Measurement 116 (2018) 543–555

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

White blood cell (WBC) counting analysis in blood smear images using various color segmentation methods

Syadia Nabilah Mohd Safuan, Mohd Razali Md Tomari*, Wan Nurshazwani Wan Zakaria

Research Centre for Applied Electromagnetics (EMcenter), Department of Mechatronic and Robotic Engineering, Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 86400, Parit Raja, Batu Pahat, Johor, Malaysia

ARTICLE INFO

Keywords:
White blood cell segmentation
Colour space analysis
Hough transform

ABSTRACT

White blood cells (WBCs) is closely related to human immunity system which is useful to fight viruses and bacteria. Cancer therapy effectiveness and some of the blood-related diseases can be determined from the count of cell region. Traditionally, WBC count is done manually which yields inaccurate results as the blood sample increases. Moreover, even though hematology counter device is fast and accurate, it is very costly. These problems has led to the invention of cost effective Computer Aided System (CAS) which analyzes the blood smeared images obtained from microscope. In CAS, the most important step is segmentation and any failure at this stage will cause inaccuracy in the subsequent stages. Realizing the importance of segmentation, this paper investigated various segmentation methods for the purpose of WBC counting based on color band thresholding procedure. Initially, color space correction based on L∗a∗b∗ color space was applied to standardize the image color intensity. Next, segmentation process was conducted to prune out the WBC region from the background by combining color analysis of RGB, CMYK and HSV with Otsu thresholding. Morphological filter was employed as the segmented image contained noises would affect the system performance. Henceforth, Connected Component Labelling (CCL) was done to distinguish the small particles that still existed in the image. Eventually, Circle Hough Transform (CHT) was applied to identify and count the WBC including the one in the clump region. Overall system performance was accessed and it was found that by using S color component of HSV, color space provided the highest WBC segmentation accuracy which was 96.92%. Meanwhile, for the combination of two color bands; S-C produced 96.56% of WBC counting accuracy. Another interesting finding was that the usage of nucleus based detection was superior compared to cytoplasm-based detection for WBC counting purpose.

1. Introduction

White Blood Cell (WBC) or leukocyte is one of the elements in human blood besides Red Blood Cell (RBC) and platelet. Every blood cell has its own function. The RBC for an example, helps to transport oxygen from the lung to each part of the body [1]. The WBC functions to fight diseases, viruses and bacteria [2]. We are surrounded by different kind of people every day. Under this circumstances, we are also exposed to the possibility of getting contracted by viruses. Therefore, a strong antibody system is crucial and having a number of WBC can greatly help to build such system. A normal WBC count is in the range of 4500–10,000 (μl) while an abnormal WBC count has either a lower or higher than the normal reading. A patient with a low WBC count can possibly be diagnosed of having diseases such as HIV and Lymphoma while high WBC count can be indicator or one of the symptoms of diseases such as Anemia, Leukemia and tissue damage. Basically, if a patient complains about their continuous sickness, the first test that will be performed by a doctor is the blood test since it will definitely explain a lot about patient's health condition. Subsequently, an early treatment can be made to prevent any serious illness. Apart from that, the WBC count can also contribute to the information of the effectiveness of a cancer therapy. If the patient is recovering, it indicates that the therapy works but if the condition is worsened, the therapy has to be stopped. Other than that, by counting the amount of WBC in the body, an individual's level of s immunity system can be determined. Even though blood test is not a part of any treatment, it is very useful to diagnose sickness that has infected the patient. Using such hints, doctors can take further actions regarding the diseases and prevent the viruses from spreading all over the body.

Traditionally, methods of detecting and counting WBC were conducted manually by a pathologist [3,4]. However, this manual practice could potentially create problems as it yields inaccurate result which very much depends on the pathologist's skills and experiences [5]. Different pathologists might achieve different kinds of results which

* Corresponding author.
E-mail addresses: ge160070@siswa.uthm.edu.my (S.N. Mohd Safuan), mdrazali@uthm.edu.my (M.R. Md Tomari), shazwani@uthm.edu.my (W.N. Wan Zakaria).

https://doi.org/10.1016/j.measurement.2017.11.002
Received 2 May 2017; Received in revised form 31 October 2017; Accepted 2 November 2017
Available online 10 November 2017
0263-2241/

will create confusion. Other than that, as the sample increases, the workload that the pathologist has to perform will proportionally increase [6]. Since there are thousands of cells lie in the blood smeared image, it will be more challenging for the pathologist to analyze it manually. Inevitably, the whole manual practice process is time consuming and sometimes requires a day for the result to come out. Nowadays, industry has come out with a hematology counter which promises a more accurate and faster result. However, these types of machines are costly [7] and cannot be afforded by some developing countries to be placed in each of their hospitals. The machines are also not portable which makes it hard to conduct blood tests at rural areas. Thus, a solution is needed to overcome these problems that will certainly be a great help to pathologists.

One of the efficient and cost effective WBC counting processes is based on Computer Aided System (CAS) concept in which the system can be tuned to work as good as human eyes and sometimes even better [8]. In the CAS framework, the processes can be divided into two main parts which are WBC identification and WBC counting. Prior to processing WBC, the CAS must be able to localize and segment the respective WBC region accurately. However, such task is challenging due to the complexity of WBC shape that consists of cytoplasm and nucleus region in which nucleus is reside inside the cytoplasm with the same color type but different level of intensity. Segmentation is reported to be a very important task as the accuracy of WBC counting is highly dependent on this process [9]. It can be divided into five main categories which are threshold-based methods, learning-based methods, active-contour-based methods, metaheuristic-based methods and saliency-based methods [2]. All these categories can be used for segmentation purposes based on different kind of image types. As for uniform type of image like blood image, a threshold-based method is reported to be the best and produce reliable performance with high running speed [2]. One of the works proposed is Otsu thresholding application in the green channel of Red Green Blue (RGB) color analysis to segment the RBC region [10]. Other than that, histogram thresholding is used to find the optimal values of threshold to segment the region of WBC [11]. While in [12], K-mean clustering is applied for initial segmentation of the cell nucleus. As for [13], the RGB image is converted to Cyan Magenta Yellow Key (CMYK) color analysis and only the Yellow (Y) color component is taken as the region of interest is more contrast in Y color space. Watershed seed region growing method and application of a threshold value of 90 to set the internal marker of WBC nucleus is proposed in [14]. Some works used Saturation (S) and Value (V) color component of Hue Saturation Value (HSV) together with Zack thresholding as applied in [15] to identify the parasites and RBC. Grayscale image is threshold to binaries the image and prune out the object of interest as proposed by [16].

For the WBC classification task, the nucleus and cytoplasm regions are needed. However, for the counting purpose, only information from one region is sufficient to predict the number of WBC including in the clump area. In this paper we investigate the suitable WBC information that can be used as a hint for the counting requirement based on the analysis of various color band information. The outcome will be very useful to the CAS-WBC framework to speed up the processing time based on simple thresholding strategy. The rest of the paper is organized as follows: Section 2 presents the architecture of the proposed system with various kind of methods; Next, Section 3 shows experimental results with discussion; and finally, the conclusions and future research are presented in Section 4.

2. System overview

Fig. 1 shows block diagram overview of the framework for WBC segmentation and counting. It contains of four main blocks which are color space correction, WBC segmentation and extraction, morphological filter and lastly, WBC counting. The input image is obtained from a light microscope and also known as blood smear image. Initially, to cater the issue regarding the image acquisition variability condition, the image's color intensity needs to be transformed to a standard color characteristic. The color correction process using L∗a∗b∗ color space correction [17] is done to the original image. Next, the WBC segmentation is done by removing RBC and background using Otsu thresholding and combination of RGB, CMYK and HSV color space analysis. Morphological filter is used to remove the remaining noises after segmentation process. Lastly, WBC region including the overlapping cell is counted using Circle Hough Transform (CHT) method. All the processes are done and the performance is evaluated by collecting and comparing the system's result with ground truth data to calculate the WBC counting accuracy.

The whole system and process focus on the images that are obtained from same instrument, with similar resolution and magnification factor. For application to other resolution and magnification factor, a constant multiplier need to be tuned to obtain appropriate image ratio setting. For this project, the analysis was applied to ALL-IDB database [23] that has been made public by Università degli Studi di Milano. The first 30 images in ALL-IDB (1) datasets were chosen since they have consistent magnification factor of 400 with the image resolution of 2592 × 1944. The images consist of RBC, WBC and background and they were acquired by capturing the blood sample using optical microscope and a Canon PowerShot G5 camera. All input images are in JPG format and the resultant images were also converted to the same format as well.

2.1. Color space correction

The blood smeared images taken from various light microscopes may produce a slight difference in terms of their color intensity as the image acquisition process condition might be different. Lighting condition and different device used while taking the samples is part of the problems that need an attention to be fixed as it is important to standardize the image's intensity in the system. One of the standardization methods is based on Color space correction since it can be used to transform the acquired input image to a standard color characteristic template. In this system, L∗a∗b color space is used to serve such purpose as it minimizes the correlation between channels for many natural scenes [17]. The proposed method maps the original image, which can be over or under the stained image to the correctly stained image for the next process. Basically, the method works by matching the original image color intensity with the targeted image color intensity. In this system, four template images that produce a good WBC counting are used as targeted image as shown in Fig. 2 and utilize as a benchmark for all input images during standardization process.

