

# Adaptive Hybrid Blood Cell Image Segmentation

T. Zalizam T. Muda<sup>1,\*</sup>, Rosalina Abdul Salam<sup>2</sup>, and Suzilah Ismail<sup>3</sup>

<sup>1</sup> School of Multimedia Technology and Communication, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia

<sup>2</sup> Faculty of Science and Technology, Universiti Sains Islam Malaysia, 71800 Nilai, Negeri Sembilan, Malaysia

<sup>3</sup> School of Quantitative Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia

**Abstract** – Image segmentation is an important phase in the image recognition system. In medical imaging such as blood cell analysis, it becomes a crucial step in quantitative cytophotometry. Currently, blood cell images become predominantly valuable in medical diagnostics tools. In this paper, we present an adaptive hybrid analysis based on selected segmentation algorithms. Three designates common approaches, that are Fuzzy c-means, K-means and Mean-shift are adapted. Blood cell images that are infected with malaria parasites at various stages were tested. The most suitable method will be selected based on the lowest number of regions. The selected approach will be enhanced by applying Median-cut algorithm to further expand the segmentation process. The proposed adaptive hybrid method has shown a significant improvement in the number of regions.

## 1 Introduction

Image segmentation is a process that divides an image into different regions. In medical study, segmentation of blood cell images has great potential area helping the expert to diagnose diseases. Generally, segmentation of blood cell images can be seen as a mechanism to assemble area of interest based on certain features such as colour, texture, and shape.

Manually segmentation methods are not suitable for a large amount of data, not efficient and consume time. To overcome this problem, an automated cell segmentation system will be a great tool for researchers and those involved in medical areas. The segmented images will be employed in the next process of object recognition and definitely help the experts to recognise the disease quickly. Various of common practical applications of image segmentation are image processing, medical imaging, computer vision, face recognition, digital libraries, image, and video retrieval, etc[4]. Generally, image segmentation methods can be categorised into five methods; pixel-based [1], region-based [2], edge-based, edge, and region-based hybrid, and clustering based segmentation [3].

The archaic segmentation methods range from thresholding to more complex techniques including the methods based on local features such as median, intensity variance, and intensity gradient.<sup>1</sup> The number of clusters which optimised this measure is the optimum number of the

cluster in the data set. Lately, soft computing components, for instance, Artificial Neural Networks[1,2], Fuzzy Logic[3, 4], and Genetic Algorithms had been employed in the area of medical image segmentation.

In this paper, the Means-shift, Fuzzy c-means and K-means methods are presented for evaluating the segmentation outcome of the blood cell images by comparing the number of regions.

## 2 Proposed Methodology

In this experiment, we used infected blood with malaria parasites and human blood is taken by using finger prick. This blood sample will be placed on a glass slide and we used Giemsa flooded stain method before we could observe under a microscope. The image taken under the microscope is approximately 40x10 magnifying from its normal size.

The image acquisition system consists of an inverted light microscope, a digital camera, and a computer. In this research there are two ways to get images: first, the object in the glass slide is captured on a typical digital camera that connected to the microscope. Second, the object is captured or recorded directly from a computer connected to both the camera and the microscope. The images used in this work are blood cells images on the glass slides which have two techniques thick and thin. The blood sample on the glass slide will leave dry on air under room temperature about 24 degrees Celsius.

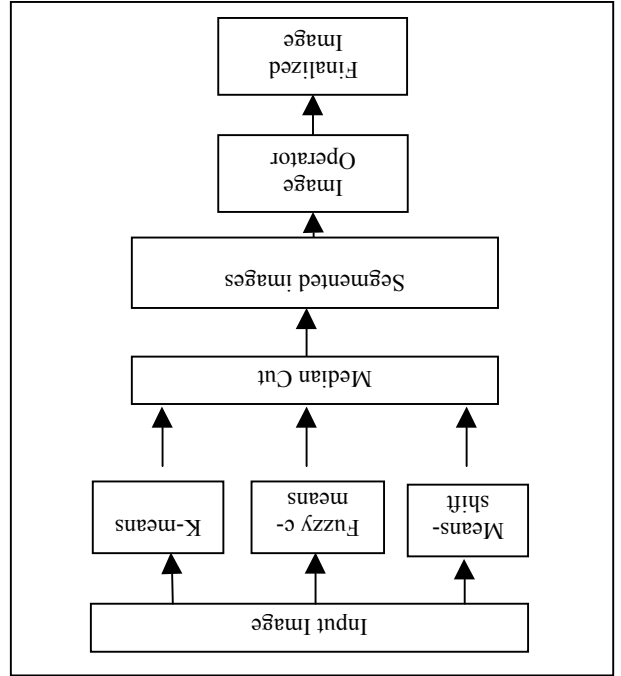
The segmentation methodologies for blood cell images are using the following algorithms: Mean-shift, Fuzzy c-means, and K-means. Next, we apply Median-cut method to improve the output as shown in **Fig. 1**. We do employ image

\* Corresponding author: [zalizam@uum.edu.my](mailto:zalizam@uum.edu.my)

operator to get the finalised image that contains only objects of interest.

