Segmentation of Malaria Parasite Candidate from Thickblood Smear Microscopic Images using Watershed and Adaptive Thresholding

Umi Salamah[1, 2], Riyanarto Sarno[1], Agus Zainal Arifin[1], Sarimuddin[1], Anto Satriyo Nugroho[3], Ismail Ekoprayitno Rozi[4], and Puji Budi Setia Asih[4]
[1]*Department of Informatics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia.*
[2]*Departement of Informatics, Universitas Sebelas Maret, Surakarta, Indonesia.*
[3]*Center for Info. and Comm. Tech. Agency for the Assessment and Application of Technology, Indonesia.*
[4]*Eijkman Institute for Molecular Biology, Indonesia.*
umi14@mhs.if.its.ac.id

***Abstract*—Segmentation of malaria parasite on thick blood smear is a critical intermediate step in automation process of malaria detection. Most of the thick blood smear have low quality that characterized by high noise, the low-intensity difference between background and foreground, and the presence of artifacts. This situation makes the segmentation process becomes difficult. In this paper we proposed a new segmentation strategy for microscopic images of malaria parasite obtained from thick blood smear using watershed and adaptive thresholding. The proposed method consists of two main stages: image enhancement and segmentation. Enhancement process used Low-pass filtering and contrast stretching. Meanwhile, the segmentation used combination watershed segmentation and adaptive thresholding. The performance was evaluated on 253 parasite candidates, cropped from 22 thick blood smear microphotographs. The experimental results showed that the average segmentation accuracy of the proposed algorithm was 95.2%. Further analysis showed that the nucleus and cytoplasm of the malaria parasite were successfully extracted, thus the method is suitable for being used on detection of malaria parasites.**

***Index Terms*—Adaptive Thresholding; Low Quality; Malaria Parasite; Thick Blood Smear; Watershed.**

I. INTRODUCTION

Malaria is a serious infectious disease in tropical country. From World Malaria report 2014, there are recorded greater than 5 hundred thousand deaths from about 200 million cases of malaria were reported in 2013 [1]. Because of the high mortality rate, malaria should be handled as soon as possible.

One important step is to make a proper diagnosis of malaria patients. It can be obtained by conducting a series of tests on blood samples. Manual microscopy examination of blood smears is the gold standard in the diagnosis of malaria [WHO]. It should be observed at least 100 microscopic fields from thick blood smear image with a high magnification [2, 3]. The aim of this process is to detect the presence of malaria parasites in the patient's blood. This process is tedious and tiring. Therefore, it required computer-aided diagnosis to overcome this problem.

Segmentation is key that determines the success or failure of the process of detection and identification of malaria. Watershed method is widely used in medical image segmentation, such as in leukemia, nucleus, or objects granular form. Advantages of this method are that it can provide a natural growth of the region corresponding to each object shape and size independently, and automatically provides a closed contour and computing efficiency [4]. However, over-segmentation may exist in case of watershed-based segmentation. Not only over segmentation but also because of by image noise, image artifacts, and discrete nature of image intensity distribution. Therefore, watershed needs to combine with other methods to overcome this drawback [4, 5, 6, 7]. Several studies perform segmentation on a thick blood smear. Kaewkamnerd et al. used adaptive threshold that used to segment image derived from information V-Value histogram [8]. Elter et al. used comparisons green and blue components of the smear to highlight the object that contains the core. The proportion of green and blue components used as the basis for determining fixed threshold in the segmentation process [9]. Arco et al. do quantify malaria parasite on thick blood smear. In the segmentation process using adaptive threshold to identify existing blood components [10].