The mapping process starts by obtaining the value of mean (μ) and standard deviation (σ) of L∗a∗b color space for both template and input image. Basically, the template images and original image are in RGB form which are then transformed into L∗a∗b representation. These values of mean and standard deviation are substituted in the Eq. (1) below. *Im* is the single color representation of l, 'a' and 'b'. The equation is to find the corrected value of l, 'a' and 'b' to be applied to the original image and produce a new image with a different color intensity.

$$Im_{corrected} = \frac{Im_{source} - \mu_{Imsource}}{\sigma_{Imsource}} \sigma_{Imsource} + Im_{template} \quad (1)$$

However, in this research, instead of using the image itself as the benchmark, the standard or fixed value of mean (μ) and standard deviation (σ) of the template images were obtained by calculating the average for both mean (μ) and standard deviation (σ) for four template images. Each template image's value of mean (μ) and standard deviation (σ) was collected and the average for these parameters was calculated. The obtained average mean (μ) values of l, a and b were 81.9863, 4.5579 and 1.2176 respectively while the average value of standard deviation (σ) for l, a and b were 4.7984, 4.7248 and 3.0947. The difference between original image and color corrected image results is as shown in Fig. 3 where Fig. 3(a) shows the original image

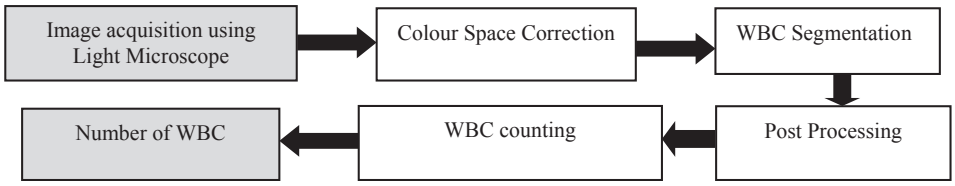

Fig. 1. General Overview of the proposed framework.

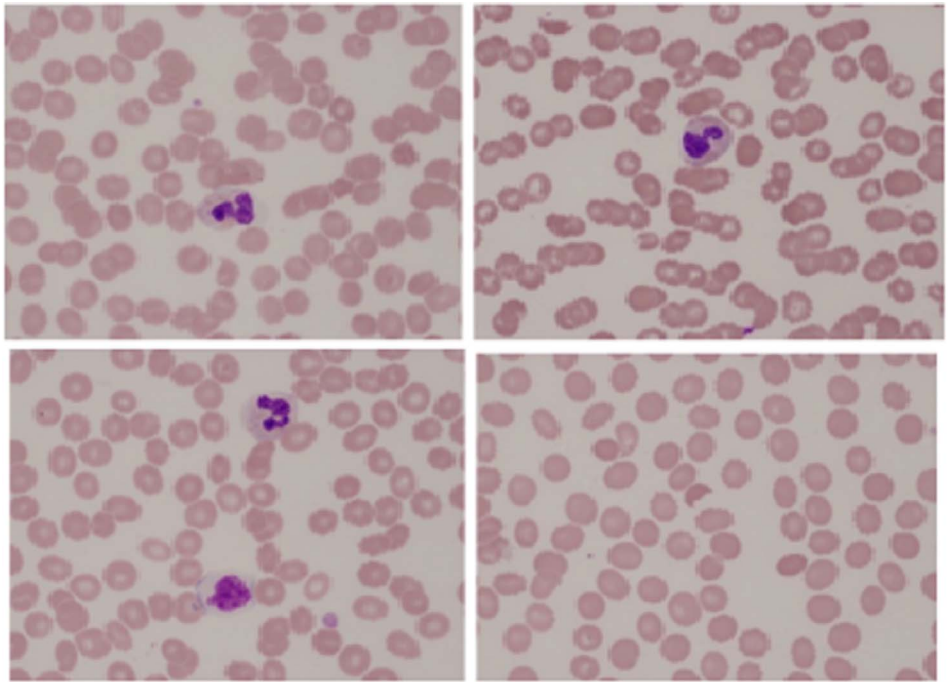

Fig. 2. Template images.

while Fig. 3(b) shows the image after color correction process. It can be seen that there was no difference between the two images in terms of the shape and structure but there was a difference in terms of its color intensity. The image's intensity of the color corrected image was slightly different and the appearance of background and RBC was reduced.

2.2. WBC segmentation and extraction

After the color correction phase, the image needs to be segmented to localize the respective WBC region. The localization of WBC is accomplished by eliminating other regions and particles such as RBC, platelets and background. However, segmenting the WBC region is quite challenging as its morphological features is not consistent. This is because WBC consists of two parts which are nucleus and cytoplasm. Both parts can be differentiated by the color intensity level as nucleus has higher color intensity than the cytoplasm and nucleus is the inner part of WBC while cytoplasm is the outer part. In this research, segmentation process was done by applying the Otsu thresholding on a color space analysis of RGB, CMYK and HSV. On top of that, two different kinds of approaches were used to segment the WBC region. The first approach was by using the single band color analysis and the second approach was by using a combination of two chosen single bands to determine the best result of WBC counting. Single color band is useful in locating the nucleus while the combination of color band can be used to detect both the nucleus and cytoplasm area which will be useful for WBC identification.

2.2.1. Single color band analysis

There were ten color space components which were *R*, *G*, *B*, *C*, *M*, *Y*, *K*, *H*, *S* and *V* investigated to prune out the area of WBC. All of these color spaces went through the thresholding process using Otsu method

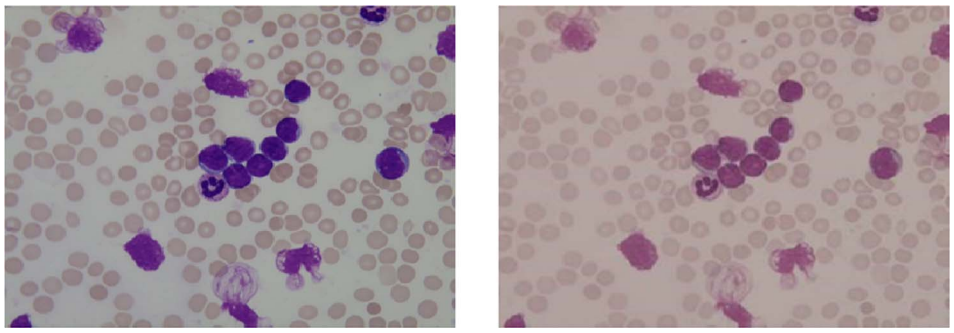

Fig. 3. (a) Original source image. (b) Colour corrected image using the template images parameters.

Fig. 4a. Sample of ground truth data for: (i) Nucleus area identification (ii) Cytoplasm area identification.

[18] and the performance was initially assessed qualitatively with respect to the minimal appearance of cytoplasm and noise. Sample of the thresholding result is depicted in Fig. 4(b). The comparison between ground truth which is Fig. 4(a)(i) and the thresholding result in Fig. 4(b) is assessed. From the figure, a few conclusions could be basically drawn which are as follows:

a. The thresholding of color band of *R, G, S, C, M* and *K* was suitable for locating the nucleus region since information about the *RBC* was totally eliminated.
b. Segmented of *B* color band highlighted the region of the *WBC* (nucleus) and *RBC* with a good quality.
c. *H* and *Y* color bands were able to prune out the whole region of *WBC* (cytoplasm and nucleus) with a good accuracy The *RBC* region was also highlighted along the way.
d. The formula of *B* – (*color band in a*) could be used for detecting the *RBC* region only.
e. The whole WBC region (cytoplasm and nucleus) could be highlighted using formula of *H* – (output of color band in d) or *Y* – (output of color band in d).

