White Blood Cell Segmentation Techniques in Microscopic Images for Leukemia Detection

Biji G[1], Dr.S.Hariharan[2]

[1.] Assistant Professor, Dept of EEE, College of Engineering, Thiruvananthapuram
[2.] Professor, Dept of EEE, College of Engineering, Thiruvananthapuram

Abstract: *Leukemia is a group of hematological disease which usually affects blood, bone marrow, lymph nodes which characterized by overproduction of abnormal white cells which are unable to fight infection. Careful microscopic examination of stained blood smear or bone marrow aspirate is the only way to effective diagnosis of leukemia. Early detection of the disease is necessary for proper treatment management. Recently many scientists have performed tremendous research in helping the hematologists in the issue of segmenting the blood cells in the early of prognosis. This paper aims to segment the blood cell images of patients suffering from acute leukemia using a Standard K-Means (SKM) clustering, Fuzzy C-Means (FCM) and Adaptive K-Means (AKM) clustering algorithm. The integrated clustering techniques have produced comprehensive output images with minimal filtering process to remove the background scene.*

Keywords : *Image segmentation, clustering, Standard K-means (SKM), Fuzzy C Means (FCM), Adaptive K-Means (AKM), acute leukemia cells*

I. Introduction

Complete Blood Count (CBC) is an important and primary blood test that is required by many physicians to get an overall view about patients with many diseases. The blood examination can indicate many diseases such as cancer, HIV/AIDS, diabetes, anemia, and coronary heart disease which are popular diseases [1].No doubt that the manual microscopic evaluation is essential when there a suspicious abnormality found in the blood smear [2]. The CBC gives a measure of the concentration of the Red Blood Cells (RBC), White Blood Cells (WBC), and platelets. However, the manual microscopic inspection is time consuming and requires technical knowledge.

Automated methods to provide these counting which are required to facilitate this process and to reduce the work load on scientists and lab technicians. One of the most significant benefits of the automation of visual sample inspection is to aid pathologists to recognize abnormalities in blood samples competently, accurately, and faster.

There are five main processes composing the automation which are: image acquisition, image preprocessing, image segmentation, image post-processing, and image analysis. Blood cells segmentation is used for the detection and classification of the Sickle Cell Anemia [3]. On the other hand, WBC segmentation is a prerequisite for the detection of Leukemia in human blood samples as Leukemia's main symptom is WBC with giant nuclei [3]. Leukemia is a cancer that begins in the bone marrow and can be treated when it is detected in its early stages.

Digital image processing becomes more significant in biomedical diagnosis and health care due to the increasing usage of it in medical diagnostics. Image segmentation is among of the image processing methods that has been, and yet still a relevant area in digital image processing due to its wide spread usage and applications. Image segmentation is a complex process which is commonly used for images segmentation in medical analysis. The goal of image segmentation is to partition of an image into a set of image regions, which is corresponding to certain properties or characteristics, for object identification, classification and processing [3].

Image segmentation can be categorized into two types, supervised and unsupervised. Practically in hematologists' lab, the most common method for evaluating the effectiveness of a segmentation method is a human supervision and comparison with the segmented results for separate segmentation algorithms. However, this process is a tedious and inherently limits the depth of evaluation to a relatively small number of segmentation comparisons over a predetermined set of images [2].An unsupervised method provides more effective and accurate results of the segmented images. This method are fully automated and uses different kind of automated algorithm such as region or boundary based [3], edge based and thresholding [4]. It has become a great attention for clinical researcher especially for hematologist to analyze the human blood and classify the area of interest such as texture, shape or color.

II. Literature Study

Many systems are reported for the classification of WBC into different categories related to the type of Leukemia in the literature. Techniques are based on feature extraction and then applying a classifier.

Supported Vector Machine (SVM) [1][5] is used with multispectral imaging. Fractal dimension and shape features are used with SVM in the work proposed in [6].

The Artificial Neural Network (ANN) is used in [7] to classify the Leukemia cells in addition to a set of features (size, shape, and color based features). Geometric features are also used for the detection and classification of WBC in [8].

Arithmetic operation and minimum filter are combined with global threshold techniques to segment cell nucleus in [9]. In automated imaging technique, the existing techniques for image recognition and visualization [5], and object based image compression [6] is highly depend on the segmentation results. Image segmentation is typically used to locate objects and boundaries in images.

