AN EVOLUTIONARY SNAKE ALGORITHM FOR THE SEGMENTATION OF NUCLEI IN HISTOPATHOLOGICAL IMAGES

M. A. Roula, A. Bouridane and F. Kurugollu

School of Computer Science, Queen's University of Belfast
18 Malone Road BT7 1NN Belfast, UK
m.roula, a.bouridane, f.kurugollu@qub.ac.uk

ABSTRACT

This paper addresses the problem of automatic segmentation of nuclei in histopathological images. A novel method, inspired from active contour models is proposed. An evolutionary based approach, which guarantees convergence to global minimum energies has been used to solve the combinatory optimization problem of snakes. The computation complexity, often associated with evolutionary approaches, has been reduced by short cutting the natural evolution step by the mean of replacing standard mutation with an oriented stochastic mutation process. Results have shown the efficiency of this method both in terms of accuracy and fast computation.

1. INTRODUCTION

The segmentation of nuclei is an important step in any automatic system for quantitative assessment of pathology specimens. This is because nuclear chromatin patterns provide important information on the cell state [1]. These patterns include morphological, texture, color and spectral features [2]. In order to measure these features, the nucleus needs to be accurately extracted from the background. However, this task can be very difficult for some types of cells such as those encountered in Epithelium layers that surrounds lumen glands in prostatic tissue sections. The identification and classification of these tissue sections is a crucial step for the diagnosis of prostate cancer, which remains one of the few cancer types, which still rely on biopsies analysis for conclusive diagnosis [1][2].

Substantial experimental work, which has widely been reported in the literature, has been achieved in the area of classification of nuclei. However, most of these studies rely on hand based segmentation which, in addition of providing a non-accurate segmentation, limits the possibilities of analyzing the large number of nuclei usually encountered in real samples and therefore limits its applicability in routine pathology.

A number of problems have to be addressed while attempting to automatically segment prostatic nuclei such as:

- nuclei may overlap
- the general appearance of nuclei can differ between cells, some nuclei are of type "Full" where they look like a mono-textured dark object, while others may look like an empty dark circle with discrete condensation of chromatin inside
- the shape of nuclei is sometimes not complete, and the incomplete part has to be interpolated from the complete ones, which is particularly difficult to quantify and implement in an automated system.

For this type of problems, active contour models or "snakes" [3][4][5] are powerful methods initially developed for constraint based image segmentation problems. They utilize *a priori* known information about the object in the segmentation process. Typically, a snake is a curve that attempts to fit the contour of an object by taking its known shape into consideration. To achieve this task an energy function E_{snake} is usually associated with the snake as follows:

$$E_{snake}= E_{internal}+ E_{external} \quad (1)$$

where: $E_{internal}$ is an internal energy, which depends on internal properties of the snake (e.g., roundness, elongation, size…etc) and $E_{external}$ is an external energy, which is measured from the image. The definition of these two terms depends on the problem at hand.

In practice snakes are not continuous curves but a set of points connected to each other by lines. Therefore, the aim is to find a combination of points that minimizes the above energy function. The snake algorithm is then reduced to a combinatory optimization problem. The optimization methods often used are: gradient descent, greedy algorithm and dynamic programming. The first two, although relatively easy to implement, do not guarantee the convergence to the global minimum. On the other hand, dynamic programming method guarantees the convergence to global minimum, but at the expense of a burdensome computation cost and large memory consumption.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

Histopathology applications require fast algorithms because, often, a large number of nuclei need be segmented at the same time. Moreover, and due the very nature of the images, the chances of falling into a local minimum are very high, because of the numerous DNA condensations inside the nuclei that can be mistaken for boundaries. Therefore, an algorithm that guarantees the global minimum and requires low computation resources at the same time is required.

To overcome this difficulty, we propose an evolutionary algorithm that searches the solution space in a stochastic manner. The search space is considerably reduced by defining a fitness function that takes null values for all points of the search space that do not satisfy strict eligibility criteria (to become contour points). Instead of being totally random, the mutation depends on this function so as to allow an oriented evolution and therefore to ensure a fast convergence of the algorithm.

