

Automated Status Identification of Microscopic Images Obtained from Malaria Thin Blood Smears

Dian Anggraini^{#1}, Anto Satriyo Nugroho^{*2}, Christian Pratama^{##3}, Ismail Ekoprayitno Rozi^{##4},

Aulia Arif Iskandar^{#5}, Reggio Nurtanio Hartono^{#6}

[#]Swiss German Univesity

Campus EduTown, BSD City Tangerang 15339 Indonesia

¹anas.anggraini04@gmail.com

^{*}Center for Information and Communication Technology

Agency for the Assessment and Application of Technology (PTIK-BPPT)

Jalan M.H. Thamrin No.8 Jakarta 10340 Indonesia

²asnugroho@ieee.org

^{##}Eijkman Institute of Molecular Biology

Jalan Diponegoro no. 69 Jakarta 10430 Indonesia

Abstract — Development of an accurate laboratory diagnostic tool, as recommended by WHO, is the key step to overcome the serious global health burden caused by malaria. This study aims to explore the possibility of computerized diagnosis of malaria and to develop a novel image processing algorithm to reliably detect the presence of malaria parasite from *Plasmodium falciparum* species in thin smears of Giemsa stained peripheral blood sample. The algorithm was designed as an expert system based on the method used by medical practitioner performing microscopy diagnosis of malaria. Digital images were acquired using a digital camera connected to a light microscope. Prior to processing, the images were subjected to gray-scale conversion to decrease image variability. Global thresholding were implemented to obtain erythrocyte and other blood cell components in each image. The segmented images were further processed to obtain possibly infected erythrocyte and the components of parasite inside the corresponding erythrocyte using multiple threshold. These parasite's constituents (nucleus and cytoplasm) were used as the preliminary basis for parasite/non parasite classification. Malaria samples prepared and provided by Eijkman Institute of Molecular Biology Indonesia were used to test the proposed algorithm.

Keywords— malaria, thin blood smears, image segmentation, thresholding

I. INTRODUCTION

Malaria is an infectious disease that has been reported to be a serious global health problem, causing between 1.5 and 2.7 million of deaths every year in more than 90 countries [1]. It is caused by intracellular single-celled parasite that belongs to genus *Plasmodium* [2]. Up to this date, 5 species are discovered to infect human by entering bloodstream, namely *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium malariae*, *Plasmodium ovale*, and *Plasmodium knowlesi* [3]. The impact of malaria on clinical and public health is geographically variable and it has been reported that the

highest transmission rate is found in tropic and subtropical countries that have enormous rate of population growth including Indonesia.

Diagnosing malaria is the first step to control the spread of the disease. Currently, there are several methods employed to pronounce malaria. Among these techniques, visual evaluation of Giemsa stained blood film by means of light microscopy is by far the most widely used in developing countries, not to mention Indonesia. The scheme that entails examining both thick and thin blood smears for the presence of Plasmodia is believed to be the most sensitive and specific one [4].

Beside numerous advantages, the use of light microscope in diagnosing malaria also has some drawbacks. The quality of smears is not homogenous and the condition of the slide is highly influenced by time and storage. It also relies heavily on the expertise of medical practitioner in the field. In addition, confirming negative status of a malaria slide take considerable time and efforts. As a consequence, these disadvantages bring many difficulties in Mass Blood Screening (MBS) and become a burden in controlling the spread of malaria especially in rural areas. This condition is mainly observed in eastern part of Indonesia where the number of expert microscopists are limited while the prevalence of malaria is considerably high. Therefore, an automated image analysis system would improve the performance of microscopy by circumventing its main limitation in term of dependency on the ability of medical practitioner to diagnose blood image accurately, thus providing a milestone for fast and accurate diagnosis of malaria in Indonesia remote areas.