In the above studies, parasite determination is based on the parasitic nucleus. As in Arco et al. [10], the assumption of blood components on a thick blood smear are only parasites and white blood nucleus. In fact, there are four components in the thick blood smear i.e. the nucleus of white blood, parasites, platelets, and artefact. In the above studies, parasite determination is based on the parasitic nucleus. As in Arco et al. [10], the assumption of blood components on a thick blood smear are only parasites and white blood nucleus. In fact, there are four components in the thick blood smear i.e. the nucleus of white blood, parasites, platelets, and artefact. The last two components are possible to have sizes and shapes similar to the nucleus parasite. Moreover, the method in previous study ineffective if used images have low difference intensity between object and the background. Meanwhile, the characteristics of low quality of thick blood smears are noisy, foreground intensity which is similar to the background, and the possibility of the appearance of artefacts (mold, dirt that occurs in the process of manufacture/ storage smears, and other bacteria). The main difficulty is that the parasite cytoplasm has a very similar color to the background. Whereas, cytoplasm is a distinguishing feature of parasites with other components. Therefore, it required a strategy that can handle the noise and low difference intensity between background and foreground, so the segmentation results can

be used later as a basis for determining whether an object is a parasite or not.

In this paper, we proposed a new segmentation strategy for malaria parasite from microscopic images using integration of watershed and adaptive thresholding to overcome enormous noise on the segmentation process of low quality of thick blood smear. Segmentation consists of two stages namely the initial segmentation stage which aims to extract the core candidates parasite from thick blood smear and advanced segmentation which aims to extract the core and parasite citoplsma of candidate parasite obtained in the previous stage. This paper focuses on the second segmentation. While the initial segmentation is discussed in another paper.

II. Materials and Method

A. Materials

Microphotographs of the blood smears were prepared in Eijkman Institute for Molecular Biology, Indonesia. The 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 10×100 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.

The data used are 253 candidates malaria parasite size 34× 34 pixels (Figure 1b). The images are obtained from 22 thick blood smear images (Figure 1a). It looks very small and difficult to recognize by the human eye. Therefore, the candidates must be enlarged to make it easier to observe. In this paper, enlargement candidate uses Gaussian pyramid and it is done three times. Each expands operation doubles the size of the initial image. The formulation is defined [11] as follows.

$$G_{l,k}(i,j) = 4\sum_{m}\sum_{n} G_{l,k-1}\left(\frac{2i+m}{2}, \frac{2j+n}{2}\right) \quad (1)$$

where $G_{l,k}$ is obtained image by expanding G_l k times, while m and n are initial size image.

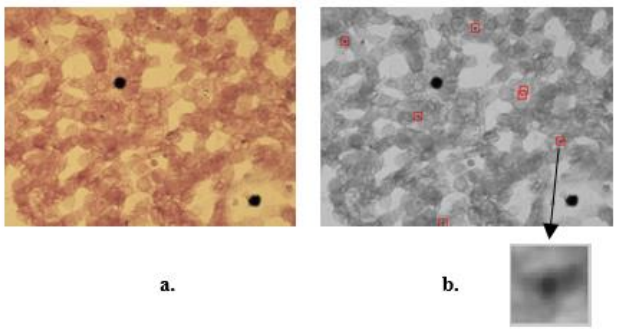

Figure 1: Thick blood smear image: a. Example image, b. Candidate parasite that cropped from graylevel image of thick blood smear

B. Method

The proposed method is formed of four steps: preprocessing to reduce noise and increase the contrast of image, Otsu thresholding, determine region minima, and segmentation (Figure 2).

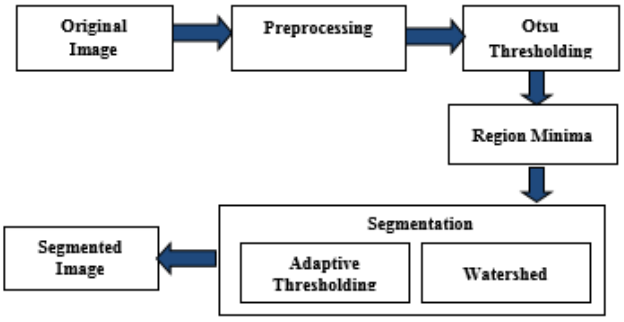

Figure 2: Flowchart of Proposed Method

a. Preprocessing

The aim of the preprocessing stage is to obtain images with more contrast than the original image. This process consists of two operations to increase the contrast of the raw acquired images: image filtering to mitigate noise and image enhancement, in the form of contrast stretching. In the first stage, a Gaussian low-pass filter is applied to the input image to get clear signal regions and suppress the influence of noise. The form of these filter, $H(u, v)$ [12] in two dimensions is given by

