AUTOMATED CELL NUCLEUS SEGMENTATION USING IMPROVED SNAKE

Min Hu, Xijian Ping, Yihong Ding

College of Information Engineering, Information Engineering University, Zhengzhou, P.R.China

ABSTRACT

Accurate cell nucleus segmentation is crucial for the development of automated cytological cancer diagnosis system. This paper presents a novel cell nucleus segmentation method for esophageal cell image. Firstly the ultimate erosion is used to detect the localizations of nuclei. Then we use an improved active contour model to isolate each cell nucleus. A growing energy based on region similarity is added to the energy function to overcome the initialization problem of conventional snake. During energy minimization, the contour points are restricted to move along the radial directions, which reduces the computation cost. The presented method has been tested on a number of cell images obtained from esophageal smear slide and the results are encouraging. It has been shown that the proposed method performs well on both well-separated nuclei and some overlapped nuclei.

1. INTRODUCTION

Cell segmentation is a fundamental subject of quantitative analysis of cytological and histological images. Among various types of cell images, the ones obtained from biopsy are more difficult to segment, because of the diversity of the structures contained in the images, the intense variation of background caused by uneven staining, and overlapped cell clusters. As most of the malignant characteristics of cell are contained in cell nucleus, the isolation of cell nucleus is an important part of segmentation for this kind of cell images. There has been several cell nucleus segmentation techniques presented in the past studies [1]-[4].

As a global boundary based technique, active contour model or snake has been extensively studied and widely used in medical image analysis. Traditional snake has mainly two drawbacks. One is that the initial position of contour should be placed near the real boundary of the object to converge it to the real boundary. The other one is that the process of energy minimization is time-consuming. Aiming at these problems, various improvements on snake have been made [5]-[6].

Based on the shape characteristics of cell nucleus in H&E (haematoxylin and eosin) stained smear image, this paper proposes a new technique to extract the cell nucleus boundary. After detecting the seed point of each cell nucleus with ultimate erosion, an improved active contour model is built for each nucleus to track its boundary. The proposed snake imposes a growing energy based on sub-block intensity similarity to expand the contour from the nucleus center. To simplify the computation, the contour points of this snake are limited to move along the radial directions. The new segmentation method has been tested on a number of esophageal cell images with very encouraging results.

2. THE LOCALIZATIONS OF CELL NUCLEI

The cell images in this paper are microscopic images of esophageal smears stained using H&E (haematoxylin and eosin) method. For the sake of analytical simplicity, the original RGB images have been transformed into 8-bit grey-level images and then smoothed with a Gaussian filter.

The first step of our method is to locate each cell nucleus in the image by detecting a seed point. Fig. 1 shows a cell image with overlapped nuclei. The cell image mainly consists of nucleus regions, cytoplasm regions and background regions. The three kinds of regions distribute at different grey ranges, so they can be roughly divided with two thresholds.

We calculate the dual thresholds by iteration. The thresholds, denoted as T_1 and T_2, are firstly initialized to trisecting the grey range of the whole image. The loop begins from the following sentence. Thresholded with T_1 and T_2, the image is divided into three parts, denoted as R_1, R_2 and R_3. The average intensities of the three parts are respectively calculated as u_1, u_2 and u_3. Then T_1 and T_2 are renewed as $T_1=(u_1+u_2)/2$ and $T_2=(u_2+u_3)/2$. Repeat the loop until T_1 and T_2 are unchanged.

Fig. 2 shows the thresholding result of Fig. 1, where nucleus regions is painted in black, cytoplasm regions painted in grey and background regions painted in white. As some cell nuclei regions are connective, the conventional region tracking method cannot exactly search all the nucleus localizations. We perform

0-7803-8554-3/04/$20.00 ©2004 IEEE.

morphological ultimate erosion on the black regions with a disk to separate the connective regions. The erosion operation is executed successively until all the regions disappear. During this process the connective regions split into separate components. The collection of the components prior to being eliminated, make up the final components. Track all the regions in the final components and search their center points as the seeds points of cell nuclei. Fig. 3 represents the final components of Fig. 2, where two nucleus regions on the border of Fig. 2 are eliminated before ultimate erosion.

