MEDICAL IMAGE MINING ON THE BASE OF DESCRIPTIVE IMAGE ALGEBRAS
Cytological Specimen Case

I. Gurevich, V. Yashina
Dorodnicyn Computing Center of the Russian Academy of Sciences, Vavilov str., Moscow, Russian Federation

H. Niemann
University of Erlangen-Nuernberg, Lehrstuhl fuer Informatik, Martensstr., Erlangen, Germany

O. Salvetti
Institute of Information Science and Technologies, CNR, 1, Via G.Moruzzi, Piza, 56124, Italy

Keywords: Image mining, image algebras, medical image analysis, pattern recognition in image understanding, information technologies, automated diagnosis.

Abstract: The paper is devoted to the development and formal representation of the descriptive model of information technology for automating morphologic analysis of cytological specimens (lymphatic system tumors). The main contributions are detailed description of algebraic constructions used for creating of mathematical model of information technology and its specification in the form of algorithmic scheme based on Descriptive Image Algebras. It is specified the descriptive model of an image recognition task and the stage of an image reduction to a recognizable form. The theoretical base of the model is the Descriptive Approach to Image Analysis and its main mathematical tools. It is demonstrated practical application of algebraic tools of the Descriptive Approach to Image Analysis and presented an algorithmic scheme of a technology implementing the apparatus of Descriptive Image Algebras.

1 INTRODUCTION

The paper is devoted to the development and formal representation of the descriptive model of the information technology for automating morphologic analysis of cytological specimens of patients with lymphatic system tumors. The main contribution are detailed description of algebraic constructions used for creating of mathematical model of the information technology and its specification in the form of an algorithmic scheme based on Descriptive Image Algebras (DIA). We specify, in particular, the descriptive model of an image recognition task and the stage of an image reduction to a recognizable form.

The theoretical base of the model is the Descriptive Approach to Image Analysis (Gurevich, 2005) and its main mathematical tools –DIA, Descriptive Image Models (DIM) and Generating Descriptive Trees (GDT).

In a sense the results are continuation, specification and extension of the previous research. In (Gurevich, et al. 2007) we presented a brief introduction into the essential tools of the Descriptive Approach (DIA, DIM, GDT), the simplified model of an image recognition task based on multi-model image representation, a descriptive model of the information technology, and the descriptive and the structural schemes of the information technology. The state of the art and motivation were presented in our previous publications (Gurevich, et al. 2003, 2006, 2007).

Section 2 illustrates a simplified descriptive model of an image recognition task based on multi-model image representation. In section 3 we introduce operands and operations (and its operational (semantic) functions) of DIAs and...
Descriptive Image Groups (DIG) necessary for constructing the algebraic model of the morphological analysis of lymphatic cell nucleuses. Section 4 presents a descriptive model of the information technology for automating morphologic analysis of cytological specimens of patients with lymphatic system tumors. The technology has been tested on the specimens from patients with aggressive lymphoid tumors and innocent tumor. The results are discussed in Section 4.

The main components of the technology are described via DIA tools and presented as an algorithmic scheme. The latter ensures a standard representation of technologies for intellectual decision making.

2 DESCRIPTIVE MODEL OF AN IMAGE RECOGNITION PROBLEM

The Descriptive Approach provides the following model for an image recognition process (Gurevich, 2005):

\[
\{I_i\}_{1..n} \rightarrow \{M_j\}_{1..s} \rightarrow \{A_{ij}\}_{1..r} \rightarrow \{P_{ij}(I)\}_{1..u} \quad (1)
\]

\[
\{I_i\}_{1..n} \text{ - a set of initial images, } \{M_j\}_{1..s} \text{ - a multimodel representation for an image recognition task, } \{A_{ij}\}_{1..r} \text{ - an algorithm combination, } \{P_{ij}(I)\}_{1..u} \text{ - a recognition result.}
\]

An algorithm combination \( \{A_{ij}\}_{1..r} \) solves an image recognition problem, if it puts a set of predicates \( \{P_{ij}(I)\}_{1..u} \) into correspondence to the set of initial images, where predicate \( P_{ij}(I)=a_{ij} \) has the values: \( a_{ij}=1 \), if an image \( I \) belongs to a class \( K_{ij} \); \( a_{ij}=0 \), if an image \( I \) does not belong to a class \( K_{ij} \); \( a_{ij} = \Delta \), if an algorithm combination does not establish membership of an image \( I \) to a class \( K_{ij} \).

