Automated Extraction of Cytoplasm and Nuclei from Cervical Cytology Images by Fuzzy Thresholding and Active Contours

Karthigai Lakshmi, G.
Department of Computer Science
V.V.Vanniaperumal College for Women
Virudhunagar,Tamil Nadu, India

Krishnaveni, K.
Department of Computer Science
Sri S.Ramasamy Naidu Memorial College
Sattur,Tamil Nadu, India

ABSTRACT

In this paper, a novel method for automated diagnosis of cervical cancer by extracting cytoplasm and nuclei from cervical cytology images is described. The background is removed by preprocessing methods like Edge sharpening and Adaptive Histogram Equalization. Fuzzy thresholding and Active contours are used for extracting the region of interest containing the cytoplasm and nuclei. The nuclei are separated from the cytoplasm using linear contrast stretching. The nucleus to cytoplasm ratio is used to determine the stage of cancer.

General Terms

Image Processing, Cervical Cancer

Keywords

Cervical Cancer, Cervical Cytology images, Linear Contrast Stretching, Adaptive Histogram Equalization, Fuzzy thresholding, Active Contours

1. INTRODUCTION

Cervical cancer is the fifth deadliest cancer in humans and second deadliest in women. In developing countries and under developed countries, the awareness of the causes and effects and of cervical cancer is far less than developed countries. Cervical cancer kills 280,000 women every year. In India, cervical cancer accounts for 27% (77,100) of the total cervical cancer deaths in the year 2008. [1] In 2010, 33,400 women died of cervical cancer. 16 per 0.1 million, women are affected per year by cervical cancer. Cervical cancer can be cured, if proper diagnosis is done at an early stage. Pap smear test is one of the popular methods of diagnosing cervical cancer using microscopic cervical cytology images. Features extracted from these images can serve to diagnose the stage of cancer. Manual screening of cervical cytology image obtained from Pap smear test is error prone due to several reasons as uneven dyeing, poor contrast, blood stain and it is a time consuming process. Differentiation of types of cells as neoplastic and dysplastic can be automated to reduce human errors and improve diagnosis. Malignant cells of cervical cancer have immature cytoplasm, abnormal features in nucleus, increased nucleus to cytoplasm ratio.

In this paper, Pap smear cervical cytology images in RGB color space are acquired, preprocessed to enhance quality of the image and to remove background. Nuclei are extracted by linear contrast stretching. Fuzzy thresholding and active contours are used in separation of cytoplasm from the image.

The paper is organized as four sections. Section 2 deals with the literature review. Section 3 gives an overview and detailed description of the proposed method. Experimental analysis and results are given in section 4.

2. LITERATURE REVIEW

Human Papilloma Virus (HPV) causes cancer in the cervix. In 1941, Georgios Nikolaou Papanikolaou developed Pap Smear test to diagnose the stage of cervical cancer.[2] Several methods of automated detection of cancerous cells from cervical cytology images have been developed. Features of cytoplasm and nuclei of cytology images are used to classify the cells into benign and malignant cells. The level of malignancy as Low Grade Squamous Intraepithelial Lesion (LSIL) and High Grade Squamous Intraepithelial Lesion (HSIL) is also determined based on these features. Region growing, K-means clustering, fuzzy c-means clustering and watershed segmentation are the various methods available for segmentation of cytology image into its components. Shys-Fan Yang-Mao et al [3] proposed an edge enhancement method that uses Alpha trimmed filter, bi-group enhancer and contour for cytoplasm and nuclei detection. This method consumes a lot of time and can work only on single cell images. Nazahah Mustafa et al [4] proposed a Seed Based Region Growing algorithm for automated multicells segmentation of Thin Prep Image. This method uses k-means clustering process for splitting image into background, cytoplasm and nuclei. A seed pixel selected using moments is used for region growing. This method works out well for non overlapping cell images. M.E.Plissiti et al [5]-[6] have used Fuzzy c-means clustering reconstruction techniques. Selection of threshold for H-minima transform to eliminate regional minima is crucial. Creation of marker image for morphological reconstruction, to locate position of candidate nuclei involves geodesic dilation which is an iterative process. The number of iterations of geodesic dilation is image dependent. Chanho Jung's [7] method uses H-minima transform and watershed segmentation. Elliptical modelling of contours around nuclei eliminates the jagged contours created by watershed segmentation. K-means clustering [8]-[10] and moving K-means clustering methods need an initial value for number of colors in the image to determine the regions of interest. In Pap smear images, the number of colors depends on the nature of cells and staining method. Gradient Vector Flow (GVF) Snakes and Active Contour models have been proposed in many research works. Initial contour positioning is imperative in determining the efficiency of these methods. Support Vector Machine (SVM) method requires a supervised training for classification of Pap Smear Images. Fuzzy thresholding is found to be a better method for removing the background. If initial contour selection is done correctly, proper extraction of cytoplasm can be done. So this paper employs two methods as fuzzy thresholding and active contours.

