Fuzzy Mathematical and Shape Theoretic Approach to Cervical Cell Classification

L. B. Mahanta
Institute of Adv. Study in Science and Technology
Guwahati – 35, P.O- Gorchuk
Assam, India

C.K Nath
Institute of Adv. Study in Science and Technology
Guwahati – 35, P.O- Gorchuk
Assam, India

S. Karan
Institute of Cybernetics and Information Technology
Salt Lake, DL – 205
Kolkata, India

D. C Nath
Dept. of Statistics
Gauhati University, Guwhati - 14
Assam, India

D. D. Majumdar
Institute of Cybernetics and Information Technology,
Salt Lake, DL – 205
Kolkata, India

J. D. Sharma
Dept. of Pathology
Dr. B. Borooah Cancer Hospital, Guwhati
Assam, India

ABSTRACT
We applied traditional fuzzy mathematical approach with enhanced initialization procedure to segment Pap smear images of cervical cells. The segmented images of the cervical cells were analyzed with the help of shape theory to classify them accordingly to the presence of abnormality in the morphological behavior of the cells.

General Terms
Fuzzy mathematical approach, Shape theory.

Keywords
Cervical cells, Pap smear, Fuzzy c-means, nucleus and cytoplasm.

1. INTRODUCTION
Automated classification of cervical cells is a challenging task in the field of medical sciences. Classification of cervical cells is the preliminary criteria for finding any abnormality occurring in the cervical region. Manual observation of cells in Pap smear slide is prone to error due various factors such huge number of cells in a slide, fatigue and tidiness in observation of large number of slides, different expertise level of the health personnel etc. An automated intelligent method tends to give uniform and accurate result.

Feeding of morphological and other information regarding the cells into various intelligent methods is the common paradigm of several researchers in automated cell classification process. Morphological information of the cells like size and shape are computed in order to create the dataset for classification purpose. A set of information regarding the morphology of the cells and colour intensity is computed from the Pap smear images using commercially available image analysis software and is classified using traditional classifiers such Minimum Distance classifier and advanced classifier such as Nearest Neighbour with GA Feature Selection, Nearest Neighbour with TABU Search Feature Selection and Ant Colony optimization [11]. *Jantzen et al.* made a comparative study on the performance of Adaptive Network based Fuzzy Inference system, Fuzzy c–mean (FCM) and Gustaffson-Kessel (GK) clustering algorithm in the classification task of cervical cells [12]. The authors in [7] described a method of classification of cervical cells using a Rank M-type radial basis function neural network which is trained with a dataset of some morphological information of the cells. A comparative study is made among fuzzy based classifier, Genetic algorithm and second order neural network based on their performance in the classification of the cervical cell data in [8]. Measurement of the DNA content of the cell nucleus is another way to classify the cells [13]. Another approach carried out by *Funes et al* [7] tries to mark the objectionable areas of the slides by cataloging the eye-fixation of the screeners using a sophisticated eye tracking device. These markers are used to find the maxima by applying Wavelet transform on the hue and a saturation/value combined component of the images. The slides are stained with liquid based ThinPrep method which allows only singular cellular layer of the slide. Pap smear slides stained with traditional Papanicloaou method cannot be analyzed with this method. Multilayer neural network with back-propagation technique was used to classify cervical cells after calculating some morphological and distributive data of the cells in [21]. The authors in [5] described a combined method of gray scale method and energy method to classify Pap smear images. Apart from the Pap smear images, Cervigram images are also analyzed using computer aided technology [1]. Fluorescence spectroscopy based radial basis function neural network was used to classify the cervical tissue spectra in [16]. Another challenging task in cytological image analysis is the segmentation of the images into different segments mainly cytoplasm and nucleus. Filter and morphological operations are applied in all three colour channel to detect the nuclei in and a deformable model is used to extract the nucleus boundary [14]. But in [14], the threshold for finding the intensity valleys has to be determined manually. Active contour, statistical modeling and deformable templates are widely used for segmentation of cytological images [17],

[20], [2]. Commercially available image processing software is also used for segmentation which also provides numerical data regarding the cell characteristics [13], [12], [11].

