especially for noisy IR test images) quality characteristics of CMSC (2) better than mutual information characteristics. Thus overall average difference in calculation speed between geometrical correlation coefficients and mutual information is about 20%.

References

1. Maes F., Collignon A., Vandermeulen D., Marchal G., and Suetens P., 1997. Multimodality Image Registration by Maximization of Mutual Information. IEEE Transactions on Medical Imaging, Vol. 16, No. 2, April, pp.187-198.
2. Pyt'ev Yu., 1993. Morphological Image Analysis. Pattern Recognition and Image Analysis. V.3. No 1, pp. 19-28.
3. Vizilter, Y. V. and Zheltov, S. Y.: Geometrical Correlation and Matching of 2D Image Shapes, ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., I-3, 191-196, doi:10.5194/isprsannals-I-3-191-2012, 2012.
4. Vizilter, Y. V., Rubis A.Y. Metric space of image shapes // Intellegent Information Processing IIP 9 pp. 406-409 (In Russian).

2D FIGURES SHAPE COMPARISON USING MORPHOLOGICAL PATTERN SPECTRA AND EMD METRICS[1]

Y.V. Visilter[2, 3], S.V. Sidyakin[2, 4]

[2] FGUP "State research institute of aviation systems", Russia, 125319, Moscow, Viktorenko street, 7, [3]viz@gosniias.ru, [4]sersid@bk.ru

This paper demonstrates in some experiments the possibility of successful solution for the tasks of classification of two-dimensional figures and binary images in real time on the basis of the comparison of discrete-continuous morphological Maragos spectra using the Earth Mover's Distance (EMD) with L^1 ground distance.

Introduction

In Maragos paper [1] the technique was proposed for describing flat figures and images with the morphological shape-size pattern spectra computed by Serra mathematical morphology [2]. In the paper [3] was proposed a computationally efficient way to build morphological spectra of shapes and images, based on the use of continuous skeletal morphology of Mestetskiy [4]. This made possible the use of morphological spectra in the real time vision systems, and therefore, the question of the potential areas of their practical application became up-to date.

Today in most of the well-known practical applications, morphological spectra are mostly used for texture analysis. Examples of the usage of morphological spectra as a descriptor object forms in recognition problems are extremely rare.

In this paper, it will be shown experimentally that the comparison of the morphological spectra in image recognition tasks can be successfully implemented with the usage of EMD-metrics [5]. However, in practice, there can be uninformative images along with a variety of "rich" images in terms of thickness. In this case, morphological pattern spectrum could be used as an additional instrument in shape recognition tasks.

Maragos discrete morphological spectrum

Let us given a figure X and a structuring element B, including the origin (0,0) of the image plane P. And there is a parametrically scalable structuring element B(r), having the form of figure B:

$B(r) = \{rb \mid b \in B\}$, $r \geq 0$, $b = (x_b, y_b) \in P$,

where r is a scale, b is a point with coordinates (x_b, y_b). Consider the binary image $X \subseteq P$. Let us define the discrete morphological pattern spectrum (PS) of X as a function [1]:

$$PS_{X,B,\Delta r}(r_i) = -\frac{S(X \circ B(r_i)) - S(X \circ B(r_{i+1}))}{r_i - r_{i+1}}, r \geq 0 \quad (1)$$

$$PS_{X,B,\Delta r}(-r_i) = \frac{S(X \bullet B(r_i)) - S(X \bullet B(r_{i+1}))}{r_i - r_{i+1}}, r < 0 \quad (2)$$

where $S(X \circ B(r_i)) = \|X \circ B(r_i)\|_{L1}$ - area of opening (closing if $S(X \bullet B(r_i))$), $r_i = i\Delta r$, $i \in Z$, Δr – a sampling step of the scale. The expressions (1) and (2) specify the spectrum, respectively, for positive and negative parts of the r axis.

In this paper we propose to use EMD metric for comparison of Maragos morphological pattern spectra.

EMD metric and the histogram comparison task

In paper [5] the concept of EMD-metrics (Earth Mover's Distance) was introduced. They are used for comparison of «similar histogram» descriptions, represented by a finite set of pairs $\langle P_i, h_{P_i} \rangle$, where P_i is «the cluster representative», and h_{P_i} is «the weight of the cluster». It is considered, that the objects belong to some set O, and weight - nonnegative real numbers. If there is a known ground metric d_E, that allows pairwise

[1] This work was supported by grants 12-07-31218-mol_a, 12-07-92695-IND_a, №11-08-01039-a.

comparing of objects from O, on the basis of the ground metric can be defined EMD metric to compare two similar histogram descriptions P and Q:

$$d_{EMD}(\mathrm{P},\mathrm{Q}) = \quad (3)$$
$$= \min_{\{h_{ij}\}} \sum_{j=1,..,m} \sum_{i=1,..,n} h_{ij}\, d_E(P_i, Q_j),$$
$$\sum_{j=1,..,m} h_{Qj} = 1,\ \sum_{i=1,..,n} h_{Pi} = 1,$$
$$\sum_{j=1,..,m}\ \sum_{i=1,..,n} h_{ij} = 1,$$
$$\forall i,j:\ h_{ij} \ge 0,\ h_{Pi} = \sum_{j=1,..,m} h_{ij},$$
$$h_{Qj} = \sum_{i=1,..,n} h_{ij}.$$

It is easy to verify that when the «objects» P_i and Q_j are the values of X_i and Y_j of some random variables X and Y, and, accordingly, the description of P and Q represent a discrete histogram of their distribution (for example, the halftone image histogram). Distance (3) is a special case of 1-th order Wasserstein distance [6]. And the 1-th order Wasserstein distance is a special case of Monge-Kantorovich distance [7]. EMD-distance estimates the minimum transformation cost of the description P in the description Q. In the general case, the distance (3) is calculated by solving the so-called «transportation problem» by the linear programming method.

In the special case when the random variables are one-dimensional: $x,y \in R$, histograms $P(x_i)$ and $Q(y_j)$ are one-dimensional discrete arrays. If the basic metric on R is L^1 and $\sum_{j=1,..,m} Q(y_j) = \sum_{i=1,..,n} P(x_i)$, as shown, in particular, in [8], EMD-L^1-distance can be calculated by the following formula:

$$d_{\mathrm{EMD-L^1}}(\mathrm{P},\mathrm{Q}) = \int_{-\infty}^{+\infty} |F(x) - G(x)|\,dx, \quad (3')$$

where F(x) and G(x) are the distribution function for $P(x_i)$ and $Q(y_j)$:

$F(x) = \int_{-\infty}^{x} P(t)dt,\ G(x) = \int_{-\infty}^{x} Q(t)dt$

Thus, there exists an effective algorithm for calculation of EMD-L^1 distance.

As is described below, the discrete morphology spectrum (1) and (2) is a histogram of the discrete Maragos transformation (thickness map) of the binary image. So EMD-L^1 metric can be used for comparison of the binary image spectra.

Maragos transformation and its relationship with the morphological spectra

Let us given a binary figure X, that fully fits on the frame K: $X \subseteq K$. Denote $X^{C(K)} = K \backslash X$ - complement or background of X on the frame K. Then the binary image can be associated with the figure X:

$$f_X(x,y) = \begin{cases} \mathbf{1},\ if\ p = (x,y) \in X; \\ \mathbf{0},\ if\ p = (x,y) \in X^{C(K)}. \end{cases}$$

Let us choose a binary figure B(q,r) with the center at the point q=(x,y)=B(q,0) and the scale parameter r as a translatable and scalable structuring element (in the particular case, it could be a disk).

Maragos transformation (thickness map) of a binary image with structuring element B is real-valued image defined on the frame K, in each point of which the maximum size of its covering structuring element is stored. This covering structuring element is fully inscribed in the figure X, or in complement of it (in the last case, the value of the scale parameter is negative):

$$T_B(f_X) = t_{X,B}(x,y) = \quad (4)$$
$$= \begin{cases} -\max_{r \in R} \{(x,y) \in B(q,r) \subseteq X^{C(K)}\}: \\ \qquad\qquad : (x,y) \in X^{C(K)}; \\ \mathbf{0} : (x,y) \in \partial X = X^{C(K)}; \\ \max_{r \in R} \{(x,y) \in B(q,r) \subseteq X\} : (x,y) \in X. \end{cases}$$

Let's establish the relationships between the transformation (4) and the Serra morphology operations of opening, closing [2]:

$X \circ B(r) = \{(x,y) : t_{X,B}(x,y) \ge r\};$

$X \bullet B(r) = \{(x,y) : t_{X,B}(x,y) \ge -r\}.$

Let's define the function $\chi_{X,B}(x,y,r) = \{1 : t_{X,B}(x,y) \ge r;\ 0 - \text{otherwise}\}$ and then introduce a measure:

$$\mu_{X,B}(r) = \left\| \chi_{X,B}(x,y,r) \right\|_{L^1} = \iint \chi_{X,B}(x,y,r)\,dxdy$$

Because $\mu_{X,B}(r) = \begin{cases} \left\| X \circ B(r) \right\|_{L^1} : r \ge 0; \\ \left\| X \bullet B(-r) \right\|_{L^1} : r < 0, \end{cases}$

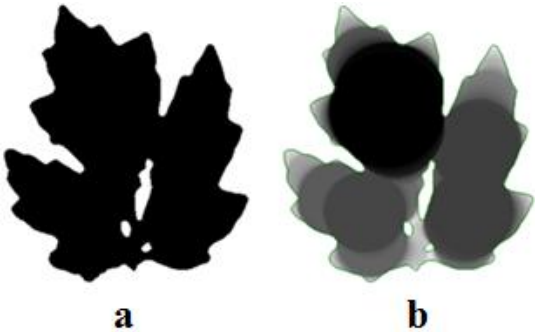
this measure allows to give an alternative definition of continuous morphological spectrum:

$$PS_{X,B}(r) = \partial \mu_{X,B}(r) / \partial r. \quad (5)$$

Thus, morphological spectrum (1)-(2) is the distribution density function of Maragos transformation values $T_B(f_X)$.

