

UNSUPERVISED SEGMENTATION OF MICROSCOPIC WHITE BLOOD CELLS IMAGES USING HISTOGRAM EQUALIZATION TECHNIQUE

Fatin A. Dawood

Department of Computer Science, College of Science, University of Baghdad. Baghdad-Iraq.

Abstract

In this paper, a new method is presented to segment the microscopic images of White Blood Cells (WBC) into cell and non-cell regions depending on the image histogram. The proposed method involves of: first, separating the captured image into its three color channels (Red, Green, and Blue) to produce three grayscale files one for each. Second, apply the histogram equalization upon the three files to obtain the white blood cells image to be segmented into white background, black nucleus with color stains inside and surrounded by cytoplasm. Experimental results show that the proposed unsupervised segmentation method have given reliable results comparable with thresholding method.

الخلاصة

يقدم هذا البحث طريقة جديدة لتقسيم الصور المجهرية لكريات الدم البيضاء الى مناطق الخلية وغير الخلية بالاعتماد على الجدول التكراري للصورة. هذه الطريقة تتضمن : اولاً، فصل الصورة المكتسبة الى ثلاث قنوات لونية (احمر، اخضر ، وازرق) والنتيجة هي ثلاث ملفات رمادية واحدة لكل منهم. ثانياً ، تطبيق عملية equalization للجدول التكراري على الملفات الثلاثة للحصول على صورة كريات الدم البيضاء التي قسمت الى خلفية بيضاء ، نواة سوداء بلطخات لونية داخلها ومحاطة بسايتوبلازم . أظهرت النتائج بان الطريقة المقترحة للانقسام غير المشرف عليه أعطت نتائج موثوقة مقارنة بطريقة Thresholding.

Introduction

Image segmentation is important in many vision and image processing applications. The aim of image segmentation is to divide the scene into significance regions. The adjacent similar pixels will be grouped together into the same region. Then the interested objects will be isolated from each other. Therefore the segmented image can be applied to widely application such as feature classification, object detection, motion estimation, data compression, etc [1].

The human blood has five major types of White Blood Cells (WBC) currently of interest. They can be divided into two major groups, distinguished by the presence or absence of granules in the cytoplasm (cell body). There are two major types of WBC without granules; these are **Lymphocytes** and **Monocytes**. The other three major types are **Neutrophils**, **Basophils**

and **Eosinophils** which differ in the way their cytoplasm granules are affected by various stains [2]. Most WBC types consist of one or more areas of nuclear matter, surrounded by an area of cytoplasm and that an area of nuclear matter cannot be shared between two or more cells. This nuclear material can be recognized visually by the fact that it is distinctly darker and purple in color. Maximum and minimum areas of cytoplasm and nucleus are set depending on the magnification of the image acquisition system [3].

The analysis of microscopic blood slides is a powerful tool in determining the health status of an individual and could detect several diseases. The count and shape, lineage and maturity level of white and red blood cells (RBC) could aid in the diagnosis of diseases like "leukemia" [4] [5]. The image segmentation can be used to divide the blood image into regions that presumably

correspond to structural units such as WBC, RBC, platelets and others. Many automated techniques were proposed to overcome the time consuming task of human experts in counting and classifying WBC.

Segmenting and classifying WBC was shown to be a difficult task due to various reasons including cell touching, close cell/ background intensities. In many of the researches presented in literatures, automatic cell segmentation was avoided to decouple the error due to segmentation with that of classification [6][7]. The main steps of classifying WBC include segmentation of such cells; extract features for each type, then classify the WBC type based on features detected. Various techniques were used for the segmentation stage including: background subtraction, histogram manipulation and thresholding [8].

The aim of the proposed method is to develop an automatic segmentation system for detecting and defining the boundaries of WBC from microscopic blood slides as distinct area. The segmentation process is based on using histogram equalization to detect every element in the blood slide and cut out the WBC segment from all other elements by color concentration.

