Descriptive Approach to Medical Image Mining. An Algorithmic Scheme for Analysis of Cytological Specimens

I. B. Gurevich[a], V. V. Yashina[a], I. V. Koryabkina[a], H. Niemann[b], and O. Salvetti[c]

[a] *Dorodnicyn Computing Center, Russian Academy of Sciences, ul. Vavilova 40, Moscow, 119333 Russia*
e-mail: igourevi@ccas.ru, werayashina@mail.ru, irinakoryabkina@bk.ru
[b] *University of Erlangen-Nürnberg, Lehrstuhl f@ur Informatik, Martensstr., Erlangen, Germany*
e-mail: niemann@informatik.uni-erlangen.de
[c] *Institute of Information Science and Technologies, CNR, 1, Via G. Moruzzi, Pisa, 56124 Italy*
e-mail: Ovidio.Salvetti@isti.cnr.it

Abstract—The present paper is devoted the development and formal representation of a descriptive model for an information technology to automate the morphological analysis of cytologic preparations (a tumor of the lymphatic system). The theoretical basis of the model is a descriptive approach to image analysis and understanding and its main mathematical tools. Practical application of the algebraic tools of the descriptive approach is demonstrated, and the algorithmic scheme of the technology is described in the language of descriptive image algebras.

DOI: 10.1134/S1054661808040032

1. INTRODUCTION

This paper is devoted the development and formal description of a descriptive model of an information technology for automated morphological analysis of cytologic preparations for patients with tumors of the lymphatic system [4].

Analysis of the morphology of blood cells is a fundamental problem in diagnosing tumors of the blood system (hemoblastoses) and researching the patterns of tumor progression. Hemoblastoses are recognized by informal, nonquantitative details of the morphology of tumor cells (mainly their nuclei). Such an analysis essentially depends on the skill level of the expert and his experience.

In connection with the necessity of increasing the effectiveness and exactness of diagnosing hemoblastoses by morphological analysis of blood cells, the task of creating an automated diagnostic system using the morphological analysis of cellular preparations is urgent.

In the present work, we describe the technology of automated morphological image analysis of nuclei of lymphoid cells of diseased hemoblastoses; this technology can serve as the basis for such a system.

The main tasks of the paper are, first, to build the structure of an information technology, and, second, to describe it using the mathematical tools of a descriptive approach to image analysis and understanding (DAIAU) [3]. The developed mathematical model should guarantee uniform representation of the algorithm for solving the task essential for programming and useful for comparing various information technologies created for solving one and the same task.

The theoretical basis of this mathematical model is the DAIAU [3] and its main tools—descriptive image algebras (DIAs) [5, 7], descriptive image models (DIMs) [13], and generating descriptive trees (GDTs) [6].

In the section DESCRIPTIVE APPROACH TO IMAGE ANALYSIS AND UNDERSTANDING, we introduce the main theoretical information necessary for understanding the suggested methods and resources for describing the information technology. This section contains five subsections. In the subsection *Algebrization of Image Analysis*, the history of algebraization is briefly described. The subsection *Descriptive Image Algebras* contains a short description of the mathematical language used in describing the scheme of the information technology. In the subsection *Descriptive Images Models*, we introduce the main definitions of image models whose construction from input data is necessary in order to apply them for recognizing algorithms. The subsection *Generating Descriptive Trees* contains a description of the concept of constructing multimodel representations by means of specialized trees. In the subsection *Descriptive Model of a Pattern Recognition Task*, a simplified descriptive model of the pattern recognition task is described based on multimodel representations of images.

The section DESCRIPTIVE MODEL OF MORPHOLOGICAL ANALYSIS OF CELL NUCLEI OF THE LYMPHATIC SYSTEM contains an interpretation of the main theoretical tools of DAIAU for their usage in solving the task of a morphological analysis of

Received June 26, 2008

nuclei. This section also contains five subsections. In the subsection *Statement of the Task*, input data are described and a specific task statement is formulated. In the subsection *Descriptive Image Algebras,* operations and operands (and their semantic functions) of DIAs and descriptive image groups (DIGs) necessary for constructing the algebraic model for morphological analysis of cell nuclei of the lymphatic system are introduced. In the subsections *Descriptive Image Models* and *Generating Descriptive Trees*, DIMs and GDTs are introduced, respectively, necessary for constructing the algebraic model. In the subsection *Algorithmic Scheme of the Morphological Analysis of Cell Nuclei of the Lymphatic System*, we describe the descriptive model of the information technology for automating morphological analysis of cytologic preparations for patients with lymphatic tumors. The described information technology has been tested on preparations of patients with malignant and benign tumors. In the section, *Algorithmic Scheme of the Morphological Analysis of Cell Nuclei of the Lymphatic System*, the results of testing of the described technology are discussed.

The basic components of the technology are described with DIA and presented in the form of an algorithmic scheme DIA. The latter guarantees standard representation of technologies for making intellectual decisions.

2. DESCRIPTIVE APPROACH TO IMAGE ANALYSIS AND UNDERSTANDING

2.1. Algebraization of Image Analysis

The main distinctive singularity of the given paper is that the tools of the descriptive approach are applied to the description of algorithms used for solving the applied task.

Algebraization of pattern recognition and image analysis has attracted and continues to attract the attention of many researchers. Appreciable attempts to create a formal apparatus ensuring a unified and compact representation for procedures of image processing and image analysis were inspired by practical requirements for effective implementation of algorithmic tools to process and analyze images on computers with specialized architectures, in particular, cellular and parallel.

The idea of constructing a unified language for concepts and operations used in image processing appeared for the first time in works by Unger [21], who suggested to parallelize algorithms for processing and image analysis on computers with cellular architecture.

Mathematical morphology, developed by G. Materon and Z. Serra [16], became a starting point for a new mathematical wave in handling and image analysis. Serra and Sternberg [18, 19] 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 [15] that it was precisely Sternberg who introduced the term "image algebra" [17] in the current standard sense. (We note that U. Grenander used this concept in the 1970s; however, he was talking about another algebraic construction [2]). Within the limits of this direction, an array of works continues to be written, devoted to the development of specialized algebraic constructions implementing or improving upon methods of mathematical morphology.

The concept of an algebraization has been successfully implemented in the case of classical pattern recognition [25]. Passage from the algebra of pattern recognition algorithms to an algebra of image recognition algorithms requires a choice, first, of algorithms used as elements of algebra, and second, of algebraic representations of images that make it possible to formalize the task of choosing descriptors. It is expedient to select representations taking into account the possibility of combining the initial information and algorithms of different types. For the first time, the idea of a combination of qualifiers with optimization of their operation by algebraic correction was suggested and justified by Yu.I. Zhuravlev [25]. The complex of mathematical methods related to synthesis and research of such qualifiers is known under the common title "Algebraic Approach to Tasks of Recognition, Classification, and Prediction." In the English-Language literature for the designation of qualifiers, the term Multiple Classifiers [20] is used. Recently, quite interesting results have been achieved in the field of theoretical-informational analysis of combined qualifiers [13], developments of specific strategies for merging algorithms [14], and usage of methods of code theory in tomography [24].

The algebraization of pattern recognition and image analysis, thus, has a long history: Unger, Sternberg, Serra, Zhuravlev, Grenander, Ritter, and others, but according to our information, algebraic methods as a whole and the descriptive approach in particular have never been applied to solving the task of analysis and recognition of medical images.

The main objective of the theoretical apparatus of image models [3, 13] is the structuring of various methods, operations, and representations used in analysis and image processing. The ultimate goal of image models is automation of knowledge extraction from images: (a) an automatic choice of methods and algorithms for understanding, estimations, and pattern recognition; (b) an automatic testing of quality and applicability of input data for solving the task of pattern recognition.

The main tools are image models, DIAs [5, 7], DIMs [13], and GDTs [6].

DIA is a mathematical language for the description, comparison, and standardization of analysis algorithms, processing, and pattern recognition. DIA makes it possible to achieve flexibility and standardization in the creation and application of algorithmic schemes of extracting information from images.

Tasks, objects, and conversions of "Image Mining" are set by hierarchical structures built by means of application of DIA operations to a set of primitive tasks, primitives of the image, and the main basis conversions.

In such an approach, there exists the possibility of changing the methods for solving a subtask using operations of image analysis as elements of DIA, and thus to save the entire scheme of the technology for the extraction of information from images.

DIMs are mathematical objects designed for representation of information extracted from an image and its legend, in the form adopted by the recognizing algorithm. The image legend can contain information on a subject domain, on a scene, on the illumination intensity, on the means by which it is obtained, on sensors and on the imaging system, about the position of the observer used for registration of the image, and other additional semantic and contextual information. DIM has been introduced to formalize the scheme of pattern recognition [4, 8, 9] in such a manner that the step of adapting images to a form convenient for recognition is one in which the source images are transferred to the corresponding model of the source image leading to the recognition algorithm.

One of concepts of DAIAU is that of operational characterization of an image, that is, construction of the formal description of the image model of the image, which is considered as implementation in the image of a certain system of conversions by which an image can be constructed from an image and other objects of a simpler nature, i.e., nonproductive elements and selected objects in the image. With such an approach, it is suggested not only to formalize image models by introducing the main classes of descriptive image models, but also to introduce hierarchies of formal image descriptions and multilevel multidimensional models allowing, in the course of recognition, to select and change the required degree of detail of the description of the object of recognition. The given hierarchies of formal descriptions of images and multilevel multidimensional models are described by means of special classes of trees (GDTs) which generate multimodel and multidimensional representations of images.

In the development of image models, a descriptive model of the task of recognition is constructed, which would schematically describe the pattern recognition task. On the basis of the descriptive model of the recognition task, the mathematical settings of the pattern recognition task can be formulated, which are mathematical bases for automating the extraction of knowledge from images.

Zhuravlev [28] proposed the first mathematical statement of the task of pattern recognition. He also introduced the determined pattern recognition algorithm. Gurevich [3] introduced the formal description of image analysis—pattern recognition task. Later, he set the task of pattern recognition based on equivalence classes of images [10].

2.2. Descriptive Image Algebras

In [5, 7], the main results obtained in the course of researching the new class of algebras are described. Definitions DIA, base DIA, and DIA with one ring were introduced; a classification was obtained, expressing how the authors imagined the hierarchy of modern algebras and the place of DIA in this hierarchy; a method of checking the properties of operands and operations for constructing a DIA with one ring was demonstrated (the given method was obtained in studying the specificity of new algebras of images); operations of the Ritter algebra for images [17] were selected for constructing DIAs (the given task was solved in a study of the operations of a DIA with one ring); the means are demonstrated of constructing DIMs using special classes of DIA (it is necessary to formalize algorithmic schemes of image analysis); necessary and sufficient conditions for generation of DIAs with one ring have been identified (the given conditions were obtained in studying the main operands and operations of DIAs).

Let us recall the definitions of a DIA:

Definition 1 [7]. The algebra is called a *descriptive image algebra* if its operands consist of image models (including the fact that both an image and a group of values and characteristics related to the image can be selected as the model) or operations over images, or both simultaneously.

Definition 2 [7]. A descriptive image algebra is called a *base descriptive image algebra* if its operands consist of only image models or only operations over images.

Definition 3 [7]. Ring U, which is a finite-dimensional vector space over some field P, is a *descriptive image algebra with one ring* if its operands are either image models, or operations over images.

A DIA with one ring, thus, should satisfy the properties of a classical algebra. A DIA with one ring is a base DIA since it contains one ring of elements of one nature, that is, either a ring of image models or a ring of operations over images.

2.3. Descriptive Image Models

In [5, 13], the main results obtained during studies of DIMs are described. The main types of image models have been defined. It has been shown what types of image models are generated by the main types of DIAs with one ring. We will recall some definitions and statements of image models.

Any image I can be unambiguously put into correspondence with the totality of sets ($\{\tilde{I}_0\}$, $\{\tilde{O}\}$, $\{\tilde{M}\}$), where $\{\tilde{I}_0\}$ is the set of initial information, $\{\tilde{O}\}$ is the set of transformations applicable to the set of initial information, and $\{\tilde{M}\}$ is the set of resules of applying transformations to the initial information.

In image models, four classes of DIMs have been introduced. We need three classes of image models for the description of morphological image analysis of cytologic preparations: image models, procedural image models, and parametrical image models. We will recall the definitions of the given three classes of image models.

Definition 4. *An I-model* of an image is any element I' of a set of $\{\tilde{I}'\}$ of image realizations.

Definition 5. *An image representation* $\mathfrak{R}(I$ is a formal scheme intended for deriving the standardized formal description of surfaces, point configurations, forms making up the image, and relationships between them.

