PATTERN RECOGNITION BASED SEGMENTATION VERSUS WAVELET MAXIMA CHAIN EDGE REPRESENTATION FOR NUCLEI DETECTION IN MICROSCOPY IMAGES OF THYROID NODULES

D. Glotsos*, S. Tsantis*, Jan Kybic**, A. Daskalakis*, P. Ravazoula***, I. Kalatzis****, N. Dimitropoulos*****, D. Cavouras**** and G. Nikiforidis*

* University of Patras, Department of Medical Physics, Patras, Greece
** Czech Technical University, Department of Cybernetics, Prague, Czech Republic
*** University Hospital of Patras, Department of Pathology, Patras, Greece
**** Technological Institute of Athens, Department of Medical Instrumentation Technology, Athens, Greece
***** EUROMEDICA Medical Center, Department of Medical Imaging, Athens, Greece

dimglo@med.upatras.gr

Abstract: To guarantee correct results in computer-assisted microscopy, reliable segmentation of nuclei images is required. Nuclei encode significant diagnostic and prognostic information, that if quantified can potentially allow the prediction of the disease course. Thus, accurate nuclei segmentation is of crucial importance. In this study, two methods have been developed and comparatively evaluated for the task of nuclei segmentation in cytological images of thyroid nodules. The first method uses concepts of pattern recognition and a Support Vector Machine classifier to discriminate pixels belonging to nuclei from pixels of surrounding background. The second method investigates wavelet multi-scale edge representation and the importance of local maxima as possible descriptors of nuclei boundaries. Results for both methods were compared to the manual segmentation of a histopathologist for 30 microscopic images of thyroid nodules. With the first method, 91% of nuclei were correctly delineated, whereas with the second method 89%. The proposed algorithms could be of value for computer-assisted microscopy systems that negotiate quantitative interpretation of microscopic images.

Introduction

Microscopic examination of thyroid diseases has become the 'golden standard' in clinical routine, since ultrasounds scans on their own cannot provide definite and reliable diagnosis [1]. Based on the appearance of cell nuclei, physicians classify patients into clinically meaningful categories (such as normal, abnormal, hashimoto etc), in order to more accurately select patients who will undergo surgery [2]. However, the clarity of microscopic images has been long questioned, since it depends on the complicate nature of nuclei staining, which is used to enhance contrast between cells, nuclei and surrounding background [3]. Stains do not always adhere solely to cells and nuclei, as expected, due to variations of the biological process of labeling, resulting in contamination of microscopic images with noisy regions. The existence of these regions promotes sources of diagnostic misinterpretations and interobserver reproducibility [4, 5]. In order to objectify the diagnostic process, computer-assisted microscopy systems have been introduced [6, 7]. These systems have been designed to investigate whether morphological and textural features of cell nuclei may be used to quantitatively assess thyroid diseases. Thus, accurate computation of nuclear features affects the success of computer-assisted microscopy systems, making accurate nuclei segmentation of crucial importance [6].

In this study, two methods have been developed and comparatively evaluated for the task of nuclei segmentation in cytological images of thyroid nodules. The first method uses concepts of pattern recognition and a Support Vector Machine (SVM) classifier. The second method investigates multi-scale edge representation and the importance of local maxima as possible descriptors of nuclei boundaries.

Materials and Methods

Thirty Hematoxylin-Eosin stained Fine Needle Aspiration (H&E-FNA) biopsies were collected from the University Hospital of Patras and the EUROMEDICA center of Athens in Greece. From regions pre-specified by a histopathologist (P.R.), images were digitized (1300x1030x8bit) using a light Zeiss Axiostar Plus microscope connected to a Leica DC 300F color CCD camera.

Pattern recognition segmentation method (PBS): From each image, forty (40) 5x5-pixel windows were manually sampled from nuclei and surrounding background respectively. From each sampled window, three textural features were extracted: the spread, cross

relation, and sum of the autocorrelation function [8]. The autocorrelation function is defined as:

$$A_F(m,n) = \sum_j \sum_k F(j,k)F(j-m,k-n) \quad (1)$$

where *F(j,k)* is the image intensity at the pixel *(j,k)*. The features *spread* and *cross relation* are defined as