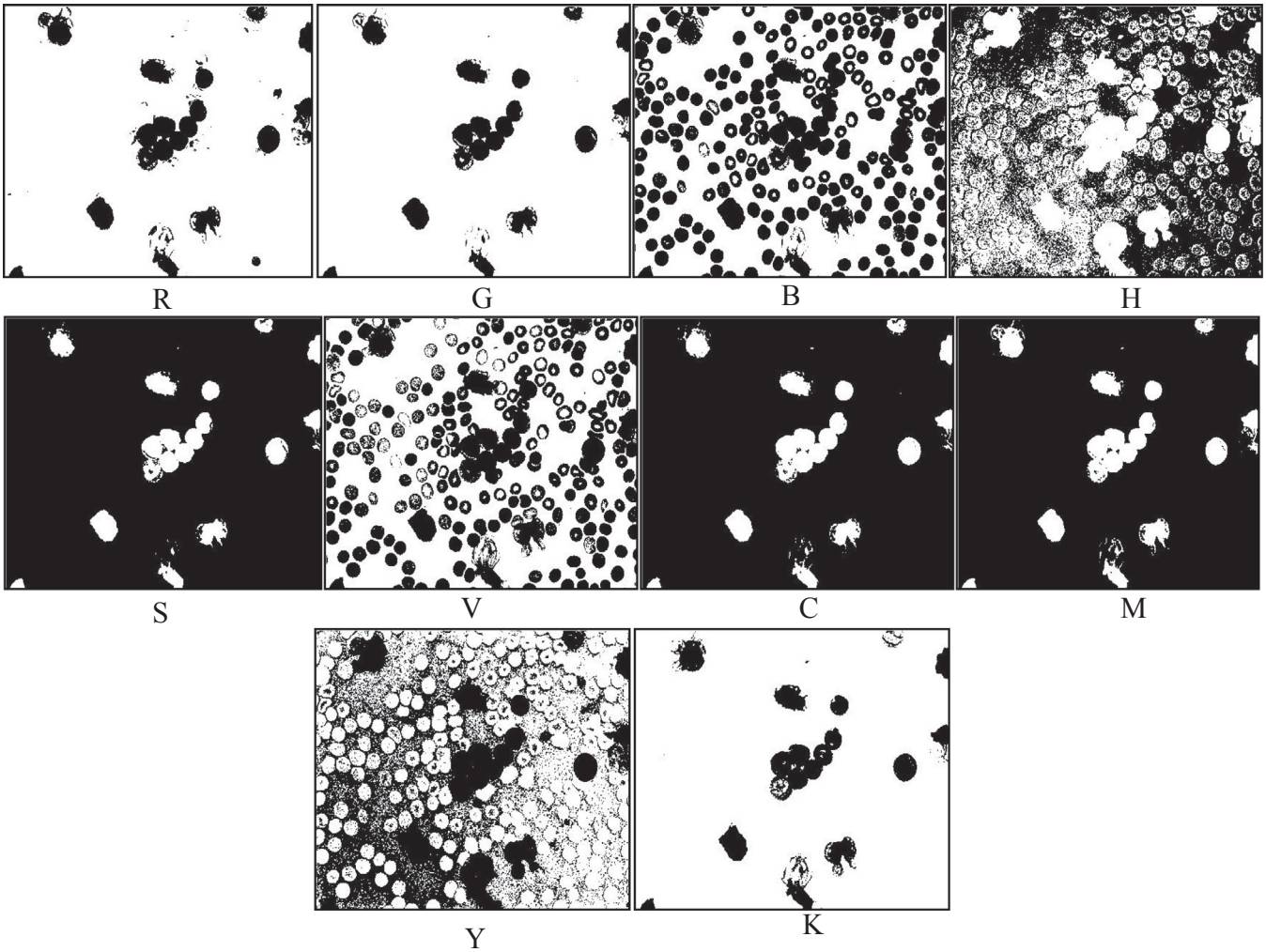

Fig. 4b. Result of color channel segmentation of green (*G*), red (*R*), blue (*B*), hue (*H*), saturation (*S*), value(*V*), cyan(*C*), magenta(*M*), yellow (*Y*) and key(K).

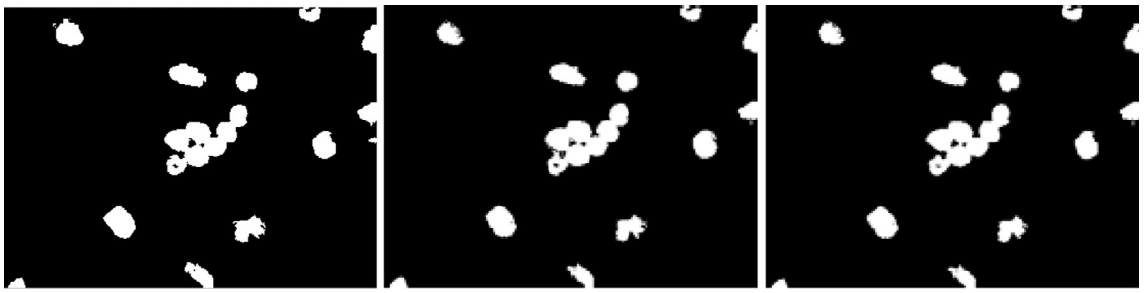

Fig. 5. (a) S-G subtracted image (b) C-G subtracted image (c) S-C subtracted image.

2.2.2. Nucleus detection based on combination of two color bands analysis

The second part of WBC nucleus segmentation was the combination of two single bands. The combinations were chosen based on the performance of a single band in Fig. 4(b). Result of segmentation in Fig. 4(b) is compared to the ground truth of nucleus segmentation in Fig. 4(a)(i). The selection of color space is based on the qualitative analysis which is the most equivalent to the ground truth. These combinations were done to vary the ways of segmenting the nucleus and also to find the best method to achieve the highest accuracy of WBC counting. Based on a single band in Fig. 4(b), it can be qualitatively depicted that the best nucleus segmentation outcome was derived from channel G of RGB, C of CMYK and S of HSV. The G of RGB gave the best contrast than R and B color space. Among H, S and V color space, S showed the best nucleus segmentation while for CMYK, C eliminated most of the WBC cytoplasm and left with only the WBC nucleus which was crucial to guarantee a convincing WBC counting result. Three unique combinations could be made by using these three channels which were GS, GC and SC. They were made by subtracting that individual component from another single color component.

2.2.2.1. S and G subtraction. For the first try, the combination of S of HSV and G of RGB was used. As stated before, S had by far given the best result of nucleus segmentation among all the color space components. It also eliminated the region of RBC and also the background. The white colored area was the WBC nucleus while G channel provided the best contrast and contained wide histogram distribution. G channel was almost the inverted version of S channel. Fig. 5 (a) shows the result of subtraction image of both color components. It can be seen that the cytoplasm area was removed and eliminated for all selected color spaces.

2.2.2.2. C and G subtraction. The next combination was C and G channel. G channel was subtracted from the C channel and the result is depicted in Fig. 5 (b). The subtraction result was almost the same as the S and G subtracted image. However, it had a slight difference in terms of the nucleus region. The nucleus region of C and G subtraction was bigger as the cytoplasm was not completely diminished. This happened because compared to S channel segmentation, C channel showed a bigger area of WBC nucleus than S channel image. Due to this reason, there were still some cytoplasm appearances in the subtracted image of C and G. However, the excessive cytoplasm region could be reduced by applying morphological filter prior counting the region.

2.2.2.3. S and C subtraction. Lastly, the combination of S and C channel was made. Between these three color channels, it can be seen that S and C color channels were better than G color channel in terms of the nucleus segmentation. Both color channels showed the minimum area of cytoplasm. For that reason, it could be said that this combination was better than SG and CG color space combination. This can be proven by the result of S and C subtraction as shown in Fig.5 (c). The subtracted image shows that the cytoplasm region and appearance was almost completely removed and eliminated. This was needed for the counting process as it would enhance the accuracy of WBC counting. A better nucleus segmentation result enables a higher WBC counting accuracy to be obtained.

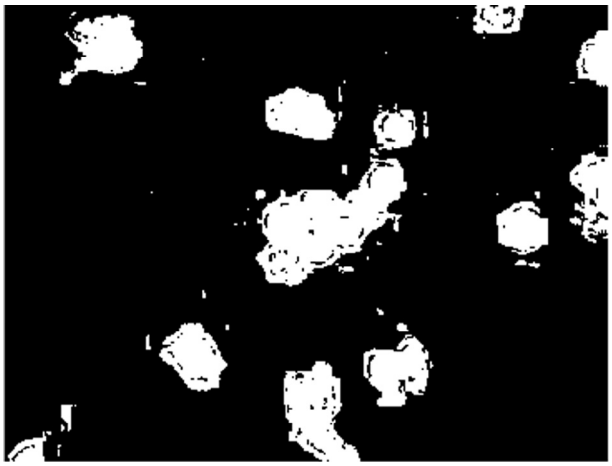

Fig. 6. H-Y subtracted image.

2.2.3. Cytoplasm detection based on combination of two single bands

The third part which is the last part of WBC segmentation was the cytoplasm segmentation. Cytoplasm is basically the whole WBC area including the outer part of the WBC region. In order to segment this area, combination of two single bands was applied. The color space analysis for cytoplasm detection purposes can be chosen qualitatively based on the ground truth of cytoplasm segmentation shown in Fig. 4(a) (ii). A single color band depicted in Fig. 4(b) could not be used for such purpose since there was no single segmentation results show the best cytoplasm segmentation. However, the segmented color spaces that showed the largest area of detected cytoplasm were H color component of HSV and Y color component of CMYK. Both color spaces were almost the inverted version of each other. In H color component, the whole WBC area was in white color region while in Y color component, the black color region was the whole WBC area. In this case, H and Y color components were combined and produced a segmentation method by subtracting the Y channel from the H channel. Fig. 6 demonstrates the subtracted result. It can be seen that after subtraction, the whole WBC area was detected which included the cytoplasm. RBCs and platelets were completely removed and eliminated from the subtracted image.