Definition 6. *An image model* $M(I)$ is a formal description constructed by implementing of an image representation $\mathfrak{R}(I)$.

Definition 7. *A realization of an image representation* is the application of the representation to realizations of the original image with particular parameter values specified for the transformations involved in the representation.

Definition 8. *The T-representation* $\mathfrak{R}_T(\tilde{\eta}, \tilde{\mu})$ of the image I is the formal scheme designed to obtain the standardized formal description of the image and constructed using context and semantic information $\{B\} \subset \{\tilde{B}\}$, procedural transformations $\{O_T(\tilde{\eta})\} \subset \{\tilde{O}_T\}$ and structuring elements $\{S(\tilde{\mu})\} \subset \{\tilde{S}\}$ ($\tilde{\eta}$, $\tilde{\mu}$ are the parameters of procedural transformations and structuring elements, respectively).

Definition 9. *A* realization *of the T-representation* $\mathfrak{R}_T(\tilde{\eta}, \tilde{\mu})$ of the image I is the process of applying the representation $\mathfrak{R}_T(\tilde{\eta}, \tilde{\mu})$ with chosen values ($\tilde{\eta} = \tilde{\eta}_0$, $\tilde{\mu} = \tilde{\mu}_0$) of parameters of transformations involved in the representation to implementations of the initial image $\{I'\} \subset \{\tilde{I}'\}$.

Definition 10. *A P-representation* $\mathfrak{R}_P(\tilde{\eta}, \tilde{\mu})$ of the image I is a formal sceme designed to obtain the standardized formal description of the image and constructed, using the context and semantic information $\{B\} \subset \{\tilde{B}\}$, parametric transformations $\{O_P(\tilde{\eta})\} \subset \{\tilde{O}_P\}$ and structuring elements $\{S(\tilde{\mu})\} \subset \{\tilde{S}\}$ ($\tilde{\eta}$, $\tilde{\mu}$ are the parameters of parametric transformations and structuring elements, respectively).

Definition 11. *A realization of the P-representation* $\mathfrak{R}_P(\tilde{\eta}, \tilde{\mu})$ of an image I is the process of applying the representation $\mathfrak{R}_P(\tilde{\eta}, \tilde{\mu})$ with chosen values ($\tilde{\eta} = \tilde{\eta}_0$, $\tilde{\mu} = \tilde{\mu}_0$) of parameters of transformations involved in the representation to implementations of the initial image $\{I'\} \subset \{\tilde{I}'\}$.

Statement 1. The setting of the value of the parameters of parametrical (procedural) conversions and structuring elements ensures generation of a set of *P-models (G-models) of the image* by any *P*-representation (G-representation).

Statement 2. Any *T*-model of image $M_T \in \{\tilde{M}_T\}$ generates some implementation of image I', that is, an *I*-model of image $M_I \equiv I' \in \{\tilde{M}_T\} \equiv \{\tilde{I}'\}$.

2.4. Generating Descriptive Trees

The introduction of axiomatics of DAIAU [13] and definition of three classes of DIM has led to the introduction of a new mathematical object for structuring representations of images and generation of image models.

Three types of appropriate conversions generating rules and a source image are necessary for constructing the three classes of representations of images (procedural, parametrical, and generating representations of images). The source image is described by means of a set of its implementations and by means of context-sensitive and semantic information.

According to the introduced axiomatics and definitions of various classes of representations of images, in this way, for merging and combination of various properties of image models, it is necessary to introduce the following hierarchies: the hierarchy of possible implementations of images; the hierarchies of semantic and context-sensitive information in images; the hierarchies of parametrical, procedural, and generating conversions; and hierarchies of generating rules. It is suggested to implement such structures in the form of special trees.

Specialization of the concept of a tree on the whole is related to specialization of nodes of a tree. As nodes we will select objects, operations, or rules of image analysis tasks used to construct different image models. Such nodes are called GDT descriptors. The definitions of parent, calculated, fixed, objective, and abstract GDT descriptors have been introduced, but will not be dwelt on in this work.

The concept of GDT first arose in [8]. In [8], in the definition of GDT, the concept of a descriptor was not explicitly introduced. In the given paper, this concept is treated and the statement that each descriptor is an image model has been refuted.

Definition 12. *The generating descriptive tree (GDT)* is the structure intended for classification and automated generation of image models and it possesses the following properties:

(1) GDT descriptors are GDT nodes;

(2) Every GDT combines the descriptors of one type; that is, GDTs represent the same type of properties of an image;

(3) Each GDT element can be united with another element to generate new partial multiaspect image models;

(4) Descriptors are linked among themselves by parent–daughter relationships;

(5) Each descriptor has a relationship with a unique parent descriptor and can have some links with derived descriptors. If the descriptor has no parent, it is called a radical GDT. If the descriptor has no derived descriptors, it is called a leaf.

Note that parametrical GDTs are GDTs intended for classification and automation of the generation of parametrical image models. A parametrical GDT, thus, contains GDT descriptors describing the properties of parametrical conversions, leading to an evaluation of features of images. A procedural GDT is a GDT intended for classification and automation of the generation of procedural image models. A procedural GDT, thus, contains the GDT descriptors describing the properties of procedural conversions.

2.5. Descriptive Model of a Pattern Recognition Task

Let us introduce, on the basis of setting a pattern recognition task, a simplified model of a pattern recognition task.

Let $\{I_i\}_{i=1,\ldots,n}$ be a set of source images, and let $\{K_j\}_{j=1,\ldots,m}$ be a set of classes (in some cases, sets of equivalence classes of images). Algorithm A solves the pattern recognition task if it puts in correspondence to a set of source images a set of predicates $\{P_j(I_i)\}_{m\times n}$, where predicate $P_j(I_i) = \alpha_{ij}$ takes following value: $\alpha_{ij} = 1$ if image I_i belongs to class K_j; $\alpha_{ij} = 0$ if image I_i does not belong to class K_j; $\alpha_{ij} = \Delta$ if algorithm A cannot set the belonging of image I_i to class K_j. It is possible to say, in such a manner, that the pattern recognition task is described by the following scheme:

$$\{I_i\}_{i=1,\ldots,n} \xrightarrow{A} \{P_j(I_i)\}_{m\times n}. \quad (1)$$

It is obvious that not every recognizing algorithm correctly solves the pattern recognition task—this has led to the idea of a combination of several various recognition algorithms. Zhuravlev [28] was the first to suggest and justify the idea of combining qualifiers with optimization of their work by algebraic correction. In the English-language literature, such qualifiers are termed Multiple Classifiers [22].

Taking into account the possibility of combining recognition algorithms, the pattern recognition task is described by the following scheme:

$$\{I_i\}_{i=1,\ldots,n} \xrightarrow{\{A_k\}_{k=1,\ldots,l}} \{P_j(I_i)\}_{m\times n}. \quad (2)$$

Multialgorithmic qualifiers are applied to various images of a learning set for customization of internal parameters. Trained multialgorithmic qualifiers are applied to a recognized set of images, and they generate a set of predicates taking a value from the set (0, 1, recognition failure) for each class of recognition and each recognized image.

Let us consider the fact that image models are always constructed during the solution of the pattern recognition task (image reduction to a recognizable form is a necessary step of any recognition task). Then, scheme (1) is transformed to the following:

$$\{I_i\}_{i=1,\ldots,n} \xrightarrow{M} \{M(I_i)\}_{i=1,\ldots,n} \xrightarrow{A} \{P_j(I_i)\}_{m\times n}, \quad (3)$$

where M is some image model.

Let us expand the idea of multiplicity from algorithms (multialgorithmic qualifiers) on a model: instead of one model, we will build multimodel representations of images.

Normally, various formal representations of images include only part of the data that the image carries in itself: the image model does not reflect all the specifics of images. For the description of images via its multiaspect characteristics and features, we offer the following approach for generating multimodel representations of images:

(1) a GDT set is constructed;

(2) every GDT generates one or several formal representations of the images most precisely reflecting necessary for solution of a task in view of recognition of properties of images;

(3) all these representations are combined in one or several formal multimodel representations, which are used for all source images or for some classes of source images in the given recognition task.

Scheme (3), on the basis of multiplicity of models, will be transformed to the following scheme, where at the first step, some combination of models $(M_1, \ldots, M_s)$ instead of a specific model M is used:

$$\{I_i\}_{i=1,\ldots,n} \xrightarrow{\{M_y\}_{y=1,\ldots,s}} \{M_y(I_i)\}_{s\times n} \xrightarrow{A} \{P_j(I_i)\}_{m\times n}. \quad (4)$$

The combination of schemes (2) and (4) leads to the following scheme:

$$\{I_i\}_{i=1,\ldots,n} \xrightarrow{\{M_y\}_{y=1,\ldots,s}} \{M_y(I_i)\}_{s\times n} \xrightarrow{\{A_k\}_{k=1,\ldots,l}} \{P_j(I_i)\}_{m\times n}. \quad (5)$$

As the given scheme does not consider the learning step of the recognizing algorithm, the following three-stage modification is introduced: (1) construction of an image representation; (2) learning—customization of the selected recognition algorithms for learning a set of images; (3) recognition—consecutive application of pattern recognition algorithms to each recognized image. Stage 1 of the construction of a multimodel representation is divided into two substeps: (a) construction of multimodel representations of a learning set; (b) construction of multimodel representations of the recognized set. Depending on the selected recognition algorithms, stage 1b is fulfilled together with stage 1a (in the case of identical multimodel representations being built for the learning and recognizing set), or after the end of stage 1a (then, at stage 1a, the choice of multimodel representations being built for the recognizing set), or after the end of stage 2. In the latter case, the result of the performance of the recognizing algorithm can influence the choice of multimodel representation of the recognized image.

In the case of recognition with a learning set, the recognition scheme will assume the following form (Scheme 6):

$$\begin{array}{l} \{I_i\}_{i=1,\ldots,\left[\frac{n}{2}\right]} \xrightarrow[1(a)]{\{M_y^1\}_{y=1,\ldots,s_1}} \{M_y^1(I_i)\}_{s_1\times\left[\frac{n}{2}\right]} \xrightarrow[2]{\{A_k(p)\}_{k=1,\ldots,l}} \{P_g(p)(I_i)\}_{r\times\left[\frac{n}{2}\right]} \xrightarrow{INFO} p_0 \\ \{I_i\}_{i=\left[\frac{n}{2}\right]+1,\ldots,n} \xrightarrow[1(b)]{\{M_y^2\}_{y=1,\ldots,s_2}} \{M_y^2(I_i)\}_{s_2\times\left(n-\left[\frac{n}{2}\right]\right)} \xrightarrow[3]{\{A_k(p_0)\}_{k=1,\ldots,l}} \{P_g(I_i)\}_{r\times\left(n-\left[\frac{n}{2}\right]\right)} \end{array} \tag{6}$$

3. DESCRIPTIVE MODEL OF THE MORPHOLOGICAL ANALYSIS OF CELL NUCLEI LYMPHATIC SYSTEM

3.1. Setting the Task

Input data. (1) Microphotographs of lymphatic organs of patients with three diagnoses: B-cellular chronic lymphoid leukosis (CLL); sarcomatous transformation of a B-cellular chronic lymphoid leukosis (TCLL); primary B-cellular lymphosarcoma (LS).[1] The prints of the material were obtained within 20–30 min after biopsy and were stained according to Gimza. Microphotographs were obtained on an original digital video microscopic setup mounted on a Leica DMRB microscope with Planapohromat ×100/1.3 oil-immersion lenses. The area of one pixel in the digital microphotographs constitutes 0.0036 or 0.0064 μm^2. Photos of preparations were saved in the form of 24-bit color images 1792 × 1200 or 1536 × 1024 pixels in size in TIFF format. In Fig. 1, a monochrome photo of a preparation of a lymph node is presented. Dark roundish areas on a light background represent lymphocyte nuclei. Light, close to white impregnations in a dark nucleus represent a chromatin nucleus. (2) Contours, determined by expert morphologists, of diagnostically valuable lymphoid cell nuclei (nuclei characteristic of three diagnoses are presented in Fig. 2).

Setting the task. To construct an information technology for referring new images of cell nuclei of the lymphatic system to one of the classes of images corresponding to the three established diagnoses: malignant tumors (LS, TCLL), or benign tumors (CLL).

Separate aspects of solving this task have been surveyed in [4, 11], the results of which are partially used in this section and in posing the task.

[1] According to the classification of A. Bobrova and M. Brilliant [24].