The results of new method are compared with those of thresholding method. The way used in thresholding to produce a binary bitmap image is:

$$G(x,y) = \begin{cases} 1 & \text{if } f(x,y) \geq T \\ 0 & \text{if } f(x,y) < T \end{cases} \quad (1)$$

Where $f(x, y)$ is the gray level of point (x, y) , and T denotes the threshold value. Thus pixel labeled '1' (i.e. white) or any other convenient intensity level corresponds to objects, whereas pixel labeled '0' corresponds to the background. This technique can have problems where the lighting level varies from one image to another [9].

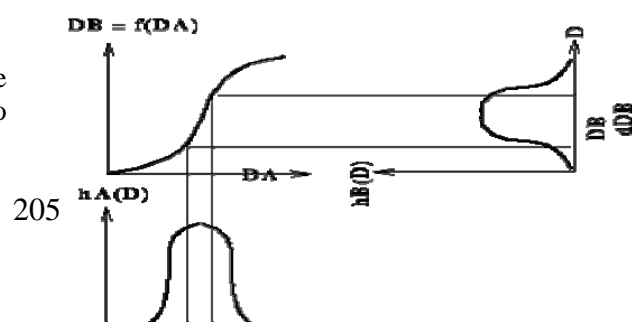
The rest of this paper is organized as follows: section two presents the description of the histogram equalization, section three illustrates the proposed method, section four presents and discusses some results and finally section five demonstrates the work conclusions.

Histogram Equalization

In an image processing context, the histogram of an image normally refers to

a histogram of the pixel intensity values. This histogram is a graph showing the number of pixels in an image at each different intensity value found in that image. For an 8-bit grayscale image there are 256 different possible intensities, and so the histogram will graphically display 256 numbers showing the distribution of pixels amongst those grayscale values. Histograms can also be taken for color images either individual histogram of red, green and blue channels can be taken, or a 3-D histogram can be produced, with the three axes representing the red, blue and green channels, and brightness at each point representing the pixel count [10].

Histogram modeling techniques (i.e. histogram equalization) provide a sophisticated method for modifying the dynamic range and contrast of an image by altering that image such that its intensity histogram has a desired shape. Unlike contrast stretching, histogram modeling operators may employ *non-linear* and *non-monotonic* transfer functions to map between pixel intensity values in the input and output images. Histogram equalization employs a monotonic, non-linear mapping which re-assigns the intensity values of pixels in the input image such that the output image contains a uniform distribution of intensities (i.e. a flat histogram). This technique is used in image comparison processes (because it is effective in detail enhancement) and in the correction of non-linear effects introduced by, say, a digitizer or display system. Histogram modeling is usually introduced using continuous, rather than discrete, process functions. Suppose that the image of interest contains continuous intensity levels (in the interval $[0,1]$) and that the transformation function f which maps an input image $A(x, y)$ onto an output image $B(x, y)$ is continuous within this interval. Further, it will be assumed that the transfer law (which may also be written in terms of intensity density levels, e.g. $D_B = f(D_A)$) is single-valued and monotonically increasing (as is the case in histogram equalization) so that it is possible to define the inverse law $D_A = f^{-1}(D_B)$. An example of such a transfer function is illustrated in Figure (1).



intensity levels). These effects are likely to decrease as the number of pixels and intensity quantization levels in the input image are increased [11].

The proposed method

The proposed method to segment WBC can be stated as a series of stages, as shown in Figure (2):

All pixels in the input image with densities in the region D_A to $D_A + dD_A$ will have their pixel values re-assigned such that they assume an output pixel density value in the range from D_B to $D_B + dD_B$. The surface areas $h_A(D_A)dD_A$ and $h_B(D_B)dD_B$ will therefore be equal, yielding:

$$h_B(D_B) = h_A(D_A) / d(D_A) \tag{2}$$

where $d(x) = df(x) / dx$.