3. Automated Extraction of Cytoplasm and Nuclei from Cervical Cytology images by Fuzzy Thresholding and Active Contours – An overview

The proposed image segmentation technique consists of three phases as follows:

i) Preprocessing
ii) Extraction of cytoplasm
iii) Extraction of nuclei

The schematic diagram of the proposed method is given in Figure 1. The acquired RGB image is smoothened by Gaussian filtering for uniform distribution of intensity. The edges of smoothened image are sharpened by applying edge sharpening filter. The features of the image are enhanced by Contrast Limited Adaptive Histogram Equalization (CLAHE). CLAHE finds histogram of small tiles of image and combines them to equalise the intensity of whole image. Cytoplasm separation is achieved by fuzzy thresholding and active contours. Nuclei extraction employs linear contrast stretching.
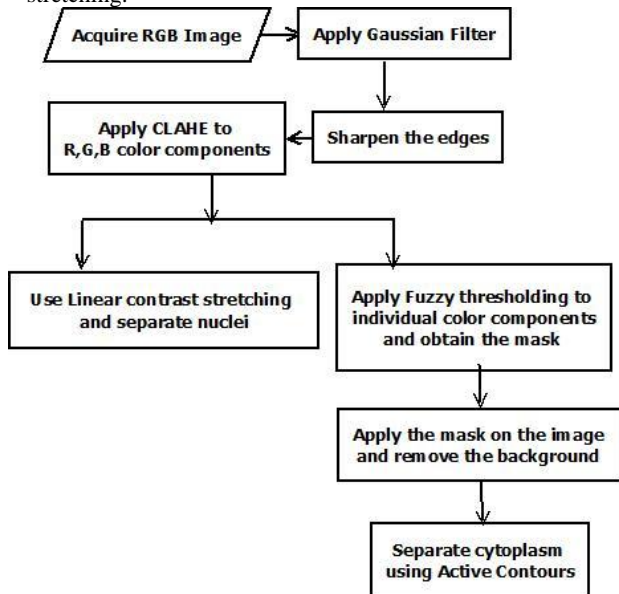

Figure 1. Schematic diagram of the proposed method

3.1. PREPROCESSING STEP

The cervical cytology image in RGB format is given as input. The input image may have non-uniform distribution of intensity due to staining method. Gaussian filter is applied to normally distribute the intensity over the image. 2D Gaussian filter convolved over the image I(x, y), with a standard deviation σ is expressed as a mathematical function G (x,y) .

$$G(x,y) = \frac{1}{2\pi\sigma^2} \left[e^{\wedge}(-(x^2 + y^2)/2\sigma^2) \right] \qquad - \quad (1)$$