Information technology. The technology is based on processing and image analysis methods and pattern recognition methods. Processing and image analysis methods are used in solving problems of obtaining images of cytologic preparations and segmentation of cellular nuclei, as well as obtaining data necessary for constructing a description of the features of nuclei. Pattern recognition methods are used for constructing a description of the features of nuclei, patients, and classification of patients.

The information technology includes the following stages:

(1) Creation of an archive of images of preparations of lymphatic organs with selected nuclei of lymphocytes for patients with benign and malignant tumors of the lymphatic system.

(2) Image processing for the purpose of eliminating differences in luminance and staining of preparations.

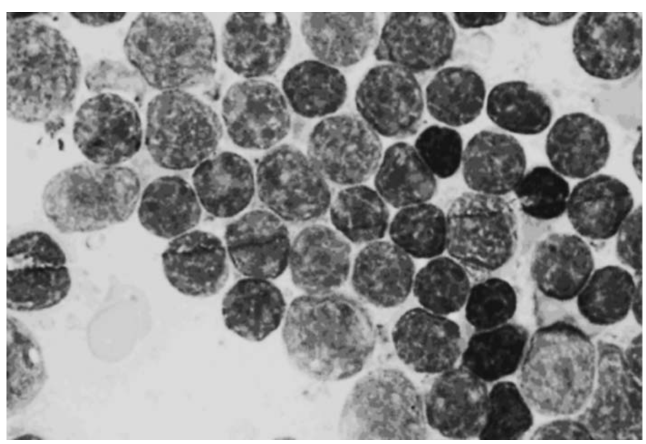


Fig. 1. Monochrome microphotograph of the print of a lymph node. Lens ×100.

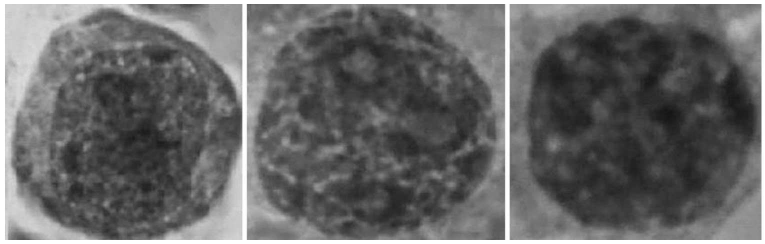


Fig. 2. Cell nuclei of patients with diagnoses of LS, TCLL, and CLL (from left to right).

(3) The choice and evaluation of features reflecting morphological characteristics used in diagnostics of lymphoid cell nuclei.

(4) Qualitative and statistical analysis of the calculated features and an estimation of their self-descriptiveness.

(5) Factor analysis of features.

(6) Creation of a description of features of the patient on the basis of the results of cluster analysis of the patient's nuclei;

(7) Experiments on classification (diagnosing) of patients.

The purpose of the given paper is to construct a **descriptive model of an information technology** for morphological analysis of cytologic preparations of people with tumors of the lymphatic system. The descriptive model should guarantee uniform representation of the algorithm for solving the task.

3.2. Descriptive Algebras of Images

Let us introduce the operands and operations (and their functions) of DIA and DIG necessary to construct the algebraic model for morphological analysis of nuclei of lymphatic cells.

DIA 1 is a set of color images.

Operands: Set $U \equiv \{I\}$ is a set of images of the sort $I = \{\{(r(x, y), g(x, y), b(x, y)), r(x, y), g(x, y), b(x, y) \in [0\dots M-1]\}, (x, y) \in X\}$, where $M = 256$ is the value of maximum intensity of color components, n is the number of source images, and X is the set of pixels.

Operations consist of algebraic operations of (1) vector addition by module M, (2) a vector product by module M, and (3) the operation of taking an element of the whole part of multiplication of module M by an element from the field of real numbers of a vector components of the image at each point of the image:

(1) $I_1 + I_2 = \{\{((r_1(x, y) + r_2(x, y)) \bmod M, (g_1(x, y) + g_2(x, y)) \bmod M, (b_1(x, y) + b_2(x, y)) \bmod M), r_1(x, y), r_2(x, y), g_1(x, y), g_2(x, y), b_1(x, y), b_2(x, y) \in [0\dots M - 1]\}, (x, y) \in X\}$;

(2) $I_1I_2 = \{\{((r_1(x, y)r_2(x, y)) \bmod M, (g_1(x, y)g_2(x, y)) \bmod M, (b_1(x, y)b_2(x, y)) \bmod M), r_1(x, y), r_2(x, y), g_1(x, y), g_2(x, y), b_1(x, y), b_2(x, y) \in [0\dots M-1]\}, (x, y) \in X\}$;

(3) $\alpha I = \{\{([\alpha r(x, y) \bmod M], [\alpha g(x, y) \bmod M], [\alpha b(x, y) \bmod M]), r(x, y), g(x, y), b(x, y) \in [0\dots M-1], \alpha \in R\}, (x, y) \in X\}$.

DIA 1 is applied to the description of source images, and the operation of product DIA 1 is applied to the description of the segmentation stage of diagnostically important nuclei in images.

DIG 1 is the set of operations $sb((U, C) \longrightarrow U')$ of the obtaining of binary masks corresponding to selected nuclei of lymphatic cells, C is information on contours of selected nuclei, and set U' is a subset of set U. If the image point (x, y) belongs to a selected nucleus, then r components $(x, y) = g(x, y) = b(x, y) = 1$; if the point (x, y) belongs to the background of a nucleus, then $r(x, y) = g(x, y) = b(x, y) = 0$.

Operands: Elements DIG 1 are operations $sb((U, C) \longrightarrow U') \in SB$.

Addition and product **operations** are introduced to the set of functions sb as operations of consecutive obtaining of binary masks and their addition and product, respectively:

(1) $sb_1(I, C) + sb_2(I, C) = SB_1 + SB_2$;

(2) $sb_1(I, C)sb_2(I, C) = SB_1SB_2$.

DIG 1 is applied to the description of the segmentation process.

DIG 2 is set U' of binary masks.

Operands: Elements DIG 2 are binary masks $SB = \{\{(r(x, y), g(x, y), b(x, y)), r(x, y), g(x, y), b(x, y) \in \{0,1\}, r(x, y) = g(x, y) = b(x, y)]\}, (x, y) \in X\}, M = 256\}$.

Operations of addition and product are operations of combination and intersection, respectively:

(1) $SB_1 + SB_2 = \{\{(r_1(x, y) \vee r_2(x, y), g_1(x, y) \vee g_2(x, y), b_1(x, y) \vee b_2(x, y)), r_1(x, y), r_2(x, y), g_1(x, y), g_2(x, y), b_1(x, y), b_2(x, y) \in \{0, 1\}\}, (x, y) \in X\}$;

(2) $SB_1SB_2 = \{\{(r_1(x, y) \wedge r_2(x, y), g_1(x, y) \wedge g_2(x, y), b_1(x, y) \wedge b_2(x, y)), r_1(x, y), r_2(x, y), g_1(x, y), g_2(x, y), b_1(x, y), b_2(x, y) \in \{0, 1\}\}, (x, y) \in X\}$.

DIG 2 is applied to the description of binary masks of color images.

DIA 2 is a set grayscale images.

Operands: Set $V \equiv \{J\}$ is set of images $J = \{\{\mathrm{gray}(x, y)\}_{(x, y) \in X}, (x, y) \in [0, \dots, M-1]\}$.

Operations—Algebraic operations of addition module M, products of module M, and the operation of taking of the whole part of the element of the product on an element of the field of real numbers of module M of the grayscale function of an image at each point of the image:

(1) $J_1 + J_2 = \{\{(\mathrm{gray}_1(x, y) + \mathrm{gray}_2(x, y)) \bmod M, \mathrm{gray}_1(x, y), \mathrm{gray}_2(x, y) \in [0\ldots M-1]\}, (x, y) \in X\}$;

2) $J_1J_2 = \{\{(\mathrm{gray}_1(x, y)\mathrm{gray}_2(x, y)) \bmod M, \mathrm{gray}_1(x, y), \mathrm{gray}_2(x, y) \in [0\ldots M-1]\}, (x, y) \in X\}$;

3) $\alpha J = \{\{[\alpha \mathrm{gray}(x, y) \bmod M], \mathrm{gray}(x, y) \in [0\ldots M-1], \alpha \in \mathrm{R}\}, (x, y) \in X\}$.

DIA 2 is applied to description of the selected nuclei in images.

DIA 3 is set F of operations $f(U \longrightarrow V)$, conversions of elements of the set of color images into elements of the set of grayscale images.

Operands: Elements of DIA 3 are operations $f(U \longrightarrow V) \in F$; such conversions can be used for smoothing the difference in luminance and in the color range of images.

Operations of addition, product, and multiplication by an element of the field of real numbers are introduced to the set of functions f as serial operations for obtaining grayscale images and their addition, product, and multiplication by an element of the field of real numbers, respectively:

(1) $f_1(I) + f_2(I) = J_1 + J_2$;

(2) $f_1(I)f_2(I) = J_1J_2$;

(3) $\alpha f(I) = \alpha J$.

DIA 3 is applied to smoothing the difference in luminance and in the color range of cell images.

DIA 4 is set G of operations $g(V \longrightarrow P_1)$ to evaluate the features of grayscale images.

Operands: **DIA 4** is a ring of functions $g(V \longrightarrow P_1) \in G$, where P_1 is a set of P-models (parametrical models).

Operations of addition, product, and multiplication by a field element are introduced into the set of functions g as operations of consecutive evaluation of appropriate P-models and their addition, product, and multiplication by a field element.

(1) $g_1(J) + g_2(J) = p_1(J) + p_2(J)$;

(2) $g_1(J)g_2(J) = p_1(J)p_2(J)$;

(3) $\alpha g(J) = \alpha p(J)$.

DIA 4 is applied to computation of values of features.

DIA 5 is set P_1 of P-models.

Operands: Set P_1 of P-models $p = (f_1, f_2, \ldots, f_n)$, where $f_1, f_2, \ldots, f_n$ are the features calculated in grayscale images, an n is the number of features.

Operations:

(1) Addition of two P-models is an operation of combining numerical descriptions of images: $p_1 + p_2 = (f_1^1, f_2^1, \ldots, f_{n1}^1) + (f_1^2, f_2^2, \ldots, f_{n2}^2) = (f_1^3, f_2^3, \ldots, f_{n3}^3)$, where n_3 is the number of features of P-model p_1 a plus the number of features of P-model p_2 a minus the number of conterminous features of P-models p_1; p_2, $\{f_1^3, f_2^3, \ldots, f_{n3}^3\} \subset \{f_1^1, f_2^1, \ldots, f_{n1}^1, f_1^2, f_2^2, \ldots, f_{n2}^2\}$ are not all repeating features from P-models p_1 and p_2 of grayscale images.

(2) The product of two P-models is the operation of obtaining addition of numerical descriptions of images: $p_1p_2 = (f_1^1, f_2^1, \ldots, f_{n1}^1)(f_1^2, f_2^2, \ldots, f_{n2}^2) = (f_1^4, f_2^4, \ldots, f_{n4}^4)$; n_4 is the number of essential features of the incorporated P-model of models p_1 and p_2; $f_1^4, f_2^4, \ldots, f_{n4}^4$ are essential features of P-model p_1; and P-models p_2, $f_1^4, f_2^4, \ldots, f_{n4}^4$ cannot belong to set $\{f_1^1, f_2^1, \ldots, f_{n1}^1, f_1^2, f_2^2, \ldots, f_{n2}^2\}$ and can consist of a combination of features.

(3) Multiplication by a field element is an operation of multiplication of an element of a P-model from set P_1 by a field element: $\alpha p = \alpha(f_1, f_2, \ldots, f_n) = (\alpha f_1, \alpha f_2, \ldots, \alpha f_n)$.

DIA 5 is applied to the choice of the most informative features. Addition is applied to constructing the incorporated parametrical model of the image. The product is applied to bringing the set of features of images to the set of essential features . Multiplication by an element of the field of real numbers is applied to normalization of the vector of features.

DIA 6 is set P_2 of P-models (P_2 includes the vector of features of the same length).

Operands: Set P_2 of P-models $p(J) = (f_1(J), f_2(J), \ldots, f_n(J))$, where n is the number of features; $f_1(J), f_2(J), \ldots, f_n(J)$ are features of grayscale images of features; and $f_1(J), f_2(J), \ldots, f_n(J) \in R$.