Many current research efforts in developing image processing algorithm, as addressed in [5], [6], [7], [8], have been focused on, not only, diagnosing the positive or negative status of malaria but also determining the species of infecting parasites. These studies used either pixel based parasite detection [5], [8], detection based on morphological

processing of segmented parasites [6], or extraction of image features from segmented cells [7]. Among these approach only in [7] the algorithm is designed in much the same way as human operator performing microscopy

The aim of this study is to propose a new algorithm for automated malaria status identification based on the standard routine used by medical practitioner performing microscopy diagnosis of malaria. The system is developed to segment the image into parts, i.e. to separate blood cells from the background, and to detect the parasites infecting them by identifying the components of the parasite. The algorithm is expected to provide a positive and negative diagnosis of malaria with comparable sensitivity and specificity to conventional microscopy.

This paper is organized in 5 sections. In Section 2, the materials and methods to obtain and standardize the blood sample slides are explained. Section 3 describes the theoretical and practical aspect of the development of the algorithm. Furthermore, the implementation of proposed algorithm and experimental results are presented and analyzed in Section 4. Finally, Section 5 provides the conclusions.

II. MATERIALS AND METHOD

Malaria samples were prepared by experts from Eijkman Institute of Molecular Biology Indonesia. All samples were prepared to be used as teaching slides for the purpose of training medical practitioners to diagnose malaria parasites using light microscopy. Of the five *Plasmodium* species known to infect human, *Plasmodium falciparum* was used as source for parasite samples in this study due to its prevalence in Eastern Indonesia as well as its major contribution to deaths associated with malaria.

Each slide was stained using standard Giemsa staining protocol. All slides had been subjected to manual microscopy examination by expert microscopists, who had verified and given a species-specific and parasitemia diagnosis.

Following several studies in automated malaria diagnosis by Di Ruberto et al, Sio et al, Frean, Proudfoot et al [6], [9], [10], [11], thin blood smears were used in this study. In preparing thin blood smears, blood materials were fixed with methanol and thus preventing hemolysis. Consequently, unlike with thick blood smears, there is no loss of parasite during staining; limited artifacts and overlapping cells; parasites are observable in their natural location and shape within the red blood cell since their morphology is conserved.

To generate the images, slides were examined under oil-immersion microscope with 10x100 times magnification and images were captured using a built-in digital camera microscope.

Images were then analysed manually. The analysis calculated the number and determined the stage and types of parasite, calculated and determined the type of white blood cells, and platelets as well as counted the number of red blood cells present in each photo. Two copies of each image were made. One copy was used to manually analyse the positive or negative status, while the other copy was used to test the algorithm designed in this study. Datasets, specifically

dedicated for this study, were then created. In total, 110 were acquired from 3 training slides while 12 images were used in this preliminary study.

III. ALGORITHM DEVELOPMENT

In this study, the image processing algorithm was designed to be as comparable as possible to the approach utilized by lab technician performing microscopy diagnosis. Each step was built according to standard microscopic procedure and confirmed by experts from Eijkman Institute of Molecular Biology Indonesia.

A. Image Acquisition

Images were captured using a 5-megapixel Nikon digital sight DS 5Mc specifically designed and built-in for the light microscope. However, in this study, the images were generated using only 1.2 megapixel resolution. Utilizing neither optical nor digital zoom, the camera captured the image with 10x100 magnifications. The slides were examined under oil immersion in order to adjust the refractory index. Images were saved in the JPEG format in 1280 x 960 pixels size.

B. Proposed Algorithm Based on Manual Microscopy Image Analysis

In Fig.1, the method used by human operators to analyze blood smear image as well as the corresponding image processing algorithm are summarized. In this study, the design of algorithm will focus on identification of parasite from the species *Plasmodium falciparum*.

In human body, malaria parasites are present generally in two different phases namely trophozoite and gametocyte. Every phase has its own specific recognizable features. Therefore, recognizing the key element of both trophozoite and gametocyte phase is the fundamental point in identifying malaria parasite [4].

