White blood cell counting analysis of blood smear images using various segmentation strategies

Cite as: AIP Conference Proceedings **1883**, 020018 (2017); https://doi.org/10.1063/1.5002036
Published Online: 14 September 2017

Syadia Nabilah Mohd Safuan, Razali Tomari, Wan Nurshazwani Wan Zakaria, et al.

White Blood Cell Counting Analysis of Blood Smear Images Using Various Segmentation Strategies

Syadia Nabilah Mohd Safuan[1,b)] Razali Tomari[1,a)] Wan Nurshazwani Wan Zakaria[1] and Nurmiza Othman[1]

[1]*Faculty of Electrical and Electronic Engineering*
Universiti Tun Hussein Onn Malaysia
86400, Parit Raja, Batu Pahat, Johor, Malaysia

[a)]Corresponding author: mdrazali@uthm.edu.my
[b)]ge160070@siswa.uthm.edu.my

Abstract. In white blood cell (*WBC*) diagnosis, the most crucial measurement parameter is the *WBC* counting. Such information is widely used to evaluate the effectiveness of cancer therapy and to diagnose several hidden infection within human body. The current practice of manual *WBC* counting is laborious and a very subjective assessment which leads to the invention of computer aided system (*CAS*) with rigorous image processing solution. In the *CAS* counting work, segmentation is the crucial step to ensure the accuracy of the counted cell. The optimal segmentation strategy that can work under various blood smeared image acquisition conditions is remain a great challenge. In this paper, a comparison between different segmentation methods based on color space analysis to get the best counting outcome is elaborated. Initially, color space correction is applied to the original blood smeared image to standardize the image color intensity level. Next, white blood cell segmentation is performed by using combination of several color analysis subtraction which are *RGB, CMYK* and *HSV*, and Otsu thresholding. Noises and unwanted regions that present after the segmentation process is eliminated by applying a combination of morphological and Connected Component Labelling (*CCL*) filter. Eventually, Circle Hough Transform (*CHT*) method is applied to the segmented image to estimate the number of *WBC* including the one under the clump region. From the experiment, it is found that G-S yields the best performance.

INTRODUCTION

Everybody needs a very defensive immune system as it can help the body to fight bacteria, viruses and diseases. Human body desperately needs that kind of system to prevent it from being affected by diseases. A person with low antibody system tends to get sick and affected easily with viruses and hence a good antibody with defensive immune system is crucial. *WBC* is said to be closely related to human body immune system as the number of *WBC* can provide information about human body condition. Analyzing blood smeared images is challenging but it is a very crucial task. *WBC* count is amazingly useful to diagnose diseases, to know the effectiveness of therapy for cancer patient and most of all, it plays an important role in body immune system to fight any infections [1][2]. Most blood disease can be detected by calculating the number of *WBC* in blood smeared image. Normal *WBC* count is in the range of 4500-10,000/μL and abnormal *WBC* count can be divided into two cases which are low *WBC* and high *WBC* count. The former is considered when the count is below 4500/μL and can results from diseases such as *HIV* and Lymphoma while the latter considered when the count is more than 10,000/μL can be indicator of diseases such as Anemia, Leukemia and tissue damage.

Traditionally, *WBC* counting is done manually which yields inaccurate result [3][4]. It is also highly dependent on the pathologist skills and as the blood sample increases, it will be more challenging to analyze it manually. But in today's current practice, the industry has came out with an automated Hematology Counter that is used to analyze *WBC*. However, such machines are costly and cannot be afforded by some countries and hence researchers developed a computer aided system (*CAS*) to fulfill such task. The *CAS* components are usually consists of four