## 2.1 Means Shift Algorithm

Means shift clustering algorithm is a data clustering algorithm commonly used in computer vision and image processing. Numerous studies have been reported of applying this method, some of them are [1-3]. This algorithm is discussed as follows. For each pixel of an image (having a spatial location and a particular color), the set of neighboring pixels (within a spatial radius and a defined color distance) is determined. For this set of neighbor pixels, the new spatial center (spatial mean) and the new color mean value is calculated. These calculated mean values will serve as the new center for the next iteration. The described procedure will be iterated until the spatial and the color (or greyscale) mean stops changing. At the end of the iteration, the final mean color will be assigned to the starting position of that iteration.



**Fig. 1.** Flows of the blood cell image segmentation of the proposed method

Given  $n$  data points  $x_1, \dots, x_n$  in the  $d$ -dimensional

space  $R^d$ , the kernel density estimator with kernel function

$$\hat{f}_n(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (1)$$

where the  $d$ -variate kernel  $K(x)$  is non-negative and integrates to one. A widely used class of kernels are the radially symmetric kernels

$$K(x) = c_{k,d} k(\|x\|_2) \quad (2)$$

where the function  $K(x)$  is called the profile of the kernel, and the normalization constant  $c_{k,d}$  is the normalization

constant. Estimation of the density gradient

$$m(x) = \frac{\sum_{i=1}^n x_i \mathcal{G}\left(\frac{h}{x - x_i}\right)}{\sum_{i=1}^n \mathcal{G}\left(\frac{h}{x - x_i}\right)} \quad (4)$$

mean shift

$c_{k,g}$  is the normalization coefficient. The final term is the  $G(x)$ .  $f, h, G(x)$  is the density estimation with the kernel  $G$ .

where  $g(x) = -k'(x)$  which can in turn be used as profile to define a kernel  $G(x)$ . The kernel  $K(x)$  is called the shadow of

$$\nabla f, h, k(x) = \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^n (x_i - x) \mathcal{G}\left(\frac{h}{x - x_i}\right) - x \left[ \frac{\sum_{i=1}^n x_i \mathcal{G}\left(\frac{h}{x - x_i}\right)}{\sum_{i=1}^n \mathcal{G}\left(\frac{h}{x - x_i}\right)} \right] = c_{k,g} f, h, G(x) \quad (3)$$

Fuzzy c-means (FCM) is a technique of clustering which permits one portion of data to belong to two or more clusters. This clustering algorithm has been extensively used and a well-known unsupervised clustering technique for pattern recognition developed by Dunn in the early 70's and improved by Bezdek ten years later[5]. FCM has been applied in the process of generating fuzzy rules from data. It has also been employed with success in segmented MR images with significant modification of the original algorithm[6]. FCM partition a collection of  $n$  vector  $X_i, i = 1, 2, 3, \dots, n$ , into  $C$  fuzzy group and finds the cluster centre in each group such that a cost function of the dissimilarity measure is minimised. FCM employs fuzzy partitioning such that a given data point can belong to several groups with the degree of belongingness specified by membership grades between 0 and 1.

Step 1: Initialize the cluster centres and the membership matrix  $U$  with random values between 0 and 1 such

that the following constraints are satisfied.

$$\sum_{i=1}^C u_{ij} = 1 \quad (5)$$

Step 2: Calculate  $C$  fuzzy cluster centers  $C_i, i = 1, 2, \dots, C$

Step 3: Compute the cost functions.

$$J_m(U, Y) = \sum_{i=1}^n \sum_{j=1}^C (u_{ij})^m E_f(x_i) \quad (6)$$

Where,  $Y = \{y_j | j \in [1, C]\}$ , is the set of centers of clusters.  $E_f(x_i)$ , is a dissimilarity measure (distance or

cost) between the sample  $x_k$ , and the center  $y_j$  of a specific cluster  $j$ .

$U=[u_{jk}]$ , is the  $c \times n$  fuzzy  $c$ -partition matrix, containing the membership values of all clusters.

$M \in (1, \infty)$ , is a control parameter of fuzziness.

Stop if either  $J_m$  below a certain tolerance or it is improved over previous iteration.

Step 4: Compute a new  $U$  and repeat the steps until an optimum result is obtained.

The performance depends on initial cluster centres, thereby allowing to run FCM several times, each starting with a different set of initial cluster centers.