Operations of addition, product, and multiplication by a field element are introduced to set P_2 as an operation of addition, product, and multiplication by an element of the field of vectors:

(1) $p(J_1) + p(J_2) = (f_1(J_1), f_2(J_1), \ldots, f_n(J_1)) + (f_1(J_2), f_2(J_2), \ldots, f_n(J_2)) = (f_1(J_1) + f_1(J_2), f_2(J_1) + f_2(J_2), \ldots, f_n(J_1) + f_n(J_2))$;

(2) $p(J_1)p(J_2) = (f_1(J_1), f_2(J_1), \ldots, f_n(J_1))(f_1(J_2), f_2(J_2), \ldots, f_n(J_2)) = (f_1(J_1)f_1(J_2), f_2(J_1)f_2(J_2), \ldots, f_n(J_1)f_n(J_2))$;

(3) $\alpha p(J) = \alpha(f_1(J), f_2(J), \ldots, f_n(J)) = (\alpha f_1(J), \alpha f_2(J), \ldots, \alpha f_n(J))$.

Table 1. DIAs with one ring used for the description of the algorithmic scheme

	Ring elements	Ring operations	Purpose
DIA 1	Color images	Algebraic operations of vector addition of module M, the vector product of module M, and operation of taking the whole positive part from the multiplication of module M by an element of the field of real numbers of each point of the image	Description of source images and segmentation process
DIG 1	Operations of obtaining the binary mask corresponding to a selected nucleus on a blood preparation	Serial operations of obtaining binary masks and their addition and product	The description of the segmentation process
DIG 2	Binary masks corresponding to selected nuclei	Algebraic operations of combination and intersection	Description of binary masks
DIA 2	Grayscale images	Algebraic operations of addition of module M, the product of module M, and the taking of the positive whole part from multiplication by an element of the field of real numbers of module M of the functions describing tonal images at each point of the image	Description of selected nuclei in images
DIA 3	Operations of converting color images into tonal images	Consecutive application of operations of obtaining tonal images and their addition, product, and multiplication by an element of the field of real numbers	Smoothing of the difference in luminance and the color range of images
DIA 4	Operations of evaluating characters of images	Consecutive evaluation of appropriate *P*-models and their addition, product, and multiplication by a field element	Calculation of characters
DIA 5	*P*-models	Operations of algebra of images (combination, addition, and multiplication by an element of the field of real numbers)	Choice of informative characters
DIA 6	*P*-models	Operations of vector addition, product, and multiplication by a field element	Bringing of images to a form convenient for recognition

DIA 6 is applied to the description of images brought to a type convenient for recognition.

Table 1 shows all DIAs with one ring and DIGs used to describe the algorithmic scheme for solving the task of recognition of cytologic images.

3.3. Descriptive Image Models

Let us describe the DIMs necessary for constructing the algebraic model for morphological analysis of lymphatic cell nuclei.

Source images of nuclei of cytologic preparations are given by the set of implementations of images of $\{I_i\}_{i=1,\ldots,n} \subset \{\tilde{I}'\}$ -digital representations of images in TIFF format (n is the amount of source images); the set of information on the patient and the diagnosis $\{D_i\}_{i=1,\ldots,n} \subset \{\tilde{B}'\}$; the set of contours, selected by experts, of diagnostically important nuclei of the cytologic preparations $\{C_j\}_{j=1,\ldots,m} \subset \{\tilde{B}'\}$ (m is the number of diagnostically important nuclei selected in images of cytologic preparations).

Procedural DIM 1.

The purpose of introduction: description of the selected masks describing the diagnostically important nuclei in source images.

Input data: the set of implementations of images $\{I_i\}_{i=1,\ldots,n} \subset \{\tilde{I}'\}$ and the set of contours of diagnostically important nuclei of cytologic preparations $\{C_j\}_{j=1,\ldots,m} \subset \{\tilde{B}'\}$.

Procedural conversion: O_T^1 is given by function $sb((U, C) \longrightarrow U') \in SB\ (DIG1)$, describing the method of applying the segmentation algorithm (set U-set of source images and set U'-subset of set U, such that three color components take simultaneously either a value of 1 or 0).

DIG 1, thus, generates a set of procedural representations of images:

$$\{\Re_1^T(I_{i(j)}, C_j)(\tilde{p})\} \equiv \{sb(I_{i(j)}, C_j)(\tilde{p})\},$$

where $j = 1, \ldots, m$, $i(j) = 1, \ldots, n$, $sb(I_{i(j)}, C_j)(\tilde{p}) \in DIG1$.

By the choice and posing of specific parameters of the segmentation algorithm, a specific representation of the image is constructed, the application of which to the source information leads to the construction of procedural models of source images $\{M_T^1(I_{i(j)}, C_j)\}_{j=1,\ldots,m} \equiv \{I_j^1\}_{j=1,\ldots,m}$.

Note that procedural image models are also *I*-image models (statement 2); that is, they are in themselves implementations of source images. This means that to procedural DIMs it is possible to apply procedural, parametrical, and generating conversions over images to construct new image models.

Procedural DIM 2.

Purpose of introduction: description of diagnostically important nuclei selected in images of cytologic preparations.

Input data: the set of implementations of images $\{I_i\}_{i=1,\ldots,n} \subset \{\tilde{I}'\}$ and the set of binary masks corresponding to diagnostically important nuclei in images of cytologic preparations described by DIMs, $\{I_j^1\}_{j=1,\ldots,m}$.

Procedural conversion: O_T^2 is set by operation DIA1 of the product of appropriate source images and binary masks.

Operation of product DIA 1, thus, generates a set of procedural representations of images:

$$\{\Re_T^2(I_{i(j)}, I_j^1)\} \equiv \{I_{i(j)}I_j^1\},$$

where $j = 1, \ldots, m$, $i(j) = 1, \ldots, n$.

When the obtained representation of the image is applied to the source information, the set of procedural models of source images $\{M_T^2(I_{i(j)}, I_j^1)\}_{j=1,\ldots,m} \equiv \{I_j^2\}_{j=1,\ldots,m}$ is constructed.

Procedural DIM 3.

Purpose of introduction: representation of diagnostically important nuclei selected in images of cytologic preparations in grayscale images. Conversion of a color image into a grayscale one is carried out to eliminate the difference in color and luminance of preparations.

Input data: the set of diagnostically important nuclei selected in images of cytologic preparations described DIMs, $\{I_j^2\}_{j=1,\ldots,m}$.

Procedural conversion: O_T^3 is given by the function $f(U \longrightarrow V) \in F$ (DIA 3) conversions of elements of a set of color images into a set of elements of grayscale images.

DIA 3, thus, generates a set of procedural representations of images:

$$\{\Re_T^3(I_j^2)\} \equiv \{f(I_j^2)\},$$

where $j = 1, \ldots, m$.

When the obtained representation of the image is applied to the source information, the set of procedural models of source images $\{M_T^3(I_j^2)\}_{j=1,\ldots,m} \equiv \{I_j^3\}_{j=1,\ldots,m}$ is constructed.

Parametrical DIM 1.

Purpose of introduction: To obtain the description of diagnostically important nuclei selected in images of cytologic preparations in the form of the vector of features. The given model of the image is admissible for applying recognition algorithms.

Input data: The set of diagnostically important nuclei selected in grayscale images of cytologic preparations described DIM, $\{I_j^3\}_{j=1,\ldots,m}$.

Parametrical conversions: $\{O_P^h\}_{h=1,\ldots,q}$ (q is the quantity of parametrical conversions) are given by the functions $g(V \longrightarrow P_1) \in G$ (DIA 4) for computing the features of grayscale images.

DIA 4, thus, generates a set of parametrical representations of images:

$$\{\Re_P^1(I_j^3)\} \equiv \{\{g^h(I_j^3)\}_{h=1,\ldots,q}\},$$

where $j = 1, \ldots, m$.

When the obtained representation of the image is applied to the source information, the set of parametrical models of source images $\{M_P^1(I_j^3)\}_{j=1,\ldots,m} \equiv \{M_P^1(j)\}_{j=1,\ldots,m}$ is constructed.

Parametrical DIM 2.

Purpose of introduction: To derive the description of diagnostically important nuclei selected in images of cytologic preparations in the form of a reduced vector of features. The given model of the image is constructed to select the most informative features describing the images.

Input data: The set of vectors of the features describing diagnostically important nuclei selected in grayscale images of cytologic preparations, represented in DIM form, $\{M_P^1(j)\}_{j=1,\ldots,m}$.

Over parametrical image models, operations of DIA 5 are introduced—addition, product, and multiplication by an element of the field of real numbers—for construction of the reduced vectors of the most informative features .

DIA 5, thus, generates a set of parametrical representations of images:

$$\{\Re_P^2(M_P^1(j))\} \equiv \{(+, \cdot, \alpha)M_P^1(j)\},$$

where $j = 1, \ldots, m$.

When the obtained representation of the image is applied to the source information, the set of parametri-

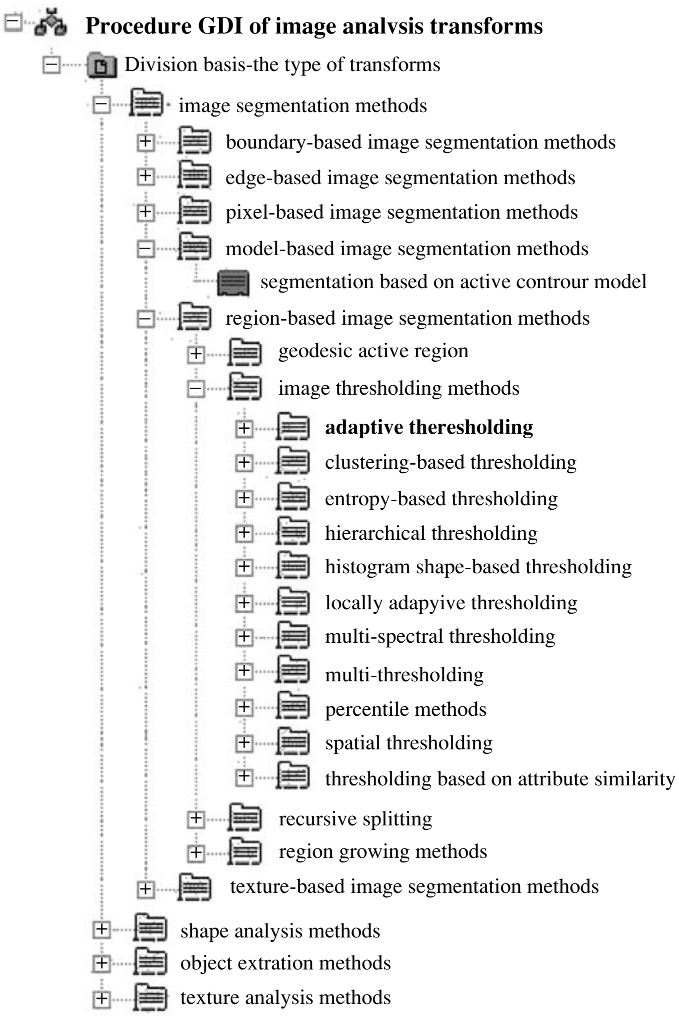


Fig. 3. Procedural GDT image analysis conversions.

cal models of source images $\{M_P^2(M_P^1(j))\}_{j=1,\dots,m} \equiv \{M_P^2(j)\}_{j=1,\dots,m}$ is constructed.

3.4. Generating Descriptive Trees

Research into parametrical and procedural GDTs have led to construction of examples of GDTs used to construct the above-described DIM. Note that procedural GDTs were constructed using the thesaurus of analysis and image processing [1].

Figure 3 presents the GDT of procedural conversions of image analysis T_1.

Figure 4 presents the GDT of conversions of image processing T_2.

According to this analysis of the literature on features, the following classifications of features [8] have been identifed:

—According to the image forming the basis for evaluation of a feature.

—According to the modeling representation forming the basis for evaluation of a feature of the image.

—According to image area on which the feature is calculated.

—According to the object forming a basis for an evaluation of a feature.

—According to a feature level.

—According to the method by which a feature is determined.

—According to the space of which the feature is an admissible element.

—According to the mathematical apparatus used to determine features.

—According to the level of regenerative ability.

—According to the invariance to certain type of conversions.

According to these classifications, it is possible to construct GDTs of character descriptions of ten types (ten base parametrical GDTs).

Figure 5 presents the GDT of parametrical conversions T_3.

The above GDTs were used to choose the conversions for construction of image models.

(1) GDT1 was used for construction of procedural DIM $\{M_T^1(I_{i(j)}, B_j)\}_{j=1,\dots,m}$.