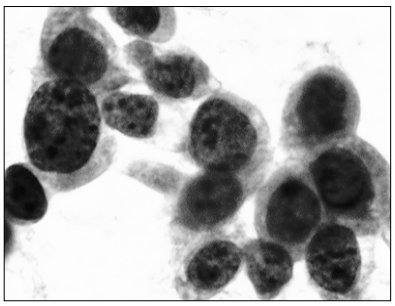

Fig.1 *Esophageal cell image*

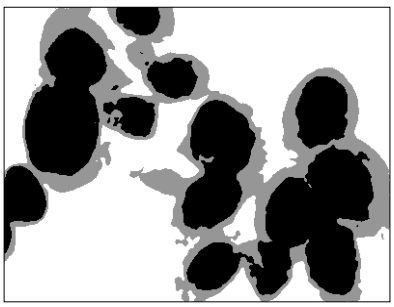

Fig.2 *Result of segmenting Fig. 1by dual thresholds*

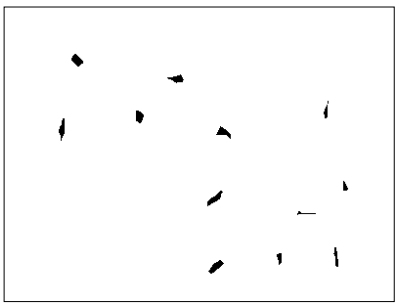

Fig.3 *Final components of ultimate erosion on Fig.2*

3. SEGMENTATION OF NUCLEI WITH SNAKE

3.1. Conventional active contour model

The base idea of snake is to search for a curve in the image where the weighted sum of internal energy and potential energy is minimum. The internal energy is defined by the curve itself to keep the model smooth during deformation. The potential energy is computed from the image data to move the curve toward an object boundary or other desired features within an image. The discrete energy function is represented as Eq.1, where $\boldsymbol{v}_i$ (i = 0, 1,…, n-1) denote contour points with n being the number of points. E_{cont}, E_{curv} and E_{image} denote respectively continuity internal energy, curvature internal energy and potential energy. α_i , β_i and γ_i are positive weighting constants.

$$E = \sum_{i=0}^{n-1}(\alpha_i E_{\text{cont}}(\boldsymbol{v}_i) + \beta_i E_{\text{curv}}(\boldsymbol{v}_i) + \gamma_i E_{\text{image}}(\boldsymbol{v}_i)) \qquad (1)$$

Various numerical implementations of snake have been proposed, such as the finite difference method, dynamic programming, simulated annealing, and greedy algorithm. In this paper, greedy algorithm is adopted to solve the proposed snake.

3.2. Improved active contour model

The prior knowledge of cell nuclei in the cell images shows that most of the nuclei have ellipse-like boundaries and there are no very large variations in the sizes of the nuclei. So it is feasible to describe the nucleus contour with discrete polar coordinates (ρ,θ) whose origin point is the detected seed point. The quantization step sizes of ρ and θ are ρ_s and $\theta_s(=2\pi/n)$ respectively. The nucleus contour are depicted with $v_i=(\rho_i,\theta_i)$ (i=0,1,2,…, n-1), where $\theta_i = i\times\theta_s$ and θ_i is invariable during deformation. This description simplifies the contour from 2-dimensional vector function to 1-dimensional scalar function.

The energy function of our model is presented as Eq.2, where a growing energy is added to the energy function of Eq.1.

$$E = \sum_{i=0}^{n-1}(\alpha_i E_{\text{cont}}(\boldsymbol{v}_i) + \beta_i E_{\text{curv}}(\boldsymbol{v}_i) + \gamma_i E_{\text{image}}(\boldsymbol{v}_i) + \varepsilon_i E_{\text{grow}}(\boldsymbol{v}_i)) \qquad (2)$$

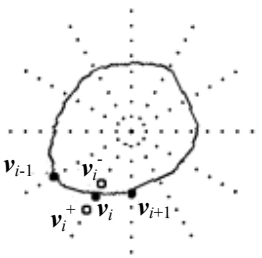

Fig.4 *Improved snake in polar coordinates*

The implementation of this model is based on greedy algorithm as follows. The first step is to initialize the contour points. Then for each point $\boldsymbol{v}_i$ (i=0,1,2,…,n-1), calculate the energies at $\boldsymbol{v}_i$, $\boldsymbol{v}_i^-$ and $\boldsymbol{v}_i^+$ and move $\boldsymbol{v}_i$ to the point with the minimum energy among $\boldsymbol{v}_i$, $\boldsymbol{v}_i^-$ and $\boldsymbol{v}_i^+$, where $\boldsymbol{v}_i^-$ and $\boldsymbol{v}_i^+$ are the two discrete points adjacent to $\boldsymbol{v}_i$ at the radial direction, as shown in Fig. 4. Iteratively

perform this operation until the number of moved contour points is sufficiently small or the iteration times exceeds a predefined threshold.