$$a(1) = \sum_{m=0}^{S} \sum_{n=-S}^{S} (m-n_m)(n-n_n)A_F(m,n) \quad (2)$$

and

$$a(2) = \sum_{m=0}^{S} \sum_{n=-S}^{S} (m-n_m)^2(n-n_n)^2 A_F(m,n) \quad (3)$$

where *a(1) is the spread and a(2) the cross relation of the* autocorrelation function A_F inside an SxS window, *m, n* represent all possible displacements inside the *SxS* window,

$$n_m = \sum_{m=0}^{S} \sum_{n=-S}^{S} mA_F(m,n) \quad (4)$$

and

$$n_n = \sum_{m=0}^{S} \sum_{n=-S}^{S} nA_F(m,n) \quad (5)$$

The 40 three-dimensional feature vectors fed an SVM classifier [9], which was trained to identify texture variations between nuclei and surrounding tissue regions. The discrimant function of the SVM classifier is the following:

$$g(x) = sign\left(\sum_{i=1}^{N} \alpha_i y_i K(\mathbf{x},\mathbf{x}_i) + b\right) \quad (6)$$

where x_i training data belonging to either class $y_i \in\{+1,-1\}$, N the number of training samples, α_i, b weight coefficients and K the kernel function.

The SVM classifier was implemented in custom-made software using MATLAB. As kernel function *K,* the Gaussian radial basis function was designed, which has the form:

$$K_{RBF}(x,x_i) = \exp\left(\frac{-\|x-x_i\|^2}{2\sigma^2}\right) \quad (7)$$

where σ is the spread, that was experimentally determined.

The optimization problem of finding coefficients a_i was solved by using the routine *quadprog* [10] provided with the MATLAB optimization toolbox.

After training, the SVM classifier was applied to all possible 5x5 windows centered at all possible pixels. The result of the classification process was a binary image with white and black regions representing pixels classified as nuclei or surrounding background respectively (see figure 2.1). Binary images were further corrected using fill-holes and morphological operations (see figure 2.2) [11] (open and close, with 5x5 circular structuring element).

Wavelet maxima chain edge detection (WMC): The Redundant Wavelet Transform (WT) [12] was applied to each image till the 4th scale. WT is based on a wavelet function with compact support, which is the first order derivative of cubic spline function. The gradient's vector modulus and angle of the wavelet transform along the horizontal and vertical orientation were computed for each scale 2^j:

$$M_{2^j} f(x,y) = \sqrt{\left|W^1{}_{2^j} f(x,y)\right|^2 + \left|W^2{}_{2^j} f(x,y)\right|^2} \quad (8)$$

$$A_{2^j} f(x,y) = \arctan\frac{W^1{}_{2^j} f(x,y)}{W^2{}_{2^j} f(x,y)} \quad (9)$$

where M_2^j is the Modulus and A_2^j the gradient vector across each scale 2^j.

Multi-scale edge points were considered as those modulus points that were local maxima of the magnitude along the direction given by the angle value. Maxima chains were computed by chaining together only those neighboring maxima with similar magnitude and angle values [13]. The maxima that their amplitude remained relatively constant across scales were retained and were considered as nuclei boundaries; if not, the maxima were discarded. The WMC algorithm was implemented in custom-made software using MATLAB.

Evaluation: To evaluate the performance of the each segmentation method, an experienced histopathologist manually delineated nuclei boundaries in all 30 images, which were subsequently segmented using the *PBS* and *WMC* methods. The area, roundness and concavity of manually and automatically segmented nuclei were calculated and compared to give an estimation of each algorithm's segmentation accuracy.

Results

Figure 1 demonstrates a typical cytological image of an H&E-FNA-stained thyroid biopsy sample. Figures 2.1-2.4 illustrate each step of the segmentation process using the *PBS* algorithm, whereas figures 3.1-3.3 show the result of the *WMC* method. According to the evaluation process described in the previous section, on average 91% of nuclei were correctly delineated with the *PBS* algorithm, whereas with the *WMC* algorithm 89%.