The choice of segmentation algorithm is performed by means of procedural GDT conversions of image analysis T_1 shown in Fig. 3. The boldface type denotes the peak of tree $T_1[\alpha]$ corresponding to selected segmentation algorithm $sb((U, C) \longrightarrow U') \in SB\ (DIG1)$

(2) GDT2 was used for construction of procedural DIM $\{M_T^2(I_{i(j)}, I_j^1)\}_{j=1,\dots,m}$ and $\{M_T^3(I_j^2)\}_{j=1,\dots,m}$.

The choice of operation of the product of two images (for construction of DIM $\{M_T^2(I_{i(j)}, I_j^1)\}_{j=1,\dots,m}$) and operations of converting elements of a set of color images into elements of a set of grayscale images (for construction of DIM $\{M_T^3(I_j^2)\}_{j=1,\dots,m}$) is carried out by means of procedural GDT conversions of image processing T_2, shown in Fig. 4. The italic type denotes the peak of tree $T_2[\beta]$ corresponding to the operation of the product of images. The boldface type denotes the peak of tree $T_2[\chi]$ corresponding to the operation of converting $f(U \longrightarrow V) \in F(DIA\ 3)$.

(3) GDT3 was used for construction of parametrical DIM $\{M_P^1(j)\}_{j=1,\dots,m}$.

The choice of features of the images used for construction of parametrical DIM $\{M_P^1(j)\}_{j=1,\dots,m}$ is carried out by means of the parametrical GDT of features of images T_3, shown in Fig. 5. The italic type denotes the following nodes of trees: the area of the object in

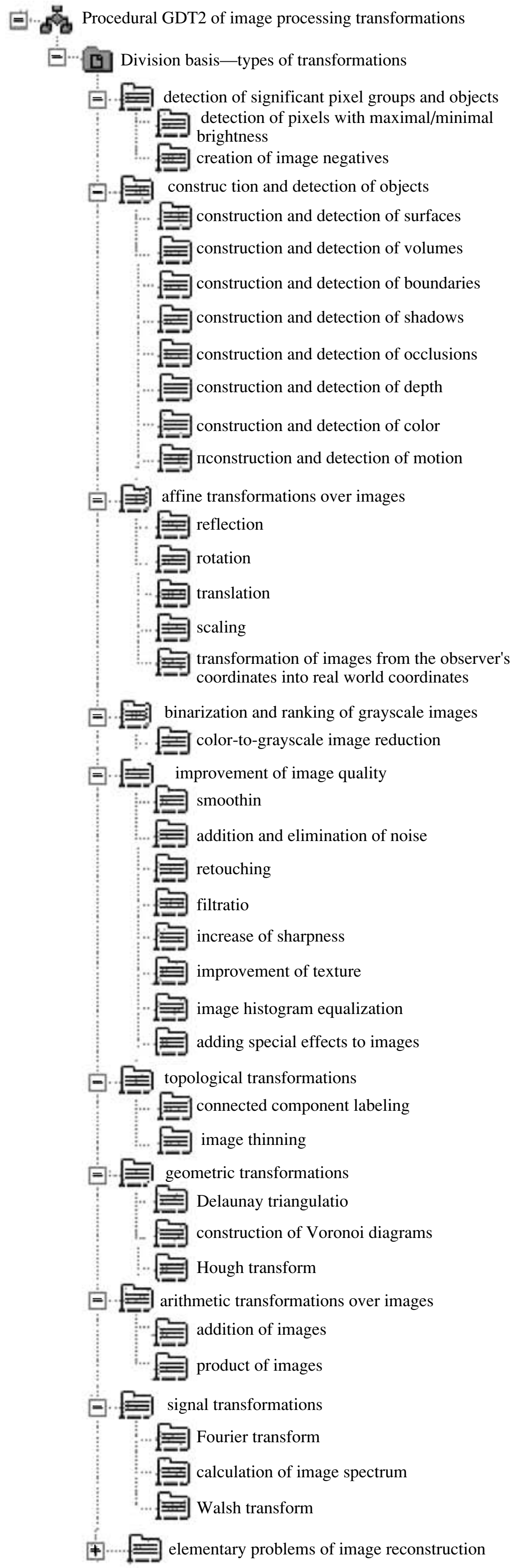


Fig. 4. Procedural GDT image processing conversions.

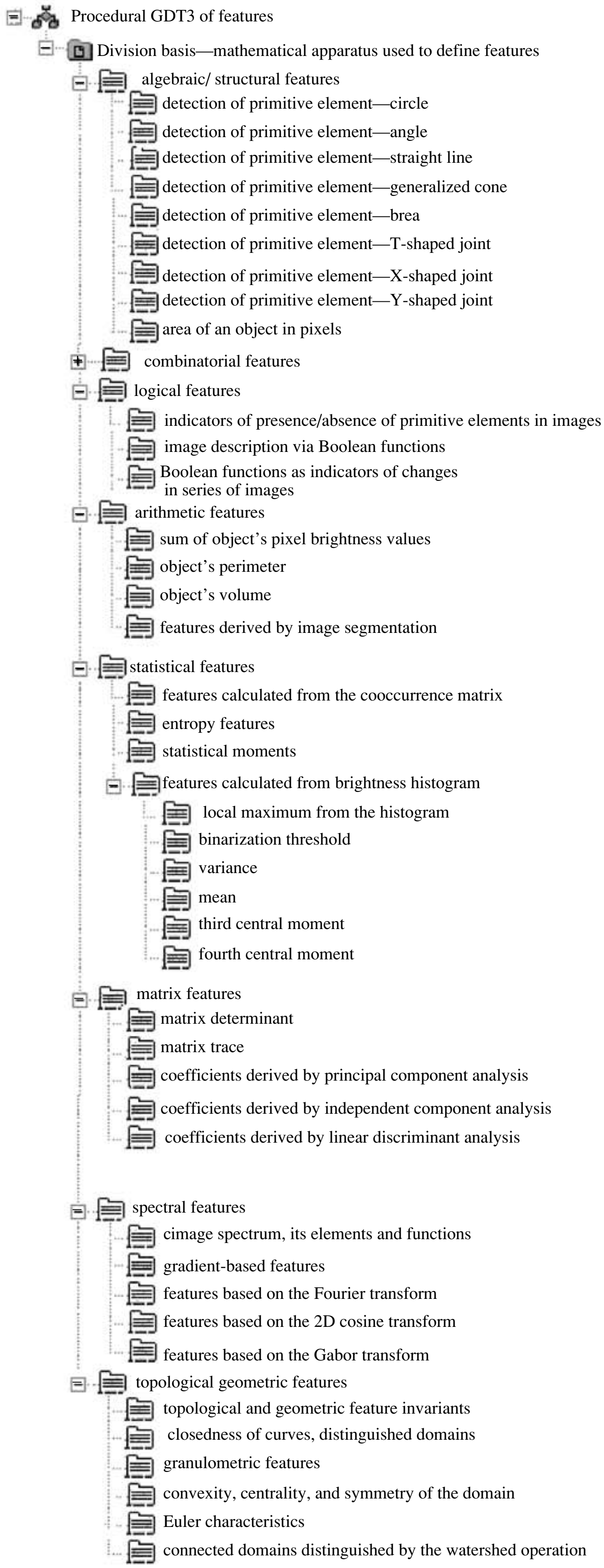


Fig. 5. Parametrical GDT characters of images.

pixels $T_3[\delta]$; four statistical features (variance $T_3[\varepsilon]$, average $T_3[\phi]$, the third central moment $T_3[\varphi]$, and the fourth central moment $T_3[\gamma]$); the features based on Fourier transform $T_3[\eta]$ (26 features); and granulometric features $T_3[\mu]$ (16 features).

3.5. Algorithmic Scheme of Morphological Analysis of Cell Nuclei of the Lymphatic System

3.5.1 Input data. To solve the task of recognition of lymphoid cell nuclei, an archive containing 1830 images of preparations from 43 patients has been created. The archive contains both images of preparations and contours of diagnostically valuable nuclei of lymphoid cells (Table 2).

Note that the source images of nuclei of the cytologic preparations are given by the set of implementations of images $\{I_i\}_{i=1,\dots,n} \subset \{\tilde{I}'\}$ (n is the quantity of source images), the set of information on the patient and diagnosis $\{D_i\}_{i=1,\dots,n} \subset \{\tilde{B}'\}$, and the set of contours of diagnostically important nuclei of cytologic preparations $\{C_j\}_{j=1,\dots,m} \subset \{\tilde{B}'\}$ (m is the number of the diagnostically important nuclei selected in images of cytologic preparations).

The set of implementations of images $\{I_i\}_{i=1,\dots,n} \subset \{\tilde{I}'\}$ is described by DIA 1.

Source images shared by two groups: learning set of images $\{I_i\}_{i=1,\dots,\left[\frac{n}{2}\right]}$ and recognized set of images $\{I_i\}_{i=[n/2]+1,\dots,n}$.

The descriptive model of the developed information technology for morphological analysis of cell nuclei of the lymphatic system is described by means of algorithmic scheme (3), which was interpreted by means of specialized DIAs, DIMs, and GDTs.

In accordance with scheme 3, the descriptive model consists of three stages: stage 1 of bringing images to a form convenient for recognition; stage 2 of training the recognition algorithm; stage 3 of applying the customized recognition algorithm to models of source images.

3.5.2. Image reduction to a recognizable form. Stage 1 is schematically described as follows:

$$\{I_i\}_{i=1,\dots,n} \xrightarrow{\{M_y\}_{y=1,\dots,s}} \{M_y(I_i)\}_{s\times n}.$$

For a detailed description of the given stage, its six steps 1.1–1.6 will be assigned as follows: obtaining masks of diagnostically important nuclei; segmentation of diagnostically important nuclei in images; convergence of color images to grayscale images; computation of features by the constructed models of a learned set; choice of informative features; and computation of features by the constructed models of a recognized set. Each step is presented as follows: (1) description of the step; (2) step operands; (3) step operations; (4) results of application of step operations. Letters "a" and "b" denote in what place processing of the learning and recognized images differs.

Table 2. Database statistics

Diagnoses	No. patients	No. images	No. nuclei
LS	18	986	1639
TCLL	12	536	1025
CLL	13	308	2497
Total	43	1830	5161

Step 1.1. Obtaining masks of diagnostically important nuclei. Application of the segmentation algorithm is described by DIG 1 operands $sb((U, C) \longrightarrow U') \in SB$. The algorithm $sb((U, C) \longrightarrow U') \in SB$ is applied to source images to derive the appropriate mask (Scheme 7).

$$\underset{DIA1}{\{I_i\}_{i=1,\dots,n}, \{C_j\}_{j=1,\dots,m}} \xrightarrow[1.1]{sb\in DIG1} \underset{DIG2}{\{M_T^1(I_{i(j)}, C_j)\}_{j=1,\dots,m}}. \quad (7)$$

Step operands: Source images $\{I_i\}_{i=1,\dots,n}$ and the contours of diagnostically important nuclei on cells of the lymphatic system $\{C_j\}_{j=1,\dots,m} \subset \{\tilde{B}'\}$.

Step operation: The operation described by means of DIG 1. Such description of the operation by means of the whole set of operands of DIG 1 gives flexibility in choosing the segmentation algorithm applied at the given step. During testing of the descriptive model of the information technology for the given step, with the help of GDT 1, peak $T_1[\alpha]$ was selected, which corresponds to the algorithm of threshold segmentation with an automatically customized threshold.

Results of application of step operations consist of the binary masks $\{M_T^1(I_{i(j)}, C_j)\}_{j=1,\dots,m} \equiv \{I_j^1\}_{j=1,\dots,m}$ described by means of operands of DIG 2. These binary masks are procedural models of source images (T-DIM 1).

Step 1.2. Segmentation diagnostically important nuclei in images. Multiplication of the mask obtained at step 1 by a source image gives an image of a diagnostically important nucleus of the source image (Scheme 8).

$$\underset{DIA1}{\{I_i\}_{i=1,\dots,n}, \{I_j^1\}_{j=1,\dots,m}} \xrightarrow[1.2]{(\cdot)DIA1} \underset{DIA2}{\{M_T^2(I_{i(j)}, I_j^1)\}_{j=1,\dots,m}} \equiv \{I_j^2\}_{j=1,\dots,m}. \quad (8)$$

Step operands: Source images $\{I_i\}_{i=1,\dots,n}$ and the binary masks described by operands of DIG 2, obtained at step 1.

Step operation is the operation DIA 1 of the product of the two operands of DIA 1. All source images have been multiplied by the appropriate binary masks. The product operation has been selected by means of GDT 2 and corresponds to peak $T_2[\beta]$ of the tree.

Results of application of the step operation: Procedural models of source images $\{I_j^2\}_{j=1,\dots,m}$ (T-DIM 2).