This result can be written in the language of probability theory if the histogram h is regarded as a continuous probability density function p describing the distribution of the (assumed random) intensity levels:

$$p_B(D_B) = p_A(D_A) / d(D_A) \tag{3}$$

In the case of histogram equalization, the output probability densities should all be an equal fraction of the maximum number of intensity levels in the input image, D_M (where the minimum level considered is 0). The transfer function (or point operator) necessary to achieve this result is simply:

$$d(D_A) = D_M * p_A(D_A). \tag{4}$$

Therefore, a digital implementation of histogram equalization is usually performed by defining a transfer function of the form:

$$f(D_A) = \max(0, \text{round}[D_M * (n_k / N^2)] - 1) \tag{5}$$

Where N is the number of image pixels and n_k is the number of pixels at intensity level k or less. In the digital implementation, the output image will not necessarily be fully equalized and there may be 'holes' in the histogram (*i.e.* unused

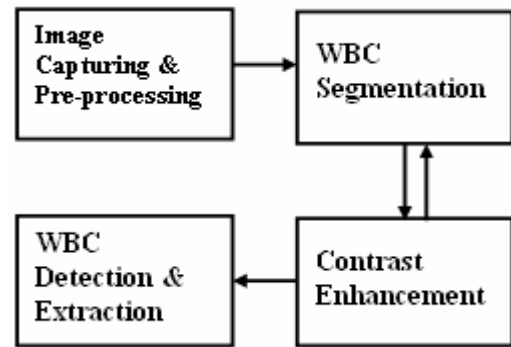


Figure (2): WBC segmentation stages

• **Image Capturing & Pre-processing**

The data acquisition of microscopic images consists of many steps: smearing blood onto microscopic slides, cover slipping them, viewing the resultant slides and capturing images including WBC from them. The captured image requires a pre-processing that separates it into its three color channels (Red, Green and Blue) and the result was three grayscale files one for each, as shown in Figure (3).

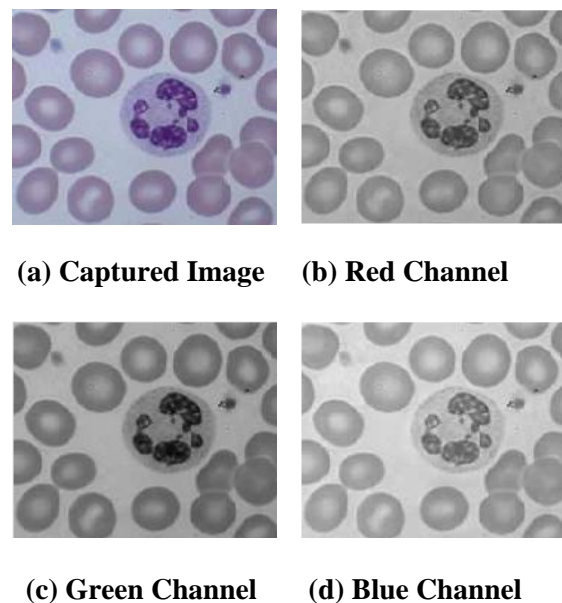


Figure (3): Separating Captured RGB Image

As it can be seen from figure, the green component of the RGB input image gives the best contrast between the background and the blood cells components including RBC, cytoplasm and nucleus of WBC.

• WBC Segmentation

To segment the desired WBC object from the background, apply the histogram equalization upon the three grayscale files. The histogram equalization process consists of three steps as follows:

1. Find the running sum of the histogram values for each file (i.e. channel).
2. Normalize the values (of step 1) by dividing them over the total number of pixels.
3. Multiply the values (of step 2) by the maximum gray level value and round.

Histogram Equalization (HE) process will be repeated twice. In the first iteration, the original image is segmented into white background, dark nucleus and cytoplasm with green color. In the second iteration, the resultant image has a black nucleus with color stains inside and surrounded by white cytoplasm.