### 2.3 K-Means

K-means clustering is one of the simplest unsupervised clustering techniques introduced by MacQueen[7]. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume  $k$  clusters) fixed a priori[16]. The main idea is to define  $k$  centroids, one for each cluster. These centroids should be placed in cunning way because of different location causes different results. So, the better choice is to place them as much possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early group is done. At this point, we need to re-calculate  $k$  new centroids as barycentres of the clusters resulting from the previous step. After we have these  $k$  new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop, we may notice that the  $k$  centroids change their location step by step until no more changes are done. In other words, centroids do not move anymore. Finally, this algorithm aims at minimizing the objective function, in this case, a squared error function. The objective function as follows

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2 \quad (7)$$

where  $\|x_i^{(j)} - c_j\|^2$  is a chosen distance measure between a data point  $x_i^{(j)}$  and the cluster center  $c_j$ , is an indicator of a distance of the data  $n$  data points from their respective cluster centers.

The algorithm is composed of the following steps:

Step 1: Place  $K$  points into the space represented by the objects that are being clustered. These points represent initial group centroids.

Step 2: Assign each object to the group that has the closest centroid.

Step 3: when all objects have been assigned, recalculate the positions of the  $K$  centroids.

Step 4: Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

### 2.4 Median Cut

Median cut algorithm is the color quantization algorithm approach for the image applying colour histogram to select

the median and separate the largest color cube along the median. The idea of the Median-cut algorithm is to make every color at the Color Quantization Table represent approximately the same amount of pixels of original images. The first step is to use the smallest cuboid bounding box, at the 3D RGB space, to enclose all the colors occurring in an image, and then divide the box by means of recursion and self-adaptive [8]. The second step is to make the color in the box sequencing along the component direction corresponding to the cuboids, and then make the box divided with the middle point of the direction. These two steps recur until the amount of the box comes up to the required amount of color, and then the average color value of each box will be calculated to build up the Color Quantization Table( i.e. the color palette). As a result, a clearer border between every region will be produced. With these methods, the cell contour can easily be obtained as well as removing unwanted edges. The results from this step will be used for the feature extraction stage.

## 3 Experimental Results

In this section, we present our experimental results on segmenting blood cell images based on the selected algorithm such as Mean-shift, Fuzzy c-means, and K-Means. For producing the result of the segmented images, we use ImageJ application tools for edge detection, JIU utilities for pre-processing and Java Advance Programming together with EDISON to segment those cell images by applying those techniques. Explanations of the results are based on the outcomes from analysis through these applications and number of regions. What we want to show here is the best approach and obtains the promising results. From the outcomes, we can compare them and find the best result for segmentation of the parasites in the red blood cells images. Finally, we used the median cut approach on the best algorithms to reduce the number of regions to the optimum level.

As shown in **Fig. 2**, a collection of color blood cell images of size 1280 x 960 taken by a light microscope at projection 40 x 10 = 400, and reduced to size 210 x 160. Five cell images have been used for this experiment and most of them have been infected with malaria parasites at various stages. The first column is the original test images and the second column is the segmentation results by using the Mean-shift algorithm.

In the third column shows the segmentation results produced by the FCM algorithm while the fourth column illustrates the outcomes of K-Means. The fifth column indicates the adaptive hybrid of the selected algorithm and the Median Cut. Lastly, the final column shows the regions of interest.

First, we look at the first cell image in the first column and row, in the figure shows the original image with 620 regions. After applying the Means-shift algorithm, the number of segmented regions is 546. While using K-means number of regions is reduced to 495. FCM remains unchanged.

In the second image, Means shift algorithm manages to reduce the region inside the image to 462 out of 544. K-Means reduces the regions inside the image about 352 and FCM method manages to get 409 regions. As we can see in

the figure, the region is well segmented while some minor features for the cell's contour are vanished.



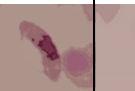


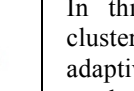
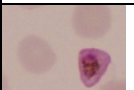
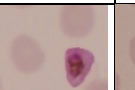
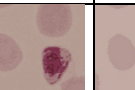
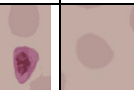

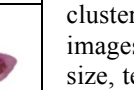
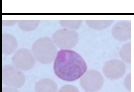

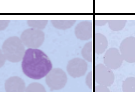


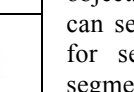

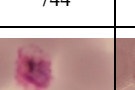



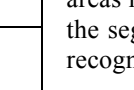
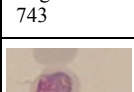
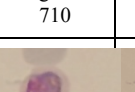



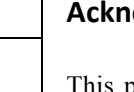
In the third image, the K-means method demonstrates the best result of segmenting the image regions by generating 653 out of 877. Means-shift algorithm produces 744 regions while FCM about 653 regions.

In the fourth image, the original image contains 743 regions. By using Means-shift, it produces 710 regions while K-means managed to construct 594 regions and 506 regions for FCM.

