ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 1, July 2013

Fuzzy Clustering Based Image Segmentation of Pap smear Images of Cervical Cancer Cell Using FCM Algorithm

Jyotismita Talukdar, Chandan Kr. Nath, P.H.Talukdar
Asian Institute of Technology, Thailand, Gauhati University, Assam

Abstract*— The FCM clustering algorithm is advantageous over hard clustering methods. The FCM clustering technique, applied on colour image, preserves the colour and a chance of data loss is minimal. Here, we used two types of random number generators to form the membership matrix for each pixel. In this study, three validity measures are used; partition coefficient (PC), partition entropy (PE), and compactness and separation function (SC), which measures the goodness of the clustering result. Shape analysis is carried out to correlate the clustering result to the clinical paradigms of the Pap smear images. The method produces a reasonable result, with some common drawbacks of the FCM clustering In this paper, an attempt has been made to highlight the different shortcomings and usefulness of FCM clustering algorithm.

***Index Terms*— fuzzy clustering, segmentation, partition coefficient (PC), partition entropy (PE), compactness and separation function (SC), Pap smear images.**

I. INTRODUCTION

In most practical application, there are many cases, in which the clusters are not completely disjointed and data could not be classified as belonging to one cluster. Therefore, the separation of the clusters becomes a fuzzy notion, and the representations of real data structures can then be more accurately handled by fuzzy clustering methods. The fuzzy c-means (FCM) algorithm is the best known and the most widely used technique. The objective function of FCM algorithm is very suitable for image processing. The complexity involving the uncertainty and vagueness in digital image can be prudently handled FCM algorithm. The use of membership values provides more flexibility and makes the clustering results more useful in practical applications. Now let us consider the FCM clustering method to segment the image. At each pixel, we have two membership values, one representing the degree of certainty of a pixel belonging to background and the other representing the degree of certainty of a pixel belonging to foreground. Since the sum of two membership values must be equal to one, they are not independent of each other.

II. PREVIOUS WORK ON PAP SMEAR USING FCM ALGORITHM

There exist many literatures suggesting various methods Pap smear image segmentation as well as classification. From conventional digital image processing techniques to hybrid intelligent methods, various methods are suggested based on the requirement. The first attempt to detect and segment cells in cervical microscopic images was based on image thresholding techniques [1]. Pixel classification [2], morphological watersheds [3], [4] were some other methods. The boundaries of the structuring elements of the cells can be detected by applying methods based on active contours [3], template fitting [5], genetic algorithms [6], region growing with moving K-means [4], and edge detectors [7]. The method proposed by Lezoray and Cardot [8] is based on pixel-classification techniques for the detection of the nuclei markers, in order to avoid the over segmentation that the watershed algorithm may produce. Pap smear images exhibit great complexity and the number of pixel classes is not obvious. The rough assumption that all the pixels of the image are distributed into two classes, such as nuclei pixels and other pixels, would produce noisy results. Plissiti *et al.* [9] proposed a method of automated cell nuclei detection in Pap smear images using morphological reconstruction and two types of clustering techniques. Unsupervised FCM and supervised Support Vector Machine (SVM) clustering are used in the study. The preprocessing of the images involved application of adaptive histogram equalization with Otsu's method [10]. The process is refined with the application of a 3x3 flat structuring element. Gray scale morphological reconstruction and h-minima transform are used to determine the centroids of the candidate nuclei. The problem of this method is requirement of prior knowledge of the problem domain. Kim et al. [11] proposed a method of segmentation and classification of Pap smear images using HSI model and FCM algorithm. The extracted characteristics of the cell nuclei are morphemetric feature, densitometric feature, colorimetric feature, and textural feature. The acquired images are preprocessed with a three layer processing: conversion to gray scaled image, removal of noise using brightness information, and application of fuzzy morphological operations. Muhimmah *et al.* [12] proposed a method of detection of epithelial cells in Pap smear images using a combined framework of distance metric and FCM clustering algorithm. The preprocessing of the Pap smear images is done by applying adaptive histogram equalization and global thresholding. Three binary images are created from each of the Pap smear images using Otsu's method [10]. Logical OR operation is used to obtain a binary mask from the binary image. The connected components with less than 500 pixels are removed. For detection of candidate nuclei,

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 1, July 2013

morphological reconstruction is applied. The extra nuclei markers are removed using an edge detection method. *Ghafar* et al. [13] proposed a method of Pap smear image segmentation using stretching and clustering technique. Cebron *et al.* [14] put forward a method of classification cell assay image of Pap smear test through FCM clustering algorithm and learning vector quantization. The method involved unsupervised classification of the cell assay image supplemented by user-defined parameters. The cell assay images contain a large number of images of cervical cells. A trainee neural network followed by region growing method does segmentation of the images.