Multi-model representation is generated by the set of GDT. Different ways for constructing multi-aspect image representations may use different types of GDT. An image representation becomes a multi-model one, if it is generated by different types of GDT.

This model including a training stage is as follows:

\[
\{I_i\}_{1..n} \xrightarrow{\times_{\Delta}} \{M_j\}_{1..s} \xrightarrow{\Delta} \{A_{ij}\}_{1..r} \xrightarrow{\Delta} \{P_{ij}(I)\}_{1..u} \quad (2)
\]

3 DESCRIPTIVE IMAGE ALGEBRAS

In this section we introduce operators and operations (and its operational functions) of DIAs and DIGs necessary for constructing the algebraic model of the morphological analysis of lymphatic cell nucleuses. 

**DIA I** is a set of color images. **The operators:** a set \( U \) of \( \{I\} \) - a set of images \( I = \{(r(x,y), g(x,y), b(x,y)), r(x,y), g(x,y), b(x,y) \in [0..M-1]\}, (x,y) \in X\), \( M=256 \) - the value of maximal intensity of a color component, \( n \) - a number of initial images, \( X \) - a set of pixels. **The operations** are algebraic operations of vector addition module \( M \), vector multiplication module \( M \) and taking an integral positive part of multiplication module \( M \) by an element from the field of real numbers in each image point: 

1. \( I_1 + I_2 = \{((r_1(x,y)+r_2(x,y)) \mod M, (g_1(x,y)+g_2(x,y)) \mod M, (b_1(x,y)+b_2(x,y)) \mod M), r_1(x,y), r_2(x,y), g_1(x,y), g_2(x,y), b_1(x,y), b_2(x,y) \in [0..M-1], (x,y) \in X\} \)
2. \( I_1 I_2 = \{((r_1(x,y)r_2(x,y)) \mod M, \)
(g_1(x,y)·g_2(x,y)) \mod M, (b_1(x,y)·b_2(x,y)) \mod 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...M-1\}, (x,y) \in \mathbb{X}; 3) a \ell = [[f_1(x,y)] \mod M], [a g(x,y) \mod M], [a b(x,y) \mod M], a(x,y), g(x,y), b(x,y) \in \{0...M-1\}, a \in \mathbb{R}, (x,y) \in \mathbb{X}. D I A 1 is applied to describe initial images and the multiplication operation of D I A 1 is applied to describe segmentation of diagnostically important nucleus on images.

D I G 1 is a set of operations \( s_b((U,C) \rightarrow U') \) for obtaining a binary mask corresponding to an indicated lymphocyte cell nucleus, \( C \) - the information about the contours of indicated nucleus, a set \( U' \) - a subset of a set \( U \). If an image point \((x,y)\) belongs to indicated nucleus then \( r(x,y)\) belongs to nucleus background, \( r(x,y)\) belongs to nucleus background, \( r(x,y)=g(x,y)=b(x,y)=0 \). The operands: Elements of D I G 1 are operations \( s_b((U,C) \rightarrow U') \in \mathbb{B} \). The operations of addition and multiplication are introduced on the set of functions \( s_b \) as sequential operations for obtaining a binary mask and their addition and multiplication correspondingly: 1) \( s_b((C)+s_b((C)=B_1+B_2 \); 2) \( s_b((C)+s_b((C)=B_1+B_2 \). D I G 1 is applied to describe a segmentation process.

D I G 2 is a set of binary masks. The operands: Elements of D I G 2 are binary masks \( B=\{(r(x,y), g(x,y), b(x,y))\}, r(x,y), g(x,y), b(x,y) \in \{0,1\}, (x,y) \in \mathbb{X}\} \). The operations of addition and multiplication are operations of union and intersection correspondingly: 1) \( B_1+B_2=\{(r(x,y) \lor r_1(x,y), g(x,y) \lor g_1(x,y), b(x,y) \lor b_1(x,y))\}, r(x,y), g(x,y), b(x,y) \in \{0,1\}\}; 2) \( B_1 \land B_2=\{(r(x,y) \land r_1(x,y), g(x,y) \land g_1(x,y), b(x,y) \land b_1(x,y))\}, r(x,y), g(x,y), b(x,y) \in \{0,1\}\}. D I G 2 is applied to describe binary masks.

