Automated segmentation of cell nuclei in PAP smear images

M.E. Plissiti, A. Charchanti, O. Krikoni and D.I. Fotiadis

***Abstract*—In this paper an automated method for cell nucleus segmentation in PAP smear images is presented. The method combines the global knowledge about the cells and nuclei appearance and the local characteristics of the area of the nuclei, in order to achieve an accurate nucleus boundary. Filters and morphological operators in all three channels of a color image result in the determination of the locations of nuclei in the image, even in cases where cell overlapping occurs. The nucleus boundary is determined by a deformable model. The results are very promising, even when images with high degree of overlapping are used.**

I. Introduction

The automated detection and segmentation of cell nuclei in PAP smear images is one of the most interesting fields in cytological image analysis. The high degree of cell overlap, the presence of more than one nucleus in a cell and the lack of homogeneity in image intensity pose a challenge for any method in order to overcome the complexity of conventional cervical cell images and to achieve a correct segmentation. Furthermore, the accurate definition of the nucleus boundary is a crucial task because the nucleus is a very important structure within the cell and it presents significant changes when the cell is affected by a disease. The identification and quantification of these changes in the nucleus morphology and density contribute in the discrimination of normal and abnormal cells in PAP smear images.

The segmentation of nuclei in cytological images has been studied by several researchers [1-6]. An unsupervised nucleus segmentation method based on a water immersion algorithm and a dual active contour was presented in [1,2]. In addition, methods that take advantage of the expected similarity in nuclei shapes and they are based on Hough transform have also been introduced [3, 4]. A combination of the generalized Hough transform and deformable models is used in [5], in order to find a set of templates specific to nuclei shape. Furthermore, a fuzzy logic engine has been proposed for the cell nuclei segmentation [6].

Several segmentation techniques have been proposed and applied in tissue microscopic images [3, 6], where cells are extended in a cell grid and there is no overlapping. Some methods are also proposed for the segmentation of isolated cells in microscopic images [1, 2]. Our work aims at the definition of nuclei boundary in conventional PAP stained cervical cell images. These images contain many cluttered objects and high degree of cell overlapping which result in difficult identification of nuclei borders. Moreover, our method is fully automated, although it is based on deformable models. The initial estimation of the deformable contour is obtained automatically and no user interaction is required.

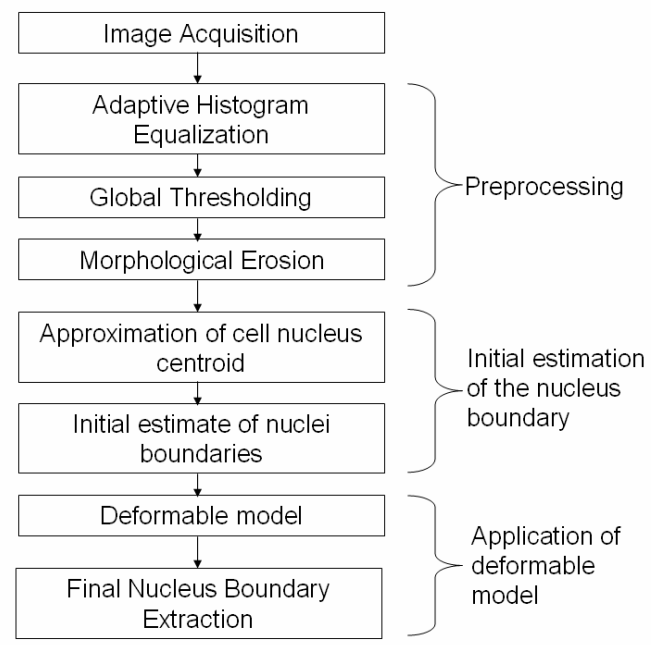


Fig 1: Schematic of the proposed method.

The proposed method is summarized in Fig.1. A preprocessing step is first applied for the exclusion of the background and the definition of the region of interest, which consists of the parts of the image where cell clusters are located. In each derived part, we locally search for the centroid of each nucleus and we apply morphological operators, which produce a rough boundary of the cytoplasm and the nucleus of every cell in the image. Based on this initial estimation of the nucleus boundary, an active contour is then used for the refinement of the detected boundary. The resulted contour corresponds to the accurate nucleus boundary.