In early stage, the parasites that infect the red blood cell (RBC) always develop through trophozoite phase or generally known as ring stage, and thus make it the most widely seen (see Fig.2a). With proper Giemsa staining, the nucleus of the parasite can be identified as one dark dot (indicated by number 1 in Fig.2a) inside the RBC while the "ring" part will be easily distinguished as ring-shaped structure with more intense color compared to its surrounding (indicated by number 2 in Fig.2a). Therefore, the evaluation of thin blood smears start with identification of blood cells including RBC, white blood cell (WBC), and platelets (see Fig.3). It proceeds further with recognition of RBC or erythrocytes. For each erythrocyte, medical practitioner will try to identify suspected parasite which is indicated by the presence of dark spot inside the RBC (indicated by number 4 in Fig.3). When no black dot is found, then the erythrocyte is classified as clean erythrocyte as labeled with 1 in Fig.3. Erythrocyte with black dot inside it will be analyzed further. Furthermore, in order to be identified as infected erythrocyte, the dot must be accompanied by cytoplasm of the parasite that appears as lighter area surrounding the nucleus with ring-shaped structure around it as clearly indicated by number 3 in Fig.2a.

The cytoplasm of the parasite has wide variety of shapes, from a fine ring to an irregular or amoeboid shape [4]. The presence of both nucleus and cytoplasm, as the main constituent of living cell, is the most important prerequisite in validating the positive or negative status of blood smear image. When this requirement is not met, then the black dot is simply an artifact.

It is possible that the cytoplasm has very dark or intense color and resembles a banana-shaped structure (see Fig.2b). When this is the case, then the parasite can be surely confirmed as *Plasmodium falciparum* in gametocyte stage, which is a preparation for the sexual phase in the female Anopheles mosquito vector. Otherwise, presence of Maurer's dot (see Fig. 2a numbered with 4) will be the main attribute to confirm that the parasite is *P.falciparum* in ring stage.

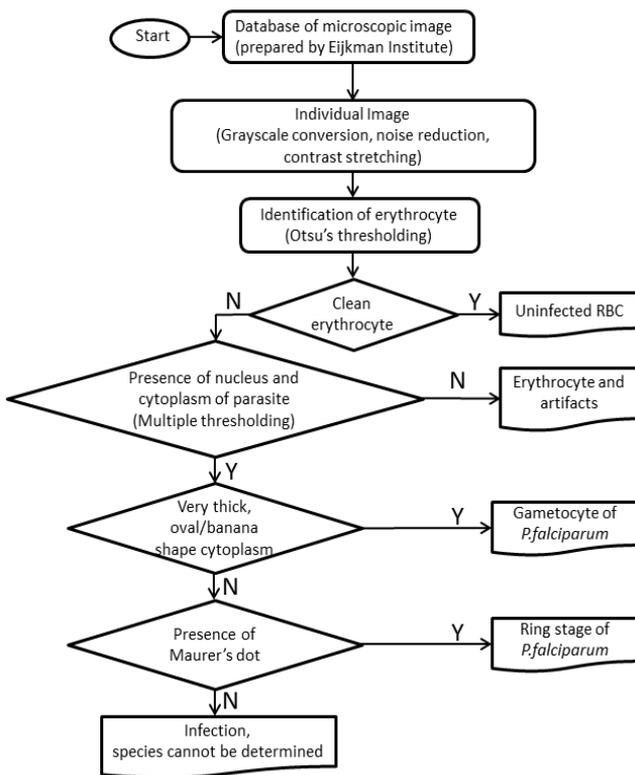


Fig. 1 Standard procedure of parasite identification of *P.falciparum* using manual microscopy

