Myelogenous Leukemia Cell Image Preprocessing for Feature Generation

Tomasz Markiewicz, Stanisław Osowski, Leszek Moszczyński, Robert Sałat1

Abstract —**The paper presents the methods of preprocessing the leukemic blast cells in order to generate the features well characterizing different types of cells. The solved problems include: the segmentation of the bone marrow aspirate by applying the watershed transformation, selection of individual cells, texture analysis and description of the geometrical nuclear features description of the cell.**

I. Introduction

The recognition of the blast cells in the bone marrow of the patients suffering from myelogenous leukemia is a very important step in the recognition of the development stage of the illness and proper treatment of the patients [3,4]. The acute leukemia is a disease of the leukocytes and their precursors. It is characterized by the appearance of immature, abnormal cells in the bone marrow and peripheral blood. The aspirated marrow is found to be infiltrated by abnormal cells. We can find different cell lines in the bone marrow: the megacaryocytic series, erythrocytic series, monocytic series and granulocytic series. The most known and recognized abnormal cells include monoblasts, promonocytes, monocytes, myeloblasts, promyelocytes, myelocyte, metamyelocytes, proerythroblasts, basophilic erythroblast, polychromatic erythroblast, eosinophilic erythroblast, megacaryoblasts, promegacaryocytes and megacaryocytes [3,4].

The percentage of blasts is a major factor at defining various subtypes of acute myeloid leukemia. According to French-American-British (FAB) standard 8 acute leukemia subtypes are classified as follows [4]:

M0-M5: >30% myelo/monoblasts determined by morphology or immunophenotyping

M6: >50% erythroid precursors, > 30% blasts in non-erythroid cells

M7: >30% megacarioblasts present.

Thus the proper treatment of leukemia patients requires not only recognition of different stages of the development of the blasts but also tedious calculation of their quantity in the aspirated bone marrow. Automatic segmentation of the image of the aspirate, leading to the selection of different types of cells, can accelerate the recognition process and at the same time make it more accurate.

This paper is dedicated to the task of the blast cell recognition. The first section describes the segmentation of the image that will lead to the selection of the individual blast cells of the aspirate. The next section describes the preprocessing of the images of individual cells in order to generate the features, characteristic for different types of cells in a way enabling high recognition rate among them. The results of numerical experiments conducted on different blasts will be presented and discussed through the whole paper.

II. Automatic Segmentation of the Image for Separation of Individual Cells

Image segmentation is a division of the image into different regions, each having certain properties. In a segmented image, the picture elements are no longer the pixels, but connected set of pixels, all belonging to the same region. We will use segmentation techniques to separate the individual cells from the set of cells creating the image.

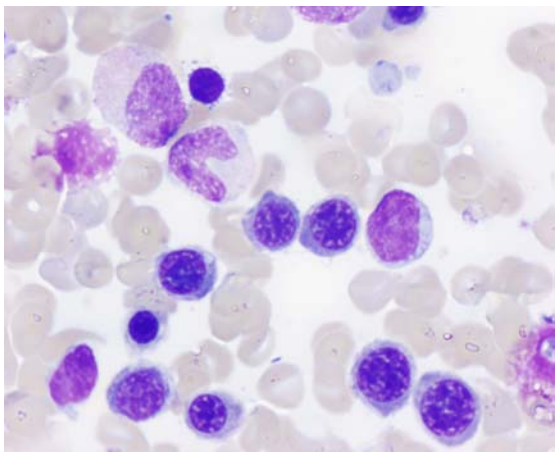

Fig. 1 The image of the bone marrow aspirate of the acute leukemia patient (most cells are myeloblasts)

The recognition and separation of individual cells from the image of the blood cells is a very difficult task, since different regions are of little grey level variations and the borders of individual cells are hardly visible.

[1] T. Markiewicz and S. Osowski are with the Institute of the Theory of Electrical Engineering and Electrical Measurements, Warsaw University of Technology, S. Osowski is also with Military University of Technology, Warsaw, email: sto@iem.pw.edu.pl

L. Moszczyński is with the Institute of Material Technology, Warsaw University of Technology.

R Sałat is with Agriculture University of Technology, Warsaw

The work is supported by KBN as the research grant 2002 - 2005

Fig. 1 presents the exemplary image of the blast cells.