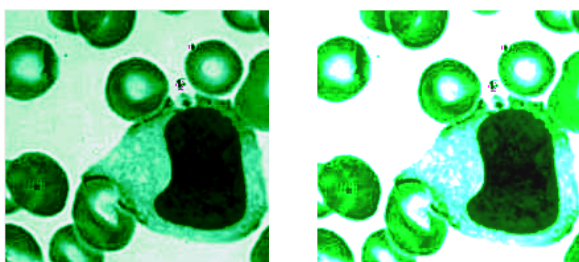
• Contrast Enhancement

To enhance the brightness of the resultant image produced from the previous stage (i.e. 1st & 2nd iterations of histogram equalization); multiply each pixel value by a factor greater than one. This process illustrated by the following transformation:

$$V_{out} = V_{in} * B \quad B > 1 \quad (6)$$

$$V_{out} = V_{maxvout} \quad \text{if} \quad V_{out} > V_{maxvout}$$

where V_{in} is the intensity of the pixel, B is the factor which controls the brightness, and V_{out} is the resulted enhanced pixel. Figure (4) shows the effect of applying contrast enhancement.



(a) Before

(b) After

Figure (4): Effect of Contrast Enhancement

• WBC Detection & Extraction

At this stage a decision has to be made in order to cut out the WBC from the resultant image produced by the second stage. WBC has a distinctive area size that depends on color stains inside the nucleus when compared with other elements. Therefore, the detecting and extracting for WBC can be achieved as follows:

1. Calculate the color differences of each pixel (i.e. differences between the three color channels for each pixel (| red level – blue level | , | blue level – green level | , | green level – red level |)).
2. Locate all pixels with the largest color differences, and mark it as the Region Of Interest (ROI).
3. Define and extract the nucleus of WBC by using both the statistical mean (i.e. center of nucleus) of the x and y position values of ROI and the standard deviation for that mean to use it as a diameter to the largest roughly circular area entirely within the cell. This ROI can be extracted as a square in size and cut out from the rest of the original image to produce the segmented WBC image.

Experimental Results & Discussions

As mentioned before, the microscopic blood color images contain five major types of WBC. The initial database contains: 20 Lymphocytes, 11 Monocytes, 16 Neutrophils, 9 Basophils and 5 Eosinophils which obtained from Blood Cells Observation [12] and BloodLine Image Atlas [13]. A comparison between the images of WBC shown in the Blood Cells Observation and those from BloodLine Image Atlas shows substantially lower detail in the images. In particular, it will be noted that, except for basophilic granulocytes (basophile), little evidence is seen of the cytoplasm granularity that is a key feature of granulocytes in general (basophils, eosinophils and neutrophils) which could be an important part for segmentation system.

The proposed method results are obtained without interaction of the user (i.e. unsupervised) and similar to the thresholding results in which the original image is segmented into white background and dark nucleus but it differ from thresholding method in that the

nucleus is more obvious since it contains color stains that indicates the position of the region of interest ROI (i.e. nucleus). To determine the largest circle around this ROI, compute the arithmetic mean and standard deviation of the x and y position values of color stains pixels within nucleus that have the largest color differences which found. Then, cut away this ROI as a small image that contains the desired segmented WBC. Experimentally it is found that this operation gives a reliable result, as shown in Figure (5).

