

A Parallel Genetic Algorithm for Cell Image Segmentation

Tianzi Jiang^{1,2}, Faguo Yang³, and Yong Fan⁴

*National Laboratory of Pattern Recognition, Institute of Automation
Chinese Academy of Sciences, Beijing 100080, P. R. China*

David J. Evans⁵

*Department of Computing, Nottingham Trent University
Nottingham, NG1 4BU, UK*

Abstract

In this paper, we propose a parallel genetic algorithm for cell image segmentation under severe noise. Our contribution aims at overcoming the drawback of the slow convenience of the traditional genetic algorithm, which was used in our previous work. A priori knowledge about cell shape is incorporated in our method. That is, an elliptical cell contour model is introduced to describe the boundary of the cell. We firstly obtain the gradient image using Canny edge detector; and then use kernel-based dynamic clustering to find out the image points that have a high probability belonging to each cell. Finally a parallel genetic algorithm is used to adjust the parameters of the cell contour model to find a best matching. The segmentation results of noisy human thyroid and small intestine cell images demonstrate that the proposed method is very successful in segmenting images of elliptically shaped cells.

1 Introduction

Pathologists often make diagnostic decision by observing the specimen cells and the geometric parameters of the cell such as the area, radius and the circumference [1]. It is very useful using computer-aided system to accurately

¹ This work was partially supported by Hundred Talents Programs of the Chinese Academy of Sciences, the Natural Science Foundation of China, Grant No. 697908001, and State Commission of Science and Technology of China, Grant No. G1998030503.

² Email: jiangtz@nlpr.ia.ac.cn

³ Email: fgyang@nlpr.ia.ac.cn

⁴ Email: yfan@nlpr.ia.ac.cn; yong.fan@ntu.ac.uk

⁵ Email: dj.evans@ntu.ac.uk

measure the geometric parameters of the cell. To get the parameters, accurate segmentation of cell images is required.

In the past years, many methods for the segmentation of cell images have been presented [1-7]. These methods include region-based methods, threshold-based methods and so on. Region-based methods employ region growing and region splitting and merging to segment the image. Threshold-based method is a common segmentation scheme, which can be regarded as the pixel classification. A feature value such as gray level is associated with each pixel. The value is compared to the threshold to classify a pixel into an object or the background. The threshold can be obtained automatically according to the histogram of the image [8, 9]. These methods mainly use the gray information of an image to segment the image and do not use any priori knowledge of an image. To some simple images without noise or with low noise, these methods may work very well. However, cell images are generally complex and have a lot of noise. It is well known that microscopic images have a short range of gray levels; There are many cluttered objects that make microscopic images segmentation difficult; Traditional staining techniques introduce a lot of inhomogeneity into the microscopic images, where not all of the parts of the same tissue are equally stained. In the presence of noise, clutter and occlusion, the segmentation of the cell image is an ill-posed problem [10]. But the segmentation performance can be greatly improved by incorporating a priori knowledge about the specific type images being segmented. In fact, one of the most challenging issues in medical image segmentation is to extend traditional approaches to segmentation and object classification in order to include shape information rather than merely image intensity.

In this paper, we not only use the edge information but also the shape information of the cell images to accurately segment the cell images. We also apply a parallel genetic algorithm to segment the cell images. Our method can be concluded as follows: firstly use Canny edge detector to detect the possible edges of the cell. Secondly locate the position of the cells approximately and find out the image points, which most likely belong to each cell boundary. Finally construct a cell contour model characterized by five parameters to accurately detect the cell contour and to eliminate the influence of the image noise. We can accurately describe the cell contour by adjusting the parameters of the cell boundary model. So the image segmentation problem is finally transformed to an optimum problem. A sequential genetic algorithm has been used this issue [15]. The goal of this paper is to parallelize the traditional genetic algorithm to improve the convergence speed of the algorithm. Compared with the sequential genetic algorithm, parallel genetic algorithms can be run on a distributed environment to reduce the computational time, and parallel genetic algorithms represent a better model of nature. So, global optimizing solution is more easily to be found out.

This paper is arranged as follows. Section 2 introduces the mathematical model used to describe the boundary of a cell. Section 3 describes a method

by which we can obtain the edge information and the possible image points belonging to each cell boundary. In section 4, we introduce the parallel genetic algorithm optimizing the parameters of the mathematical model to match the cell contour. Section 5 devotes to the experimental results. Section 6 is a conclusion and the future work.

