Segmentation of Nucleus and Cytoplasm of White Blood Cells Using Gram-Schmidt Orthogonalization and Deformable Models

S. H. Rezatofighi*, R. A. Zoroofi
Control and Intelligent Processing Center of Excellence, Department of Electrical and Computer Engineering, Faculty of Engineering, University of Tehran, Tehran 14395-515, Iran
*h.tofighi@ece.ut.ac.ir

R. Sharifian
Hematology-Oncology and BMT Research Center, Tehran University of Medical Sciences, Tehran, Iran
sharifianR@sina.tums.ac.ir

H. Soltanian-Zadeh
Control and Intelligent Processing Center of Excellence, Univ. of Tehran, Tehran, Iran & Henry Ford Hospital, Detroit, MI, USA.
hamids@rad.hfh.edu

Abstract

Automatic recognition of white blood cells in hematological can be divided into four major parts: preprocessing, image segmentation, feature extraction and classification. Due to the multifarious nature of these cells and uncertainty in the hematological images, segmentation of white blood cells is one of the most important stages in this process. A scrupulous segmentation obviously reduces errors of next stages. In this paper, we introduce a novel method based on Gram-Schmidt process and parametric deformable models for segmenting the nucleus and cytoplasm. Also, we propose a new preprocessing method for improving the results of cytoplasm segmentation. Moreover, for finding the initial contour for parametric deformable model, an automatic scheme is defined. Experimental results show that our proposed method is capable of segmenting the white blood cells in the hematological images. To evaluate the proposed algorithm quantitatively, we compare its results with the manual segmentations by a hematologist. This study shows robustness of the proposed method. Another feature of the proposed method is that it is simple to implement.

1. Introduction

Automatic recognition of white blood cells in hematological images usually consists of four major steps: preprocessing, image segmentation, feature extraction, and classification. The segmentation step is crucial because the accuracy of the subsequent feature extraction and classification steps depends on the correct segmentation of the white blood cells. It is also a difficult and challenging problem due to the complex nature of these cells and uncertainty in the hematological images [1]. Therefore, this step is the most challenging in many applications and improvement of nucleus and cytoplasm segmentation is the most widespread effort in many researches.

The most common segmentation methods use edge detection [2] based on HSI model of a color image. In [3]-[5], they segment the nucleus and cytoplasm based on selecting color features and thresholding of the histogram. In another method [6], initial classification is done by the Bayes classifier on the feature vectors extracted from the RGB intensities and final segmentation of nucleus is obtained by the C-means fuzzy clustering method.

In most of the hematological images analyzed for recognition of white blood cell types, the color of nuclei is often violet with different intensities and saturation levels. Also, by changing the blood samples, the variations of intensities and saturations are increased; especially, when these images are acquired by different cameras and microscopes. Therefore, methods based on thresholding, HSI model and other methods that directly use color features are not capable of segmenting the nuclei accurately. Furthermore, cytoplasm of many types of white blood cells is colorless and their edges are barely invisible. Therefore, cytoplasm also cannot be segmented by these methods.

For segmenting the white blood cells, multi-spectral imaging technique is a new method introduced in [7]. Results from this paper show that the method is very efficient for this purpose. But, execution of such system is very cost-consuming and it is not easy for making prevalent in all laboratories. Besides the mentioned methods, region growing [8] and watershed methods [9] are used for this purpose. Finally, an unsupervised GVF snake approach is utilized in [10]. For initializing the snake boundary, an appropriate threshold is selected based on the image histogram for finding the approximate region of nucleus. Then, a boundary near the boundary of nucleus is selected after morphological operations as the initial boundary for snake. This method is very efficient for this purpose. However, the resulting boundary may not always be suitable as the initial boundary. This is due to the fact that some white blood cells have multiple separated nuclei that are named polymorphonuclears [11]. Thus, multiple boundaries are obtained for a white blood cell which is not reasonable.
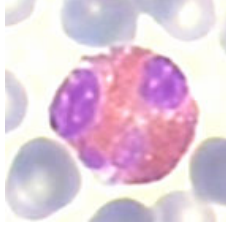

Fig. 1. A polymorphonuclear white blood cell.

In this paper, two methods are introduced based on Gram-Schmidt orthogonalization [12] and parametric deformable models [13] for segmentation of nucleus and cytoplasm of white blood cells. Before segmenting the cytoplasm, a preprocessing step is applied for improving the results. Also, the initial contour for the parametric deformable model is obtained automatically by an algorithm which depends on the number of separate parts of nucleus in a white blood cell. For validation, results of our proposed method are compared with manual segmentation by an expert. Experimental results demonstrate that this method is very efficient for the segmentation of nucleus and cytoplasm of different white blood cells.

