1st International Conference on Experiments/Process/System Modelling/Simulation/Optimization
1st IC-EpsMsO
Athens, 6-9 July, 2005

NUCLEI SEGMENTATION OF MICROSCOPY IMAGES OF THYROID NODULES VIA ACTIVE CONTOURS AND K-MEANS CLUSTERING

Dimitris Glotsos[1], Antonis Daskalakis[1], Ioannis Kalatzis[2], Panagiota Spyridonos[1], Panagiota Ravazoula[3], Nikos Dimitropoulos[4], George Nikiforidis[1] and Dionisis Cavouras[2]

[1] Medical Image Processing and Analysis Group, Laboratory of Medical Physics,
University of Patras
26500 Patras, Greece
e-mail: dimglo@med.upatras.gr, web page: http://mipa.med.upatras.gr

[2] Medical Image and Signal Processing Laboratory, Department of Medical Instruments Technology,
Technological Institute of Athens, Greece
email: cavouras@teiath.gr, web page: http://medisp.bme.teiath.gr

[3] Department of Pathology,
University Hospital of Patras, Greece

[4] Department of Medical Imaging
EUROMEDICA Medical Center
11 527 Athens, Greece

Keywords: segmentation, thyroid nodules, active contours, k-means clustering.

Abstract. *Purpose: In this work the problem of automatic segmentation of nuclei in cytological images of thyroid nodules is investigated. Materials and Methods: 50 Fine Needle Aspiration images of thyroid nodules were digitized (768x576x8 bit) using a light microscopy imaging system comprising a Zeiss Axiostar plus microscope connected to a Leica DC 300 F color video camera. An automatic segmentation algorithm was designed combining the k-means clustering algorithm and active contours. The k-means algorithm is used to classify all pixels in a given image into nuclei or surrounding tissue, in order to provide an initial gross estimation of nuclei regions. Starting from the boundaries of these regions, an active contour snake fired and propagated till converging to nuclear boundaries. Results: On average, 94% of all nuclei were correctly delineated according to a histopathologist evaluation. Conclusions: The algorithm might be of value for computer-assisted microscopy systems, since accurate nuclei segmentation enables the accurate quantification of DNA content that may potentially allow the prediction of the disease course.*

1 INTRODUCTION

Raw microscopic cytological samples cannot be visualized under the microscope, because they are transparent and need to be properly prepared and stained [1]. Staining aims to enhance contrast between the region of interest (usually nuclei) and the background. A stain is a substance that adheres to nuclei, giving the cell nucleus color. However, stains undesirably concentrate to a variety of structures outside nuclei due to variations of the biological process of labeling, contaminating the microscopic images with noisy regions [1]. The latter significantly complicates accurate nuclei segmentation, which is the most important pre-processing stage of any computer-assisted microscopy system [2-3].

Many studies have been presented to solve the demanding task of medical image segmentation [4-5], but only a few focus in microscopy image segmentation [2-3, 6-8]. Among these studies, in [6] a method has been presented for the automatic segmentation of nuclei in breast fine needle aspiration images using the Hough transform. In [7] images from Papanicolaou stained smears have been automatically segmented using Neural Networks and active contours respectively. In [8], a method for segmentation of immunohistochemically stained nuclei has been presented, using a supervised pixel based classification algorithm and watershed operations. In [3], a supervised pixel based classification algorithm, using neural networks and morphological filters, was presented for segmentation of Hematoxylin and Eosin stained images of urinary bladder cancer. In this work, we investigated and developed a pixel-based clustering method combined with active contours for the automated segmentation of microscopy images of thyroid nodules.

2 MATERIAL AND METHODS

Fifty (50) Fine Needle Aspiration (FNA) obtained biopsies of thyroid nodules were collected from the EUROMEDICA medical center, Athens, Greece. Raw cytological biopsy samples were stained with Hematoxylin and Eosin (H&E). For each biopsy, a histopathologist (P.R.) specified the most representative region. From this region, images (Fig. 1) were digitized (768x576x8 bit), using a light microscopy imaging system, comprising a Zeiss Axiostar plus microscope, connected to a Leica DC 300 F color video camera.

2.1 Segmentation Methodology

5x5 pixel windows were considered around each image pixel. At the location of each window, three textural features were extracted: the energy of the first and second detailed images of the Mallat redundant discrete wavelet transform [9]. In this way, the differences in texture between nuclei, surrounding tissue, and boundaries were encoded, since coarser regions (i.e. regions of surrounding tissue, boundaries) resulted in higher values for both features, whereas smoother textured regions (i.e. interior of nuclei) resulted in lower values.

Feature vectors, denoted by $x_1, \ldots, x_N$, N = number of all image pixels, were then fed to a k-means clustering algorithm [10]. The k-means algorithm clusters the N feature vectors into K different partitions by minimizing the following criterion:

$$J = \sum_{j=1}^{K} \sum \left| x_n - \mu_j \right|^2 \qquad (1)$$

where x_n is the N^{th} feature vector and μ_j is the centroid of each of the k clusters. The algorithm firstly starts by assigning all data randomly to one of the predefined clusters (the nuclei or the background cluster). Then, the centroid of each cluster is computed and the procedure is iteratively performed until no further change in the assignment of the feature vectors occur [10].