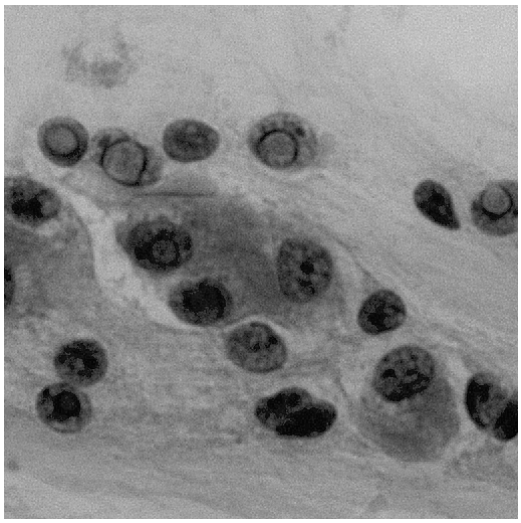

Figure 1: Image from an H&E stained FNA cytological image of a thyroid nodule

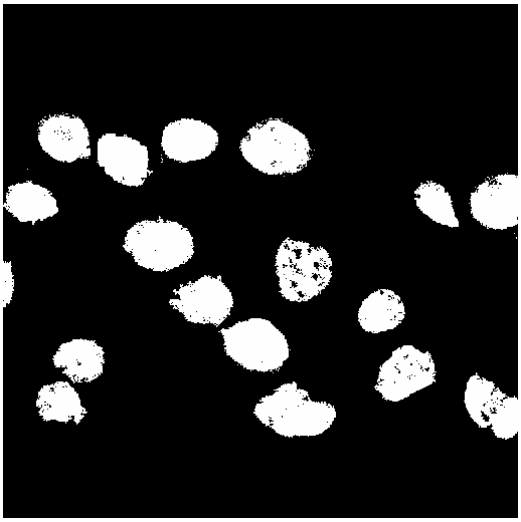

Figure 2.1: SVM pixel segmentation

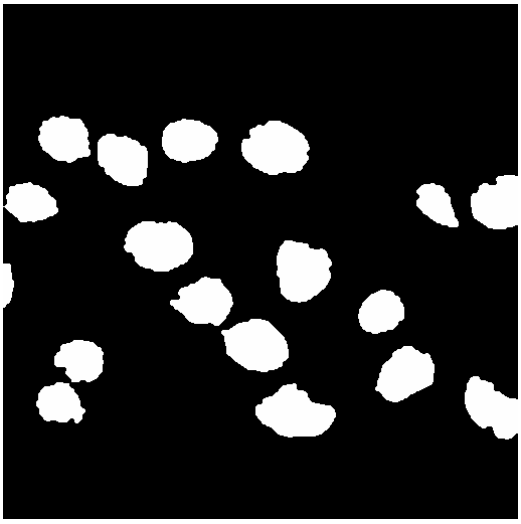

Figure 2.2: Nuclei refinement using fill-holes and morphological operations

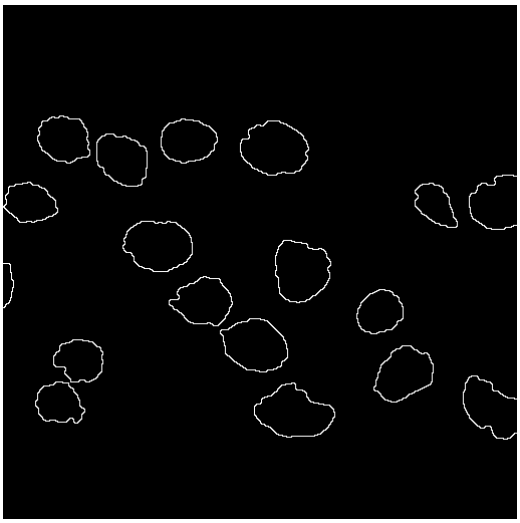

Figure 2.3: Retaining only the boundaries of segmented nuclei by using Roberts filtering

