AUTOMATIC WHITE BLOOD CELL SEGMENTATION BASED ON IMAGE PROCESSING

Juma Al-Muhairy, Yousef Al-Assaf

School of Engineering, American University of Sharjah
juma.almuhairy@dwtc.com ; yassaf@ausharjah.edu

Abstract: This paper presents the initial work in developing a system to segment and classify white blood cell (WBC) of human peripheral blood smear samples. Segmentation of WBC from other blood elements is achieved using various image processing techniques. Applying thresholding and watershed algorithm on the red channel of the acquired images resulted in producing appropriate segmentation of the blood elements within a smear. Comparing the area, color features of the segmented areas against those of typical WBC singled out the segments that contained WBC for further processing. The proposed segmentation method has given promising segmentation accuracy.

Keywords: image enhancement, image analysis, image segmentation, medical applications and medical systems.

1. INTRODUCTION

The analysis of blood slides is a powerful tool in determining the health status of an individual and could detect several diseases (Markiewicz and Moszczynski, 2004). The count and shape, lineage and maturity level of white and red blood cells (RBC) could aid in the diagnosis of diseases that range from inflammatory to leukemia (Markiewicz and Moszczynski, 2004; Hengen et al.,2002; Park and Keller, 2001; Solomon and Somasunudaram,2001; Stamatoyannopoulos et al, 2001).

Many automated techniques were proposed to overcome the tedious and time consuming task of human experts in counting and classifying white blood cells (Theera-Umpon and Gader, 2002; Ray et al., 2002; Wu 1998).The main steps of classifying WBC include segmentation of such cells, extract features for each type then classify the WBC type based on such features. Various techniques were used for the segmentation stage including background subtraction, histogram manipulation and thresholding (Park and Keller, 2001; Solomon and Somasunudaram, 2001; Theera-Umpon and Gader, 2002; Ray et al., 2002; WieLin, 1994). Statistically based classifiers, learning vector quantization (LVQ) and feed forward neural networks were utilized for WBC classification (Park and Keller, 2001; Solomon and Somasunudaram, 2001; Theera-Umpon and Gader, 2002; Ray et al., 2002; Gauch, 1999).

Segmentation is one of the most critical steps in the process of reducing images to information. In blood image processing this could correspond in dividing the image into regions that presumably correspond to structural units such as WBC, RBC, platelets and others. 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 literature automatic cell segmentation was avoided to decouple the error due to segmentation with that of classification (Theera-Umpon and Gader, 2002; Ray et al., 2002). Umpon and Grader (Theera-Umpon and Gader, 2002) have done manual segmentation for all the acquired images to individually get WBC. Although manual segmentation of WBC is relatively easy, automating such knowledge faces various challenges. These include the presence of RBC that could appear as one big unit, non-uniform staining of images, various shapes of WBC. Park and Keller (Park and Keller, 2001) have combined snakes and watershed algorithm for better object boundary extraction to segment WBC. Although, their work produced adequate results in terms of finding the borders of each WBC from blood slides, locating the initial snake has to be placed near the cell. Otherwise, the snake will be attracted by other forces presented in the image. Ray and Ley (Ray et al., 2002) have enhanced the gradient vector flow of the active contour algorithm to track the flow of WBC on a video stream frames. Adequate results in segmenting WBC were achieved.

The aim of our project is to develop an automatic classification system for segmenting, counting and classifying WBC from slides of peripheral blood. In addition, expert's knowledge in relating such classification to certain diseases will be established and automated. This paper presents part of the first stage of this long term project in which segmenting four types of WBC is considered. The segmentation process is based on using thresholding and watershed to enclose every element in the blood slide in a distinct area. Further, typical features of WBC such as area and color concentration and weighting are used to single out the WBC segment from all other elements.

2. METHODOLOGY

Fig. 1 presents the methodology adopted in this work to segment WBC.

