Focused Color Intersection for Leukocyte Detection and Recognition System

Lina, Arlends Chris, and Bagus Mulyawan

***Abstract*—The proposed system aims to automatically detect and recognize the leukocyte types from microscopic images. Thedeveloped system detects the leukocyte area within an input image and classifies the leukocyte type in one of the five categories (neutrophils, eosinophils, basophils, lymphocytes, and monocytes). The focused color intersection method is performed based on the color information of the blood cell images. Three main processes in the proposed system are: 1) the color histogram construction, 2) the focus region determination, and 3) the active search process. The system starts with constructing the Red Green Blue (RGB) color histogram and finding the histogram intersection. Next, the system selects the focus regions and does active searching of the focus regions which contain parts of the image which have the highest probability of the histogram intersection. The focused color intersection method enables the system to automatically detect leukocyte areas and determine the leukocyte types from microscopic images.**

***Index Terms*—Leukocyte detection, leukocyte recognition, white blood cell, focus color intersection, active search.**

I. Introduction

A hematology analyzer is one of the most helpful devices in a medical laboratory. It helps hematologists to diagnose the hematopoietic system disorders efficiently, compares to the manual screening and evaluation. In order to identify the hematopoietic system disorders, the hematologists will need to detect blood disorder and perform the leukocyte count from human blood.

The blood elements include erythrocytes (red cells), leukocytes (white cells), and platelets [1]. In contrast to red cells, normal white blood cells are nucleated and include neutrophils, lymphocytes, monocytes, eosinophils, and basophils [2]. In order to be able to produce the blood cell report, the hematologists will have to carefully determine each cell type and find its location in the microscopic images.

Some attempts to help hematologists to classify and count the blood cell by developing an automatic cell recognition system have been proposed. Markiewicz used the Support Vector Machine method for classifying the blood cells [2], [3], Colunga proposed the automatic cell recognition system using EM algorithm [4], and some previous works used Neural Network methods as cell classifiers [5]-[7].

Manuscript received December 24, 2012; revised March 1, 2013. This work was supported by the Indonesian ministry of Education under HibahUnggulanPerguruanTinggi Grant (2012-2014).

Lina is with the Faculty of Information Technology, Tarumanagara University, Jakarta, Indonesia (e-mail: lina@untar.ac.id).

A. Chris is with the Faculty of Medicine, Tarumanagara University, Jakarta, Indonesia (e-mail: arlends.chris@untar.ac.id).

B. Mulyawan is with the Faculty of Information Technology, Tarumanagara University, Jakarta, Indonesia (e-mail: bagus@untar.ac.id).

In this paper, an automatic system that can detect the white blood cell and classify its type is developed. The proposed system works based on the similarity of image color using the focused color intersection method and the active search method. Three main processes in the proposed system are: 1) the color histogram construction, 2) the focus region determination, and 3) the active search process.

In the first process, the system starts with constructing the Red Green Blue (RGB) color histogram, determining the number of histogram bins, and finding the histogram intersections. Next, the system selects the focus regions by constructing a small window and shift-scanned it to the rest of the image. Finally, the active search algorithm is used to determine the focus regions which have the highest values of histogram intersections with the model.The searching locations are reduced by implementing the focus region technique; Therefore, the proposed detection system will be more efficient than the conventional searching method.

The remainder of this paper is organized as follows. In Section II, the methods used in the proposed leukocyte detection and recognition system are described, i.e. the color histogram which contains the color quantization, normalization, and histogram intersections methods, the focus region method, and the active search with upper bound method. Section III presents the focused color intersection algorithm, while Section IV describes the active search algorithm. Section V explains the experimental setup and results. Finally, the conclusion is presented in Section VI.

II. Leukocyte Detection and Recognition System

A. Color Histogram

Color histogram is obtained by counting all discrete intensity values of each color in the image array [8]. The histogram concept only considers the color distribution from an object and ignores the shape and texture features. Histogram is invariant to translation, rotation, view and scale changes, and occlusions [8].

The steps for applying color histogram function are as follows:

1) Color quantization

Color quantization is a procedure to reduce the possibility of color numbers. Every microscopic image within RGB color domain will be quantized by its intensity values.

2) Normalization

Normalization is calculated by dividing the number of actual pixel with the total number of pixel in an image. Normalization step will make the histogram no longer dependent to the number of pixel, but to color distribution.