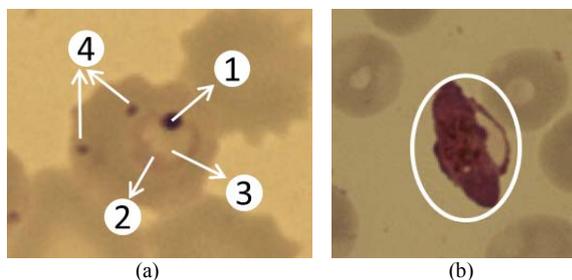


Fig.2 Erythrocyte infected by ring stage or gamete stage parasite

This paper proposes an image processing algorithm to embrace the mechanism discussed above. The scheme of the algorithm was initiated by implementing pre-processing stages that includes removal of unwanted noise and enhancement of parasites' components. The processed images were then investigated further to separate blood cells, parasites and other components from background in a microscopic field of a thin blood smear. This was accomplished by applying several global thresholding methods and visually compares the results to qualitatively determine which technique yields the best result. The binary image was then subjected to hole filling morphological operator and applied as marker to label blood cells. From each identified cell (RBC and WBC), constituents of the parasite (nucleus and cytoplasm) were extracted using multiple threshold.

Substantially, this paper will address the process up to attaining the nucleus and cytoplasm of the parasite whereas the positive or negative status of the image is given based on the extracted components of the parasites.

IV. IMPLEMENTATION OF PROPOSED ALGORITHM AND EXPERIMENTAL RESULTS

A. Pre-processing

The pre-processing stage was designed to remove unwanted effects from the image and to adjust the image as necessary for further processing [12]. This stage is comprised of converting the color image to gray-scale, filtering the obtained gray-scale image with 3x3 median filter, and subjecting each image to contrast stretching.

In Figs. 3-5, several results of the process are presented. In Fig.3, the initial sample image, some constituents of blood cells including, labeled in ascending order, RBC, WBC, and platelets, are shown. In addition, parasites, along with their nucleus and cytoplasm, are clearly observed inside the RBC (indicated by number 4 in Fig.3). Prior to any image processing method, the color image was converted to gray-scale image in order to reduce complexity without compromising the importance of parasite details.

The gray-scale image was then filtered using 3x3 median filter. The basic idea of this step is replacing the value of every pixel in the image by the median value of the intensity level in the neighborhood that will result in reduced "sharp" transitions, which is the indication of random noise, in intensities [13]. A 3x3 mask was used because, unlike average filter and 5x5 median filter, it does not remove parasite details while performing well in removing the noise from the image [12].

Histogram of the smoothed gray-scale image, Fig.4, shows that one problem with the image is that it has low contrast indicated by a narrow histogram located typically toward the middle of the intensity scale. In order to enhance the details of parasite components and to obtain a cleaner and brighter image background, each image was processed through a contrast stretching. This step will expand the intensity level in an image so that the histogram cover a wider range of intensity scale and the image exhibits a large variety of gray

tones. [13] The result of the pre-processing step can be observed in Fig.5.

B. Segmentation of Blood Cells Components

The algorithm was further pursued with image segmentation that identifies and segments potential parasites and erythrocytes from the background. All foreground objects are separated to extract the infected erythrocytes and further identify the parasites components. Reference [14] and [15] used edge detection and watershed algorithm for blood cell segmentation that is commonly used in general image segmentation. However, in this study, the algorithm is designed to heavily rely on thresholding that convert the gray-scale image into a binary image by selecting a threshold value that maximally divide the image into two classes of intensity: *C1* that contains intensity of background and *C2* that contains intensity of objects.

When the intensity distribution of objects and background pixels are significantly distinct, it is possible to use a single global threshold over the entire image. However, this is not the case in this study. Even after contrast stretching, the intensity distribution between blood cell components and background is not sufficiently different. To overcome this burden, Otsu's Method [16] was utilized to automatically select threshold level that maximizes the between-class variances of the histogram. In principle, Otsu's method calculates the separability of the two classes, *C1* and *C2*, which depends on the probability of class *C1* occurring, average intensity up to the threshold level and average intensity of the entire image. It iteratively computes the value of between-class variance for different value of threshold and selects the one that give the highest value of between-class variance.