The rest of the paper is organized as follows. In Section 2, we will introduce segmentation of nucleus based on Gram-Schmidt orthogonalization. In Section 3, preprocessing and finding the initial boundary is discussed. Parametric deformable model is the title of the next section. Experimental Results are outlined in Section 5. Finally, Paper is concluded in Section 6.

2. Segmentation of Nucleus Based on Gram-Schmidt Orthogonalization

To apply Gram-Schmidt orthogonalization [12] for the segmentation of nucleus of white blood cells, a 3-D feature vector is defined for each pixel using the RGB components of the images. Then, according to the Gram-Schmidt method, a weighting vector w is calculated for amplifying the desired color vectors and weakening the undesired color vectors. As shown in Fig. 2(b), the inner product of the weighting vector and the pixel vectors defined from the original image produce a composite image which has maximum intensity in regions with violet color and minimum intensity in other regions.

Next, by choosing an appropriate threshold based on the histogram information, we segment the image. The resulting segmentation is noisy and contains both of nucleus areas and platelets. Because these platelets are smaller than the nuclei of the white blood cells, by removing small components, the nuclei are segmented.
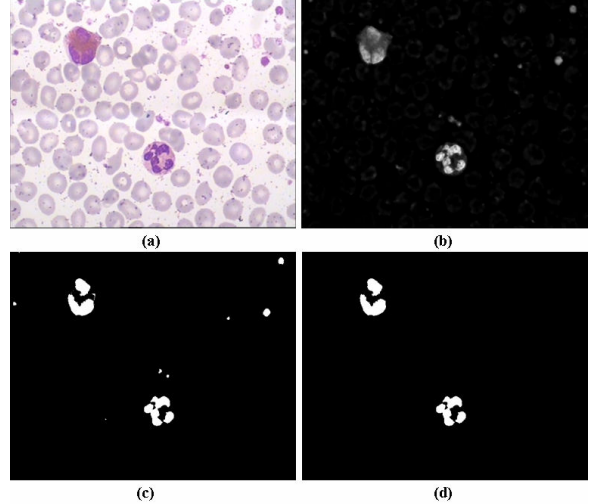

Fig. 2. (a) Original image, (b) Gram-Schmidt image, (c) Image obtained after thresholding, (d) Image obtained after removing small components.

As mentioned in the introduction, the color of nuclei is violet with different intensities and saturation levels. Also, by changing the blood samples, the variations of intensity and saturation may increase. Due to this large variation of color, three different weighting vectors are calculated for the Gram-Schmidt orthogonalization. For each vector, an image which is similar to Fig. 2(b) is obtained and an appropriate threshold is calculated for each image. These thresholds are computed based on weighted mean and maximum of their histograms.
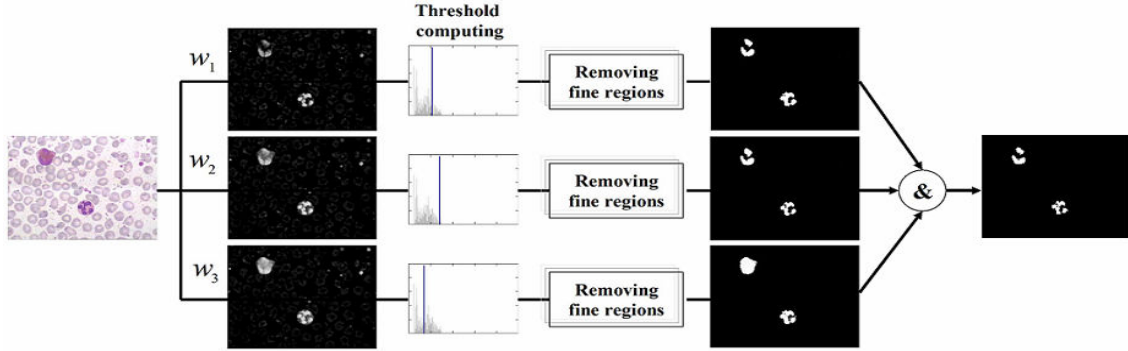
Fig. 3. Block diagram for the proposed scheme.

Finally, after thresholding and removing small (fine) components, we apply the logical "AND" operation to the results. Fig. 3 shows a flow chart for our proposed scheme.

3. Preprocessing and Finding the Initial Boundary

Since the nucleus zone is found in the previous stage, each white blood cell image is cropped using a window with an appropriate size centered at each nucleus. Due to high accumulation of the red blood cells, they may touch the cytoplasms of the white blood cells. Thus, the boundary between the cytoplasms and the red blood cells may not be distinguished when the color image is changed into gray-scale. So, we first enhance the image by color histogram equalization. Next, the enhanced image is transferred into the Hue-Saturation-Intensity (HSI) space. The final image is obtained by extracting the saturation plate from the HSI image. Based on this idea, we have a gray-scale image that has good discrimination between the boundaries of cytoplasm and red blood cells. Then, the image is smoothed by a Gaussian kernel to eliminate the holes and roughness from the cytoplasm (Fig. 4).

