ISSN 2319-2720
Volume 1, No.1, July – August 2012
International Journal of Computing, Communications and Networking
Available Online at http://warse.org/pdfs/ijccn07112012.pdf

Robust Fuzzy C-Mean algorithm for Segmentation and analysis of Cytological images

C. K. Nath*, Jyotismita Talukdar# and P. H. Talukdar*
*Department of Instrumentation & USIC, Gauhati University
Guwahati – 14, Assam, India
#Asian Institute of Technology
Bangkok, Thailand
charock4@gmail.com

ABSTRACT

In this paper, we are proposing a method for segmentation of PAP (Papinocolaou) smear images using Fuzzy C-Mean Algorithm (FCM) and analysis of the segmented images based on shape and size criteria. The traditional FCM algorithm used in the present study is modified by replacing the Euclidean distance metric by Mahalanobis distance metric. Further, the computation of the cluster center is modified by including the gray level distribution i.e. the histogram of the image. For shape and size analysis, the cell nuclei distribution based on area, compactness and eccentricity of the cell nuclei are computed. It is found that the proposed Mahalanobis distance metric enhances the ability of FCM algorithm to detect clusters of arbitrary shapes. Further, the inclusion of histogram in the computation of cluster center reduces the computation time significantly. The shape and size analysis provides more specific information for classification of the images more accurately and efficiently.

Keywords : PAP smear Image, Cluster Center, Histogram, Cell Nuclei Distribution, Shape and Size Analysis, Compactness, Eccentricity.

1. INTRODUCTION

Image segmentation is the most basic and important part of image processing which segments an image into meaningful areas according to some characteristics such as gray level, spectrum, texture, colour, and so on. Fuzzy clustering is a branch of cluster analysis. Fuzzy C mean (FCM) algorithm, proposed by Dunn [5] and generalized by Bezdek [4], is based on Euclidean distance. FCM algorithm can only detect spherical shape cluster. Mahalanobis distance is based on correlation between variables by which different patterns can be identified and analyzed. It differs from Euclidean distance in that it takes into account the correlation of the data set and is scale invariant. The first order statistics of an image is the histogram of the image. It gives the information of the distribution of the gray level values in a dynamic range. It is the global statistics of an image.

A Papanicolaou test or Pap smear is a medical screening method that can help prevent cervical cancer. Shape and size analysis is the main factor for detecting abnormality in cervical cells. The automatic analysis of Pap smear microscopic images is one of the most interesting fields in medical image processing.

2. THE FUZZY C MEAN (FCM) ALGORITHM

The Fuzzy C mean (FCM) algorithm just has the function to describe the fuzzy classification for the pixels by calculating the fuzzy membership value. The objective function used in traditional FCM algorithm [4] is

$$\sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik})^{m}(d_{ik})^{2} = \sum_{k=1}^{n}\sum_{i=1}^{c}(u_{ik})^{m}\left\|x_k - v_i\right\|^{2} \quad (1)$$

'c' is the number of cluster, m is the fuzzifier, m>1, which controls the fuzziness of the method. u_{ik} is the membership value of the pixels and v_i is the cluster center in the subset *i* of the feature space. U is the fuzzy partition. The term, $\left\|x_k - v_i\right\|^2$ is the squared Euclidean distance between a data object x_k and cluster center v_i. The steps in FCM algorithm are as follows:

1. Choose c, the number of cluster, $2\le c\le n$, choose m, the fuzzifier, $1<m<\alpha$ (say m=2 for converging error, $\varepsilon>0$ such as $\varepsilon = 0.001$), initialize U^0, the initial membership matrix such that $u_{ik} = 1$, i = 1, 2,…c, k = 1, 2, …n are not equal.

2. Calculate the cluster center, *vi*

$$vi = \frac{\sum_{k=1}^{n} (u_{ik})^{m} X_k}{\sum_{k=1}^{n} (u_{ik})^{m}} \tag{2}$$

3. Calculate the new membership value as follows

$$uik = \frac{1}{\sum_{j=1}^{c} \left(\frac{\|x_k - v_i\|}{\|x_k - v_j\|} \right)^{\frac{2}{m-1}}} \qquad where\ i = 1,2,\ldots c,\ j = 1,2,\ldots n \tag{3}$$

4. Compare U^{k+1} and U^k, if | U^{k+1}- U^k|≤ε then stop, otherwise go to step 2

3. THE IMPROVED FCM ALGORITHM

3.1. Efficient Cluster detection

The FCM algorithm, based on Euclidean distance, can only detect spherical shaped clusters leading to inefficient

$$\sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^{m} \left\| (x_k - v_i)^T S^{-1} (x_k - v_i) \right\|^2 \tag{4}$$

segmentation. This problem is solved by replacing Euclidean distance by Mahalanobis distance in the objective function, (1) and it becomes

3.2. Minimization of computation time

To decrease the computing time the cluster center calculation is enhanced by inducing histogram of the image denoted by *h(k)* which gives the information of the occurrence of the (L-1) gray level. Equation (2) becomes

$$vi = \frac{\sum_{k=0}^{L-1} (u_{ik})^{m} h(k) k}{\sum (u_{ik})^{m} h(k) k} \tag{5}$$

With the new objective function, (4) and cluster center, (5), the computation is carried out.