Our work mainly focuses on the segmentation of the cell nuclei and the cytoplasm of the microscopic Pap smear images of uterine cervix and classification of normal and abnormal cells. Fuzzy c-mean algorithm used to segment the images and generalized shape theory is applied to classify the images. Other necessary image processing tasks are also carried out in our work. For the convenience of the pathologists, an automatic report generation method is developed where the report contains the necessary information of the cells regarding their shape, size, nucleus/cytoplasm ratio and so on.

2. ACQUISITION OF PAP SMEAR IMAGE

Pap smear is a cytological test of the uterine cervix. In Pap smear test, cervical cells are collected from the surface of the cervix using a brush or spatula and are smeared to into a slide. The slide is then stained Papaniclolaou method. The staining makes it possible to observe the characteristics of the cells under a microscope. A stained slide when looked upon under a microscope shows the cell nuclei and the cytoplasm along with the background. The cytologists look for any cellular change in the slides. The factors that indicate any abnormality are the change in the shape and size of the cell nuclei, increasing nucleus-cytoplasm area ratio (N/C ratio), and increasing chromatin content of the nucleus etc. We mounted a high resolution camera on microscope through an adapter. The 400x magnified microscopic images are shot by the camera and passes into computer memory for analysis. The image files are stored in TIFF format.

3. IMAGE COMPUTATION

3.1 Image Computation using Fuzzy c-mean (FCM) algorithm

We consider colour Pap smear images in RGB colour channel for analysis. 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). Hence to find the different areas inside the image, we consider the image as a set of data having pixel values R, G and B. This dataset $(x_1, x_2,....x_n)$ is classified using Fuzzy c-mean (FCM) clustering algorithm to distinguish the different regions inside image namely nucleus and cytoplasm. To initiate the clustering process we generate random numbers corresponding to each of the R, G and B value. Three kinds of random number generator are used; general random number generator, chaos function and chaos function with timestamp. 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.

3.2 Cluster Validation

In practical applications, we need a cluster validity method to measure the quality of the clustering result. Several cluster validity measures have been developed in the past. In this section, we describe three of these measures: partition coefficient, partition entropy and compactness and separation validity function. The partition coefficient is defined as

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

where $U=[u_{ik}]$ is a $c \times n$ matrix , where u_{ik} is i^{th} membership value of the k^{th} input sample X_k c = number of cluster, n = number of datasets, u_{ik} = i^{th} membership value of the k^{th} input sample X_k .Suppose that Ω_c represents the clustering result, then the optimal choice of c is given by

$$\max_c\left\{\max_{\Omega_c}F(U,c)\right\}$$
$$c=2,...,n-1$$

The partition coefficient measures the closeness of all input samples to their corresponding cluster centers. If each sample is closely associated with only one cluster, that is, if for each k, μ_{ik} is large for only one I value, then the uncertainty of the data is small, which corresponds to a large F(U , c) value. The partition entropy is defined as

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

The optimal choice of c is given by

$$\min_c\left\{\min_{\Omega_c}H(U,c)\right\}$$
$$c=2,...,n-1$$

When all μ_{ik} 's have values close to 0.5, which represents a high degree of fuzziness of the clusters, H (U, c) is large and thus indicates a poor clustering result. On the other hand, if all μ_{ik} 's have values close to 0 or 1, H (U, c) is small and indicates a good clustering result. The compactness and separation validity function is defined as:

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

The optimal choice of c is given by

$$\min_c\left\{\min_{\Omega_c}S(U,c)\right\}$$
$$c=2,...,n-1$$

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. A good cluster procedure should make all input samples as close to their cluster centers as possible and all cluster centers separated as far as possible. The compactness and separation validity function seems to work better for image segmentation problems..

