# Blood Cell Image Segmentation Using Unsupervised Clustering Techniques

**T. Zalizam T. Muda[a], Rosalina Abdul Salam[b]**

*[a]College of Arts and Sciences*
*Information Technology Building*
*Universiti Utara Malaysia, 06010 UUM Sintok, Kedah*
*Tel : 04-9284639, Fax : 04-9284753*
*E-mail : zalizam@uum.edu.my*

*[b]School of Computer Science,*
*Universiti Sains Malaysia, 11800 USM, Penang*
*Tel : 04-88282486, Fax : 04-88286557*
*E-mail : rosalina@cs.usm.my*

**ABSTRACT**

*In blood cell image analysis, segmentation is an indispensable step in quantitative cytophotometry. Blood cell images have become particularly useful in medical diagnostic tools for cases involving blood. The aim of our research is to develop an effective algorithm for segmentation of the blood cell images. In this paper, we present a framework of comparison cell images segmentation by using unsupervised clustering techniques with the purpose of acquiring the best method to segment the cell images. We use blood cell images infected with malaria parasites as cell images for our framework. Methods that involved in this comparison framework are Fuzzy C Means, K Means and Means Shift Analysis. The outcome from these methods will help to identify which technique or algorithm is the best for cell segmentation. Results from the segmented cell will be use for further classification and recognition.*

**Keywords**

*Segmentation, Cell Images, Fuzzy C Means, K Means, Means Shift Analysis*

## 1.0 INTRODUCTION

Computer and information technology has been widely used in almost all aspect of human life including agriculture, medicine, manufacturing, business and finance. Automated segmentation of images into anatomical tissues, fluids, bloods, and bone structures is an interesting field in medical analysis. Segmentation of blood cell images is a way to group area of interest based on features such as colour, texture and shape. Segmentation of blood cell images has great potential in clinical practice for the study of diagnostic pathology. In other words, helping the experts to detect disease. Automated blood cell image segmentation is the most critical part and yet complicated phases in blood cell image analysis. Operated-assisted segmentation methods are not practical for large amounts of data, and also inefficient. Fully automatic cell segmentation from blood cell images is a great instrument for researcher and those involve in clinical studies. Based on the segmented blood cell images, enable the experts to identify the infections rapidly.

Several studies on cell image segmentation have been proposed (Awasthi, Doolitle, Parulkar, and McNally, 1994; Berns and Berns, 1982; Gautheir, Levine, and Noble, 1991; Haralik and Shapiro, 1985; Sahoo and Saltani, 1988; Weska, 1978 ). There are many different methods of segmenting blood cell images, such as clustering, region growing and probabilistic methods. The classical segmentation methods range from thresholding to more complex techniques including the methods based on local features such as median, intensity variance and intensity gradient. The number of clusters which optimised this measure is the optimum number of cluster in the data set. Recently, soft computing tools, such as Artificial Neural Networks, Fuzzy logic, and Genetic Algorithms had been applied in the field of medical image segmentation (Pham et al., 1996). In this paper the K Means, Fuzzy C Means and Means Shift methods are presented for evaluating the segmentation outcome of the blood cell images.

## 2.0 Material and Methods

### 2.1 Biological 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 microscope. The image taken under microscope is approximately 40x10 magnifying from its normal size.

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. These experiments are developed in Institute of Medical Research, Kuala Lumpur (IMR) and we received the materials as Jpeg-Images.

### 2.2 Imaging System

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 on the microscope. Second, the object is captured or recorded directly from a computer connected to both the camera and the microscope.

## 3.0 PROPOSED METHODOLOGY

The segmentation methodologies for blood cell images are using Fuzzy C Means, K Means and Means Shift as shown in Figure 1.

