Jurnal Teknologi

Full Paper

AUTOMATED WHITE BLOOD CELLS COUNTING SYSTEM FOR ACUTE LEUKEMIA BASED ON BLOOD IMAGES

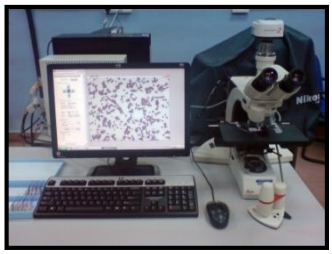
N. H. Harun[a*], M. Y. Mashor[b], H. N. Lim[b], R. Hassan[c]

[a]School of Computing, College of Arts and Science Universiti Utara Malaysia, Kedah, Malaysia
[b]Electronic and Biomedical Intelligent System (EBItS) Research Group, School of Mechatronics Engineering, Universiti Malaysia Perlis, 02600 Pauh, Perlis, Malaysia
[c]Department of Hematology, Universiti Sains Malaysia, KubangKerian, Kelantan, Malaysia

Article history
Received
22 July 2015
Received in revised form
20 December 2015
Accepted
27 January 2016

*Corresponding author
hazlyna@uum.edu.my

Graphical abstract



Abstract

Leukemia is a blood cancer disease that contributes to the increment of mortality rate in Malaysia each year. Acute leukemia can be divided into two types which are acute myelogenous leukemia (AML) and acute lymphocytic leukemia (ALL). The production and development of acute leukemia cells happen rapidly and uncontrollable. Identification of technique for acute leukemia could be done fast and effective in the early stage, the proper treatment could be delivered. Hematologists or technologists will screen for leukemia using microscopic blood image. Screening and diagnosis with computer aided system becomes more popular choice amongst medical image processing researchers as the existing equipments which use other method than blood image analysis are too expensive to own. This paper presents automated counting procedures, including a series of image processing techniques such as segmentation, noise elimination and extraction to count the region of interest, in this case the white blood cells (WBC). The performance of the system is tested by comparing the results between manual counting and the total cells that the system detects. The proposed system is able to yield an average of 99.09% based on sensitivity.

Keywords: Color thresholding, median filter, seed region growing, acute leukemia, HSI color space

1.0 INTRODUCTION

Computer aided screening system has been a popular choice for researchers in image processing technology field in an effort to improve and enhance the medical diagnosis efficiently. Counting, detection and classification of white blood cells (WBC) based on peripheral blood or bone marrow samples are able to achieve this objective. This paper presents a computer aided screening system attached with an image capturing process using 10*X* magnification lens, a series of image processing techniques such as segmentation, noise elimination and extraction to count the region of interest, in this case the WBC. The focus of this study is to analyze the medical images morphologically as the technique is the cheapest compared to other diagnosis method such as cytogenetic and polymerase chain reaction (PCR).

During the screening process, blood images will be classified by determining the total number of WBC in a normal or abnormal blood samples. The microscopic blood image of a leukemic patient shows a high and significant number of acute leukemia cells as compared to one from a healthy person. Clinically, WBC detection is more than red blood cells in microscopic images as most blood diseases have a

close connection to WBC such as leukemia, lymphoma and myelody splastic syndrome [1]. The nucleated cells in human blood can only be WBCs [2]. Duan& Yu stated in their study that if the nucleus is detected, indirectly it detects the WBC [3]. In this case, if the nucleus is detected in abnormal blood images, it also detects the acute leukemia cells. There are two types of acute leukemia, which are acute myeloid leukemia (AML) and acute lymphoid leukemia (ALL). These two types of acute leukemia are considered in this study.