Now let scale parameter r can take only discrete values from the set RZ = {$r_i = i\Delta r$, $i \in Z$} with with some fixed $\Delta r>0$. Such a variety of sizes generates the discrete Maragos transformation (discrete-valued thickness map) (6) (Fig. 1):

$$T_{B,\Delta r}(f_X) = t_{X,B,\Delta r}(x,y) = \quad (6)$$
$$\begin{cases} -\max\limits_{r\in RZ(K)}\{(x,y)\in B(q,r)\subseteq X^{C(K)}\}: \\ \qquad :(x,y)\in X^{C(K)}; \\ \mathbf{0}:(x,y)\in \partial X = X^{C(K)}; \\ \max\limits_{r\in RZ(K)}\{(x,y)\in B(q,r)\subseteq X\}: \\ \qquad :(x,y)\in X. \end{cases}$$



a **b**

Fig. 1. Example of figure thickness map. a) – the binary figure X: b) - the thickness map for X.

Thickness map can be represented in the «mosaic» form [9]:

$$t_{X,B,\Delta r}(x,y) = \sum_{i=-i_{\max}}^{i_{\max}} r_i \chi_{X_i}(x,y), \text{ and}$$

accordingly, normalized histogram of the thicknesses map can be defined

$$H(t_{X,B,\Delta r}) = \left\{ \begin{array}{l} \left\langle r_i, h_{X_i} \right\rangle : h_{X_i} = \\ = \dfrac{\iint \chi_{X_i}(x,y)dxdy}{S} \end{array} \right\},$$

$i = -i_{\max}, \ldots, i_{\max}$.

Proposition. Discrete morphological pattern spectrum (1) - (2) is proportional to the normalized histogram of the discrete Maragos transformation (6):

$$PS_{X,B,\Delta r(r_i)} = \frac{S}{\Delta r} h_{X_i}.$$

Proof. For $r_i \geq 0$:

$$PS_{X,B,\Delta r}(r_i) = -\frac{S(X\circ B(r_i)) - S(X\circ B(r_{i+1}))}{r_i - r_{i+1}} =$$
$$= -\frac{\|X\circ B(r_i)\|_{L^1} - \|X\circ B(r_{i+1})\|_{L^1}}{r_i - r_{i+1}} =$$
$$= -\frac{\mu_{X,B}(r_i) - \mu_{X,B}(r_{i+1})}{-\Delta r} = \frac{\|\chi_{X_i}(x,y)\|_{L^1}}{\Delta r} =$$
$$= \left(\frac{\iint \chi_{X_i}(x,y)^{dxdy}}{S} \right) \left(\frac{S}{\Delta r} \right) = \frac{S}{\Delta r} h_{X_i}, QED$$

By analogy for $r_i<0$.

So EMD-distance (3) between morphological spectra of the binary image figures of *X* and *Y* could be evaluated by formula (3).

Truncated spectra. The invariance to a shift and rotation of figures

It is obvious, that the positive part of the Maragos transformation and the positive part of the morphological pattern spectrum remain unchanged when it is shifted or rotated. So the truncated from the bottom discrete morphological pattern spectrum is defined as:

$$PS_{X,B,\Delta r,r0+}(r_i) = \begin{cases} PS_{X,B,\Delta r}(r_i), \text{ if } r_i > r_0; \\ \sum\limits_{j=-i_{\max}}^{r_0} PS_{X,B,\Delta r}(r_j), \text{ if } r_i = r_0; \\ \mathbf{0}, \text{ if } r_i < r_0. \end{cases} \quad (7)$$

In this paper experiments $r_0=0$. Such truncated spectra without the negative part allow comparing not only the figure images, but also the shapes of figures, invariant to shift and rotation. In practical applications, a truncation parameter r_0 can be selected more or less than 0, which enrich or impoverish the spectrum.

EMD-L^1 comparison of truncated morphological spectra for pattern classification

The dataset Myth [10] of the silhouettes was considered in our first experiment. The three classes of beings were taken into account: men, centaurs, and horses. Thus, the part of the figures within the class was subjected to

deformation or new objects were added to the figure. Evaluated experiment results showed that the value of the EMD-L^1 distance between the morphological spectra monotonously increases from the base image of a man figure to the image of the centaur (horse) figures, and, what is more, this metric considers that the most similar figure images are such images, which shapes are the most similar (Fig. 2).

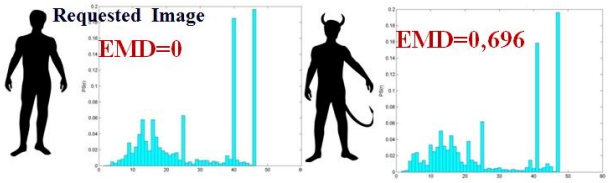


Fig. 2. Comparison of the figures according to their morphological spectra with EMD- L^1.

There was calculated the table of average distances between the classes from the Myth (Tab. 1). The another experiment with Tools 2D dataset [10] showed the same result.
They testify to the fact that the EMD-L^1 metric spectrum comparison of figures in this experiments divides well the objects of different classes. This is consistent with our intuitive perception of forms similarity.

Table 1. Average EMD-L^1 distances between the classes for Myth.

	people	centaurs	horses
people	0,812	3,054	6,31
centaurs	3,054	0,69	2,25
horses	6,31	2,25	0,78

The conducted experiments have demonstrated that the comparison of the morphological spectra by EMD-metrics is an adequate way to compare figures in the case if the compared figures have quite a good variety in terms of thickness. The EMD-L^1 comparison is stable to small deformations of flexible objects, to their orientation modification on the plane and to slight noise at the borders of the figures.
The experiments were carried out on a PC with Core i5-2320. The results (tab. 2) confirmed the computational efficiency of the considered approach: (a) an average skeletonization time for one image, ms; (b) the average time of morphological spectrum construction, ms; (c) the average time of calculation of EMD-L^1 distances in comparison of spectra, ms; (d) the overall average time, ms.

Table 2. The average comparison performance in millisecond.

Image sets	(a)	(b)	(c)	(d)
Tools	4,8	3,2	less 0,1	8
Myth	6,8	15	less 0,1	21,8

Conclusion

The method was proposed for comparison of discrete morphological spectra of binary images and flat figures using EMD-metrics. The concept of thickness map was introduced. The relationship between Maragos morphological pattern spectrum and histogram of thickness map was proven. It was experimentally shown, that the comparison of truncated morphological spectra with the usage of EMD-L^1 metric can be used successfully in the shape object classification tasks.

References

1. Maragos P. Pattern Spectrum, Multiscale Shape Representation.// IEEE Trans.on pattern analysis, machine intelligence, Vol, II, No 7, July 1989.
2. Serra J. Image Analysis and Mathematical Morphology. // Academic Press. London – 1982.
3. Vizilter Yu. V., Sidyakin S.V. Computation of Morphological Spectrum for Grayscale Images // Vestn. Komp'yut. Inf. Tekhnol., No. 4, 8–17 (2012). [in Russian].
4. L. M. Mestetskii, Continuous Morphology of Binary Images: Figures, Skeletons, Circulars (Fizmatlit, Moscow, 2009) [in Russian].
5. Y. Rubner, C. Tomasi, and L. J. Guibas. "The EaTRh Mover's Distance as a Metric for Image Retrieval", International Journal of Computer Vision, 40(2):99-121, 2000.
6. S. S. Vallender, "Calculation of the Wasserstein Distance Between Probability Distributions on the Line", Theory Probab. Appl., 18(4), 784–786.
7. Rachev, S.T. "The Monge-Kantorovich mass transference problem and its stochastic applications. Theory of Probability and its Applications", XXIX(4):647–676, 1984
8. Deza E.I., Deza M.M. "Encyclopedic Dictionary of distances" // Nauka, Moscow, 2008) , pages 444 . [in Russian].
9. Yu. P. Pyt'ev, "Problems on Image Morphological Analysis," // Mathematical Methods for Researching the Earth Resources from the Space (Nauka, Moscow, 1984), pp. 41–83 [in Russian].
10. Dataset, http://tosca.cs.technion.ac.il/book/resourcesdata.html

USE OF COLOR INFORMATION AND GRADIENT OPERATOR IN INFORMATIVE AREA SEARCH AND DETERMINATION OF SCANNING WINDOW DIMENSIONS IN THE PROBLEM OF CONSTRUCTION OF DISPARITY MAP

A. Volkovich[1]

United Institute of Informatics Problems, National Academy of Sciences, Minsk, Belarus, Surganov str 6, 220012, anvolkovich@gmail.com

The paper addresses issues of construction of disparity maps in the problem stereo reconstruction. Proposed the approaches to improve the quality of repair and optimize the execution time.