To give a full explanation of the energy function in Eq.2, we define $\Omega_i = \{\boldsymbol{v}_i^-, \boldsymbol{v}_i, \boldsymbol{v}_i^+\}$, where $\boldsymbol{v}_i^- = (\rho_i - \rho_s, \theta_i)$, $\boldsymbol{v}_i^+ = (\rho_i + \rho_s, \theta_i)$. The two contour points adjacent to $\boldsymbol{v}_i$ are denoted as $\boldsymbol{v}_{i+1} = (\rho_{i+1}, \theta_{i+1})$ and $\boldsymbol{v}_{i-1} = (\rho_{i-1}, \theta_{i-1})$ respectively. For any $\boldsymbol{v}_j = (\rho_j, \theta_j) \in \Omega_i$, the energy function is computed as the following.

- E_{cont}

Williams has defined $|\bar{d} - |\boldsymbol{v}_j - \boldsymbol{v}_{i-1}||$ as the continuity internal energy of $\boldsymbol{v}_j$ [6], where $\bar{d}$ is the average distance of contour points. In this paper, we enhance this energy by adding the radial distance difference.

$$E_{\text{cont}}(\boldsymbol{v}_j) = |\bar{d} - |\boldsymbol{v}_j - \boldsymbol{v}_{i-1}|| + |\bar{\rho} - |\rho_j - \rho_{i-1}|| \quad (\boldsymbol{v}_j \in \Omega_i) \quad (3)$$

where $\bar{d} = \sum |\boldsymbol{v}_t - \boldsymbol{v}_{t-1}| / n$, $\bar{\rho} = \sum |\rho_t - \rho_{t-1}| / n$.

The continuity energy is then normalized in Ω_i as

$$E_{\text{cont}}(\boldsymbol{v}_j) = E_{\text{cont}}(\boldsymbol{v}_j) / \max_{\boldsymbol{v}_j \in \Omega_i} E_{\text{cont}}(\boldsymbol{v}_j) \quad (\boldsymbol{v}_j \in \Omega_i) \quad (4)$$

- E_{curv}

Like the continuity energy, the proposed model defines the curvature energy by adding a second-order difference of radial distances of contour points to the curvature energy defined by Williams, shown as

$$E_{\text{curv}}(\boldsymbol{v}_j) = |\boldsymbol{v}_{i+1} - 2\boldsymbol{v}_j + \boldsymbol{v}_{i-1}|^2 + |\rho_{i+1} - 2\rho_j + \rho_{i-1}|^2 \quad (\boldsymbol{v}_j \in \Omega_i) \quad (5)$$

It is then normalized in Ω_i as

$$E_{\text{curv}}(\boldsymbol{v}_j) = E_{\text{curv}}(\boldsymbol{v}_j) / \max_{\boldsymbol{v}_j \in \Omega_i} E_{\text{curv}}(\boldsymbol{v}_j) \quad (\boldsymbol{v}_j \in \Omega_i) \quad (6)$$

- E_{image}

Generally the potential energy is designed to be in inverse proportion to the gradient of the point. In this paper, the contour points are designed to move along the radial directions. So we define the potential energy in inverse proportion to the difference between the average intensities of R discrete points outside $\boldsymbol{v}_j$ and R discrete points inside $\boldsymbol{v}_j$ at the radial direction, denoted as P_{image}

$$P_{\text{image}}(\boldsymbol{v}_j) = \frac{1}{R}\sum_{r=1}^{R} I(\rho_j + r \times \rho_s, \theta_j) - \frac{1}{R}\sum_{r=1}^{R} I(\rho_j - r \times \rho_s, \theta_j) \quad (\boldsymbol{v}_j \in \Omega_i) \quad (7)$$

where $I(\rho, \theta)$ is the grey-level intensity function of position (ρ, θ). The normalized potential energy is defined as

$$E_{\text{image}}(\boldsymbol{v}_j) = (P_{\min} - P_{\text{image}}(\boldsymbol{v}_j)) / (P_{\max} - P_{\min}) \quad (\boldsymbol{v}_j \in \Omega_i) \quad (8)$$

where $P_{\min} = \min_{\boldsymbol{v}_j \in \Omega_i} P_{\text{image}}(\boldsymbol{v}_j)$, $P_{\max} = \max_{\boldsymbol{v}_j \in \Omega_i} P_{\text{image}}(\boldsymbol{v}_j)$