Advances in Electrical and Electronic Engineering: From Theory to Applications
AIP Conf. Proc. 1883, 020018-1–020018-8; doi: 10.1063/1.5002036
Published by AIP Publishing. 978-0-7354-1563-8/$30.00

main modules which are preprocessing, segmentation, post processing and counting [4][5][6]. In the modules, the most important components are *WBC* segmentation and counting. Segmentation is responsible to prune out the object of interest in the image which in this case, is *WBC* region and it contributes to the accuracy of *WBC* detection stage. Segmentation method can be divided into several categories which are threshold based, learning based, active contour based, metaheuristic based and saliency based [2]. However, threshold based method is reported to be the best and reliable to segment the uniform image and fast. In [4], histogram thresholding is used to find the lower and upper threshold value of its saturation image. Other than that, [5] proposed Otsu thresholding to the *G* channel of *RGB* and found that it is able to detect the red blood cell (*RBC)* region. Otsu is said to be the most accurate threshold to segment the object of interest. While in [7], authors threshold the *RGB* image to *Y* component of *CMYK* color model as it gives the best contrast of *WBC* region.

Segmented image somehow will produce noise or unwanted region together with the region that is needed. This problem can be overcome by applying morphological filter which contains two operators which are dilation and erosion. It is an undeniably needed process to achieve the accurate counting of *WBC* region. There are several methods used for counting purposes that consider a clumped regions. Some of the methods are distance transform [8], edge detection [9] and Circle Hough Transform (*CHT*) [5][6]. Among this method, *CHT* is mentioned to perform better for estimating number of cell in the clumped regions [10].

In this paper, a comprehensive analysis of *WBC* segmentation that can provide an accurate *WBC* counting is elaborated. The analysis is conducted using various color space subtraction method. There are several selected single color channel representations from *RGB, CMYK* and *HSV* color analysis. Three subtraction analysis have been made for performance comparison purposes in terms of *WBC* counting. The rest of the paper is organized as follows: Section 2 explains about the system framework; Next, Section 3 discusses the result and analysis; Lastly, Section 4 presents the conclusion and recommendation for this system.

SYSTEM OVERVIEW

Figure 1 shows the block diagram of the proposed framework for rapid *WBC* counting analysis which consist of four main modules namely color space correction, *WBC* segmentation and extraction, morphological filter and *WBC* region counting. Firstly, the input image color is transformed using *L*a*b** color space in order to ensure the color intensity of an input image is corrected to a standard color characteristic. Next, the segmentation process is executed to prune out the *WBC* region from the background by using Otsu thresholding and combination of *RGB* and *CMYK* color space. The noises and unwanted foreground cells are removed by applying morphological filter. Lastly, CHT method is implemented to count the number of *WBC* including the overlapping cell.

All the mentioned processes and methods are applied and implemented on the first 30 images of ALL_IDB1 image database [11]. The outcome of the counting is compared with the ground truth to determine the feasibility of the proposed method.

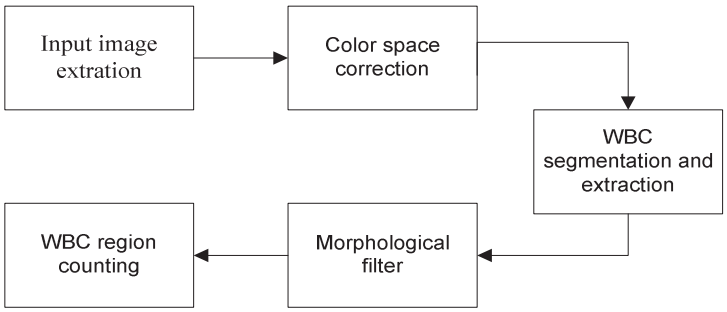


FIGURE 1. Overview of proposed framework

Color Space Correction

The input image that is obtained may differ from one another in terms of its color intensity. It is because of the various acquisition conditions while the blood smeared images were taken. Different microscope settings can lead to different conditions, lighting and color intensity of the image. This problem needs to be overcome and reduced by

correcting the color of input image to a one standard color. The standardize color can be obtained by manipulating the targeted image. There are targeted images that were used to lead the original source image. In this paper, there are four targeted images which are also known as template image. This approach maps the colour distribution of an over or under stained image to that well stained target image. This method represents the matching colour distribution of the source image and the target image by using the *L*a*b** colour space. *L*a*b** colour space is used in this step because it minimizes correlation between channels for many natural scenes [9].