The significant difference of total nucleus in the normal and abnormal blood images becomes the criteria in developing computer aided acute leukemia screening system. Rezatofighi *et al.* and Nazlibilek *et al.* proposed the automatic recognition of WBCs [4, 5]. Their approach was solely on the five classes of white blood cells which are basophil, eosinophil, neutrophil, lymphocyte and monocyte. Scotti presented ALL cells identification in peripheral blood microscopic images [6]. In his study, the leukocytes were individuated in the blood image from the other blood cells composed by adaptive filtering and segmentation algorithms. It is useful if the image contains a low number of cells yet in most leukemia cases, the peripheral blood image has a high volume of cells. Sadeghian *et al.* proposed a segmentation framework that segment the WBC to its two dominant elements: nucleus and cytoplasm [7]. The nucleus segmentation part is based on morphological analysis and the cytoplasm segmentation is based on pixel-intensity thresholding. The research was focused only on the segmentation scheme and the sample used was ALL subtype L2 microscopic images. The work of Theera-Umpon *et al.* and Lim *et al.* were observed briefly as in the case of acute leukemia images studies [8, 9]. The application of their studies was on bone marrow samples, which has complex and hypercellular contents. As bone marrow aspiration is a tedious and stressful procedure for the patients, peripheral blood samples are preferable. Most of the studies conducted focus on the identification of blood cells, yet they mostly acquitted the images manually as well as manual processing of the image. Nevertheless, the previous works include case study of diseases but does not involve normal samples. This research aims in developing a counting system for normal cases and acute leukemia cases as the procedure is a part of the standard diagnosis steps.

In this paper, images with 10*X* magnification is used to count the number of WBC in normal images and blasts in AML and ALL images. Figure 1 shows the apparatus that is facilitated in this research: the motorized microscope *Leica DMLA 12000* controlled by the user-friendly screening system software, camera Infinity 2 and computer. Furthermore, the framework of acute leukemia counting system is illustrated in Figure 2. The motorized microscope was controlled using the graphical user interface (GUI) in our system to automatically capture the images and focusing on the morphology zone of the slide. The red box stating automated classification system using 40*X* magnification will be the future work of this system. The implementation of artificial intelligent (AI) will be emphasized in this work for classification purposes.

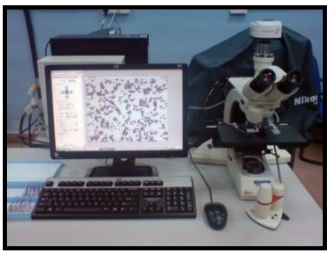


Figure 1 Camera Infinity 2 mounted on microscope Leica DMLA 1200 and attach to computer

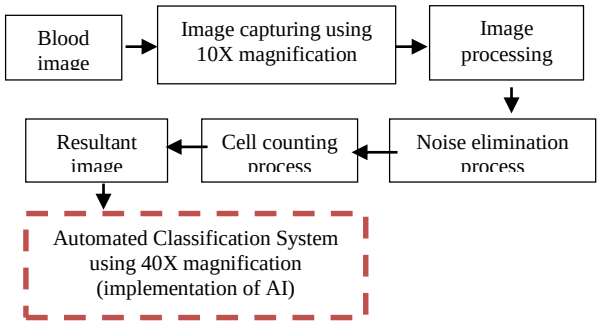


Figure 2 Development of screening system of acute leukemia

2.0 MATERIALS AND METHODS

This study proposes the effective counting procedures in order to achieve the objective significance. As shown in the Figure 3, there are three necessary steps to perform the acute leukemia counting process: Image capturing under 10X magnification using Infinity 2 camera mounted on motorized microscope Leica DMLA 1200, S-component based on HSI (hue, saturation, intensity) color space, color thresholding, 5x5 median filter and seed region growing (SRG) algorithm which will be further described in the following sections. The resulting images from this counting process can be used in the future work such as classification in determining the type of acute leukemia.