In the grey-level cell image, the average intensity within cell nuclei is lower than that of the outer cytoplasm and background. So the value of P_{image} at the nucleus boundaries is larger. That is the corresponding E_{image} takes minimum value along the radial direction, which can push the contour point toward the nucleus boundary. The attraction range of the nucleus boundary is in proportion to the size of R. But the excessively large value of R may lead to wrong segmentation results by dragging the contour point to the border between cytoplasm and background.

- E_{grow}

The proposed model improves snake by designing a growing energy to expand the contour from the center of nucleus to its boundary. The growing energy is defined as

$$E_{\text{grow}}(\boldsymbol{v}_j) = \begin{cases} e & \text{if } \boldsymbol{v}_j = \boldsymbol{v}_i^+ \text{ and } |\bar{I}_{\boldsymbol{v}_j} - \bar{I}_{\text{origin}}| < T \\ 0 & \text{else} \end{cases} \quad (9)$$

$$\text{and } \bar{I}_{\boldsymbol{v}_j} = \frac{1}{k \times k}\sum_{\boldsymbol{v}_t \in \Psi_{\boldsymbol{v}_j}} I(\boldsymbol{v}_t),\ \bar{I}_{\text{origin}} = \frac{1}{k \times k}\sum_{\boldsymbol{v}_t \in \Psi_{\text{o}}} I(\boldsymbol{v}_t)$$

where e is a negative constant, and T is a threshold. $\Psi_{\boldsymbol{v}_j}$ and Ψ_{o} denote two sub-blocks of the image with their center points being respectively $\boldsymbol{v}_j$ and the seed point. The size of the sub-blocks is $k \times k$ (k is an odd number).

So the energy at $\boldsymbol{v}_i^+$ decreased when the sub-block at $\boldsymbol{v}_j$ is similar to the sub-block at the seed point in the intensity, which gives the priority to move the contour point outside, and this priority will disappear when the sub-block intensity contrast between the contour point and the nucleus center is great.

The absolute value of e needs to be chosen appropriately. If it is too small, the contour cannot overcome the shape restriction caused by the internal energy to expand outward. And if it is too large, the contour may get over the attraction of the potential energy to exceed the boundary.

4. RESULTS AND CONCLUSION

The initial positions of the contour points are set on a circle whose radius must be small enough to ensure the initial contour within the cell nucleus. In this paper, it is chosen to be 10. The constant n is set at 100 ($\theta_s = 0.02\pi$). The step ρ_s is set at 1. The sizes of R and k are 5. The weighting parameters are chosen to be constants as $\alpha_i = \beta_i = 0.5$, $\gamma_i = 1$ and $\varepsilon_i = 1$. The value of e is chosen to be –0.7 and T is 40. Fig. 5 gives the nucleus segmentation results of Fig. 1, where the seed point of each nucleus is marked by white dot and the nucleus boundary is drawn with white curve connecting all the contour points in turn.

Comparisons to conventional snake are made in Fig.6, where Fig.6(b) and Fig.6(e) are the results of Fig.6(a) and Fig.6(d) segmented using snake with conventional gradient-based potential energy and imposing no grow energy. Fig.6(c) and Fig.6(f) are the results obtained by the snake proposed in this paper. It is showed that the

boundary attraction range is broadened in our model.

For overlapped nuclei, a contour model is build separately for each nucleus. The iteration of each model is performed independently. The proposed method has acquired satisfactory results on some overlapped nuclei as shown in Fig.7. The segmentation performance is better for the nuclei whose overlapped parts remain more gradient information.

Conclusively, this paper presents a new cell nucleus segmentation method for esophageal cell image. Based on the seed point of the nucleus detected with ultimate erosion, an improved active contour model is build to extract the boundary of the nucleus. The presented model has the advantages of short computation time and broadened boundary attraction range. The contour need not be initialized near the real object boundary. The experiment results show that the proposed method has strong anti-noise and boundary detection abilities.

