Journal of Fundamental and Applied Sciences

ISSN 1112-9867

Available online at http://www.jfas.info

Research Article

Special Issue

FUSION7×7 MEDIAN FILTER AND SEEDED REGION GROWING AREA EXTRACTIONALGORITHMS FOR EFFECTIVEDETECTION OF ACUTE LEUKEMIA BASED ON BLOOD IMAGES

H. N. Hazlyna[1,*], H. H. Aini[1], S. Fadzilah[1], K. Nadia[2], M. Y. Mashor[2], R. Hassan[3]

[1]School of Computing, College of Arts and Science, Universiti Utara Malaysia, Sintok, Kedah, Malaysia

[2]School of Mechatronics Engineering, Universiti Malaysia Perlis, Pauh, Perlis, Malaysia

[3]Department of Hematology, Universiti Sains Malaysia, Kubang Kerian, Kelantan, Malaysia

Published online: 24 February 2018

ABSTRACT

In the most medical image analysis tasks, image segmentation is a crucial step where one of the usual segmentation methods is based on clustering algorithms. Clustering algorithms have been employed as a digital image segmentation method in various fields, including engineering, computer and mathematics. This study aims to detect acute leukemia cells based on blood images with effective methods. Normally, although the segmentation process is completed, there is still the existence of unwanted noise and background in acute leukemia blood images. Thus, this study proposes a fusion of median filter and seeded region growing area extraction algorithm to overcome this problem. Finally, the results demonstrate that the proposed method has successfully been distinguished and segmented between acute leukemia cells from the unwanted noise or complicated background in blood images. The proposed method is able to yield an average of 98.26% based on the percentage of accuracy.

Keywords: Acute leukemia, median filter, Seeded Region Growing Area Extraction, k-Means

Author Correspondence, e-mail: hazlyna@uum.edu.my

DOI: http://dx.doi.org/10.4314/jfas.v10i3s.62

I. INTRODUCTION

A. What is Leukemia?

Leukemia is one of the blood type cancer that comes from bone marrow disorder and it is detected through condition and amount of the white blood cells. It usually happens if an abnormal white blood cell begins to duplicate itself, accumulate and then it prevents the production of other normal blood cells in the marrow, leading to recurrent infections, bleeding and anemia. From time to time, leukemic cells proliferate through the bloodstream where they further form tumors and destruct essential organs such as the liver and kidney. Basically, leukemia can be divided into acute leukemia (which progresses fast) and chronic leukemia (which progresses in a slow manner). The classification of acute leukemia comes in two types namely Acute Myelogenous Leukemia (AML) and Acute Lymphoblastic Leukemia (ALL) according to the French-American-British (FAB). AML primarily affects adults, but it can happen in children and adolescents while ALL is most common in children.

From 2004 to 2010, the five-year relative survival rates have shown that 66.3% children and adolescents not older than 15 years were diagnosed with AML and 93% children not older than 5 years and 91.8% children and adolescents not older 15 years diagnosed with ALL in United States[1]. According to the National Cancer Institute report in 2009, for 17 SEER (Surveillance, Epidemiology and End Results) geographic areas, it is estimated that 12,330 people (6,590 men and 5,740 women) will be diagnosed with AML in 2010 [2]. While, the statistic it is also estimated that 5,330 people (2,180 women and 3,150 men) will be diagnosed with ALL in 2010. Based on these statistics, the importance of cost-effective and fast production of blood cell count reports is paramount in the healthcare industry. Thus, several methods of automated leukemia detection have been developed due to its high consumption in medical diagnostics. These brought to the implementation of several algorithms and techniques to make a system powerful to achieve the best detection. However, some techniques have their own weaknesses. For blood cell images, segmenting and counting them remain a challenging task owing to the complex nature.

B. Problem and Motivation

The main problem in many cases of white blood cell segmentation is that the edge detection becomes poor on blood cell images. Previously, there are several techniques that have been suggested for the leukemia blood cell segmentation. For example, [3] ulitized the morphological analysis of microscopic images to identify the leukemia cells. [4] proposed Hausdorff Dimension and contour signature to classify a lymphocytic cell in the blood image into normal lymphocyte or lymphoblast (blasts). They also extracted the shape and texture features to get better classification. [5] utilized the adaptive k-Means clustering together with the mean shift algorithm to segment the acute leukemia blood cell images. Many methods tend to generate comprehensive output images with minimal filtering process for the purpose of removing the background scene. However, it is also difficult to get the entire edge information and locate cells accurately and not all boundaries are sharp. The main factor that contributed to the problem is the appearance of noise. The results of image processing like segmentation, feature extraction, and classification normally depend on the results of the noise removal.