$$H(u,v) = e^{-D^2(u,v)/2D_0^2} \quad (2)$$

where D(u, v) is distance from center of frequency rectangle and D_0 = cut off frequency.

The next stage in preprocessing is contrast stretching. This operation spreads out intensity values along the total range of values in order to achieve higher contrast. It changes the distribution and range of the digital numbers assigned to each pixel in an image. This method is useful when an image has low contrast, such as images in which both the background and foreground are similar intensities. The form of these contrast enhancement [12] is given by

$$O_{ij} = \frac{1}{1+\left(\frac{m}{I_{ij}+\varepsilon}\right)^E} \quad (3)$$

where:
M = midline of switching from dark value to light value
O_{ij} = intensity of output pixel (i, j)
I_{ij} = intensity of input pixel (i, j)
E = slope function
ε = constant, distance between 1 and the next largest number

b. Otsu Thresholding

Parasite candidates image is complemented that aims to change the image of the object which is considered as lighter colored and dark-colored background then threshold by Otsu. This method [12] is defined by

$$\sigma_b^2(t) = \sigma^2 - \sigma_\omega^2(t) = \omega_1(t)\omega_2(t)[\mu_1(t) - \mu_2(t)] \quad (4)$$

Inter-class variance is expressed in ω_i class probability and average class μ_i, can be updated iteratively. Intra-class variance ($\sigma_\omega^2(t)$) can be defined as the sum of the variance of the class object and the background in accordance with Equation 5.

$$\sigma_{\omega}^2(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t) \quad (5)$$

weights ω_i is the probability of two classes separated by a threshold and σ_i^2 variance of classes i (i = 1 for the background and i = 2 for the foreground).

Implementation of Otsu thresholding aims to separate the object and background (Figure 3c). The results will be the basis for forming the watershed line (Figure 3d).

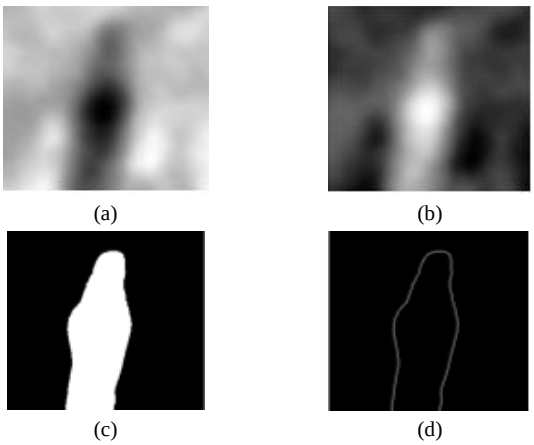

Figure 3: Otsu Thresholding process (a) Input Image (b) Complement Image (c) Otsu result (d) Watershed line

c. Determine region minima

Determination of region minima is taken from a point that satisfies the condition, in which the collection point will form a catchment basin. While the collection of points that do not meet the conditions will serve as the watershed line. The example below form 3 catchment basin marked by M1, M2, M3 and limit or watershed line is the highest area in the catchment basin meeting (Figure 4).

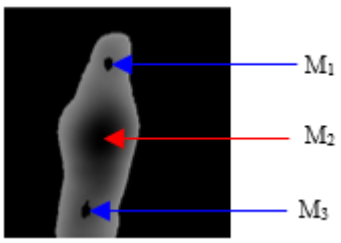

Figure 4: Determination of region minima

d. Segmentation

In the segmentation process, there are two parts of segmentation. First, image segmented with Watershed Distance Transform Methods. Due to the limited space a brief step by step explanation is provided in [9, 12].