2 Contour Model of Cell Image

Most of the cells in the human body usually have ellipse-like boundaries as shown in Fig 1 (a) and Fig 2 (a), from which we can see that the gray levels of the cells are lower than those of the background. Moreover, there is a lot of noise, clutter, and occlusion in the image. In order to get the geometric parameters of the cells, we must segment the cells accurately from the background. The boundaries of the cells are of ellipse-like shapes. In order to improve the performance of boundary detection methods, a priori knowledge of the boundary shape always needs to be incorporated into such a method. If a cell boundary can be analytically found out, we can easily solve the segmentation problem. Therefore, the cell image segmentation can be cast as a parametric optimization problem. If the model parameters of a cell boundary are determined, we can reconstruct the segmented image, which is used to extract meaningful geometric parameters for pathologists. In order to use a priori knowledge that the cells have an ellipse-like boundary, we use an ellipse equation to describe the boundary of a cell. The mathematical model of the cell boundary can be described as follows:

$$(1) \quad \frac{[(x - x_0) \cos \theta + (y - y_0) \sin \theta]^2}{a^2} + \frac{[(x - x_0) \sin \theta + (y - y_0) \cos \theta]^2}{b^2} = 1$$

The Eq. (1) denotes an ellipse in the (x, y) domain, and x_0 and y_0 determines the center of the ellipse, θ indicates the orientation of the ellipse, a and b determines the size of the ellipse. As we know, an ellipse can be determined by five different points, which are on the ellipse. Two parameters are needed to determine the location (x_0, y_0) , and three are needed to determine its size and orientation (a, b, θ) . So, our segmentation task is cast to choose five different points from the image points to determine an ellipse matching the cell boundary best.

3 A Novel Approach to Detection of Image Points in a Cell

3.1 Localization of Cells

Traditionally, edge-based segmentation has been divided into two independent stages: edge detection and edge linking. Under Marrs paradigm, boundary extraction is conventionally treated as a set of independent problems, where each of them has input information, a method to process them and output

information. This one-way flow of the information may yield wrong results because of the error propagation. So we adopt a different scheme in this paper. In our scheme, we first find those image points that have a very high probability belonging to a cell. Then we can make use of the image points to find out the approximate location of each cell. The location information is relatively stable. This process can be viewed as getting high semantic information from low semantic information. After getting the approximate location of each cell, we take advantage of this high level semantic information (the approximate location of the cell) to process the low semantic information further to determine whether other image points belong to the cell boundary or not. In our scheme, information can flow in both directions: from low semantic level to high semantic level and vice versa. So, error propagation is effectively suppressed. We use Canny's edge detector [11], a well-known algorithm, to detect the image edges. Because of the influence of the noise, there are a lot of false edges in the gradient image (see Fig 1. (b) and Fig 2 (b)). Some of these edge points are connected. From the original cell images (see Fig.1 (a) and Fig. 2 (a)), we can see that the variation of the gray levels in the location near the cell boundary is immense. So the edges formed due to image noise have small number of connected image points. We can use a threshold to decide whether an edge indicates a cell boundary. If the number of the connected image points is greater than the threshold, we think the edge represents a cell boundary. Otherwise, we think the edge is formed due to image noise. The threshold can be selected by experience.

3.2 Kernel-Based Dynamic Clustering for Detection of Image Points in Cell

After detecting the approximate location of the cells, we use kernel-based dynamic clustering method to find out the image points possibly belonging to each cell. Kernel-based dynamic clustering method is a no-monitor clustering method. In this method, a kernel k_j , which represents a cluster, is defined. A kernel can be a function, an image point sets or other models. In order to determine whether a sample point belongs to a cluster, a measurement $\Delta(y, k_j)$, which describes the similarity between sample point y and cluster k_j , is also defined. The steps involved in the method are as follows:

- (i) Determine an initial kernel k_j for each cluster.
- (ii) For all sample points, follow the following rules to classify them: If $\Delta(y, k_j) = \min_l \Delta(y, k_l)$, and $\Delta(y, k_j) < t$, then $y \in \Gamma_j$, where y denotes a sample point, t is a threshold, and Γ_j represents the j th cluster.
- (iii) Update the kernel k_j . If all of the kernel do not change again, stop the process, otherwise return to the step 2.

