

# Patch-Based White Blood Cell Nucleus Segmentation Using Fuzzy Clustering

Nipon Theera-Umpon, Non-member

## ABSTRACT

Cell segmentation is one of important steps in the automatic white blood cell differential counting. In this paper, we propose a technique to segment single-cell images of white blood cells in bone marrow into two regions, i.e., nucleus and non-nucleus. The segmentation is based on the fuzzy C-means clustering and mathematical morphology. The segmentation results are compared to an expert's manually segmented images. The initial investigation of the use of the derived segmented images in the cell classification is also performed by using the Bayes classifier.

**Keywords:** Automatic white blood cell segmentation, White blood cell differential counts, Mathematical morphology, Fuzzy C-means, Bayes classifier

## 1. INTRODUCTION

The counts of different types of white blood cells in bone marrow, the so-called differential counts, provide invaluable information to doctors in diagnosis of diseases such as AIDS, leukemia or cancers. The traditional method for an expert to achieve the differential counting is very tedious and time consuming. An automatic counting system will save time and will let the expert to perform other jobs those are more important.

White blood cells in bone marrow are classified according to their maturation stages. Even though, the maturation is a continuous variable, white blood cells are classified into discrete classes. In the myelocytic series, they can be classified into six classes, i.e., myeloblast, promyelocyte, myelocyte, metamyelocyte, band, and polymorphonuclear (PMN) ordered from the youngest to the oldest cells [1-2]. Figure 1 shows samples of white blood cells in the myelocytic series.

In this paper, we propose a technique to segment nucleus of bone marrow white blood cells. The fuzzy C-means (FCM) algorithm is applied to overly segmented cells. The patches in each oversegmented image are combined so that only two segments - nucleus

and non-nucleus - are achieved using the FCM centers. The segmentation errors are evaluated by comparing the automatic segmented images to the images segmented by an expert. An initial application of the automatic segmented images to the cell classification problem is also performed.

Some researches on automatic white blood cell counting are briefly described in the next section. Section 3 introduces the fuzzy C-means clustering, mathematical morphology, and Bayes classifier. The white blood cell data set is described in section 4. The experimental frameworks including the proposed technique and the experimental results are shown and discussed in section 5. Section 6 concludes this paper.

## 2. RELATED RESEARCH

Previously, most proposed methods followed the traditional manual maneuver, i.e., detecting a cell, extracting its features, classifying the cell, and then updating the count [3-7]. Even though several attempts have been made to solve the blood cell counting, they are applied to peripheral blood only. The counting problem in bone marrow is much more difficult due to the high density of cells. Moreover, there are many types of bone marrow white blood cells that may not be found in the blood. Our previous works were all applied to the problem in bone marrow, but were based on an assumption that the hand-segmented images are available [8-12]. We found that, in many cases, only nucleus information is adequate to classify a cell.

In other works, we developed the mixing theories of the mathematical morphology and applied them to the problem [8-9]. We also developed a new training algorithm for neural networks in order to count numbers of different cell classes, without classification [10,11]. There are several researches on cell segmentation in literature. Some examples of common techniques used in cell segmentation are thresholding [13,14], cell modeling [14-16], filtering and mathematical morphology [17], watershed clustering [16], fuzzy sets [18], etc. It should be noted that only the technique performed in [18] is applied to bone marrow whereas the other mentioned segmentation techniques are applied to peripheral blood.

## 3. METHODOLOGY

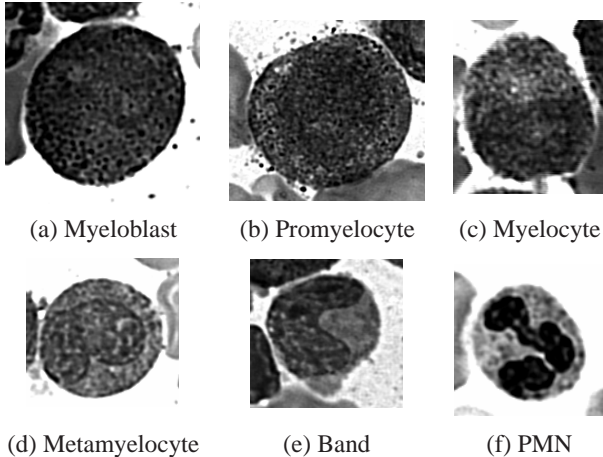
In this research, we use the fuzzy C-means (FCM) algorithm to overly segment each cell image. Because

---

05PSJ02: Manuscript received on January 20, 2005 ; revised on March 7, 2005.