The region minima obtained in the previous step made flooding to form catchment basin or area around the region minima. The search area is limited by the presence of dams that prevent merging of catchment basin. This dam boundary line relates to the division of the watershed region (Figure 5a).

Second, the region that formed by the segmentation process is performed by adaptive threshold. Intensity of the watershed region returned to its complement intensity and the intensity of the area outside the watershed region is equated with the intensity boundary watershed region (Figure 5b). Furthermore, the average image is calculated, that is, the image is convolved with a mean filter (50 × 50 mask). Then, the average image is compared with the value of each pixel in the input image [10]. Therefore, if the pixel in the original image is greater than T% of the pixel in the average image, then the pixel is labeled as 1 and it therefore belongs to the background. Conversely, if the pixel in the original image is less than T% of the pixel in the average image, then the new value of this pixel is labeled as 0. Therefore, it belongs to the signal.

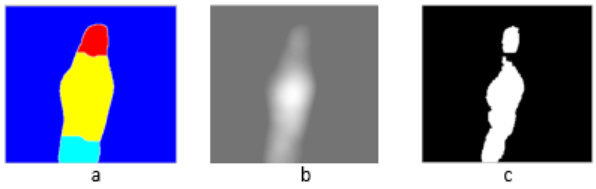

Figure 5: Segmentation Process (a) Formation of watershed region (b) Adjusted image intensity (c) Segmented result of adaptive threshold

C. Performance Measurement

Performance measurement of enhancement used Mean square error (MSE) which is used to calculate the measure of change in quality between the original image and preprocessed image and Peak Signal to Noise Ratio (PSNR) which is used to calculate the ratio of the maximum possible power of an image signal and the noise which affects the fidelity of the representation [13]. These measure defined as

$$MSE = \frac{1}{MN}\sum_{x=1}^{M}\sum_{y=1}^{N}(g(x,y) - f(x,y))^2 \quad (6)$$

where $g(x, y)$ is image before enhancement and $f(x,y)$ is image after enhancement.

$$PSNR = 10\log_{10}\left(\frac{(Max\ I)^2}{MSE}\right) \quad (7)$$

Performance measurement of segmentation is done by comparing the ground truth and the result of segmentation. The ground truth was built based on Malaria Diagnosis Guideline by Eijkman Institute for Molecular Biology and expert instruction. This segmentation accuracy, Acc, given by

$$Acc = \left[1 - \left(\frac{\text{no of pixels misclassified}}{\text{no of pixels in consideration}}\right)\right] \times 100\% \quad (8)$$

III. Result and Discussion

In the present work, enhancement and segmentation methodologies have been applied to 253 malaria parasite candidate images, but we show only 10 candidates in Figure 7. The result and discussion of enhancement and segmentation methodologies are described in detail as follows.

A. Enhancement

Tests performed on 10 data parasite candidate with the value parameter of the process low pass filtering D = 10. The contrast stretching parameters are m = 0.5, E = 4, and variations ε = 0.4, 0.5, and 0.6. Enhancement results as shown in Figure 6.

MSE and PSNR values of the images on enhancement process above are calculated by equation 6 and 7 respectively. The result presented in Table 1.