2.3. Morphological filter

After the segmentation process, it can be seen that the presence of noises and unwanted region still existed and these needed to be diminished. The noises were produced because of the portion of RBC or background that was not fully eliminated from the segmented image. For some color channels, the intensity level of RBC and background was almost similar to the WBC level and hence generated the unwanted noise. For such a case, a filtering method is crucial to be introduced to

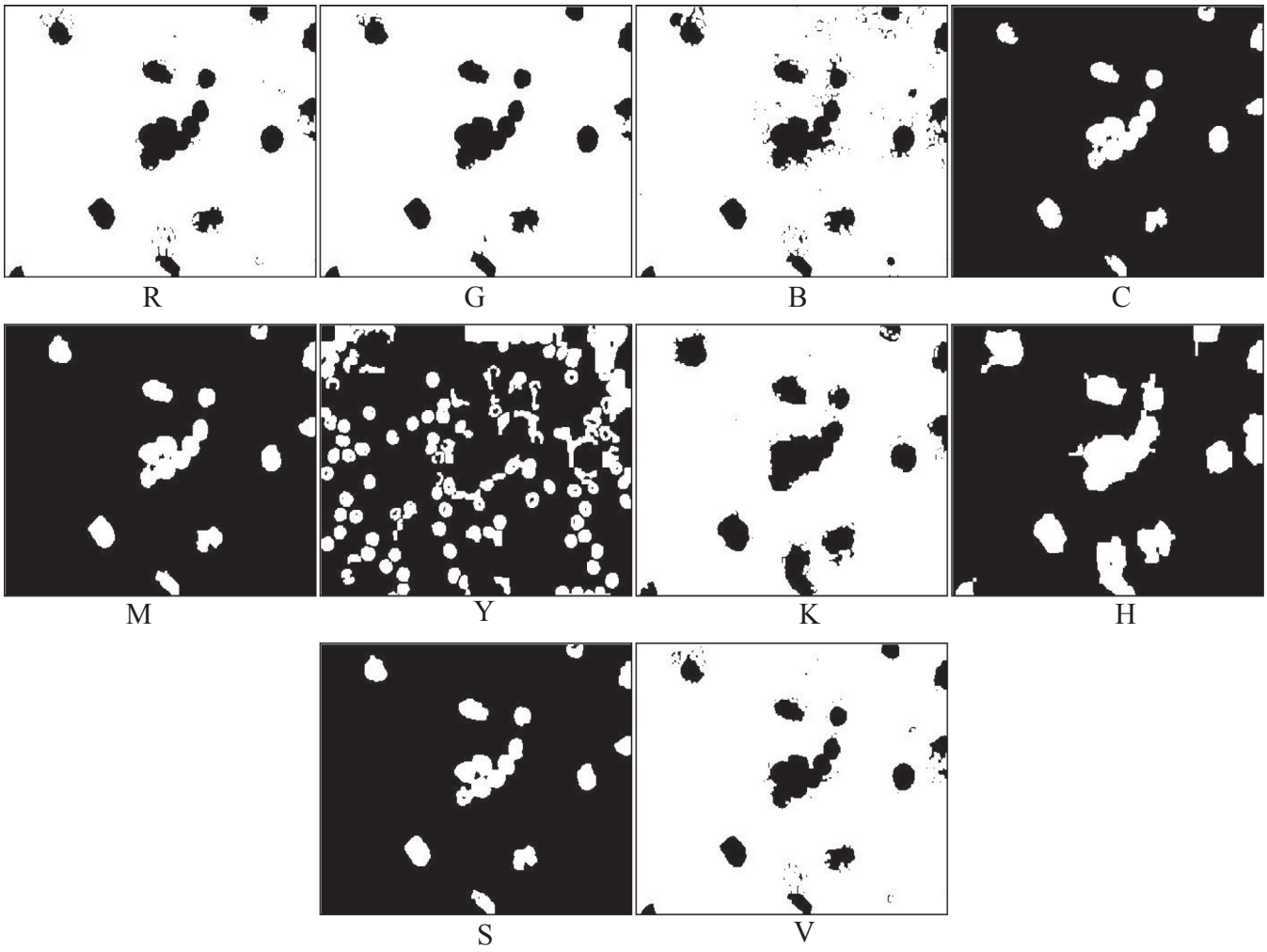

Fig. 7. Sample result of morphological filter for color channel of green (*G*), red (*R*), blue (*B*), hue (*H*), saturation (*S*), value(*V*), cyan (*C*), magenta (*M*), yellow (*Y*) and key (*K*).

diminish the noise. If the noise remains, it will affect the performance of the next process and thus affect the final result of the system. However, if the noise dimension is at its peak as a single WBC region area, the system will be unable to detect it as a noise. The noises and unwanted regions in the image need to be eliminated and removed completely from the image as it can disrupt the WBC counting performance in the system as the system might mistook the noise as another cell region while it is actually not.

In this paper, two types of methods were used to eliminate the noises in the segmented image which were morphological operation and Connected Component Labelling (CCL). Elimination of noises was undeniably a needed process as the noises would interfere the counting performance later.

These approaches were done on both the single band components and combination of two single bands. Firstly, the morphological operation that was used in the system was the erosion and dilation process. Morphological operation such as erosion and dilation is widely applied for filtration purposes. Erosion is meant to shrink and erode the objects in the image while dilation is used to add pixels and grow the objects in image. As for this project, both erosion and dilation were done once per image. First, the erosion was applied, which shrunk the cells and removed some small unwanted particles in the image. The application of dilation brought back the shrunken cells to its original structure and size. The removed particles would not appear again. However, there were still some unwanted cells left in the image even after morphological operation was applied to the image. To overcome this problem, Connected Component Labelling (CCL) was applied to the image which had set the pixels needed in the image and removed the unwanted ones. In this project, the standard pixels needed to be in the image was set which was more than 150 pixels. So, any objects that had less than 150 pixels were removed from the image. Connected Component Labelling (CCL) removed the unwanted pixels that had been set in the system without changing the other object's structure and shape. This process incredibly cleared out the small particles in the image. The final result after CCL process and after morphological filter was applied can be seen in Fig. 7 for single band color components while for combination of two single bands' morphological filter result is depicted in Fig. 8. The noises were removed and the image after morphological filter was better than the segmented image in terms of the nucleus identification. Fig. 9 shows the image of cytoplasm detection after morphological filter was applied to the subtracted image. Almost all the noises were eliminated.

2.4. WBC counting

This is the crucial part where the performance of segmentation and identification of WBC methods are compared. It can be classified as a challenging task due to the shape of WBC region which is not consistent. Other than that, some of the cells are clumped and overlapping with each other which make it harder to detect the individual cell. For the WBC counting purpose, it is important to consider the clumped cells to achieve a high accuracy outcome. Basically, such requirement can be satisfied by using several methods such as distance transform [19], edge detection [20] and Circle Hough Transform (CHT) [21]. CHT method is

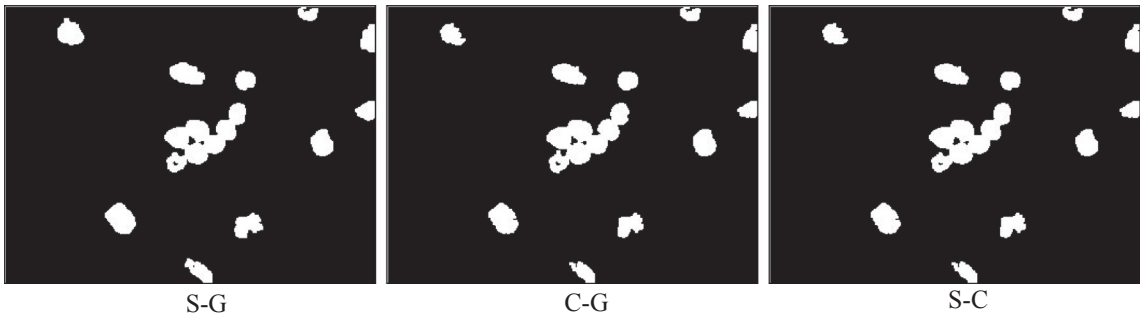

Fig. 8. Sample result of morphological filter for combination of two single bands S-G, C-G and S-C.

reported to be able to identify and count the overlapping cells with minimal complexity [22]. In this research, the segmentation performance was obtained by the accuracy of WBC region counting. Since various methods were elaborated for the segmentation process, the analysis was conducted to obtain the optimal segmentation method for the purpose of WBC counting. In the experiments, the image that was fed into the counting process was the final image after filtration by morphological filter and CCL filter. For the counting purposes, Circle Hough Transform (CHT) method was used. This method was meant to detect and count the circles in the image. This was done by finding the circle using the series range of radius from the minimum to the maximum. The radius range was set from 10 to 12 pixels. CHT considers the clumped and overlapping cells in its process and hence suit the purpose of this research as some images contained cells that were overlapping with each other and needed to be counted as well. CHT find the best intersection points based on an Eq. (2) below where a_i and b_i is defined as the center of circle. X_i and Y_i will give the parameter of the circle, θ is the angle through 360° and r_i is circle's radius.

$$Xi = ai + ricos\Theta \qquad Yi = bi + risin\Theta \tag{2}$$

The sample result of WBC counting is indicated in Fig. 10 and Fig. 11 respectively. In Fig. 10, WBC counting of single band color analysis is depicted while for the three combinations of S-G, C-G and S-C outcome is as shown in Fig. 11. It can be seen that for come color space component, CHT was able to detect the clumped and overlapping cells as well. The counting performance of all method was observed and the result and analysis will be discussed in the next section. Meanwhile, the result image of WBC cytoplasm counting, is as depicted in Fig. 12. In the image, it can be seen that some single cells were miscounted as a clumped cell because of its wide area of detected WBC region.

Fig. 9. Sample result of morphological filter for cytoplasm detection of H-Y color space.