Introduction

The problem of the depth reconstructing on the base of several images captured at the same time, is extremely important and necessary for many applications. One of the areas of practical application is the 3D- reconstructing of the object on its images taken from different angles. This fundamental problem of many researches in computer graphics, but also in cognitive science and neuroscience.

Reconstruction of volume information on stereo images

The main stage of constructing a three-dimensional model is the task of matching point searching on the two rectified images (drawing a map of disparity). In this case the disparity can be understood as the pixel offset of one image point relative to its corresponding point on the other.
Disparity map building is a complex computational problem. The problem of correspondence search completely unresolved. There is a wide range of functions used to determine the similarity measure regions, but as a commonly used input brightness information (monochrome image processing.) The disadvantage of this method is that the multiplicity of interpretations of colors for the image points with the same value of brightness. Thus, in order to improve the adequacy of search results matching seems possible to use color component image pixels. In addition, there is a problem processing homogeneous areas of images that have a small amount of information needed to determine the correspondence between the points of stereo images. Hence the need for a system which allows itself to change the size of a scanning window or screen out non-informative area.

The use of color characteristics in corresponding point searching

In world practice, when working with images in stereo recostruction problems, only intensity information is usually used as a criterion for comparing image points. A disadvantage of this approach is the multiplicity of interpretations of colors for points with the same intensity value. In addition, nonuniformity in the perception of color and monochrome images should be taken into account. This feature is considered in methods of image color model degradation to 256 shades of gray by introducing coefficients applied to the appropriate channels.
Given that most images are originally created in color by a color sensor, in order to enhance the effectiveness of the project, the authors modified the algorithm to enhance its color information processing.
When moving to three components, images can be represented as "clouds" of points in the three-dimensional space with the axes corresponding to color channels of the image. However, the RGB space is not orthogonal because of the specificity of the human's visual analyzer, which has a

different number of rods and cones sensitive to a particular color.

Since Euclidean distance, which is applicable to the orthogonal systems, is used as a measure for comparing points in the three-dimensional space, the RGB space should be orthogonalized into the XYZ space. The representation of basic RGB colors as recommended by ITU in the XYZ space is as follows:

Red: x = 0,64 y = 0,33;
Green: x = 0,29 y = 0,60;
Blue: x = 0,15 y = 0,06.

After reduction of spaces, operations used in the case of orthogonal systems can be applied to the points.

Thus the measure of similarity of points can be represented as

$$L = \sqrt[2]{(X_2 - X_1)^2 + (Y_2 - Y_1)^2 + (Z_2 - Z_1)^2}$$

where Xn,Yn,Zn– are the respective components of compared points.

The maximum length of the vector in the space constructed for the 8-bit color channel is about 441 units, which in terms of informativeness is 1.7 times higher than the comparison of points in a halftone representation of the image with a maximum distance between points of 256 units. In turn, after transition to the color comparison measure, the number of unique values describing points increases by 65 000 times in comparison to monochrome.

Numerical experiments showed a significant improvement in the adequacy of reconstruction by the modified algorithm, which makes it possible to more clearly define the boundaries of objects with similar visual characteristics, objects that visually blend into the background, and objects with fine details.

In turn, there is still the problem of processing of image areas that are not in focus, objects with large homogeneous regions, and an infinitely distant object (for example, the sky). Such image areas have practically no discontinuities (jumps) in intensity that carry the maximum amount of information used in image processing. Obviously, the problem of finding such areas arises. This will make it possible to exclude uninformative areas from processing and, hence, to reduce the number of errors and iterations of the algorithm.

Use of a gradient operator to filter out uninformative areas and search-window size determination

Despite an increase in the range of values being compared by using color information remains a problem processing portions of images that are not in focus, objects with large homogeneous regions, as well as infinitely distant objects (such as the sky). Such image area have practically no discontinuities (swings) in the brightness, which in turn carry the maximum amount of information used in image processing. This obviously arises the problem of finding sites that will exclude uninformative treatment area and thus reduce the number of errors and the number of iterations.In order to classify a point located on the intensity jump, the variation in intensity associated with this point should be significantly greater than the intensity variation at the background point. Due to the nature of local computations, the method of determining "significant" values consists in establishing a threshold. In turn, the concept of the first and second derivatives is used to quantify intensity variations.

Determination of the image point as the intensity jump occurs if its two-dimensional derivative of the first order exceeds a predetermined threshold. In accordance with a predetermined connectivity criterion, the connected set of such points is the intensity jump and the prolonged intensity jump is the contour.

One approach to connecting points of the contour is analysis of the characteristics of pixels within a small neighborhood (3×3 or 5×5) of each point (x,y) of the image that has been marked as a contour point (jump point). All points that are similar according to some predefined criteria are connected and form a contour consisting of pixels which satisfy these criteria. In this analysis, the following two main options for establishing the similarity of contour pixels are used, i.e., the gradient operator response and the direction of the gradient vector.

Calculation of the first derivative of the digital image is based on different discrete approximations of a two-dimensional gradient. The direction of the vector gradient coincides with the direction of the maximum rate of change in function f at point (x,y).

B Calculation of the image gradient consists in obtaining partial derivatives $Gx = df / dx$ and

$Gy = df / dy$ for each point. One method of finding the first partial derivatives Gx and Gy at a particular point consists using the following Sobel gradient operator:

$$Gx = (z7 + 2 * z8 + z9) - (z1 + 2 * z2 + z3)$$

$$Gy = (z3 + 2 * z6 + z9) - (z1 + 2 * z4 + z7)$$

For the Sobel operator, which detects horizontal and vertical contours (intensity jumps), it is necessary to determine the appropriate masks for convolution with the original image. It is also possible to change the given formulas so that they provide the maximum response for diagonally directed contours. The sum of coefficients of each mask is zero, so these operators will give a zero response in constant intensity areas, which is characteristic of a differential operator.

Processing of the image by the gradient operator, as well as subsequent binarization of the results, makes it possible to construct the map of calculations.

In most cases occurring in practice, the Sobel operator is applied to a monochrome version of the image. However, as mentioned above, conversion of the color model can distort data in informative areas. Therefore, the authors proposed to process each color channel and their ensemble by the Sobel operator when constructing the map of calculations..

Map computations can be used for image processing using dynamic scanning windows. This process is based on the gradual increase of the window size in the event that it is scanned in the area is insufficient information.

This approach reduces the number of processing errors homogeneous areas, however it should be noted that as the window size is a "swelling" of the boundaries of the object which occurs when a window size 50x50, which distorts the real shape of the object.

Also note averaging the comparison result based on the specific areas of the functions of similarity measures, leading to a reduction in the processing efficiency when the number of pixels in the scanning window begins to exceed the dimension of the range of possible values X Y of the similarity function.

Thus, we can say that it is possible to automatically determine the dimensions of the window in a range that is determined by the bit depth input data based on the requirements of inequality for the converted *XYZ* color space:

$$S_{scan} \le \sqrt{((\max(X) - \min(X))^2 + ((\max(Y) - \\ - \min(Y))^2 ((\max(Z) - \min(Z))^2}$$

where S_{scan} – area scanning window;

$\max(X \dots Y \dots Z)$ –maximum value of the component;

$\min(X \dots Y \dots Z)$ – minimum possible value of the component.

In the case where the dimension of the window does not reach the maximum possible "brought together" value in the scanned area sufficient information should make this point screening, due to the fact that information about its position in the space is ambiguous.

Sifting of uninformative areas during computer simulation made it possible to reduce the number of errors in the disparity map and processing time by 20% (including calculation time of local gradient values) compared with the full processing of 860 000 image points.

Conclusion

During the research analyzed the possibility of use a color image information in searching the matches between the pixels in the image problem of the construction of disparity maps. Investigated the specificity of the use of RGB-model and bring it to the orthogonal XYZ-model. Computational experiments on real images that showed a significant improvement in the results of constructing maps of disparity compared to methods based on using only the intensities of the pixels of the image.

Also produced by the potential use of the gradient operator to search for informative areas of the images and determine the dimensions of the scanning window in the construction of disparity maps.