Recently many segmentation tools have been proposed and develop to produce better segmentation on medical images such as clustering [7], active contour [8], thresholding [9] and region-based [10,11]. Thresholding is among the initial techniques developed for image segmentation due to its simplicity and intuitive properties that provides image thresholding a central position [9].

The active contour model known as snake model introduced by Kass in 1988 using the deformable model to extract features, had also been used to segment white blood cells in bone marrow [8]. Model based algorithm is proposed to solve the cluster-separation problem in leukocytes cluster using moving interface models and model-based combinatorial optimization scheme [10].

A combination technique is used to segment white blood cells on color space images using feature space clustering techniques, scale-space filtering for nucleus extraction, and watershed clustering for cytoplasm extraction [11]. Many automated segmentation methods are based on two basic properties of the pixels related to their local neighborhood, which is boundary-based methods, and region-based methods. Unfortunately, both techniques, boundary-based and region-based often fail to produce accurate segmentation results [9, 10]. Recently there are some approaches that have been developed to perform automated detection for leukemia cells which utilized Otsu method combine with artificial intelligence which includes Cellular Automata and heuristic search [10] and thresholding technique [12].

III. Proposed Work

The ultimate goal of blood cell segmentation is to extract the cells from complicated background and to segment every cell into morphological components such as nucleus, cytoplasm, and some others. In recent years there has been a growing interest in developing effective methods for image clustering. Clustering techniques classifies the pixels with same characteristics into one cluster, thus forming different clusters according to coherence between pixels in a cluster. Image clustering is a means for description of image content. Because of its simplicity and efficiency, clustering approaches were one of the first techniques used for the segmentation of (textured) natural images [10].

Image clustering provides an efficient retrieval algorithms and the creation of a user-friendly interface to the database. The quality of clustering depends on the method and implementation measure which able to discover hidden patterns. A high quality clusters consist of high intra-class similarity and low inter-class similarity. K-Means clustering algorithm is also one of the recent techniques that have been proposed in the area of blood cells analysis. K-Means algorithm is an unsupervised clustering algorithm that classifies the input data points into multiple classes based on their minimum distance. In medical imaging, K-Means clustering has been proven to give good segmentation image performance due its performance in clustering massive datasets [11]. The final clustering result of the K-Means clustering algorithm is highly dependable on the correctness of the initial centroids, which are selected randomly. In 2011, Filipczuk used a thresholding method prior k-means algorithm to distinguish nuclei from red blood cells and other objects. Recent paper suggested hybrid K-Means merging with median-cut algorithms for blood cell image segmentation to produce better segmented image of the blood cells. In 2011, Moving K-Means is used to segment the blast cell in acute leukemia blood samples.

The clustering was performed after applying the threshold method using saturation component formula [9]. Latest work of using an adaptive K-Means which imposed an efficient way of choosing the initial centroid during the initialization step for better segmentation of the blast cells. This method shows that the proposed method would yield better segmentation as compared to standard K-Means [11].

This paper is to apply the adaptive K-Means method integrated with mean shift algorithm to remove the background noise for further computer processing on pattern recognition and classification. Initialization step for better segmentation of the blast cells. This method shows that the proposed method would yield better segmentation as compared to standard K-Means [11].This paper is to apply the adaptive K-Means method to remove the background noise for further computer processing on pattern recognition and classification.

IV. Methodology

Clustering is partitioning a group of data points into a set of non-overlapped clusters. The resultant image of these techniques is compared and analyzed. The resultant image using the best performance technique among the three is applied to the mean shift algorithm to remove the background noise for better segmentation without applying any filtering process. All processes have been applied with canny edge detector to highlight the cell boundaries for better viewing.

The datasets used in this study consist of six samples taken from patients suffering from acute leukemia type AML. The size of the image is 1280 by 960 pixels

Three clustering technique has been used to compare the performance of the image segmentation, which are classic K-Means (CKM), Fuzzy-C-Means (FCM) and adaptive K-Means(AKM). The segmentation is compared prior applying the second clustering technique to remove the background scene.

Standard K-Means

K-means (MacQueen, 1967) is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The main idea of the learning process is based on the center based clustering method. The method is divided into 2 phases: first phase is defining the k-centroids, one for each cluster which contribute to the initial steps for the whole process. Then each point from data set will be mapped to the nearest centroids until all points are assigned based on minimum Euclidean Distance. Second phase observed the updates each of the points.