The result of the clustering process was a binary image with white regions, those pixels that were assigned to the nuclei cluster, and black regions, those pixels assigned to the background cluster. Using morphological and edge detection filters [2, 11], only the boundaries of nuclei were retained and were considered as the initialization coordinates for an active contour (snake) model. Starting from these coordinates, the snake was generated for each nucleus and propagated until converging to the final nuclei boundaries. Active contours are algorithms designed to detect boundaries around objects by generating curves (snakes) that can be deformed, based on information extracted from the objects themselves. In this work, the gradient vector flow (GVF) model was utilized [12-13].

To evaluate the performance of the segmentation algorithm, a histopathologist examined each pair of original and segmented images and noted the number of wrongly identified and/or missed nuclei against the number of correctly identified ones. Cases of wrongly identified nuclei were overlapped, 'missed' and corrupted nuclei.

3 RESULTS

In figure 1a, a FNA H&E stained microscopic image of a thyroid nodule is illustrated. The performance of the k-means clustering algorithm in locating nuclei regions is illustrated in figure 1b. Using morphological (Fig. 1c) and Roberts filtering (Fig. 1d), only nuclei boundaries were retained and superimposed to the original image (Fig. 1e). The snake was designed for each nucleus; it was initiated from the boundaries, detected after applying the Robert filter, and then converged to the final nuclei boundaries (Fig. 1f). Following the nuclei segmentation evaluation procedure described above, the accuracy of the correctly identified nuclei ranged between 83-95%, with an average of 94%.