Thus in this paper, the effectiveness of fusion 7x7 median filters and seeded region growing area extraction (SRGAE) method used to remove unwanted noise from acute leukemia images. The advantages of this fusion method are that it is more robust and effective to remove salt and pepper. Other than that, it can reduce large noise between a dark and a light area. Highlighting the organization of the rest of the paper, Section 2 describes the methodology of the image segmentation, whereas Section 3 describes the experimental result and in the final section, the conclusion.

II. METHODOLOGY

The aims of this proposed method is to provide an automated segmentation process by combining the removal noise methods, hence producing an effective effect to resultant images. The proposed technique comprises of the saturation component based on the HSI color space, k-means algorithm, the fusion of 7x7 pixels median filter and SRGAE algorithms. Furthermore, the proposed framework of this study is illustrated in Figure. 1. The dotted line box illustrates the implementation of the noise removal methods for improved resulting images after completing the segmentation process.

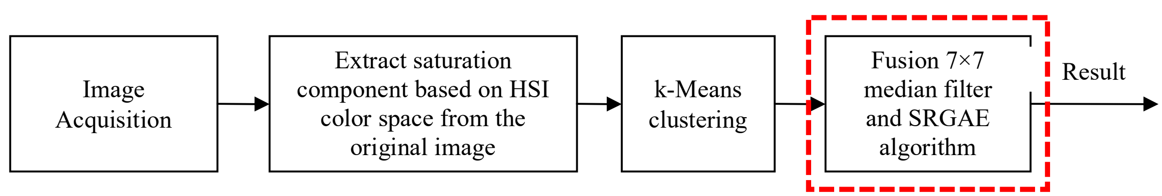
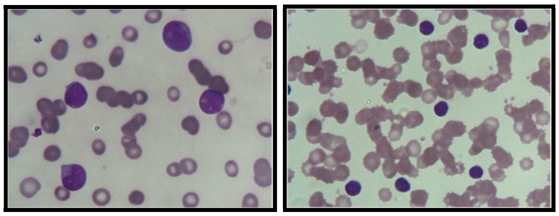


Fig.1. Automated segmentation using the fusion of 7×7 Median filter and SRGAE algorithm.

A. *Image Acquisition*

The samples of acute leukemia were collected from Universiti Sains Malaysia Hospital (HUSM) Kubang Kerian Kelantan. Leica microscope at 40 magnifications was utilized to analyze the acute leukemia blood slide images. To capture the images, *Infinity* 2 camera was used and saved into (. *bitmap) format with 800×600 resolution. The validation and confirmation on the total number of cells and its group in both the normal and abnormal cases were done by the hematologists. Total data used in this study are 100 acute leukemia images (type AML = 50 images and type ALL = 50 images). Figure 2 (a) and 2 (b) present the original images of AML and ALL, respectively.



(a) AML image

(b) ALL image

Fig.2. Samples of original blood images for AML and ALL types.

B. Threshold Selection Methods

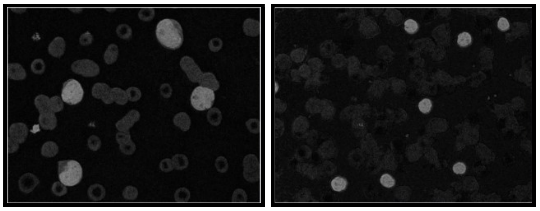
1) Extracting S-component based on HSI color space

RGB color space serves to be one of the color image segmentation techniques that was conducted in [6] to identify acute leukemia blood cells. Nevertheless, one of the issues when using RGB color space is that the pixel threshold value did not present well because of the image's color's inconsistency. This is subject to the blood concentrations and the preparation for the staining procedure (operator-dependent). Additionally, the threshold value is also recognized having referred to the changes in the acute leukemia blood cells image. Therefore, to reduce this problem, segmentation using HSI color space was used. Figure 3 shows the resultant AML and ALL images after applying the saturation formula.

The acute leukemia cells (blast) are clearly seen and highlighted in the S- component image based on the observation of the blood image. Meanwhile, less saturation component image was seen in the visibility for the other particles and red blood cells. Thus, the saturation component was selected to facilitate the clustering process and lessen the computational effort. Equation (1) shows that the saturation formula was employed particularly for the original acute leukemia blood images.

$$Saturation = 1 - \left(\frac{3}{M+H+B}\right)\min(M, H, B) \qquad (1)$$

where M, H and B indicate the pixel components of red, green and blue of the white blood cell in the blood image.



(a) AML image (b) ALL image

Fig.3. Sample resultant images for AML and ALL types after applying S-component based on HSI colour space.