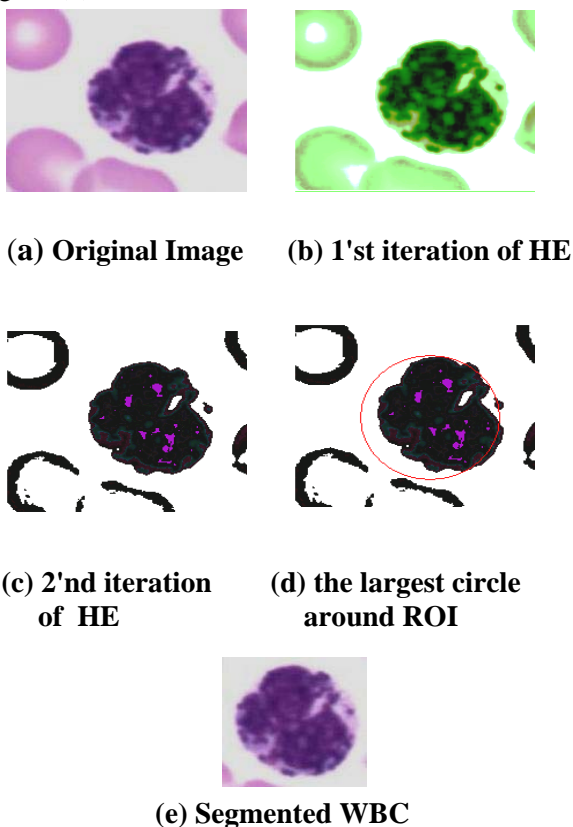


Figure (5): Unsupervised segmentation of Microscopic WBC image

To have a measure of segmentation accuracy, WBC is considered to be segmented accurately if the segment obtained by the proposed method contains all the nucleus and cytoplasm. Table (1) shows the segmentation accuracy of the proposed method.

Table (1): Segmentation accuracy for WBC

WBC Type	No. Of WBC	No. Of WBC correctly segmented	Accuracy
Lymphocytes	20	17	85%
Monocytes	11	9	82%

Neutrophils	16	13	81%
Basophils	9	7	78%
Eosinophils	5	3	60%
Total	61	49	80%

As it can be seen from table, the Lymphocytes is better segmented due to the fact that cytoplasm and nucleus are clustered and cover all the WBC making it easy to contain it in one defined ROI. Figure (6) shows the correctly segmented Lymphocytes.

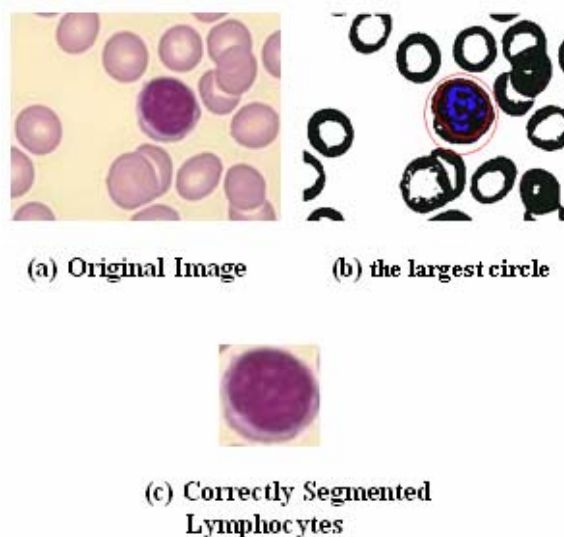


Figure (6): Segmentation of Lymphocytes using the proposed method

In other cases, incorrectly segmentation is achieved for some Eosinophils. As it can be seen in Figure (7), a major portion of the WBC is omitted. This could be due to the fact that, Eosinophils have red color WBC cytoplasm.