Application of Gaussian filter has a smoothening effect on the image. To sharpen the boundaries of cytoplasm and nuclei, an edge sharpening filter is used. This is achieved by subtracting a scaled unsharp version of the image from the original. The resultant image is split up into its RGB components. Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied to each component. CLAHE divides the image into smaller tiles. Histogram of intensity values is found for each tile. The histogram values are then used to redistribute the lightness values of the image. The local contrast of the image is improved. The features of the image are enhanced by CLAHE. The cytoplasm and nuclei can be differentiated from the background and the image has been preprocessed.
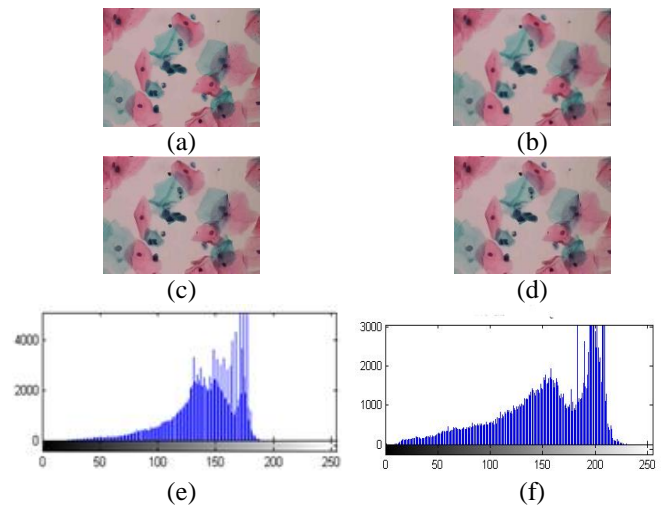

Figure 2 (a) Input image   (b) Gaussian filtered image
(c) Edge Sharpened image (d) Histogram equalized image
(e) Histogram of (c)   (f) Histogram of (d)

3.2 EXTRACTION OF CYTOPLASM

Normal cervical cells have large cytoplasmic area and small, nearly circular, dense nuclei. Pre-cancerous cells have increased nucleus to cytoplasm ratio. Malignant cells occur in clumps and have irregularly shaped nuclei and cytoplasm. To discriminate cervical cells as normal, pre-cancerous, LSIL and HSIL cells, it is imperative to calculate the ratio of area of nuclei to that of cytoplasm (N/C ratio). The stage of cancer is also decided based on shape of nuclei and distance between nuclei. The separation of the cervical cytology image into its components as nuclei and cytoplasm is obligatory.
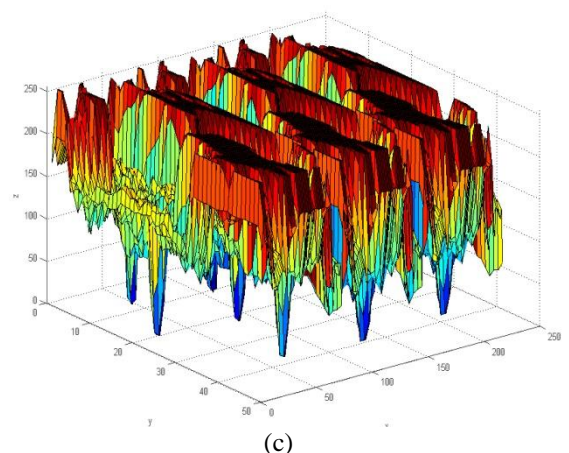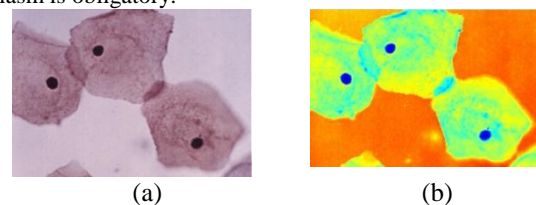