Since the cell has an ellipse-like boundary, we use Gauss function as the kernel, which can be formulated as

$$(2) \quad K_j(y) = \frac{1}{2\pi\|\Omega_j\|} \exp \frac{1}{2}(y - m_j)\Omega_j^{-1}(y - m_j),$$

where m_j is the mean value of the samples, Ω_j is the covariance matrix.

The similarity between sample point y and cluster j is defined as follows:

$$(3) \quad \Delta(y, k_j) = \frac{1}{2}(y - m_j)\Omega_j^{-1}(y - m_j) + \frac{1}{2}\log\|\Omega_j\|.$$

Using the above method, we can find out the image points that have a high probability belonging to each cell respectively. After the image points possibly belonging to a cell are detected, we can search an ellipse that best matching the cell boundary in a relatively small space. So, the time spent in finding the solution is decreased greatly.

4 Extracting Cell Boundary Using a Parallel Genetic Algorithm

4.1 Genetic Algorithms

Genetic algorithm as the most favorite optimization method for ill-posed and nonlinear object function due to its excellent global optimization ability has been heavily studied [12,13,16]. As a relatively novel member of the genetic algorithm family, Parallel Genetic Algorithm has demonstrated to be more successful in the optimization than classical genetic algorithm [16]. By incorporating migration operation, parallel genetic algorithm makes itself more nature where several isolated subpopulations self-evolve in parallel, periodically exchanging their best individuals by migration with the neighboring subpopulations. The basic algorithm for parallel genetic algorithm can be described as below:

STEP 1: Define a suitable representation and genetic operators, generate randomly a population of candidate solutions and partition it into several subpopulations; decide a migration strategy for share individuals between the subpopulations.

STEP 2: Each subpopulation executes the step 3 and 4.

STEP 3: Self-evolve based on the chosen genetic operators: selection, crossover, mutation, local hill-climbing.

STEP 4: Send the best individuals to the neighboring subpopulations based on the migration strategy, receive their best ones and replace the worst ones of the subpopulation.

STEP 5: Determine whether the stopping criteria are satisfied. If satisfied, stop the iteration; otherwise go step 2.

According the above basic algorithm, different migration strategies and different strategies of subpopulations self-evolution produce different instantiations for parallel genetic algorithm. For the strategy of subpopulation self-evolution, we can apply the successful and popular accepted genetic algorithm. As to the migration strategy, the most popular models are the island model and the stepping stone model [17]. In the island model, best individuals are allowed to

be sent to any other subpopulation, i.e. all subpopulations are neighbors. In a sense, the island model represents a better model of nature at the cost of a high communication overhead is introduced in it. In order to reduce communication overhead, the stepping stone model defines a close chain neighborhood structure in which any subpopulation is merely able to share individuals with its two neighbors. Below, we present a parallel genetic algorithm for extraction of cell boundary with a different migration strategy.

4.2 Extraction of Cell Boundary with Parallel Genetic Algorithm

The input information is N possible edge points of the cell. Each of these image points has an associated index, which is a number from 1 to N . We can select five of them to represent an ellipse. Let $I = (I_1, \dots, I_5)$ denote the index of the five image points selected. For the sake of convenience, we assume that $I_i < I_j$ for $i < j, i, j = 1, \dots, 5$. We define an object function $f(I)$ as follows [14]:

$$(4) \quad f(I) = \sum (r_{I_i}^2),$$

where s is the step function: $s = 1$ if r_j is greater than or equal to the template width, and $s = 0$ otherwise. The objective function counts the number of points within a fixed distance of the ellipse. Let P denotes the population, which contains M chromosomes $I_1, I_2 \dots I_M$. Let L denotes the size of the population. Let NUM represents the number of the generations. Let N_r denotes the number of chromosomes replaced each generation. For convenience, let N_r be an even number. The parallel genetic algorithm based method for cell boundary extracting can be described as follows:

Part one: Parallelization

Step 1: Parameters Setting: The number of sub-populations N_{pop} , the number of individuals of a sub-population N_{size} need to be set.

Step 2: Send initialization messages to sub-populations

Step 3: Get population from sub-populations

Step 4: Send step messages to sub-populations

Step 5: Get population from sub-populations and determine whether repeat or stop, if repeat, continue step 6.