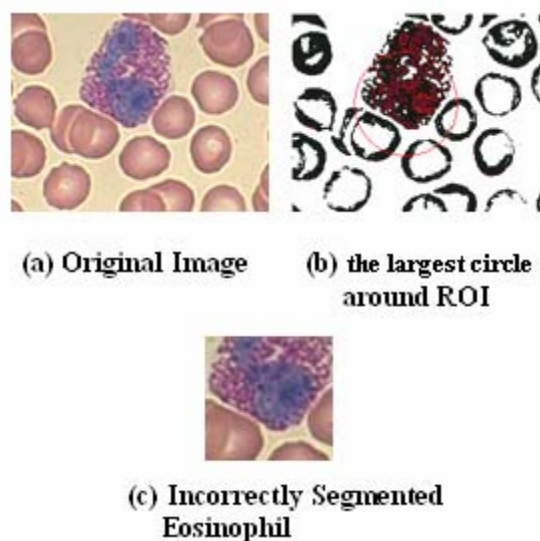


Figure (7): Incorrectly Segmentation of

Eosinophil

Additionally, the proposed method is more accurate than the thresholding. The thresholding process includes three main steps. First, extract the green channel from the original image. Second, select a threshold value that equal to 128 (i.e. the midpoint of the histogram). The researcher has to test several threshold values to convert the RGB image into a binary image and then select the appropriate one visually to give a reliable result. Third, apply the two basic morphology operators: 'erosion' followed by 'dilation' on the resultant image as a pre-processing for determining the region of interest to remove platelets that may be found in nuclear area. Here, thresholding used in opposite way. This means pixel labeled '0' (black) corresponds to objects, whereas pixel labeled '255' (white) corresponds to the background. Figure (8) depicts some results of the proposed method.

Conclusions

From the results, one can conclude the following:

- The proposed method can be categorized as unsupervised method, while the thresholding method is supervised.
- The proposed method is more appropriate than thresholding method especially to overcome the time-consuming in counting and segmenting white blood cells that contains more than one nucleus.
- The proposed method can be used for other types of white blood cells (e.g. abnormal white blood cell). Blast cells, for instance, are characteristic of certain types of **Leukemia** and can be obtained some results that similar to the thresholding method.

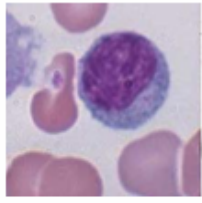
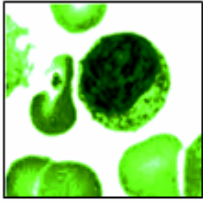

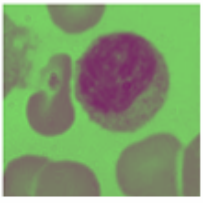

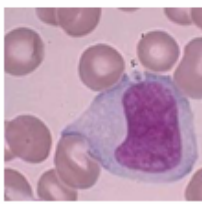
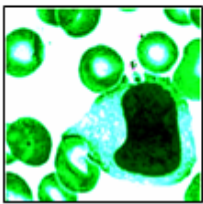

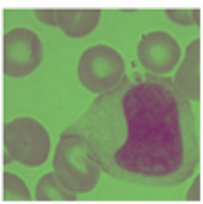

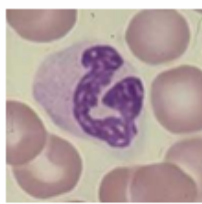
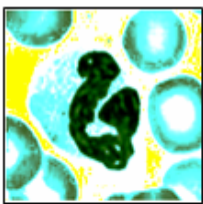

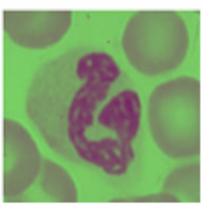

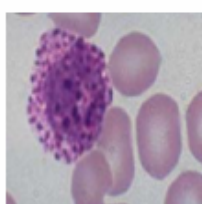
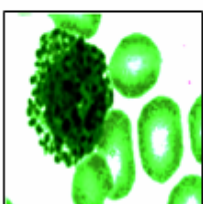
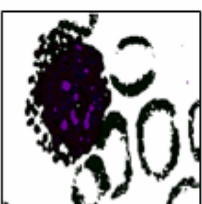
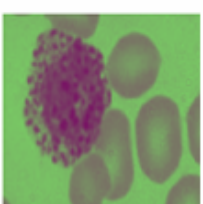

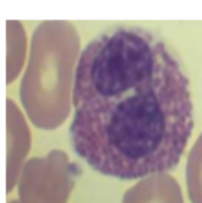
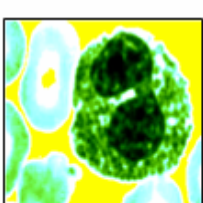
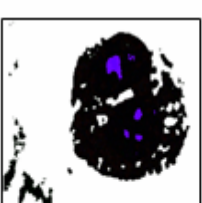
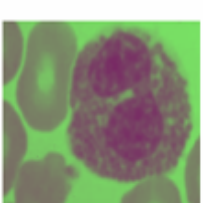