References

1. Linda G. Shapiro and George C. Stockman. Computer Vision. Prentice Hall. 2001, 580 p. ISBN 0-13-030796-3.
2. Lyakhovsky V.V, Volkovich A.N. Zhuk D.V, Tuzikov A.V. Automatic reconstruction of three-dimensional scenes on several images. Proceedings of the V Belarusian Space Congress, 25-27 October 2011., Mn.: UIIP NASB, 2011, Volume 2, p. 129-133. (in Russian)
3. Tuzikov A.V., Scheinin S.A., Zhuk D.V. Mathematical Morphology, moments stereoprocessing: selected issues of processing and analysis of digital images. Minsk, Belarus. science, 2006.-198p (in Russian)

CONSTRUCTION OF FAULT-TOLERANCE SIGNAL FEATURES SUBSETS[1]

A.E. Yankovskaya[2,3], R.V. Ametov[4]

[2] Tomsk State University, Lenin Ave. 36, Tomsk, 634050, Russia
[3] Tomsk State University of Architecture and Building, Tomsk, Russia, 2, Solyanaya Sq., Tomsk, 634003, 8(3822)65-33-51, ayynkov@gmail.com, yank@tsuab.ru4 rin@tsuab.ru

It is proposed to use new kinds of regularities, fault-tolerant subsets of signal features in intelligent systems based on the test pattern recognition. The changing of these features values points to the possibility of the transition of object states from one pattern to another. The algorithms of fault-tolerance signal features revealing are given and the subsystem that implements it is described. The developed subsystem is based on test methods of patterns recognition and is included into intelligent instrumental tool IMSLOG. The examples of the different problem areas for which urgent problem of revealing subsets of signal features that are tolerant to measurement errors are given.

Introduction

One of the main problems of intelligent systems creation is the problem of revealing regularities in data and knowledge [1-3]. The concept of signal features was introduced in 2007 [2]. Hereinafter we use the definitions of the regularities given in [3] and introduce a new kind of regularities named subset of signal features that are tolerant to measurement errors (fault-tolerant subset of signal features).

The regularities are subsets of features with easy in order to interpret certain properties that affect the distinguishability of objects from different patterns. These regularities are consistently observed for the objects from the training sample and manifested at other objects of the same nature, as well as the weight coefficients of features describing their individual contribution to the distinguishability of objects from different patterns. Signal features of the 1st kind are the minimal subset of characteristic features that distinguish objects belonging to the two different patterns.

Signal features of the 2nd kind are the minimal subset of characteristic features that distinguish the description of the object under study belonging to one pattern from describing objects belonging to another pattern.

The transition process of the objects description from one pattern to another is qualitative and is particularly important in the areas of knowledge such as medicine, psychology, ecology, psychiatry, education, etc. Therefore, great importance is accounting the measurement errors in the signal features values that define this transition and the construction of fault-tolerant subsets of signal features.

Mathematical foundation of fault-tolerant subsets of signal features

Matrix model of data and knowledge representation [4-6] includes matrix of descriptions (Q) of objects in the characteristic features space and matrix of discrimination (R) of the objects in the classifying features space. Each element of the matrix Q defines value of the j-th feature for the i-th object. Element r_{ij} of the matrix R defines whether i-th object belongs to one of the selected classes by the j-th mechanism of classification.

The set of the nonrecurrent rows of matrix R is associated with the set of specified patterns.

[1] This research is funded by Russian Foundation for Basic Researches grants (no. 13-07-00373 and no. 13-07-98037) and by Russian Humanitarian Scientific Foundation grant (no. 13-06-00709).

Elements of the pattern (class) are objects, which are presented by rows of matrix *Q*, and correspond to similar rows of matrix *R*. In case when the only mechanism of classification is given then the discrimination matrix degenerates to the column and this case corresponds to the traditional knowledge representation in pattern recognition problems [7].

This model allows one to represent not only data, but also the knowledge of experts, since one row of matrix *Q* can determine in interval form a subset of the objects of the same decision determined by the row of matrix *R* [4,5].

The diagnostic test is called minimal when it contains minimal number of features [7].

Submatrix *Q*' of the matrix *Q* and submatrix *R*' of the matrix *R* present compressed description of the patterns [4,5,8]. Columns of matrix *Q*' correspond to characteristic features, that belong to all the minimal tests, and from the compressed description of the objects are deleted all the similar rows as well as all their correspondent rows from matrix *R*. Also weights values of remaining rows in matrix *Q*' should be corrected.

Constructed on the bases of matrices *Q* and *R* binary irredundant matrix of implications *U*' defines discrimination of the objects from different patterns [4,5]. There are no absorbing rows in matrix *U*'.

It is necessary to create subsystem of signal features revealing based on test patterns recognition methods [4,6,9] and include this system into the instrumental intelligent tool IMSLOG [10].

Solution of this problem requires the following:

1. For each pair of compressed descriptions of patterns it is necessary to select all subsets of signal features of the 1st kind.

This problem is reduced to the search of all the minimal tests that discriminate objects from each pair of patterns.

2. Using the object under investigation, belonging to one of the patterns, and the given compressed description of the patterns it is necessary to reveal all the subsets of signal features of the 2nd kind.

This problem is reduced to the problem of selection of all the minimal tests that discriminate object under investigation from the object of the given pattern.

Revealing of fault-tolerant subsets of signal features

We will use irredundant matrix of implications (U') [4,5] for revealing fault-tolerant minimal subsets of signal features. This matrix defines distinguishability of objects from different patterns (classes at fixed classification mechanism). All pairs of objects of the two selected patterns will be used at U' matrix construction for revealing fault-tolerant minimal subsets of 1-st kind signal features. Pairs which consists of the object under investigation and every objects from selected pattern will be used at U' matrix construction for revealing fault-tolerant minimal subsets of 2-nd kind signal features.

Distinguishability of objects from different patterns is determined by using a vector-function of distinguishing $f(a,b)$ [10] which values are calculated by the following formula:

$$f(a,b) = z(a)\bar{z}(b) \vee \bar{z}(a)z(b), \qquad (1)$$

where $\mathbf{z}(r)$ equals *r*-th row of **Q'** matrix, $r \in \{a,b\}$, *a, b* – objects from different patterns (classes at fixed classification mechanism). Note that if the features are 0,1, then operations $\vee$, $\wedge$, $\neg$, are used, and if features are "–", then result of operations is equal 0.

The value of the vector-function of distinguishing on the basis of the i-th feature is calculated by the formula:

$$f_i(a,b) = z_i(a)\bar{z}_i(b) \vee \bar{z}_i(a)z_i(b) \qquad (2)$$

The dimension of the vector-functions of distinguishing equal to *m*.

Theorem 1. A necessary and sufficient condition for constructing fault-tolerant subsets of signal features (tolerant to a number not exceeding *t* errors of measurement (input) of the *i-th* feature (all features) values) is ensuring the following condition in the matrix Q':

$|f_i(a,b)| \geq h$ $(|f(a,b)| \geq h)$,

where $h = 2t + 1$ for the *i-th* feature (all features) at the calculation vector-functions of distinguishing for all couples object-object from different patterns (classes at a fixed mechanism of classification).

Theorem 2. A sufficient condition for constructing fault-tolerant subsets of signal features (tolerant to a number not exceeding *t* errors of measurement (input) of the *i-th* feature (all features) values) is ensuring the following condition in the matrix Q':
$|f_i(a,b)| \geq h \ (|f(a,b)| \geq h)$,
where $h = 2t + 1$ for the *i-th*- feature (all features) at the calculation vector-functions of distinguishing for all pairs pattern-pattern, pattern-object, object-object from different patterns (class-class, object-class, object-object for different classes at a fixed mechanism of classification).

Consequence: If the condition of theorems *1*, 2 does not satisfy for the row *Uj'*, then it is impossible to provide a fault-tolerant subsets of signal features on the features associated with different from *0* elements of the rows of the matrix *Uj'*.

On the basis of the **Theorem 1** and **Theorem 2** is constructed irredundant matrix of implications *U'*. In the process of a matrix *U'* construction the absorbing rows are removed.

To construct the minimal subsets of signal features that are resistant to the number *t* of errors of measurement (input) values of features in the description of the object, it is necessary and sufficient that each row of the matrix U' containing at least *h (h = 2t +1)* values equals to 1 [11].

In order to ensure condition of sustainability to measurement errors (conditions of fault-tolerance) it is suggested a modification of matrix *U'* constructing algorithm that was given in [1]. Unlike the algorithm [1], during the construction of the matrix *U'* are removed only covering rows in which the number of *1* values at least *h*. Furthermore, the deleted covering rows in which the number of *1* values is less than *h*. In the latter case, the condition for fault-tolerance features included in the covered rows will not be provided. If condition of fault-tolerance is satisfied, then the problem of constructing fault-tolerant subsets of the signal features is reduced to the problem of finding the minimal h-fold column-coverings of the matrix *U'*.