The individual cells are close to each other and the borders among them are hardly visible. The task of segmentation of the image is focused on the automatic recognition and separation of each cell for further processing, in order to obtain the stable features, useful in recognition of the type of the cell. In solving the segmentation task we have used the morphological operations

Morphological operations aim at extracting relevant structures of the image by probing the image with another set of a known shape called structuring element (SE), chosen as the result of prior knowledge concerning the geometry of the relevant and irrelevant image structures. The most known morphological operations include: erosion, dilation, opening and closing [1].

The morphological approach to image segmentation combines regions growing and edge detection techniques. It groups the pixels around the regional minima of the image. The boundaries of adjacent grouping are precisely located along the crest lines of the gradient image. In our experiment we have accomplished this through an operation called the watershed transformation.

The watershed transformation takes its origin from the topographic interpretation of the gray scale image. According to the law of gravitation, the water dropped on such surface will flow down until it reaches a minimum. The whole set of points of the surface, whose steepest slope paths reach a given minimum, constitutes the catchment's basis associated with this minimum. The watersheds are the zones dividing adjacent catchment's basins.

In numerical implementation of the watershed algorithm the original image is transformed so, as to output an image whose minima mark relevant image objects and whose crest lines correspond to image object boundaries. In this way the image is partitioned into meaningful regions that may correspond for example to the individual blast cells. In our experiments we have used the watershed algorithm implemented by using Matlab platform]8].

The applied procedure of the image segmentation and cell separation is as follows:

transformation of the original color image into gray scale (subtraction the green color from the red)

transformation of the gray image to binary image by applying the biased segmentation

application of closing and erosion operations to smooth the contours and to eliminate the distortions

generation of the map of distances from the black pixel to the nearest white pixels

application of the watershed algorithm to image segmentation.

The results of such image preprocessing are the image segmented into the regions corresponding to individual cells, with the visible borders between cells.

Fig. 2 presents the results of the segmentation of the exemplary image of the aspirated bone marrow. The dark regions represent different blasts. All of them have been separated as individual regions. Most cells have been separated properly. Some small errors are visible, especially at the border of chromatine, but the number of such errors is very limited. The blast cells are composed mostly of nuclei. Their shapes are easily visible and have been reconstructed by the algorithm. The slight problem may be in reconstruction of the chromatin shapes, since the color of chromatin is very light and at the packed blasts the real border between chromatin of different cells is really unknown.

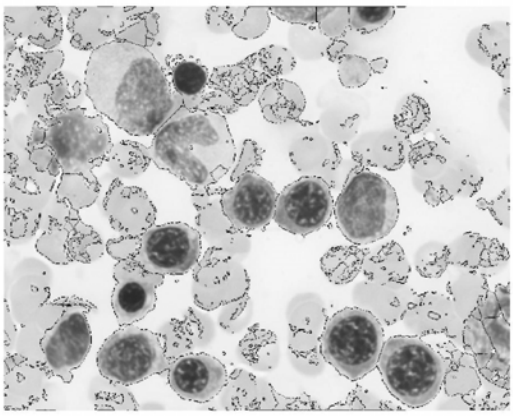

Fig. 2 The segmented image of the aspirated bone marrow presented in Fig. 1.

III. TEXTURE FEATURE DESCRIPTION

Texture refers to the arrangement of the basic constituents of a material. In digital image the texture is depicted by the interrelationships between spatial arrangements of the image pixels. They are seen as changes in intensity patterns, or gray tones.

The efficient recognition of the texture images requires the preprocessing of them in order to extract the features, characterizing the image in a way suppressing the differences within the same class and enhancing the differences between textures belonging to different classes. There are many different techniques of texture preprocessing for extraction of such features [5,6]. The most common include: the Haralick gray level co-occurrence matrix features (GLCM), Markov random field features, Unser sum and difference histograms, Gabor transformation features, wavelet or fractal descriptions, etc [5,6]. Here we will limit ourselves to the first four preprocessing methods.

The Haralick description of texture is based on the co-occurrence matrix, which counts the number of pixel pairs in an image that have values g and g' and are separated by a pixel distance d in a relative direction α.

Using α=0o, 45o, 90o and 135o we get the so called 4 neighbors of a pixel. Using the co-occurrence matrix the following coefficients are calculated: the angular second momentum, contrast, correlation, sum of squares, inverse difference moment, sum average, sum variance, entropy, sum entropy, difference variance, information measure of correlation, maximum correlation coefficient. In all , there are 14 features generated on the basis of the co-occurrence matrix.