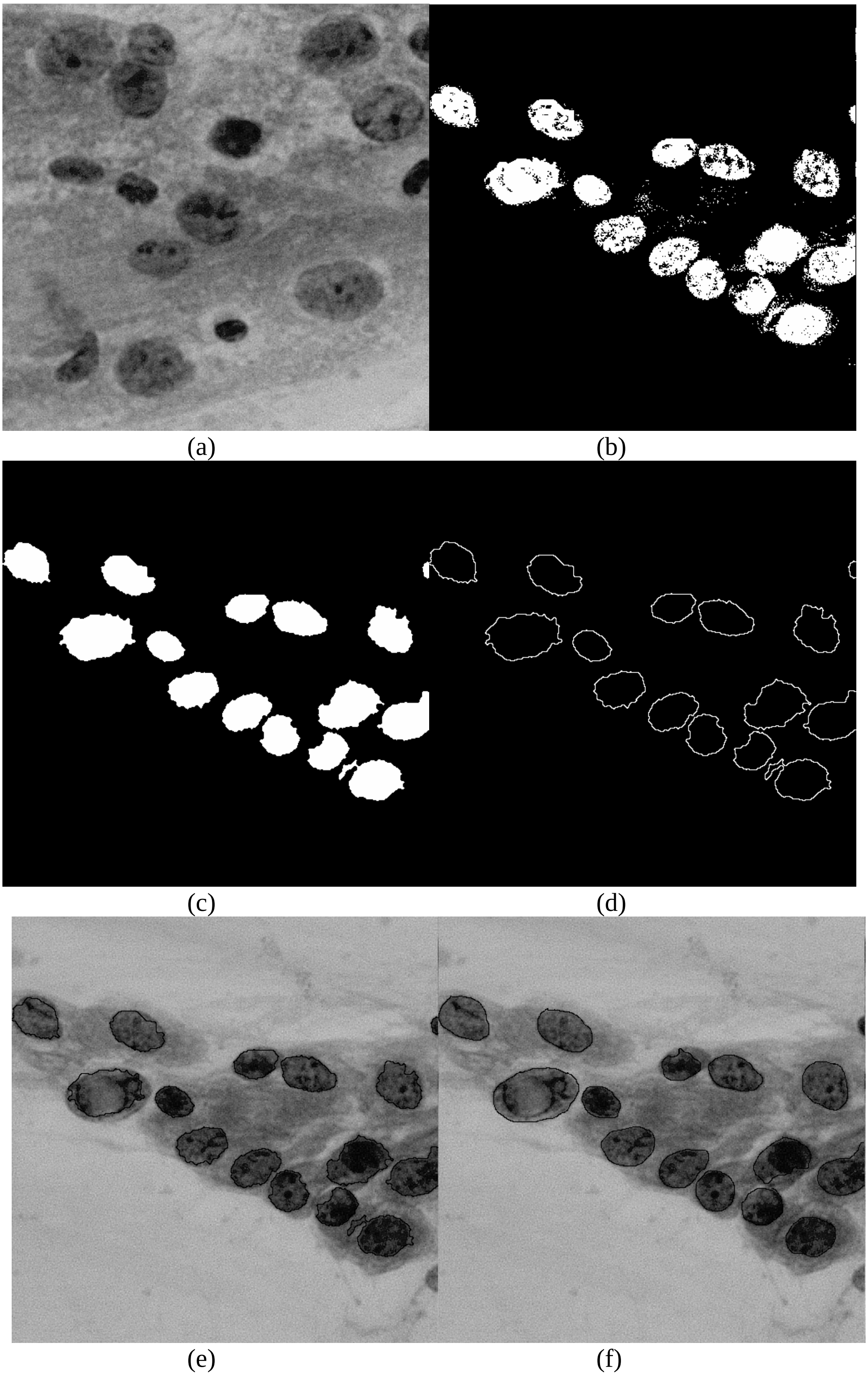

Figure 2. (a) Original image, (b) Result of k-means clustering, (c) Morphological filtering, (d) Edge detection using Roberts filtering, (e) Initialization coordinates for the GVF snake and (f) final nuclei boundaries

4 DISCUSSION

Accurate segmentation of nuclei is of crucial importance to guarantee correct results in computer-assisted microscopy [2-3]. Nuclei encode significant diagnostic and prognostic information that, if quantified, it can potentially allow the prediction of the disease course [3]. Previous studies that have investigated the demanding task of nuclei segmentation [6-8], have reported relatively high segmentation accuracies, such as 85% [6] and 89% [7]. Segmentation results in this study may be regarded as most promising considering that the H&E staining protocol is not as accurate in staining nuclei as other specialized protocols used in previous studies [2, 6-8]. It has to be stressed that the aim of the segmentation procedure was to extract a representative number of accurately segmented nuclei from every set of images describing each case (patient), in order to compute nuclear features. Under this perspective and considering that segmented nuclei ranged from 220 to 340 for each case-patient, the misclassification error of 6% may be regarded as of limited significance. The latter can be furthermore supported by the fact that it has been shown that even 200 correctly segmented nuclei per case are adequate for extracting nuclear features [14].

Concluding, in this study a method for the automatic segmentation of FNA H&E stained images of thyroid nodules was introduced by developing a k-means clustering algorithm and an active contour snake model. The method could be of value for computer-assisted microscopy systems, since accurate segmentation is of valuable importance.

ACKNOWLEDGEMENTS

The present research was carried for the project “Computer-based system for the automatic diagnosis of thyroid nodule cancer” co-funded by 75% from the European Union and 25% from the Greek Government under the framework of the Education and Initial Vocational Training Program – Archimedes.

REFERENCES

[1] Leboffe, M. J., Pierce, B. E. (2002). *Microbiology A Laboratory Theory and Application.* Morton Publishing Company.

[2] Mouroutis, Roberts and Bharath, (1998), “Robust cell nuclei segmentation using statistical modelling”, Bioimagning, Vol. 6(2), pp. 79-91.

[3] Spyridonos P., Cavouras D., Ravazoula P., Nikiforidis G. (2002), “Neural Network based segmentation and classification system for the automatic grading of histological sections of urinary bladder carcinoma”, Analytical and Quantitative Cytology and Histology Vol. 26, pp.317-324.

[4] Xu H., Liao M. (1998), **“**Cluster-based texture matching for image retrieval”, International Conference on Image Processing, Vol. 2, pp. 766 – 769.

[5] Saha P. K., Udupa J. K., Odhner D., (2000), “Scale-Based Fuzzy Connected Image Segmentation: Theory, Algorithms, and Validation”, Computer Vision and Image Understanding Vol. 77, pp. 145-174.

[6] Lee K-M. and Street N. (1999), “A fast and robust approach for automated segmentation of breast cancer nuclei”, *Proceedings of the 2nd IASTED International Conference on Computer Graphics and Imaging, Spain, Malaga, September 9-12,* pp. 42-47.

[7] McKenna S.J. (1994), *Automated analysis of papanicolaou smears*, Ph.D., University of Dundee.

[8] Ranefall P., Egevad L., Nordin B., and Bengtsson E. (1997), “A new method for segmentation of colour images applied to immunohistochemically stained cell nuclei”, Analytical Cellular Pathology, Vol. 15, pp. 145-156.

[9] Shensa M J (1992), “Wedding the `a trous and Mallat algorithms” IEEE Trans. Signal Proc. Vol. 40 pp. 2464–82.

[10] Bishop, C. M. (1995), *Neural Networks for Pattern Recognition.* Oxford, England: Oxford University Press.

[11] Roberts L. (1965), *Machine Perception of 3-D Solids*, Optical and Electro-optical Information Processing, MIT Press.

[12] C. Xu and J.L. Prince: Snakes, Shapes, and Gradient Vector Flow. IEEE Transactions on Image Processing 1998; 7:359-369.

[13] http://iacl.ece.jhu.edu/projects/gvf/

[14] Sallinen P., Sallinen S., Helen T., Rantala I., Rautiainen E., Helin H., Kalimo H., Haapsalo H. (2000), “Grading of diffusely infiltrating astrocytomas by quantitative histopathology, cell proliferation and image cytometric DNA analysis”, Neuropathology and Applied Neurobiology Vol. 26, pp. 319-331.