Figure 3 – (a) Normal cells (b) Their intensity distribution (c) Surface plot showing sparsely spaced intensity valleys
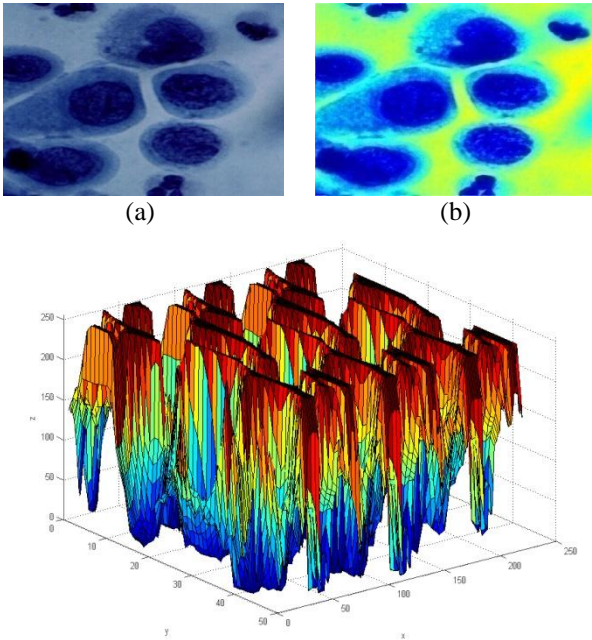
(a) (b)

Figure 4 - (a) Malignant cells (b) Their intensity distribution (c) Surface plot showing closely spaced intensity valleys

The features of cytoplasm can be studied by extracting it from the background. Otsu's global thresholding is a popular method used in image segmentation. In this paper, Fuzzy thresholding [11] is used to separate cytoplasm and background. Gaussian fuzzy thresholding adapts fuzzy logic to determine whether a pixel belongs to background or foreground. Fuzziness index of the image is calculated assuming a normal intensity distribution of pixels. Gaussian index of fuzziness f(z, T) of an image with intensity values in the range 0 to L is given by

$$f(z,T) = \sum_{z=0}^{L-1} h(z)G(\mu_I(z,T)) \text{, where } 0 < T < L-1 \quad - (2)$$

$$G(\mu_I(z,T)) = e^{-\frac{1}{2}\left[\frac{\mu_{I(zT)}-c}{\sigma}\right]^2} \text{ if } 0 < \mu_A < 1 \quad - (3)$$

$$G(\mu_I(z,T)) = 0 \text{ if } \mu_A = 0 \text{ or } \mu_A = 1 \quad - (4)$$

where h(z) is the histogram value of an intensity level z, L refers to maximum intensity. $\mu_I(z,T)$ is the membership function value for a fuzzy set I. The values chosen for the parameters are given below:

Mean c=0.5, Standard Deviation σ =0.15

This fuzziness index f is used as a criterion to decide whether the pixel belongs to a class of background pixels or foreground pixels. Using a generalized Gaussian distribution μ_I (z, T) is given by

$$\mu_I(z,T) = \begin{cases} \dfrac{1}{1 + \dfrac{|(z-m_b(T)|^{2b}}{D}} \\ \dfrac{1}{1 + \dfrac{|(z-m_o(T)|^{2b}}{D}} \end{cases} \quad - (5)$$

where z is the intensity of a pixel, T is threshold, D is the maximum possible intensity value, m_b is mean intensity of background and m_o is the mean intensity of foreground. The parameter b decides the sharpness of fuzzy set. A value of b=0.3 is used here.

$$m_b(T) = \frac{\sum_{z=0}^{T} zh(z)}{\sum_{z=0}^{T} h(z)} \quad - (6)$$

$$m_o(T) = \frac{\sum_{z=T+1}^{L-1} zh(z)}{\sum_{z=T+1}^{L-1} h(z)} \quad - (7)$$

Maximum f *(z, T)* value is used to decide threshold T which separates background intensity and foreground intensity values. Fuzzy thresholding is applied to each color component and binary masks *bwr*, *bwg* and *bwb* are obtained. A binary *and* operator on *bwr*, *bwg* and *bwb* is used to form a single binary mask for the image. This mask is applied on the image to remove the background. The results are shown in Figure 5.
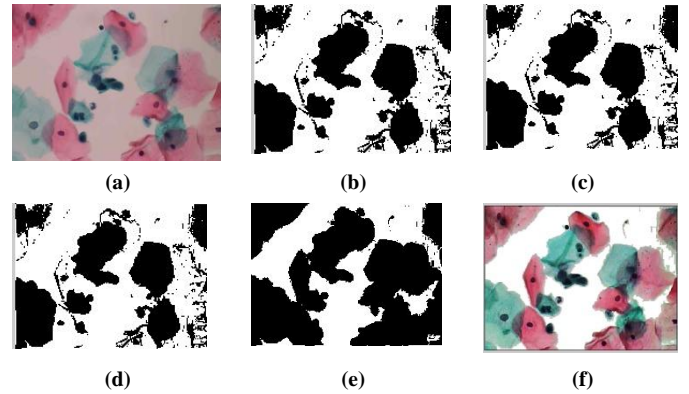