III. FCM ALGORITHM IN PAP SMEAR IMAGE SEGMENTATION

An image is nothing but a collection of pixels; each pixel having a particular value for the three colour channels red (R), green (G) and blue (B). To initiate the clustering process random numbers are generated corresponding to each of the R, G and B value. Two types of random number generator are used; general random number generator and random number based on Chaos theory. These random numbers constitute the membership value (μ) of the pixels. The μ value of the pixels are compared with that of the cluster center and classified accordingly. The clustering process segments the images namely into three classes: cytoplasm, nucleus and the background.

IV. CLUSTER VALIDITY FOR SEGMENTATION OF CERVICAL CELL

Several cluster validity measures have been developed in the past. In this section, three of these measures are described: partition coefficient, partition entropy, compactness and separation validity function. The partition coefficient (PC) is defined as [15]

$$F(U,c)=\frac{1}{n}\sum_{i=1}^{c}\sum_{k=1}^{n}(\mu_{ik})^2$$

The partition coefficient measures the closeness of all input samples to their corresponding cluster centers**.** The partition entropy (PE) is defined as [15]

$$H(U,c)=-\frac{1}{n}\sum_{i=1}^{c}\sum_{k=1}^{n}\mu_{ik}\log(\mu_{ik})$$

The compactness and separation (SC) validity function is defined as [16], [17]

$$S(U,c)=\frac{\frac{1}{n}\sum_{i=1}^{c}\sum_{k=1}^{n}\mu_{ik}^2\left|x_k-v_i\right|^2}{\min_{i,j}\left|v_i-v_j\right|^2}$$

S(U,c) is the ratio between the average distance of input samples to their corresponding cluster centers and the minimum distance between cluster centers.

V. SHAPE ANALYSIS OF PAP SMEAR IMAGE

The chain code is chosen among 8 selected points on the boundary of the object as shown in Fig.1.The angle between any consecutive lines connecting two consecutive pair of points is 45°. Any two opposite points can be taken as starting points of boundary. Suppose point 1 and point 5 are taken as the starting point of the boundary and let the respective boundaries be *b1* and b5, then we can write *b5=b1+p*, where *p* is a real number.

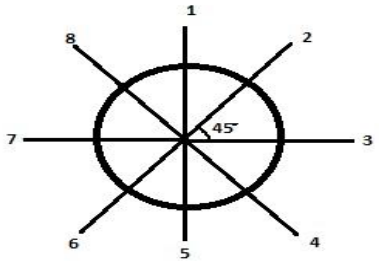

Fig.1: Direction code with 8 points

VI. RESULTS

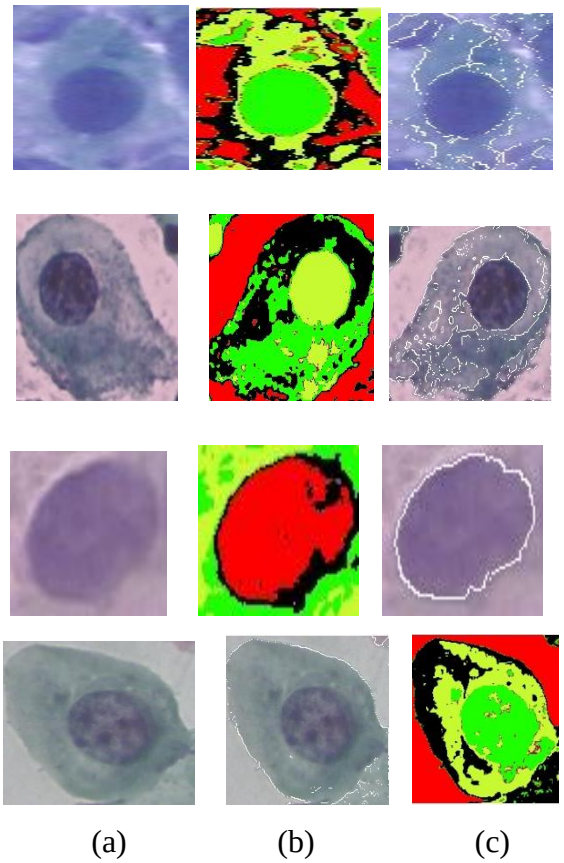

Fig.2 (a) Original image (b) Segmented image (c) Traced cell nucleus boundary

Fig.2 shows the result of boundary tracing and segmentation of Pap smear images. The nucleus is segmented and it can be separately studied. The simple chain code traces the contour of the cell nuclei, which gives an idea about the shape of the cell nuclei. The shape plays an important part in cervical cancer diagnosis.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 1, July 2013

Table 1: Cluster validity measure values

Partition Coefficient (PC)	Partition Entropy (PE)	Compactness and Separation index (SC)
0.48	0.72	0.79
0.31	0.75	1.05
0.49	0.83	1.12
0.51	0.89	1.44

Table 1 gives the cluster validity measure values for the four Pap smear images. We can see that the values are closer to the best possible values indicating good clustering result.