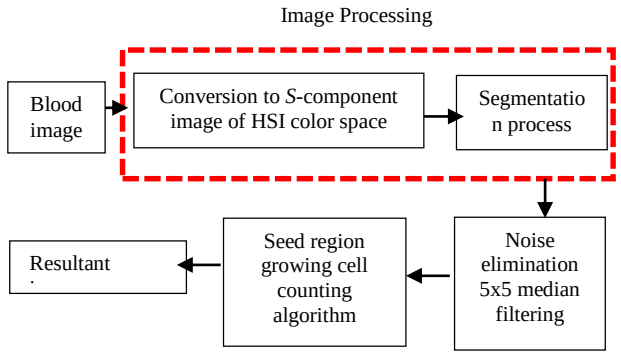


Figure 3 Segmentation procedures for acute leukemia image

2.1 Automated Image Acquisition under 10*X* Magnification

Peripheral blood samples from normal and acute leukemia cases were provided from the Hematology Department of Hospital Universiti Sains Malaysia (HUSM), Kubang Kerian, Kelantan. 42 blood slide samples were obtained which consisted of 25 normal blood samples and 17 acute leukemia blood samples. The images used in the screening process were captured using *Infinity 2* camera mounted on *Leica DMLA 1200* motorized microscope under 10*X* magnification with 800 x 600 resolution and were saved in bitmap format (*.bmp). Validation and confirmation on the total number of cells and its group in both the normal and abnormal cases were done by the hematologists.

2.2 Color Thresholding Technique Using *S*-component based on HSI Color Model

The color image carries a lot of information. However, most researchers prefer to convert the images into gray scale or other color models before continuing with the image segmentation process as described by Gonzalez & Woods, RGB, CMY and other similar color models [10]. The HSI color model is widely implemented in the studies of WBC segmentation such as in Dhandra *et al.* and Kang *et al.* [11, 12]. Regarding observation of blood images in this study, only the *S*-component is considered as it gives the best contrast in terms of saturation value to the cells in question. RBCs, artifacts and background regions are less saturated in *S*-component image, hence decreasing its visibility. The saturation formula can be applied to acute leukemia images by using (1):

$$Saturation = 1 - \left(\frac{3}{R+G+B}\right) \min(R,G,B) \quad (1)$$

In developing the algorithm for color thresholding, selecting the suitable threshold value is paramount so that it can effectively eliminate the unwanted regions and retain the cells in interest. The steps of determining the threshold value from the blood images are as follows:

i. Converting the original RGB image to HSI color model.
ii. Extracting the *S*-component value from the HSI color space image using equation (1).
iii. Generating *S*-plot to determine the threshold value for the image.
iv. Selecting the threshold value TH = μ from the *S*-plot.

The rules for determining the threshold value for normal and abnormal images were based on equation (2) and (3):

$$O(x,y) = \begin{cases} I(x,y), if\ TH \geq \mu \\ 255, \quad else \end{cases} \quad (2)$$

$$\mu = Display_{Saturation} / 255 \quad (3)$$

Parameter *O*(x,y) is the pixel value of the *S*-component output image while parameter *I*(x,y) is the pixel value of the input image in its original color. The output pixel value *O*(x,y) is prefixed at 255 (white) or the original color of the image is retained. From equation (2), the threshold value for the illustration purpose is set at Display Saturation = 90 and it is equivalent to *TH* = μ= 90/255 = 0.35. The value of μ, which is the average saturation value for HSI color model, is between 0 and 1. The selected μ value ensures the pixel value of the WBC or leukemia cells remains the same as its original image and converts all the unrelated value to white color.

2.3 Noise Elimination with 5x5 Median Filter

Median filter is suggested in this paper due to the effectiveness of salt and pepper noise cancellation and preserving useful detail in the image [13]. In this paper, 5 x 5 neighborhoodsis selected as it provides better end result with low time consumption. Although the larger neighborhood value yields better results, the time consumed of pixels full sorting in sliding windows is proportionally long[14].