This paper is organized as follows: section 2 describes the proposed algorithm. Results and analysis are discussed in section 3, while section 4 is concerned with the summary part.

2. THE EVOLUTIONARY SNAKES

The algorithm proposed in this paper belongs to the class of evolutionary algorithms because it uses random mutations and crossover mechanism as methods for the exploration of the solution space. However, it is not a genetic algorithm [6] in a sense that it does not use the natural selection process to select the best individual out of a population of solutions. Rather, the final population is *itself* the solution-snake. Each snake is composed of a number of radial lines from its center of gravity (a radial lines is a chromosome in the genetic analogy). The algorithm works by iteratively performing the mutation and crossover on these radial lines in a way that minimizes the global energy of the snake.

2.1 Radial snakes

The snake is defined by a central point *Cp* and a number N_R of radial lines. This type of snakes has proven to be particularly suitable for nuclei segmentation [4].

Each radial line stretches from a minimum distance (*DistMin*) to a maximum distance (*DistMax*) from the center *Cp*. We can then define $R_r(i)$ {i=1 to M=*DistMax*–*DistMin*} as a vector containing the grey levels of pixels inside a given radial line *r* where *M* is the number of pixels in each radial line. Similarly, $Grad_r(i)$ {i=1 to M}contains the gradient of the vector R_r.

Each radial line *r* has one contour point defined by index ic_r of the contour pixel in the radial line ($1<ic_r<M$}. The problem is to find the set {ic_I,… ic_M} that minimizes the global energy function of the snake *E*.

2.2 Ray fitness function

the selection of contour points ic_r is based on a fitness function F_r where each element of the $F_r(i)$ is associated with a point *i* in the radial line r. This gives us a measure of how fit the point *i* is to become to be contour point ic_r.

The fitness function $F_r(i)$ is defined in a way to implement the constraints of the problem.

Typically, the contours consist of points (pixels) which belong to the nucleus. This means that for a pixel to be considered as a contour point, its grey level value should be small while its gradient high. Also in order to have a convex shape the contour points should not be too far from the average distance of other contour points an especially the closer ones.

The fitness function is then defined as follows:

$$F_r(i) = \begin{cases} C_{th} + Grad(i) - R(i) & if\ (R(i) < C_{th}\ \&\ Grad(i) > C_{Gth}) \\ & \&\ (|i - i_{Av}| < C \times i_{std}) \\ \dfrac{(C_{th} + Grad(i) - R(i))}{|i - i_{Av}| / C \times i_{std}} & if\ (R(i) < C_{th}\ \&\ Grad(i) > C_{Gth}) \\ & \&\ (|i - i_{Av}| > C \times i_{std}) \\ 0 & else \end{cases} \tag{2}$$

where:

C_{th},$C_{Gth,}$,C are constants fixed experimentally.

ic_{ave} is the average *i* index of the contour points through all the radial lines {r=1 to r=N_R}.

ic_{std} is the standard deviation from this average.

According to the fitness function, the radial lines are divided into three areas:

- Area of null values: In order for a pixel in a radial line to be considered to become the next contour point it must have a grey level value R(i) smaller than the maximum grey level threshold C_{th} and its gradient greater than a gradient threshold C_{Gth}
- If the previous two conditions hold, then the fitness of a given pixel to become a contour point would depend on whether this pixel is within a minimum distance from the average distance of contour points from the center point *Cp*. If it is within this distance then the fitness will be proportional to the pixel gradient minus its grey level value (the darker the pixel and sharper the edge the higher fitness value.
- If the pixel is not within this minimum distance, then the fitness is sensibly reduced after being divided by a factor, which depends on the actual distance. This is the internal constraint which guarantees that the shape of the snake will tend to be convex.

2.3 The energy function

The energy function of a snake to be minimized is simply the inverse of the sum of fitness values of contour points:

$$E = \left(\sum_{r=1}^{N_R} F_r(ic_r) \right)^{-1} \tag{3}$$

By changing the location of a one contour point the fitness values of all other points will be affected since ic_{ave} is and ic_{std} will also be modified as shown by equation (2). Therefore, an efficient multivariate search technique is required to explore the $M^{(N_R)}$ solutions space. We propose an oriented mutation method to overcome this problem as described below.