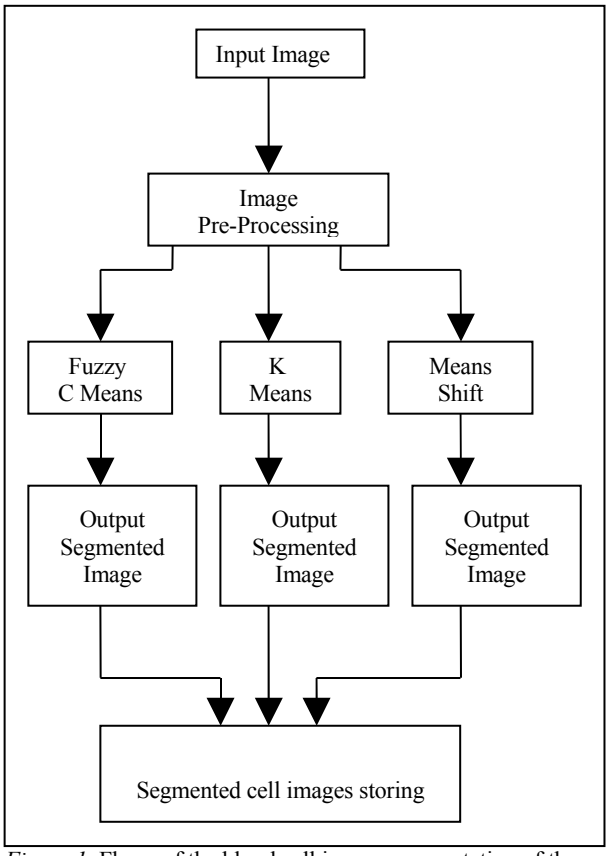



*Figure 1*: Flows of the blood cell image segmentation of the proposed method

### 3.1 Pre-processing Filtering

Intensity inhomogenities, caused by both imperfections in imaging devices and blood sample on glass slide using giemsa flooded stain reduced the quality of the cell images. Image enhancement by using mathematical morphology such as dilation, erosion and other available approaches may provide better quality of the images (Gonzalez & Woods, 2002).

### 3.2 Clustering Techniques

Clustering refers to identifying the number of subclasses of c cluster in a data universe X comprised of n data samples, and partitioning X into c clusters (2<=c<n). One of the simplest similarity measures is a distance between pair of feature vectors in the feature space. The clustering methods defines optimum partition through a global criterion function that measures the extent to which candidate partitions optimise a weighted sum of squared errors between data points and cluster centres in feature space.

#### 3.2.1 Fuzzy C Means Clustering Algorithm

The fuzzy c-means (FCM) clustering algorithm has been extensively used and a well known unsupervised clustering techniques for pattern recognition (Bezdek, 1973). FCM has been applied in the process of generating fuzzy rules from data. It has also been employed with success in the soft segmentation of MR images and for the estimation of partial volumes (Jiang and Yang, 2003).

FCM partition a collection of n vector X*i*, *I* = 1,2,3,…….,*n.* into C fuzzy group and finds the cluster centre in each group such that a cost function of 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.

The FCM algorithm is simply an iterative procedure. In a batch mode operation FCM determines the cluster centres C*i* and the membership matrix U using the following steps (Kannan, 2005).

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.

$$\sum_{i=1}^{c} uij = 1 \qquad (1)$$

Step 2: Calculate C fuzzy cluster centres C*i*, i = 1, 2,…….,C
Step 3: Compute the cost functions.

$$J_m(U,Y) = \sum_{k=1}^{n} \sum_{j=1}^{n} (u_{jk})^m E_j(x_k) \qquad (2)$$

Where, $Y = \{y_j \mid j \in [1,c]\}$, is the set of centres of clusters.

$E_j(x_k)$, is a dissimilarity measure (distance or cost) between the sample $x_k$, and the centre $y_j$ of a specific cluster *j*.

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

M ε (1, ∞), 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 different set of initial cluster centres.

### 3.2.2 K-Means Clustering Algorithm

The K-means clustering is one of the simplest unsupervised classification techniques (MacQueen, 1967). It is also one of the most popular unsupervised learning algorithms due to its simplicity. The procedure follows a simple and easy way to classify a given data set through s certain number of clusters (assume k clusters) fixed a priori. 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 groupage 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 any more.

Finally, this algorithm aims at minimising objective function, in this case a squared error function. The objective function

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \| x_i^{(j)} - c_j \|^2 \qquad (3)$$

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

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 closet 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.

Although it can be proved that the procedure will always terminate, the k-means algorithm does not necessarily find the most optimal configuration, corresponding to the global objective function minimum. The algorithm is also significantly sensitive to the initial randomly selected cluster centres. The k-means algorithm can be run multiple times to reduce this effect. K-means is a simple algorithm that has been adapted to many problem domains.

