The Effectiveness of Colour Constancy on Segmenting Leukemia Cells Using Unsupervised Clustering Technique

Nurul Hazwani Abd Halim* MohdYusoffMashor* Aimi Salihah Abdul Nasir* RoslineHassanb

*aElectronic & Biomedical Intelligent Systems (EBItS) Research Group, School of Mechatronics Engineering, University Malaysia Perlis, 02600 KampusPauh Putra, Perlis, Malaysia.

bDepartment of Hematology, School of Medical Science, University Science Malaysia, KubangKerian, Kelantan, Malaysia.

*aCorresponding author: nurul.hazwani43@yahoo.com.

Abstract. Nowadays, colour image segmentation plays important tasks in medical field application such as brain, eyes, glaucoma and so on. This paper aims to compare the performance between different colour constancy of gray world (GW) and white patch (WP) that will be fusion with clustering algorithm to segment the nucleus of blast cells. First, both of the colour constancy algorithms are utilized on leukemia images to improve the microscopic quality of the blast cells. An unsupervised clustering, namely moving k-means algorithm is utilized on the various colour constancies of GW and WP for the purpose of segmentation. Both colour constancy algorithms have been analysed to identify which one can give the good segmentation performance. Since there is some noise in the segmented nucleus, the combination of median filter and region growing technique plays important tasks in reducing noise. The results show that WP is the best colour constancy in segmenting the nucleus as compared to Gray World.

1. Introduction
Leukemia is a cancer of blood which is white blood cells and its early diagnosis can prevent the rapid progression of the disease. Acute leukemia can be classified into two categories: acute lymphoblastic leukemia (ALL) and acute myelogenous leukemia (AML) [1]. ALL occurs for children of age 2-5 years old [1] while AML occurs in adults which effect in men than women [2]. The recent development in diagnosing of the leukemia are fluorescent microscopy, polymerase chain reaction (PCR), cytochemistry, cytogenetic and molecular [3]. Despite these advances, light microscopy remains the golden standard in developing countries. This is because the method is simple, rapid, and low cost. However, microscopic examination has several drawbacks, which are depending on the subjectivity in human perception, labor intensive and time-consuming [3]. Image processing method is one of the most important steps in preparing a way for a successful implementation of existing technologies for leukemia diagnosis.

In digital images, colour constancy plays important role in removing unreal colour from the image, which is caused by the acquisition condition [4]. There are several algorithms have been proposed by researchers such as Gray World [5], White Patch [6] and Iterative White Balance algorithm [6].
general, the gray world will use the average of the pixels as their illuminant, while the maximum values in the image will be used in white patch method [6]. Many researchers worked on blood cell segmentation and have been proposed various techniques to segment the region of interest, WBC. Thresholding technique is a most common conventional algorithm and usually depend on the histogram in order to select the threshold value correctly. Green component of the RGB has been extracted and then thresholding technique has been applied for this authors in order to segment the WBC [7]. Optimal thresholding and watershed transformation are other approaches that have been introduced by them in order to segment WBC [8]. Lim proposed a double K-means clustering and watershed [9]. Thus the current study proposed a combination between colour constancy (Gray World and White Patch) and moving k-mean clustering algorithm. The best segmentation performance was analysed either WP or Gray World algorithm that can segment nucleus successfully.

2. Methodology
The methodology for the proposed segmentation method for acute leukemia images are shown as in Figure 1.

![Figure 1](image)

**Figure 1.** The proposed method for segmentation of acute leukemia.

2.1. Image Acquisition
A total of 100 leukemia images that were obtained from Hospital Universiti Sains Malaysia (HUSM), were captured under 100X magnification and at a resolution of 1280×960 pixels. Samples of captured images of acute leukemia cells can be seen in Figure 2.

![Figure 2](image)

**Figure 2.** Sample images of acute lymphoblastic leukemia (ALL).

2.2. Colour Constancy Algorithms
Based on the gray world theory, the average colour of the pixels will be used in order to estimate the illuminant [6]. Assume \(x \times y\) pixels are the resolution of leukemia image. Let the image \(s(x, y)\) with the size of image \(M \times N\) and the average colour can be calculated as below:

\[
R_{\text{avg}} = \frac{1}{M \times N} \sum_{x=1}^{M} \sum_{y=1}^{N} S_R(x, y)
\]  

(1)
\[ G_{avg} = \frac{1}{M \times N} \sum_{x=1}^{M} \sum_{y=1}^{N} S_G(x, y) \]  
(2)

\[ B_{avg} = \frac{1}{M \times N} \sum_{x=1}^{M} \sum_{y=1}^{N} S_B(x, y) \]  
(3)

Where: \( S_R(x, y) \), \( S_G(x, y) \) and \( S_B(x, y) \) are the red, green and blue channel for the image respectively.