Each resulting binary mask of blood cell components was then processed with morphological operator so that the holes inside blood cells can be removed. Conditional dilation [13] is generally used to remove holes in a particular image. However, in this study, hole filling process was developed based on extraction of connected components algorithm [12], [13].

The process, which leads to all holes in an image being filled, is shown in Fig.7 and can be described as follow. Fig.6 shows a binary image of blood smear in which some erythrocyte contains hole, i.e. part of foreground object that has the same intensity as the background. Since the principle of CCL is finding foreground point [13], the complement of the image was first made before subjecting it to CCL. Following the CCL stage, the image was reconstructed by setting an intensity of 0 to the labeled objects otherwise the intensity was set to 255. Fig.8 shows the reconstructed image. Hole filling process was finalized by making the complement image of the reconstructed image (see Fig.9). This process worked well for all 12 images.

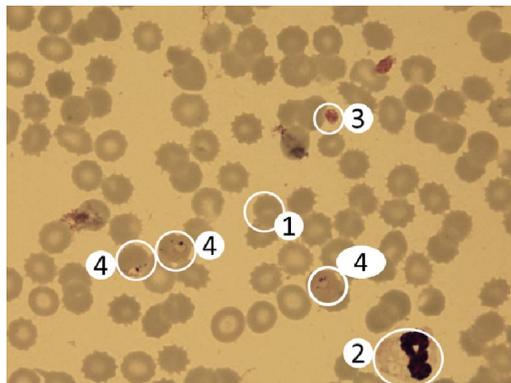


Fig. 3 A sample malaria blood image with blood cell components circled in white and numbered