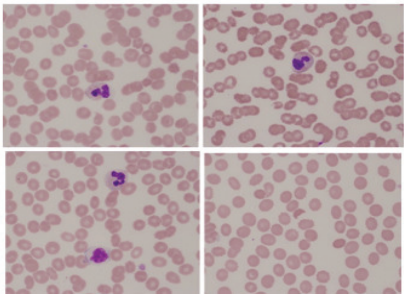
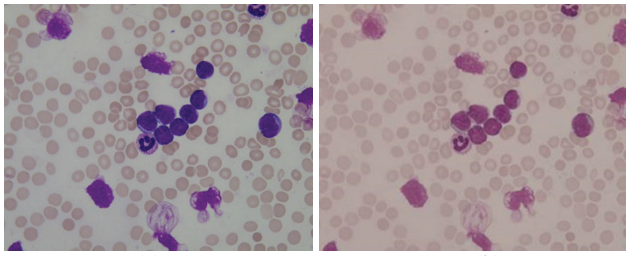


FIGURE 2. Template images

The template images are as shown in Fig. 2. The value of *mean* (μ) and *standard deviation* (σ) of *l, a* and *b* of the template images and original source image was obtained and the values were substituted in the equation below. *Im* is the single color representation of channel *l, a* and *b*. The value of *l, a* and *b* of corrected image is computed by executing the equation (1).

$$\mathrm{Im}_{corrected} = \frac{\mathrm{Im}_{source} - \mu_{\mathrm{Im}\,source}}{\sigma_{\mathrm{Im}\,source}} \sigma_{\mathrm{Im}\,template} + \mathrm{Im}_{template} \tag{1}$$



(a) (b)

FIGURE 3. (a) Sample of original source image (b) Result of color space correction

In this paper, instead of extracting the template image, the standard template value was obtained and applied on the original source image. The template value was obtained by computing the average value of *mean* (μ) and *standard deviation* (σ) of every channel of *l, a* and *b*. The average value of *mean* (μ) for *l, a* and *b* is 81.99, 4.56 and 1.22 respectively. While the average value of standard deviation of *l, a* and *b* is 4.79, 4.72 and 3.09 respectively. All these values have been substituted in the equation to finally come out with corrected image value and the value is implemented on the source image. Fig. 3(a) shows a sample of original image that was used for color correction and the result is as shown as Fig. 3(b).

WBC Segmentation and Extraction

After the color of blood smeared image has been standardized, it is easier for the segmentation process which is the background and red blood cell elimination. However, it is a challenging task to segment the *WBC* as it has various shapes and contains both cytoplasm and nucleus. Cytoplasm has slightly different color compared to nucleus. In order to segment and do the counting for *WBC*, only nucleus is matter. By including the cytoplasm for

segmentation, it yields inaccurate result of counting process as the size is bigger and the system might count it as overlapping cell. In this paper, the segmentation process was done by thresholding the image using Otsu thresholding to find the color components of *CMYK* and *RGB*. Combination of *CMYK* and *RGB* is meant to prune out the *WBC* region only with minimal appearance of its cytoplasm and eliminates other blood components.

The samples of every color component analysis of *R,G,B,C,M,Y,K,H,S* and *V* were taken. In this paper, there are only several color analysis that have been taken to be compared to each other which are *G, H, Y* and *S* as shown in Fig. 4. Color component *G* of *RGB* shows the segmentation of *WBC* nucleus with minimum appearance of its cytoplasm and gives the best contrast while color component *Y* of *CMYK* shows the segmentation of whole *WBC* region including the cytoplasm. The combination of two color analysis to be subtracted was done and result of subtraction between two color component *H* and *Y*, *G* and *Y*, *G* and *S*, is as shown as Fig. 5(a), 6(a) and 7(a) respectively. It can be seen that the image's background and red blood cell has been eliminated and left only the *WBC* region. However, there are some noises and unwanted cell region in the image that needs to be distinguished. Hence, morphological filter was applied which uses both dilation and erosion once. Color Component Labelling (*CCL*) was also implemented to remove the unneeded small particles. Particles which below 150 pixels were removed from the images and leaves only the needed particles which is the *WBC* region without any noises and unwanted particles. These improved segmented results are important in order to achieve high counting accuracy. In order to get the accurate counting result, segmentation process must be analyzed to achieve the best method for the process.