3) Histogram intersection

Histogram intersection shows the number of pixel in

histogram model which can be found from an image. The histogram intersection is defined as follows [8]:

$$\sum_{j=1}^{n} \min (I_j, M_j) \quad (1)$$

where *I*is the histogram of an image, *M* is the histogram model, and n is the number of the histogram bin.

Histogram intersection finds the minimum value of the histogram bins between the image and model. The histogram bin consists of pixels with the same color intensity. The higher the intersection value means the more similar the image with the model.

B. Focus Region

Histogram result from an image usually has a certain level of similarity with the histogram model [9]. However, it is not necessary for processing the whole image at once. The process can be done by selecting some parts of the image and match them with the model. These parts of an image are named as focus region.

Fig. 1 shows the illustration of selecting the focus regions of an image. First, construct a small window with *w* pixelx*w* pixel. Then, this window is shifted *s* pixel to one direction and continues for the rest of the image.

C. Active Search with Upper Bound

As neighboring focus regions tend to have a similar color histogram, the active search algorithm will concentrate to the focus region which has the highest probability of histogram intersectionwith the model [9]. The searching locations are reduced by applying the upper bound technique to the histogram intersection. The upper bound $\hat{S}$ of a focus region B to a model M can be calculated as follows [5]:

$$\hat{S}(B, M) = \frac{\min(S(A,M)|A|,|A \cap B|) + |B - A|}{|B|} \quad (2)$$

with $|A|$ is the pixel number in the focus region A,$|B|$ is the pixel number in the focus region B, $|A \cap B|$ is the pixel number that is belong to A and B, and $|B - A|$ =is the pixel number in B which is not belong in A.

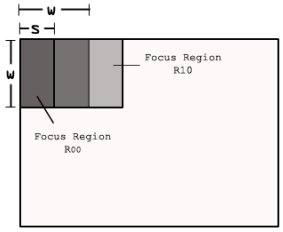


Fig. 1. The shifting process of a focus region in an image.

III. Focus Color Intersection

The algorithm of the Focused Color Intersection is as follows [9]:

Step 1. Set $M_c = M$and$R_c = R$, where *R* is the focus regions and *M* is the models. Meanwhile, $R_c$and$M_c$are the focus regions and models which are competing.

Step 2. For every model $M \in M_c$, determine the best matching focus region R^Musing the active search algorithm. $R^M$is the focus region which has been related with model M.

Step 3. S(R^M, *M*) = $\max_{M \epsilon M_c} S(R^M, M)$. Correlate the region$R^M$ with model *M*.$S(R^M, M)$is the normalized color histogram value, which is calculated between focus region R^M and model *M*. Then,$\max_{M \epsilon M_c} S(R^M, M)$is calculated for finding the winner couple of focus region $R^M$and model *M* which has the highest value of $S(R^M, M)$.

Step 4.Masking in focus region is used to avoid rematching process of the focus region to the model. The masking process is applied to every pixel including the pixels in the focus regionR^M.

Step 5. Determine the set of the focus region $R'_c$and the set of the modelM'_c. Delete the focus regions which have less than a certain threshold value (β) from $R_c$to R'_c, and delete the model with upper bound higher than the$S(R^M, M)$and less than θ from set $M_c$to the set of M'_c.

Step 6. If M'_c orR'_c is empty, then quit.

Step 7. Set $M_c = M'_c$and$R_c = R'_c$and continue to Step 2.

IV. Active Search

The active search algorithm is used to determine the focus region which has the highest value of histogram intersection with the model. The active search algorithm is defined as follows [9]:

Step 1. Set $\theta' = \theta$, and$lub(R_{ij}^k, M) = 1.0$ for all $R_{ij}^k \in R$. The upper bound value of a focus region can be calculated more than once, with$\theta$is a threshold value, while$\theta'$is a temporary threshold value. $lub(R_{ij}^k, M)$is the temporary upper bound value between focus region$R_{ij}^k$and model *M*. This $lub(R_{ij}^k, M)$ will be saved for comparison with other upper bound values from histogram intersections with other focus regions neighbors.

Step 2. Continue with the next focus region R_{ij}^k. If$lub(R_{ij}^k, M) < \theta'$, then set $S(R_{ij}^k, M) = 0$and go to step 5.

Step 3. Calculate $S(R_{ij}^k, M)$ as a normalized color histogram intersection value between a focus regionR_{ij}^k and model *M*. Set $\theta' = \max(S(R_{ij}^k, M), \theta')$.