(a) (b) (c)

(d) (e) (f)

Figure 5 – (a) CLAHE applied image (b) Mask of Red component (c) Mask of Green component (d) Mask of Blue component (e) Final mask obtained by AND operator (f) Background removed image

The cytoplasm is extracted from the background removed image using active contours. An Active Contour or Snake is a deformable continuous closed spline curve. The deformation of active contour is determined by internal and external forces of gradient vectors.

$$E_{snake} = \int_0^1 (E_{int}(V(s)) + E_{image}(V(s)) + E_{con}(V(s))ds \quad - (8)$$

The internal force $(E_{int}(V(s))$ acts as constraint on smoothness of snake. $E_{image}(V(s))$ and $E_{con}(V(s))$ are the external forces. $E_{con}(V(s))$ is the constraint energy that drives the active contour towards the features of the image. $E_{image}(V(s))$ represents the energy function of image as a combination of energies of lines, edges and terminations in the image. Active contours perform global thresholding and edge detection by incessant assessment of continuity, curvature and strength of local edge. A conventional active contour has important setbacks like initial contour positioning and initial values of parameters. In this paper, a method suggested by

Chen & Vese has been incorporated. A grid of multiple small circles of radii depending on the size of image is used as initial contour. The initial contour deforms to the cytoplasm boundary in a number of iterations. The number of iterations performed for global cytoplasm thresholding is again image dependent.
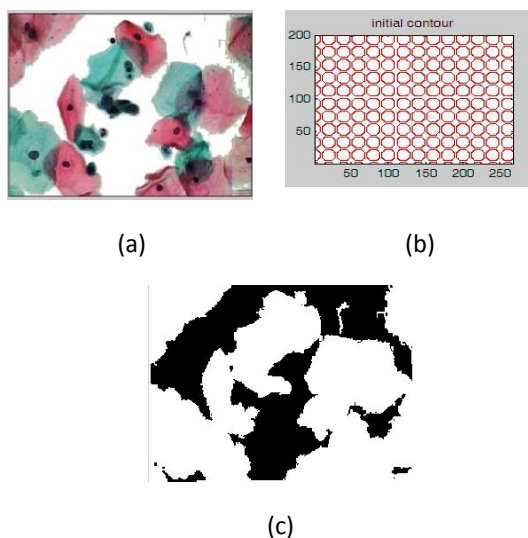

(a)   (b)

(c)

Figure 6 – (a) Background removed image (b) Initial contour (c) Cytoplasm of image

3.3. EXTRACTION OF NUCLEI

Background removed image obtained after fuzzy thresholding is used as input for nuclei extraction. Linear Contrast Stretching (LCS) is applied on this image to adjust the contrast levels and highlight low intensity regions. The LCS applied image and CLAHE applied image are added to brighten the parts of the image except nuclei. The difference of the resultant image and CLAHE applied image is added to the original contrast adjusted image to brighten the nuclei alone.
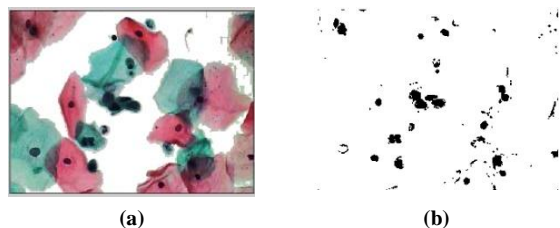

(a)   (b)