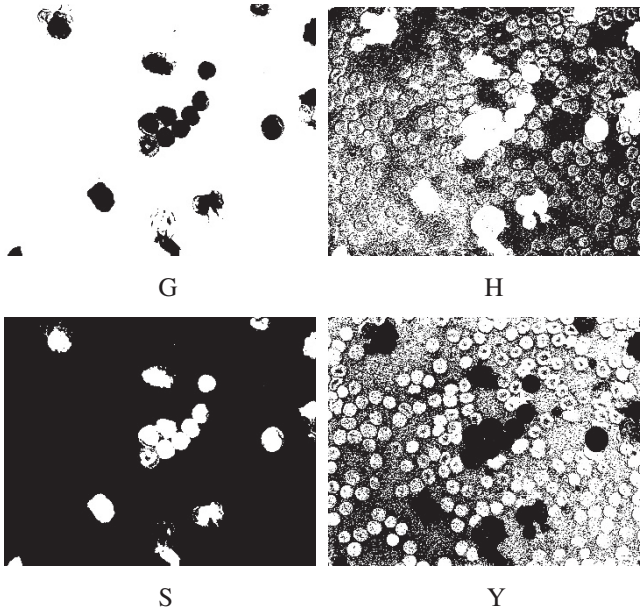
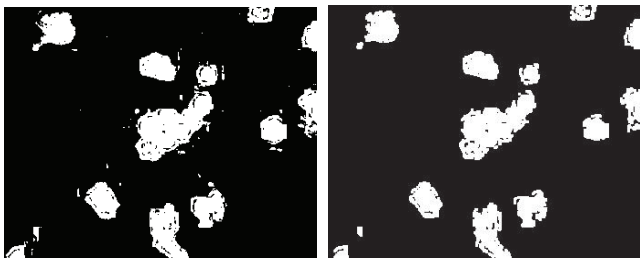


FIGURE 4. G, H, S and Y single color component

H and Y Subtraction

For the first attempt, we used *H* of *HSV* and *Y* of *CMYK* color analysis. It can be seen in Fig. 4 that color component *H* of *HSV* gives the largest area of *WBC* segmentation while the background has been eliminated from the image. While for *Y* of *CMYK* color component, the largest area of *WBC* was eliminated and left only red blood cell region and the background. The subtraction of both color component leaves the biggest region of *WBC* including the cytoplasm area is as shown as Fig. 5(a) and sample of result after morphological filter and *CCL* is as shown as Fig. 5(b). It is the image used for counting process to calculate the number of extracted *WBC* region in blood smeared image.

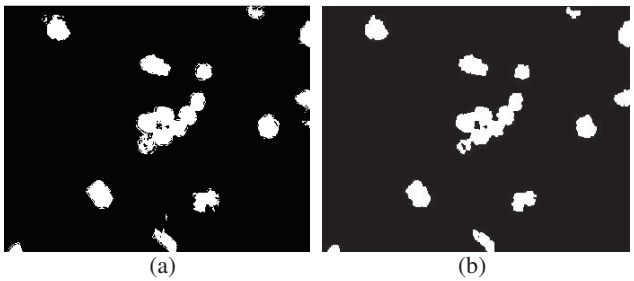


(a) (b)

FIGURE 5. (a) H-Y subtract image (b) Subtracted image after CCL

G and Y Subtraction

G of *RGB* gave the *WBC* region and eliminates all the red blood cell area as well as the background. *G* gave the best contrast as it has wide histogram. The result of *G* and *Y* subtract is as shown as Fig. 6(a) which the noises are not as much as *H* and *Y* subtract. However, there were still some noises that need to be distinguished to reduce the segmentation error. Segmentation result will eventually affect the counting result. Fig. 6(b) shows the image result after *CCL* is applied to the morphological image. The image shown is slightly different from result of *H* and *Y* subtract.