II. Automatic method for the detection of cell nuclei

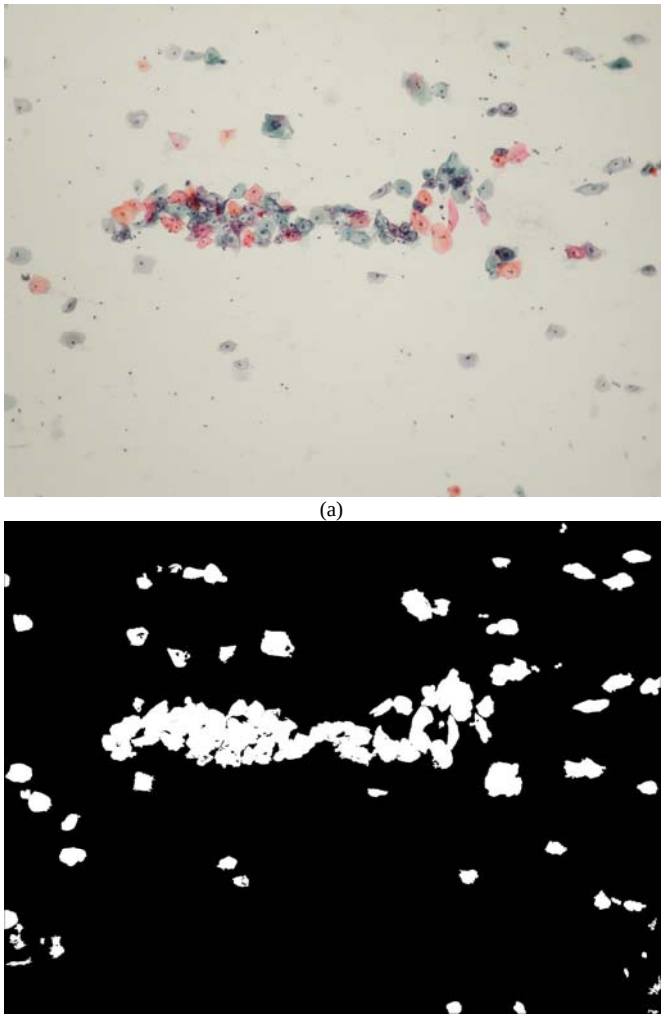
The images used in this work are conventional PAP stained cervical cell images, acquired through a CCD camera adapted to an optical microscope. We used a 10×

M.E. Plissiti is with the Department of Computer Science, Unit of Medical Technology and Intelligent Information Systems, University of Ioannina, GR 45110, Ioannina, Greece (e-mail: marina@cs.uoi.gr).

A. Charchanti is with the Department of Anatomy-Histology and Embryology, Medical School, University of Ioannina, 45110 Ioannina, Greece (tel: +30-2651-0-97832; e-mail: acharcha@cc.uoi.gr)

O. Krikoni is cytologist, member of Greek Cytological Society.

D.I. Fotiadis is with the Department of Computer Science, Unit of Medical Technology and Intelligent Information Systems, University of Ioannina, GR 45110, Ioannina, Greece and also with the Biomedical Research Institute – FORTH, GR 45110 Ioannina, Greece (tel: +30-2651-0-98803; fax: +30-2651-0-98889; e-mail: fotiadis@cs.uoi.gr).



(a)

(b)

Fig. 2: (a) The initial PAP smear image, and (b) the binary mask which is obtained in the preprocessing step.

magnification lens and images were stored in JPEG format having size 2048×1536 pixels.

A. Preprocessing

The preprocessing step is necessary for the background extraction and the definition of the regions of interest in the image. In order to reduce the searching area in the image, we first define the location of the cells (isolated or cell clusters). To do that, we create a binary image, containing the cell locations in the image. This image is obtained from the initial RGB image after the application of spatial filters and morphological operators in all three components of the color image. More specific we perform contrast-limited adaptive histogram equalization to the red, green and blue component of the image, which results in contrast enhancement and edge sharpening. We then find a global threshold for each image using the method proposed by Otsu [7], and using this threshold the intensity image is converted into a binary one. Finally all the three binary images are added using a logical OR operator. In this image we perform morphological erosion and we remove all objects with area smaller than a threshold T. This threshold is smaller than the area of an isolated cell and larger than the size of small objects in the image. In this way small artifacts are eliminated. The resulted binary image is used as a mask to indicate the regions which are covered by cell clusters in the initial image. In these regions the detection algorithm is then applied. The initial and the resulted binary image are shown in Fig. 2(a) and 2(b), respectively.

B. Cell Nuclei Detection

The parts of the image found in the preprocessing step contain either isolated cells or cell clusters. In the first case the detection of cell nuclei is a rather easy procedure, as the area of nucleus is darker than the cytoplasm. On the other hand, in cell clusters the high degree of cell overlap and the inhomogeneities in the nuclei intensity makes the detection of the nuclei difficult. Our approach to this problem consists of three steps: the approximation of the cell nucleus centroid, the initial estimation of the nucleus boundary and finally the application of a deformable model for the refinement of the boundary. These steps are described in detail below.