4. PRE-PROCESSING BEFORE SEGMENTATION

The coloured PAP smear microscopic images were converted to gray level images for simplicity. We enhanced the contrast of the images by histogram equalization.

5. SHAPE AND SIZE ANALYSIS

5.1. Cell nuclei distribution

The number of normal and abnormal cells in a Pap smear microscopic image is good factor for detecting any abnormality and size of the cell nuclei can be treated as

$$A = \sum_{i}^{M} \sum_{j}^{N} seg(i, j) \tag{6}$$

categorizing factor. We calculated the area of the cell nuclei as follows:

seg(i, j) are the pixels of the segmented object.

5.2. Compactness

It is dimensional shape feature that measures the compactness. It is defined as

$$C = \frac{A}{P^2} \tag{7}$$

A is the area and *P* is the perimeter of the cell nuclei. Perimeter is the sum pixels that form the contour of the object.

5.3. Eccentricity

The shape of the cell nuclei can be treated as ellipsoidal.

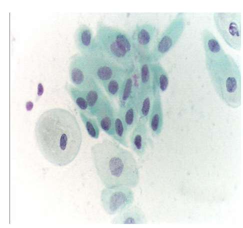
$$e = \sqrt{\frac{a^2 - b^2}{b^2}} \tag{8}$$

We calculated the length of semi-major and semi-minor axes and calculated the eccentricity as follows

a is the semi-major axis and *b* is the semi-minor axis.

6. EXPERIMENTAL RESULTS

6.1. Segmented PAP smear images



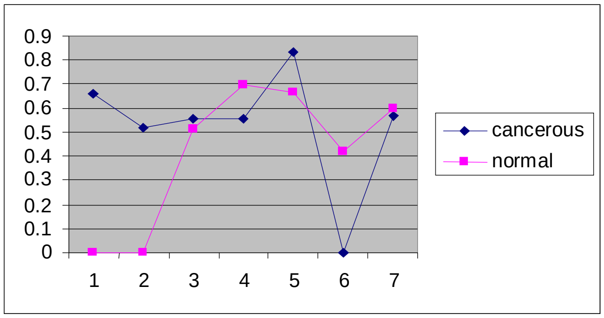
6.1.(a)

6.1.(b)

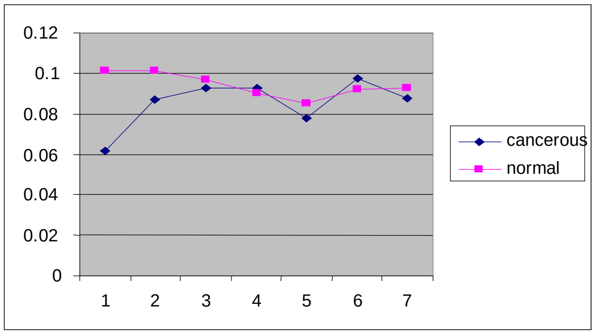
6.2. Table 1: Cell nuclei distribution

Number of Nuclei	Normal Cell	Abnormal Cell
2	13	45
2	24	30
1	21	40
1	18	42
1	26	25

6.3. Graphical comparison of eccentricity of normal and abnormal cell nuclei



6.4 Graphical comparison of compactness of normal and abnormal cell nuclei



7. CONCLUSION

The use of Mahalanobis distance enabled FCM to detect clusters of arbitrary shape and thus resulting more efficient segmentation. For an *(M x N*) image, the data set is reduced to *L* from (*M x N).* So, for an (512 x512) image size and 8-bit gray image, the computing time is improved by 1024 times when we neglect the time needed to compute the histogram and mark the pixels. Shape and size analysis gives a clear distinction between normal and abnormal cells which is very useful in categorizing the images.

REFERENCES

1. D. C. Pham and J. L. Prince. **An Adaptive Fuzzy C-Mean algorithm for image segmentation in the presence of intensity in homogeneity**, Pattern Recognition Letters, Vol. 20, pp. 57-68, 1999.

2. F. J. Funes, M. E**.** Mayorga and R. C. Santiago. **Rank M-type Radial basis function (RMRBF) Neural Network for PAP smear microscopic image classification**, Apeiron, Vol. 16, No. 4, 2009.

3. H. C. Liu, J. M. Yih, D. B. Wu and S. N. Liu. **Fuzzy C-Mean algorithm based on complete Mahalanobis distances**, Proc. 7th Int.Conf. on Machine learning and Cybernetics, Kunming, 2008.

4. J. C. Bezdek. **Pattern recognition with Fuzzy Objective function algorithm**, New York: Plenum Press, 1981.

5. J. C. Dunn. **A Fuzzy relative of the ISODATA process and its use in detecting compact, well-separated clusters**, J. Cybernets, Vol. 31. pp. 32-57, 1973.

6. L. Jinag and W. Yang. **A modified Fuzzy C-Mean algorithm for segmentation of Magnetic Resonance images**, Proc. 7th Digital Image Computing: Technologies and Applications, Sydney, 2003.

7. R. C. Gonzalez and R. E. Woods. **Digital Image Processing**, 2nd Edition, New Delhi: Pearson education India, 2003.

8. S.D. Gupta. **The Evolution of D2-Statistics of Mahalanobis**, Indian J. of Pure Math., Vol. 26, pp. 485-501, 1995.