Step 4. Calculate $\hat{S}(R_{uv}^p, M)$ untuk R_{uv}^p dalam *neighborhood*dari R_{ij}^k menggunakan persamaan (3). Set $lub(R_{uv}^p, M) = min(lub(R_{uv}^p, M), \hat{S}(R_{uv}^p, M))$.$\hat{S}(R_{uv}^p, M)$is the upper bound from R_{uv}^p. $R_{uv}^p$is the temporary focus region to save R_{ij}^k, which is the focus region from the normalized color histogram intersection.

Step 5. Continue to step 2 if there is another focus region left.

Step6 R_{uv}^p is selected so that $S(R_{uv}^p, M) = max_{R_{ij} \epsilon R} S(R_{ij}^k, M)$, with R_{uv}^p is the focus region with the highest value of histogram intersection if $S(R_{uv}^p, M) > \theta$. If $S(R_{uv}^p, M) < \theta$ then no focus region which has intersection with *M* with higher value than threshold θ.

V. Experimental Results

We conducted several experiments for evaluating the proposed leukocyte recognition system. We developed our own FTI-Untar Leukocyte database, which consists of 144 leukocyte images (90neutrophils, 8 eosinophils, 3 basophils, 44 lymphocytes, and 18 monocytes).

We used 10% of the available images for training, which means 9 neutrophils, 1 eosinophil, 1 basophil, 4 lymphocytes,

and 2 monocytes were used for training. The samples of the microscopic image which contain leukocytes are shown in Fig. 2. A 35x35 pixels window was used to scan the testing image with *10* pixels shifting value. The testing was conducted with various quantization values as follows:

1) RGB quantization Q = (2, 2, 2)
2) RGB quantization Q = (5, 5, 5)
3) RGB quantization Q = (10, 10, 10)
4) RGB quantization Q = (50, 50, 50)

The window is expected to have 95% of the total pixels value in the focus region.

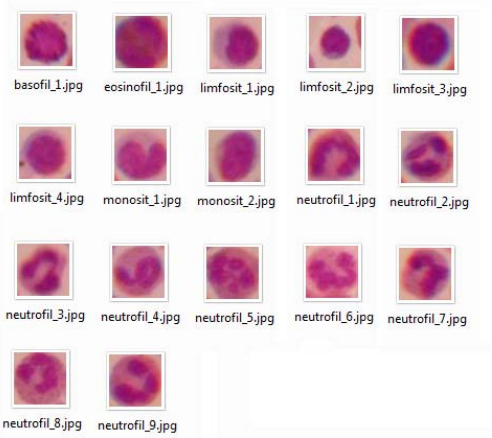


Fig. 2. Samples of leukocyte images used in the experiments.

TABLE I: RECOGNITION ACCURACIES OF THE LEUKOCYTE RECOGNITION SYSTEM USING VARIOUS QUANTIZATION VALUES

Type	# oftotal image model	Recognition accuracy (%)			
		Q (2,2,2)	Q (5,5,5)	Q (10,10, 10)	Q (50,50,5 0)
Basophils	5	0	100	100	100
Eosinophils	5	20	100	100	100
Lymphocyte s	20	10	100	100	90
Monocytes	10	0	100	100	100
Neutrophils	45	4.44	100	100	73.33

TABLE II: RECOGNITION RESULTS USING Q = (2, 2, 2), SHIFTING VALUE = 10, WINDOW SIZE = 35, BETA = 0,95

Type	Accuracy (%)	FAR (%)	FRR (%)
Neutrophils	46.7	84.4	53.3
Lymphocytes	42.3	88.3	57.7
Monocytes	15.7	94.4	84.3
Eosinophils	8.8	98.5	91.2
Basophils	22.2	99	77.8

TABLE III: RECOGNITION RESULTS USING Q = (5, 5, 5), SHIFTING VALUE = 10, WINDOW SIZE = 35, BETA = 0,95

Type	Accuracy (%)	FAR (%)	FRR (%)
Neutrophils	33.8	14.9	66.2
Lymphocytes	13.8	34.1	86.2
Monocytes	12.9	79.1	87.1
Eosinophils	0	0	100
Basophils	0	0	100