### 3.1.1 Means Shift Clustering Algorithm

The means shift clustering algorithm is a data clustering algorithm commonly used in computer vision and image processing. Numerous studies have been reported for this method, some of them are (Cheng, 1995; DeMenthon, 2002; Wand & Jones, 1995; Scott, 1992; Comaniciu & Meer, 2002; Tao, Jin & Zhang, 2007).

This algorithm is discussed as follows. For each pixel of an image (having a spatial location and a particular colour), the set of neighbouring pixels (within a spatial radius and a defined colour distance) is determined. For this set of neighbour pixels, the new spatial centre (spatial mean) and the new colour mean value are calculated. These calculated mean values will serve as the new centre for the next iteration. The described procedure will be iterated until the spatial and the colour (or greyscale) mean stop changing. At the end of the iteration, the final mean colour will be assigned to the starting position of that iteration.

Given *n* data points $x_1 \ldots\ldots\ldots x_n$ in the *d*-dimensional space $R^d$, the kernel density estimator with kernel function *K(x)* and a window bandwidth *h*, is given by (Duda, Hart & Stork, 2000; Scott, 1992, Wand & Jones, 1995).

$$\hat{f}_n(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left( \frac{x - x_i}{h} \right) \qquad (4)$$

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

$$K(x) = c_{k,d} k(\| x \|^2) \qquad (5)$$

where the function *K(x)* is called the profile of the kernel, and normalization constant $c_{k,d}k$ is the normalisation constant. Estimation of the density gradient

$$\nabla \hat{f}\, h,k(x) = \frac{2c_{k,d}}{nh^{d+2}} \sum_{i=1}^{n} (x_i - x) g\left( \| \frac{x - x_i}{h} \|^2 \right)$$

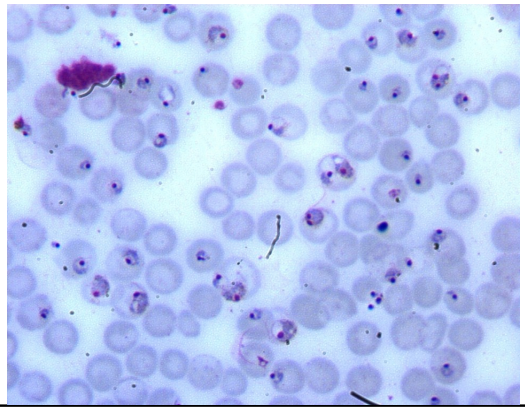
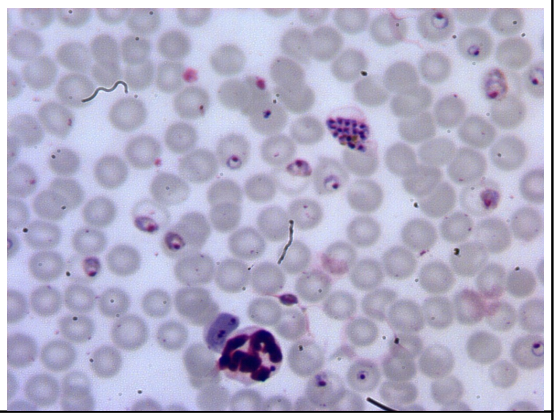
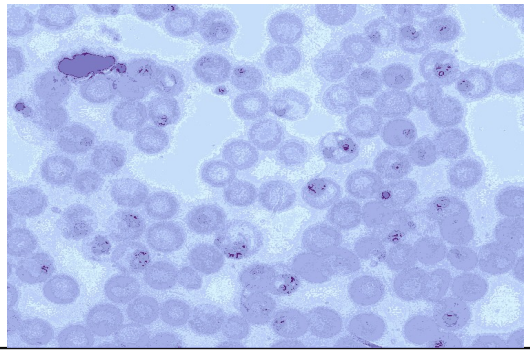
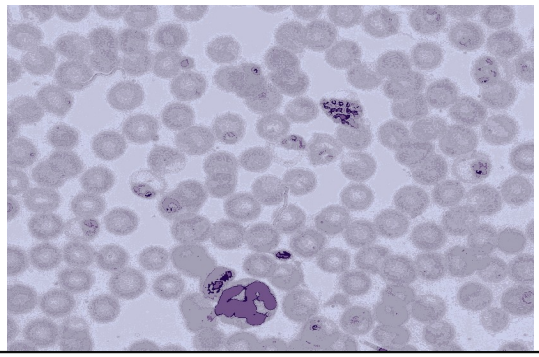
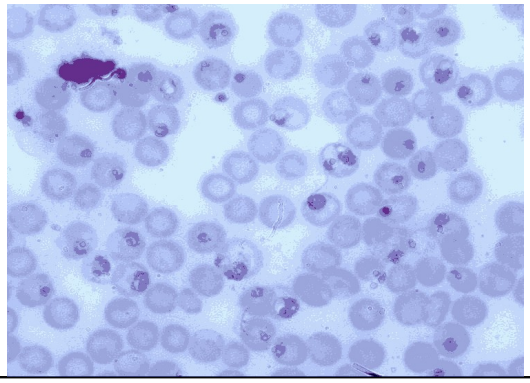
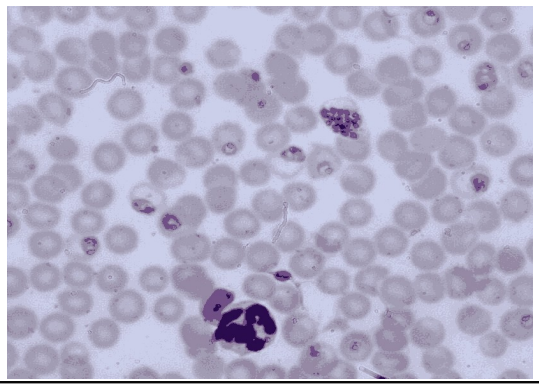
$$= c_{k,g}\, \hat{f}\, h, G(x) \left[ \frac{\sum_{i=1}^{n} x_i g\left( \| \frac{x - x_i}{h} \|^2 \right)}{\sum_{i=1}^{n} g\left( \| \frac{x - x_i}{h} \|^2 \right)} - x \right]$$