Step 6: Randomly set how many sub-populations will send individuals to others and how many best individuals will be sent. Our migration strategy is different from the island model and the stepping stone model. Who and how many sub-populations will migrate their best individuals to others are randomly determined. This strategy is a better compromise between the island model and the stepping stone model, which not only reduces the communication but also makes the parallel genetic algorithm behavior more close to nature.

Step 7: Migration between the selected sub-populations and go to step 4.

Part two: evolution in sub-populations

Step 1: Get Initialization Message: To initialize each chromosome, randomly generate five integers between 1 and N (N is the number of the possible edge image points) as the index of five image points. We repeat the process for L times to initialize the whole population. Let $k = 1$, where k is the iteration step indices.

Step 2: Get Step Message:

- (i) Selection and crossover: We use the fitness proportional model to select chromosomes to reproduce offspring. To get N_r children, we should select $N_r/2$ mother and dad respectively. Each pair of two parents has two children. The crossover operator is described as follows. Randomly generate an integer m between 1 and 5; Exchange the genes that have the index m of the two chromosomes selected as parents. So, N_r new chromosomes are reproduced.
- (ii) Mutation: For all of the newly generated chromosomes, according to a mutation probability to mutate them. The mutation method are described as follows. Randomly generate an integer m between 1 and 5 and another integer n between 1 and N . Replace the gene having the index m of the chromosome by the integer n .
- (iii) Replacement scheme: We use the newly generated N_r chromosomes to replace those having a low fitness. If the stop criterion (see Remark) is satisfied, stop the iteration; otherwise, go to (i).

Remark 1 *Firstly, we compute the fitness of each chromosome and the fitness can be viewed as a distribution function. Then we compute the entropy of the distribution function. When the entropy does not change again, we think the genetic algorithm converged. If the number of generation N is reached, we also stop the iteration.*

Step 3: Get Migration Message: Send emigrations from selected sub-populations to destinations, and delete the equivalent number of best individuals in the destination sub-populations.

5 Experimental Results

In this section, we present our experimental results on segmenting the cell images of hypothyroid and small intestine and make a comparison with the histogram-based methods.

The histogram-based method can be described as follows: (1) Calculate the histogram of the image to be segmented. (2) According to the histogram of the image, try to find a proper threshold to segment the image. (3) If the gray level of a pixel is smaller or equal to the threshold, the pixel will be classified to the object cluster; otherwise the pixel will be classified into the

background. The threshold can be found out through a minimal error method or a maximum entropy method [8,9].

In our experiment, we let the crossover probability $P_c = 0.6$, the mutation probability $P_m = 0.1$, the number of generation $NUM = 250$, the number of sub-population $N_{pop} = 10$, and the sub-population size $N_{size} = 100$. Fig.1 (c) and Fig. 2 (c) are our experimental results on segmenting the cell images of hypothyroid and small intestine respectively. The edge images are obtained using Canny's edge detector. The parameters used in Canny's edge detector are as follows: The higher threshold is 0.9 and the lower threshold is 0.8; the covariance is 1.6.

The experimental results with the threshold methods are also given out in Fig.1 (d) and Fig. 2 (d). From Fig 1 and Fig 2, we can see the advantages of our method compared to the histogram-based methods. Our approach has a strong ability to suppress noise. As long as the gray levels near the cell boundary differ a few, our method can actually detect the cell boundary and complete the segmentation task well. When two cells are located very closely, the histogram-based method cannot distinguish them, but our method can extract them successfully. So our approach has some ability to resist occlusion of the cell boundary. Our algorithm is coded in C language with MPICH in LINUX environment and tested on a 5-PC-Cluster. Because we firstly use kernel-based dynamic clustering method to find out which image points possibly belonging to each cell, the searching space is reduced greatly and the time spent in finding the optimum ellipse to match the cell boundary is decreased accordingly.

6 Conclusion and Future Work

In this paper, we have proposed a novel approach to cell image segmentation under severe noise conditions by combining kernel-based dynamic clustering and a parallel genetic algorithm. In our algorithm, we make use of not only the edge information but also the shape information of the cell contour that the cell boundary has an ellipse-like shape. Due to a priori knowledge of the cell boundary is incorporated in our approach; our method has a high ability to resist noise. Image points that possibly belonging to a cell are firstly determined using a kernel-based dynamic clustering method, so the search space is sharply reduced and the time used to optimize the objective function is decreased accordingly. By using parallel genetic algorithm as optimizer, we achieved not only a more accurate solution but also a speedup because of several processors sharing the population evaluations. The results indicate a promising direction for further research into automatic initialization, which is especially important for designing automatic algorithm in biomedical applications.