2) Segmentation using k-Means clustering algorithm

After the utilization of the saturation formula, the k- means clustering process is performed. Each pixel in the image will be sorted out automatically into three clusters where the intensity is almost the same with every cluster. The k- means algorithm can be implemented to the saturation component resultant images [7] as follows:

1. Randomly *n* points are selected and placed into the space presented by the pixels that are being clustered. These points present the initial cluster center C_c , where c=1, 2, 3…,n.
2. Calculate the Euclidean distance E_d for each pixel, x_{im} of an image

$$E_d = \|x_{im} - C_c\|^2 \tag{2}$$

3. Assign each pixel to the cluster that has the nearest center.
4. Recalculate the positions of the *n* centers when all pixels have been assigned.
5. Repeat Steps 2 and 4 until the centers no longer move.

C. Removal Noise Methods for Effectiveness Effect

1) Fusion of 7×7Median Filter and SRGAE

In this study, the fusion of the 7×7 median filter and SRGAE was proposed as the improvement method for an effective detection of acute leukemia based on blood images. Normally, after the segmentation process is completed, there are still small holes on the object of interest and the existence of noise known as salt and pepper. The use of the 7×7 median filter was proposed because of its effectiveness in reducing the salt and pepper noise while maintaining the edge. At the first step, the segmented images were filtered using the 7×7 pixels median filter. This process is able to yield a better result in the visualization of acute leukemia blood cells image by removing the unwanted noise and eliminating the background. Median filter often serves for images that have gray level intensity. As suggested by its name, it works by substituting the value of a pixel by the median of the gray levels in the pixels' neighborhood. This neighborhood can also be depicted by the square mask. The square mask for $n \times n$ pixels chosen is shown in Figure 4 where n must be odd-numbered – it could be 3, 5, 7 and so on.

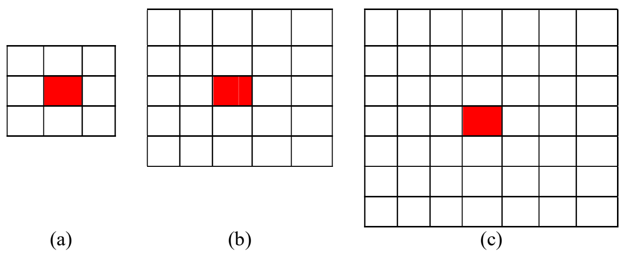


Fig.4. Square mask of $n \times n$ pixels with (a) n = 3 (b) n = 5 (c) n = 7.

Here, the neighborhood of $n \times n$ (n = 7) pixels is selected because large neighborhoods produce more severe smoothing. The steps for median filter [8] are as follows:

i. Read all the intensity values in a specified neighborhoods.
ii. Sort the values in ascending order.
iii. Determine the median among the sorted values.
iv. Assign the value to the center pixel in the neighborhood.
v. Move the neighborhood one pixel to the left.
vi. Repeat steps i-v until the end of the row.
vii. Continue with the next step.
viii. Repeat steps i-vii until all pixels in the image have been processed.

However, there are still some unwanted noises present in bigger areas which cannot be filtered by the 7 × 7 pixels median filter. Therefore, the SRGAE algorithm is proposed because of its ability to eliminate larger noise pixels which are beyond the limits in the previous filtering step. Referring to the illustrated example in Figure 5, the region of interest (ROI) is counted as the total number of pixels in k-th area as in Equation (3).

$$size[k] = total\_pixels\_in\_ROI[k] \tag{3}$$

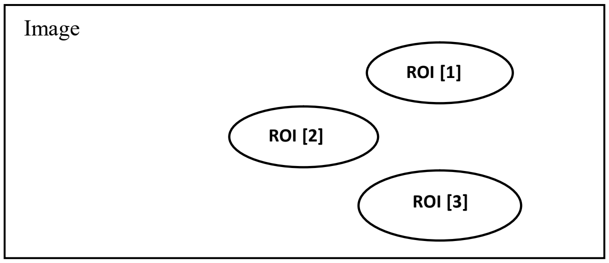


Fig.5. The region of interest (ROI) for the area extraction process.

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 (*TH*) = μ. The region growing process starts when the seed value is found. This first pixel corresponds 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 the *S*-plot during the segmentation process)

iii. The eight neighborhood pixels are revised and added to the region if they have a similar pixel value to the seed, Saturation pixel = μ. Increase size [*k*] = size [*k*] + 1 for each pixel that fulfills the growing condition.

iv. Search the size [*k*] with the conditions,

a) If size [*k*] has the pixel value more than 1000 pixels, the object will be preserved with the original color.

b) Otherwise, if size [*k*] has a value the pixel more than 1000 pixels, the object will be converted into white.

v. The growing process from the 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. Find the new seed point which does not belong to the previous region.