4. SHAPE BASED ANALYSIS

The perception of shape has been used for pattern recognition, computer vision, shape analysis, and image registration. Here we shall consider shape analysis and shape based similarity measures based on Dutta Majumder's generalized shape theory, shape distance and shape metric approach [6]. The Generalized shape theory uses co-ordinate transformation of landmark points of the region of interest (ROI) in the respective images. Where as shape metrics and shape distance uses degrees of match between the corresponding shapes of the two nucleus images.

4.1 Generalized Shape Theory

In Generalized Shape theory we align on the basis of some invariant landmark points on the boundary of the region of interest (ROI) by means of translation and/or scaling and/or rotation. We consider a geometrical figure X in R^K space containing N control points that can be represented by $X : N \times K$ matrix. Now two shapes, X and X' are of the same shape if they are related by the following rigid body transformation equation $X' = \beta X\Gamma + I_N vT$ where

$\Gamma : K \times K$ is a rotation matrix and $|\Gamma| = 1$

$I_N = N \times 1$ of one.

$v = K \times 1$ is a translation vector.

β: isotropic scaling factor and >= 0

It is possible to formulate an approximate co-ordinate transformation for mapping between two sets of landmarks in a least square sense using Taylor series expansion. For two sets of landmark points (x_m, y_m) and (x_m', y_m'), m = 1,2,…,n, one set can be expressed in terms of other as follows:

$$x' = q_0 + q_1 x + q_2 y + q_3 x^2 + q_4 xy + q_5 y^2 + \ldots$$

$$y' = r_0 + r_1 x + r_2 y + r_3 x^2 + r_4 xy + r_5 y^2 + \ldots$$

Our study has shown that even in the case of change in structural (shape) properties, the algorithm is general enough to capture the changes. With the change, re-registration may be necessary with different set of invariant points.

4.2 Shape metric and Shape distance in Cancer cell image registration

We present a theoretic approach to define shape and shape distance in our endeavor to apply the concept in cytological image registration. The shape of an object is defined as a subset X in R^2 if

(i) X is closed and bounded.

(ii) Interior of X is non-empty and connected.

(iii) Closure property holds on interior of X.

An object, Y in R^2 has the same shape as X if R^2 if it preserves translation, rotation and scaling invariance.

Shape distance between two shapes, X and Y in F is defined as follows

$$d_1(X,Y) = m_2[(X - Y) \cup (Y - X)]$$

where m_2 is Lebesgue measure in R^2 and d_1 satisfies following rules:

i) $d_1(X,Y) \geq 0$

ii) $d_1(X,Y) = 0$ *if and only if* $X = Y$

iii) $d_1(X,Y) = d_1(Y,X)$

iv) $d_1(X,Y) + d_1(Y,Z) \geq d_1(X,Z)$

For shape similarity measure we consider two cells are of same shape if and only if one of there image is translation, dilation and rotation of other. In order to register two cells images, it is necessary to normalize the form of images of interest in terms of position, size and orientation. After normalization, shape distance is measured in terms of volume of mismatch. The shape is described on the basis of its structural features using certain chain codes with respect to one reference point. Reference points are obtained from the intersection points of the major axis with the contour of the region, which is invariant under translation, rotation, and dilation of the region and the major axis is unique.

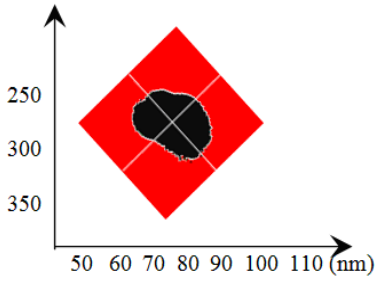

Fig 1: Contour with major axis and two intersecting points with the contour

The centroid (x_g, y_g) of the contour is given by N number of points as:

$$x_g = \frac{1}{n}\sum_{j=1}^{n} x_j \qquad y_g = \frac{1}{n}\sum_{j=1}^{n} y_j$$

