Segmentation of Overlapping Cervical Cells by Multiple Level Set Functions Optimization

P. Shanmugapriya and N. Priscilla Vilma Manorathi

Department of ECE, Saranathan College of Engineering, Trichy, India

Abstract: Segmentation of overlapping cervical cells is helpful in medical applications for classification of cytologic specimens. The presence of blood, mucus, reduces the accuracy of Pap smear test in this situation it leads to false negative results. Segmentation of the nucleus and cytoplasm of cervical cells within the clumps is very critical. The whole region representing the cellular clumps is segmented which is very difficult to analyze the individual cell within the clump, thus the amount of information conveyed for specimen classification and diagnosis is limited. The methodology addresses these issues by utilizing an optimization of multiple level set functions. Here each function represents a cell within a clump and it has both unary (intra-cell) and pair wise (inter-cell) constraints. The constraints that are based on edge strength and cell shape are unary constraints and the constraint which are based on the area of the overlapping regions are inter-cell constraints. The steps consists of cell clump detection and nuclei detection using the Maximally Stable Extremal Regions (MSER) algorithm. The second step then optimizes the initial segmentation using a level set method that utilizes multiple level set functions. The main goal of our methodology is to segment each and every cell within the clump and produce a significantly larger number of fully segmented cells to improve the classification of cytologic specimens.

Key words: Overlapping cell segmentation · Level set · Pap smear images · Support vector machines · histopathology

INTRODUCTION

The goal of segmentation is to analyze the image in an efficient manner. Image segmentation is used to identify boundaries in images. Image segmentation assigns a label to every pixel in an image. Segmentation divides the image into a set of segments that covers the entire image. Pixels in a region are similar in color with respect to each other. Segmentation of overlapping cells is one of the most challenging problems in the analysis of microscopic images. Segmentation of cervical cells in Pap smear images is critical in diagnosis. Accurate detection of the cells in cervix is essential to identity pre-cancerous cells. Due to the complexities of cell structures the fast, accurate cervical cell segmentation remains a problem [1]. The Pap smear slides [1] are examined under a microscope to find any abnormalities in the cells. Presence of any abnormalities is found by expert cytologists. The correct classification of a specimen for better treatment of portions affectedby disease is highly important. Human error leads to misclassifications in analysis [2]. An approach is developed for nuclei detection, segmentation, classification and feature computation in [3]. These techniques are used in histopathology imagery using protocols. Nuclei detection and segmentation are used in cancer diagnosis and grading. Nucleus detection is performed by Euclidean distance map and Laplacian of Gaussian method. Segmentation is performed by watershed method. Nuclei separation is done by watershed Transform and Classification is performed using SVM classifier. Watershed method can locate nuclei regions effectively, but problems are faced during segmentation of the overlapping regions which results in inaccurate nuclei boundaries. The main disadvantage of nuclei detection method is that it results in over segmentation and tiny regions are detected as nuclei. The method is not straight forward because of various evaluation methods and multiple performance metrics are involved. The approach which deals with segmentation of both nucleus and cytoplasm from isolated cells in [4]
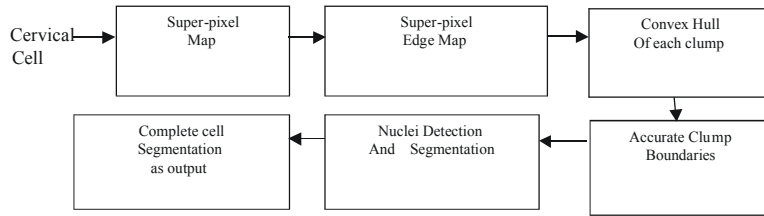

Corresponding Author: P. Shanmugapriya, Department of ECE, Saranathan College of Engineering, Trichy, India.