1) Approximation of cell nucleus centroid: Once we have found the regions of cell clusters, we search locally in each part of the image for the detection of the nuclei. For the definition of the searching area, we compute the bounding box for each white area in the binary image and then we define the corresponding subimage in the color image. Considering that nuclei are darker than the surrounding cytoplasm, in each subimage we search for intensity valleys in the red, green and blue channels of the color image. These valleys consist of pixels with intensity value lower than a specific threshold, and they are bounded by pixels whose intensity value is greater than this threshold. The areas of valleys found in the three images are joined using a logical OR operator. We then find the boundary of these valleys and we consider that the nuclei are enclosed in this boundary. At this point we do not pay attention on the accuracy of the derived boundary, but we are interested in the determination of the location of each nucleus. For that reason, we calculate the centroid of each area of the valleys, which indicates the existence of a nucleus at that pixel.

2) Initial estimation of nucleus boundary: Having found the centroid of each nucleus we automatically obtain an initial estimation for the nucleus boundary, which is necessary for the active contour. The pixel of the centroid is used as a starting point for the definition of some points at the circumference of the nucleus, which will construct a smooth elliptical curve for the initial estimation of the nucleus boundary. The image we use for the extraction of these points is obtained from the subtraction of two images. The first image is derived by the application of an averaging filter in the original image in order to avoid inhomogeneities in nuclei intensity. The second image is the result of successive erosions on the original image. With this way the

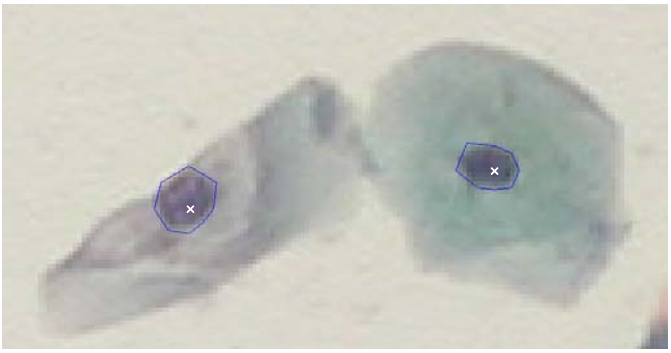


Fig 3: The nuclei markers and the initial estimation of the nuclei boundaries.

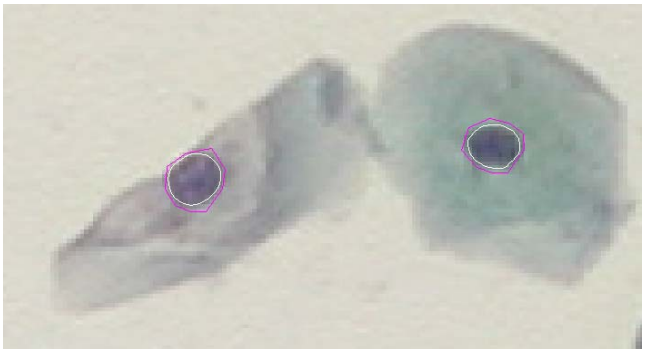


Fig 4: The initial nucleus boundary and the final contour obtained by the snake deformation.

nuclei become more pronounced (they become bigger and darker) while their area become smoother. We use a flat disk-shaped structuring element for the morphological operation. The result of the subtraction is an image which contains the boundaries of nuclei accentuated. We then obtain the grayscale image of the last image and we search in the neighborhood of cell centroids to find the points in the nuclei circumference. We use radial profiles in equal arc length intervals and we choose for the initial estimation of the nucleus points those pixels which have the highest intensity in a specific distance from the centroid. The extracted contour is an ellipse-like curve and it is used as the initial estimation for the deformable model, in order to find a more accurate estimation of the nucleus boundary. Fig. 3 shows the approximation of the nuclei centroid and the initial boundary estimation as well.