2.5. WBC identification

In this section, an assessment of WBC detection accuracy is elaborated which involve a quantitative segmentation analysis. The performance is measured by comparing the segmented image with respect to a ground truth image. For ease of explanation, the assessment is divided into two categories which are nucleus area identification and cytoplasm area identification. Identification performance is conducted using four crucial parameters which are True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). TP is a WBC region that is detected as WBC, TN is a non-WBC area that is detected as non-WBC, FP is a non-WBC region detected as WBC and FN is a WBC area that is detected as non-WBC by the system. From these values, accuracy, specificity and sensitivity of WBC detection is calculated. The accuracy of a test is its ability to differentiate the WBC from non-WBC portion correctly. Specificity defines the proportion of TN that is correctly identified by the system while sensitivity defines the proportion of TP that is correctly identified by the system. Basically, sensitivity shows how accurate the system identifies WBC and specificity shows how accurate the system identifies the non-WBC region. Accuracy, specificity and sensitivity are calculated using Eqs. (3), (4) and (5) respectively.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{3}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{4}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{5}$$

3. Result and analysis

3.1. Graphic User Interface (GUI)

The system was constructed using Matlab software that includes a Graphic User Interface (GUI) to make it user friendly as shown in Fig. 13(a). It can be seen in the GUI that the interface contains three main figures, two push buttons, one list box and one button group with three choices. The interface operates by initially choosing the appropriate color space which are S, S-C or H-Y color space. Then, the user needs to select the desired image database directory which will list all images name in the list box. Following that, the user can choose the intended image that need to be processed. Apart from that, the system also provides a rapid processing feature which will do the batch processing. To accomplish this, the user just needs to insert all images that need to be processed rapidly in the selected folder and the final image is stored in a new folder named 'Final Image' as shown in Fig. Fig. 13(b). For the GUI arrangement, the input image with its histogram is displayed on the left side while the output image with overlaid counting result is shown on the right side. In order to achieve an accurate and

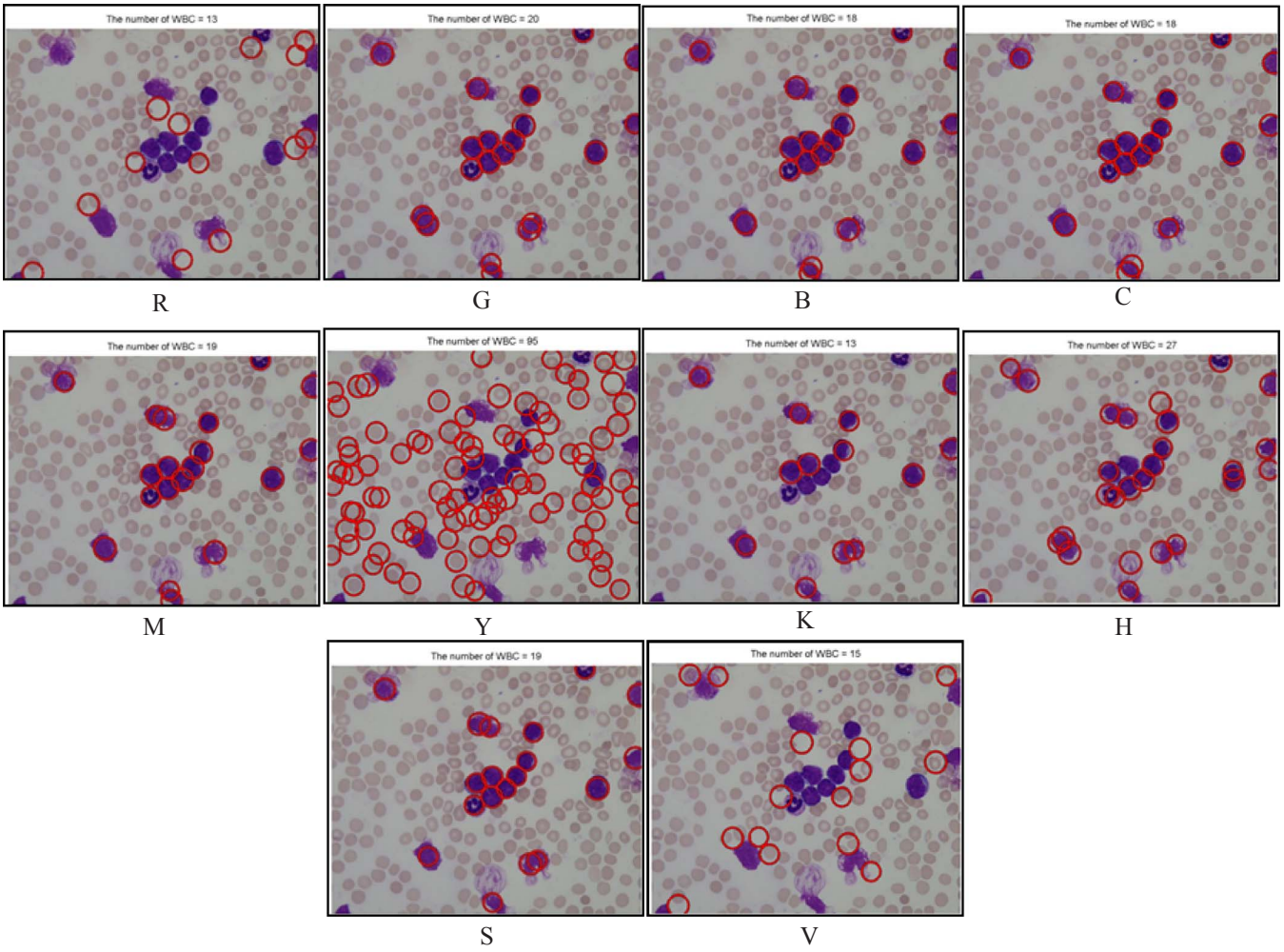

Fig. 10. WBC counting for color channel of green (*G*), red (*R*), blue (*B*), hue (*H*), saturation (*S*), value (*V*), cyan (*C*), magenta (*M*), yellow (*Y*) and key (*K*).

consistent result, the input image needs to be in the same resolution of 2592 × 1944 and same magnification factor of 400. It can be extended to various resolutions later on by adding an appropriate constant multiplier to control the image ratio.

3.2. WBC counting accuracy

In this section, performance of the developed system was assessed using sample images obtained from the IDB database [23]. Basically, the assessment is carried out based on the WBC region counting accuracy for 30 images from the database as they are in the same magnification factor. As has been mentioned in previous section, the segmentation process plays the biggest role in contributing to the higher counting accuracy. Various methods that have been used for the WBC segmentation have been elaborated in which their counting accuracy is compared in this section. The WBC counting performance was accessed based on the ground truth data in which information about how well the methods work was depicted. The result of WBC counting for single bands and combination of two single bands comparison are summarized in Table 1 and Table 2 respectively.

Table 1 summarizes the WBC counting result for single band components which were R, G, B, C, M, Y, K, H, S and V. From the table, it

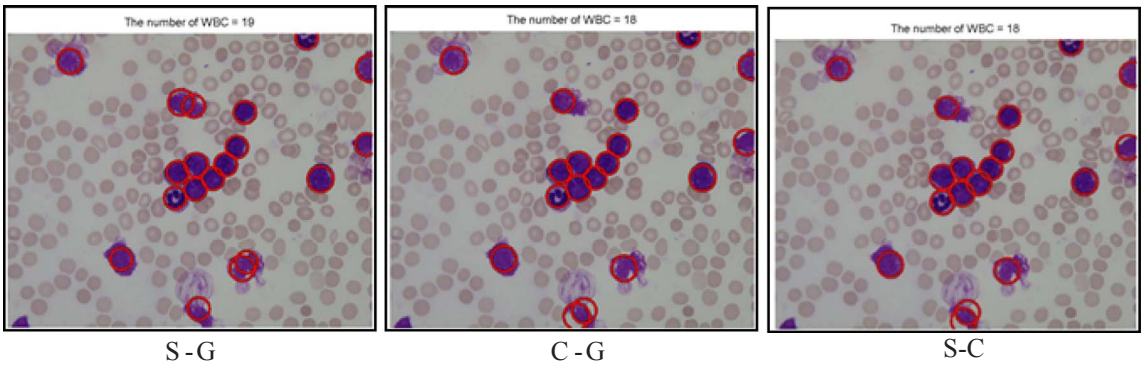

Fig. 11. Sample result of WBC counting for combination of two single bands S-G, C-G and S-C.

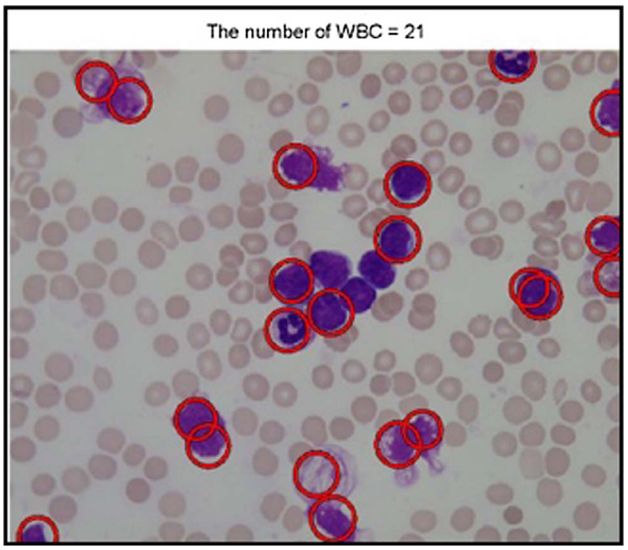

Fig. 12. Sample result of WBC counting for cytoplasm detection (H-Y).

can be distinctly seen that some of the color components produced a good result and could work on its own. Two of the color components which were C and S color component achieved the WBC counting accuracy of more than 90%. Among all the color components, S color channel showed the highest accuracy as highlighted in the table. Based on the result, it can be concluded that S channel was able to segment out the WBC nucleus area better than any other color components. The lowest WBC counting accuracy for single band component was the Y color component which produced 14% accuracy. This was because of the segmentation of Y component on its own was unable to prune out WBC region correctly as shown in Fig. 4(b). As a result, the counting process using CHT did not yield a good result as be seen in Fig. 10.