Thus, the construction of minimal irredundant unconditional diagnostic tests includes three stages:

1. ***U'*** matrix construction on the base of two selected patterns or selected pattern and the description of object under investigation.
2. Search for *h*-fold the minimal column-coverings of matrix U '.
3. Construction matrix of tests on the basis of the minimum *h*-fold column-coverings of the matrix *U'*: T^1 matrix for signal features of the *1st* kind and the matrix T^2 for signal features of 2nd kind. The rows of the matrices T^1 and T^2 are corresponding to diagnostic tests, the columns are corresponding to columns of the matrix *Q* (characteristic features). Values equal to *1* in each row of the matrix marks features that are included in the test corresponding to this row.

Conclusion

New kinds of regularities named fault-tolerance subsets of signal features are introduced and algorithms of its construction are suggested. Also actuality of its using for a number of problem areas is shown.

Ways to further investigation are as follows: algorithms of fault-tolerance subsets of signal features will be implemented as plug-in module for instrumental tool IMSLOG designed for knowledge regularities revealing and for construction applied intelligent systems.

In the first place, the algorithms will be tested for the problem area of organization stress and depression diagnostics for peoples with mental disease.

References

1. Yankovskaya A.E. Logical Tests and Means of Cognitive Graphics. Published by LAP LAMBERT Academic Publishing. – 2011. – 92 p. [in Russian].
2. A. Yankovskaya. New Kinds of Regularities in Knowledge and Algorithms of Their Revealing // 7th Open German/Russian Workshop on Pattern Recognition and Image Understanding.
3. A.E. Yankovskaya, "Decision-Making, Tolerant to Measuring Errors of Features Values in Intelligent Systems," Proceedings of the 10th International Science-Technical Conference of Artificial Intelligence, Intelligent Systems, September 28-October 3 2009, Gelendzhik, Divnomorskoe, Izd. TTI YuFU, Taganrog, 2009, pp. 127-130 [in Russian].
4. A.E. Yankovskaya. Logical tests and means of cognitive graphics in intelligent system // New Information Technologies in Investigations of

Discrete Structures: Proceedings of the 3-d All-Russian Conf. with Foreign Participants. – Tomsk: SO RAS, 2000. – P. 163-168. (in Russian).

5. Yankovskaya A.E. Logic-Combinational Probabilistic Recognition Algorithms// Pattern Recognition and Image Analysis. – 2001. – Vol. 11, No. 1. – pp. 123-126.
6. A.E. Yankovskaya, A.I. Gedike. Theoretical-methodological basis for creation of intelligent test pattern recognition systems // Tavricheskiy vestnik informatiki i matematiki. – 2004. – № 1. – P. 83-94.
7. Yu.I. Zhuravlev, I.B. Gurevitch. Pattern Recognition and Image Analysis // Artificial Intelligence in 3 books. Book 2. Models and Methods: Reference Book / By ed. D.A. Pospelov. – M: Radio and Comm. – 1990. – P. 149-191. (in Russian).
8. Yankovskaya A.E. and Gedike A.I. Construction and Evaluation of Compressed Descriptions of Patterns in an Intelligent Recognizing System// Pattern Recognition and Image Analysis. – 1999. – Vol. 9, No. 1. – pp. 124-127.
9. Yankovskaya A.E. Minimization of Orthogonal Disjunctive Normal Forms of Boolean Function to be Used as a Basis for Similarity and Difference Coefficients in Pattern Recognition Problems// Pattern Recognition and Image Analysis. – 1996. – Vol. 6, No 1. – pp. 60-61.
10. A.E. Yankovskaya, A.I. Gedike, R.V. Ametov and A.M. Bleikher, “IMSLOG-2002 Software Tool for Supporting Information Technologies of Test Pattern Recognition,” Pattern Recognition and Image Analysis, Vol. 13, No. 4, 2003, pp. 650-657.
11. Yankovskaya A.E. Distinguish functions at knowledge base analysis for intelligent systems with matrix knoowledge representation // Artificial intelligence -90. Abstracts of II union conference. T 1. – Minsk, 1990. – pp. 102-105 [in Russian].

ABOUT THE ORDERING OF FINITE SET OF POINTS ON PLANE RELATIVE A ROTATION CENTER

A. A. Zharkikh[1,2], S. M. Bychkova[1,3]

[1] Murmansk State Technical University, Musmansk, Russian Federation, 183038, box 387, [2]zharkikh090107@mail.ru, [3]LyasnikovaSM@yandex.ru

Algorithms of the ordering of finite set of points on plane relative a rotation center are presented in paper. Two options of the ordering of points are presented. The first option is based on the geometric location of points of set and does not require additional labels. The second option is required information about the mutual position of points and the introduction of some additional labels in the form of any physical parameters.

Introduction

In [1, 2] we presented the probability distributions of Euclidean distances between two copies of a finite set of points on plane. Copies of set are distinguished by random rotations or reflections. In this paper we consider only rotations. The Euclidean distance was calculated to estimate changing in position of a set of points. The Euclidean distance was introduced under the condition that coordinates of all points are known and strict the ordering of these coordinates was defined. The strict ordering can be introduced, using additional physical parameters: temperature, brightness, etc. In this paper, we investigate the possibility of the ordering of finite set of points relative a rotation center, based on their geometric arrangement. Such an approach in a number of applications can eliminate the need for any additional measurement of parameters for the ordering of points.

A review of methods of the ordering of points

In [3] simple transformations, such as parallel shift, similarity transformation, scaling transformation and etc. are considered for images transformation. Authors consider different options of images transformations and they build the code of image. This code is invariant to the considered transformation. For example, in paper the case was considered when image A was shifted and rotated some way. Obviously, the mutual position of points did not change. Authors constructed the image code $A <M_A, T_A>$, using the property of preserve of distances between points. Here M_A is function of the numbering of the set A. The set T_A is set of numbers $r(q, m)$, which are the distance between points with numbers q and m.

Approaches to ordering of grouped spatial point objects were considered in [4]. Authors noted that the primary number may be assigned at any point on standard rule, for example, in the order of point appearance, when a scene is being scanned. But this way is not invariant to rotation transformation. Therefore it is necessary to use regularities of the mutual position of points. Authors consider that needed to assign number every point of a set. This number should be invariant to a rotation or other transformation, as well as should be resistant to noises in coordinates. The procedure is performed for the original image and noisy image. Authors distinguish the following methods of ordering. The first, it is the numbering of points in the reference frame with the origin at the center of gravity. The center of gravity is identified as center of gravity of set of points. Then the bunch is constructed with pole, combined with the center of gravity. Further the minimum value of the modulus of the radius vector is selected and it is denoted as the first point, then numeration of the rest of radius vectors of the bunch is made in the natural order. "Minimal tree" is the second algorithm, which was considered by authors. The ordering of set of points is considered as the problem of minimizing the length of a network, which connects a predetermined number of points by cable of minimum length. "Spectrum" is the

third algorithm, which is considered by authors. The algorithm consists in the fact, that an arbitrary point of a set can be zero point, but for definiteness zero point is the closest point to the center of gravity. The first point is the closest to zero point. Point number 2 is the closest to the center of gravity of points with numbers 0 and 1. Point number 3 is the closest to the center of gravity of points with numbers 0, 1 and 2 and etc. Authors compared algorithms. "Spectrum" showed the best results.

In [5] author assigns informative characteristic to each point of a grouped spatial point object, so a point before and after the rotation has the same number. The author assumes that original and transformation (which should be recognized) objects are given. The transformation object is obtained by rotation and noise of the original object. Rotation parameters are unknown, and the noise is the measurement error of coordinates of points of the transformed object. It is assumed, that the noise is normally distributed with zero mean and known variance. Points of an original object were numbered and points of the transformed object are numbered arbitrarily, because parameters of a rotation are unknown. The procedure for ordering of original and transformation objects was developed by author and corresponding points of objects will be have the same numbers in result. The procedure for ordering of points is based on the procedure of recognition. The probability of correct ordering is decreased with the increase of noises in coordinates.

Examples of the ordering of set points on plane, based on the mutual position of points

Consider rotation of the plane relative a fixed point (x_0, y_0). The set of points on plane can be represented by two ways. The first way as the set C. Set C is the union of an infinite number of concentric circles with center at the point (x_0, y_0). Every element of this set is invariant to transformation of a rotation. The second way as set L. Set L is the union of an infinite number of rays, which emanate from the point (x_0, y_0). The set L is invariant to transformation of rotation. The ordering of a finite set of points on plane is constructed with the using sets C and L. Consider two examples.

Example 1. Consider an arbitrary finite subset of set C. Let it contains n concentric circles of radii R_i, $i = 1,\ldots, n$. Without loss of generality we can assume, that a sequence of radii is a monotonically increasing function of the number. Assume that only one point is defined on each of concentric circles. We can point out 2 ways of the ordering of these points by the monotonicity of a sequence of radii: on ascending of a radius R_i or on descending of radius R_i. The belonging of points to elements of set L is not matter.

Example 2. Consider any finite subset of set L. It contains m rays, emanating from a point (x_0, y_0). We find the intersection of this subset with one of elements of C, for example a circle of radius $R>0$. The result is a finite set of m points on a circle of radius R with center (x_0, y_0). Determination of the ordering of set points on plane is based on the geometric arrangement of points and is due to some difficulties. Consider some of options, when a finite set of points on a circle can be ordered from geometrical considerations.