D I A 2 is a set of gray scale images. The operands: A set \( V \) of \( [\alpha] \) - a set of images \( J=\{\{\text{gray}(x,y)\}_{(x,y)} \} \in \{0...M-1\} \). The operations are algebraic operations of gray functions addition module \( M \), multiplication module \( M \) and taking an integral positive part of multiplication module \( M \) by an element from the field of real numbers in each image point: 1) \( J_1+J_2=\{\{\text{gray}(x,y)+\text{gray}(x,y)\} \mod M, \text{gray}(x,y), \text{gray}(x,y) \in \{0...M-1\}\}, (x,y) \in \mathbb{X}\}; 2) \( J_1 \cdot J_2=\{\{\text{gray}(x,y) \cdot \text{gray}(x,y)\} \mod M, \text{gray}(x,y), \text{gray}(x,y) \in \{0...M-1\}\}, (x,y) \in \mathbb{X}\}; 3) \( a \ell=\{\{\text{gray}(x,y) \mod M\}, \text{gray}(x,y) \in \{0...M-1\}, a \in \mathbb{R}\}. D I A 2 is applied to describe separated nucleus on images.

D I A 3 - a set \( F \) of operations \( f(U \rightarrow V) \) converting elements from a set of color images into elements of a set of gray scale images. The operands: elements of D I A 3 - operations \( f(U \rightarrow V) \in \mathbb{F} \); such transforms can be used for elimination luminance and color differences of images. The operations of addition, multiplication and multiplication by an element from the field of real numbers are introduced on the set of functions \( f \) as sequential operations of obtaining gray scale images and their addition, multiplication and multiplication by an element from the field of real numbers correspondingly: 1) \( f_1(I)+f_2(I)=J_1+J_2 \); 2) \( f_1(I) \cdot f_2(I)=I_1 \cdot J_2 \); 3) \( a(I)=\alpha I \). D I A 3 is applied to eliminate luminance and color differences of images.

D I A 4 - a set \( G \) of operations \( g(V \rightarrow P_1) \) for calculation of a gray scale image features. The operands: D I A 4 - a ring of functions \( g(V \rightarrow P_1) \in \mathbb{E}, P_1 \) - a set of P-models (parametric models). The operations. Operations of addition, multiplication and multiplication by a field element are introduced on a set of functions \( g \) as operations of sequential calculation of corresponding P-models and its addition, multiplication and multiplication by a field element. 1) \( g_1(I) \cdot g_2(I)=p_1(I)+p_2(I) \); 2) \( g_1(I) \cdot g_2(I)=p_1(I) \cdot p_2(I) \); 3) \( a(g(I)) = \alpha g(I) \). D I A 4 is applied to calculate feature values.

D I A 5 - a set \( P_1 \) of P-models. The operands: a set \( P_1 \) of P-models \( p=(f_1, f_2, ..., f_n) \) - gray scale image features, \( n \) - a number of features. The operations: 1) addition – an operation of unification of numerical image descriptions: \( p_1+\alpha p_2=(f_1, f_2, ..., f_n)+\alpha = (f_1, f_2, ..., f_n) \); \( n \) - a number of features of P-model \( p_1 \) plus a number of features of P-model \( p_2 \); 2) multiplication of \( P_1 \) - a number of coincident features of P-models \( p \); 3) \( f_1, f_2, ..., f_n \) - different features and coincident gray scale image features of P-models \( p_1 \) and \( p_2 \); 2) multiplication of \( 2 \) P-models – an operation of obtaining a complement of numerical image descriptions: \( p_1+\alpha p_2=(f_1, f_2, ..., f_n)+\alpha = (f_1, f_2, ..., f_n) \); \( n \) - a number of significant features of unified P-model of models \( p_1 \) and \( p_2 \), \( f_1, f_2, ..., f_n \) - significant features obtained after analysis of features of P-model \( p_1 \) and P-model \( p_2 \), \( f_1, f_2, ..., f_n \) may not belong to \( f_1, f_2, ..., f_n \) and may consist from feature combinations; 3) multiplication by a field element - operation of multiplication of a number, a vector, or a matrix by an element of the field: \( \alpha p=(\alpha f_1, f_2, ..., f_n) \). D I A 5 is applied to select informative features. The addition is applied for constructing joint parametric image representation. The multiplication is applied for reducing a set of image features to a set of...
significant features. The multiplication by an element from the field of real numbers is applied for feature vector normalization.