Fig. 1: Block Diagram of the proposed overlapping cervical cellsegmentation

which gives a new external force for active contours for solving large problems. Snakes or active contours are used in many applications of image processing. The external force is Gradient Vector Flow (GVF) and it is computed by diffusion of gradient vectors of a gray level image. GVF has the property in moving snakes into boundary concavities. From the results of initialization, it is inferred that the traditional potential forces are too weak. To overcome this problem, a balloon model is used and the pressure forces cause the balloon to bulge outward and subjectiv contours are not reconstructed well. From the results using distance potential force, the subjective boundaries at top and bottom are reconstructed well but the snake fails to find the boundary concavities at left and right and it is not able to proceed to the U shaped object. Problems associated with initialization and poor convergence to boundary concavities limited the utility of this work.The aim of the proposed work is to solve the problems pertaining tosegmentation. The issues related to segmentation are solved by utilizing multiple level set functions. Each level set function is representing a cell within a clump. It has both intra-cell and inter-cell constraints. On the basis of edge strength and cell shape, calculation of unary constraints is done. From the area of overlapping regions, the calculation of pairwise constraints is done. Segmentation is performed in each and every cell which increases the potential in diagnosis to identify diseased portions and in the classification of specimens.

MATERIALS AND METHODS

In the proposed approach the Overlapping cells are segmented by Level set method and the nuclei are detected using Maximally Stable Extremal Regions Algorithm.The input to the overlapping cell segmentation is cervical cell, given in Fig. 1, the super-pixel map is obtained by quick-shift algorithm. The prominent edges are obtained by canny edge detection; with the help of edge map the convex of each clump is obtained. The clump boundaries are obtained using level set functions. The nuclei regions are detected using MSER (Maximally

Stable Extremal Regions) algorithm. From cell and nuclei segmentation, the complete cell segmentation is obtained.The Overlapping cell segmentation consists of the following steps namely: Scene Segmentation, Joint Level set Segmentation of Overlapping Cells.

Scene Segmentation: It involves cell clump segmentation to segment the cell clumps. The segmentation process of an image contains two regions namely Region of Interest and Background. The purpose of segmentation is to separate these two regions and further processing are carried out in the Region of Interest.

The Region of Interest is extracted by setting thresholds which discriminates Region of Interest from background. Also, it involves the following steps given below.

Super-Pixel Map Generated by Quick-Shift: The first step is computation of the Super-pixel Map by Quick-Shift algorithm. In this algorithm all the data points are arranged into a tree where the parents in the tree are the nearest neighbors who increases the density. This method is helpful in image segmentation and it is the group of pixels which has similar characteristics. For N data points the algorithm finds the Parzen Density Estimate around each point and is given by the following equation,

$$P(x) = \frac{1}{2\pi\sigma^2 N} \sum_{i=1}^{N} e^{\frac{-\|x - xi\|^2}{2\sigma^2}} \qquad (1)$$

where $P(x)$ is the Parzen Density Estimate , x and x_i are data points and σ is the variance. Each point is connected to the nearest point in the feature space which has the higher density estimate. In this method each connection has the distance associated and the set of connections for all pixels forms the tree. The root of the tree is the point which has highest density estimate. Segmentation is obtained by choosing a threshold and break all the links in the tree with the distance greater than threshold [5].

Super-Pixel Edge Map: The next step is performing edge-detection on super-pixel map which is used to locate sharp discontinuities in an image. A number of edge

detector operator is available. Canny Edge detector is prefered because it produces better results and low error Rate. The process of Canny Edge detection algorithm is given as follows: Gaussian filter is used to smooth the image in order to remove the noise. The smoothed image is obtained as,

$$f_s(x, y) = G(x, y) * f(x, y) \qquad (2)$$

where,
$f_s(x, y)$ is the smoothed image,
$G(x, y)$ is the gaussian function,
$f(x, y)$ is the input image.