The author is with Department of Electrical Engineering, Faculty of Engineering, Chiang Mai University, Chiang Mai 50200 Thailand, Phone (+66)5394-4140, E-mail: nipon@ieee.org.

the FCM algorithm is well-known and can be easily found in literature [19], we will briefly introduce it here. We also introduce the mathematical morphology and Bayes Classifier that are also applied to our proposed technique.



**Fig.1:** Cell samples in the myelocytic series.

### 3.1 Fuzzy C-Means Algorithm

Fuzzy C-means clustering method is a well-known fuzzy clustering technique [19]. Consider a set of data  $X = \{x_1, x_2, \dots, x_n\}$ , where  $x_k$  is a vector. We would like to partition the data into  $c$  clusters. Assuming that we have a fuzzy pseudopartition  $P = \{A_1, A_2, \dots, A_c\}$ , where  $A_i$  contains membership grades of all  $x_k$  to cluster  $i$ . The centers of the  $c$  clusters can be calculated by

$$v_i = \frac{\sum_{k=1}^n [A_i(x_k)]^m x_k}{\sum_{k=1}^n [A_i(x_k)]^m}, i = 1, 2, \dots, c \quad (1)$$

where  $m > 1$  is a real number that controls the effect of membership grade. In the FCM algorithm, the membership grade of the vector  $x_k$  to cluster  $i$  is defined as follows: if  $\|x_k - v_i\|^2 > 0$  for all  $i \in \{1, 2, \dots, c\}$ , then define

$$A_i(x_k) = \left[ \sum_{j=1}^c \left( \frac{\|x_k - v_i\|^2}{\|x_k - v_j\|^2} \right)^{\frac{1}{m-1}} \right]^{-1}, \quad (2)$$

if  $\|x_k - v_i\|^2 = 0$  for some  $i \in I \subseteq \{1, 2, \dots, c\}$ , then define  $A_i(x_k)$  for  $i \in I$  by any nonnegative real numbers satisfying

$$\sum_{i \in I} A_i(x_k) = 1, \quad (3)$$

and define  $A_i(x_k) = 0$  for the remaining  $i$ 's, where  $\|\bullet\|$  is some inner product-induced norm. The performance index of a fuzzy pseudopartition  $P$  is defined

by

$$J_m(P) = \sum_{k=1}^n \sum_{i=1}^c [A_i(x_k)]^m \|x_k - v_i\|^2. \quad (4)$$

The clustering goal is to find a fuzzy pseudopartition  $P$  that minimizes the performance index  $J_m(P)$ . The solution to this optimization problem was given by Bezdek in [20] and is now available in several textbooks.

### 3.2 Mathematical Morphology

Mathematical morphology was first introduced by Matheron in the context of random sets [21,22]. Morphological methods are used in many ways in image processing, for example, enhancement, segmentation, restoration, edge detection, texture analysis, shape analysis, etc. [23,24]. Morphological operations are nonlinear, translation invariant transformations. Because we consider only binary images in this research, we describe binary morphological operations only. The basic morphological operations involving an image  $S$  and a structuring element  $E$  are

$$\text{erosion} : S \ominus E = \cap \{S - e : e \in E\} \quad (5)$$

$$\text{dilation} : S \oplus E = \cup \{E + s : s \in S\}, \quad (6)$$

where  $\cap$  and  $\cup$  denote the set intersection and union, respectively.  $A + x$  denotes the translation of a set  $A$  by a point  $x$ . The closing and opening operations, derived from the erosion and dilation, are defined by

$$\text{closing} : S \bullet E = (S \oplus (-E)) \ominus (-E) \quad (7)$$

$$\text{opening} : S \circ E = (S \ominus E) \oplus E \quad (8)$$

where  $-E = \{-e : e \in E\}$  denotes the 180° rotation of  $E$  about the origin.