**DIA 6** - a set $P_2$ of P-models ($P_2$ includes feature vectors of the same length). **The operands** are: a set $P_2$ of P-models $p(J) = (f_1(J), f_2(J), ..., f_n(J))$, $n$ - a number of features, $f_1(J), f_2(J), ..., f_n(J)$ - gray scale image features, $f_1(J), f_2(J), ..., f_n(J) \in \mathbb{R}$. **The operations** of addition, multiplication and multiplication by a field element are introduced on the set $P_2$ as operations of a vector addition, multiplication and multiplication by a field element: 

1. \[ p(J_1) + p(J_2) = (f_1(J_1), f_2(J_1), ..., f_n(J_1)) + (f_1(J_2), f_2(J_2), ..., f_n(J_2)) = (f_1(J_1) + f_1(J_2), f_2(J_1) + f_2(J_2), ..., f_n(J_1) + f_n(J_2)) \]
2. \[ p(J_1) \cdot p(J_2) = (f_1(J_1), f_2(J_1), ..., f_n(J_1)) \cdot (f_1(J_2), f_2(J_2), ..., f_n(J_2)) = (f_1(J_1) \cdot f_1(J_2), f_2(J_1) \cdot f_2(J_2), ..., f_n(J_1) \cdot f_n(J_2)) \]
3. \[ \alpha p(J) = \alpha (f_1(J), f_2(J), ..., f_n(J)) = (\alpha f_1(J), \alpha f_2(J), ..., \alpha f_n(J)) \]

**DIA 6** is applied to describe images reduced to a recognizable form.

Table 1 shows all DIA with one ring used for describing the algorithmic scheme for solving the task of cytological image recognition.

### Table 1: DIAs with one ring used for describing algorithmic scheme for solving the task of cytological image recognition.

<table>
<thead>
<tr>
<th>Ring elements</th>
<th>Ring operations</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIA1</td>
<td>color images</td>
<td>algebraic operations of vector addition module $M$, vector multiplication module $M$ and taking an integral positive part of multiplication module $M$ by an element from the field of real numbers in each image point</td>
</tr>
<tr>
<td>DIG1</td>
<td>binary masks</td>
<td>operations of obtaining the binary mask that corresponds to indicated lymphocyte cell nuclei</td>
</tr>
<tr>
<td>DIA2</td>
<td>gray scale images</td>
<td>algebraic operations of gray functions addition module $M$, multiplication module $M$ and taking an integral positive part of multiplication module $M$ by an element from the field of real numbers in each image point</td>
</tr>
<tr>
<td>DIA3</td>
<td>operations reducing color images to gray scale images</td>
<td>sequential operations of obtaining gray scale images and their addition, multiplication and multiplication by an element from the field of real numbers</td>
</tr>
<tr>
<td>DIA4</td>
<td>operations of image feature calculation</td>
<td>sequential calculation of corresponding P (parametric)-models and its addition, multiplication and multiplication by a field element</td>
</tr>
<tr>
<td>DIA5</td>
<td>P-models</td>
<td>image algebra operations (union, complement, multiplication by real number)</td>
</tr>
<tr>
<td>DIA6</td>
<td>P-models</td>
<td>operations of a vector addition, multiplication and multiplication by a field element</td>
</tr>
</tbody>
</table>
Table 2: Database Statistics.

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Patient number</th>
<th>Image number</th>
<th>Nuclei number</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL</td>
<td>18</td>
<td>986</td>
<td>1639</td>
</tr>
<tr>
<td>TCLL</td>
<td>12</td>
<td>536</td>
<td>1025</td>
</tr>
<tr>
<td>CLL</td>
<td>13</td>
<td>308</td>
<td>2497</td>
</tr>
<tr>
<td>Total:</td>
<td>43</td>
<td>1830</td>
<td>5161</td>
</tr>
</tbody>
</table>

Footprints of lymphoid tissues were Romanovski-Giemsa stained and photographed with digital camera mounted on Leica DMRB microscope using PlanApo 100/1.3 objective (Fig. 1). The equivalent size of a pixel was 0.0036 mcm². 24-bit color images were stored in TIFF-format.