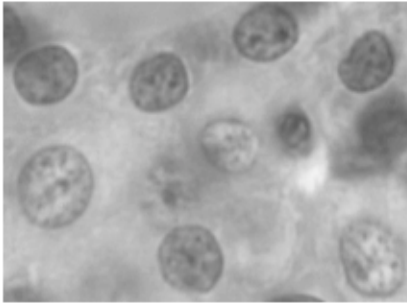
Occlusion of the cells is a very difficult problem to deal with in cell image segmentation. The puzzle is partly solved in our approach by incorporating

a priori knowledge about the shape of the cell contour. But a lot of effort is still needed to solve the occlusion problem completely.

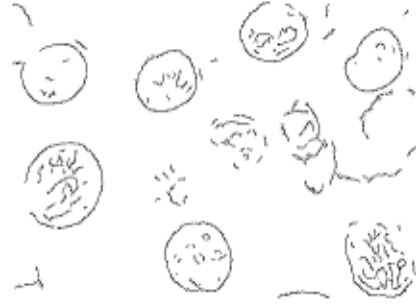
References

- [1] Wu, H. S., Barba, J., and Gil, J., A Parametric Fitting Algorithm for Segmentation of Cell Images, *IEEE Trans. on Biomedical Engineering*, **45**(1998), 400-407.
- [2] Garbay, C., Image structure representation and processing discussion of some segmentation methods in cytology, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **8**(1986), 140-147.
- [3] Garrido, A., and Perez, N., Applying deformable templates for cell image segmentation, *Pattern Recognition*, **33**(2000), 821-832.
- [4] Mouroutis, T., Roberts, S. J., and Bharath, A. A., Robust cell nuclei segmentation using statistical modelling, *BioImaging*, **6** (1998), 79-91.
- [5] Simon, I., Pound, C. R., Partin, A. W., Clemens, J. Q., and Christensbarry, W. A., Automated image analysis system for detecting boundaries of live prostate cancer cells, *Cytometry*, **31**(1998), 287-294.
- [6] Wu, H. S., and Gil, J., An iterative algorithm for cell segmentation using short-time Fourier transform, *Journal of Microscopy*, **184**(1996), 127-132.
- [7] Wu, H. S., Barba, J., and Gil, J., Iterative thresholding for segmentation of cells from noisy images, *Journal of Microscopy*, **197**(2000), 296-304.
- [8] Kapur, J. N., Sahoo, P. K., and Wong, A. K. C., A New Method for Gray-Level Picture Thresholding Using the Entropy of the Histogram. *Computer Vision, Graphics and Image Processing*, **29**(1985), 273-285.
- [9] Kittler, J., and Illingworth, J., Minimum Error Thresholding. *Pattern Recognition*, **19** (1986), 41-47.
- [10] Poggio, T., and Torre, V., Ill-posed problems and regularization analysis in early vision, "Proceedings of AARPA Image Understanding Workshop", 1984, pp.257-263.
- [11] Canny, J., A computational approach to edge detection, *IEEE Trans. Pattern Anal. Mach. Intell.*, **8** (1986), 679-698.
- [12] Goldberg, D. E., "Genetic Algorithm in Search, Optimization and Machine Learning", Addison-Wesley Publishing Company, 1989.
- [13] Rudolph, G., Convergence analysis of canonical genetic algorithm. *IEEE Trans. on Neural Networks*, **5**(1994), 96-101.
- [14] Ke, Q. Jiang, T., and Ma, S., A tabu search method for geometric primitive extraction, *Pattern Recognition Letters*, **18**(1997), 1443-1451.