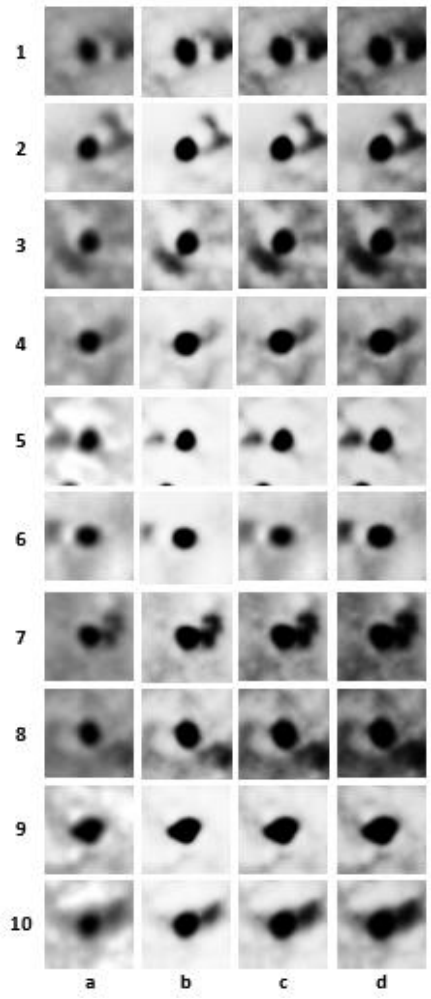

Figure 6: Three variation of parameters contrast stretching, ε values.
(a) Original image (b) ε = 0.4 (c) ε = 0.5 (d) ε = 0.6

Based on the results of the experiments above, enhancement process can improve the contrast between objects and the background so the resulting image becomes clearer than the original. Variation of ε value have an effect on the MSE and PSNR of the process. Result with ε = 0.4 produces a clearer image than others values. However, the increased image brightness must be considered. It does not only eliminate noise from the image, but also can result in the loss of information about the cytoplasm. This is because of the intensity of the cytoplasm which is sometimes similar to the surrounding background. As in the 2nd, 4th, 5th, and 6th images, there is an erosion of the citoplasma area due to its resemblance to the background. This will be bad for the next process. Result with ε = 0.5 give clear image and visible cytoplasm. It is also supported by the average of MSE and PSNR values that are better than others.

B. Segmentation

Segmentation is performed on the enhanced image of parasite candidates. Performance of proposed method is measured based on segmentation accuracy and it compares with segmentation method based Otsu thresholding and Fuzzy Similarity Measure (FSM) that proposed by Lopes et al. [14] and Pratamasunu [15]. This second approach is claimed as proper method to segment the image that has very irregular histogram because more suitable to deal with object edges and ambiguity, and it can avoid the problems involved in finding the minimum of a function. Very irregular histogram is one of the major problems in binary segmentation. It also becomes one of the problems in the segmentation of malaria parasite candidate on the thick blood smear image.

The segmentation result of candidate parasite images shows in Figure 7. Segmentation accuracy of the above three methods are calculated by Equation 7. The result summarized in Table 2.