Denote points A_k, $k = 0, \ldots, m-1$, so that the first measured point will be have the subscript 0, and further, points will be denote following numbers, under the traversing of circle counterclockwise. Denote α_k is the angle between rays, which are passing through points A_k and A_{k+1}, $k = 0, \ldots, m-2$ (α_{m-1} is the angle between rays are passing through A_{m-1} and A_0). We find minimum at k (1) for the ordering of this set points.

$$\min_k \alpha_k = \tilde{\alpha}_{\min}, \qquad (1)$$

$$\tilde{j} = \arg\min_k(\alpha_k),\ k = 0,\ldots,m-1.$$

The function of the ordering is defined by one of two ways as in Table, if this minimum is reached at a one value of k.

Table.

j	$\tilde{j}$	$\tilde{j}+1$	…	$\tilde{j}+m-1$
$\varphi_1(j)$	0	1	…	$m-1$
$\varphi_2(j)$	$m-1$	$m-2$		0

In Table, $j = (\tilde{j} + p) \bmod m$, $p = 0,\ldots,m-1$, $\varphi_1(j)$ is function determining the ordering of points on ascending numbers (counterclockwise), $\varphi_2(j)$ is function

determining the ordering of points on descending numbers (clockwise).

If the minimum is reached for several arguments, then we solve the problem on maximum (2) for the same set of angles.

$$\max_k \alpha_k = \widetilde{\alpha}_{\max}, \qquad (2)$$

$$\widetilde{j} = \arg\max_k(\alpha_k),\ k = 0,...,m-1.$$

If the maximum is reached at a one value k, then the function of ordering is determined by one of two ways as in Table.

We can begin the procedure of the ordering of points with the finding the minimum or maximum angle. If the checking of possibility of the ordering is began with the solving the problem at minimum, then the general algorithm can be described as follows. If minimum is one, then the ordering is performed by one of two ways as in Table. If the minimum is not unique, then there is needed solve the problem of the maximum. If the maximum is one, then the ordering is produced as in Table. If the maximum is not unique, then the ordering of points, based on geometrical considerations impossible. In this algorithm the order of solution of problems on the minimum or maximum is possible to interchange. If both problems (minimum and the maximum) have several solutions, then we can label one arbitrary point by some parameter for ordering of points on a circle. The number 0 can be assign for marked point, and then points are numerated clockwise or counter-clockwise.

If coordinates of points are set by an experimenter in advance, it is advisable to arrange points on a circle so that the one solution exists for one of problems on the minimum or maximum angle.

Generic variant of the ordering

Suppose, that a number of concentric circles with radii R_i is had and $m_i \geq 1$ points are arranged on each of circles. Suppose that the problem of ordering of points is solved for each of circles according to the algorithm from the previous section. It is understood, that the geometric ordering was done or physical mark was put for one point of the current circle. Thus, we can order easily all set of points according to following expressions (3) and (4).

$$k = \overline{0, m_1 - 1}, \qquad (3)$$

if points lie on a circle of radius R_1.

$$k = \overline{\sum_{s=1}^{i-1} m_s, \sum_{s=1}^{i} m_s - 1}, \qquad (4)$$

if points lie on a circle of radius R_i, $2 \leq i \leq n$.

Conclusion

In [1, 2] we investigated the probability characteristics of the Euclidean distance between two copies of a set of points on plane, where one of copies is randomly rotated or reflected. The introduction of the Euclidean distance between the original and the transformed sets of points is based on a strict ordering of points, which is defined. Presented in this paper algorithms of the ordering can reduce the number of measurements, which are required to introduce such a Euclidean distance. Results of this paper allow to point out options for the ordering of points at where the center of rotation is known. The first option is based only on the geometric arrangement of points and doesn't require additional labels. The second option doesn't allow to do purely geometrical measurements of distances and angles, and requires the introduction of additional labels. It should be noted, that the number of labels should be substantially less, than in the case, where all points are marked.

References

1. Zharkikh A. A., Bychkova S.M. On distribution of Euclidean distances between ordered sets of plane points at random rotations and reflections //Proceedings of the MSTU. – Murmansk, 2010. – Vol. 13, №3. – p. 592 – 606. [in russian]
2. Zharkikh A. A., Bychkova S.M. Distributions of Euclidean Distances between Copies of the Set of 2D Points When One Copy is Randomly Rotated or Reflected//Pattern recognition and image analysis: Pleiades Publishing, Ltd., 2012. – Vol. 22, No. 3. – P. 433 – 445.
3. Kudryavtsev V.B., Gasanov E. E., Podkolzin A. S. Introduction in the theory of intelligent systems: Textbook. – M.: Publishing Department of the Faculty of Moscow State University named. M.V. Lomonosov; MAKS PRESS, 2006. – 208 p. [in russian]
4. Introduction in contour analysis. Application to image and signal processing / J. A. Furman, A. V. Krevetsky, A. K. Peredreev, A. A. Rozhentsov, R.G. Hafizov, I. L. Egoshina, A. N. Leukhin. – M.: FIZMATLIT, 2003. – 592 p. [in russian]
5. Egoshina I. L. The procedure for ordering points of spatial grouped point objects// Methods and devices of information transmitting and processing/ Murom Institute (branch), 2011. № 13. – p. 81 – 85. [in russian]

SUBMISSION OF DIGITAL IMAGES BY MEANS OF FRACTAL MODEL[1]

A.L. Zhiznyakov[2,3], D.G Privezentsev[2,4]

[2] Murom Institute of the Vladimir State University,
Murom, Orlovskaya, 23, Russian Federation
[3]lvovich@newmail.ru, [4]dgprivezencev@mail.ru

In article the way of the description of the digital image using fractal models is considered. The described approach is based on treelike fractal representation, property of self-similarity describing display in the image.

Introduction

One of the most important aspects of use of methods of digital image processing is the solution of the tasks connected to recognition. The task of image identification consists in classification of images on the basis of certain requirements, and the images relating to one class of images, possess rather high level of closeness [1,2].

Creation of the description of the image on the basis of its representation with use of signs – is one of the most complex challenges in the course of creation of any system of recognition. Thus if within some mathematical models it was succeeded to formalize classification process, selection process of signs still remains procedure heuristic and dependent both from data domain, and from the developer.

Now, one of the developing directions in digital image processing is the fractal analysis of images. It is promoted by that fact that the majority of images somewhat can be considered as fractals or multifractals [2]. Therefore, the image possesses properties and characteristics of fractal objects, including invariance to the scale of reviewing and turn that it is expedient to use for development of new methods of fractal image processing.

In case of the description of the digital image by means of iterated functions system which are using in case of creation of fractals, formation of new signs of the images which are based on fractal submission of the image [3, 4] is obviously possible.

Tree-like representation of a fractal code of images

Image partition on rank units in the course of formation of a fractal code is carried out as follows. In the beginning images breaks into four parts – four rank units of the first level $R^1 = \{R_1^1, R_2^1, R_3^1, R_4^1\}$. Then for each rank unit R_i^1 search of the domain unit D_j and conversion w_j so that is carried out:

$$R_i^1 \approx w_j(D_j) \qquad (1)$$

In case of such approximation by the domain unit the approximation error ε takes place. If $\varepsilon \le \varepsilon_{\max}$ where $\varepsilon_{\max}$ the given maximum error of approximation, then is considered the current rank unit coded. Otherwise the rank unit breaks into four rank units of the second level $R_j^2 = \{R_1^2, R_2^2, R_3^2, R_4^2\} \in R_i^1$. Then process of approximation repeats again. Partition proceeds until the error of approximation won't meet the given conditions, or the maximum depth of partition won't be reached.

Thus, the image consists of four rank units of the first level, each rank unit of the first level consists of four rank units of the second level, etc.

Graphic representation of rank units of the test image is given in a figure 1.

[1] The reported study was partially supported by RFBR, research project No. 13-07-00825 a

Fig. 1. The source image and the rank units created for it.

Analytically rank units of the image can be provided in the form of a tree where a root element is the image, and leaves are rank units of which the image is built.

Let's provide the image f in the form of a graph $G = (V, \overline{U})$, where V - the list of rank units of the image, $\overline{U}$ - communications between rank units. Then process of formation of a tree of rank units can be painted the image as follows. There is a rank unit of zero level R^0 to which there corresponds all image, i.e. $v_1 = R^0$. It contains rank units of the first level $\{R_1^1, R_2^1, R_3^1, R_4^1\} \in R^0$, i.e.:

$$\begin{aligned} &v_2 = R_1^1, v_3 = R_2^1, v_4 = R_3^1, v_5 = R_4^1, \\ &\{\overline{u_1} = (v_1, v_2), \overline{u_2} = (v_1, v_3), \\ &\overline{u_3} = (v_1, v_4), \overline{u_4} = (v_1, v_5)\} \in \overline{U} \end{aligned} \qquad (2)$$

Having continued further formation of a fractal code, the finished tree of rank units of which the image is built turns out. The fragment of a tree of rank units is given in a figure 2.