2.4 Seeded Region Growing Area Extraction Algorithm

This algorithm is proposed because of its ability to eliminate larger noise pixels which are beyond the limits in the previous filtering step. More importantly, the application is to extract the area of the WBC cells and then used to determine the total number of the WBC in the image. Referring to the illustration example in Figure 4, the region of interest (ROI) is counted as the total number of pixels in k^{th} area as in equation (4).

size[*k*] = total pixels in ROI[*k*] (40)

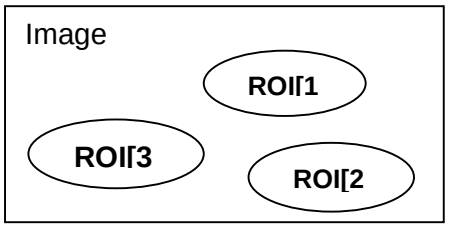


Figure 4 Example of region of interest (ROI) in an image

After the threshold value is determined and noise is eliminated, the eight seed-based region growing algorithm is executed as follows:

i. Initialize size [*k*] = 0. At the same time, set the value of *k* = 0 (where *k* is the number of current region of interest).
ii. Search for the *S*-component pixel value, Saturation = *μ*. Region growing process starts when the seed value is found. This first pixel corresponded as the seed point. The value of *k* is increased from *k* to *k* + 1 and size[*k*] = 1; otherwise, go to step (viii).
(Note: The value Saturation = *μ* is the threshold value obtained from *S*-plot during segmentation process in the earlier section 2.2)
iii. The eight neighborhood pixels are checked and are added to the region if they have similar pixel value to the seed, Saturation pixel = *μ*. Increase size[*k*] = size[*k*] + 1 for each pixel that fulfill the growing condition.
iv. Search the size[*k*] with the conditions,
    a. If size[*k*] has the pixel value in between *a*1 and less than *a*2, the total of ROI is increased by 1.
    b. Else if size[*k*] has larger value than *a*2 but less than a3, the total of ROI is increased by 2.
    c. Else if size[*k*] has larger value than *a*3 but less than *a*4, the total of ROI is increased by 3.
    d. Else if size[*k*] has larger value than *a*4 but less than *a*5, the total of ROI is increased by 4.
    e. Else if size[*k*] has larger value than *a*1 or larger value than *a*5, the total of ROI will not be increased.
    (Note: if size[*k*] has a larger pixel value than *a*5, it refers to the smudge cells or other artifacts in the acute leukemia blood images.)
v. The growing process from neighborhood pixel in step (iii) is repeated and size[*k*] = size[*k*] + 1 is increased for every pixel that satisfies the condition. The process is repeated until the ROI cannot be expanded or all the pixels in the current ROI are added to the region.
vi. Search for new seed point which does not belong to the previous region.
vii. If the seed point is found, increase *k* to *k* + 1 and size [*k* + 1] = 1. Go to step (iii); otherwise, go to step (viii).
viii. End.

The eight neighborhood seed-based region growing algorithm of Adams *et al.* has been modified and improved in this paper for the suitability for normal and acute leukemia peripheral blood images.[15]The suggested parameter values in this algorithm are shown in Table 1. The parameter *μ* refers to the threshold value from the *S*- component of HSI color space and the parameter values *a*1 to *a*5 refers to the area of pixels in the normal or abnormal individual WBC cells.

Table 1 Parameter value for SRG area extraction algorithm

No.	Parameter	Value
1	μ	0.35
2	a_1	40 pixel
3	a_2	440 pixel
4	a_3	590 pixel
5	a_4	720 pixel
6	a_5	900 pixel

3.0 DATA SAMPLE AND SYSTEM EVALUATION

In this study, the proposed procedures for counting of acute leukemia cells have been applied and tested on 100 images including normal and acute leukemia images. Figure 5(a), 6(a) and 7(a) show the original images of normal, AML and ALL respectively. In order to perform segmentation on the acute leukemia image, the *S*-component images that have been extracted from original normal, ALL and AML images are shown in Figure. 5(b), 6(b) and 7(b), respectively. Although segmentation using thresholding is able to segment the WBCs in the image, however some unwanted regions such as segmented RBCs can still be seen in the image. Therefore, these images are extras with median filter and SRG algorithms which remove the unwanted regions from the image. The results of the combined formula *S*-component based on HSI color model, the threshold method, 5 x 5 median filter and SRG algorithm can determine the total number of WBC in the blood images.