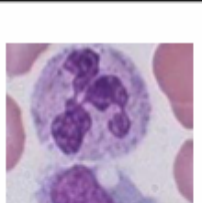
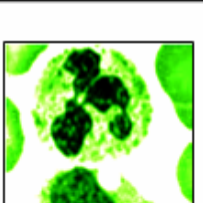
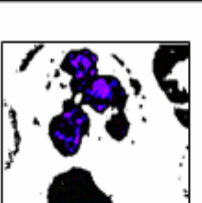
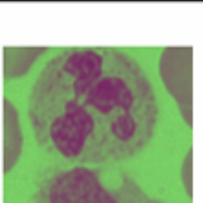

Original image	Proposed method		Thresholding method	
 <p data-bbox="271 616 391 645">Lymphocyte</p>				
 <p data-bbox="271 869 391 898">Monocyte</p>				
 <p data-bbox="271 1122 391 1151">Neutrophil</p>				
 <p data-bbox="271 1375 391 1404">Basophil</p>				
 <p data-bbox="271 1628 391 1657">Eosinophil</p>				
 <p data-bbox="271 1881 391 2033">Neutrophil -- Monocyte</p>				

Figure (8): Comparison some results between the proposed and thresholding methods.

Reference

1. Chitwong S., Cheevasuvit F., Dejhan K., Mitatha S., Nokyoo C. and Paungma T., **2000**, "*Segmentation on Edge Preserving Smoothing Image based on Graph Theory*". Proceeding of IGARSS, July.
2. Keith B. Taylor and Julian B. Schorr, **1978**, "*Blood*", Colliers Encyclopaedia, Vol. 4.
3. Harry Harms and Hans Magnus Aus, **1987**, "*Tissue Image Segmentation with Multicolor, Multifocal Algorithms*", Pattern Recognition Theory and Applications, Edited by P.A. Devjver and J. Kittler, Springer-Verlag Berlin, Heidelberg.
4. Markiewicz T. and Moszczynski, **2004**, "*Analysis of Features for Blood Cell Recognition*", VI International Workshop, Computational Problems of Electrical Engineering.
5. Hengen H., Spoor S., Pandit M., **2002**, "*Analysis of Blood and Bone Marrow Smears using Digital Image Processing Techniques*", SPIE, Medical Imaging, San Diego.
6. Theera-Umpon N., Gader D., **2002**, "*System-Level Training of Neural Networks for Counting White Blood Cells*", IEEE Transactions on systems, man, and cybernetics-Part C: Applications and reviews, Vol. 32, No.1.
7. Ray N., Acton T., Ley K., **2002**, "*Tracking Leukocytes in Vivo with Shape and Size Constrained Active Contours*", IEEE Transactions on Medical imaging, Vol. 21, No. 10.
8. Solomon C. A., Somasunudaram D., **2001**, "*Feature Extraction of White Blood Cell Images*", Electrical & Electronics Engineering Department, PSG College of Technology.
9. Gonzalez R. C. and Woods R.E., **1992**, "*Digital Image Processing*", Addison-Wesley.
10. Boyle R. and Thomas R., **1988**, "*Computer Vision*": A First Course, Blackwell Scientific Publications.
11. Bob Fisher, Simon Perkins, Ashley Walker and Erik Wolfart, **1994**, "*Hypermedia Image Processing Reference*", Department of Artificial Intelligence, University of Edinburgh, UK.
12. Daniela Tagliasacchi and Giorgio Carboni, April **1997**, "BLOOD CELLS". Available at URL:<http://www.md.huji.ac.il/gabi/blood/bloodmain.htm>.
13. Carden Jennings Publishing Co., Ltd., and Copyright **1995-2003**, "*BloodLine Image Atlas*".