### 3.3 Bayes Classifier

Bayes classifier is a traditional statistical-based classifier that analyzes discriminant functions by using Bayes' theorem. Consider a classifier, we assign an input vector  $x$  to class  $C_k$  if  $y_k(x) > y_j(x)$  for all  $j \neq k$ . By choosing  $y_k(x) = P(C_k|x)$ , this posterior probability is the probability of pattern belonging to class  $C_k$  when we observe the input vector  $x$ . Bayes' theorem yields

$$y_k(x) = P(C_k|x) = \frac{p(x|C_k)P(C_k)}{p(x)}, \quad (9)$$

where  $p(x)$  is the unconditional density and  $P(C_k)$  is the prior probability of the  $k$ th class. Assuming the conditional probability density is normal, i.e.,

$$p(x|C_k) = \frac{1}{(2\pi)^{d/2} |\sum_k|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_k)^T \sum_k^{-1} (x-\mu_k)\right), \quad (10)$$

where  $\mu_k$  and  $\sum_k$  are the mean vector and the covariance matrix of the  $k$ th class, respectively. We have

$$\begin{aligned} \ln(y_k(x)) &= -\frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln\left(\sum_k |\cdot|\right) \\ &\quad - \frac{1}{2}(x - \mu_k)^T \sum_k^{-1} (x - \mu_k) + \ln(P(C_k)) \end{aligned} \quad (11)$$

#### 4. DATA DESCRIPTION

In the experiments we use bone marrow images collected at the University of Missouri Ellis-Fischel Cancer Center. Each white blood cell image is cropped manually to form a single-cell image. Then, a single-cell image is segmented manually into nucleus, cytoplasm, and background regions. The images were manually classified by Dr. C. William Caldwell, Professor of Pathology and Director of the Pathology Labs at the Ellis-Fischel Cancer Center. The data set consists of six classes of white blood cells - myeloblast, promyelocyte, myelocyte, metamyelocyte, band, and PMN. There are 20, 9, 139, 33, 45, and 185 hand-segmented images for all six cell classes, respectively. Each hand-segmented image is composed of three regions - nucleus, cytoplasm, and background - with gray level = 0, 176, and 255, respectively. Samples of cells, their corresponding hand-segmented images, along with the automatic nucleus-segmented images of all six cell classes are shown in Figure 3.

### 5. EXPERIMENTAL FRAMEWORKS

#### 5.1 Proposed Technique

The intensity inconsistency in each region of a cell is the biggest problem in the cell segmentation and classification, particularly in gray-scale images. In this research we apply a  $15 \times 15$  median filter to ease the problem. The filtered images are then overly segmented using the fuzzy C-means clustering. As we know that when  $m$  is increasing, the FCM partition becomes fuzzier. In this case, we would like to introduce some, but not too much, fuzziness to the clustering. Therefore, we heuristically set the parameter  $m$  to 2. For the number of clusters  $c$ , we would like to segment an image into several segments, much more than 2. However, if  $c$  is too large, the computational time will increase. Therefore, we set the parameter  $c$  to 10. After overly segmentation, images containing only two regions - nucleus and non-nucleus - are derived by combining the patches in the oversegmented images. The patch combining is achieved by considering the FCM centers. If the center of the patch is less than 60% of the mean of all centers, then the patch is labeled as nucleus. Otherwise, it is labeled as non-nucleus. The morphological operators, i.e. opening following by closing, both with a structuring element of 5-pixel diameter disk, are applied in the final touch to remove the small patches and smooth the edges.