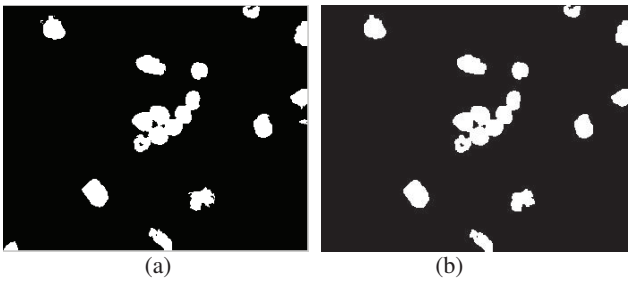


(a) (b)

FIGURE 6. (a) G-Y subtract image (b) Subtracted image after CCL

G and S Subtraction

The reason why G component of *RGB* was chosen is because it contains only the *WBC* region's area and eliminates the red blood cell and background. It also gives the best contrast compared to R and B color component. Furthermore, it has the least area of cytoplasm and focuses on segmenting the nucleus area. Next, *S* color component of *HSV* shows that *WBC* appeared in white colored area and all other components in blood smeared image has been distinguished. It is also has the least area of cytoplasm. This pair was chosen to test the subtraction result of the color component that contains least area of cytoplasm. Same goes to previous method, the subtracted image still contains noises and it is better after application of morphological filter followed by *CCL*. The *G* and *S* subtracted image and result after *CCL* image is as shown in Fig. 7(a) and 7(b) respectively.



(a) (b)

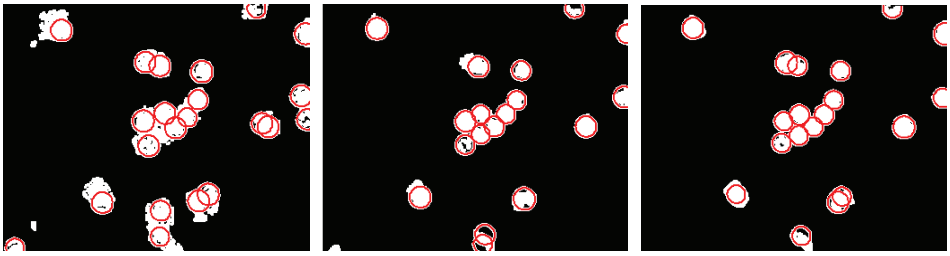
FIGURE 7. (a) G-S subtract image (b) Subtracted image after CCL

WBC Counting

Since there are various methods of subtractions that have been done in the previous process, the result was compared to each other in order to get the view of the best method to calculate the number of *WBC*. The image that was used forward for counting process is the image after the application of *CCL*. Since the images are free from noises and unwanted regions, Circle Hough Transform (*CHT*) was applied next. *CHT* helps to circle the object of interest by considering clumped or overlapped region. Currently, *CHT* method is widely used for shape detection purposes. The equation used for circle detection is as shown in the equation (2) below. The center is defined as a_i and b_i where θ is the angle through 360° range. X_i and Y_i will trace parameters of the circle. The range of radius has been set from 10 to 12 as the radius of detected *WBC* region lies in between the range. The counting process was done to all the images from *HY* subtract, *GY* subtract and *GS* subtract to be compared in the result and analysis section.