Given a set of observations $(x_1, x_2, \ldots, x_n)$, where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k sets $(k \leq n)$ S = $\{S_1, S_2, \ldots, S_k\}$, so as to minimize the within cluster sum of squares (WCSS)

The k-centroids need to be recalculating as new k-centroids and new mapping need to be done between the points and the new k-centroids. This process will give changes in k-centroids location step by step until the k-centroids do not move anymore. According to D. Malyszko [10].

The general steps of the center based clustering are:
1) Initialize step with initial k centroids value randomly.
2) For each data point x_j, compute its minimum distance with each center μ_i.
3) For each center μ_i, recomputed the new center from all data points x_j belong to this cluster.
4) Repeat steps 2 and 3 until convergence

The algorithm is based on the following objective function:

$$W(C) = \sum_{i=1}^{N} \sum_{X_j \in S_i} \left\| X_j - \mu_i \right\|^2 \quad \text{--------} \quad (1)$$

where μ_i is the mean of point S_i for ith cluster

based on the assignment C. The interest is to minimize the sum of square distance within the cluster and such assignments have to map each point to its nearest centroids.

Fuzzy C-Means

Fuzzy C-Means is a clustering method which allows single data belong to more than one clusters. This method (Dunn, 1973) is a pattern recognition based on minimization of the following objective function given as Equation (2) below:

The FCM algorithm attempts to partition a finite collection of n elements X=$\{X_1, X_2, \ldots X_n\}$ into a collection of c fuzzy clusters with respect to some given criterion. Given a finite set of data, the algorithm returns a list of C cluster centres C=$\{C_1, C_2, \ldots C_n\}$ and a partition matrix Like the K-means clustering, the FCM aims to minimize an objective function:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{c} u_{ij}{}^m \left\| x_i - c_j \right\|^2 \text{---------------}(2)$$

where:

$$u_{ij} = \cfrac{1}{\sum_{k=1}^{c} \left(\cfrac{\left\| x_i - c_j \right\|}{\left\| x_i - c_k \right\|} \right)^{\frac{2}{m-1}}} \text{-------------------}(3)$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{\,m}.x_i}{\sum_{i=1}^{N} u_{ij}^{\,m}} \text{-----------------------(4)}$$

where each element u_{ij} tells the degree to which element x_i belongs to cluster C_j .

This differs from the k-means objective function by the addition of the membership values u_{ij} and the fuzzifier m, with m≥1. The fuzzifier m determines the level of cluster fuzziness. A large m results in smaller memberships ||*|| is any norm expressing the similarity between any measured data and the center.

The algorithm is composed of the following steps:
1. Initialize U=[u_{ij}] matrix, U(0)
2. At k-step: calculate the centers vectors C(k)=[cj] with U(k)

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^{\,m} \cdot x_i}{\sum_{i=1}^{N} u_{ij}^{\,m}}$$

3. Update   U(k) ,   U(k+1)

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}}$$

4. If $\| U(k+1) - U(k)\| < \varepsilon$ then STOP; otherwise return to step 2.

Adaptive K-Means

This adaptive K-Means method which proposed recently was originally based on the standard K-Means algorithm for better segmentation [13]. During initialization step, the placement of the k-centroids is very crucial because different location will give different result. The standard K-Means method uses the randomly choose k-centroids which leads to accuracy degradation in segmentation. This may cause inconsistent result in the image segmentation.

Despite of using the normal randomly choose initial k-centroids, this adaptive method manipulates the local minimum and maximum values based on the RGB colour space during the initialization step. The enhanced initialization method returns a two-element array with minimum and maximum RGB values from the whole pixel area.

The operator computes the maximum and minimum pixel values for each band of a rendered image within the region of interest. The adaptive method is an iteration-based clustering that produces an optimal value of initial k-centroids by minimizing the objective function. The initial k-centroids can be obtained by using the following objective function:

$$\alpha = \sum_{i=1}^{c}\sum_{j=1}^{n} \partial\left(\mu_{\min}, \mu_{\max}\right)\text{--------------------(5)}$$

Where α is the Euclidean distance between minimum RGB value μmin and maximum RGB value μmax, n represent the number of image pixels and c is the total number of cluster. The whole initialization process is presented in Fig.1as shown below.
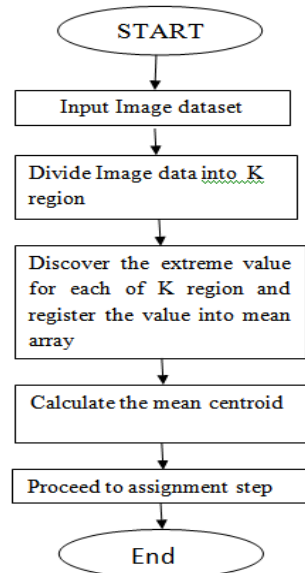
Fig 1. Method of adaptive K-Means during initialization step

V. Experimental Results

This section is to evaluate the performance of proposed method through human supervision. Fig.(a),(b),(c) shows the real sample of microscopic images with resolutions of 1280 X 960. The resultant images after applying the initial clustering technique using SKM, FCM and AKM are shown in Fig. 2 - 4(a), (b) and (c) respectively. From the result shows that the adaptive K Means algorithm give better segmentation result as compared to SKM and FCM.