In polar co-ordinate the major axis is defined as:

$$f(\theta, c) = \sum_{j=1}^{n} (c\cos\theta + r_j \sin(\theta - \theta_j))^2$$

where θ = 0° to 180°

The slope of axis (∝) is obtained from the best linear fit solution by minimizing f (θ, c) as:

$$\tan 2\alpha = \frac{2\sum_{j=1}^{n}(x_j - x_g)(y_j - y_g)}{\sum_{j=1}^{n}(x_j - x_g)^2 - \sum_{j=1}^{n}(y_j - y_g)^2}$$

$$\cos 2\alpha * \sum_{j=1}^{n}(x_j - x_g)^2 - \sum_{j=1}^{n}(y_j - y_g) + 2\sin 2\alpha * \sum_{j=1}^{n}(x_j - x_g) \geq 0$$

To extract the feature of the boundary of the ROI it is helpful to represent the closed contour with a set of direction. The direction code may be taken among "n" selected points on the contour, which has same distance between any two consecutive points. The direction d makes an angle *(i-d) 45°* with direction i, where real number *d* $\in$*1 to 8* and *i = (1,2,..8).* Let $d_m = (d_{ij})_{j=1}^{n}$ where *m = A, B,* be the contour starting from each reference point A and B and are denoted by d_A and d_B respectively. If d_2 is a rotation of d_1 then $d_2 = d_1 + \gamma$ for any real number γ. For all j we can write

$$d_2 = d_1 + \gamma \forall j$$

The distance function D, in terms of the direction code between the contour of interest and the model is defined as:

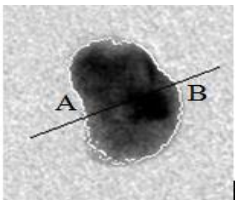

Fig 2: Two reference points A and B

$$D(d_1 d_2) = \sum_{j=1}^{n} \min((d_{1j} d_{2j}), 8 - (d_{1j} d_{2j}))$$

The normalized value of D is D/n and the shape similarity measure between the two shapes is given by

$$\mu = 1 - \frac{D}{n}$$

Smaller value of D indicates higher degree of similarity [18].

5. IMAGE COMPUTATION RESULT

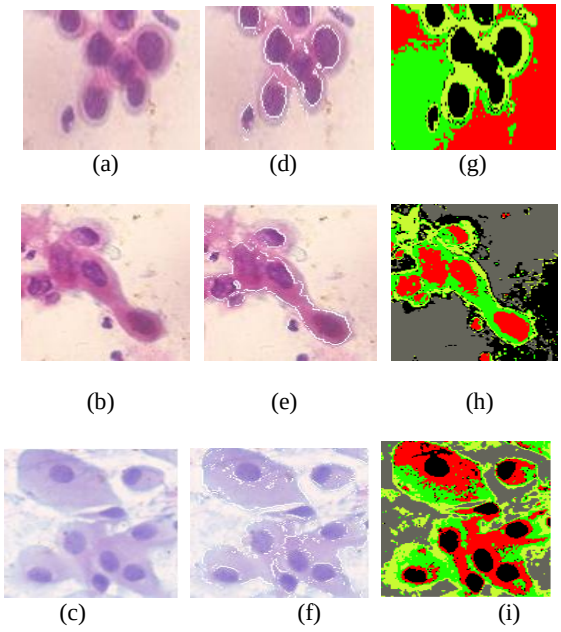

Fig 3: (a) and (b) HSIL Pap image. (c) Atrophic Pap image, (d) to (f) carcinoma tracing, (g) to (i) segmented Pap image

.

6. CONCLUSION

Our method of cervical cancer diagnosis is a robust one and is very successful in assisting the pathologists in the screening process of cervical cancer. The image data is very prudently handled by the Fuzzy c-mean (FCM) clustering algorithm. The initialization process which we have used in the clustering process ensures firm result which can be seen in the segmentation outputs. The results, we have obtained, are validated with the clinical findings and it proves to be satisfactory with some minor enhancement has to be made. One of the most important and foremost enhancement is to establish a direct relationship between the actual cell dimensions and the pixel dimension of the digital images. Only morphological information of the cells is used in the classification process; the inclusion of parameters such as chromatin content and DNA content of the cell will give more accurate results. We are trying to incorporate those features by studying the colour intensity variation and texture analysis. The statistical data of the experiment can used to flag normal sample to questionable samples.