Table 2 shows the result of WBC counting for combination of color spaces which were S-G, C-G and S-C. From the table, it can be seen that the subtraction method of S-C gave the highest counting accuracy which was 96.56%. This was because S and C themselves provided an accurate WBC counting on its own as depicted in Table 1 and hence the combination was also expected to produce more or less the same result. From the WBC localization and detection perspectives, S and C color space also identified WBC nucleus region better than any other individual color space component which apparently contributed to the higher WBC counting. For all three sets of color combinations, the average accuracy is higher than 90% and hence can be considered as exemplary method because of the well combination of individual color space component. In Table 3, the counting accuracy based on cytoplasm detection which uses combination of H and Y color components is 40.72%. The performance of WBC counting based on cytoplasm indicator did not yield a good result due to the cytoplasm had given a wide area of WBC region. As a result, it has confused the CHT and some individual cells are mistakenly confused and marked as a clump region.

All in all, for the counting assessment generally, CHT showed a good performance of detecting and counting the WBC region including the clumped cells. From our observation it failed to accomplish the task due to several reasons as follow:

- RBC region was mistakenly taken as a foreground because of its color intensity.
- CHT might consider the stains on the blood smeared image as WBC because of the stain color which was nearly similar to the WBC region.
- When the cells overlapped too close with each other, CHT was unable to locate the individual cell accurately.

3.3. Effect of color space correction to the WBC counting

Previously the motivation of using color space correction for the WBC counting purpose was elaborated. In this section, the feasibility of such method is assessed based on two criteria which were: (1) Effect of the WBC counting performance with and without using the color space correction. (2) Effect of color correction parameter estimation using template images, and combination of all tested images. For first assessment, the aim is to get a clear view whether the color correction

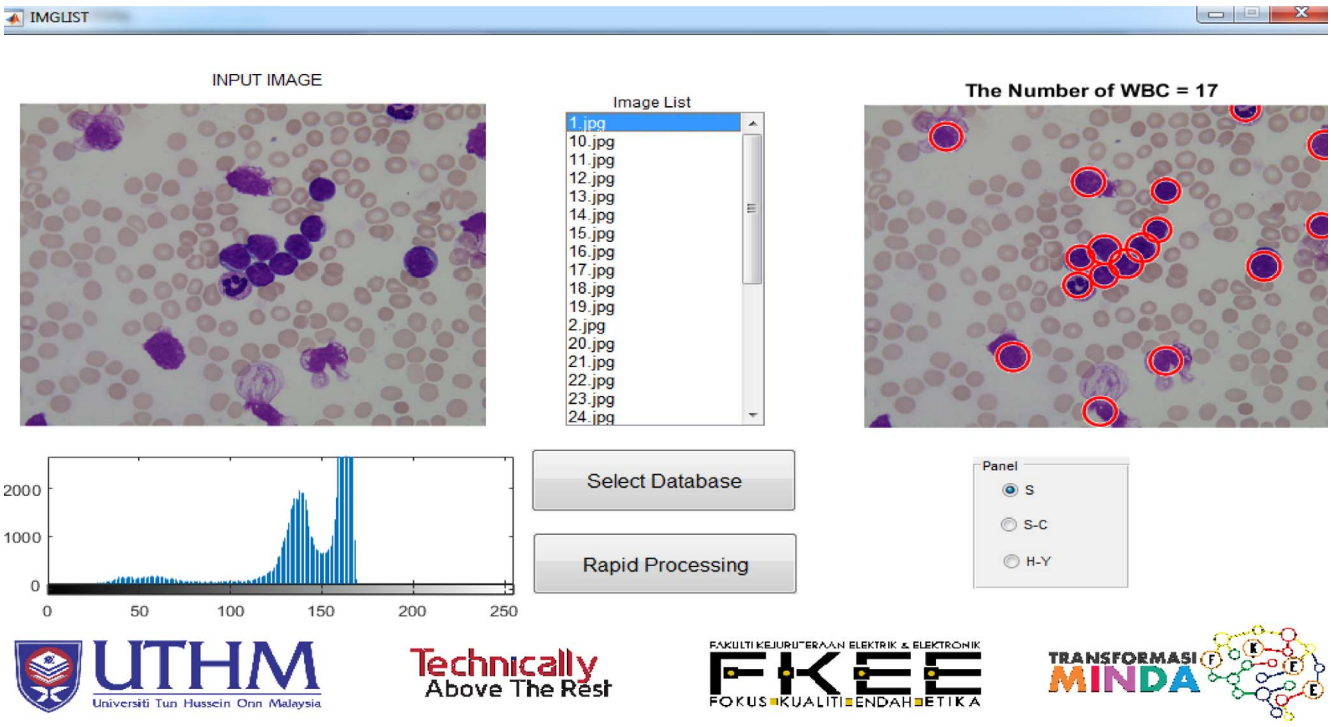

Fig. 13a. Graphic User Interface (GUI) for automated WBC counting system.

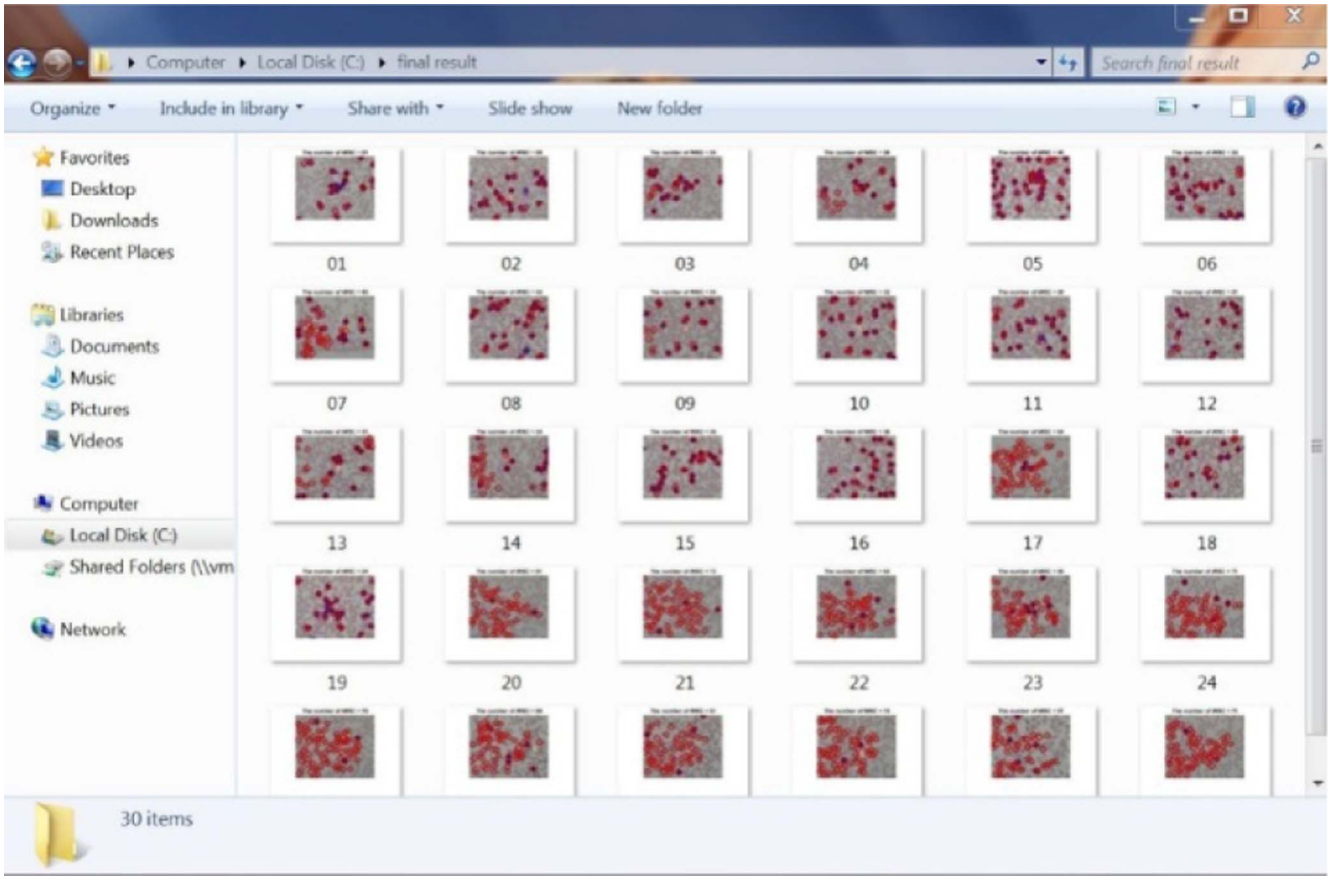

Fig. 13b. Final image folder.

Table 1
Accuracy of WBC counting for color channel of RGB, CMYK and HSV.

Color Space Component	Average Accuracy (%)
R	63.95
G	59.39
B	25.34
C	95.13
M	74.34
Y	14.28
K	45.18
H	37.47
S	96.92
V	60.92

Table 2
Accuracy of WBC counting for combination of two single bands S-G, C-G and S-C.

Color Space Combination	Average Accuracy (%)
S-G	96.43
C-G	94.33
S-C	96.56

Table 3
Accuracy of WBC counting for cytoplasm detection.

Color Space Combination	Average Accuracy (%)
H-Y	40.72

Table 4
Color space correction analysis.