In the fifth image, the Means shift generates 507 out of 539 regions from the original image. In the K-Means algorithm, it produces 480 and FCM method shows the best result of 457 regions.

Based on the segmented images, most of the background features have not been removed completely. We can see that Means-shift failed to segment the blood cell properly compared to FCM and K-means technique. And again, in those images, both algorithms demonstrated the ability to perform well in segmenting cell images.

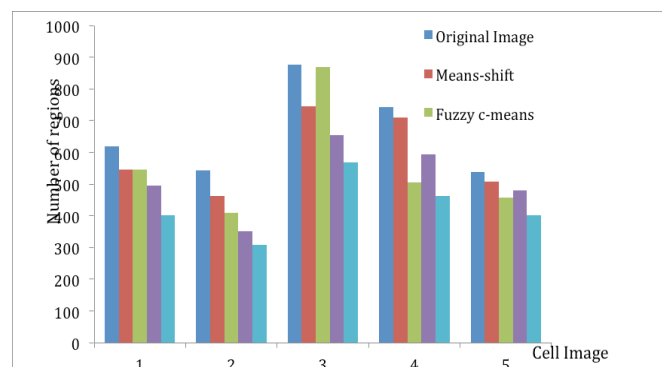
In the final stage, we applied the median-cut algorithm to reduce the number of regions to the optimum level. Later on, we applied image operator to get back the original form of the image and manage to get the area of interest of segmented images. In **Fig. 3**, showing a bar graph of the experimental results. The bar represents the number of regions of the images. As we can see that the adaptive hybrid algorithms with Median-cut manage to lower the number of regions making the combination perfect for segmented images.

Original Image	Mean-shift	Fuzzy C-means	K-means	K-means & Median-cut	Region of Interest Image
					
Region = 620	Region = 546	Region = 546	Region = 495	Region = 402	
					
Region = 544	Region = 462	Region = 409	Region = 352	Region = 307	
					
Region = 877	Region = 744	Region = 868	Region = 653	Region = 569	
					
Region = 743	Region = 710	Region = 506	Region = 594	Region = 516	
					
Region = 539	Region = 507	Region = 457	Region = 480	Region = 462	

**Fig. 2.** Test result of segmentation in Red Blood Cells Images

**Table 1.** Results

Original Image	Means-shift	Fuzzy c-means	K-means	Adaptive Hybrid
620	546	546	495	402
544	462	409	352	307
877	744	868	653	569
743	710	506	594	461
539	507	457	480	401



**Fig. 3.** Graph number of regions

## 4 Conclusions

In this paper, we present three popular methods of clustering algorithm to segment blood cell images. An adaptive hybrid method of the algorithms and Median-cut produced a better result of parasite segmentation in color images. We have demonstrated that the adaptive hybrid clustering algorithm is better to segment the blood cell images. Human blood cells are normally in the same shape, size, texture, and color. If one of them is changing in those features, means that cells have been infected with foreign objects such as parasites. According to the experiment, we can see that different techniques have different outcomes for segmented images. Based on the results of the segmentation we can identify the objects, for example in segmented blood cell images we can see that the infected areas have been grouped into several clusters. The result of the segmentation can be used for further classification and recognition.

## Acknowledgment

This paper presents work that is funded by the Universiti Utara Malaysia under University Grant. The authors cordially thank for this funding. Special thanks to Dr. Raden Sharmila Hisan from the Institute of Medical Research, IMR, Kuala Lumpur, Malaysia who provides the images and valuable input.

## References

1. Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, pp. 790-799, 1995.
2. Yong-mei Zhou, Sheng-yi Jiang, and Mei-lin Yin, "A Region-based Image Segmentation Method with Mean-Shift Clustering Algorithm," in *Fifth International Conference on Fuzzy Systems and Knowledge Discovery, 2008. FSKD '08 Jinan, China, 2008*.
3. Oncel Tuzel, Fatih Porikli, and Peter Meer, "Kernel Methods for Weakly Supervised Mean Shift Clustering," in *IEEE 12th International Conference on Computer Vision, ICCV 2009 Kyoto, Japan, 2009*.
4. R. O. Duda., P. E. Hart, and G. Stork, *Pattern Classification*. New York: John Wiley & Sons, 2000.
5. J. C. Dunn, "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters," *Journal of Cybernetics*, vol. 3, pp. 32-57, 1973.
6. Ping Wang and H. Wang, "A Modified FCM Algorithm for MRI Brain Image Segmentation," in *International Seminar on Future BioMedical Information Engineering, 2008. FBIE '08 Wuhan, China, 2008*.
7. J.B MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," in *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, California, 1967, pp. 281 – 297.
8. G. H. Geng, M. Q. Zhou, "Characters Analysis of Current Algorithms for Quantization." *Mini-micro Systems*, 1998, 19(9):46-49.