2.4 Oriented stochastic mutation

Mutation is the mechanism by which the contour points are moved to a new position after each generation. This process, although stochastic, is not totally random. Only the points having a non-zero positive value of the fitness function can become contour points. The mutation is only applied on radial lines whose contour points are farther than $C \times ic_{std}$ from the average contour point ic_{ave} of the snake (considered to be too different from the other contour points to be a valid contour point). A contour point ic_r will be mutated to a random position with a probability depending on its fitness function (See Fig. 1). Since the fitness function is a non-continuous function of i and with a lot of null elements, the mutated points will move only to areas which are likely to be contour points and as such an efficient search is achieved.

The probability of pixel i to become the next contour point ic_r^{n+1} on a radial line r is given as follows:

$$P\left(ic_r^{n+1} = i\right) = \frac{F_r(i)\,\Phi_r(i)}{\sum_{j=1}^{M} \left(F_r(j)\,\Phi_r(j)\right)} \tag{4}$$

and

$$\Phi_r(i) = \begin{cases} 1 & if\,(ic_r^n - ic_{Ave})(ic_r^n - i) > 0 \\ 0 & if\,(ic_r^n - ic_{Ave})(ic_r^n - i) < 0 \end{cases} \tag{5}$$

where n is the generation index, $\Phi_r(i)$ is a directional constraint that causes the mutation to move always towards the average value of the contour points (and not in the opposite direction). This causes the snake to be more compact and smooth.

The mutation jumps allow the diversity in the solutions obtained and also prevent the solution to converge to a local energy minimum that can happen in the case of a continuous movement of contour points such as in a standard active contour model [3].

By not being totally random, the oriented stochastic mutation presented in this paper replaces the natural selection process. The computation involved in the calculation of the fitness function for each ray and each generation is largely compensated by the gain in computation time due to the removal of the natural selection process because there is no need to create a population of solutions.

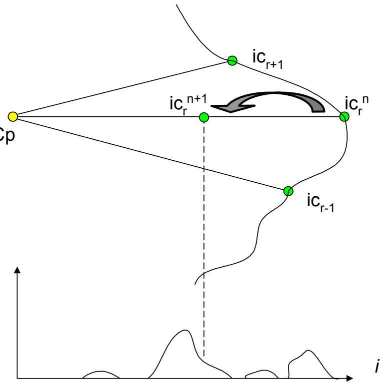

Figure 1:
Contour points mutation.

2.5 Crossover

On the other hand, the crossover brings regularity into the position of contour points. after crossover, the new position of a given contour point will be at a distance from the snake's center Cp, which is the average distance of the contour points of adjacent radial lines (see Fig.2). The crossover is applied randomly on snake radial lines with a fixed probability P_c. It has the effect of making the shape more regular and smooth. It also interpolates areas where the nucleus boundaries are not complete.

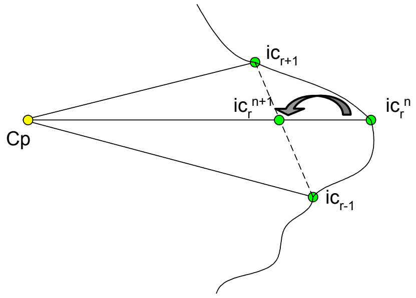

Figure 2:
Contour points Crossover.