Unser algorithm of feature generation is the simplified version of the Haralick method. The sum and difference histograms of gray levels of the neighboring pixel are created for different directions, usually 0o, 45o, 90o and 135o. The features used in texture representations are the mean, second order moment, contrast and entropy. They are calculated independently for the sum and difference at 4 mentioned above directions. This makes 32 features altogether [5].

In Markov random field method the texture is characterized by determining the mean gray level of the neighborhood of each pixel. Each image is characterized by the global set of parameters approximating the gray levels of the whole texture. In this way we create 11 features of the texture image [5].

In Gabor descriptors we measure the similarity between the so-called Gabor mask (sinusoidal wave modulated by gaussian function) and its neighborhood. The mask is generated for different wavelength, angle, phase shift and variance of the gaussian function. For different combinations of these parameters we calculate the energy of the image. The energy values form the features of the texture. Taking for example the angles equal 0o, 45o, 90o and 135o and 3 different wavelengths we get 12 features, used in our experiments.

Different methods of feature generation can produce different sets of features. Some of them represent the textures of different blast cells in less or more distinctive way. Usually after generating the features for the set of images, the best are chosen and the rest is discarded.

IV. THE NUCLEAR GEOMETRICAL FEATURES.

The important information concerning the blast cells is contained in the geometrical shapes and parameters associated with them. It is known from the observation that various cells differ greatly with the size. For example the eosinophilic erythroblasts have the size of 8-10 micrometer, while megakcariocyte may be up to 100 micrometer. The shapes of different blasts are either round, oval or kidney-shaped. We have used the following geometrical features of the cells, limiting to their nuclei:

- radius –measured by averaging the length of the radial line segments defined by the centroid and the border points
- perimeter - the total distance between consecutive points of the border
- area – the number of pixels on the interior of the cell and adding one-half of the pixels on the perimeter
- compactness – measured by the formula: perimeter2/area
- concavity – the severity of concavities or indentations in a cell (the extend to which the actual boundary of a cell lies on the inside of each chord between non-adjacent boundary points)
- concavity points – the number of concavities, irrespective of their amplitudes
- symmetry – the length difference between lines perpendicular to the major axis to the cell boundary in both directions.

These features have been implemented in Matlab, by applying the Image Processing Toolbox.

V. THE RESULTS OF NUMERICAL EXPERIMENTS

In numerical experiments we have preprocessed different sets of blasts in an attempts to find the best representation of the features. Below we will present the numerical results for 4 different blasts: neutrophilous band, basophilous erythroblast, polychromatic erythroblast and mono/myelo blast. The selected typical representations of 4 blasts under considerations are presented in Fig. 3

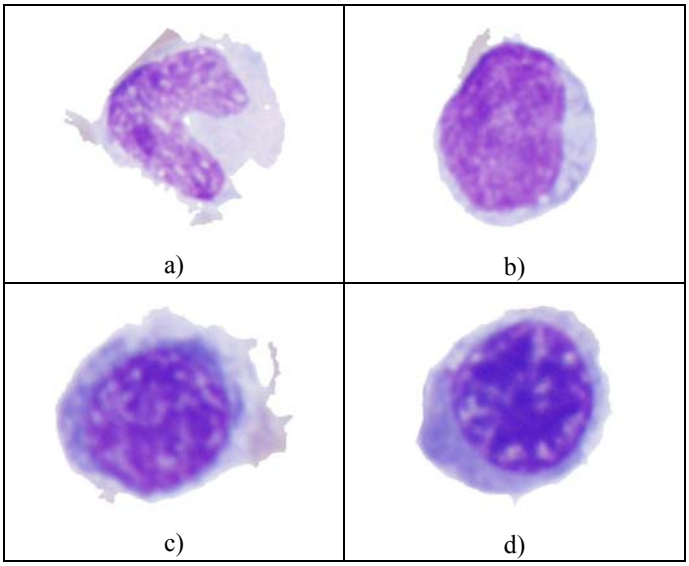

Fig. 3 The blast cells used in numerical experiments of feature generation. Each picture represents different blast type: a) neutrophilous band, b) mono/myelo blast c) basophilous erythroblast d) polychromatic erythroblast

. They have been selected from the image of Fig. 2. As you can see the cells are of different shapes and different area. The most prominent part of the cell forms the nucleus of very special texture and shape (the dark

part of the cell). The chromatin corresponds to the light color of the cell.