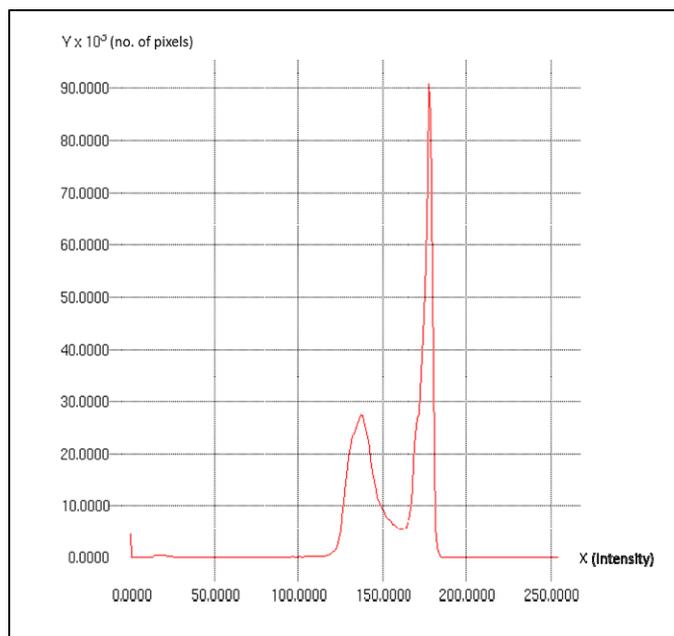


Fig.4 Histogram of smoothed image

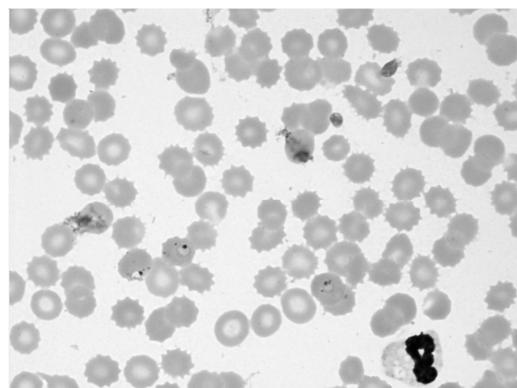


Fig.5 Image after pre-processing

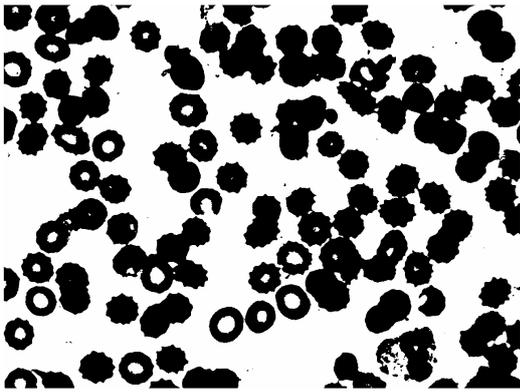


Fig.6 Binary image obtained with Otsu's method

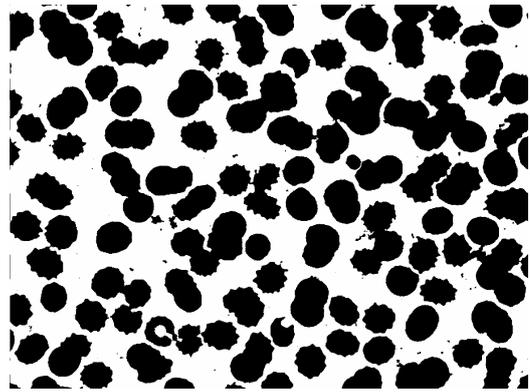


Fig.9 Binary image after hole filling process

C. Extraction of Parasite Components

As mentioned before, there are two important components, parasite's nucleus and cytoplasm, that become the prerequisite for status identification of thin blood smear image. Using the knowledge that parasite's cytoplasm appear lighter while parasite's nucleus appear darker than the cytoplasm of the erythrocyte, it is possible to locate the infected cell by first segmenting each cell into 3 regions by mean of a multiple threshold [13].