Step 1.3. Convergence of color images to grayscale images. To compensate the difference in the conditions of luminance and the color range of preparations, color images have been converted into grayscale images (scheme 9).

$$\underset{DIA1}{\{I_j^2\}_{j=1,\dots,m}} \xrightarrow[1.3]{f\in DIA2} \underset{DIA3}{\{M_T^3(I_j^2)\}_{j=1,\dots,m}} \equiv \{I_j^3\}_{j=1,\dots,m}. \tag{9}$$

Step operands: Image models $\{I_j^2\}_{j=1,\dots,m}$ (T-DIM 2).

Step operations are described by the elements of DIA 2. Such a representation of the step of conversion of color images to grayscale images gives flexibility in choosing the processing algorithm for processing color images. During testing of the descriptive model of the information technology, with the help of GDT 2, node $T_2[\chi]$ was chosen, which corresponds to the operation of converting $f(U \longrightarrow V) \in F$ (element DIA 2), which has the following form ($I = \{\{(r(x, y), g(x, y), b(x, y)), r(x, y), g(x, y), b(x, y) \in [0 \dots M-1]\}_{(x, y)\in X}\}$): $f(I) = J = \{\{\text{gray}(x, y)\}_{(x, y)\in X}, (x, y) \in [0\dots M-1]\}$, $\text{gray}(x, y) = g(x, y)\dfrac{2B}{M}$, where B is the average luminance of the dark blue component of the initial RGB image. The green hue in this case is selected as the most informative.

Results of application of step operation are T-models $\{I_j^3\}_{j=1,\dots,m}$ (T-DIM 3).

Step 1.4a. Computation of features by the constructed models of the learned set. To calculate the various features, the learning set of source images was processed by means of various operations of DIA 4 (Scheme 10) (m_1 is the number of segmented nuclei of the learning set).

$$\underset{DIA3}{\{I_j^3\}_{j=1,\dots,m_1}} \xrightarrow[1.4a]{\{g_1, g_2, \dots\}\in DIA4} \underset{DIA5}{\{M_P^1(I_j^3)\}_{j=1,\dots,m_1}} \equiv \{M_P^1(j)\}_{j=1,\dots,m_1}. \tag{10}$$

Step operands: Image models $\{I_j^3\}_{j=1,\dots,m_1}$ (T-DIM 3).

Step operations are described by elements of DIA 4. Such a description of the step gives flexibility in choosing features for obtaining P-models $M_P^1(j)$ (elements of DIA 5). During testing of the descriptive model of the information technology, with the help of parametrical GDT 3, the following were obtained: 47 features, calculated in each source image of a nucleus; the area of the object in pixels $T_3[\delta]$; four statistical features (variance $T_3[\varepsilon]$, average $T_3[\phi]$, the third central moment $T_3[\varphi]$, and the fourth central moment $T_3[\gamma]$); features based on Fourier transform $T_3[\eta]$ (26 features); granulometric features $T_3[\mu]$ (16 features). Model $M_P^1(j)$ is a vector of dimension 47, calculated on each procedural model I_j^3, $j = \dots m_1$.

Results of application of step operations are P-models $\{M_P^1(j)\}_{j=1,\dots,m_1}$ (P-DIM 1).

Step 1.5a. Choice of informative features. This is an additional step for abbreviating the dimension of the image model to which the recognition algorithm will be applied. As will be shown below, the recognition algorithm is applied like it is to complete parametrical model $M_P^1(j)$ ($j = m_1 + 1\dots m$) and to reduced model $M_2^P(j)$ ($j = m_1 + 1\dots m$) (also, there is matching of results of application of the same recognition algorithms to different image models). At the given step, the constructed descriptions of images of the learning set were studied in terms of separation of the most informative features from them (Scheme 11).

$$\underset{DIA5}{\{M_P^1(j)\}_{j=1,\dots,m_1}} \xrightarrow[1.5a]{(+,\,\bullet,\,\alpha\bullet)DIA5} \underset{DIA6}{\{M_P^2(M_P^1(j))\}_{j=1,\dots,m_1}} \equiv \{M_P^2(j)\}_{j=1,\dots,m_1}. \tag{11}$$

Step operands: Parametrical image models $\{M_P^1(j)\}_{j=1,\dots,m_1}$.

Step operations are described by elements of DIA 5. Operations of DIA 5 of addition and product are introduced for uniform convergence of the set of features of images to the set of essential features of images. Multiplication by an element of the field of real numbers is introduced to normalize the vector of features. Such a description of operations on separation of essential features gives flexibility in various methods of analyzing vectors of features to obtain a reduced set of features. During testing of the descriptive model of the information technology, factor analysis [11] was used for feature analysis.

The factor analysis model assumes that the structure of links between observable variables (features) can be explained by the fact that these variables depend on a smaller number of not immediately observable "hidden" variables, called factors. The purpose of factor analysis is to represent a character in the form of a linear combination of factors:

$$F_{ij} = a_i^1 p_j^1 + a_i^2 p_j^2 + \ldots + a_i^r p_j^r,$$
$$i = \overline{1, n}, \quad j = \overline{1, m},$$

where F_{ij} is value of the ith character in the jth character description; n is the number of features; m is the number of character descriptions; r is number of factors; $r < n$, a_i^1, a_i^2, … a_i^r are factor loads subject to determination of the ith character, taking a value from –1 to 1; p_j^1, p_j^2, …, p_j^r are the factors for the jth nucleus. This factor consists of the factor loads counted over all features. Each addend $a_i^k p_j^k$ reflects the degree of influence of the kth factor on variable F_{ij}. The obtained factors together are called the factor solution.

The vector of the features describing the learning set were studied by means of factor analysis (information on how selected nuclei corresponded to diagnoses was used). It was explained in [11] that for the factors of one significance level obtained on different groups of nuclei, appropriate feature sets with high loads are not equal. Therefore, for different groups of patients, the diagnostic value of the same features varies. Three groups of diagnostically valuable features have been selected: in the first group of features, F1 (the area of the nucleus in pixels), F15, F16 (granulometric features) in combination with individual features from groups F22–F29 (the features based on evaluation of the Fourier spectrum); in the second group of features, F2 (a Fourier character) was selected; and the third group features, F42 and F45 (Fourier features) were selected. By means of factor analysis, 14 of the most important features (which have the greatest loads in the first and second factors) were thus selected.

Results of application of step operations are parametrical models $\{M_P^2(j)\}_{j = 1, \ldots, m_1}$ (P-DIM 2), of vector dimension 14, counted on each procedural image model $I_j^3, j = 1 \ldots m_1$.

Step 1.6b. Evaluation of features on the constructed models of the recognized set. By means of steps 1.4 and 1.5, multimodel representations were obtained for the learning set (multimodelibility arises from the use of several image models in the course of obtaining a definitive one). Step 1.6 is one of evaluation of features of images for the recognized set of images (Scheme 12).

$$\{I_j^3\}_{j = m_1 + 1, \ldots, m} \xrightarrow[1.6b]{\{g_1, g_2, \ldots\} \in DIA4} \{M_P^1(I_j^3) \vee M_P^2(M_P^1(I_j^3))\}_{j = m_1 + 1, \ldots, m} \equiv \{\Psi(j)\}_{j = m_1 + 1, \ldots, m}. \quad (12)$$

Step operands: Image models $\{I_j^3\}_{j = m_1 + 1, \ldots, m}$.

Step operations are described by means of elements of DIA 4. During testing of the descriptive model of the information technology for the given step on a set of recognized images, the evaluation of either 47 or 14 features of images is supposed.

Results of application of step operations are parametrical models $\{\Psi(j)\}_{j = m_1 + 1, \ldots, m}$ (at the given step, multimodel representations of recognized images are calculated).

3.5.3. Training and recognition. As the recognition algorithm, the algorithm for computing estimations (ACO) was selected as an algorithm conveniently represented by means of algebraic tools [11].

Input data. DIA 6 and its operands, $\Psi(j) \equiv M_1^P(I_j^2) \vee M_2^P(M_1^P(I_j^2))$ ($j = 1 \ldots m$), describe input data for recognizing algorithm A; ($\Psi(j) = (\psi_1, \psi_2, \ldots, \psi_n)$ is the vector of features of dimension $n = 47$ or $n = 14$; $\{\Psi(j)\}_{m_1 + 1 \ldots m}$ is the information on the recognized set; $\{\Psi(j)\}_{1 \ldots m_1}$ is the information on the learning set; $\{P_g(I_j')\}_{r \times m_1} = \{a_{gj}\}_{r \times m_1}$ is the information on the learning set of images belonging to classes $\{K_g\}_{1 \ldots r}$ ($a_{gi} \in \{0, 1\}$, $r = 3$); $\{I_j'\}_{1 \ldots m}$ are source images of preparations of one image on each selected nucleus (since in one image more than one nucleus can be selected, this set of source images can contain conterminous images). The recognition algorithm $A(\{\Psi(j)\}_{1 \ldots m_1}, \{a_{gi}\}_{r \times m_1}, \{\Psi(j)\}_{m_1 + 1 \ldots m}) = \{a_{gi}\}_{r \times (m - m_1)} \in \{A_y\}_{1 \ldots l}$ solves the pattern recognition task, and $\{a_{gj}\}_{1 \ldots r}$ is the informational vector obtained by means of algorithm A ($j = m_1 + 1 \ldots m$).

The recognition algorithm was applied like it was to complete image models $M_1^P(j)$ ($j = 1 \ldots m$, 47 features) and to reduced image models $M_2^P(j)$ ($j = 1 \ldots m$, 14 features).

The task of the ACO model. Any ACO algorithm is defined as $A = R_A r_A$, where R_A is the recognition operator (it calculates real estimations) and r_A is a solution rule.

The first step of the task of algorithm A in the ACO model is to indicate the system Ω_A of subsets of the set $\{1, 2, \ldots, n\}$. Elements of Ω_A are called reference sets of the algorithm, and system Ω_A is the system of reference subsets of algorithm A. In the given work, system Ω_A of all nonempty subsets of set $\{1, 2, \ldots, n\}$ was used. To each subset $\Omega = \{i_1, i_2, \ldots, i_k\}$, the characteristic Boolean vector $\tilde{\omega} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ can be compared one to one, where $\alpha_{i_1} = \ldots = \alpha_{i_k} = 1$, and the remaining coordinates are equal to 0.

The second step of the task of algorithm A is to define the proximity function $B(\tilde{\omega} I_u', \tilde{\omega} I_t') = B(\tilde{\omega}\Psi_u,$

$\tilde{\omega}\Psi_t$) of two subdescriptions of objects I'_u and I'_t. A description $\tilde{\omega}\Psi_u$ is called a subdescription I'_j of an object with a description $\Psi_u = (\psi_1^u, \psi_2^u, \ldots, \psi_n^u)$. We will give the proximity function by a system of inequalities with parameters. Let $\varepsilon_i \geq 0$, $i = 1, 2, \ldots, n$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)$. We will write a system inequalities: $\rho_i(\psi_i^u, \psi_i^t) \leq \varepsilon_i$, i = 1, 2, …, k, where the ρ_i is the metrics in set M_i of the value of the ith character, $i = 1, 2, \ldots, n$. We will juxtapose a pair of admissible objects I'_u and I'_t with descriptions Ψ_u and Ψ_t to a characteristic vector $\tilde{\delta} = (\delta_1, \delta_2, \ldots, \delta_n) = \tilde{\delta}(I'_u, I'_t)$ according to the following rule: if $\rho_i(a_i, b_i) \leq \varepsilon_i$, $\delta_i = 1$; if $\rho_i(a_i, b_i) > \varepsilon_i$, $\delta_i = 0$, $i = 1, \ldots, n$. We will designate $\|\tilde{\delta}(I'_u, I'_t)\|$ the number of elements in $\tilde{\delta}(I'_u, I'_t)$. We will examine the proximity function $B(\tilde{\omega}I'_u, \tilde{\omega}I'_t)$, which is equal to 1 if and only if individual coordinates ω in $\tilde{\omega}$ contain, among individual coordinates, $\tilde{\delta}(I'_u, I'_t)$ (all inequalities $\rho_\omega \leq \varepsilon_\omega$ are fulfilled). In the given ACO model, the recognition operator R_A for each new object I' calculates the estimation $\Gamma_j(I') = \frac{1}{\mu(W_j^1)} \sum_{I'_i \in W_j^1} (2^{\|\tilde{\delta}(I', I'_i)\|} - 1)$, where $j = 1, \ldots, r$; $W_j^1 = K_j \cap \{I'_1, I'_2, \ldots, I'_m\}$, $\mu(W_j^1)$ is the number of elements in W_j^1, and m is the number of objects in a class. Parameters of the recognizing operator are elements of vector $\boldsymbol{\varepsilon}$. To the calculated estimations r_A, a decision rule is applied for decision-making on an object belonging to a class. In the given ACO model, a standard decision rule is used:

$$\alpha_i(I') = \begin{cases} 1, & \sum_{j=1}^{r} \delta_j^i \Gamma_j(I') \geq \delta_{r+1}^i \\ 0, & \sum_{j=1}^{r} \delta_j^i \Gamma_j(I') \leq \delta_{r+2}^i. \end{cases}$$

The case of the presence of two or more elements is interpreted as "the object possibly belongs to several classes." The case when the binary vector consists only of zeros implies that the object is not similar to one of classes. The parameters of the decision rule are vectors $\tilde{\delta}^i = (\delta_1^i, \ldots, \delta_r^i, \delta_{r+1}^i, \delta_{r+2}^i)$, $i = 1, \ldots, r$.