As shown in the Figure 5(c), 6(c) and 7(c) value's stray pixels appear in the output image due to the pixel value similarity between the wanted and unwanted region. Meanwhile, Figure 5(d), 6(d) and 7(d) show results of images after filtering the resultant thresholding using 5x5 median filter algorithm. By filtering the image, most of the small background pixels have been removed. However, some unwanted regions which are bigger in size can still be seen in the image. In order to give better visualization results, the SRG algorithm has been further applied as shown in Figure 5(e), 6(e) and 7(e).The unwanted pixels in this case are referred to as salt and pepper noise which are eliminated in the next step. The selected *μ* value ensures the pixels value of the WBCs or acute leukemia cells remains the same as its original image and converts all the unrelated value to white color. In human blood image, only WBCs have a nucleus. Therefore, if the nucleus can be detected, the WBC count was found.

To evaluate the suitability of the developed counting procedures, 100 images in total were used to validate the system, which consists of 40 normal images, 30 AML and 30 ALL images. The total number of WBC cells and the acute leukemia cells were counted, differentiated and confirmed manually by the trained experts from HUSM, Kubang Kerian, Kelantan, Malaysia.

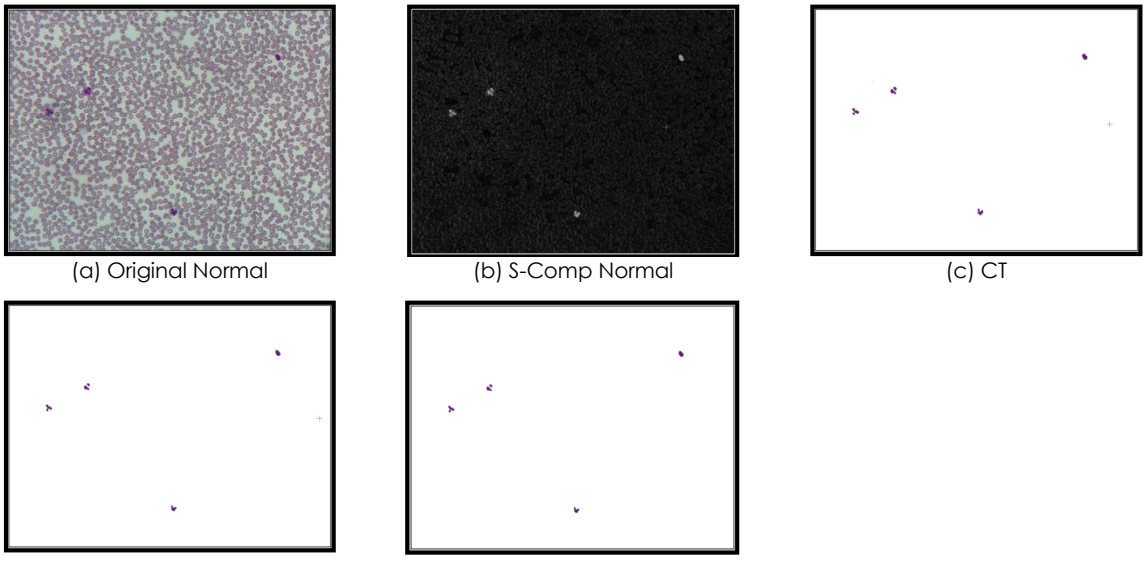
Next, the manual counted cells were used to compare with the total number of cells detected by the developed system. The performance of the system was evaluated based on the sensitivity percentage using the equation (5) as follows:

$$Sensitivity\% = \left(\frac{TP_s}{TP_s + FN_s}\right) \times 100\% \quad (5)$$

The sensitivity percentage evaluates how well the system is to count and differentiate the normal or abnormal WBC cells correctly depending on the total number of cells. The parameter *TPs* refers to the total number of WBC components that can be detected by the system in the normal or abnormal blood images, also known as 'true positive'; *FNs* refers to the condition where the system fails to detect the WBC, also known as 'false negative'. The results of the analysis on system performance are shown in the following tables. Table 2 shows the analysis on 40 normal images, while 60 acute leukemia images on AML (30 images) and ALL (30 images) are shown in Table 3 and Table 4 respectively. The overall performance analysis on normal, AML and ALL images is shown in Table 5. Another parameter, FP_s which are shown in the following tables, refers to the non-WBC cells detected as WBC cells by the system.