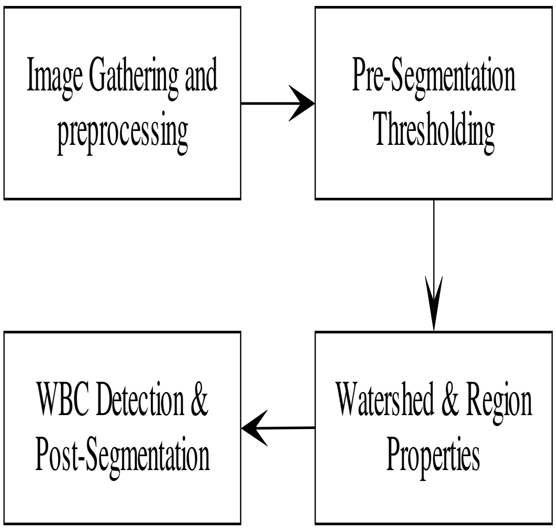


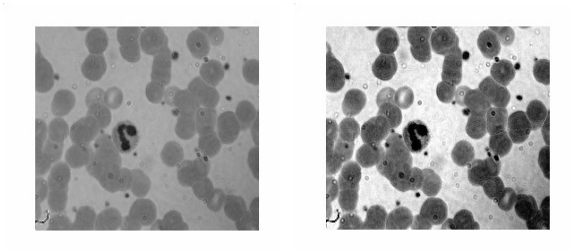
Fig. 1. WBC segmentation methodology

The following sections introduce the steps of the methodology adopted.

2.1 Image Gathering and Pre-Processing

Peripheral blood samples from 10 patients are collected and processed by Twam Hospital haematology Department at Al-Ain in UAE. The hospital is mainly concerned with the diagnoses and treatment of cancer. Images are captured from slides by an Olympus BX50 microscope, equipped with an Olympus color camera DP11, and recorded serially by an 8 bit PC interface card. For each image the hematologist expert classification of WBC is recorded to build up the database. Patients included females and males with ages ranging from 24 to 55 years. Patients reported to the hospital suffering from various degrees of illness. The resolution of the images is 1712x1368 pixels and all images acquired have been resized to a 200 by 250 matrix to ensure standardization, save memory and processing time.

During image acquisition, images are saved in JPEG format to use less computational requirements and to ensure not to have over-segmentation in watershed algorithm as image will have more details. As it is commonly the case, acquired images have all blood elements colors close to background color, red blood cells are clustered with white blood cells and the presence of noise and stain in the blood slides is significant (Hengen et al., 2002). To overcome or reduce the effect of such factors, the images posteriori standardized by increasing their contrast. Fig. 2 shows the effect of applying contrast enhancement.

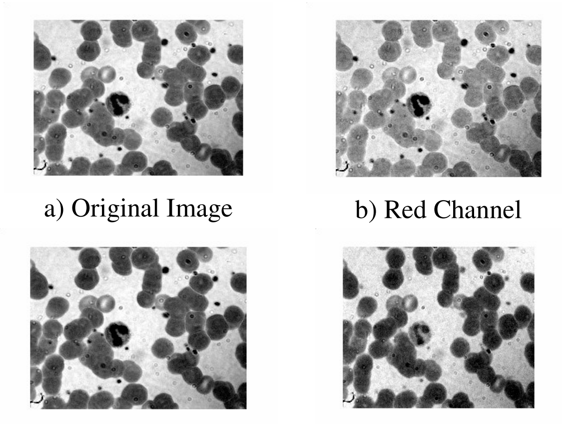


a) b)

Fig. 2. a) A typical image as taken b) after intensities adjustment

2.2 Pre-Segmentation and Thresholding

WBC has cytoplasm and nucleus with different colors that vary from class to another. To segment the desired WBC object from the background it is found that the red component of the RGB input image gives the best contrast between the background and the blood cells components including WBC, RBC and platelets as shown in Fig. 3. However, when the blue channel is used, WBC fades out while with the green channel the WBC cytoplasm color is close to the RBCs color as shown in Fig. 3. In order to produce a representative binary image, Otsu's adaptive thresholding algorithm (Otsu, 1979) is then applied on the red channel. Fig. 4 shows the output binary image produced corresponding to that shown in Fig. 3.