In this experiment we make a comparison using k-centroids value of k=6 with 10 iterations due to the convergence occurs with less than 10 For fuzziness parameter, selected the degree of fuzziness, m = 2 and threshold, = 0.01
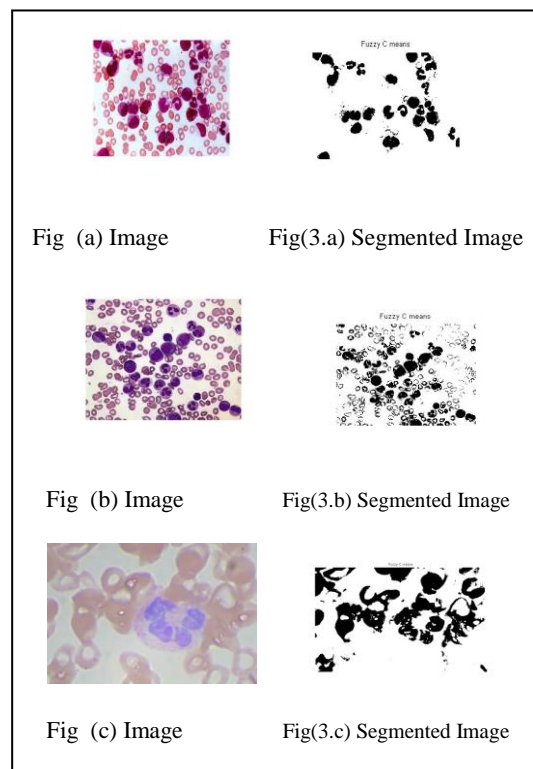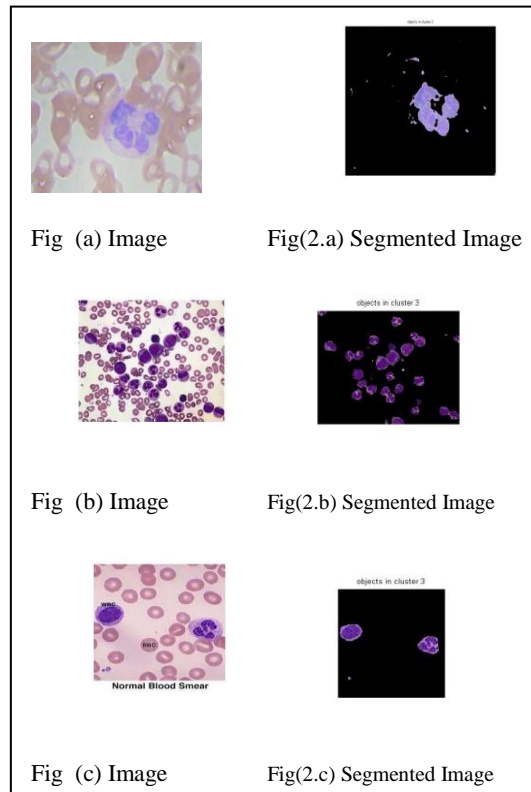