As shown in the Table 5, the average sensitivity of the system in recognizing and counting the number of WBC from 40 normal blood images is 98.46%, while 100% for AML and 98.88% for ALL images. Overall, 5951 WBC cells were counted and confirmed manually by the trained experts in the 100 images used (40 normal images, 30 AML and 30 ALL images) and the system detected 6231 cells. The system was able to detect 5897 WBC cells out of 6231 cells, which belongs to the *TPs* group. Moreover, the system identified 334 non-WBC cells (*FPs*) as WBC cells and 54 WBC cells were not detected (*FNs*). Performance wise, all running programs were shut down except the system to ensure there was no disturbance of any kind towards the system processing performance.

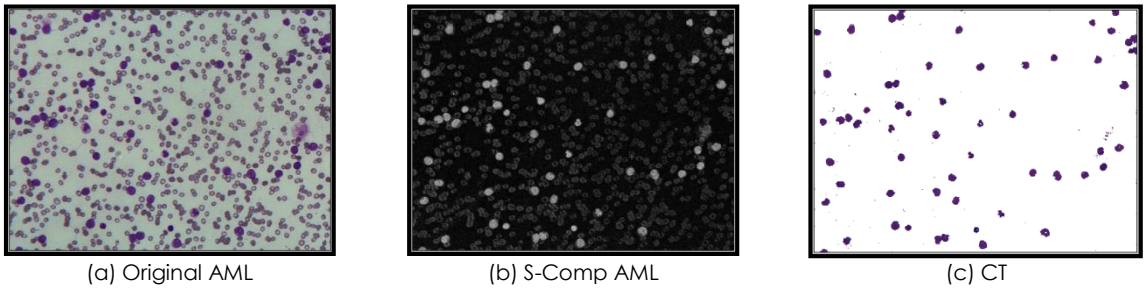
The time the system took to run the complete cells screening process was no more than six seconds, which means the duration did not exceed ten minutes for every 100 images tested. Therefore, the computer aided counting system for acute leukemia is able to help reduce the time needed to manual cell counting by the experts / specialists.



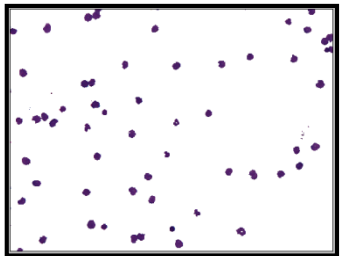
(a) Original Normal (b) S-Comp Normal (c) CT