Generally, white patch will use the determined white point as the maximum values, \( R_{max} \), \( G_{max} \) and \( B_{max} \) based on the presumption that there is always white in the image. The formula can be calculated as below [7]:

\[ R_{max} = \max \{ S_R(x, y) \} \]  
(4)

\[ G_{max} = \max \{ S_G(x, y) \} \]  
(5)

\[ B_{max} = \max \{ S_B(x, y) \} \]  
(6)

2.3. Conversion of RGB to HSI Colour Models

Three features can be determined based on HSI colour model which is hue, saturation and intensity. The following equations were used to transform RGB to HSI colour model [10].

\[ H = \begin{cases} \cos^{-1} \left( \frac{1/2(R-G)+G-R)}{|(R-G)^2+(R-B)(G-B)|^{1/2}} \right) & \text{for } B \leq G \\ 2\pi - \cos^{-1} \left( \frac{1/2(R-G)+(G-R)}{|(R-G)^2+(R-B)(G-B)|^{1/2}} \right) & \text{for } B > G \end{cases} \]  
(7)

\[ S = 1 - \frac{3}{(R+G+B)} \times \min(R, G, B) \]  
(8)

\[ I = \frac{1}{3} \times (R+G+B) \]  
(9)

2.4. Moving k-Means Clustering Algorithm

Recently, the most commonly clustering algorithms used by researchers are \( k \)-means clustering and fuzzy \( c \)-means clustering. However, both techniques have several drawbacks such as centre redundancy, dead centre and the centre trapped in local minima which can be leading a poor result. Thus, a modified version of \( k \)-means clustering called moving \( k \)-means clustering [11] will be implementing in this research. In this study, the saturation component (from the HSI colour model) was selected to be segmented using moving \( k \)-means clustering as the blast cells can be seen clearly rather than other colour component as shown in Figure 3(c).
Assume a leukemia image with resolution of $x \times y$ pixels to be clustered into $n_c$ regions. Let $s(x, y)$ as an input pixel to be clustered and $c_j$ is $j^{th}$ centre (cluster) ($x = 1,2,..X$, $y = 1,2,..Y$ and $j = 1,2,..n_c$). The procedures for MKM clustering are as follows [11]:

1. The centres and $\alpha_a$ is initialized and set $\alpha_a = \alpha_b = \alpha_o$ (where $\alpha_o$ is a small constant value, $0 < \alpha_o < 1/3$and should be chosen to be inversely proportional to the number of centres. $\alpha_a$ and $\alpha_b$ are small constants).
2. The Euclidean distance between the pixel and all the centres is calculated. All pixels are assigned as the member of the centre with the smallest Euclidean distance.
   \[ c_j = \frac{1}{n_j} \sum_{x \in c_j, y \in c_j} s(x, y) \]  
   (10)
3. The fitness of each centre is checked using the equation:
   \[ f(c_j) = \sum_{x \in c_j, y \in c_j} \left(\left\| s(x, y) - c_j \right\| \right)^2 \]  
   (11)
4. Find $c_s$ and $c_l$, the centre that has the smallest and the largest value of $f(.)$.
5. If $f(c_s) < \alpha_a f(c_l)$,
   (a) The members of $c_s$ and $c_l$ are assigned, if $s(x, y) < c_s$, where $x, y \in c_l$, and leave the rest of the members to $c_l$.
   (b) The positions of $c_s$ and $c_l$ are recalculated based on:
   \[ c_s = \frac{1}{n_s} \sum_{x \in c_s, y \in c_s} s(x, y) \]  
   (12)
   \[ c_l = \frac{1}{n_l} \sum_{x \in c_l, y \in c_l} s(x, y) \]  
   (13)
   Note: $c_s$ will give up its members before step 5(a) and, $n_s$ and $n_l$ in (7) and (8) are the number of the new members of $c_s$ and $c_l$ respectively, after the reassigning process in step 5a.
6. $\alpha_a$ is updated according to $\alpha_a = \alpha_a - \alpha_a/n_c$ and steps 4 and 5 are repeated until $f(c_s) \geq \alpha_a f(c_l)$
7. All pixels are reassigned to the nearest centre and the centre positions are recalculated using (5).
8. $\alpha_a$ and $\alpha_b$ are updated based on $\alpha_a = \alpha_a$ and $\alpha_b = \alpha_b - \alpha_b/n_c$ respectively, and steps 3 to 7 are repeated until $f(c_s) \geq \alpha_b f(c_l)$.