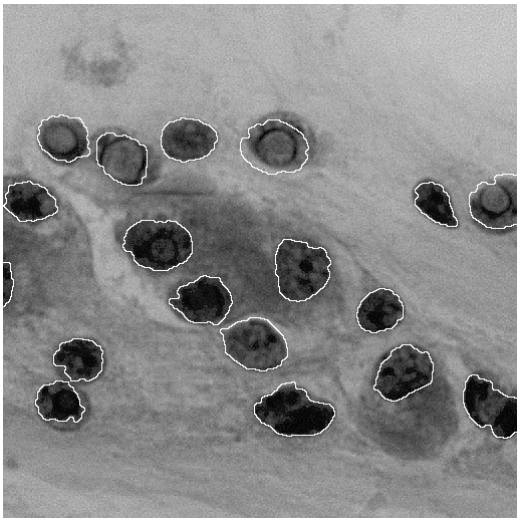

Figure 2.4: Final segmented image

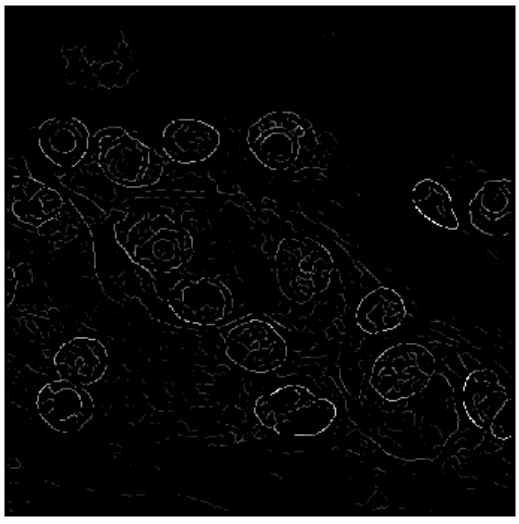

Figure 3.1: Local maxima representation

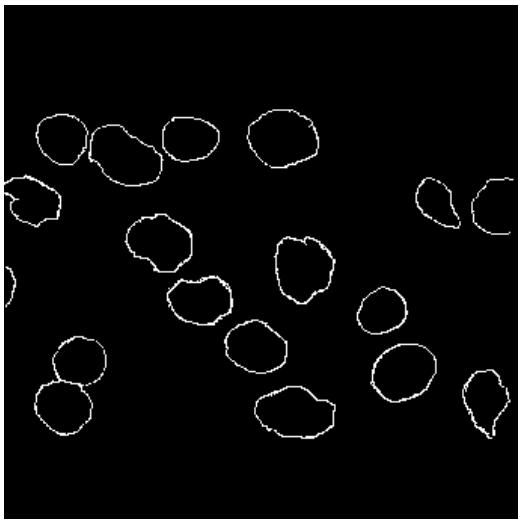

Figure 3.2: Retaining the maxima that their amplitude remained relatively constant across scales, and using connecting component analysis [14] to find connected components in the binary image

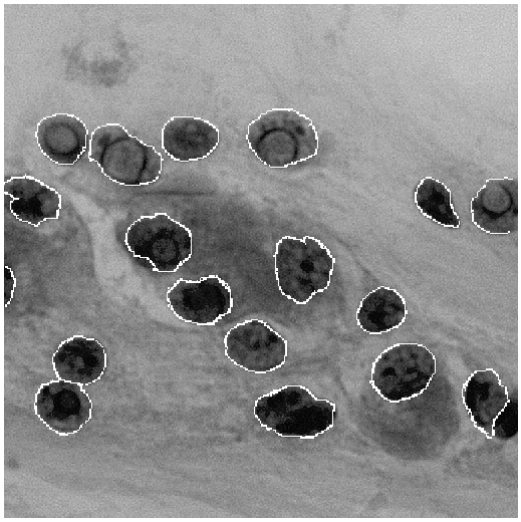

Figure 3.3: Final segmented image using the WMC algorithm

Discussion

A number of segmentation techniques have been proposed for nuclei delineation in microscopy images. Among most popular algorithms are considered the Hough transform, which has been applied to cytological images of breast lesions [15] and active contours, which have been designed for papanicolaou stained microscopy smears [16]. Watershed techniques have also been introduced, giving promising results, with applications in immunohistochemically stained nuclei images [17]. The success of these algorithms greatly depends on initialization conditions, such as for example the a priori estimation of nuclei shape for the Hough transform, and the starting coordinates of active contour models.

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 [15, 16, 18].

The *WMC* method was faster and more accurate in retaining nuclei shape, whereas the *PBS* method was more effective in separating closely located nuclei. Both methods could be of value to computer-based systems designed to objectively interpret cytological images, since provide means for accurate nuclei segmentation, thus reliable computation of nuclear features.

Acknowledgements

We thank European Social Fund (ESF), Operational Program for Educational and Vocational Training II (EPEAEK II) and particularly the Program IRAKLEITOS for funding the above work.