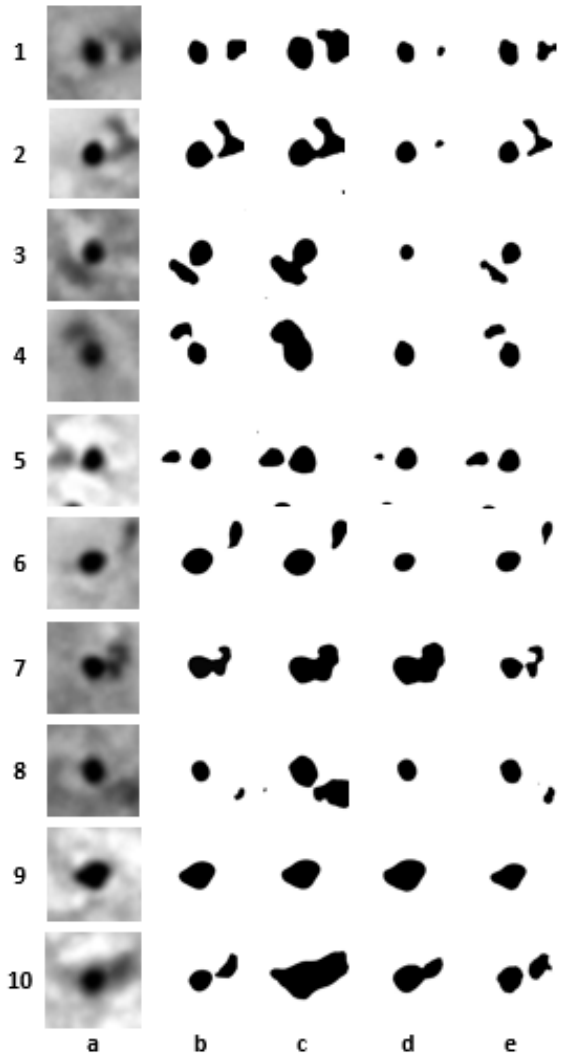

Figure 7: Segmentation result of candidate malaria parasite
(a) Input image (b) Groundthruth, (c) Segmentation result of Otsu, (d) FSM (e) proposed method

Visually, the segmentation result by Otsu method is over segmentation because the area is bigger than groundtruth which means non-object pixels (noncore and cytoplasm) are extracted too. Meanwhile, segmentation results using FSM have under segmentation because the resulting area is smaller than groundtruth. Some parasite images even lose their cytoplasm, as in the 3rd, 4th, 6th, and 8th candidates. The result of the segmentation using the proposed method tends to be under segmentation but the form is more similar to groundtruth than the other two methods. In addition, citoplasma can still be extracted. The existence of cytoplasm in the segmentation result is very important because it distinguishes a parasite nucleus with another object.

Table 1
MSE and PSNR Values of Enhancement Image

Candidate	e=0.4		e=0.5		e=0.6	
	MSE	PSNR	MSE	PSNR	MSE	PSNR
1	2792.4	13.67	488.0	21.25	1284.3	17.04
2	1844.3	15.47	417.9	21.92	626.2	20.16
3	2616.1	13.95	405.5	22.05	1497.1	16.38
4	2603.1	13.98	451.9	21.58	638.0	20.08
5	816.4	19.01	217.2	24.76	519.0	20.98
6	1598.0	16.10	343.9	22.77	396.5	22.15
7	2777.7	13.69	415.3	21.95	1410.0	16.64
8	2377.2	14.37	493.2	21.20	1841.6	15.48
9	1303.8	16.98	313.2	23.17	389.3	22.23
10	1393.5	16.69	406.4	22.04	751.1	19.37
Average	2012.3	15.39	395.3	22.27	935.3	19.05

Table 2
Segmentation Accuracy

Candidate	Acc (%)		
	Otsu	FSM	Proposed Method
1	89.6	94.9	96.6
2	92.2	89.6	93.0
3	92.7	90.0	93.9
4	89.1	94.7	96.1
5	92.5	95.9	96.5
6	96.1	91.5	93.9
7	84.6	83.0	89.0
8	84.7	95.7	95.1
9	96.6	95.4	96.8
10	81.9	93.3	94.2
Average	90.6	93.0	95.2

Quantitatively, the segmentation performance of the three methods is presented in Table 2. It is seen that the proposed method is superior to the other two methods on most images with an average accuracy of 95.2%. The strange thing happened to the 8th Candidate segmentation result. Although FSM obtains the highest accuracy, visually the segmentation result is incorrect because they cannot extract the cytoplasm.

Based on the results of the analysis visually and quantitatively, the proposed method can give more precise and accurate result than other method. Therefore it is proper to segment parasites candidate of malaria.

IV. CONCLUSION

A new segmentation strategy for low quality of thick blood smear is presented in this paper. There two stage i.e. enhancement and segmentation process. The use of low-pass filtering and contrast stretching on enhancement process can reduce noise and improve the contrast between malaria parasite and the background. Combination Watershed and adaptive thresholding are very effective to segment malaria parasites. It is characterized by obtaining the nucleus and cytoplasm of parasites on segmentation results. The proposed method has precise and accurate results in a segmentation of malaria parasites from the thick blood smear and achieves highest average segmentation accuracy than other method i.e. 95.2%.