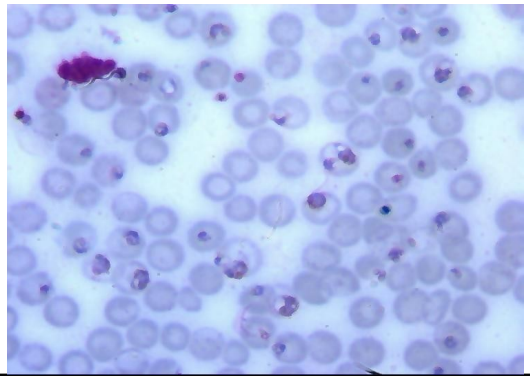
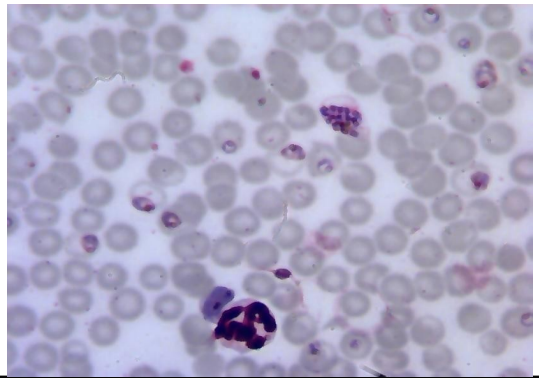
where *g(x) = -k'(x)* which can in turn be used as profile to define a kernel *G(x)*. The kernel *K(x)* is called the shadow of *G(x)* (Cheng, 1995). $\hat{f}\, h, G(x)$ is the density estimation with the kernel *G*. $c_{k,g}$ is the normalisation coefficient. The final term is the mean shift

$$m(x) = \frac{\sum_{i=1}^{n} x_i g\left( \| \frac{x - x_i}{h} \|^2 \right)}{\sum_{i=1}^{n} g\left( \| \frac{x - x_i}{h} \|^2 \right)} - x \qquad (7)$$

The algorithm is precise as follows(DeMenthon, 2002):
Step 1: computation of the mean shift $m(x^k)$
Step 2: Updating the current position $x^{k+1} = x^k + m(x^k)$,
until reaching the stationary point which is the candidate cluster centre.