Stage 2 "Learning of recognition algorithms" is schematically described as follows:

$$\{M_y^1(I_i)\}_{s_1 \times \left[\frac{n}{2}\right]} \xrightarrow[2]{\{A_k(p)\}_{k=1,\ldots,l}} \{P_g(p)(I_i)\}_{r \times \left[\frac{n}{2}\right]} \xrightarrow{INFO} p_0.$$

Step 2.1. Learning of the recognition operator. At the given step, recognition operator R_A is applied to the learning set to compute the estimations $\{\Gamma_j(I'_i)\}_{r \times m_1}$.

$$\underset{DIA6}{\{\Psi_j\}_{j=1,\ldots m_1}} \xrightarrow[2.1]{} \underset{R}{\{\Gamma_i(j)\}_{r \times m_1}}. \quad (13)$$

Step operands: Image models of learning set $\{\Psi_j\}_{j=1,\ldots,m_1}$.

Step operation is described by the formula $\Gamma_j(I') = \frac{1}{\mu(W_j^1)} \sum_{I'_i \in W_j^1} (2^{\|\tilde{\delta}(I', I'_i)\|} - 1)$.

As a result of **application of the operation**, the set of models of the learning set is converted into a set of real estimations.

Step 2.2. Decision rule training. At the given step, to the set of calculated estimations of the learned set, the decision rule r_A is applied, and on true informational vectors, parameters $\tilde{\delta}^i = (\delta_1^i, \ldots, \delta_r^i, \delta_{r+1}^i, \delta_{r+2}^i)$, $i = 1, \ldots, r$ and $\boldsymbol{\varepsilon}$ are customized.

$$\begin{aligned} &\{\Gamma_i(j)\}_{r \times m_1}, \\ &\underset{R}{\{D_{i(j)}\}_{j=1,\ldots,m_1}} \xrightarrow[2.2]{\alpha_i(I')} \underset{R}{(\boldsymbol{\varepsilon}, \tilde{\delta}^1, \ldots, \tilde{\delta}^r)}. \end{aligned} \quad (14)$$

Step operands: Valid estimations $\{\Gamma_i(j)\}_{r \times m_1}$ and semantic information on diagnoses of patients, whose images of cytologic preparations enter into the learning set $\{D_{i(j)}\}_{j=1,\ldots,m_1} \subset \{\tilde{B}'\}$ (let the present information be given in the form of true informational vectors; then vector elements are also real numbers).

Step operations: Operation of application of the decision rule, operation of estimation of the results in accordance with true informational vectors, and operations of customization of parameters of the decision rule and the recognition operator.

Results of application of step operation will be customized parameters of recognition algorithms $(\boldsymbol{\varepsilon}, \tilde{\delta}^1, \ldots, \tilde{\delta}^r)$.

Stage 3 "Application of recognition algorithms" is schematically described as follows:

$$\{M_y^2(I_i)\}_{s_2 \times \left(n - \left[\frac{n}{2}\right]\right)} \xrightarrow[3]{\{A_k(p_0)\}_{k=1,\ldots,l}} \{P_g(I_i)\}_{r \times \left(n - \left[\frac{n}{2}\right]\right)}.$$

Step 3.1. Application of the recognition operator. At the given step, recognition operator R_A is applied to the recognized set to compute the estimations $\{\Gamma_j(I'_i)\}_{r \times (m - m_1)}$ with parameter $\boldsymbol{\varepsilon}$ obtained at the previous step.

$$\underset{DIA6}{\{\Psi_j\}_{j=m_1+1,\ldots,m}} \xrightarrow[3.1]{} \underset{R}{\{\Gamma_i(j)\}_{r \times (m - m_1)}}. \quad (15)$$

Step operands: Image models of the recognized set $\{\Psi_j\}_{j = m_1 + 1, \ldots, m}$.

Step operation is described by formula $\Gamma_j(I') =$

$$\frac{1}{\mu(W_j^1)} \sum_{I'_i \in W_j^1} (2^{\|\tilde{\delta}(I', I'_i)\|} - 1).$$

As a result of application of the operation, the set of models of the learning set is converted into a set of real estimations.

Step 3.2. Application of the decision rule. At the given step, to the set of calculated estimations of the recognized set, the decision rule r_A with the parameters $\tilde{\delta}^i = (\delta_1^i, \ldots, \delta_r^i, \delta_{r+1}^i, \delta_{r+2}^i)$, $i = 1, \ldots, r$ is applied.

$$\underset{R}{\{\Gamma_i(j)\}_{r \times (m - m_1)}} \xrightarrow[3.2]{} \underset{\{0, 1, \Delta\}}{\{\alpha_{ij}\}_{r \times (m - m_1)}}. \tag{16}$$

Step operands: Valid estimations $\{\Gamma_i(j)\}_{r \times (m - m_1)}$.

Step operation: Operation of application of the decision rule.

Results of application of operation are informational vectors $\{\alpha_{ij}\}_{r \times (m - m_1)}$.

Algorithmic scheme. In this section, we have described the main steps and elements of an information technology for automation of diagnostic analysis of cytologic preparations of patients with tumors of the lymphatic system (Figs. 6, 7):

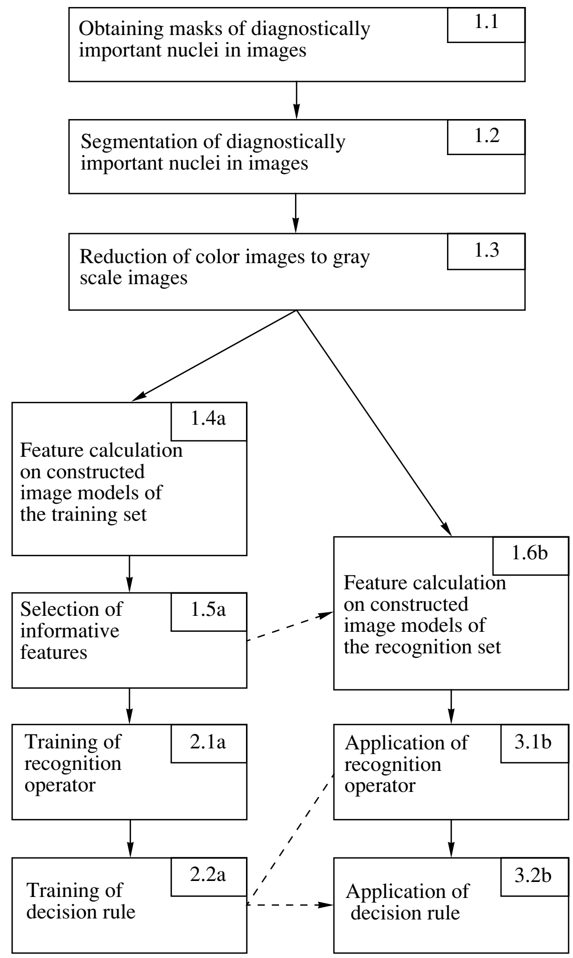


Fig. 6. Descriptive model of the information technology.

3.5.4. Results of testing of descriptive model of an information technology. For testing of the descriptive model of the information technology at the stage of bringing images to a form convenient for recognition, the program system Black square [9] was used. The given system was used for preprocessing of images, segmentation of nuclei of lymphoid cells in images of cytologic preparations, and computation of the geometrical, brightness, granulometric, and spectral features of nuclei.

In the program, the elementary algorithm of threshold segmentation of an image with morphological processing of the obtained image of nuclei is implemented. The segmentation module can be augmented by other procedures of automatic segmentation of images. The module for computation of features can be augmented by procedures for evaluating other features of images.

At the level of learning and recognition, for an experimental research, the program system Recognition 1.0 [29] was used, which includes effective implementation of ACO methods and allows applying them to solve practical tasks. Practice has shown that the best results are achieved by voting over all possible reference sets, and that during autodetection of the power of reference sets and the assigning of a fixed power of the reference set, the accuracy decreases.

When the entire set of features is used, the accuracy of recognition is 86.75%, and for various classes, the accuracy of recognition differs (Table 3). The high accuracy of recognition in diagnosing CLL is probably linked to the fact that CLL is a benign affliction, while LS and TCLL are malignant. Thus, cells in the diagnosis of CLL have strongly pronounced differences from cells of other diagnoses, and among themselves cells of diagnoses of LS and TCLL are more similar.

Table 3. Accuracy of recognition when using feature description from 47 characters

Diagnosis	No. correctly recognized cells	Total no. cells	Accuracy of recognition, %
LS	693	820	84.51
TCLL	325	513	63.35
CLL	1221	1248	97.84
For the entire set of cells	2239	2581	86.75

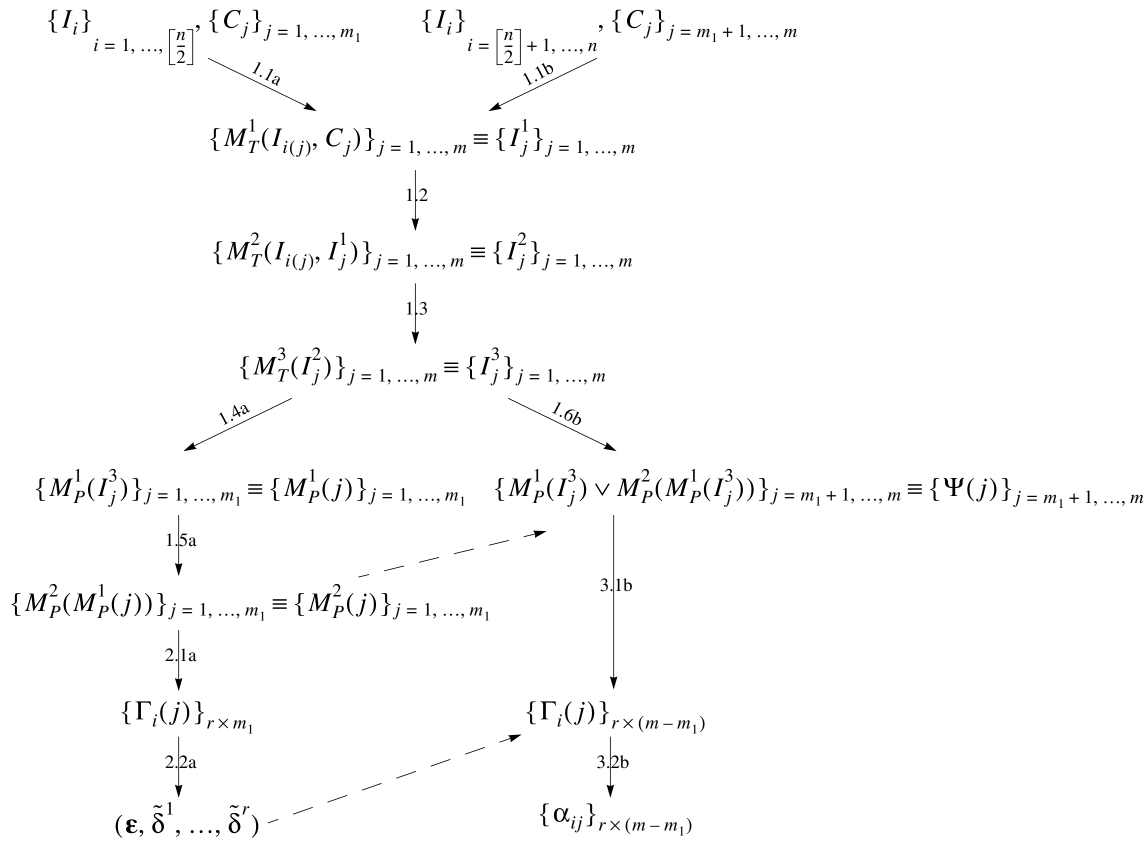


Fig. 7. Algorithmic scheme of the information technology.