4.2 Reducing an Image to a Recognizable Form

The initial images were divided into 2 groups: training image set \( \{ I_j \}_{j=1..n} \) and recognition image set \( \{ I_j \}_{j=n/2+1..n} \). The steps 1.1-1.6 of stage 1 “Reducing an image to a recognizable form”) are described below as follows: description, step operands, step operations, results of step operation applying. It will be highlighted by letters ‘a’ and ‘b’ where processing of training and recognition sets differs.

Step 1.1: Obtaining Masks of Diagnostically Important Nucleus on Images. Application of segmentation algorithm is described by operands \( sb((U,C) \rightarrow U') \in B \) of DIG1. An algorithm \( sb((U,C) \rightarrow U') \in B \) is applied to initial images in order to obtain corresponding mask (equation 3).

\[
\{ I_j \}_{j=1..n} \xrightarrow{sb(DIG1)} \{ B_j \}_{j=1..n} \tag{3}
\]

Step operands are initial images \( \{ I_j \}_{j=1..n} \) and contours of lymphocyte cell nucleus. Step operation is an operation described by DIG1. Such description gives flexibility for using different kind of segmentation algorithms. The applied algorithm of threshold segmentation was supplemented by morphological processing of derivable nuclei images in order to obtain a corresponding mask.

Results of operation applying are binary masks \( \{ B_j \}_{j=1..n} \) represented as operands of DIG2.

Step 1.2: Segmentation of Diagnostically Important Nucleus on Images. The mask multiplication by an initial image gives indicated nuclei image (equation 4).

\[
\{ I_j \}_{j=1..n} \{ B_j \}_{j=1..m} \xrightarrow{\text{operation}} \{ M_{T} \}_{j=1..m} \tag{4}
\]

Step operands are initial images \( \{ I_j \}_{j=1..n} \) and binary masks represented as operands of DIG2.

Step operation is an operation of multiplication of 2 operands of DIA1. All initial images were multiplied by corresponding binary masks.

The results of the operation are T(transformational)-models \( \{ I'_j \}_{j=1..m} \) of initial images.

Step 1.3: Reducing Color Images to Gray Scale Images. To compensate different illumination conditions and different colors of stain the specimen images were processed before feature values calculation (equation 5).

\[
\{ I_j \}_{j=1..m} \xrightarrow{f} \{ M_{T} \}_{j=1..m} \tag{5}
\]

Step operands are image models \( \{ I'_j \}_{j=1..m} \). Step operations are described by the elements of the DIA 2. Such representation gives flexibility for using different kinds of processing operations. Here the function \( f(U \rightarrow V) \in F \) (DIA 2 element) has a form \( (I=\{(r(x,y),g(x,y),b(x,y)),r(x,y),g(x,y),b(x,y)\in[0..M-1]\}_(x,y)) \rightarrow f(I)=\{\text{gray}(x,y)\}_{(x,y)\in X,(x,y)\in[0..M-1]}, \text{gray}(x,y)=g(x,y) \frac{2B}{M}, B \text{ - an average brightness of a blue component of an initial RGB-image. The green tone in this case is the most informative.}

The results of the operation are T-models \( \{ I'_j \}_{j=1..m} \).
Step 1.4a: Feature Calculation on Constructed Image Models of the Training Set. To calculate different features the training set were processed by different operations of DIA 4 (equation 6) \( (m_1 \text{ equals to a number of segmented nucleus in training set}) \).

\[
\{I_j^2\}_{1\cdots m_1} \xrightarrow{\text{DIA4}} \{M^p_1(I_j^2)\}_{1\cdots m_1} = \{M^p_1(j)\}_{m_1}
\]

(6)

Step operands are image models \( \{I_j^2\}_{1\cdots m_1} \).

Step operations are described by the elements of DIA 4. To describe each image \( 47 \) or \( 14 \) features were selected for describing each of the images: the size of nucleus in pixels, 4 statistical features calculated on the histogram of nucleus intensity, 16 granulometric and 26 Fourier features of nucleus. \( M^p_1(j) \) is the vector with dimension 47 for each image model \( I_j^2, j=1\cdots m_1 \).