a) Original Image   b) Red Channel

c) Green Channel   d) Blue Channel

Fig. 3. a) Original RGB image b) red c) green d) blue channels

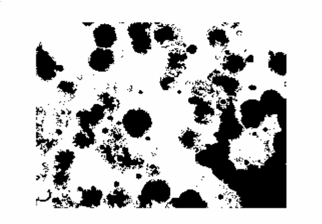


Fig. 4. Blood Image after applying Otsu's adaptive thresholding algorithm

2.3 *Watershed Transformation and Region Properties*

Watershed distance transform process is applied on the binary image obtained by Otsu's thresholding. The idea behind the watershed transform is to spot the local minimum which is the result of applying the inverse of distance transform operation. Therefore, what so called a "valley" is created and surrounded by the zeros which are called "watershed pixels/lines" (Roerdink and Meister, 2000). These watershed pixels are the boundary of each object. Then, each region/valley is filled with a different color after initiated by watershed algorithm from the center of the predefined valleys. As the regions grow, the nearby valley ran into overlapping. However, an elevated watershed lines is created to prevent overflow. This step ensures that each region is filled with different colors (Park and Keller, 2001; Roerdink and Meister, 2000). As a result, each group of nonzero values will be detected and will differentiate each region separately. At the center of the region it makes high pixel values and the pixel values converge to zero as the boundaries are approached.

Using watershed on blood slides, different objects (including WBC, RBC, platelets and stain), are extracted from the original image. Fig. 5 shows the results of the watershed algorithm when applied on the image shown in Fig. 3. As it can be seen, a considerably large number of segments are achieved while the image has only one WBC.

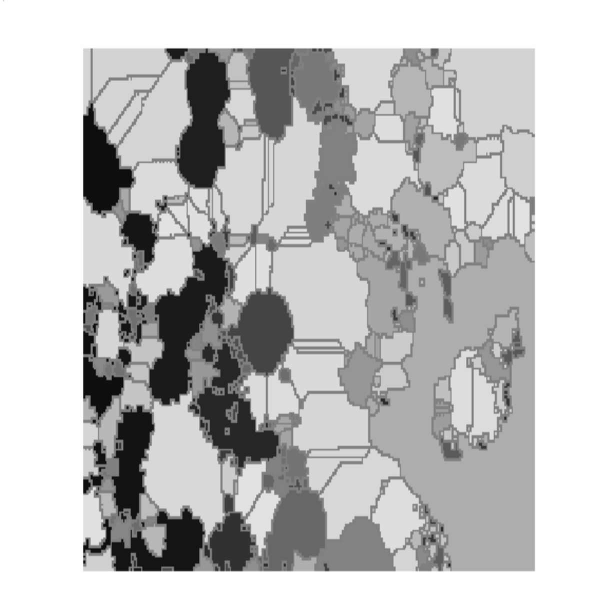


Fig. 5 Watershed output of the image shown in Fig.3

2.4 *WBC Detection and Post-Segmentation*

At this stage a decision has to be made in order to Figure out which masks, of the many obtained from watershed, represent WBC. Each area obtained with the watershed was singled out by masking it with the original image. Furthermore, a bounding box, to reduce the background part and better prepare the segment object for further processing, is created. The first stage of WBC classification is based on the area of the bounding box representing each element obtained from watershed stage. Since WBC has distinctive area size when compared with other elements a threshold value is obtained, using trial and error, below which a segment is rolled out of being a WBC.

The other two features that are used a second stage to classify WBC from other objects were color averaging (A) and the aspects ratio (AR) of the bounding box given as:

$$A = \frac{\sum_{x=0}^{N}\sum_{y=0}^{M} R(x,y) + \sum_{x=0}^{N}\sum_{y=0}^{N} G(x,y) + \sum_{x=0}^{N}\sum_{y=0}^{M} B(x,y)}{3} \quad (1)$$

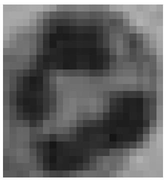
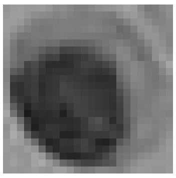
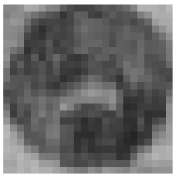
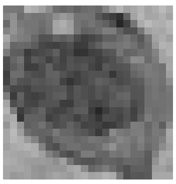
$$AR = \frac{N}{M} \quad (2)$$

Where N and M are the width and height of the bounding box in pixels of the region under consideration. R, G and B refer to RGB individual matrix values. Threshold values for the above two features are obtained for typical WBC, and any bounding box that does not meet these thresholds conditions is then disregarded as being a WBC.

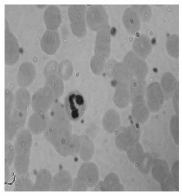
3. RESULTS & DISCUSSIONS

As mentioned before, four types of WBC are considered in this work as shown in Table 1.