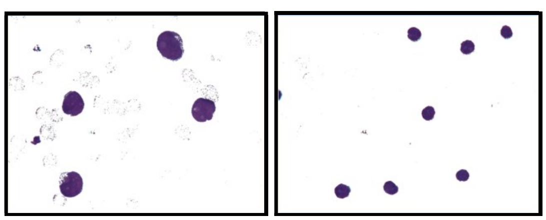
vii. If the seed point is found, increase *k* to *k* + 1 and size [*k* + 1] = 1. Go to step (iii); otherwise, go to step (viii).

viii. End.

The eight neighbourhood seed-based region growing algorithm of [9] has been modified for the suitability with acute leukemia images. The modification for this seed-based region growing algorithm regarding detection is based on the size of the acute leukemia cell itself. From the analysis of most of the images, the saturation pixel value (TH= μ) proposed in this study was 0.19 for acute leukemia images. Besides, through observations, it shows that size [k] for ALL and AML leukemia cells is more than 1000 pixels. Therefore, the pixel value below 1000 will not be considered as ALL and AML cell leukemia.

III. RESULTS AND DISCUSSIONS

The 100 images of acute myeloid leukemia (AML) and acute lymphoid leukemia (ALL) obtained from 42 slides of blood samples were gathered using the method proposed. From the results shown in Figure 6 (a) for AML images and Figure 6 (b) for ALL images, we have observed that the k-means clustering algorithm can cluster the region of interest for the acute leukemia blood images to the regions namely cytoplasm, nucleus and background. Unfortunately, as shown in Figure 6 (a) and 6 (b) the stray pixels still appear in the resultant image due to the pixel value similarity between the normal and abnormal cells in the region of interest.

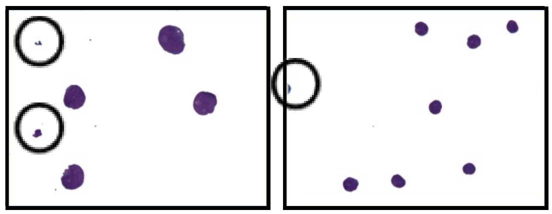


(a) AML image

(b) ALL image

Fig.6. The resultant images using S-component obtained from HSI color space and k-means algorithm (without the removal noise methods).

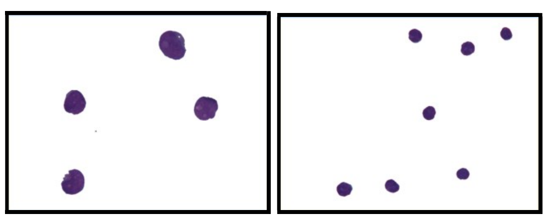
In spite of the fact that the segmentation using the k-means clustering algorithm can well segment the blast in the image, however some unwanted regions such as segmented RBCs can still be seen in the image. Therefore, in order to eliminate the unwanted regions from the image, these images are further processed with the 7×7 median filter and SRGAE algorithms. Figure 7 (a) and 7 (b) show the results of images after the segmented resultant images are filtered using 7×7 pixels median filter. By filtering the image, most of the small background pixels have been removed. However, some unwanted regions which are bigger in size (highlighted circle) can still be seen in the images. Thus, the SRGAE algorithm aims to give a better visibility of results.



(a) AML image (b) ALL image

Fig.7. Resultant image after using the 7x7 median filter algorithms

Figure 8(a) and 8(b) represent the resultant images after applying the SRGAE algorithm. By using the SRGAE algorithm, it can be seen that this algorithm can remove unwanted regions which area is less than 1000 pixels from the image, as compared to the resultant images that have been filtered using 7×7 pixels median filter. Therefore, a segmented acute leukaemia image was obtained by applying the fusion of 7×7 median filter and SRGAE algorithms.



(a) AML image (b) ALL image

Fig.8. The Resultant images using *S*-component, *k*-means with fusion 7x7 median filter and SRGAE algorithms

Making the effort to ascertain the effectiveness of the segmented image using the proposed method, the assessment is compared with the ground truth image by considering the resultant segmented image based on pixels' similarity.

TABLE 1. OVERALL PERFORMANCE FOR THE AUTOMATED SEGMENTATION PROCESS.