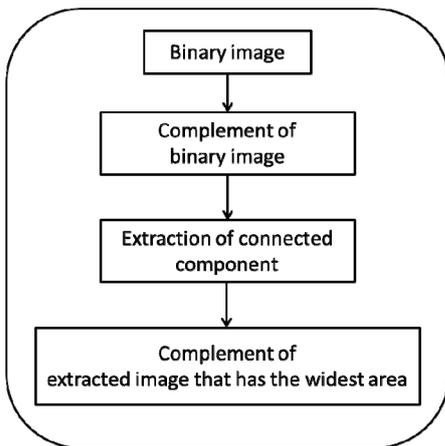


Fig.7 Hole filling process

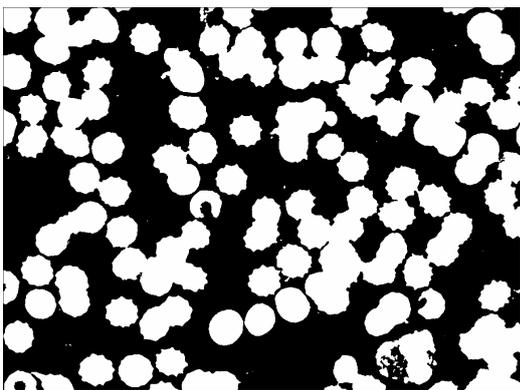


Fig.8 Reconstructed image

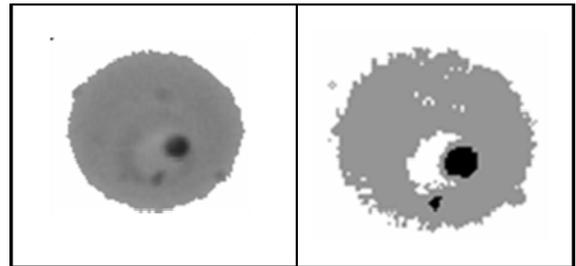


Fig.10 Infected erythrocyte and result after multiple threshold

This approach can be illustrated as follow. Fig.10a shows an image of infected RBC taken from one of the images used in this study. The objective is to segment the image into 3 classes of intensity: the area of erythrocyte's cytoplasm, the cytoplasm of the parasite and the dark spot which is the nucleus of the parasite. It is evident from the histogram that two threshold should successfully solve this case. The multiple threshold procedure resulted in threshold 101 and 137, divides the histogram into 3 regions of intensity and can be noted from Fig.11. The process resulted in separation of the infected RBC into 3 regions (see Fig.11 and 10b).

The similar concept was applied to all 12 blood smear images. It was found that it successfully detected the parasites' components in 10 out of 12 images. In those 10 images, the infected RBCs have parasite in late ring stage and thus both nucleus and cytoplasm are much more apparent (see Fig.12).

V. CONCLUSIONS

This study provides a good basis for those who are aiming to investigate the automated blood film analysis for screening of malaria or other infectious disease. Unlike prior algorithm, the method proposed in this paper was developed based on the routine manual microscopy and each step of the development has been confirmed by expert microscopists from Eijkman Institute of Molecular Biology Indonesia. As a result, the malaria blood image processing can significantly reduce the main detractions of microscopy, which are its dependence on the skill and experience of human technicians and the time and labor intensive work of conventional blood smear analysis.

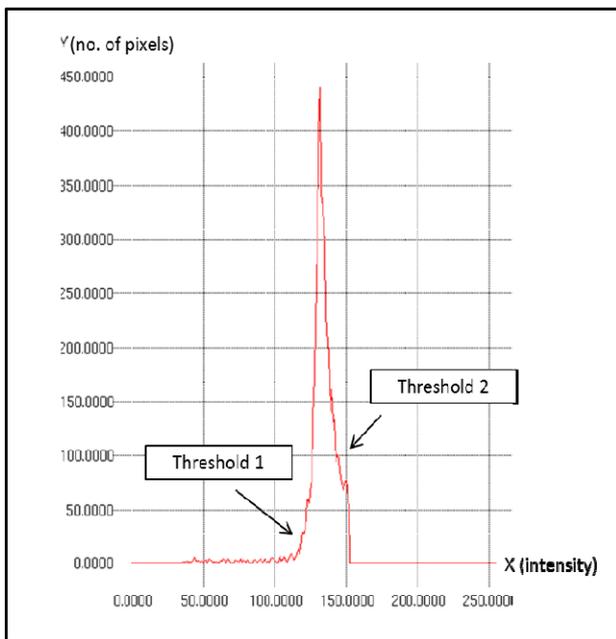


Fig.11 Histogram of infected cell

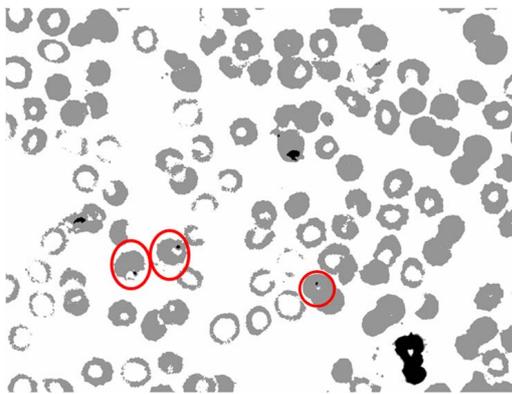


Fig.12 Parasites' detection after multiple threshold

The main goal of this research was accomplished by solving two main problems by mean of image segmentation: separation of blood cell components from background and identification of parasites' constituents, in this case the nucleus and cytoplasm, inside the infected erythrocyte. At first, the algorithm employed automatic threshold selection and successfully segmented blood cell components from background. It was further pursued by novel adaptation of CCL and connected component extraction to perform hole filling morphological operator. Automatic multiple threshold selection completes the algorithm by successfully identified the nucleus and cytoplasm of the parasites.