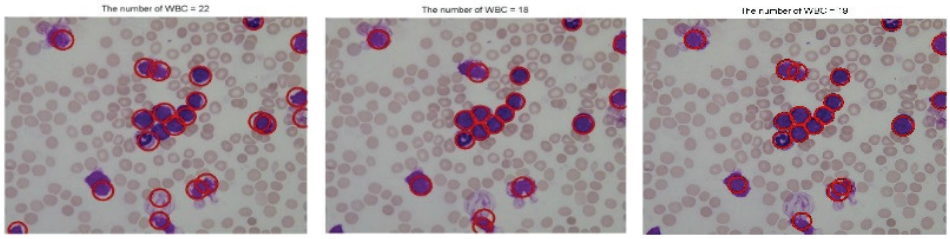
$$X_i = a_i + r_i \cos(\theta) \qquad Y_i = b_i + r_i \sin(\theta) \tag{2}$$

Fig. 8(a), 8(b) and 8(c) shows circle detection for the improved segmented image, while Fig. 9(a), 9(b) and 9(c) shows samples of counted *WBC* region in blood smeared image of *H-Y, G-Y* and *G-S* subtraction method with the appearance of *WBC* number. The results were obtained for 30 images in ALL-IDB1 database. The counting performance for each subtraction method was observed and computed. The result is discussed in result and analysis section.



(a) (b) (c)

FIGURE 8. (a) CHT on H-Y subtract (b) CHT on G-Y subtract **(c)** CHT on G-S subtract



(a) (b) (c)

FIGURE 9. (a) Final result of H-Y subtract (b) Final result of G-Y subtract (c) Final result of G-S subtract

RESULT AND ANALYSIS

The various methods that have been used for *WBC* segmentation was compared in terms of its accuracy of the *WBC* counting. The correctness of *WBC* detection contributes to the counting accuracy. Graphic User Interface (*GUI*) has been developed to make the system a user friendly system. The *GUI* contains two axes to display figures and a push button to analyze the *WBC* rapidly. It will call all the images that have been stored in a database and process the images to get the final result of *WBC* counting. The final images were stored in one folder to make it easier for the

pathologist to do the review of the result. The *GUI* window is as shown in Fig. 10 below with the image number and result image was displayed side by side.

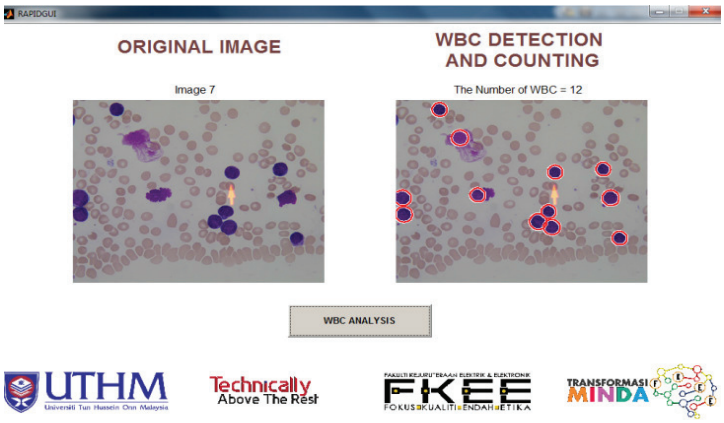


FIGURE 10. GUI window for rapid WBC analysis

Quantitative analysis of counting was taken using *CHT* which also considers the clumped and overlapping cell's region. This analysis shows the best method between the three methods and performance of the whole system. It was tested on images that are in a database called ALL-IDB1 database. There are 108 images in total in the database but only the first 30 images were used because of their consistent magnification factor and color intensity. Both magnification factor and color consistency are considered in the selection of images. All the 30 images gone through the same processes for each subtraction methods. The result of counting for each subtraction method has been compared to the ground truth data which is the manual counting in order to get the differences and traces the best subtraction method. The performance was assessed based on 30 images in total. Next, it has been shown in Table 1 that *G-S* subtraction method is able to get the highest accuracy in terms of *WBC* counting and able to segment and detect the *WBC* region well by gaining 97.79% average accuracy of 30 images. It can be seen that by applying *G-S* subtraction method, the counting accuracy is higher compared to other subtraction methods on most images. This is because *G* and *S* subtraction method does not include the area of cytoplasm as it can confuse Hough Transform to count the region as clumped or overlapping region. By taking only the nucleus area and minimum area of cytoplasm, the counting accuracy can be increased. By analyzing the data and result, it can be concluded that in order to achieve better accuracy of *WBC* counting, it is important to only include the nucleus segmentation. Cytoplasm segmentation should not be considered in *WBC* counting purposes because of its shape and area irregularities.