3) Application of a deformable model: For every nucleus found in the image we apply a deformable model [8] using as initial estimation the curve obtained at the previous step. The snake is defined using the points in the circumference of the nucleus as $v(s)=(x(s),y(s)), s\in[0,1]$ and it deforms under the influence of internal and external forces in order to minimize its energy functional:

$$E_{snake}=\int_0^1 \left(E_{int}\left(v(s)\right)+E_{ext}\left(v(s)\right)\right)\, ds \qquad (1)$$

As in most conventional snake models, the internal energy is a function of the first and second order derivatives of the curve, and can be expressed as:

$$E_{int}=\alpha(s)\ \left|v'(s)\right|^2+\beta(s)\ \left|v''(s)\right|^2 \qquad (2)$$

We use constant weighting parameters α,β. For the external energy we adopt the gradient vector flow (GVF), as it is described in [9]. The GVF field is the vector $w(x,y)=(u(x,y),\upsilon(x,y))$ which minimizes the energy functional

$$\mathrm{E}=\iint \mu\left(u_x^2+u_y^2+\upsilon_x^2+\upsilon_y^2\right)+\left|\nabla f\right|^2\left|w-\nabla f\right|^2 dxdy \qquad (3)$$

where $f(x,y)$ is an edge map of the initial image, which have larger intensity values near the image edges. In our work we use the Canny edge detector to find the edges in the image. The deformable model is quite flexible and it is attracted by the nucleus boundaries. Fig. 4 shows the initial estimation and the final contour obtained by the deformation of the model.

III. Results

Our method has been applied in several PAP smear images containing 462 cell nuclei, defined by an expert observer. The preprocessing step is a fast procedure which results in the reduction of the region of interest in the image, since it excludes all the background and leaves for further processing the parts of the image which contain isolated cells or cell clusters. The step for the detection of the cell nucleus centroid has shown that the resulted points of the image indicate the area of the nuclei, as it is confirmed by the expert observer. The method has successfully identified 94.15% of nuclei and its sensitivity is 95.94%. Then, the initial boundary we obtain is a rough approximation of the nucleus boundary, which becomes more accurate after the application of the deformable model.

IV. Discussion

The proposed method is fully automated and its application was performed without any observer interference. As it is verified by the results, the method is suitable for the detection and the correct segmentation of cell nuclei in PAP smear images. However, several parameters must be defined for the extraction of acceptable results. First of all, according to the threshold we use for the selection of the intensity valleys, a number of false positive pixels occur. This is a consequence of the complexity of PAP smear images, where the overlapping of the cells produces dark areas which have common characteristics with cell nuclei. For the choice of the threshold we performed several tests using different values, and we choose the threshold, which balance the accuracy of the method and the number of false positive pixels. Furthermore, the distance from the cell centroid, in order to select the points in the nucleus circumference, plays a crucial role in the determination of the initial contour estimation. Based on a priori knowledge of the normal and

abnormal nucleus size, and considering the nucleus as a circle, we use a radius which is slightly bigger than the radius of an abnormal cell nucleus. Finally, the parameters that enter in the computation of the energy of the snake have been defined empirically.

V. Conclusion

We propose a robust and accurate segmentation method for the detection of cell nuclei in conventional PAP smear images. The method is fully automated and it is suitable for cell images with high degree of cell overlapping, as it can detect not only the nuclei of isolated cells but also nuclei in cell clusters with high sensitivity (95.94%). It can be used as the basis for further processing of cell images, which is a tedious and time consuming procedure made by expert observers, in order to localize abnormal or malignant cells.

References

[1] P. Bamford, B. Lovell, “Unsupervised cell nucleus segmentation with active contours”, Signal Processing 71(2), pp. 203-213, 1998.

[2] Bamford P. B. Lovell “A water immersion algorithm for cytological image segmentation”, Proceedings of the APRS Image segmantation workshop, pp. 75-79, University of Technology Sydney, Sydney 1996.

[3] T. Mouroutis, S. J. Roberts, “Robust cell nuclei segmentation using statistical modelling”, IOP Bioimaging, 6, pp. 79-91, 1998.

[4] A. Garrido, N. Perez de la Blanca, “Applying deformable templates for cell image segmentation”, Pattern Recognition 33, pp. 821-832, 2000.

[5] Lee K.M., Street W.N., “Learning shapes for automatic image segmentation”, Proc. INFORMS-KORMS Conference, pp. 1461-1468, Seoul, Korea, June 2000.

[6] Begelman G., Gur E., Rivlin E., Rudzsky M., Zalevsky Z., “Cell nuclei segmentation using fuzzy logic engine”, International Conference on Image Processing, Vol. 5, pp 2937-2940, October 2004.

[7] Otsu, N., "A Threshold Selection Method from Gray-Level Histograms," IEEE Transactions on Systems, Man, and Cybernetics, Vol. 9, No. 1, pp. 62-66, 1979.

[8] M. Kass, A. Witkin, D. Terzopoulos, “Snakes: Active contour models”, International Journal of Computer Vision 1, 321-331, 1987.

[9] C. Xu, J. L. Prince, “Snakes, Shapes, and Gradient Vector Flow”, IEEE Transactions on Image Processing, 7(3), pp. 359-369, March 1998.