When the set of features is reduced to 14, selected by factor analysis, the accuracy of recognition is reduced to 83.18 % (Table 4).

4. CONCLUSIONS

In the given paper, practical application of the algebraic instruments DAIAU is demonstrated: we have shown how to build, by means of DIA, the model of a technology for automating diagnostic analysis of cytologic preparations of patients with tumors of the lymphatic system. This model has been used for the creation of software for application of this technology, its testing, and comparison of results.

Table 4. Accuracy of recognition with character description from 14 characters

Diagnosis	No. correctly recognized cells	Total no. cells	Accuracy of recognition, %
LS	626	820	76.34%
TCLL	300	513	58.48%
CLL	1221	1248	97.84%
For the entire set of cells	2147	2581	83.18%

The main contribution of the given paper is construction of a model for a method ensuring a unified representation of the technology, instead of development of a method for solving a medical task. This work, thus, solves a dual task: first, it represents a technology in the form of a well-structured mathematical model and, second, shows how DIA can be used in an image analysis task.

In the future, DAIAU and its main instruments—DIA, DIM and GDT—will be applied to constructing models of an information technology for automation of diagnostic analysis of medical images in other areas of medicine.

ACKNOWLEDGMENTS

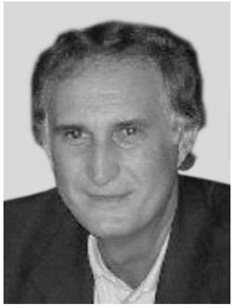
This study was partially supported by the Russian Foundation for Basic Research (grants no. 08-01-00469, 07-07-13545, 06-07-89203); the mathematical sciences division of the Russian Academy of Sciences ("Descriptive Image Algebras with One Ring over Image Models" project within the "Algebraic and Combinatorial Methods of Mathematical Cybernetics" program for fundamental research), and the Russian Academy of Sciences presidium (project no. 2.14 within the

"Fundamental Problems of Computer Science and Information Technology" program).

REFERENCES

1. V. N. Beloozerov, I. B. Gurevich, N. G. Gurevich, et al., Thesaurus for Image Analysis: Basic Version, Pat. Rec. Image Anal. Adv. Math. Theory Appl. **13** (4), 556–569 (2003).
2. U. Grenander, *General Pattern Theory: A Mathematical Study of Regular Structures, Clarendon*: (Oxford, 1993).
3. I. B. Gurevich, The Descriptive Framework for an Image Recognition Problem, in Proc. 6, 1989 pp. 220–227.
4. I. Gurevich, D. Harazishvili, I. Jernova, et al., Information Technology for the Morphological Analysis of the Lymphoid Cell Nuclei, in Proc. 13, 2005, pp. 367–374.
5. I. B. Gurevich and V. V. Yashina, Descriptive Image Algebras with One Ring, in Pat. Recog. Image Anal.: Adv. Math. Theory Appl. **13** (4) (2003).
6. I. B. Gurevich and V. V. Yashina, Generating Descriptive Trees, in Proc. 10, 2005, pp. 367–374.
7. I. B. Gurevich and V. V. Yashina, Operations of Descriptive Image Algebras with One Ring, Pat. Recog. Image Anal. Adv. Math. Theory Appl. **16** (3), 298–328 (2006).
8. I. B. Gurevich and I. V. Koryabkina, Comparative Analysis and Classification of Features for Image Models, Pat. Rec. Image Anal. Adv. Math. Theory Appl. **16** (3), 265–297 (2006).
9. I. B. Gurevich, A. V. Khilkov, I. V. Koryabkina, et al., An Open General-Purpose Research System for Automating the Development and Application of Information Technologies in the Area of Image Processing, Analysis, and Evaluation, Pat. Rec. Image Anal. Adv. Math. Theory Appl. **16** (4), 530–563 (2006).
10. I. B. Gurevich and V. V. Yashina, Computer-Aided Image Analysis Based on the Concepts of Invariance and Equivalence, *Pat. Rec. Image Anal. Adv. Math. Theory Appl*. (MAIK "Nauka/Interperiodica", Pleiades Publishing, Inc., 2006), Vol. 16, No. 4, pp. 564–589.
11. I. B. Gurevich, D.V. Harazishvili, O. Salvetti, et. al., Elements of the Information Technology of Cytological Specimen Analysis: Taxonomy and Factor Analysis, Pat. Rec. Image Anal. **16** (1), 113–115 (2006).
12. I. B. Gurevich and V. V. Yashina, Descriptive Analysis of Image Data: Basic Models, in Image Mining Theory and Applications: Proc. 1, pp. 3–15 (2008).
13. I. Gurevich, V. Yashina, H. Niemann, and O. Salvetti, Medical Image Mining on the Basis of Descriptive Image Algebras: Cytological Specimen Case, in *Proc. Int. Conf. on Health Informatics–HEALTHINF* (2008).
14. J. Grin, J. Kittler, P. Pudil, and P. Somol, Information Analysis of Multiple Classifier Fusion, in Multiple Classifier Systems: Proc. Second Int. Workshop, Springer, pp. 168–177 (2001).
15. J. Kittler and F. M. Alkoot, Relationship of Sum and Vote Fusion Strategies, in Multiple Classifier Systems: Proc. Second Int. Workshop, Springer (2001).
16. P. Maragos, Algebraic and PDE Approaches for Lattice Scale-Spaces with Global Constraints, in International Journal of Computer Vision **52** (2/3), 121–137 (2003).
17. G. X. Ritter and J. N. Wilson, Handbook of Computer Vision Algorithms in Image Algebra, CRC, 2 (2001).
18. J. Serra, *Image Analysis and Mathematical Morphology* (Academic, London, 1982).
19. S. R. Sternberg, *Language and Architecture for Parallel Image Processing, in Proc. Conf. On Pattern Recognition in Practice* (Amsterdam, 1980).
20. S. R. Sternberg, *An Overview of Image Algebra and Related Architectures, in Integrated Technology for Parallel Image Processing,* Ed. by S. Levialdi (Academic, London, 1985).
21. S. R. Sternberg, Grayscale Morphology, Comp. Vis., Graph. Image Proc. **35** (3), pp. 333–355 (1986).
22. D. M. J. Tax and R. P. W. Duin, Combining One-Class Classifiers, in Multiple Classifier Systems (Proc. Second Int. Workshop, Springer, 2001).
23. S. H. Unger, A Computer Oriented toward Spatial Problems, in Proceedings IRE **46**, 1744–1750 (1958).
24. *Atlas of Tumors of the Lymphatic System*, Ed. by A. I. Vorobyev (Hematological Scientific Center, Russian Academy of Medical Sciences, 2001).
25. B. L. Waerden, *Algebra I* (Springer, Berlin, 1971).
26. D. Winbridge and J. Kittler, Classifier Combination as a Tomographic Process, in Multiple Classifier Systems, Proc. Second Int. Workshop, Springer, pp. 248–258 (2001).
27. Yu. I. Zhuravlev and I. B. Gurevich, Pattern Recognition and Image Recognition, Pat. Rec. Image Anal. Adv. Math. Appl. **1** (2), 149–181 (1991).
28. Yu. I. Zhuravlev, An Algebraic Approach to Recognition and Classification Problems, Pat. Rec. Image Anal. Adv. Math. Theory Appl. (1998).
29. Yu. I. Zhuravlev, V. V. Ryazanov, et al., . RECOGNITION: A Universal Software System for Recognition, Data Mining, and Forecasting, Pat. Recog. Image Anal. Adv. Math. Theory Appl. **15** (2), 476–478 (2005).

Heinrich Niemann. Obtained degree of Dipl.-Ing in Electrical Engineering and Dr.-Ing. from Technical University Hanover, Germany. Worked at Fraunhofer Institut für Informationsverarbeitung in Technik und Biologie, Karlsruhe, and at Fachhochschule Giessen in the department of ElectricalEngineering. In 1975 became Professor of Computer Science at the University of Erlangen-Nürnberg, where he was dean of the engineering faculty of the university from 1979–1981. From 1988–2000 headed the research group Knowledge Processing at the Bavarian Research Institute for Knowledge-Based Systems (FORWJSS). From 1998 to 2005, speaker of a "special research area" (SFB) entitled "Model-Based Analysis and Visualization of Complex Scenes and Sensor Data," funded by the German Research Foundation (DFG). His fields of research are speech and image understanding and the application of artificial intelligence techniques in these fields. He is on the editorial board of *Signal, Image and Video Processing*, *Pattern Recognition and Image Analysis*, and *Journal of Computing and Information Technology.* He is the author or coauthor of 7 books and about 550 journal and conference contributions, as well as editor or coeditor of 34 proceedings volumes and special issues. He is a member of GI and IEEE, a honorary member of DAGM, a Fellow of IAPR.



Ovidio SaIvetti. Director of Research at the Institute of Information Science and Technologies (ISTI) of the Italian National research Council (CNR), in Pisa. Works in the field of theoretical and applied computer vision. His fields of research are image analysis and understanding, pictorial information systems, spatial modeling and intelligent processes in computer vision.

He is coauthor of 4 books and monographs and more than 300 technical and scientific articles; he is also owner of ten patents regarding systems and software tools for image processing. He has been scientific coordinator of several National and European research and industrial projects, in collaboration with Italian and foreign research groups, in the fields of computer vision and high-performance computing for diagnostic imaging. Member of the erditorial boards of the international journals *Pattern Recognition and Image Analysis* and *G. Ronchi Foundation Acts*. He is at present the CNR Contact Person at ERCIM (the European Research Consortium for Informatics and Mathematics) for the Working Group on Vision and Image Understanding, Member of IEEE and of the Steering Committee of a number of EU Projects. He is Head of the ISTI Signals and Images Laboratory.

Irina V. Koryabkina. Born 1978. Graduated from the Faculty of Computational Mathematics and Cybernetics of Moscow State University in 2000. Received her candidate's degree in technical sciences (theoretical foundations of computer science) in 2006. Researcher at the Dorodnicyn Computing Center of the Russian Academy of Sciences, head of the Secretariat of the Russian Association for Pattern Recognition and Image Analysis of the Russian Academy of Sciences. Scientific interests: mathematical theory of image analysis, image features, and medical applications. Author of 20 publications in refereed journals and conference abstracts. Awarded by diplomas for the best young scientist papers presented at four international conferences and workshops. Scientific secretary at the 6th and 7th International Conferences on Pattern Recognition and Image Analysis: New Information Technologies and the 6th Open Russian–German Workshop on Pattern Recognition and Image Understanding. Investigator and principal investigator of several research projects in the framework of national and international programs.

Igor' B. Gurevich. Born 1938. Received his Dr.-Eng. degree (Diploma Engineer, Automatic Control and Electrical Engineering) from Moscow Power Engineering Institute in 1961. Received his doctor's degree (Mathematical Cybernetics) from Moscow Institute of Physics and Technology in 1975. At present, he is the head of department at the Dorodnicyn Computing Center of the Russian Academy of Sciences; an assistant professor at the Computer Science Faculty, Moscow Lomonosov State University. He has worked as an engineer and researcher in industry, medicine, and universities since 1960 till now and in the USSR/Russian Academy of Sciences since 1985. Area of expertise: image analysis, image understanding, mathematical theory of pattern recognition, theoretical computer science, medical informatics, applications of pattern recognition and image analysis techniques in medicine, nondestructive testing, process control, knowledge-based systems. Author of 2 monographs (in co-authorship), 205 papers in peer reviewed journals and proceedings, and 5 patents. He is the Scientific Secretary of the National Committee for Pattern Recognition and Image Analysis of the RAS, member of the IAPR Governing Board (representative from RF), IAPR Fellow. He has been the PI of many R&D projects as part of national and international research programs. Vice Editor-in-Chief of Pattern Recognition and Image Analysis Journal, member of Editorial Boards of several international scientific journals, member of the committees of many international scientific conferences.

Vera V.Yashina. Born 1980. Received her diploma mathematician from Moscow Lomonosov State University in 2002. At present, she is a researcher at the Dorodnicyn Computing Center of the RAS. Scientific expertise: mathematical theory of image analysis, image algebras, descriptive image models and medical informatics. Author of 17 papers in peer reviewed journals, conference and workshop proceedings. She was awarded several times for the best young scientist papers presented at several international conferences—7th International Conference on Pattern Recognition and Image Analysis: New Information Technologies, 7th Open German-Russian Workshop on Pattern Recognition and Image Analysis, 8th International Conference on Pattern Recognition and Image Analysis: New Information Technologies.