TABLE I

MARKOV FEATURES DESCRIPTION OF FOUR DIFFERENT BLAST TYPES

C1	C2	C3	C4
-0.0184	0.0090	-0.0353	-0.0157
-0.0960	-0.0508	-0.0118	-0.0253
0.0420	-0.0467	-0.0342	-0.0372
-0.0103	0.0130	0.1000	0.0518
0.2520	0.3613	0.2542	0.2758
-0.0279	0.0137	-0.0252	-0.0022
0.2877	0.2414	0.2734	0.2796
-0.0340	-0.0352	0.0031	-0.0273
0.0980	0.0066	0.0047	0.0122
-0.0382	-0.0098	-0.0258	-0.0085
1.2585	0.8826	0.6735	0.5789

The experiments have been carried out for the nuclei of these chosen cells. Table 1 shows the set of typical texture Markov features describing each blast.

The column vectors is the set of features. We used Markov features as the most representative for the task. C1 is related to neutrophilous band, C2 – to mono/myelo blast, C3 – to basophilous erythroblast and C4 – to polychromatic erythroblast .

As we can see the features are different in each type of blast cells. Especially high differences can be observed in the last 11th feature. Such distribution of the features means usually good recognition capabilities of different types of blasts.

Table 2 depicts differences among geometrical nuclear features for the same set of blast cells. The following rows of the table represent: radius, perimeter, area, concavity, concavity points, symmetry and compactness of the nuclei (all given in pixels).

TABLE II

THE GEOMETRICAL FEATURES DESCRIBING FOUR DIFFERENT BLAST TYPES

Feature	C1	C2	C3	C4
Radius	89.35	96.42	85.37	81.12
Perimeter	2105	1005	1080	1067
Area	31053	29094	22867	21824
Concavity	945	32	200	211
Concavity points	101	117	71	93
Symmetry	9014	1717	1632	1578
Compactness	142.70	34.72	50.99	52.17

As we can see there is also great variation between the values of these features for different blast types. Combining together the texture and geometrical nuclear features we get the set of features capable of good recognition among different types of blasts. The final recognition of blast cells, relied on these sets of features, may be performed easily using either neural, Bayes of competitive classifiers [9,10].

VI. CONCLUSIONS

This paper has presented efficient method of myelogenous leukemia cell image preprocessing in order to obtain features characteristic for individual cells in a way enabling good recognition among them. The solved problems include:

- application of watershed transformation for segmentation of the bone marrow aspirate image into individual blast cells
- different transformations of the cells for separation of the chromatin and nucleus
- generation of texture features describing the cell in a most distinctive way
- comparison of different sets of features characterizing the blast cells.

Results presented in this paper have shown, that method of image preprocessing as demonstrated in our experiments may find practical application in automatic computer aided blast cells recognition.

REFERENCES

[1] P. Soile, Morphological image analysis, principles and applications, Springer, 2003, Berlin
[2] W. Wolberg, W. N. Street, O. L. Mangasarian, Machine learning to diagnose breast cancer from image-processed nuclear features of fine needle aspirates, 1994, Internal report of University of Wisconsin
[3] K. Lewandowski, A. Hellmann, Atlas hematologiczny, Medyczne Wydawnictwo Multimedialne, Gdańsk, 2001
[4] K. Janicki, Hematologia kliniczna, PZWL, Warsaw, 1991
[5] T. Wagner, Texture analysis, (in Jahne, B., Haussecker, H., and Geisser P., (Eds.), Handbook of Computer Vision and Application), Academic Press, 1999, pp. 275-309
[6] T. Reed, J. Buf, A review of recent texture segmentation and feature extraction techniques, CVGIP: Image Understanding, 1993, vol. 57, pp. 359-372
[7] S. Osowski, Tran Hoai Linh, ECG beat recognition using fuzzy hybrid neural network, IEEE Trans. on Biomedical Engineering, vol. 48, pp. 1265-1271, 2001
[8] Matlab user manual, Natick, 1999
[9] S. Osowski, Sieci neuronowe do przetwarzania informacji, Oficyna Wydawnicza, 2001
[10] S. Haykin, Neural networks, comprehensive foundation, Prentice Hall, 1999, New Jersey