As further enhancement, it is planned to classify the phase of the parasite based on the shape and size of the parasites' cytoplasm as well as identification of Maurer's dot. Moreover, limitation of this study will be its focus on the species of *Plasmodium falciparum* and thus might not be optimally applicable to other species. Therefore, further work will also involve some improvements to accommodate the diagnosis of other species of parasites.

ACKNOWLEDGMENT

This research has been supported by the Ministry of Research and Technology, Government of Indonesia under Grant Inseintif Riset RT 2011 – 2311. All supports are, therefore, hereby acknowledged. Distinctive gratitude is due to Eijkman Institute of Molecular Biology Indonesia for preparing the materials for this research and granting exceptional access to the datasets. The authors also wish to thank Swiss German University for giving opportunity to conduct this research and publish the paper in ICEEI 2011. Equally important, the authors received help from many and specially would like to deliver the best gratitude to Andree Ang Surya and Teresa Vania Tjahja from Swiss German University for the supports and scholarly inputs.

REFERENCES

- [1] World Health Organization, *World Malaria Report*, Geneva: WHO, 2009
- [2] P.R.Murray, *Medical Microbiology*, 6th ed., USA: Mosby, 2008.
- [3] W.E.Collins and J.W.Barnwell, "*Plasmodium knowlesi*: Finally being recognized", *Journal of Infectious Disease*, vol. 199, pp. 1107-1108, 2009.
- [4] World Health Organization, *Basic Malaria Microscopy. Part1. Learner's Guide*, 2nd ed., Geneva: WHO, 2010
- [5] G. Diaz, F. Gonzales, and E. Romero, "Infected cell identification in thin blood images based on color pixel classification: comparison and analysis", Springer Berlin, pp. 812-821, 2007.
- [6] C. Di Ruberto, et al, "Morphological image processing for evaluating malaria disease", *IWVF4*, pp. 739-748, 2001.
- [7] N.E. Ross, et al, "Automated image processing method for the diagnosis and classification of malaria on thin blood smears," *Medical and Biological Engineering and Computing* 44, pp. 427-436, 2006.
- [8] F.Tek, A.Dempster, and I.Kale, "Malaria parasite detection in peripheral blood image", *Proceeding of British Machine Vision Conference*, pp. 347-356, 2006.
- [9] S.W.S. Sio, et al, "Malaria Count: an image analysis-based program for accurate determination of parasitemia", *Journal of Microbiological Methods*, vol. 68, pp. 11-18, 2006.
- [10] J. Frean, "Microscopic determination of malaria parasite load: role of image analysis", *Microscopy: Science, technology, Applications, and Education*, pp. 862-866, 2010
- [11] O. Proudfoot, et al, "Investigation of a novel approach to scoring Giemsa stained malaria infected thin blood films", *Malaria Journal*, vol.7:62, April 2008. Available : <http://www.malariajournal.com/content/7/1/62>
- [12] Wu Q., Merchant F., Castleman K. *Microscope Image Processing*. 2008. New Jersey: Academic Press.
- [13] R.C. Gonzales, R.E. Woods, and S.L. Eddin, *Digital Image Processing*, 3rd ed., USA: Prentice Hall, 2008
- [14] Ritter, N. and J. Copper, "Segmentation and border identification of cells in image of peripheral blood smear slides", *Proceeding of Thirtieth Australasian Computer Science Conference*, pp. 161-169, 2007.
- [15] F.Tek, A.Dempster, and I.Kale, "Blood cell segmentation using minimum area watershed and circle radon transformation", *Mathematical Morphology: 40 Years On*, pp. 441-454, 2005.
- [16] N. Otsu, "A threshold selection method from grey level histogram", *IEEE Transactions on Man, System, and Cybernetic*, vol.9, no.1 pp. 62-66, 1979.