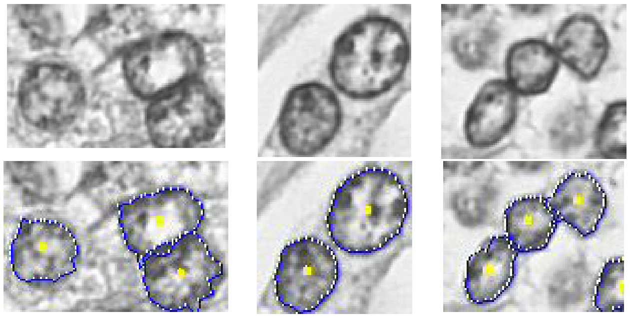

Figure 3. Segmented nuclei using evolutionary snakes

3. RESULTS

The initial central points of the snakes are set either manually -by a mouse click at any point inside the nuclei- or by using an automatic method, such as the one given in [Mak99]. However, it is worth mentioning that the performance of the automatic method tends to degrade with large numbers of nuclei. As such, the analysis has been carried out using the manual method, so as to assess the performance of the proposed algorithm independently of the accurate location of the initial central points. The parameters of the algorithm were set as follows: C_{th}=180, C_{Gth}=10, C =1.5, N_r=64 or 32, and M=40. Generally, the algorithm works well as most nuclei are clearly and accurately segmented. A good performance is obtained for separated, adjacent or slightly overlapping nuclei as shown in Figure 3. In most cases, the nuclei segmentation occurs within a fraction of second after a mouse-click, which makes the technique suitable for use in routine pathology. This is due to the algorithm's fast convergence, usually taking less than 15 generations. This algorithm was applied for nuclei identification in prostatic tissue samples which have three main types of nuclei: Benin Prostatic Hyperplasia BPH, Prostatic Intraepithelial Neoplasia (PIN) and Prostatic Carcinoma PCa [1] (in increasing order of malignancy).

Results performed on a number of 150 representative nuclei from each class are summarized in Table 1.The following criteria were used for segmentation assessment:

- *Perfect segmentation*: All snake contour points are actual boundaries of nuclei.
- *Good overall segmentation*: more than 90% of snake contour points are actual boundaries of nuclei.
- *Poor segmentation*: between 70% and 90% of snake contour points are actual boundaries of nuclei.

It is worth mentioning that for all nuclei (including the poorly segmented), more than 70% of contour point were accurately segmented. This algorithm however, performs less well when the scene contains degenerating nuclei, and when nuclei overlap, because pixel boundaries are not defined even by human pathological assessment.

	Good segmentation			Poor segmentation	
	Perfect segmentation	Good overall segmentation	%		%
BPH	34	12	**92%**	4	**8%**
PIN	33	11	**88%**	6	**12%**
PCa	28	15	**86%**	7	**14%**
Overall	**95**	**38**	**89%**	**17**	**11%**

Table 1 Segmentation results

4. CONCLUSION

This paper has presented a novel algorithm for the segmentation of nuclei in pathology imaging. The main contribution is based on the proposed oriented stochastic mutation, which uses the constraints of the problem to bypass the inherently slow natural selection process. Its main feature is its suitability for real quantitative pathology applications, due to the fast computation time.

The computation involved in the calculation of the fitness function for each radial line is largely compensated by the gain in computation time due to the removal of the natural selection process. Currently, work is underway to enhance the algorithm performance by introducing a repulsion constraint and competition between adjacent snakes to overcome the problem of overlapping snakes.

5. REFERENCES

[1] F. K. Mostofi, I. A. Sesterhe and C. J. Davis, 'A Pathologist's view of Prostatic Carcinoma", Cancer,71, pp. 906-932, 1993.

[2] M. A. Roula, J. Diamond, A. Bouridane, Paul Miller and A. Amira, "A Multispectral Computer Vision System for Automatic Grading of Prostatic Neoplasia," IEEE International Symposium on Biomedical Imaging, ISBN:0-7803-7585-8, pp.193-196. 2002.

[3] M.Kass, A. Witkin, and D.Terzopoulos, "Snakes: active contour models", International Journal of Computer Vision, vol. 1pp. 321-331, 1988.

[4] M. Takahashi and M. Nakano "A novel segmentation method for nuclei in histopathologic images", European Biological and Medical Conference EMBEC 2002, vol. 2, pp.960-961.

[5] S. Heynen and J. H. Price. "Evaluation of SNAKES for an Objective Image Segmentation Standard on Cell Nuclei", Int. Soc. Analyt. Cytol. XIX Int. Congress , Feb-Mar, 1998.

[6] D. E. Goldberg, *Genetic algorithms*. Addison-Wesley, 1989.

[7] J. Makita and M. Takahashi, "Extraction of position and shape of nuclei from hepathocellular images", National convention of information processing society of Japan, 4W-4, 1999.