For finding a boundary as the initial contour for snake, the morphological dilation operation is applied on the binary nucleus image. The structuring element for the dilation operation is a square with adaptive size based on the size of the nucleus. In the next step, we determine the number of separate parts of the nucleus. If there is one zone, the convex boundary of this zone is selected as the initial contour. Otherwise, a circle with adaptive radius based on the size of nucleus is chosen. Using this algorithm, we define an initial boundary for any kind of white blood cells (Fig. 5).
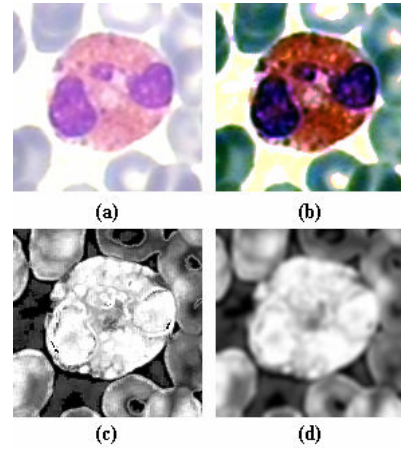

(a)   **(b)**

(c)   **(d)**

Fig. 4. (a) Cropped image, (b) Color image after histogram stretching,,(c) Saturation plate, (d) Smoothed image by a Gaussian kernel.

4. Parametric Deformable Models

Parametric deformable models or snakes are energy-minimizing splines guided by the internal constraint forces and influenced by the image forces that pull it towards distinct features (e.g., lines and edges). It was first proposed by Kass *et al* in 1987 [13]. The snake algorithm is an iterative process used for the segmentation of objects whose edges are not well-defined. By taking into account the quantitative energy values for each of the points that make up the snake, the algorithm seeks to find the position of the various snake points that minimize the overall energy of the snake. The energy function for the snake defined by Kass *et al* is represented by a vector, v(s) = (x(s), y(s)) having arc length, s, as parameter:

$$E_{snale}(v) = \int_0^1 (E_{int}(v(s)) + E_{ext}(v(s)))ds \qquad (1)$$

where E_{int} represents the internal energy of the snake due to bending or discontinuities and E_{ext} is the image forces. The image forces can be due to various events, such as lines, edges, and terminations [14]. The algorithm starts an iterative process from the initial set of points defined in the previous step. The parameters of this algorithm were set as: α =2, β = 5, γ =0.7, and ρ =0.4. The snake algorithm ends when the no snake points are moved to new positions for four consecutive iterations.
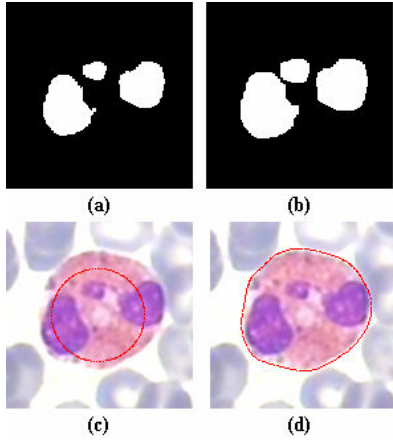

Fig. 5. (a) Segmented nucleus area, (b) Dilated image by an adaptive structure, (c) Initial boundary for poly region after dilation operation, (d) Result of deformable program.

5. Experimental Results and Discussion

The proposed method was tested on approximately 251 blood smear slide images acquired by light microscope from stained peripheral blood using the Digital Camera-Sony-Model No. SSC-DC50AP with magnification of 100. The resolution of the images is 720*576 pixels. The digital images were classified by a hematologist into the normal leukocytes: basophil, eosinophil, lymphocyte, monocyte, and neutrophil. Also, areas related to the nuclei were manually distinguished by an expert. The automatic segmentation results were quantitatively evaluated using the manual segmentation results using the similarity measure given in Eq. (2).

$$Ts = \frac{A_{program} \bigcap A_{\exp ert}}{\max(A_{program}, A_{\exp ert})} \qquad (2)$$

Here, $A_{program}$ is the segmented area by the proposed method and $A_{\exp ert}$ is the area determined by the expert hematologist. Table 1 presents the similarity measures for the segmentation of each type of white blood cells. The average values are computed using weighted averaging based on the ratio of the areas of the nucleus and the cytoplasm.