Color Space Analysis	With Color Space Correction	Without Color Space Correction
C	95.13%	94.12%
S	96.92%	93.55%
S-C	96.56%	93.63%
S-G	96.43%	94.10%
H-Y	40.72%	17.55%

method is necessary for the system. A comparison is made using two highest accuracy result obtained from Table 1 and Table 2, and result in Table 3. The finding is as depicted in Table 4 which indicates that for each color space analysis, the counting accuracy slightly deteriorated with an average of 2.41% for four cases which were C, S, S-C and S-G. For the case of H-Y, the performance significantly reduced from 40.72% to 17.55% without correction. Hence, it can be deduced that it is important to include color correction method in the WBC counting framework to obtain a high WBC counting accuracy.

Initially the Color Space Correction (CSC) parameters (mean and standard deviation of L∗a∗b color space) were generated from an average of 4 template images as shown in Fig. 2. These parameters were utilized to map any incoming image to a standard color representation. In the second assessment, an effect of changing the CSC values were investigated based on using the average of L∗a∗b color information from the entire dataset of images. For the evaluation purpose, the performance of the WBC counting accuracy of S color space component was analyzed since it provided the highest accuracy compared to other components. Fig. 14 shows the qualitative comparison of the original image, the color corrected image by

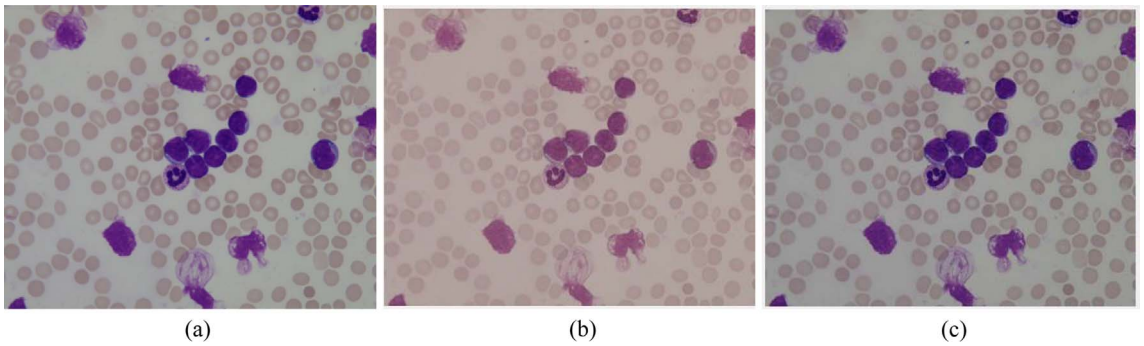

Fig. 14. (a) Original image. Color corrected image using (b) 4 template images (c) whole dataset.

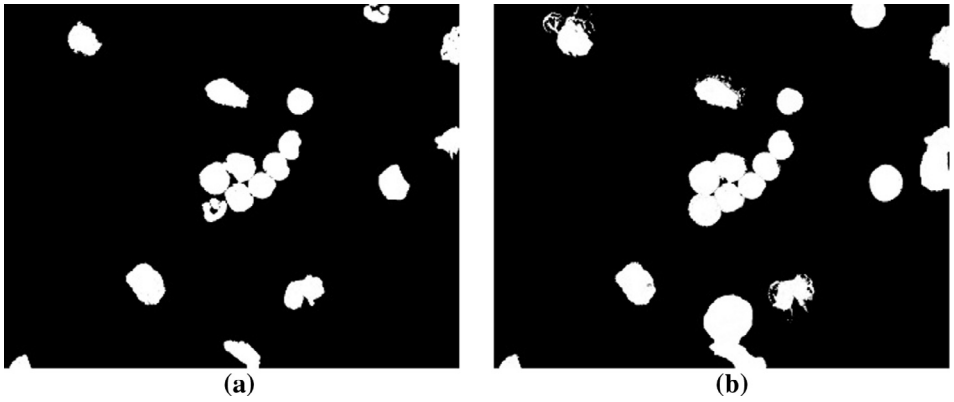

Fig. 15. Sample of ground truth data for: (a) Nucleus area identification (b) Cytoplasm area identification.

using 4 template images, and the whole dataset. It can be seen that by using 4 template images, the color of the WBC region became more visible and the RBC region turned faded. This representation is very useful to prune out WBC region from others in the blood smeared image. As for the WBC counting assessment, quantitatively the performance reduced from 96.92% to 86.49% by using entire dataset for the CSC parameter estimation. Such finding shows a huge reduction in counting accuracy and hence indicated that 4 template images 2 were more suitable to be used as the CSC parameter estimation.

3.4. WBC detection accuracy

In this section, the performance of image segmentation process was assessed based on its ability to identify and detect WBC region correctly. S color component from nucleus based WBC counting was selected since it provided the highest accuracy and H-Y color component from the cytoplasm based in this analysis. Three measurements which were accuracy, specificity and sensitivity were calculated for this assessment. This was done by comparing the result of the threshold image with respect to the ground truth data. The ground truth of nucleus and cytoplasm area had a slightly difference shape and area as depicted in Fig. 15. It can be seen in the figure below that the shape and area of cytoplasm was wider than the nucleus area. The ground truth data was constructed by manual editing of the original image using professional editing software which only included the part that was important to be identified.

Figs. 16 and 17 show the trend of accuracy, specificity and sensitivity of 30 images for nucleus and cytoplasm based area identification respectively. It shows that the performance of nucleus area detection was more consistent compared to the cytoplasm area detection. Another

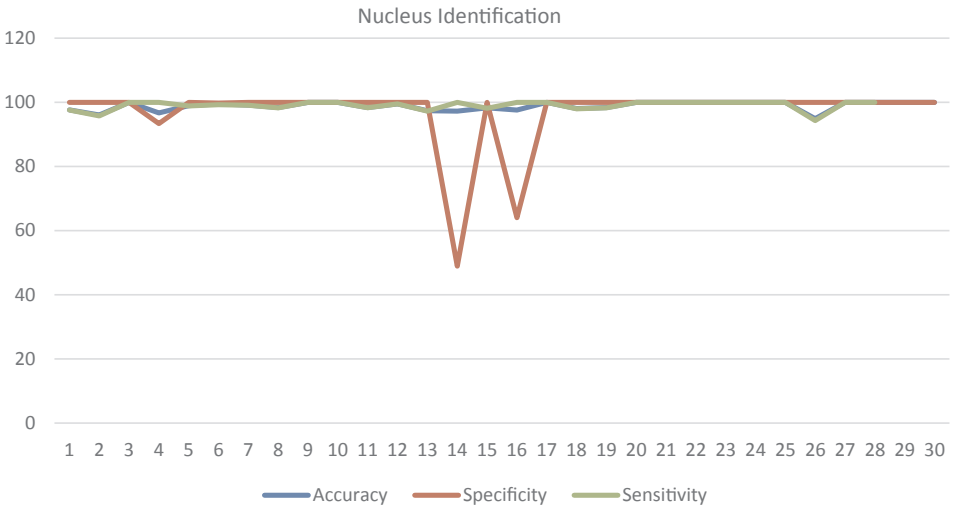

Fig. 16. Accuracy, Specificity and Sensitivity trend for nucleus identification.

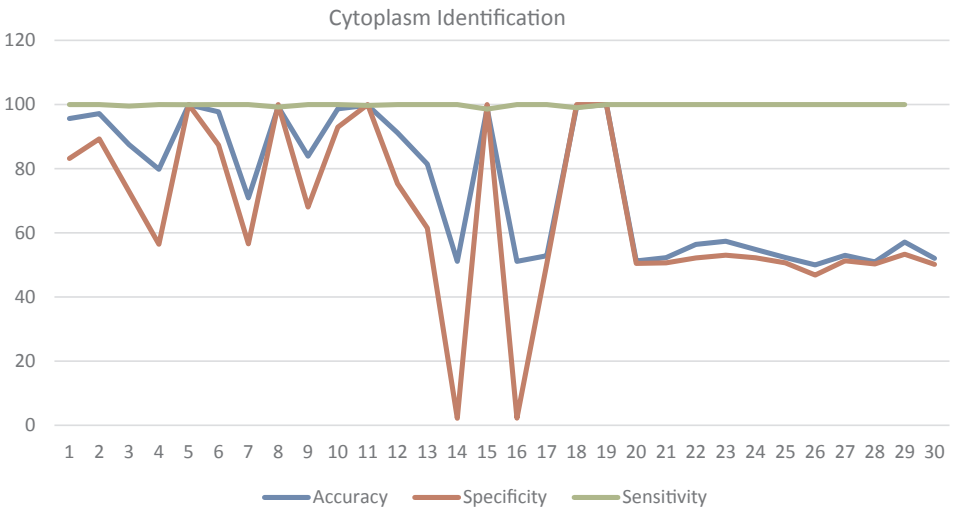

Fig. 17. Accuracy, Specificity and Sensitivity trend for cytoplasm identification.