References

[1] KOIKE, E. (2001): 'Effect of Combining Ultra Sonography and Ultrasound-Guided Fine-Needle Aspiration Biopsy Findings for the Diagnosis of Thyroid Nodules', *Eur. J. of Sur*, **167**, pp. 656-661

[2] KLEMI, P. J., JOENSUU, H. and NYLAMO, E. (1991): 'Fine Needle Aspiration Biopsy in the Diagnosis of Thyroid Nodules', *Acta Cytol*, **35**, pp. 434-8

[3] MICROSCOPY AND STAINING, Internet site address: http://www.mansfield.ohio-state.edu/~ sabedon /black03.htm

[4] EINSTEIN, A. J., GIL, J., WALLENSTEIN, S., BODIAN, C. A., SANCHEZ, M., BURSTEIN, D. E., WU, H. S. and LIU, Z. (1997): 'Reproducibility and Accuracy of Interactive Segmentation Procedures for Image Analysis in Cytology', *J Microsc*, **188 (Pt 2)**, pp. 136-48

[5] PRAYSON, R., AGAMANOLIS, D., COHEN, M., ESTES, M., KLEINSCHMIDT-DEMASTERS, B., ABDUL-KARIM, F., MCCLURE, S., SEBEK, B. and VINAY, R. (2000): 'Interobserver Reproducibility among Neuropathologists and Surgical Pathologists in Fibrillary Astrocytoma Grading,' *J. of the Neur. Sc.*, **175**, pp. 33-39

[6] ELDAR, S. (1998): 'Computer-Assisted Image Analysis of Small Cell Lymphoma of the Thyroid Gland,' *Comp. Med. Im. and Grap.*, **22**, pp. 479-488

[7] RORIVE, S., EDDAFALI, B., FERNANDEZ, S., DECAESTECKER, C., ANDRE, S., KALTNER, H., KUWABARA, I., LIU, F. T., GABIUS, H. J., KISS, R. and SALMON, I. (2002): 'Changes in Galectin-7 and Cytokeratin-19 Expression During

the Progression of Malignancy in Thyroid Tumors: Diagnostic and Biological Implications,' *Mod Pathol*, **15**, pp. 1294-301

[8] FAUGERAS, O. and PRATT, W. (1980): 'Decorrelation Methods of Texture Feature Extraction,' *IEEE Trans. Pattern Analysis and Machine Intelligence*, **14**, pp. 323-332

[9] BURGES, C. (1998): 'A Tutorial on Support Vector Machines for Pattern Recognition,' *Data Min. and Kn. Disc*, **2**, pp. 121-167

[10] MAHTHWORKS-OPTIMIZATION TOOLBOX, Internet site address: http://www.mathworks.com/products/optimization/?BB=1

[11] GONZALEZ, R. and WOODS, R. (2002),'Some Basic Morphological Algorithms',in GONZALEZ, R. (2nd ed): 'Digital Image Processing', (Addison-Wesley Pub, New York), pp. 518-28.

[12] SHENSA, M. (1992): 'The Discrete Wavelet Transform: Wedding the `a Trous and Mallat Algorithms,' *IEEE Trans. Signal Proc*, **40**, pp. 2464-82

[13] MALLAT, S. (1992): 'Characterization of signals from multi-scale edges'**,** *IEEE Trans. Patt. Anal. Mach. Intell*, **14**, pp. 710-732

[14] HARALICK, R. M. and SHAPIRO, L. G. (1992),'Computer and Robot Vision', Addison-Wesley, pp. 28-48.

[15] LEE, K.-M. and STREET, N. (1999): 'A Fast and Robust Approach for Automated Segmentation of Breast Cancer Nuclei,' Proc. of *Proceedings of the Second IASTED International Conference on Computer Graphics and Imaging*, Palm Springs, CA, USA, 1999, p. 42-47

[16] MCKENNA, S. J., 'Automated Analysis of Papanicolaou Smears,' Ph.D thesis (University of Dundee), 1994

[17] RANEFALL, P., EGEVAD, L., NORDIN, B. and BENGTSSON, E. (1997): 'A New Method for Segmentation of Colour Images Applied to Immunohistochemically Stained Cell Nuclei,' *Anal Cell Pathol*, **15**, pp. 145-56

[18] RANEFALL, P., WESTER, K. and BENGTSSON, E. (1998): 'Automatic Quantification of Immunohistochemically Stained Cell Nuclei Using Unsupervised Image Analysis,' *Analytical Cellular Pathology*, **16**, pp. 29-43