(d) MF (e) SRG

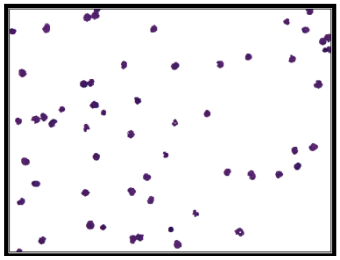
Figure 5 Normal Image



(a) Original AML (b) S-Comp AML (c) CT

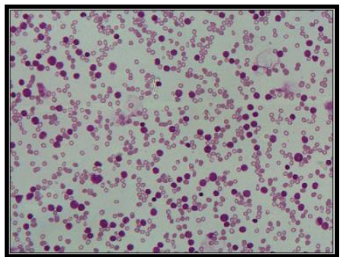


(d) MF

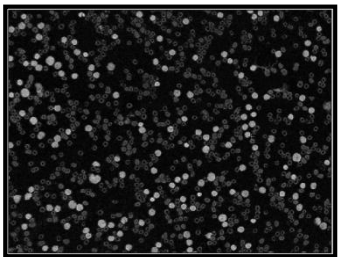


(e) SRG

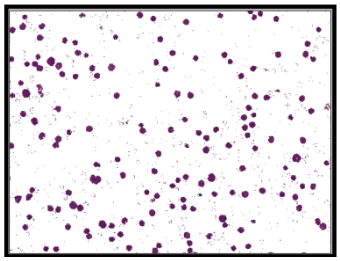
Figure 6 AML Image



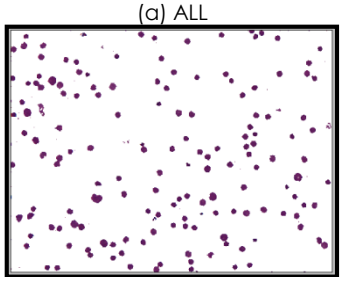
(a) ALL



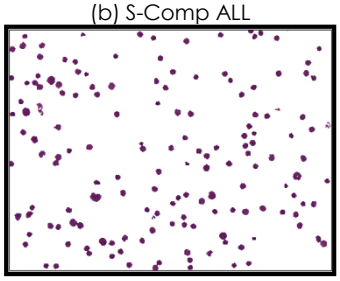
(b) S-Comp ALL



(c) CT



(d) MF



(e) SRG

Figure 7 ALL Image

Table 2 System performance on normal blood images

Normal Image Set	Manual Count	System Count	TP	FP	FN	Sensitivity (%)
Image 1 – 10	44	48	43	5	1	98.33
Image 11 – 20	40	42	40	2	0	100
Image 21 – 30	59	66	59	7	0	100
Image 31 – 40	55	53	49	4	6	90.23
Total	198	209	191	18	7	97.21

Table 3 System performance on AML images

AML Image Set	Manual Count	System Count	TP	FP	FN	Sensitivity (%)
Image 1 – 10	536	604	536	68	0	100
Image 11 – 20	465	525	466	60	0	100
Image 21 – 30	550	637	550	87	0	100
Total	1552	1767	1552	215	0	100

Table 4 System performance on ALL images

ALL Image Set	Manual Count	System Count	TP	FP	FN	Sensitivity (%)
Image 1 – 10	1608	1578	1577	1	31	98.09
Image 11 – 20	1703	1696	1687	9	16	99.09
Image 21 – 30	890	981	890	91	0	100
Total	4201	4255	4154	101	47	99.06

Table 5 Overall system performance on 100 images

Types	Total Images	Manual Count	System Count	TP	FP	FN	Sensitivity (%)
Normal images	40	198	209	191	18	7	98.46
AML images	30	1552	1767	1552	215	0	100
ALL images	30	4201	4255	4154	101	47	98.88
Total	100	5951	6231	5897	334	54	99.09

3.0 CONCLUSION

Generally, this paper recommends a new procedure by integrating several techniques in image processing to count the WBC and acute leukemia cells using image capturing tool with 10*X* magnification. The advantage of the computer aided acute leukemia counting system is important as an assisting facility to the hematology department that is able to count and identify WBC or acute leukemia cells effectively without being influenced by operator fatigue.

The combination of *S*-component based on HSI color model, thresholding, 5 x 5 median filter and seeded region growing in this paper has proven its ability to determine the total number of WBC nucleus in the normal and acute leukemia blood images. To ease the blood cell segmentation process which consists of WBCs, RBCs and background, *S*-component of HSI color model is proposed. 5 x 5 median filter is implemented to remove salt and pepper noise after thresholding. For a more sophisticated noise elimination process, seeded region growing algorithm was introduced. The algorithm is also used to count the total number of WBC and acute leukemia cells based on the size of the cell detected in the blood image. Besides that, the resultant image from the new procedure makes it easier for the experts to analyze images compared to original images with various cells and artifacts.

Acknowledgement

The authors would like to acknowledge and thank the members of the leukaemia research team at Universiti Malaysia Perlis (UniMAP) for making this research achievable, Hospital Universiti Sains Malaysia (HUSM) for providing the acute leukaemia blood samples and Universiti Utara Malaysia for supporting this research.