The Gaussian filter kernel of size (2k+1)×(2k+1) is given by,

$$H_{ij} = \frac{1}{2\pi\sigma^2} \exp\left(\frac{\exp - (i - k - 1)^2 + (j - k - 1)^2}{2\sigma^2} \right) \qquad (3)$$

where H_{ij} is Gaussian filter kernel and σ is the variance. The intensity gradients of the image are found by,

$$G = \sqrt{G_x^2 + G_y^2} \qquad (4)$$

$$\alpha(x, y) = \tan^{-1} G_x/G_y \qquad (5)$$

The intensity gradient in the horizontal direction (G_x) and vertical direction (G_y) and direction $\alpha(x, y)$ is computed in the above equation. The non-maximum suppression is used to get rid of spurious response to edge detection. Double threshold is used to detect the potential edges. Finally the edge is tracked by hysteresis which detects the edges finally, by suppressing all the other edges that are weak [6].

Convex Hull of Each Clump: From super-pixel edge map as input the convex hulls are obtained. There are two methods to compute the convex hull, to differentiate the object portions from background portions in an image. The Union Convex Hull method is used to find the hull of all foreground objects . This method treats all objects as a whole. The object Convex Hull method is used to find the convex hull of each connected component of image individually.

The output of which it contains the hull of each connected component. In overlapping cell segmentation the convex hull of each connected component of image individually is to be determined, so, object convex hull method is preferred [7].

Accurate Clump Boundaries:From convex hull the clump boundaries are obtained using GMM (Gaussian Mixture Model) and Level set Functions. For d dimension input, $x = (x^1, x^2, \ldots\ldots x^d)$ the Gaussian distribution of a vector is given by,

$$N(x \mid \mu, \Sigma) = \frac{1}{2(\pi)^{\frac{d}{2}} \sqrt{|\Sigma|}} \exp\left(\frac{-1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu) \right) \qquad (6)$$

where u is the mean and Σ is the covariance of the Gaussian. The probability density function given in a mixture of K Gaussians is given by,

$$p(x) = \sum_{j=1}^{k} W_j . N\left(x / \mu_j, \Sigma_j \right) \qquad (7)$$

where W_j is the prior probability (weight) of the jth Gaussian.

$$\sum_{j=1}^{k} Wj = 1 \text{ and } 0 \leq W_j \leq 1 \qquad (8)$$

The likelihood of $p(x/\theta)$ of the data with regard to the model parameters are maximized using the optimization of the following function.

$$\theta = \underset{\theta}{argmax}\, p(x/\theta) = \underset{\theta}{argmax} \sum_{i=1}^{N} p(x_i/\theta) \qquad (9)$$

To maximize the likelihood of the data, the Expectation maximization algorithm is used.From the cell clumps, the nuclei portion is detected and segmented using MSER (Maximally Stable Extremal) algorithm. The MSER method detects blobs in images. A series of thresholds are applied one for each gray scale level. The pixels below a given threshold are white. Those pixels which are above or equal are set as black, which results in black and white images. One extreme is all white and the other extreme is all black, between these regions blobs grow and merge. These regions are kept as features. With the help of cell segmentation and nuclei segmentation the complete cell segmentation is obtained [8].

Level Set Method of Segmentation of Overlapping Cells: The joint level set segmentation uses as input the initial segmentation obtained from scene segmentation. The overlapping cell segmentation is performed with the initial scene segmentation followed by detailed segmentation which involves several level set functions [9].

Consider N represents the number of cells which are detected, then the level set function is given by, $\{\varnothing_i\}_{i=1}^{|N|}$. The energy function to be minimized is given by the following equation,

$$\varepsilon\left(\{\varnothing_i\}_{i=1}^{|N|}\right) = \sum_{i=1}^{N}\varepsilon_u(\varnothing_i) + \sum_{i=1}^{N}\sum_{j\in N(i)}\varepsilon_b(\varnothing_i,\varnothing_j) \tag{10}$$

The first term in equation 10 represents the unary term and the second term represents the binary term. The main level set function is given by,

$$\frac{\partial\phi}{\partial t} = \nabla_\phi E_{image} + \lambda.\delta_\in(\varnothing).div\left(\frac{\nabla(\phi)}{|\nabla(\phi)|}\right) \tag{11}$$

In the above equation,

$\frac{\partial\phi}{\partial t}$ Represents the segment curve,

∇_ϕ is the Gradient Flow and E_{image} is Curvature Energy,

$\delta_\in(\phi)$ is the Kernel Density,

$div\left(\frac{\nabla(\phi)}{|\nabla(\phi)|}\right)$ is the Shape priority operators divergence.