Fig 2: Image and segmented Image using Standard K Means

Fig (a) Image   Fig(2.a) Segmented Image

Fig (b) Image   Fig(2.b) Segmented Image

Fig (c) Image   Fig(2.c) Segmented Image

Fig 3: Image and segmented Image using fuzzy C Means

Fig (a) Image   Fig(4.a) Segmented Image

Fig (b) Image   Fig(4.b) Segmented Image

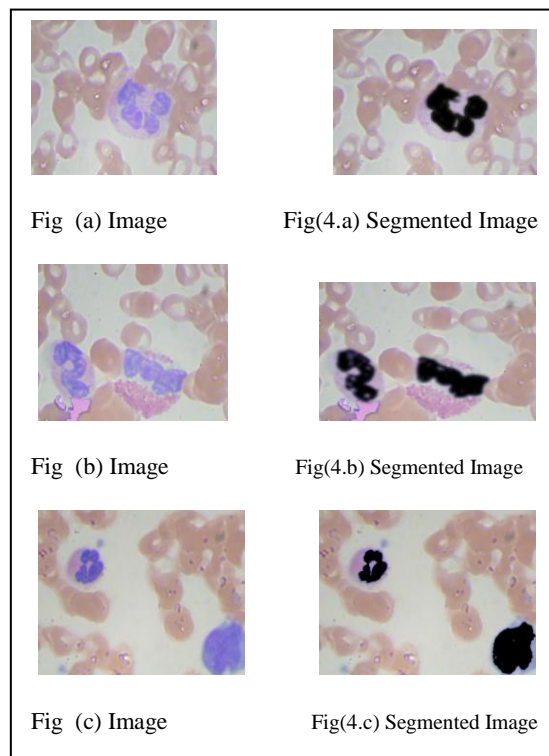Fig (c) Image   Fig(4.c) Segmented Image

Fig 4: Image and segmented Image using Adaptive K Means

VI. Conclusion

This paper presents Standard K Means(SKM), Fuzzy C Means(FCM) and adaptive K-Means(AKM) clustering algorithm for segmenting acute leukemia blood cells images without applying any filtering method to remove the background scene. Experimental results shows better performance of segmentation images using the proposed adaptive system as compare to standard K-Means and Fuzzy C-Means. For future works, this research will continue to perform image recognition and classification for acute leukemia blood cells image.

Refernces

[1]   P. Filipczuk, M. Kowal, A. Obuchowicz, "Automatic Breast Cancer Diagnosis Based on K-Means Clustering and Adaptive Thresholding Hybrid Segmentation", Image Processing and Communications Challenges. 3, 295-302, 2011.

[2]   T. Z. Muda, R. A. Salam, "Blood cell image segmentation using hybrid K-Means and median-cut algorithms" IEEE International Conference on Control System, Computing and Engineering (ICCSCE), 2011.

[3]   A. S. Samma and R. A. Salam, "Adaptation of K-Means Algorithm for Image Segmentation", International Journal of Information and Communication Engineering 5, 2009.

[4]   N. H. Harun, M. Y. Mashor, and R. Hassan, "Automated Blasts Segmentation Techniques Based on Clustering Algorithm for Acute Leukaemia Blood Samples", Journal of Advanced Computer Science and Technology Research 1. 96-109, 2011.

[5]   F. H . A. Jabar, W. Ismail, K. A Rahim, and R. Hassan,"K-Means Clustering For Acute Leukemia Blood Cells Image", Proceedings of the International Conference on Soft Computing and Computational Mathematics (ICSCCM), 2013.

[6]   H. Zhang, J. E. Fritts, S.A. Goldman, "Image Segmentation Evaluation: A survey of unsupervised methods", IEEE proceedings on Computer Vision and Image Understanding, Volume 110, Issue 2, Pages 260- 280, May 2008.

[7]   J. Freixenet, X. Munoz, D. Raba, J. Marti, X. Cufi, "Yet another survey on image segmentation: Region and boundary information integration." ,Notes in Computer Science Volume 2352, pp 408-422, 2002.

[8]   M. Kunt, M. Benard, and R. Leonardi, " Recent results in high compression image coding" IEEE Transactions Circuits System, vol.34, 1987, pp.1306–1336.

[9]   F. H . A. Jabar, R. Hassan," Image Segmentation Using an Adaptive Clustering Technique for the Detection of Acute Leukemia Blood Cells Images", Proceedings of the International Conference on Advanced Computer Science Applications and Technologies 2013.

[10]   M. Veluchamy, K. Perumal, and T. Ponuchamy, "Feature extraction and classification of blood cells using artificial neural network," American Journal of Applied Science, vol. 9, no. 5, pp. 615–619, 2012

[11]   F. Kasimn, A. S. Prabuwono, and A. Abdullah, "Detection of leukemia in human blood sample based on microscopic images: a study," Journal of theoretical and Applied Information Technology, vol. 46, no. 2, 2012.

[12]   N. H. Harun, M. Y. Mashor, and R. Hassan, "Automated Blasts Segmentation Techniques Based on Clustering Algorithm for Acute Leukaemia Blood Samples", Journal of Advanced Computer Science and Technology Research 1. 96-109, 2011.

[13]   D. Comaniciu and P. Meer, " Mean shift: A robust approach toward feature space analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence,24:603–619, 2002.