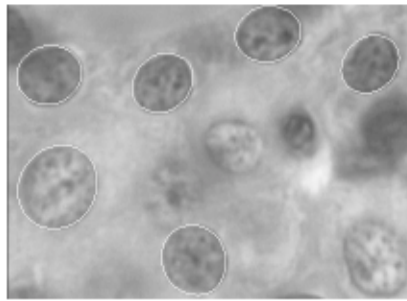
- [15] Yang, F., and Jiang, T., Cell Image Segmentation with the Kernel-Based Dynamic Clustering and an Ellipsoidal Cell Shape Model, *Journal of Biomedical Informatics*, **34**(2001).
- [16] Mühlenbein, H., Schomisch, M., and Born, J., The parallel genetic algorithm as function optimizer, *Parallel Computing*, **17**(1991), 619-632.
- [17] Wilkinson, B., and M. Allen, “Parallel Programming: Techniques and Applications using networked workstations and Parallel Computers”, Prentice Hall, New Jersey, 1999.



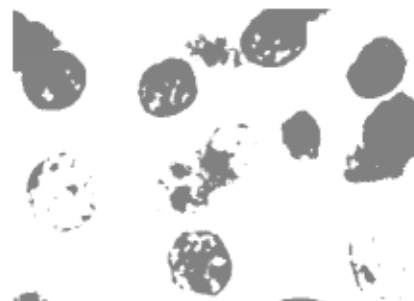
(a)



(b)



(c)



(d)

Fig 1. Segmentation results of hypothyroid image: (a) Original image, (b) Edge image, (c) Our approach, and (d) Histogram based method

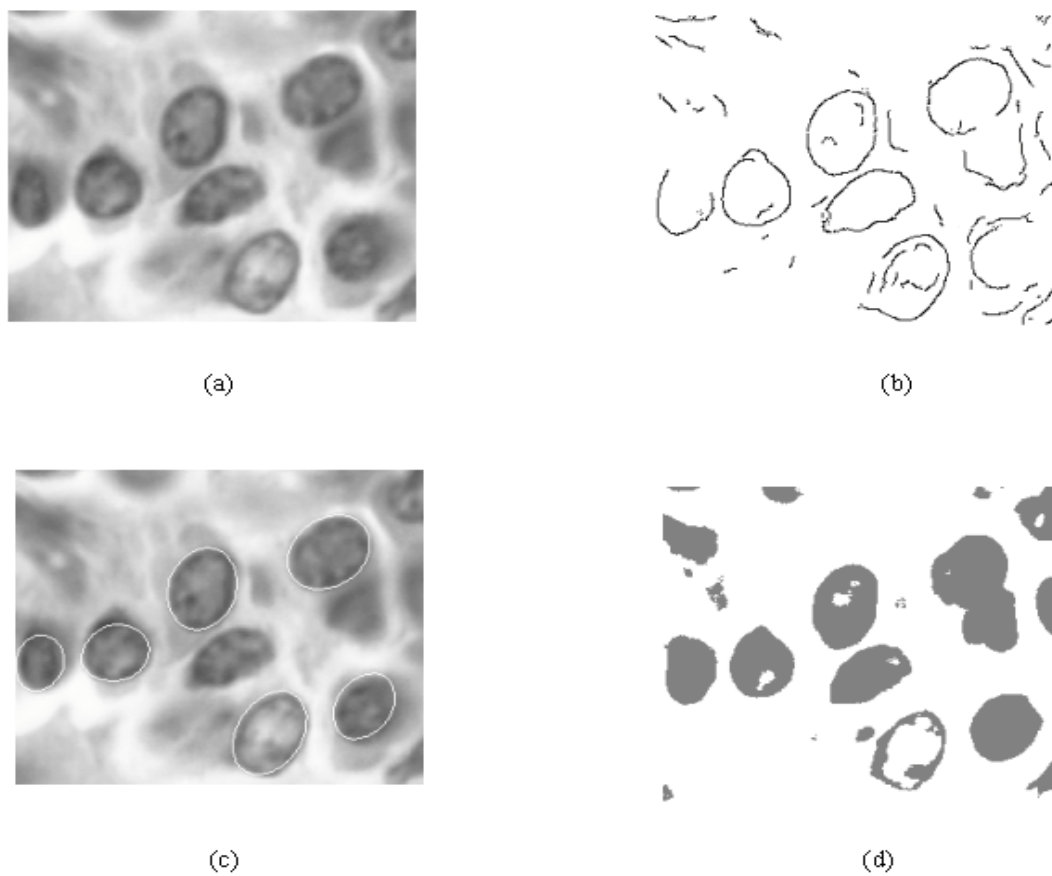


Fig 2. Segmentation results of small intestine image: (a) Original image, (b) Edge image, (c) Our approach, and (d) Histogram based method