Probability Density Estimation Based on Kernel:

The Probability Density Estimation based on kernel is given below,

$$p_{in}(z,\varnothing) = \int_\Omega \frac{K(z-l(x))H_\in(-\varnothing)}{H_\in(-\varnothing)}dx \tag{12}$$

$$p_{out(z,\varnothing)} = \int_\Omega \frac{K(z-l(x))H_\in(\varnothing)}{H_\in(\varnothing)}dx \tag{13}$$

where $K(z-l)(x)$ is the specified kernel and Heaviside Step Function [10] is given by,

$$H_\in(\varnothing) =$$

$$\begin{cases} 1, & \varnothing > \in \\ 0, & \varnothing < -\in \\ (1/(2))\left((1 + (\varnothing/\epsilon) + \frac{1}{\pi}\sin(\pi\varnothing/\epsilon)\right) & otherwise \end{cases} \tag{14}$$

and Kernel function is given by,

$$\delta_\in(\varnothing) = \begin{cases} 0 & \varnothing > \in \ and \ \varnothing < -\in, \\ (1/(2\,\epsilon))\,(1 + \cos(\pi\varnothing/\epsilon)), & otherwise \end{cases} \tag{15}$$

Level set method produces irregularities during the evolution which results in errors thereby affecting the performance, so, reinitialization is applied. The energy is designed in such a way that it achieves a minimum, is obtained by solving level set evolution. The penalty term is used to maintain the signed distance property. The level set evolution is obtained as the gradient flow which minimizes the energy function with distance regularization term. With the help of potential function, the distance term is defined such that the desired level set evolution has both forward and backward diffusion effect which helps to maintain the shape of level set function. Hence to maintain this, a signed distance function is maintained near the level set function [10]. The edge indicator function consists of a Gaussian Kernel and Standard deviation which helps to reduce noise. The advantages of Level set method is that they handle changes like merging in an efficient manner. These features make them suitable for image segmentation. The segmented image is helpful for further process like feature extraction and classification, which plays an essential role in medical diagnosis.
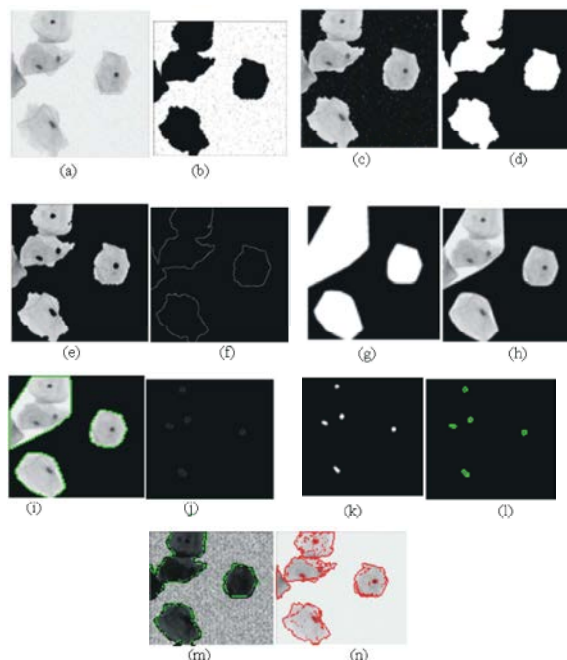
Experimental Results and Discussion