TABLE IV: RECOGNITION RESULTS USING Q = (10, 10, 10), SHIFTING VALUE = 10, WINDOW SIZE = 35, BETA = 0,95

Type	Accuracy (%)	FAR (%)	FRR (%)
Neutrophils	27.2	0	72.8
Lymphocytes	8.2	56.8	91.8
Monocytes	7.1	37.5	92.9
Eosinophils	0	0	100
Basophils	0	0	100

The first experiment was conducted to evaluate the accuracy of the leukocyte recognition system using four level of quantization: (2, 2, 2), (5, 5, 5), (10, 10, 10), and (50, 50, 50) with the testing images were also used for training. Table I shows the recognition accuracy for each quantization values. The highest recognition results were 100%, achieved by using the quantization level (5, 5, 5) and (10, 10, 10) for all types of leukocytes. Meanwhile, the lowest results were achieved by the system using (2, 2, 2) quantization level. This failure was caused by the wide range of the histogram bin values, which classify windows with high color differences in the same histogram bin. In the opposite, quantization level (50, 50, 50) force a narrow range of the histogram bin values, which caused a very sensitive classification of windows into the histogram bin.

Table II, Table III, and Table IV show the recognition results using Q = (2, 2, 2), Q = (5, 5, 5), and Q = (10, 10, 10), respectively. Each table presents the accuracies, the false acceptance rate (FAR), and the false rejection rate (FRR) of the system in recognizing all leukocyte types. It is clearly seen from Table II, Table III, and Table IV that the recognition accuracies of the leukocyte images were very low, with high false acceptance rates. The highest recognition accuracy for neutrophils was 46.7%, 42.3% for lymphocytes, 15.7% for monocytes, 8.8% foreosinophils, and 22.2% for basophils, all using Q = (2, 2, 2). The low recognition accuracies of the leukocytes were possibly due to the small number of learning images input to the system. Also, the testing images consist of leukocyte images with incorrect cropping, as a result from an automatic detection system.

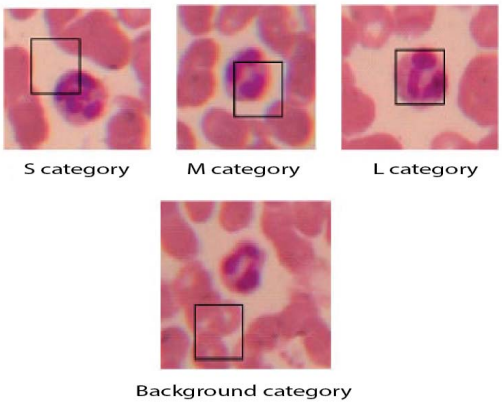


Fig. 3. Samples of four categories of detection used in the experiment

TABLE V: DETECTION RESULTS USING Q = (5, 5, 5)

Type	# of recognition	Detection result (%)				Accuracy (%)
		S	M	L	Background	
Neutrophils	32	14.29	58.39	27.33	0	85.09
Lymphocytes	8	43.9	53.66	2.44	0	65.85
Monocytes	9	13.95	79.07	6.98	0	20.93
Eosinophils	0	0	0	0	0	0
Basophils	0	0	0	0	0	0

TABLE VI: DETECTION RESULTS USING Q = (10, 10, 10)

Type	# of recognition	Detection result (%)				Accuracy (%)
		S	M	L	Background	
Neutrophils	22	2.73	59.09	38.18	0	100
Lymphocytes	7	43.24	56.76	0	0	43.24
Monocytes	2	0	62.50	37.50	0	62.50
Eosinophils	0	0	0	0	0	0
Basophils	0	0	0	0	0	0

Besides the experiments for recognizing the leukocyte types, we also conducted the experiments for detecting the leukocytes positions in the microscopic images. Fig. 3 shows the samples of four categories used for classifying detection results in the experiments. The four categories of detection results are:

S category: when the focus region contains 1 – 33.33% of the leukocyte area.

M category: when the focus region contains 33.33 – 66.67% of the leukocyte area.

L category: when the focus region contains 66.67 – 100% of the leukocyte area.

Background category: when the focus region contains 0% of the leukocyte area.

Table V and Table VI show the detection result from the proposed system. It is clearly seen from both tables that the highest detection accuracy was achieved for detecting the neutrophils locations, with 85.09% for Q = (5, 5, 5) and 100% for Q = (10, 10, 10). For lymphocytes, the highest detection result was 65.85% achieved by the system with Q = (5, 5, 5), while for monocytes the highest detection result was obtained by the system with Q = (10, 10, 10) with 62.50%. However, the system failed to detect both eosinophils and basophils areas. The system could detect leukocyte area in a small focus region, however misclassification of the leukocyte type was still a major problem.