According to Table 1, the result of nucleus segmentation for the lymphocyte class is lower than the other classes. The main reason is that the color of the cytoplasm is very similar to the color of the nucleus in many of the lymphocytes samples. Therefore, the segmentation error for this type of the white blood cell is lager than the others. Due to the inherent nature of the basophiles, their cytoplasm and nuclei appear together (they are not separable). As such, both areas are segmented by the Gram-Schmidt method as nuclei. Hence, we do not have significant values to report for the cytoplasm segmentation. Also, since the vitreous cytoplasm of monocytes is colorless, even the deformable models are unable to find the cytoplasm boundaries accurately. Therefore, exact segmentation of the cytoplasm may be considered as the weakness of the proposed method.

6. Conclusion

In this paper, we proposed a new technique to segment nucleus and cytoplasm of the white blood cells in hematological images. This technique applies orthogonality principles using the Gram Schmidt method and weak boundary segmentation using parametric deformable models. For improving the results, preprocessing and automatic definition of the initial contours are added to the proposed algorithm. While applicable to the segmentation of all types of the white blood cells, the proposed method is easily implementable.

Table.1. Similarity measures for the segmentation of different types of white blood cells.

	Basophil	Eosinophil	Lymphocyte	Monocyte	Neutrophil	Total
Nucleus	94.7%	90.81%	88.86%	96.7%	94.05%	**93.02%**
Cytoplasm	-	95.55%	93.05%	81.23%	97.25%	**91.79%**
Average	**94.7%**	**93.22%**	**90.01%**	**91.23%**	**96.23%**	**93.09%**

References

[1] J. WU, P. Zeng, Y. Zhou, and C. Oliver, "A Novel Color Image Segmentation Method and Its Application to White Blood Cell Image Analysis", *International Conference on Signal Processing*, vol. 2, Beijing, November 2006, pp. 245-248.

[2] H. Hengen, S.L. Spoor, and M.C. Pandit, "Analysis of Blood and Bone Marrow Smears Using Digital Image Processing Techniques", *in: M. Sonka, J.M. Fitzpatrick, (Eds.), Proceedings of the SPIE, Medical Imaging 2002: Image Processing*, vol. 4684, 2002, pp. 624–635.

[3] I. Cseke, "A Fast Segmentation Scheme for WBC Images", *IEEE Conf. Image, Speech and Signal Analysis*, 1992, pp.530-533.

[4] S.F. Bikhet, A.M. Darwish, H.A. Tolba, and S.I. Shaheen, "Segmentation and Clustering of WBC", *IEEE Conf. Acoustics, Speech, and Signal Processing*, vol. 4, 2000, pp. 2259-2261.

[5] Q. Liano and Y. Deng, "An Accurate Segmentation Method for White Blood Cell Images", *IEEE Conf. Biomedical Imaging*, 2002, pp. 245-248.

[6] N.T. Umpon, "Patch-based White Blood Cell Nucleus Segmentation Using Fuzzy Clustering", *ECTI Trans. Electrical. Electronic. Communications,* vol. 3, no. 1, 2005, pp. 5–10.

[7] N. Guo, L. Zeng, and Q. Wu, "A Method Based on Multi-spectral Imaging Technique for White Blood Cell Segmentation", *Computers in Biology and Medicine,* vol. 37, 2006, pp. 70-76.

[8] J.M. chassery and C. Garbay, "An Iterative Segmentation Method Based on Contextual Color and Shape Criterion", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, no. 6, 1984, pp. 795-800.

[9] K. Jiang, Q.M. Liao, and S.Y. Dai, "A Novel White Blood Cell Segmentation Scheme Using Scale-space Filtering and Watershed Clustering", *Proceedings of the Second International Conference on Machine Learning and Cybernetics,* 2003, pp. 2820-2825.

[10] F. Zamani and R. Safabakhsh, "An Unsupervised GVF Snake Approach for White Blood Cell Segmentation Based on Nucleus," 8th *International Conference on Signal Processing*, vol. 2, Beijng, 2006, pp.-.

[11] Guyton, C. and E. Hall, *Text Book of Medical Physiology*, Elsevier Saunders, Pennsylvania USA, 2006.

[12] H. Soltanian-Zadeh, D.J. Peck, and D.O. Hearshen, and Renee R. Lajiness-O'Neill: "Model-Independent Method for fMRI Analysis." *IEEE Trans. Med. Imag.*, vol. 23, no. 3, March 2004, pp. 285-296.

[13] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *Int. J. Comp. Vision*, vol. 1, no. 4, 1987, pp. 321-331.

[14] C.C.H. Lean, A.K.B. See, S.A. Shanmugam, "An Enhanced Method for the Snake Algorithm", *International Conference on Innovative Computing, Information and Control (ICICIC'06)*, vol. 1, no. 30, August 2006, pp. 240-243.