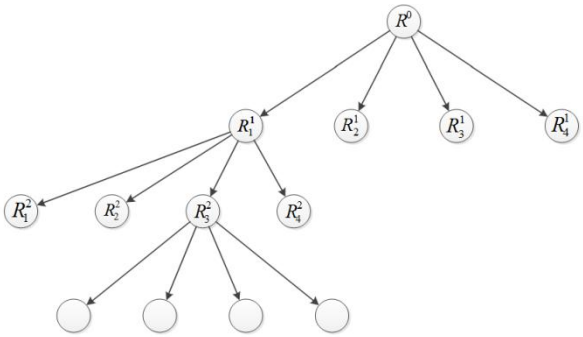


Fig. 2. Fragment of a tree of rank units.

Submission of halftone images

The operator of formation of a fractal code, in particular a fractal tree, we will call direct fractal conversio:

$$F(f, \mathbf{D}) = \Phi, \qquad (3)$$

where Φ - a received fractal code of the image by means of the operator. Besides, an additional statement parameter is the list of domain units of the image participating in formation of a code.

Except direct conversion there is an inverse fractal transformation which creates the image on a fractal code:

$$f' = F^*(\Phi, \mathbf{D}) = F^*(F(f, \mathbf{D})). \qquad (1)$$

For the correct restoration of the image on its fractal code, it is important that in conversions the identical list of domain units **D** was used.

In initial option the created fractal code is full. It means that each peak of a graph has one parent and four descendants, but length of a descending chain doesn't exceed depth kvadro-tree which is set at the initial stage of creation of a code.

If for image restoration by the operator (5) as parameter to use a full tree, then the image f' it will be most similar to the source image.

For the purposes of obtaining modified images incomplete trees of rank units are used. For formation of an incomplete tree the operator of a look is used:

$$\Phi^* = \Gamma(\Phi). \qquad (5)$$

The form of the operator Γ depends on required result. So, for example, having removed some branches of a fractal tree, it is possible to achieve obtaining the image on which sections for which remote branches were responsible, will be rounded or "hardened", and remaining sections accurate.

For kratnomasshtabny representation from a fractal code tree leaves are deleted. As a result the smoothed, "hardened" images turn out.

In fig. 3 the example of similar decomposition of the test image is given. The received set of images, allows to carry out more detail analysis of the source image. Transition to other scales allows to depart from fine and accidental details, it is better to reveal "internal" structure. In this regard, reviewing of the structural elements selected at different scales of submission of the image [8] is of interest.

Fig. 3. fractal submission of the test image.

Conclusion

Thus, the fractal code of the image can be provided in the form of graph G in which peaks of V are rank units R, and arcs – the relations of an enclosure of rank units. Thus the structure of a code remains invariable, this approach changes only the form of representation and mathematical apparatus of research.

On the basis of such representation approach to fractal submission of images which can be used for formation of the feature set, reflecting an inner pattern of the image as fine details are discarded is based. Further on the basis of the offered description algorithms of separation of circuits, segmentations of images, skeletization and др can be constructed

References

1. Methods of Computer Image Processing, Ed. by A. V. Soifer (Fizmatlit, Moscow, 2003) [in Russian].
2. Newest Methods of Digital Image Processing, Ed. by A. A. Potapov (Fizmatlit, Moscow, 2008) [in Russian].
3. Fractal Image Compression: theory and application, Ed. by Yaval Fisher (Springer, New York, 1995).
4. Fractal and Wavelet Image Compression Techniques, Ed. by S. Welstead (SpiePress, Moscow, 1999) [in Russian].
5. A. L. Zhiznyakov and D. G. Privezentsev, "Fractal digital image model," Algorithms, Meth. Syst. Data Processing, No. 15, 147--152 (2010).
6. A. L. Zhiznyakov and D. G. Privezentsev, "Distribution of self-similarity of the digital image proceedings of the higher education institutions," Instrumentation, No. 2, 37--32 (2012).
7. A. L. Zhiznyakov and D. G. Privezentsev, "Allocation of local self-similarity features of digital images," Meth. Devices Transl. Processing Inf., No. 12, 54--58 (2010).
8. A. L. Zhiznyakov, D. G. Privezentsev, and A. A. Fomin, "Image classification based on local features of self-similarity," Polzunovsk. Vestn., No. 3-1, 12--14 (2011).

METHOD FOR ADAPTIVE MULTIDIMENSIONAL MEASUREMENTS PROCESSING BASED ON IGIS TECHNOLOGIES

N. Zhukova[1]

[1] **Saint-Petersburg Electrotechnical University, Saint-Petersburg, Prof. Popova street, 5, 197376, 234-14-27, nazhukova@mail.ru**

In the paper the problem of providing information space and business logic adapted for each end user's needs is regarded from the point of view of measurements time series processing and analyses. A general method for measurements processing and analyses based on harmonization, integration and fusion concept implemented using means provided by IGIS technologies is proposed. Results of the proposed method application to ocean data processing are discussed.

Introduction

Modern information systems, in particular decision making support systems (DMSS), have become one of the key elements that provide solutions of end-users tasks in various spheres of human activities. Well founded solutions of highly complicated tasks such as situation retrieval, assessment, awareness, prediction and management are to be supported. The main reason of the complexity of end users tasks is a huge amount of heterogeneous data and information received from multiple distributed sources that is to be gathered, processed, analyzed, explored and managed in real time. Special attention is to be paid to measurements processing, as measurements are not stationary time series that contain noise, outliers and gaps, they are not coordinated in time and space. Processing of measurements time series significantly depends on behavior of time series and the context of their processing.

It is reasonable to build DMSS on the base of intelligent geographical information systems (IGIS). IGIS is a distributed multilevel heterogeneous system that due to integrated means of artificial intelligence and adaptive process management is able to provide solutions of end users problems in conditions of varying environment state, amount, quality and type of information, accessible information sources, user requirements and implemented processing tools [1]. Intelligent GIS have a set of highly appreciated features such as support of complex geospatial processes, calculation of complicated specialized mathematical tasks, consumption of information about both technical and environmental objects and etc. High efficiency of IGIS technologies are demonstrated by three real work systems, that were developed both for military and civilian purposes - Ontomap, V1 and V2 [2] and are now widely used for supporting of decision-making by the naval personnel at different levels of C2 systems.

However, tendencies of the exponential increase of number of the solved tasks and volume of data that is to be processed that have been observed for several years already are the main reason of constantly increasing complexity of information systems. Along with that many new users start working with information systems. Users are oriented on solving a set of specialized tasks and most of them are not experts in data processing, especially measurements processing. So for each user it is necessary to provide an adaptive information space and business logic that will meet the user's requirements and correspond to the user's skills. In the paper the main attention is focused on the problem of automatic adaptive multidimensional measurements processing that is an important element of adaptive business logic organization.

Concept of adaptive measurements processing based on IGIS technologies

To provide adaptive measurements processing in DMSS systems based on IGIS technologies that will be easily adapted to end users tasks it is necessary to define a general concept and a set of methods and tools that are sufficient to process, analyze and manage data in various contexts in the defined subject domain.

The concept is to be defined taking into account the following principal requirements: i) support of extended business logic - business logic that assumes application of complicated mathematical-based methods is needed to solve modern user's tasks with the desired quality and in the appropriate time; ii) support of processing in automatic mode - users tasks at the level of measurements processing must be solved in automatic mode as most users have not enough time and qualification to manage the process; iii) support of DMSS systems reorganization and reconfiguration on the fly - the set of tasks solved by the users as well as contexts and conditions in which tasks are solved are permanently changing; iv) low requirements to computational resources - various ways of system implementation must be supported, special attention is to be paid to decisions based on mobile technologies.

Main features of measurements processing in modern DMSS are the following: i) all types of measurements of both technical and environmental objects parameters must be processed; ii) measurements of different quality that are not coordinated in time and space, as well as uncertain and incomplete measurements must be analyzed; iii) to provide required quality of measurements processing results, processing must be organized taking into account all available knowledge to the system; iv) several types of measurements processing must be supported – real-time, operational and delayed mode processing, as well as measurements based information retrieval.

Modern state of results of researchers made by Jeffrey D. Scargle (NASA), Eamonn Keogh (UC Riverside) and many others in domain of data and measurements processing can be characterized in the following way: i) a large amount of separate mathematical based methods and algorithms are developed that are oriented on processing measurements of the defined type and behavior; ii) a part of the existing models and methods need to be improved as they are not able to deal with information that can increase the accuracy of the provided solutions; iii) methods and algorithms are implemented in different libraries and toolboxes.

The concept of adaptive measurements processing based on IGIS technologies supposes building processes for measurements processing that are able to solve complicated end users tasks in dynamics as a combination of existing methods and algorithms according to the observable context and using all available knowledge. Information about the context as well as instruments and tools for knowledge based organization and management of processes are provided by IGIS. Existing methods and algorithms are used for measurements processing due to the IGIS universal mechanism for interaction with external systems, tools and libraries. Improvement of existing methods and tools is organized using means of artificial intelligence that are a part of IGIS and a unified information model (UIM) [3] that contains complete actual information about subject area.