Table 1: *Blood Cell Images before and after Segmentation*

| | | |
|---|---|---|
| Original Image |  |  |
| Fuzzy C-Means |  |  |
| K-Means |  |  |

| | | |
|---|---|---|
| Means Shift |  |  |

## 4.0 EXPERIMENTAL RESULTS

In this section, we present our experimental results on segmenting cell images, and compare among Fuzzy C-Means, K-Means and Means Shift clustering algorithms. For producing the result of the segmented images, we use ImageJ and Java Advance Imaging(JAI) programming to segment those cell images by applying those techniques. Explanations of the results are based on the outcomes from analysis through the ImageJ and JAI programming. What we want to show here are the basic approaches and general results. From the outcomes we can compare them and find the best result for cell images.

Table 1 shows two blood cell images of size 960 x 722 pixels taken by a light microscope at projection 40 x 10 = 400, and 24 bits depth. Two samples of cell images have been used for this experiment and both of them have been infected with malaria parasites, one sample with various stages and the second sample at trophozoite and schizont stage.

First we look at the original cell images, in the Table 1 shows blood cell image infected with malaria parasites at the various stages. In the second image, blood cells have been infected with malaria parasites at trophozoite stage and schizont stage. We applied those three techniques by using ImageJ and Java Advance Imaging programming to segment the blood cells. The segmented images by using FCM technique are not segmented very well compared to K-Means. But, Means Shift technique produced a better segmented cell images in the first image.

In the second image, we can see that FCM failed to segment the blood cell properly compared to K-Means technique. And again, in this second image, Means Shift demonstrated the ability to perform very well in segmenting cell images.

## 5.0 CONCLUSION

In this paper we present a general method of unsupervised clustering to segment blood cell images. We have demonstrate that means shift clustering algorithm is better to segment the cell images. Human blood cells are normally in the same shape, size, texture and colour. If one of them is changing in those features, means that cells has 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 area have been grouped into several clusters. The result of the segmentation can be used for further classification and recognition.


## ACKNOWLEDGEMENT

The authors wish to thank Dr. Rohani Ahmad from Institute of Medical Research, IMR, Kuala Lumpur, who provide images which are used to this experiment.


## REFERENCES


Awasthi, V., Doolitle, w., Parulkar, G., & McNally, J. G. (1994). *Cell Tracking using a Distributed Algorithm for 3D Image Segmentation*. Bio Imaging, USA.

Berns, G. S., & Berns, M. W. (1982). *Computer-based tracking of living cells.* Experimental Cell Research, vol 142, page 103-109, USA.

Bezdek, J. C. (1973). *Fuzzy mathematics in pattern classification*. Ph.D diss., Cornell Univ., Ithaca, New York, USA.

Cheng, Y. (1995). *Mean shift, mode seeking, and clustering.* IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 17, no. 8, pp. 790-799.

DeMenthon, D. (2002). *Spatio-temporal segmentation of video by hierarchical mean shift analysis.* Statistical Methods in Video Processing Workshop, Copenhagen, Denmark.

Duda, R. O., Hart, P. E., & Stork, G. (2000). *Pattern Classification,* John Wiley & Sons, New York, USA.

Gauthier, D., Levine, M. D., & Noble, P. B. (1991). *Principles of object detection for an automated cell tracking systems*. Image Analysis in Biology(D.P. Hader, ed.) CRC Press, Florida, USA.

Gonzalez, R. C., & Woods, R. E. (2002). *Digital Image Processing.* Prentice Hall, New Jersey, USA.

Jiang, L., & Yang, W. (2003). *A modified Fuzzy C-Means Algorithm for Segmentation of Magnetic Resonance Images.* Proc. VIIth Digital Image Computing: Techniques and Applications, Sydney, Australia.

Kannan, S. R. (2005). *A New Clustering Algorithm for Segmentation of Magnetic Resonance Images Using Fuzzy C-Means and Computer Programming.* ICGST International Journal on Graphics, Vision and Image Processing, vol 2.

MacQueen, J. B. (1967). *Some Methods for Classification and Analysis of Multivariate Observations,* Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 1:281 – 297, Berkeley, California, USA.

Prince, P. D., Dagher A. P., & Xu, C., (1996) *An automated technique for statistical characterisation of brain tissues in magnetic resonance Imaging.* Technical Report JHU/ECE 96-13

Tao, W., Jin, H., & Zhang, Y. (2007) *Colour Image Segmentation Based on Means Shift & Normalized Cuts,* IEEE Trans. On Systems, Man, & Cybernetics Vol:37 no. 5, pp. 1083-4419.

Wand, M. P., & Jones M. C. (1995). *Kernel Smoothing,* Chapman & Hall, London, England.