The results of the operation are P-models \( \{M^p_1(j)\}_{1\cdots m_1} \).

Step 1.5a: Selection of Informative Features. This is an additional step of image model reduction. As it will be shown below the recognition algorithm was applied to both a full model \( M^p_1(j) (j=1\cdots m_1) \) and a reduced model \( M^p_2(j) (j=m_1+1\cdots m) \). At this step the constructed descriptions of images from the training set are studied for selecting the most informative features (equation 7).

\[
\{M^p_1(j)\}_{1\cdots m_1} \xrightarrow{\text{DIA4}} \{M^p_2(M^p_1(j))\}_{1\cdots m_1} = \{M^p_2(j)\}_{m_1}
\]

(7)

Step operands are image models \( \{M^p_1(j)\}_{1\cdots m_1} \).

Step operations are described by the elements of DIA 5. Operations of addition and multiplication are introduced for unificating and for reducing sets of image features to a set of significant features. Operation of multiplication by an element from the field of real numbers is introduced for normalization of feature vectors. Such representation gives flexibility for using different kinds of feature analysis to obtain a reduced set of features. Application of factor analysis to training image set detected 14 features with the largest loads in the first and second factor (Gurevich, 2006).

The results of the operation are P-models \( \{M^p_2(j)\}_{1\cdots m_1} \) - a the vector with dimension 14 for each of image models \( I_j^2, j=1\cdots m_1 \).

Step 1.6b: Feature Calculation on Constructed Image Models of the Recognition Set. The steps 1.4 and 1.5 obtain a multi-model representation for training set. The step 1.6 is the step of feature values calculation for a recognition set (equation 8).

\[
\{I_j^2\}_{m_1+1\cdots m} \xrightarrow{\text{DIA4}} \{M^p_1(I_j^2)\} \cup \{M^p_2(M^p_1(I_j^2))\}_{m_1+1\cdots m} = \{\Psi(j)\}_{m_1+1\cdots m}
\]

(8)

Step operands are image models \( \{I_j^2\}_{m_1+1\cdots m} \).

Step operations are described by the elements of DIA 4. To describe each image 47 or 14 features were selected.

The results of the operation are P-models \( \{\Psi(j)\}_{m_1+1\cdots m} \) (note that the multi-model representation of images was constructed).

4.3 Training and Recognition

The class "Algorithms Based on Estimate Calculations" (AEC-class) were chosen as recognition algorithms since they can be conveniently represented by algebraic tools (Zhuravlev, 1998).

Initial Data. DIA 6 and its operands \( \Psi(j) = M^p_1(U_j^2) \cup M^p_2(M^p_1(U_j^2)) \) \( (j=1\cdots m) \) describe initial data for recognition algorithm \( A \) \( (\Psi(j) = \{\psi_1, \psi_2, \cdots, \psi_n\}) \) - feature vector with a dimension \( n=47 \) or \( n=14 \), \( \{\Psi(j)\}_{m_1+1\cdots m} \) - information about recognition set, \( \{\Psi(j)\}_{m_1+1\cdots m} \) - information about training set, \( \{P^r_j\}_{m_1} = \{a_{gr}\}_{m_1} \) - information about memberships of training set images to classes \( \{K_{s1\cdots r}\}_{r=3} \ (a_{gr} \in \{0,1\}, \ r=3, \ \{I_j^2\}_{m_1} \) - initial specimen images, one image for each indicated nucleus). Recognition algorithm
\( \mathcal{M}(\Psi(j))_{m\times n}, \{a_{gj}\}_{n\times m}, \{\Psi(j)\}_{m\times n} = \{a_{gj}\}_{r\times(n-m)} \), solves an image recognition problem, \( \{a_{gj}\}_{r\times(m-n)} \) - an information vector of image model \( I_j \) calculated by algorithm A (\( j=m_1+1\ldots m \)).

The algorithms were applied to both full image models \( M^f_1(j) \) (\( j=1\ldots m \), 47 features) and reduced image models \( M^r_1(j) \) (\( j=1\ldots m \), 14 features).