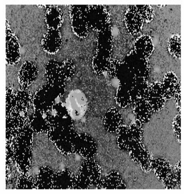
Table 1 Types of WBC considered.

SN	WBC Type	Image
1	Neutrophils	
2	Lymphocytes	
3	Eosinophils	
4	Monocytes	

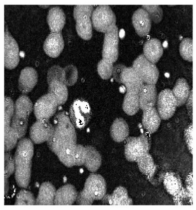
The initial database contains 120 Neutrophils, 96 Lymphocytes, 67 Eosinophils and 76 Monocytes WBC. Although other work in the area have incidentally found that the use of the intensity channel of HSI color model space gives adequate segmentation results (Hengen et al.,2002), our work have shown that using the red component of the RGB model would give better uniformity for the background color as can be seen in Fig.6. This would improve background eliminations using Otsu's adaptive thresholding algorithm.



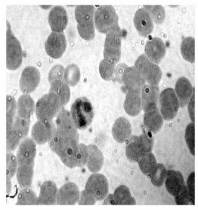
a) Red Channel of the RBG



b) Hue Image

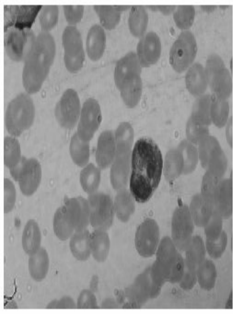


c) Saturation Image



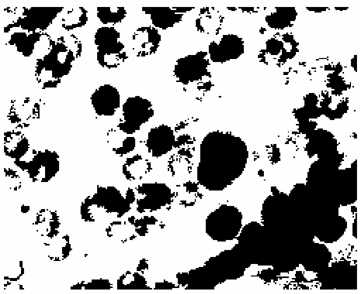
d) Intensity Image

Fig. 6. RGB (Red Channel) and HSI Model Space channels



a) Original Image for WBC

b) Using Adaptive Thre sholding Algorithm



b) Using Average Thresholding Algorithm

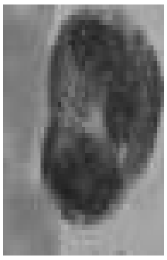
Fig. 7. Using the Normal and the adaptive algorithm

Many threshold techniques are tested to convert the RGB image into a binary image however Otsu's method is found to give appropriate thresholding since each image had different characteristics for many factors including store time of slide, lighting level and stain degree. Fig. 6 shows the resulted images using average and Ostu's adaptive thresholding. As it can be seen adaptive thresholding was capable of eliminating (or reducing the size) of many not WBC segments.

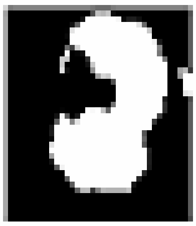
Our work has shown that in some cases the watershed produced segments which contained a WBC with a major portion of a neighbouring RBC as shown in Fig. 8a. To eliminate this unwanted object, the S component of the YES model (Wittman, 2000) is modified and used to enhance the contrast of the nucleus and cytoplasm of the WBC. Experimentally it is found that the following S component gives adequate contrast:

$$f(x,y) = 0.25R(x,y) + 0.25G(x,y) - 0.45B(x,y) \quad (3)$$

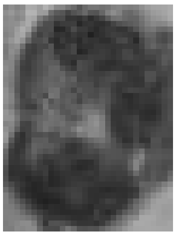
For the image shown in Fig. 8a applying the above component on each pixel of the image would produce the mask shown in Fig 8b. Multiplying the original image and its mask eliminates the neighbouring unwanted elements as shown in Fig.8c.



a) The Pre-Segmented image



b) Resulting Mask From F(x, y) function



c) The Segmented image after applying F(x, y) algorithm

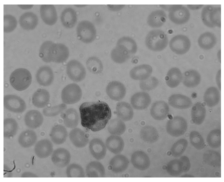
Fig. 8. Comparison between Segmented image before and after applying F(x, y) algorithm, where the segmentation is further enhanced.

To have a measure of segmentation accuracy, a WBC is considered to be segmented accurately if the segment obtained by the proposed algorithm contains all the nucleus and cytoplasm and does not contain any full RBC or other objects. Table.2 shows the segmentation accuracy using the above discussed methodology.