Figure 7 (a) Input image (b) Nuclei obtained by linear contrast stretching

4. EXPERIMENTAL ANALYSIS AND RESULTS

38 Pap smear cervical cytology images were taken as experimental samples. When ratio of the area of nucleus to that of cytoplasm is less than 0.4 the cells are normal. When the ratio[12] lies between 0.4 and 0.5, then the cells are classified as Cervical Intraepithelial Neoplasia1 LSIL. When the ratio lies between 0.5 and 0.67 then the cells are classified as Cervical Intraepithelial Neoplasia2 HSIL (CIN2 HSIL) When the ratio lies above 0.67, then the cells are categorized as Cervical Intraepithelial Neoplasia3 HSIL (CIN3 HSIL). Some results are shown in Table 1.

Table 1 – Implication from Nucleus to Cytoplasm ratio Images 1 and 2 are normal cells. Images 3,4 and 5 are cancerous cells

No.	Input Image	Cytoplasm	Nucleus/Nuclei	N/C Ratio
1				0.0445
2				0.0365
3				0.4286
4				0.8613
5				1.7683

5. CONCLUSION

The proposed method separates both the nucleus and cytoplasm from the Pap smear cervical cytology images using linear contrast stretching, fuzzy thresholding and active contours. This is a pilot study that considers the cluster of cells as a whole. The results obtained are found to be satisfactory. Further work will be carried out by attempting to separate individual cells from the clusters and analyse the nature of those cells individually and the stage of cancer.

6. REFERENCES

[1] Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E. and Forman, D. (2011), Global cancer statistics. *CA: A Cancer Journal for Clinicians*, 61: 69–90. doi: 10.3322/caac.20107

[2] Papanicolaou GN, Traut HF. The diagnostic value of vaginal smears in carcinoma of the uterus, *American Journal of Obstetrics and Gynecology*. 1941; 42:193.

[3] Shys-Fan Yang-Mao, Yung-Kuan Chan, and Yen-Ping Chu. Edge Enhancement Nucleus and Cytoplast Contour Detector of Cervical Smear Images, *IEEE Transactions on Systems, Man and Cybernetics—Part B: Cybernetics, Vol. 38, No. 2, April 2008*

[4] Nazahah MUSTAFA, Nor Ashidi MAT ISA and Mohd Yusoff MASHOR. Automated Multicells Segmentation of ThinPrep Image Using Modified Seed Based Region Growing Algorithm, *Biomedical Soft Computing and Human Sciences, Vol. 14, No.2, pp. 41-47(2009)*

[5] M.E.Plissiti, E.E. Tripoliti, A.Charchanti, O. Krikoni and D.I. Fotiadis. Automated Detection of Cell Nuclei in PAP Stained Cervical Smear Images using Fuzzy Clustering

[6] Marina E.Plissiti, Christophoros Nikou and Antonia 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 2011*

[7] Chanho Jung and Changick Kim. Segmenting Clustered Nuclei Using *H*-minima Transform-Based Marker Extraction and Contour Parameterization, *IEEE Transactions on Biomedical Engineering, Vol. 57, No. 10, October 2010*

[8] Chang Wen Chen*, Jiebo Luo, and Kevin J. Parker. Image Segmentation via Adaptive K-Mean Clustering and Knowledge-Based Morphological Operations with Biomedical Applications, *IEEE Transactions on Image Processing, Vol. 7, No. 12, December 1998*

[9] R. Ghafar, N.A. Mat Isa, U.K. Ngah, M.Y. Mashor, N.H. Othman. Segmentation of Stretched Pap Smear Cytology Images using Clustering Algorithm

[10] Ghafar, R., Mashor, M. Y. & Othman, N. H. 2002. Segmentation of Pap smear Slides Images Using K-Means Clustering, *In Proceedings of Kuala Lumpur* International Conference on Biomedical Engineering. 41-43. Kuala Lumpur, Malaysia.

[11] Manuel G.F. and Oswaldo.R. New Formulation in Image thresholding using Fuzzy Logic, *Memorieas Congreso Latinoamericano de Ingenierio Biomedico, Habana 2001, La Habana, Cuba*

[12] www.eurocytology.eu