**Algorithmic Scheme.** We described the main steps and elements of an algebraic model of information technology for automation of diagnostic analysis of cytological specimens of patient with lymphatic system tumors (Fig. 2):

\[
\begin{align*}
(I_j)_{m\times n} & \xrightarrow{\mathcal{B}_j} \{M^r_1(j)\}_{n\times m} \\
\{M^r_1(j)\}_{m\times n} & \xrightarrow{\mathcal{B}} \{M^f_1(j)\}_{m\times n} \\
\{M^f_1(j)\}_{m\times n} & \xrightarrow{\mathcal{B}_{P_2}} \{\Psi_2\}_{m\times n} \\
\{\Psi_2\}_{m\times n} & \xrightarrow{\mathcal{B}_{P_2}} \{a_{gj}\}_{r\times(m-n)}
\end{align*}
\]

Figure 2: Algorithmic scheme of information technology.

**Discussion of the Results.** The elements of the technology were tested via software system «Recognition 1.0» (Zhuravlev, et al., 2005) including AEC-algoritms. It appeared that the best results are achieved by voting using all possible support sets, while automatic selection of support set cardinality and selection of support sets of fixed cardinality give lower precision.

Recognition rate for full feature set (Table 3) is 86.75%, while the rates differ for different recognition classes. High recognition rate for CLL (97.84%) is probably connected with innocent nature of CLL as opposed to CL (63.35%) and TCLL(84.51%) - the malignant cases. Thus, cells of CLL have evident distinctions from cells of other diagnoses, and cells of CL and TCLL are more similar to each other.

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>The number of correctly recognized cells</th>
<th>Total number of cells</th>
<th>The recognition rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL</td>
<td>693</td>
<td>820</td>
<td>84.51%</td>
</tr>
<tr>
<td>TCLL</td>
<td>325</td>
<td>513</td>
<td>63.35%</td>
</tr>
<tr>
<td>CLL</td>
<td>1221</td>
<td>1248</td>
<td>97.84%</td>
</tr>
<tr>
<td>Total cell set</td>
<td>2239</td>
<td>2581</td>
<td>86.75%</td>
</tr>
</tbody>
</table>

The recognition rate reduced feature set (14 features) decreased to 83.18% (Table 4). This feature set includes the following features: size of nucleus in pixels, average by intensity histogram (statistic feature), numbers of elements with typical and minimal size of nuclei (granulometric features), 9 Fourier-features of nucleus.

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>The number of correctly recognized cells</th>
<th>Total number of cells</th>
<th>The recognition rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL</td>
<td>626</td>
<td>820</td>
<td>76.34%</td>
</tr>
<tr>
<td>TCLL</td>
<td>300</td>
<td>513</td>
<td>58.48%</td>
</tr>
<tr>
<td>CLL</td>
<td>1221</td>
<td>1248</td>
<td>97.84%</td>
</tr>
<tr>
<td>Full cell set</td>
<td>2147</td>
<td>2581</td>
<td>83.18%</td>
</tr>
</tbody>
</table>

The software system «Recognition 1.0» (Zhuravlev, 2005), used for experimental investigation, includes effective realization of AEC methods and allows to apply them for practical task solution. It was experimentally verified that the best results are achieved by voting using all possible support sets, while automatic definition of support set capacity and definition of fixed support set capacity give lower precision.

**5 CONCLUSIONS**

The paper demonstrates practical application of algebraic tools of the Descriptive Approach to Image Analysis - it is shown how to construct a model of a technology for automation of diagnostic analysis on...
images using. It is presented an algorithmic scheme of a technology implementing the apparatus of DIA. The paper solves a dual task: it presents technology by well-structured mathematic model it shows how DIA can be used in image analysis application. The described techniques and tools will be used for creating software implementation of the technologies, its testing and performance evaluation.

ACKNOWLEDGEMENTS

This work was partially supported by the Russian Foundation for Basic Research Grants Nos. 05-01-00784, 06-01-81009, 07-07-13545, by Collaborative project “Image Analysis and Synthesis: Theoretical Foundations and Prototypical Applications in Medical Imaging” within agreement between CNR (Italy) and the RAS, by the project “Descriptive Algebras with one ring over image models” of the Program of Basic Research “Algebraic and Combinatorial Techniques of Mathematical Cybernetics” of the Department of Mathematical Sciences of the RAS, by the project no. 2.14 of the Program of the Presidium of the Russian Academy of Sciences “Fundamental Problems of Computer Science and Information Technologies”.

REFERENCES