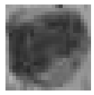
Table2: Segmentation accuracy for various WBC

WBC Type	Number of WBC	Number of WBC correctly segmented	Percent
Neutrophils	120	107	90%
Lymphocytes	96	96	100%
Eosinophils	67	30	45%
Monocytes	76	60	79%
Total	359	287	80%

As it can been seen from the above table, the Lymphocytes is better segmented due to the fact that cytoplasm and nucleus are clustered and cover all the WBC making it easy for the watershed algorithm to contain it in one defined region. Fig. 9 shows correctly segmented lymphocytes.



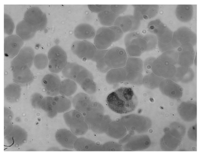
a) Original Image



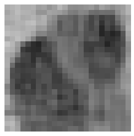
b) Segmented lymphocytes

Fig. 9. Segmentation of Lymphocytes using the proposed method

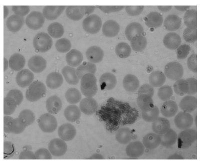
However, inadequate performance is achieved for Eosinophils segmentation.
As it can be seen in Fig. 10, adequate segmentation is achieved for some eosinophils (Fig. 10a, b). However, in other cases a major portion of the WBC is omitted (Fig. 10c, d). This could be due to the fact that, eosinophils have red color WBC cytoplasm. Hence, when going through watershed the WBC would be segmented to many regions. Further work on this issue is needed.



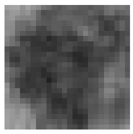
a) Original Image



b) Good Segmented Eosinophil



c) Original Image



c) Inadequately Segmented Eosinophil

Fig. 10. Eosinophils comparison between good and inadequate segmentation

4. CONCLUSION AND FUTURE WORK

In this paper our preliminary results in segmenting WBC from peripheral blood samples slides are presented. Thresholding, watershed and basic features comparison techniques are used to reach the desired segmentation. Outcomes have shown the adequate segmentation is achieved using the combination of these techniques. Currently polynomial classifiers and neural networks techniques are developed as an alternative method not only for WBC segmentation but also to distinguish their classes.

5. ACKNOWLEDGMENT

The authors would like acknowledge the help of Mr. Wessam Khiwi and Mrs Mayam Al-Shamsi from Twam Hospital for their support in image acquiring.

REFERENCES

Anoragaingrum D. (1999), "*Cell Segmentation with median Filter and Mathematical morphology Operation*" 0-7659-0040-4199, IEEE.

Gauch J. (1999) "*Image Segmentation and Analysis via Multi-scale Gradient Watershed Hierarchies*", *IEEE Trans. Image Processing*, **Vol.8**, No.1.

Hengen H., Spoor S., Pandit M. (2002), "*Analysis of Blood and Bone Marrow Smears using Digital Image Processing Techniques*", *SPIE, Medical Imaging*, San Diego.

Markiewicz T. and Moszczynski (2004), "*Analysis of Features for Blood Cell Recognition*", VI International Workshop, *Computational Problems of Electrical Engineering*.

Otsu N. (1979), "*A Threshold Selection Method from Gray-Level Histogram*", *IEEE Trans. System, Man and Cybernetics*, **Vol. 8**, no. 1.

Park J., Keller M. (2001), "*Snakes on the Watershed*", IEEE Transactions on pattern analysis and machine intelligence, **Vol.23**, No.10.

Ray N., Acton T., Ley K. (2002), "*Tracking Leukocytes in Vivo with Shape and Size Constrained Active Contours*", *IEEE Transactions on Medical imagining*, **Vol. 21**, No. 10.

Roerdink J.B. and Meijster A. (2000), "*The watershed transform: Definition, algorithm and parallelization strategies*", *Fundamental information* (41).

Solomon C. A., Somasunudaram D., "*Feature Extraction of White Blood Cell Images*", Electrical & Electronics Engineering Department, PSG College of Technology.

Stamatoyannopoulos G., Majerus P., Perlmutter R., Varmus H. (2001), "*The Molecular Basic of Blood Diseases*", W.B. Saunders Company, 3rd Edition.

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.

Wittman T.(2000), "*Face Detection in Crowded Images*", Department of Mathematics, University of Minnesota.. http://www.math.umn.edu/~wittman/faces/main.html.

Wu. S.(1998), "*Optimal Segmentation of Cell Images*" IEEE Proc.-vis. Image Signal Processing.**Vol.145**,No.1