VII. CONCLUSION

Two types of random number generators are used to form the membership matrix for each pixel. The use of right random number generator is very important for initialization of the FCM algorithm. In this study, three validity measures are used; partition coefficient (PC), partition entropy (PE), and compactness and separation function (SC). Compactness and separation validity function (SC) is more in practical application such as image segmentation. The physical interpretation of SC closely resembles the visual perception of digital image. The shape analysis gives the contour of cell nuclei, which is a critical criterion for cervical cancer diagnosis. The application of chain code is easy and simple. The method produces a reasonable result, which suffers from some common drawbacks of the FCM clustering and intricacies of high dimensional image data. The obvious drawback of FCM algorithm is that it cannot detect clusters of arbitrary shape except spherical shaped clusters. In case of colour image segmentation, where very subtle and uncertain change in pixel properties occurs within a large data space, FCM fails to detect all the valid clusters. The tracing of the cell nuclei works well in case of images with less complexity, such as minimal colour variation, but fails to trace the region of interest (ROI) where the boundary is vague. The feature enhancement tasks include enabling the FCM algorithm to detect arbitrary shaped clusters, minimizing the effect of high dimensionality of Pap smear images, and to generalize the shape analysis with some concrete and invariant shape

REFERENCES

[1] H. S. Wu, J. Gil, and J. Barba, “Optimal segmentation of cell images”, IEE Proceedings on Visual Image Signal Processing, Vol.-145, No.-1, pp. 50–56, February, 1998.

[2] E. Bak, K. Najarian, and J. P. Brockway, “Efficient segmentation framework of cell images in noise environments”, Proceedings of 26th International Conference IEEE Engineering Medical Biology, Vol.-1, pp. 1802–1805, September, 2004.

[3] P. Bamford and B. Lovell, “Unsupervised cell nucleus segmentation with active contours”, Signal Processing, Vo.-71, No.-2, pp. 203-213, 1998.

[4] N. A. Mat Isa, “Automated edge detection technique for Pap smear images using moving K-means clustering and modified seed based region growing algorithm”, International Journal of The Computer, the Internet and Management, Vol.-13, No.-3, pp. 45–59, 2005.

[5] H. S. Wu, J. Barba, and J. Gil, “A parametric fitting algorithm for segmentation of cell images”, IEEE Transactions on Biomedical Engineering, Vol.-45, No.-3, pp. 400–407, March, 1998.

[6] N. Lassouaoui and L. Hamami, “Genetic algorithms and multifractal segmentation of cervical cell images”, in Proceedings of 7th International Symposium on Signal Processing Application, Vol.-2, pp. 1–4, 2003.

[7] C. H. Lin, Y. K. Chan, and C. C. Chen, “Detection and segmentation of cervical cell cytoplast and nucleus”, International Journal of Imaging Systems Technology, Vol.-19, No.-3, pp. 260–270, 2009.

[8] O. Lezoray and H. Cardot, “Cooperation of color pixel classification schemes and color watershed: A study for microscopic images”, IEEE Transactions on Image Processing, Vol.-11, No.-7, pp. 783–789, July, 2002.

[9] M. E. Plissiti, C. Nikou, and A. Charchanti, “Automated Detection of Cell Nuclei in Pap smear Images using Morphological Reconstruction and Clustering”, IEEE Transactions on Information Technology in Biomedicine, Vol.-15, No.-2, March, 20

[10] N. Otsu, “A threshold selection method from gray-level histogram”, IEEE Transactions on Systems, Man, Cybernetics, SMC-8, pp. 62-66, 1978.

[11] K. Kim, S. Kim, and G. Kim, “Nucleus Classification and Recognition of Uterine Cervical Pap-Smear using FCM Clustering algorithm”, ICANNGA 2007, Part II, LNCS 4432, pp. 290-299, Springer-Verlag Berlin Heidelberg, 2007.

[12] I. Muhimmah, R. Kurniawan, and Indrayanti, “Automatic Epithelial Cells Detection of Pap smears images using Fuzzy C-means Clustering”, 4th International Conference on Bioinformatics and Biomedical Technology, Vol.-29, IACSIT Press, Singapore, 2012.

[13] R. Ghafar, M. Y. Mashor, and N. H. Othman, “Segmentation of Pap smear slides Images using k-means Clustering”, Proceedings of Kuala Lumpur International Conference on Biomedical Engineering, pp. 41-43, 2002.

[14] N. Cebron and M. R. Berthold, “Adaptive prototype-based fuzzy classification”, Fuzzy Sets and Systems, Vol-159, pp. 2806-2818, 2008.

[15] J. Bezdek, R. Ehrlich and W. Full “FCM: Fuzzy C-Means Algorithm”, Computer and Geosciences, Vol.-10, pp 191-203, 1984.

[16] X. L. Xiu and G. Beni “A validity measure for fuzzy clustering”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.-13, No.-4, pp 841-846,1991

[17] N. Zahid, M. Limouri, and A. Essaid, "A new cluster-validity for fuzzy clustering", Pattern Recognition, Vol.-32, pp.1089–1097, 1999.