ACKNOWLEDGMENT

This research was supported by Ministry of Research, Technology and Higher Education of the Republic of Indonesia under Grant PDD 2017 with contract number is 623/UN27.21/PP/2017. Authors thanks to Eijkman Institute for Molecular Biology Indonesia for preparing the materials for this research and granting exceptional access to the datasets.

REFERENCES

[1] WHO, World Malaria report 2014, Switzerland [Online]. Available: http: //www.who.int/malaria/publications/world_malaria_report_2014 /en/

[2] WHO, Basic Malaria Microscopy, Switzerland, 2010.

[3] D. Syafruddin, P.B Asih, I.E.P. Rozi, K. Chand, and S. Wangsamuda. *Diagnosis Mikroskopik Malaria*. Lembaga Biologi Molekuler Eijkman, 2010.

[4] H. Yang and N. Ahuja, "Automatic segmentation of granular objects in images: Combining local density clustering and gradient-barrier watershed", *Pattern Recognition* vol. 47, pp. 2266–2279, 2014.

[5] E. A. Mohammed, M. A. Mohamed, C. Naugler, and B. H. Far, "Chronic Lymphocytic Leukemia Cell Segmentation from Microscopic Blood Images Using Watershed Algorithm and Optimal Thresholding", in *26th IEEE Canadian Conference Of Electrical And Computer Engineering (CCECE)*, 2013.

[6] L. Xu, H. Lu, and M. Zhang, "Automatic segmentation of clustered quantum dots based on improved watershed transformation", *Digital Signal Processing*, vol. 34, pp. 108–115, 2014.

[7] A. Mouelhia, M. Sayadia, F. Fnaiecha, K. Mradb, and K. B. Romdhaneb, "Automatic image segmentation of nuclear stained breast tissue sections using color active contour model and an improved watershed method", *Biomedical Signal Processing and Control*, vol. 8, pp. 421– 436, 2013.

[8] S Kaewkamnerd, et al., "Detection and Classification Device for Malaria Parasites in Thick-Blood Films", in *The 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, September*, 2011.

[9] M. Elter, E. Haßlmeyer, and T. Zerfa, " Detection of malaria parasites in thick blood films", in *33rd Annual International Conference of the IEEE EMBS*, Boston, August 2011.

[10] J. E.Arco, J. M. Górriz, J. Ramírez, I. Álvarez, and C. G. Puntonet, "Digital Image Analysis for Automatic Enumeration of Malaria Parasites using Morphological Operations", *Expert Systems with Applications*, vol. 42, pp. 3041–3047, 2015.

[11] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden *Pyramid methods in image processing*. RCA Engineer, 1984.

[12] R. C. Gonzales and R. E. Woods, Digital Image Processing – Third Edition. Prentice Hall. Upper Saddle River, New Jersey, 2010.

[13] M. Ghosh, C. Chakraborty, A. Konar, and K. R.Ray," Development of hedge operator based fuzzy divergence measure and its aplication in segmentation of cronic myelogenous leukocytes from microscopic image of periperal blood smear", *Micron*, vol. 57 pp. 41 – 55, 2014.

[14] N. V. Lopes, P. A. Mogadouro, H. Bustince, and P. Melo-pinto, "Automatic Histogram Threshold Using Fuzzy Measures," *IEEE Trans. IMAGE Process.*, vol. 19, no. 1, pp. 199–204, 2010.

[15] G. Q. O. Pratamasunu, A. Z. Arifin, D. A. Navastara, A. Y. Wijaya, and W. N. Khotimah, "Image Thresholding Based on Index of Fuzziness and Fuzzy Similarity Measure," in *IEEE 8th International Workshop on Computational Intelligence and Applications, 2015*.