Fig. 2: (a) Input Cervical Cell Image (b) Cell Image is converted to Binary (c) Cell with Nuclei Image with unwanted contents (d) Cell with Nuclei Image with unwanted contents removed (d) Cell with Nuclei Image subtracted from nuclei only image (e) Canny Edge Detection (f) Convex hull as mask (g) The mask applied to input image (h) Boundaries are

found using built-in border function (i) Canny edge detection in nuclei image (j) Nuclei detection (k) Convex hull is applied to nuclei image (l) Convex Hull in nuclei regions (m) Clump Boundaries (n) Complete overlapped segmentation. The input to the overlapping cell segmentation given in fig. 2 is the cervical cell image of size 512x512 pixels and the images are synthetic images. This work is implemented using MATLAB2013A environment. From the results, it is inferred that, each and every cell within the clumps are segmented and individual nucleus are identified

CONCLUSION

The method of overlapping cervical cells addresses the problem of segmenting each individual cells nucleus and cytoplasm from a clump of cervical cells. The method is based on an optimization of several level set functions performs well on clumps. The cervical cells are segmented and individual nucleus is identified, thereby improving the classification of cytologic specimens. The performance for the threshold value 0.9 is calculated. The Dice Coefficient is measured and found to be 86.5%. The true positive rate is found to be 90.4%. The false negative proportion is found to be 9.6%. The false positive rate is found to be 27.0%.Segmentation of overlapping cervical cells can be extended for the classification of cytologic specimens. This proposed method can be extended to classification of affected regions, identification of disease and analysis of disease in cervical cells in efficient manner. The main advantage of the above technique improves the classification of cytologic specimens which finds an role in the area of medical diagnosis in identification of affected regions in diseased persons.

REFERENCES

1. Plissiti, M.E., C. Nikou and A. Charchanti, 2011. Automated detection of cell nuclei in Pap smear images using morphological reconstruction and clustering," IEEE Trans. Inf. Technol. Biomed., 15(2): 233-241.

2. Li, K., Z. Lu, W. Liu and J. Yin, 2012. "Cytoplasm and nucleus segmentation in cervical smear images using radiating GVF snake," Pattern Recognit., 45(4): 1255-1264.

3. Irshad, H., A. Veillard, L. Roux, D. Racoceanu, 2014. Methods for nuclei detection, segmentation and classification in digital histopathology: a review - current status and future potential". IEEE Reviews in Biomedical Engineering, 7: 97-114.

4. Chenyang Xu, X. and J. Prince, 1998. "Snakes, shapes and gradient flow", IEEE TIP, 7: 359-369.

5. Vedaldi and S. Soatto, 2008. "Quick shift and kernel methods for mode seeking," in Computer Vision (Lecture Notes in Computer Science), D. Forsyth, P. Torr and A. Zisserman, Eds. Heidelberg, Germany: Springer-Verlag, pp: 705-718.

6. Shrivakshan, G.T. and C. Chandrasekar, 2012. "A Comparison of various Edge Detection Techniques used in Image Processing", IJCSI International Journal of Computer Science Issues, 9, Issue 5, No 1, September 2012.

7. Dillencourt, Michael B., Hannan Samet and Markku Tamminen, 1992. "A general approach to connected-component labeling for arbitrary image representations". Journal of the ACM (JACM), 39(2): 253-280.

8. Matas, J., O. Chum, M. Urban and T. Pajdla, 2002. "Robust wide baseline stereo from maximally stable extremal regions," in Proc. BMVC, Cardiff, U.K., pp: 384-396.

9. Li, C. Xu, C. Gui and M.D. Fox, 2010. "Distance regularized level set evolution and its application to image segmentation, " IEEE Trans. Image Process., 19(12): 3243-3254.

10. Gao, X., B. Wang, D. Tao and X. Li, 2011. "A relay level set method for automatic image segmentation," IEEE Trans. Syst., Man, Cybern. B, Cybern., 41(2): 518-525.