#### 5.2 Evaluation Measure

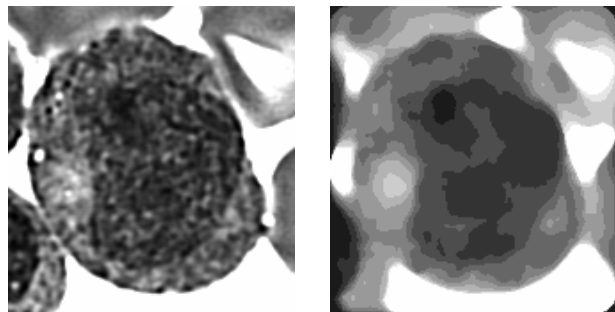
To evaluate the segmentation technique, we use the measure shown in Eq.(12).

$$E_{seg} = \frac{N_1 + N_2}{\text{Total number of pixs in the image}} \quad (12)$$

where  $N_1$  is the number of pixels in which the algorithm's decision is "Non-Nucleus" but the expert's decision is "Nucleus" and  $N_2$  is the number of pixels in which the algorithm's decision is "Nucleus" but the expert's decision is "Non-Nucleus". Basically, this measure provides the percentage of the number of the pixels in an image that are labeled differently by the expert and the algorithm.

#### 5.3 Experimental Results

##### 5.3.1 Segmentation Results



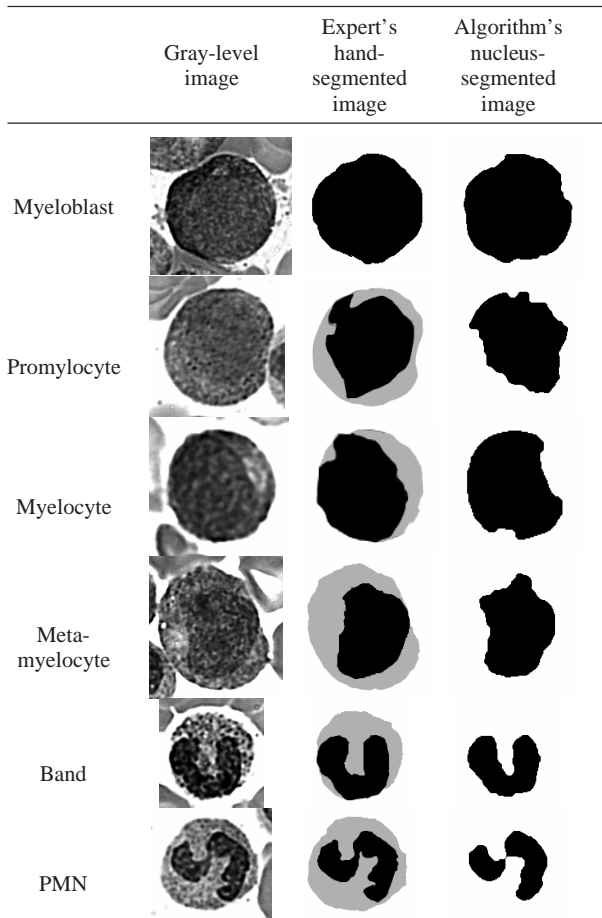
**Fig.2:** Sample gray scale and corresponding oversegmented images of a white blood cell (Metamyelocyte).

**Table 1:** Class-wise segmentation error (%).

	Blast	Pro	Myelo	Meta	Band	PMN
Segmentation error	9.23	16.07	14.73	10.21	8.60	7.01

An example of the oversegmented image is shown in Figure 2. Its corresponding final segmented image is illustrated in Figure 3 (metamyelocyte in row 4). The examples of segmentation results of all cell classes are shown in Figure 3. We compute the segmentation error of each segmented image. The overall segmentation error is calculated by averaging those of all 431 cell images. From the experiment, we achieve the overall segmentation error of 10.20%. We also calculate the class-wise segmentation errors by averaging the errors in each class. The class-wise segmentation errors are shown in Table 1.