Results & Discussions

Comparisons between the gray world and white patch algorithms have been performed to validate the best colour constancy in segmenting the nucleus of blast cells. Input images of each colour constancy have been utilized for MKM clustering. Figure 4 (a), (b) and (c) are examples of acute leukemia images that will be conduct in this paper. Both of the original image and after enhancement image have been analysed based morphological features of blast cells. The results after implementing the gray world and white patch can be seen in Figure 5 and Figure 6 (a), (b) and (c), respectively. By using white patch algorithm, the contrast of cytoplasm and blast cells were significantly increased rather than the gray world. The output of the gray world image was tended to brighter and appear likely to original images.
A saturation component from HSI colour model was used as input for MKM clustering. Figure 7 and Figure 8 are shown the saturation colour component was extracted from GW and WP images, respectively. Based on the resultant results, the blast cells appear as the brightest areas when saturation component was used on both gray world and white patch but RBC and background areas can be seen more clearly using white patch algorithm.

In Figure 9 and Figure 10 are shown the MKM segmentation results using saturation component for the gray world and white patch, respectively. From the images, MKM will be cluster into three categories such as blast cell, RBCs and background.
Figure 9. Results after applying moving $k$-means clustering on saturation component images for gray world

(a) ALL1
(b) ALL2
(c) ALL3

Figure 10. Results after applying moving $k$-means clustering on saturation component images for white patch

Figure 11 and Figure 12 indicated the retrieved images based on the gray world and white patch, respectively. Our proposed technique is not working very well on ALL2 for gray world compared to white patch due to having the similarities the pixels’ value between blast cells and also RBCs.

Figure 11. Results after moving $k$-means clustering image has been retrieved based on gray world image.

(a) ALL1
(b) ALL2
(c) ALL3

Figure 12. Results after moving $k$-means clustering image has been retrieved based on white patch image.
There are some small areas of RBCs that could not be eliminated using the clustering process since it has similar pixel values with blast cells. To overcome this problem, the fusion of the $7 \times 7$ pixels' median filter with region growing were processed for smoothing the image, and to get rid of any large unwanted regions from the image, respectively. Non-blast cells were eliminated from the images since it has less than 1000 pixels. The results after implementing the region growing technique and median filter are shown in Figure 13 and Figure 14. The segmented nucleus has been successful obtained by using WP since it still maintains the features of the nucleus rather than Gray World algorithm.

3. Conclusion
Segmentation using various colour constancy algorithms and MKM clustering has been presented. In this study, comparison between GW and WP has been making to analyse the significant of it in segmenting the acute leukemia images. It is shown that WP algorithm has proven a better outcome in segmenting nucleus from acute leukemia images compared to GW algorithm.

Acknowledgments
The authors gratefully acknowledge to the team members and University Science Malaysia (USM).

4. References
[1] S. Rajendran, “Image Acquisition and Retrieval Systems for Leukaemia Cells” Master’s thesis, Department of Computer Science, UM, 2007.
[2] J.N. Jameson, L.K. Dennis, T.R. Harrison, E. Braunwald, A.S. Fauci, S.L. Hauser, D.L. Longo, “Harrison's principles of internal medicine”, New York: McGraw-Hill Medical Publishing Division, 2005.
[3] F. Thung, and I.S. Suwardi, "Blood parasite identification using feature based recognition," in International Conference on Electrical Engineering and Informatics (ICEEI), 2011, pp.1-4.
[4] R.A.A.Raof,M.Y. Mashor ,R.B.Ahmad, S.S.Md.Noor , “Image Enhancement Using White Balance Technique in Ziehl-Neelsen Sputum Slide Images”,Proc. of International Postgraduate Conference on Engineering(IPCE 2010),16-17 October 2010.
[5] G. Buchsbaum,” A special processor model for object color perception “, in Journal of Franklin Institute, Vol 310,pp.1-26, 1980.
[6] V. Cardei, B. Funt, and K. Barndad, “White point estimation for uncalibration images,” in Proceeding of the Seventh Color Imaging Conference (CIC’99), pp 97-100, 1999.

[7] M. Mohamed, and B. Far, "An enhanced threshold based technique for white blood cells nuclei automatic segmentation," in e-Health Networking, Applications and Services (HealthCom), 2012 IEEE 14th International Conference on, vol., no., pp. 202-207, 10-13 Oct. 2012.

[8] M. Mohamed, B. Far and C. Naugler, "Chronic lymphocytic leukemia cell segmentation from microscopic blood images using watershed algorithm and optimal thresholding," in Electrical and Computer Engineering (CCECE), 2013 26th Annual IEEE Canadian Conference on, Regina, SK, 2013, pp. 1-5.

[9] L. HueyNee, M.Y. Mashor, N.Z. Supardi, R. Hassan, "Color and morphological based techniques on white blood cells segmentation," in Biomedical Engineering (ICoBE), 2015 2nd International Conference on, vol., no., pp. 1-5, 30-31 March 2015.

[10] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3rd ed., Prentice Hall, 2007.

[11] M.Y. Mashor, International Journal of the Computer 8, (2), pp. 50–65 (2000).