TABLE 1. Average counting accuracy for each method

SUBTRACTION METHOD	H-Y	G-Y	G-S
ACCURACY	42.41%	93.77%	97.79%

CONCLUSION AND FUTURE WORK

In this paper, various subtraction methods were used to calculate the accuracy of *WBC* counting. This automated *WBC* counting was developed under computer vision system for image processing. Firstly, original input image is obtained and the color intensity was transformed to a standardize color characteristic. The images were rescaled to speed up the processing time. Larger image took longer time to process than the smaller scale image. The standard values of *mean* and *standard deviation* of *L*a*b** color space correction was determined. The mean values of *l, a* and *b* are 81.99, 4.56 and 1.22 respectively while standard deviation values are 4.79, 4.72 and 3.09. This color corrected image is needed to ensure the correctness of *WBC* segmentation. Next, the image was segmented using the combination of *RGB, CMYK* and *HSV* color analysis together with Otsu thresholding. There are three subtraction methods applied through the process which are the subtraction of *H-Y, G-Y* and *G-S* color analysis. The results are

used to be compared with respect to the accuracy of *WBC* counting performance based on various segmentation strategies. The performance can be considered as a good performance as it achieves more than 90% counting accuracy.

In future, this system can be improved by considering red blood cell count as well and obtain a Complete Blood Count (*CBC*). Furthermore, it is expected to be able to classify all five different types of *WBC* in blood smeared image.

ACKNOWLEDGEMENT

The authors would like to thank to Ministry of Education (MOE) and Universiti Tun Hussein Onn Malaysia (UTHM) for supporting this research under Fundamental Research Grant Scheme (Vot. no. 1582).

REFERENCE

1. Prinyakupt J., Pluempitiwiriyawej C. Biomedical Eng. **14**, 1-19, (2015).
2. Zhang C., Xiao X., Li X., Chen Y., Zhen W., Chang J., Zheng C., Liu Z. Sensors. **14**, 16128-16147 (2014).
3. Priyankara GP, Seneviratne OW, Silva RK, Soysa WV, De Silva CR. An extensible computer vision application for blood cell recognition and analysis. Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka. 2006.
4. Venkatalakshmi B, Thilagavathi K. *IEEE Conference on Information & Communication Technologies (ICT)*. (IEEE, 2013), pp. 267-271.
5. Tomari R., Wan Zakaria W.N., Jamil M.M.A, Nor F.M, Fuad N.F.N. J. of Proc. Comp. Science. **42**, 206-213 (2014).
6. Reddy V.H. Int. J. of Research in Advent Tech. **2**, 294-299 (2014).
7. Putzu L., Ruberto C.D. Int. J. of Medical, Health, Biomed., Bioeng. and Pharmaceutical Eng. **363**, 20-27, 2013.
8. Khan HA, Maruf GM. *International Conference Informatics, Electronics & Vision (ICIEV)*. (IEEE, 2013), pp. 1-6).
9. Asmitaba G., Pipalia D. Int. J. of Advanced Research in Electrical, Electronics and Instrumentation Eng. **5**, 3807-3812 (2016).
10. Tomari R., Wan Zakaria W.N., Ngadengon R., Abd Wahab M.H. ARPN J. of Eng. and Applied Sciences. **10**, 1413-1420 (2015).
11. R. Donida Labati, V. Piuri, F. Scotti. *Int. Conf. on Image Processing*. Brussels, 2011. (IEEE, 2011), pp. 2045-2048.