The errors in Table 1 show that the proposed nucleus-segmentation technique performs better for the older cells. This is not surprising because when a cell becomes more mature, its nucleus is darker and nucleus boundary is sharper. Therefore, the segmentation becomes easier when the cell is older. From Table 1, however, the segmentation error of myeloblasts



**Fig.3:** Sample gray scale, corresponding hand-segmented, and automatic nucleus-segmented images of white blood cell.

(youngest cells) is less than that of some older cells. This is because the entire myeloblasts are considered to be nuclei.

We know that the immature cells are normally seen only in the bone marrow [2]. This confirms that blood cell segmentation and classification in bone marrow are more difficult than that in peripheral blood.

### 5.3.2 Initial Classification Results

To demonstrate the initial application of the automatic nucleus-segmented images, we calculate the area of nucleus in each image to be the feature to the Bayes classifier. The experiments are performed using the 10-fold cross validation. The classification rates from the classifier using the nucleus area of the automatic segmented images as a feature on training sets and test sets are 59.55% and 59.63%, respectively. For a comparison, the classification rates from the classifier using the nucleus area of the hand-segmented images as a feature on the training sets and test sets are 55.09% and 55.22%, respectively. From this initial investigation, we can see that the

segmented images from the proposed technique are promising information in the cell classification.

## 6. CONCLUSION

In this research, we propose a new technique to segment nuclei of white blood cells in bone marrow. Instead of considering each pixel, we consider a group of connected pixels called a patch. The fuzzy clustering of pixels provides the oversegmentation in which several patches are generated. These patches are then combined to form two segments of nucleus and non-nucleus regions depending upon their similarities. The opening and closing operators are applied at the final stage to perform the smoothing and noise reduction in images. From the experiments, we achieve a good segmentation and promising classification performances compared to an expert's ground truth. Due to the gray-scale inconsistency in each region of a white blood cell image, the proposed patch-based segmentation technique makes more sense than the pixel-based segmentation techniques. It also mimics how humans accomplish the cell segmentation, i.e., we consider groups of connected pixels or regions rather than each pixel.

## ACKNOWLEDGMENT

This work is supported by the Ministry of University Affairs and the Thailand Research Fund under Contract MRG4680150. The author would like to thank Dr. C. William Caldwell of Ellis-Fishel Cancer Center, University of Missouri, for providing the data and the ground truth. We acknowledge the contribution of Dr. James Keller and Dr. Paul Gader through many technical discussions on this research. We also thank the reviewers for their valuable comments.

## References

- [1] L. W. Diggs, D. Sturm, and A. Bell, *The Morphology of Human Blood Cells*, Abbott Laboratories, Abbott Park, 1985.
- [2] V. Minnich, *Immature Cells in the Granulocytic, Monocytic, and Lymphocytic Series*, American Society of Clinical Pathologists Press, Chicago, 1982.
- [3] M. Beksac, M. S. Beksac, V. B. Tipi, H. A. Duru, M. U. Karakas, and A. N. Cakar, "An Artificial Intelligent Diagnostic System on Differential Recognition of Hematopoietic Cells From Microscopic Images," *Cytometry*, vol. 30, pp. 145-150, 1997.
- [4] H. Harms, H. Aus, M. Haucke, and U. Gunzer, "Segmentation of Stained Blood Cell Images Measured at High Scanning Density With High Magnification and High Numerical Aperture Optics," *Cytometry*, vol. 7, pp. 522-531, 1986.
- [5] J. Park, and J. Keller, "Fuzzy Patch Label Relaxation in Bone Marrow Cell Segmentation," *IEEE*