VI. CONCLUSION

We have presented the focused color intersection method to automatically detect leukocyte areas and determine the leukocyte types from microscopic images.The recognition and detection results of leukocyte images using focused color intersection are highly dependent on the quantization level and the threshold value. Other parameters such as shift level, window size, and beta would also affect the accuracy of the system as well as the execution time.

In the future, for improving the system's accuracy, we consider to develop a dynamic window model for detecting the leukocyte area, the use of other color domains, i.e. Hue, Saturation, and Value (HSV) and texture information in the system.

ACKNOWLEDGMENT

The authors thankNadyaPaskaria(B.Sc of Computer Science, Faculty of Information Technology, 2007) for her assistance in collecting the data for the experiments.

REFERENCES

[1] M. M. Wintrobe, *Clinical Hematology*, 12th Edition, Lippincott Williams & Wilkins, Philadelphia, 2008.

[2] M. Adjouadi, N. Zong, and M. Ayala, "Multidimensional Pattern Recognition and Classification of White Blood Cells using Support Vector Machines," *Particle & Particle Systems Characterization*, vol. 22, no. 2, 2005.

[3] T. Markiewicz and S. Osowski, "Data Mining Techniques for Feature Selection in Blood Cell Recognition," in *Proceedings of European Symposium on Artificial Neural Networks*, Belgium, 2006.

[4] M. C. Colunga, O. S. Siordia, and S. J. Maybank, "Leukocyte Recognition using EM-algorithm," in *Proceedings of 8th Mexican International Conference on Artificial Intelligence,* Guanajuato, 2009.

[5] N. T. Umpon and P. D. Gader, "Training Neural Networks to Count White Blood Cells via a Minimum Counting Error Objective Function," in *Proceedings of International Conference on Pattern Recognition*, Barcelona, 2000.

[6] M. Beksac, M.S. Beksac, V. B. Tipi, H. A. Duru, M. U. Karakas, and A. Cakar, "An Artificial Intelligent Diagnostic System on Differential Recognition of Hematopoietic Cells from Microscopic Images," *Cytometry*, vol. 30, 1997.

[7] G. P. M Priyankara, R. K. O. H. Silva, and C. R. D. Silva, "An Extensible Computer Vision Application for Blood Cell Recognition and Analysis," *IEEE Trans. on Neural Networks*, vol. 4, 1993.

[8] M. J. Swain and D. H. Ballard, "Color Indexing," *International Journal of Computer Vision*, vol. 7, no. 1, 1991.

[9] V. V. Vinod, H. Murase, and C. Hashizume, "Focused Color Intersection with Efficient Searching for Object Detection and Image Retrieval," *Pattern Recognition*, vol. 30, no. 10, 1997.



Lina was born in Padang, West Sumatera, Indonesia, in 1979. She received a B.Eng. degree in computer science from Tarumanagara University, Indonesia in 2001, and got a M.Sc. degree in computer science from the University of Indonesia in 2004. In 2005, she received the Japanese Government Scholarship as a research student at the Graduate School of Information Science, Nagoya University, Japan. Later she continued her graduate study and got her doctoral degree from Nagoya University, Japan in 2009. Currently she is working as a lecturer at Tarumanagara University, Indonesia. Her research interest is in pattern recognition and computer vision field. She is also a member of IEEE and IEICE.



Arlends Chris was born in Bogor, West Java, Indonesia, on March 21, 1977. He got his Medical Doctor from the Faculty of Medicine from Tarumanagara University, Indonesia, in 2005. Later, he obtained his master degree in Educational Psychology from Tarumanagara University in 2009. Now, he is working as a lecturer at the Faculty of Medicine, Tarumanagara University. He is also a member of the Indonesian Medical Association.



BagusMulyawan was born in Yogyakarta, Central Java, Indonesia, on 1962. He got his bachelor degree from Computer Engineering department,Gunadarma University, Indonesia, in 1992. He obtained his master degree in Information System Management from Budi Luhur University, Jakarta, Indonesia, in 2008. Now, he is working as a lecturer at the Faculty of Information Technology, Tarumanagara University.