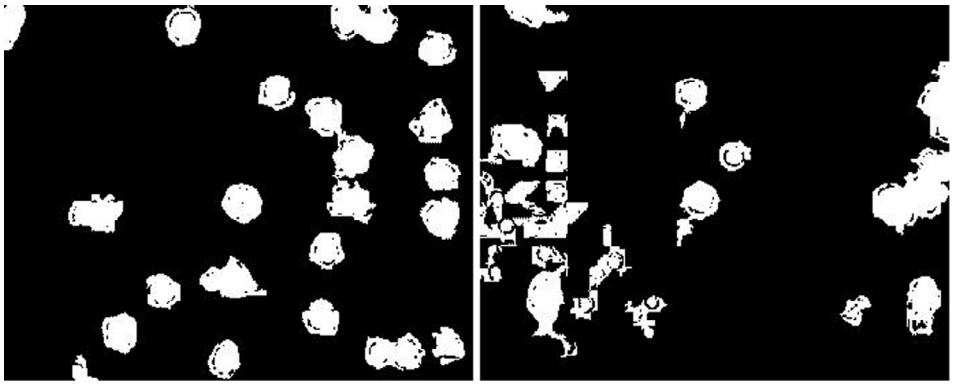

Fig. 18. Sample of high intensity level of noise image.

interesting point is that, for both types of area identification, images 14 and 16 produced low specificity and sensitivity values. Such situation was happening because of the high intensity level of noises resided in the images and caused the WBC region not to be segmented correctly as shown in Fig. 18. Even though the specificity value was low for these images, apparently the WBC counting still worked well and abled to produce high accuracy. Such finding indicates that, in a situation where the segmentation fails, the WBC can still be counted accurately with proper tuning of the CHT parameters.

The average values of accuracy, specificity and sensitivity of nucleus and cytoplasm based area detection was calculated by averaging the parameters of TP, TN, FP and FN of the 30 images. The summary of average accuracy, specificity and sensitivity result of WBC segmentation is shown in Table 5. For nucleus based area detection, the averages for all three parameters were more than 95% while for cytoplasm based area detection, the results were not as good as nucleus identification percentages. Judging from this finding, it can be generally concluded that for the purpose of WBC counting, a high accuracy can be achieved by using the method based on nucleus based segmentation which is superior compare to the cytoplasm based segmentation.

Table 5
Accuracy of WBC detection.

WBC Identification	Accuracy (%)	Specificity (%)	Sensitivity (%)
Nucleus (S)	98.87	96.87	99.10
Cytoplasm (H-Y)	74.12	65.32	99.87

4. Conclusion and future work

In this paper, an empirical framework for automatic detection and count for the number of WBC in blood smeared image is proposed. This automatic system is developed under computer vision system for image processing purposes. Several methods were compared with each other to find the best method to identify and count WBC region in the image. However, the main blocks of each method went through similar processes which consisted of color space correction, WBC segmentation, post processing and WBC counting. Firstly, the original blood smeared image's color intensity was corrected using L∗a∗b∗ color space correction based on the mean and standard deviation value of l, a and b. The mean values that were set for l, a and b were 81.9863, 4.5579 and 1.2176 while the standard deviation values were 4.7984, 4.7248 and 3.0947 respectively. Next, the corrected image was segmented using Otsu Thresholding method. In this process, various segmentations were done using color space analysis of RGB, CMYK and HSV. Three combinations of color space analysis were made which were S-G, C-G and S-C and WBC region segmented using the Otsu method. It was shown that for a single band color component, S provided the highest accuracy of WBC counting which was 96.92%. Meanwhile in Table 2, the combination of S-C color analysis produced the best result of 96.56%. As for the accuracy for WBC counting, cytoplasm detection should not be included as the result showed a very low performance. However, among all the color space analysis methods including the combinations, S achieved the highest accuracy. The overall system performance of S color space method was the best among all the method tested. Lastly, it can be concluded that with the presence of color space correction process, the WBC accuracy can be increased. Without color space

correction, the WBC accuracy will be slightly lower. This has been proven by comparing the results of several methods. Based on the comparison, it has been found that the accuracy of all methods is higher when the color space correction process is carried out.

In future, the scope in the system can be improved by considering the RBC count and develop a Complete Blood Count (CBC) automated system. Other than that, the system is expected to have a classification ability in which classifications of different types of WBC in blood smeared image can be conducted. The system is also expected to be accessible through mobile devices such as smart phone and tablet.

References

[1] Razali Tomari, Wan Nurshazwani Wan Zakaria, Rafidah Ngadengon, An Empirical Framework For Automatic Red Blood Cell Morphology Identification and Counting, ARPN J. Eng. Appl. Sci., Batu Pahat, 2013.

[2] Congcong Zhang, Xiaoyan Xiao, Xiaomei Li, Wu. Ying-Jie Chen, Jun Chang Zhen, Chengyun Zheng, Zhi Liu, White Blood Cell Segmentation by Color-Space-Based K-Mean Clustering, Sensors 14 (1424–8220) (2014) 16128–16147.

[3] Salim Arslan, Emel Ozyurek, Cingdem Gunduz-Demir, A Color and Shape Based Algorithm for Segmentation of White Blood Cells in Peripheral Blood and Bone Marrow Images, Int. Soc. Advancement Cytometry, vol. Part A 85A (2014) 480–490.

[4] Jun Duan, Le Yu, A WBC Segmentation Method Based on HSI Color Space, in: Proceedings of IEEE, Baotou, 2011.

[5] B. Venkatalakshmi, K. Thilagavathi, Automatic red blood cell counting using hough transform, in: IEEE Conference on Information and Communication Technologies, Chennai, 2013.

[6] J.B. Prof Neman, V.A. Chakkarwar, P.B. Lahoti, White blood cell segmentation and counting using global threshold, Int. J. Emerging Technol. Adv. Eng. 3 (6) (2013) 639–643.

[7] Sedat Nazlibilek, Deniz Karacor, Tuncay Ercan, Murat Husnu Sazli, Osman Kalender, Yavuz Ege, Automatic segmentation, counting, size determination and classification of white blood cells, vol. 55, no. 0263-2241 (2014) pp. 58–65.

[8] Erik Cuevas, Margarita Diaz, Miguel Manzanares, Daniel Zaldivar, Marco Perez-Cisneros, An Improved Computer Vision Method for White Blood Cells Detection, Comput. Math. Meth. Med., (2013) p. 14.

[9] Sonal Kothari, Qaiser Chaudry, May D Wang, "Automated Cell Counting and Cluster Segmentation Using Concavity Detection and Ellipse Fitting Techniques," in IEEE International Conference on Symposium on Biomedical Imaging: From Nano to Macro, Massachusetts, 2009.

[10] Razali Tomari, Wan Nurshazwani Wan Zakaria, Rafidah Ngadengon, Mohd Helmy Abd Wahab, Red blood cell counting analysis by considering an overlapping constraint, ARPN J. Eng. Appl. Sci. 10 (3) (2015) 1413–1420.

[11] Yazan M. Alomari, Siti Norul Huda Sheikh Abdullah, Raja Zaharatul Azma, Khairuddin Omar, Automatic Detection and Quantification of WBCs and RBCs Using Iterative Structured Circle Detection Algorithm, Comput. Math. Meth. Med., vol. 2014 (2014) p. 17. 10.1155/2014/979302.

[12] Subrajeet Mohapatra, Diptri Patra, Automated Cell Nucleus Segmentation and Acute Leukemia Detection in Blood Microscopic Images, Int. Conf. Syst. Med. Biol., Kharagpur, 2010.

[13] Lorenzo Putzu, Cecilia Di Ruberto, "White Blood Cells Identification and Counting from Microscopic Blood Image," International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering, pp. 20-27, 2013.

[14] Leyza Baldo Dorini, Rodrigo Minetto, Neucimar Jeronimo Leite, Semiautomatic white blood cell segmentation based on multiscale analysis, IEEE J. Biomed. Health Inform. 17 (1) (2013) 250–256.

[15] Lalit B. Damahe, R.K. Krishna, N.J. Janwe, V. Thakur Nileshsingh, Segmentation based approach to detect parasites and RBCs in blood cell images, Int. J. Comp. Sci. Appl. 4 (2) (2011) 71–81.

[16] Deepali Ghate, N. Chaya Jadhav, Usha Rani, Automatic detection of malaria parasite from blood images, Int. J. Adv. Comp. Technol. 4 (1) (2015) 129–132.

[17] Erik Reinhard, Michael Ashikihmin, Bruce Gooch, Peter Shirley, Color transfer between images, IEEE Comp. Graph. Appl. 21 (5) (2001) 34–41.

[18] Yuncong Feng, Haiying Zhao, Xiongfei Li, Xiaoli Zhang, Hongpeng Li, A multi-scale 3D Otsu thresholding algorithm for medical image segmentation, Digit. Signal Process. 60 (2017) 186–199.

[19] Haider Adnan Khan, Golam Morshed Maruf, Counting Clustered Cells Using Distance Mapping, in: International Conference on Informatics, Electronics and Vision (ICIEV), Dhaka, 2013.

[20] Gohil Asmitaba, Dhaval Pipalia, Design an algorithm to detect and count small size object using digital image processing, Int. J. Adv. Res. Electr., Electron Instrument. Eng. 5 (5) (2016) 3807–3812.

[21] D.H. Ballard, Generalizing The Hough Transform to Detect Arbitrary Shapes, in: Readings in Computer Vision: Issues, Problems, Principles, and Paradigms, Morgan Kaufmann Publishers Inc, San Francisco, 1987, pp. 714–725.

[22] Mausumi Maitra, Rahul Kumar Gupta, Manali Mukherjee, Detection and counting of red blood cells in blood cell images using hough transform, Int. J. Comp. Appl. 53 (16) (2012) 18–22.

[23] R. Donida Labati, V. Piuri, F. Scotti, ALL-IDB: the acute lymphoblastic leukemia image database for image processing, in: Proc. of the 2011 IEEE Int. Conf. on Image Processing (ICIP 2011), Brussels, Belgium, pp. 2045-2048, September 11-14, 2011.