- Intl Conf on Syst, Man, Cybern*, pp. 1133-1138, Orlando, FL, 1997.
- [6] S. S. S. Poon, R. K. Ward, and B. Palcic, "Automated Image Detection and Segmentation in Blood Smears," *Cytometry*, vol. 13, pp. 766-774, 1992.
- [7] S. Sohn, *Bone Marrow White Blood Cell Classification*, Master's Project, University of Missouri-Columbia, 1999.
- [8] N. Theera-Umpon and P. D. Gader, "Counting White Blood Cells Using Morphological Granulometries," *Journal of Electronic Imaging*, vol. 9, no. 2, pp. 170-177, 2000.
- [9] N. Theera-Umpon, E. R. Dougherty, and P. D. Gader, "Non-Homothetic Granulometric Mixing Theory with Application to Blood Cell Counting," *Pattern Recognition*, vol. 34, no. 12, pp. 2547-2560, 2001.
- [10] N. Theera-Umpon and P. D. Gader, "Training Neural Networks to Count White Blood Cells via a Minimum Counting Error Objective Function," *15th Intl Conf on Pattern Recog*, pp. 299-302, Barcelona, Sept 2000.
- [11] N. Theera-Umpon, and P. D. Gader, "System Level Training of Neural Networks for Counting White Blood Cells", *IEEE Trans. Systems, Man, and Cybern Part C: App and Reviews*, Vol. 32, No. 1, pp. 48-53, 2002.
- [12] N. Theera-Umpon, "Automatic White Blood Cell Classification using Biased-Output Neural Networks with Morphological Features", *Thammasat Intl Journal of Sci and Tech*, Vol. 8, No. 1, pp. 64-71, 2003.
- [13] I. Cseke, "A Fast Segmentation Scheme for White Blood Cell Images," *11th IAPR Intl Conf on Image, Speech and Signal Analysis*, pp 530-533, Sept 1992.
- [14] Q. Liao and Y. Deng, "An Accurate Segmentation Method for White Blood Cell Images," *IEEE Intl Sym on Biomedical Imaging*, pp. 245-248, July 2002.
- [15] B. Nilsson and A. Heyden, "Model-Based Segmentation of Leukocytes Clusters," *16th Intl. Conf. on Pattern Recognition*, pp. 727-730, Aug 2002.
- [16] K. Jiang, Q. Liao, and S. Dai, "A Novel White Blood Cell Segmentation Scheme Using Scale-Space Filtering and Watershed Clustering," *2nd Intl Conf on Machine Learning and Cybern*, pp. 2820-2825, Nov 2003.
- [17] D. Anoraganingrum, "Cell Segmentation with Median Filter and Mathematical Morphology Operation," *Intl Conf on Image Anal and Proc*, pp. 1043-1046, Sept 1999.
- [18] P. Sobrevilla, E. Montseny and J. Keller, "White Blood Cell Detection in Bone Marrow Images," *18th Intl Conf of the North American Fuzzy Info Proc Soc (NAFIPS)*, pp. 403-407, June 1999.
- [19] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, New Jersey, 1995.
- [20] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.
- [21] G. Matheron, *Random Sets and Integral Geometry*, Wiley, New York, 1975.
- [22] J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, New York, 1983.
- [23] E. R. Dougherty, *An Introduction to Morphological Image Processing*, SPIE Press, Bellingham, Washington, 1992.
- [24] E. R. Dougherty, *Random Processes for Image and Signal Processing*, SPIE Press, Bellingham, Washington, and IEEE Press, New York, 1999.



**Nipon Theera-Umpon** received his Ph.D. from the University of Missouri Columbia, U.S.A., M.S. degree from the University of Southern California, U.S.A., and B.Eng (Hons.) degree from Chiang Mai University, Thailand, all in Electrical Engineering, in 2000, 1996, and 1993, respectively. Since 1993, he has been a faculty member in the department of Electrical Engineering, Chiang Mai University, where he currently holds the rank of assistant professor. From 2000 to 2001, he received a postdoctoral fellowship and was a visiting scholar at the Center for Applied Image Recognition, University of Missouri Columbia where he conducted a research on automatic land mine detection. His research interests include computational intelligence, mathematical morphology, neural networks, medical imaging, and signal and image processing. He is a member of the IEEE Signal Processing Society and Computational Intelligence Society, Thai Robotics Society, and Biomedical Engineering Society of Thailand. His biography is listed in the Marquis Who's Who in Science and Engineering and Who's Who in the World by Marquis, New Jersey.