7. REFERENCES

[1] Amir Alush, Heyit Greenspam and Jacob Goldberg, "Lesion detection and segmentation in Uterine Cervix images using an Arc-level MRF", Proc. of the 6th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2009

[2] A. Garrido and N. Perezde la Blanca, "Applying Deformable templates for cell image segmentation", Pattern Recognition 33, 2000

[3] A. P. Kurkure and B. B Yeole, "Social inequalities cancer with special reference to South Asian Countries", Asia Pacific Journal of Cancer Prevention, Vol. 7, No. 1, 2006

[4] B. H. Yang et al, "Cervical cancer as a priority for prevention in different world regions: An evaluation using years of life lost", International Journal of Cancer, 109, 2004

[5] Chin Wen et al, "Automatic Segmentation of abnormal cell nuclei from microscopic image analysis for Cervical Cancer screening", Proc. IEEE 3rd International Conference on Neuro/Molecular Medicine and Engineering, 2009

[6] D. Dutta Majumdar and M. Bhattacharya, "Cybernetics approach to medical technologies: application to cancer screening and other diagnostics", Kybernetes, Vol. 29, No. 7/8, 2004

[7] F. J. Gallegos-Funes et al, "Rank M-type Radial Basis function (RMRBF) Neural network for Pap smear microscopic Image classification", Apeiron, Vol. 16, No. 4, 2009

[8] George Dounias et al, "Automated identification of Cancerous smears using various competitive intelligent techniques", Oncology Reports, 15, 2006

[9] G. Ritter, "Handbook of Computer Vision Algorithms in Image Algebra", CRC Press, 2001

[10] H. R. Myler and A. R. Weeks, "The Pocket Handbook of Image Processing in C", Prentice Hall, 1993.

[11] Jens Byriel, "Neuro-fuzzy classification of cells in Cervical smears", M.Sc Thesis, Technical University of Denmark, 1999

[12] Jan Jantzen and George Dounias, "Analysis of PAP smear image data", Proc. of Nature Inspired Smart Information System, NISIS 2006

[13] James H. Tucker, "An Image Analysis System for Cervical Cytology Automation using DNA content", The Journal of Histochemistry and Cytochemistry, Vol. 27, No. 1, 1979

[14] L. R. Coombes and P. F. Culverhouse, "Pattern recognition in Cervical Cytological Slide Images", 5th International Conference on Advances in Pattern Recognition (ICAPR -2003) December, 2003, Kolkata, India

[15] M. E. Plissiti et al, "Automated segmentation of cell nuclei in Pap smear images", Proc. of IEEE International Special Topic Conference on Information Technology in Biomedicine, Greece, 2006

[16] N. Ramanujam, J. Ghosh and R. Richards-Kortum, "Ensemble of Radial Basis Function Network for Spectroscopic Detection of Cervical Precancer", IEEE Trans. on Biomedical Engineering, Vol. 45, No. 8, 1998

[17] P. Bamford and B. Lovell, "Unsupervised cell nuclei segmentation with active contours", Signal Processing, Vol. 72, No. 2, 1998

[18] R. Sankaranarayana et al, "Accuracy of visual screening for cervical cancer neoplasia: Results from IARC multicentre study in India and Africa", International Journal of Cancer, 111, 2004

[19] S. E. Waggoner, "Cervical Cancer", Lancet, 361, 2003

[20] T. Mourotis and S. J. Roberts, "Robust Cell nuclei segmentation using statistical modeling", IOP Bioimaging 6, 1998

[21] Zhong Li and Kavyan Najarian, "Automated classification of PAP smear Tests using Neural Networks", Proc. on International Joint Conference on Neural Networks (IJCNN 01) vol 4, 2001, pp 2899-2901