The proposed model is particularly appropriate to segment the objects that have low variations of curvature with high computation efficiency. To achieve more general application, improvement can be made by dynamically supplementing contour points where the curvature varies rapidly. The future work can be done in this direction.

5. REFERENCES

[1] Lassouaoui, N., Hamami, L., "Genetic algorithms and multifractal segmentation of cervical cell images", *Proceedings of IEEE-EURASIP Seventh International Symposium on Signal Processing and its Applications*, vol. 2, pp. 1-4, July 1-4, 2003.

[2] Tianzi Jiang, Faguo Yang, "An Evolutionary Tabu Search for Cell Image Segmentation", *IEEE Trans. on Systems, Man and Cybernetics*, vol. 32, pp. 675-678, 2002.

[3] T. Mouroutis, S. J. Roberts, and A. A. Bharath, "Robust Cell Nuclei Segmentation Using Statistical Modeling," *Bioimaging*, vol. 6, pp. 79-91, 1998.

[4] Pascal Bamford, Brian Lovell, "Unsupervised Cell Nucleus Segmentation with Active Contours," *Signal Processing*, vol. 71, pp. 203-213, 1998.

[5] Chenyang Xu, Jerry L. Prince, "Snakes, Shapes and Gradient Vector Flow," *IEEE Transaction on Image Processing*, vol. 7, pp. 359-369, 1998.

[6] D. Williams, M. Shab, "A Fast Algorithm for Active Contours and Curvature Estimation," *Computer Vision, Graphics and Image Processing: Image Understanding*, vol. 55, pp. 14-26, 1992.

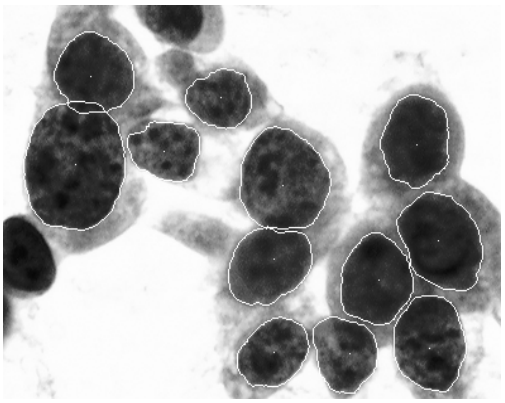

Fig.5 *Segmentation result of Fig.1*

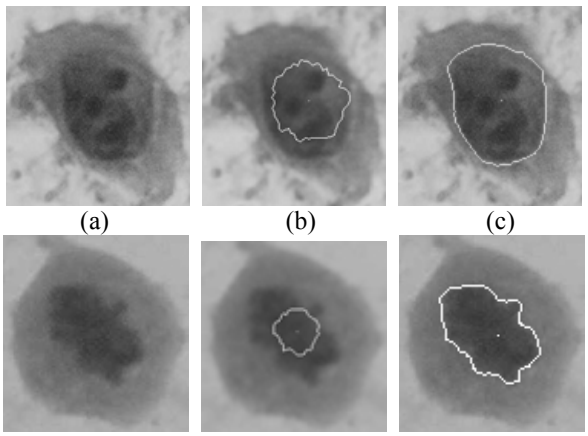

Fig. 6 *Segmenting cell nucleus with different models*

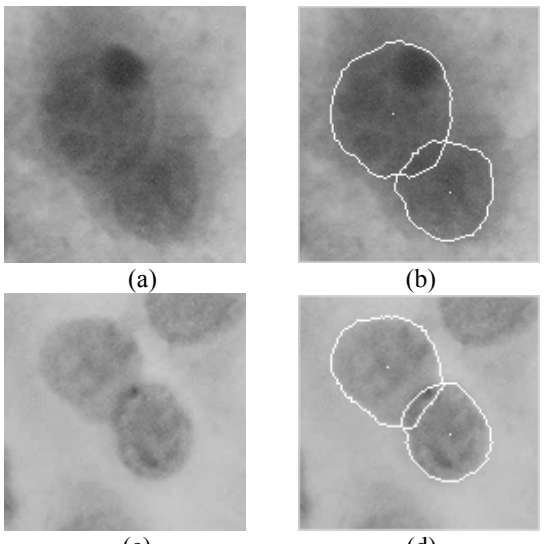

Fig.7 *Segmentation results of overlapped cell nuclei*