	Performance	Image		
		AML Image (50)	***ALL Image (50)***	***Average (50)***
S-component and k-means algorithm (without the removal noise methods)	Overall Accuracy (%)	96.82	97.05	96.93
	Sensitivity (%)	87.96	83.34	85.65
	Specificity (%)	97.49	98.14	97.81
S-component, k-means with the fusion 7x7 median filter and SRGAE algorithm	Overall Accuracy (%)	98.72	97.80	98.26
	Sensitivity (%)	87.08	81.00	84.04
	Specificity (%)	99.72	99.23	99.48

Table 1 shows the comparison of the segmentation performances for 100 segmented acute leukaemia images, with a total of 50 images for each type of AML and ALL. The results analysis was observed based on the performance without the noise removal or after being combined with the noise removal. This study has focused on the accuracy value that has been obtained. Based on the segmentation performance of 100 acute leukaemia images, the fusion method proposed with the noise removal algorithms has been proven better in producing segmented images with the overall accuracy 98.26%, 98.72% (AML accuracy) and 97.80% (ALL accuracy), respectively. However, the sensitivity value decreased after the noise removal process due to over segmentation and inability to process fine input from noise.

IV. CONCLUSION

This paper has proposed a segmentation based on the *S*-component based on HSI color space and *k*-means algorithm for segmenting the blast in AML and ALL images. Furthermore, to ensure the effectiveness of the detection of acute leukemia cells, this study also proposed fusion with 7x7 median filter and SRGAE algorithms for better visualization of resultant image. The performances of these proposed segmentation fusion with 7x7 median filter and SRGAE algorithms have been analyzed qualitatively and quantitatively. Qualitatively, it can be seen that these proposed fusion technique has proven better in obtaining the clean resultant segmented image. Quantitatively, based on the best overall accuracy value that has been obtained, a more convincing segmentation performance in segmenting the blast in acute leukaemia image has been achieved by using the fusion with the 7x7 median filter and SRGAE algorithms. Thus, it is proven that the fusion with the 7x7 median filter and SRGAE algorithms boast off a better performance for segmentation of acute leukaemia images. From the proposed method, features of AML and ALL images such as the shape and cells size are able to be revealed. These features are essential in distinguishing the difference between ALL and AML images for further analysis by the haematologists.

ACKNOWLEDGMENT

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

REFERENCES

[1] "FACTS AND STATISTICS." [Online]. Available: http://www.lls.org/facts-and-statistics/facts-and-statistics-overview.

[2] N. H. Harun, M. Y. Mashor, and R. Hassan, "Automated Blasts Segmentation Techniques Based on Clustering Algorithm for Acute Leukemia Blood Samples," J. Adv. Comput. Sci. Technol. Res., vol. 1, pp. 96–109, 2011.

[3] A. Khashman and E. Al-zgoul, "Image Segmentation of Blood Cells in Leukemia Patients," in Proceedings of the 4th WSEAS International Conference on Computer Engineering and Applications, 2009, pp. 104–109.

[4] S. Mohapatra, "Automated Leukemia Detection using Hausdorff Dimension in Blood Microscopic Images," in Emerging Trends in Robotics and Communication Technologies (INTERACT), 2010 International Conference on, 2010, pp. 64–68.

[5] F. H. a. Jabar, W. Ismail, R. A. Salam, and R. Hassan, "Image Segmentation Using an Adaptive Clustering Technique for the Detection of Acute Leukemia Blood Cells Images," in 2013 International Conference on Advanced Computer Science Applications and Technologies, 2013, pp. 373–378.

[6] N. H. Harun, M. Mashor, NR.Mokhtar, A. Salihah, R. H. A.N, R. A. A. Raof, and M. Osman, "Comparison of Acute Leukemia Image Segmentation using HSI and RGB Color

Space," in 10th International Conference on Information Science, Signal Processing and their Applications (ISSPA 2010), 2010, pp. 749–752.

[7] J. Macqueen, "Some methods for classification and analysis of multivariate observations," in Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1967, vol. 233, no. 233, pp. 281–297.

[8] R. A. A. Raof, R. B. Ahmad, M. Y. Mashor, and S. S. M. Noor, "Modified Median Filter for Noise Reduction in Ziehl-Neelsen Sputum Slide Images," in International Postgraduate Conference On Engineering (IPCE 2010), 2010.

[9] R. Adams and L. Bischof, "Seeded Region Growing," IEEE Trans. PATfERN Anal. Mach. Intell., vol. 16, no. 6, pp. 641–647, 1994.

How to cite this article:
Hazlyna H N, Aini H H, Fadzilah S, Nadia K, Mashor M Y, Hassan R. Fusion7×7 median filter and seeded region growing area extractionalgorithms for effectivedetection of acute leukemia based on blood images. J. Fundam. Appl. Sci., 2018, *10(3S), 726-737*.