Organization of adaptive measurements processing in IGIS assumes that general processes for data processing and analyses are described a priori. These processes are detailed when end users tasks are solved using all types of information available to the system at the moment. Main features of adaptive process management approach are: i) a set of defined general processes for solving both end users and system tasks describe main principles of the system behavior; they are considered as a part of information model; ii) processes have multilevel structure; each level corresponds to a certain abstraction level of process description; iv) detailing of processes is organized using a priori defined rules, results of exploration analyses and knowledge about data processing. The detailed description of the considered concept is given in [4].

One of the most complicated tasks in the concept implementation is to define general processes for solving various tasks. As a rule general processes are defined by experts.

Algorithm for multidimensional measurements processing

A general algorithm for the subject domain of adaptive measurements processing was developed. It contains a sequence of steps that can be executed consistently or in parallel and possible ways for their detailing. The algorithm assumes that both universal and specialized methods and tools can be executed; several contexts are taken into account – geospatial, time related and historical; steps of the general algorithm are defined according to the concept of data harmonization, integration and fusion.

Results are formally described using a developed hierarchy of models [5] and represented as a part of the UIM of subject domain. Below each of the three steps of multidimensional measurements processing are considered. Measurements harmonization algorithm assumes execution of the stages represented in Fig.1. Algorithm for measurements time series integration is given in Fig.2, for time series fusion in Fig.3.

Input data. Initial data in the form of binary stream
Output data. Modified UIM of subject domain
Algorithm description.
A. Binary data processing
Step A1 Binary data structure identification
Step A2 Binary data structure reveling (*not obligatory*)
Step A3 Evaluation and improvement of data structure description
B Measurements time series processing
Step B1 Conversion of binary streams to time series
Step B2 Input time series transformation
convert time series to unified formats
execute preliminary processing of time series
estimate time series quality and reliability
Step C. Description of measurements in terms of subject domain
identify the list of entities
identify relations between entities

Fig. 1. Algorithm for measurements harmonization.

Input data. Measurements time series provided by information model of subject domain
Output data. Modified UIM of subject domain
Algorithm description.
A. Specialized measurements processing
Step A1. Execution of the set of standard tests defined for data sources
Step A2. Execution of calculation procedures for elimination of known systematic errors
B. Statistical measurements processing
Step B1 Basic statistical measurements processing
identify time series types
execute the set of context dependent tests
calculate procedures defined for data sources
exclude duplicated values
Step B2 Extended statistical measurements processing
execute joint analyses of time series
estimate time series on the base of results of historical data analyses
C Task-oriented measurements processing
Step C1 Calculation of derived parameters
Step C2 Solution of the defined set of applied tasks

Fig. 2. Algorithm for measurements integration.

The desired list of algorithms for measurements harmonization includes following methods and algorithms: structure revealing, identification and verification methods based on calculation of correlation functions or graph-oriented representation of structures; algorithms for time series preprocessing, including algorithms for removing noise, outliers and filling gaps as well as algorithms for preliminary measurements quality estimation that depend on types of processed time series; a unified algorithm for data transformation to the UIM.

Algorithm for measurements integration is based on the combination of the following algorithms: a group of specialized tests and procedures for data sources defined in subject domain model; a set of standard context dependent methods and algorithms applicable for all data sources; a set of basic and extended statistical methods for time series analyses; a set of algorithms required for solving end user's tasks.

Input data. Measurements time series provided by information model of subject domain
Output data. Knowledge about processed measurements
Algorithm description.
A. Fusion of measurements
Step A1. Building measurements time series formalized descriptions
execute segmentation of measurements time series
calculate statistic characteristics of segments
execute cluster analyses of segments
define segment classes
Step A2. Identification of dependencies in time series
search for association dependencies between segment classes
search for temporal patterns in time series behavior
B. Fusion of measurements and data of all other types
Step B1 Reveal and identify complicated context dependent groups of data
reduce feature space
identify groups of data in the defined context
classify identified groups of data in the defined context
identify external factors that influence the groups structure
Step B2 Identify behavior of various groups of data
identify logical dependencies between groups of data for different contexts
identify temporal dependencies between groups of data for different contexts
compare identified dependencies in different contexts
C Exploration of historical data and knowledge retrieval
Step C1 Building local statistical based models for defined contexts
Step C2 Building complex hierarchical statistical based model
D Building regular grids of objects parameters using operational and historical measurements

Fig. 3. Algorithm for measurements fusion.

Measurements fusion algorithm is the most complicated but it is universal and independent from the input data suppliers and the set of the

solved tasks. It is based on a wide range of various methods and algorithms that are a complicated combination of traditional algorithms, statistical and data mining algorithms. Adaptation of algorithms to input measurements and context of their processing is achieved due to: i) usage of statistical methods for proper preparation of input measurements for further processing; ii) definition of algorithms optimal parameters using results of historical data processing; iii) definition of criteria for results estimation on the base of statistical algorithms; iv) retrieval of additional exploratory information about input measurements. Adaptation of processing algorithms provides possibility to apply algorithms in automatic mode, as parameters are defined automatically; to increase efficiency of algorithms due to using preprocessed data and additional information about data; to reapply algorithms with different parameters if estimations of obtained results don't satisfy users requirements.

To support algorithms adaptation it is necessary to define formalized description of each algorithm that contains: a name of the algorithm, a short formalized description, a detailed not formalized description, a description of input and output data, input parameters of the algorithm, a set of criteria estimating if the algorithm can be applied to the defined input data, a set of algorithm results estimation and history of algorithm application. Algorithms and their descriptions are defined in the UIM. All algorithms are grouped and taxonomy of algorithms is build according to applied computational methods. Methods, algorithms and procedures used at different steps of harmonization, integration and fusion algorithms are described in [6].

Conclusion

The proposed method for adaptive measurements processing was implemented in Decision Making Support System for Arctic Exploration, Monitoring and Governance for ocean data processing. Initial oceanographic data included temperature and salinity profiles received from more than 135000 various stations at the period from 1870 up to nowadays. The task was to provide operative information about water area to systems of hydro acoustic fields calculation. On the base of initial data regular grids were built using the proposed method in automatic mode and with the desired accuracy of the results. The results of system operation are described in [6].

References

1. V. Popovich, S. Potapychev, A. Pankin, S. Staida, M.N. Voronin, "Intellectual GIS in Systems of Monitoring", Tr. SPIIRAN, 3:1, 2006 http://www.oogis.ru
2. S. Potapychev, A. Pankin Universal Data Model for Intelligent GIS // Information Fusion and Geographic Information Systems / Eds. by V. Popovich et al. Springer-Verlag, Berlin Heidelberg, 2007
3. A. Pankin, A. Vitol, N. Zhukova Adaptive Multidimensional Measurements Processing Using IGIS Technologies // Information Fusion and Geographic Information Systems / Eds. by V. Popovich et al. Springer-Verlag, Berlin Heidelberg, 2013
4. A. Pankin, N. Zhukova, A. Vitol Model for knowledge representation of multidimensional measurements processing results in the environment of intelligent GIS. Conceptual Structures for STEM Research and Education // Proceedings of 20th International Conference on Conceptual Structures, 2013
5. A. Vitol, A. Deripaska, N. Zhukova, I.Sokolov Technology of Adaptive Measurements Processing, SPbETU "LETI", Saint-Petersburg, 2012
6. N. Zhukova, O. Smirnova Atmosphere and Ocean Data Processing in Decision Making Support System for Arctic Exploration// Information Fusion and Geographic Information Systems / Eds. by V. Popovich et al. Springer-Verlag Berlin Heidelberg, 2013

AUTHOR INDEX

For design of the cover the following space images of Earth are used:

Volga region. Terra satellite (MODIS), 2007.
Reception and processing: SSAU, Volga Center of Space Geoinformatics.

Samarskaya Luka. Landsat-7 satellite. 2009.
From an open source of Global Land Cover Facility (www.landcover.org)

НАУЧНОЕ ИЗДАНИЕ

11th International Conference
«PATTERN RECOGNITION and IMAGE ANALYSIS: NEW INFORMATION TECHNOLOGIES»
PRIA-11-2013

September 23-28, 2013

Samara, the Russian Federation

CONFERENCE PROCEEDINGS
Volume I

Оформление и верстка Н.М.Боргест, Д.А.Урывская, В.А.Федосеев, В.Е.Быков, О.Ю.Чичёва, С.А.Шагеев
Технический редактор В.В. Сергеев

Подписано в печать 25.08.2013. Формат издания 60x84 1/16.
Печать офсетная. Тираж 300 экз.

ИСОИ РАН
443001, Самара, Молодогвардейская ул., д. 151.

Отпечатано с готового оригинал-макета в типографии
ООО «Предприятие «Новая техника»
443010, г.Самара, ул.Фрунзе, 145
тел. (846) 332-67-81