References

[1] MeIntyre, D. 2002. *Color Blindness*. Dalton Publishing.
[2] Christine, R. 1999. The Eye of the Beholder–Designing for Colour-blind Users. *British Telecommunications Engineering*. 17: 291-295.
[3] Neitz, M. and Neitz, J. 2000. Molecular Genetics of Color Vision and Color Vision Defects. *Archieves of Ophthalmology*. 63(2): 232-237.
[4] Healy, G., Shafer, S. and Wolff, L. 1992. *Physics Based Vision: Principles and Practice, COLOR*. Boston: Jones and Bartlett.
[5] Brettel, H. and Vienot, F. 2001. *Color Display for Dichromats, Proceeding of SPIE on Color Imaging*. 4300: 199-207.
[6] Poret, S., Jony, R. D. and Gregory, S. 2009. Image Processing for Color Blindness Correction. *IEEE Toronto International Conference*. 1-6.
[7] Ohkubo, T. and Kobayashi, K. 2008. A Color Compensation Vision System for Color-blind People. *SICE Annual Conference*. The University ElectroCommunications Japan.
[8] Plataniotis, K. N. and Vinetsanopoulos. A. N. 2000. *Color Image Processing and Application*. Berlin: Springer-Verlag.
[9] McDowell, Jason. 2008. *Design of a Color Sensing System to Aid the Color Blind*. 27: 34-39.
[10] SeuttgiYmg and Yong Man Ro. 2003. *Visual Contents Adaptation for Color Vision Deficiency*. 1: 453-456.
[11] Yau-Hwang Kuo and Jang-Pong Hsu. 1996. *MCFC-R: A Fuzzy Connectionist Model for Color-blindness Plate Recognition*. 2: 718-723.
[12] Swain, M. and Ballard, D. 1991. *Color Indexing. International Journal of Computer Vision*. 7: 11-32.
[13] Birch, J. 2012. Worldwide Prevalence of Red-green Color Deficiency. *J Opt Soc Am A Opt Image Sci Vis*. 29(3): 313-320.
[14] Konstantakopoulou, E., Rodriguez-Carmona M., and Barbur J. L. 2012. Processing of Color Signals in Female Carriers of Color Vision Deficiency. *Journal of Vision*. 12(2): 1-11.
[15] Hood, S. M., Mollon, J. D., Purves, L. and Jordan, G. 2006. Color Discrimination in Carriers of Color Deficiency. *Vision Research*. 46: 2894-2900.

[16] Nathans, J., Thomas, D., and Hogness, D.S. 1986. Molecular Genetics of Human Color Vision: The Genes Encoding Blue, Green, and Red Pigments. *Science*. 232(4747): 193–202.
[17] Sharpe, L. T., Stockman, A., Jagle, H. and Nathans, J. 1999. Opsin Genes, Cone Pigments, Color Vision and Color Blindness. In Gegenfurtner K. R., Sharpe, L. T. (eds*). Color Vision.* Cambridge: Cambridge University Press.
[18] Walraven, J. and Alferdinck, J. W. 1997. Color Displays for the Color Blind. Proc. *On Color Science, Systems, and Application of 5th Color Image Conference, Scottsdale, Arizona: Society for Imaging Science and Technology*. 17–22.
[19] Bimber, Oliver, and Ramesh, Raskar. 2005. *Spatial Augmented Reality*. Massachusetts: A K Peters.
[20] Products for the Blind and Visually Impaired: Colorino retrieved November, 20, 201 from http://www.caretec.at/ColorTest_Colorino.32.0.html/.
[21] Brettel, H., Vienot, F. and Mollon, J. 1997. Computerized Simulation of Color Appearance of Dichromats. *Journal of Optical Society of America.* 14(10): 2647–2655.
[22] Solem, J. E. 2012. *Programming Computer Vision with Python*. Sebastopol: O'Reilly Media.
[23] Joseph Howse. 2013. *OpenCV Computer Vision with Python*. Birmingham: Packt Publishing Ltd.
[24] Bradsky, G. and Kaehler, A. 2008. Learning OpenCV. California: O'Reilly Media.
[25] Jeffries, B. J. 1880. *Color-blindness: Its Dangers and Its Detection.* Boston: Houghton, Osgood and Company.