34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Automated segmentation of free-lying cell nuclei in Pap smears for malignancy-associated change analysis

Ramin Moshavegh, Babak Ehteshami Bejnordi, Andrew Mehnert, K. Sujathan, Patrik Malm and Ewert Bengtsson

Abstract—**This paper presents an automated algorithm for robustly detecting and segmenting free-lying cell nuclei in bright-field microscope images of Pap smears. This is an essential initial step in the development of an automated screening system for cervical cancer based on malignancy associated change (MAC) analysis. The proposed segmentation algorithm makes use of gray-scale annular closings to identify free-lying nuclei-like objects together with marker-based watershed segmentation to accurately delineate the nuclear boundaries. The algorithm also employs artifact rejection based on size, shape, and granularity to ensure only the nuclei of intermediate squamous epithelial cells are retained. An evaluation of the performance of the algorithm relative to expert manual segmentation of 33 fields-of-view from 11 Pap smear slides is also presented. The results show that the sensitivity and specificity of nucleus detection is 94.71% and 85.30% respectively, and that the accuracy of segmentation, measured using the Dice coefficient, of the detected nuclei is 97.30±1.3%.**

I. Introduction

The Papanicloau test is the primary screening test for cervical cancer. It involves the microscopic examination of cells sampled from in and around the cervix for signs of precancerous and cancerous changes; e.g. large nucleus relative to cytoplasm. The test is labor-intensive and complex requiring the exhaustive review of tens of thousands of cells. One in every 10 to 20 positive cases is missed in routine screening [1]. There are two reasons for this [1]. The first is inappropriate interpretation due to factors such as fatigue, habituation, and inexperience. The second is due to sampling error wherein diagnostic cells do not make it onto the glass slide in the first place. Whilst automation using a computer and robotic microscope can address the issue of inappropriate interpretation it cannot address sampling error. Research suggests that the malignancy-associated change (MAC) phenomenon may be the solution. MACs are subtle sub-visual changes in the appearance of normal-looking cells from an abnormal Pap smear. The aim of MAC analysis is not to perform an exhaustive review of all of the cellular material to identify diagnostic cells but rather to look for MACs in a sub-population of cells sampled from the slide.

An essential first step in the development of an automated screener based on MACs is robust automatic segmentation of free-lying cell nuclei in digitized Pap smear images. Numerous algorithms have been published for this purpose. They can be categorized according to the primary underlying segmentation methodology used: global and adaptive thresholding [2], watershed transform [3], boundary detection algorithms and deformable models [4-7], and edge enhancement based techniques [8, 9]. The algorithms based on thresholding lack robustness to noise, uneven illumination, and variations in staining intensity. The algorithms based on the watershed transform can yield oversegmentation (when too many potential cell nuclei candidates are selected) and inaccurate boundary segmentation (because of the lack of sufficient gradient information). The algorithms based on boundary detection and deformable models are highly dependent on strong prior knowledge about the shape and location of objects in an image to guide the segmentation process. Moreover they require the initial contour to be reasonably close to the true object boundaries. As a consequence they can fail in images containing clustered and overlapping cells. The algorithms based on edge enhancement are designed to segment the cytoplasm and nucleus of free-lying cells and likewise do not perform well in images containing clustered and overlapping cells. Another criticism of many of these proposed algorithms is that their segmentation accuracies have not been objectively evaluated.

In this paper we present a new algorithm for robustly detecting and segmenting free-lying intermediate cell nuclei in bright-field microscope images of Pap smears. The novelty of the algorithm stems from a robust marker selection method for selecting candidate free-lying nuclei-like objects for subsequent marker-controlled watershed segmentation to obtain the nuclear boundaries. The algorithm also implements artifact rejection based on size, shape, and nuclear granularity to ensure only the nuclei of intermediate squamous epithelial cells are retained. The remainder of this paper is organized as follows. Section II introduces the proposed algorithm. Section III presents an empirical evaluation of its performance. Finally Section IV is the summary and conclusion.

R. Moshavegh, B. Ehteshami and A. Mehnert are with the Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden. A. Mehnert is also with MedTech West, Sahlgrenska University Hospital, Gothenburg, Sweden (e-mail: andrew.mehnert@chalmers.se).

P. Malm and E. Bengtsson are with the Centre for Image Analysis, Uppsala University, Uppsala, Sweden.

K. Sujathan is with the Department of Pathology, Regional Cancer Centre (RCC), Thiruvananthapuram, India.

II. Proposed Segmentation Algorithm

The proposed algorithm is designed to detect and segment free-lying intermediate squamous epithelial cell nuclei because these are the most abundant on a Pap smear and are the most suitable for MAC analysis [10]. The "nuclei of the intermediate cells measure about 8μm in average diameter, are round or oval, with a clearly defined nuclear membrane surrounding well-preserved homogeneous nucleoplasm" [11]. The algorithm is summarized in Table I and the individual steps discussed below.

TABLE I
PROPOSED SEGMENTATION ALGORITHM FOR FREE-LYING INTERMEDIATE CELL NUCLEI.

Input:	Gray-scale image containing a field-of-view (FOV) from a Pap smear slide.
Output:	Binary image containing connected components, each corresponding to a free-lying intermediate cell nucleus.
Steps:	
1.	Extract inner markers for free-lying nuclei-like objects (these locate the interiors of candidate objects).
2.	Apply the marker-controlled watershed transform on the FOV image with respect to the inner markers (this yields an outer marker that lies between the candidate objects).
3.	Apply the marker-controlled watershed transform on the gradient image with respect to the inner and outer markers (this yields the object boundaries/masks).
4.	Compute the area and quantitative measures of shape and texture granularity for each segmented nuclei-like object.
5.	Reject objects that are too small or large to be intermediate cells, that do not have an elliptical shape, and that do not have a granular texture.

Step 1: Extraction of inner markers

The detection of free-lying nuclei is achieved using the gray-scale annular closing operator. The operator is defined

$$\Psi_{anclo}(f, B) = (f \ominus B) \vee f \qquad (1)$$

where f is a gray-scale image, $\ominus$ denotes erosion, $\vee$ denotes pointwise maximum, and B is a symmetric structuring element which does not contain the origin [12]. The effect of this operator is to remove isolated dark spots in the gray-scale image. The operator is an algebraic closing and, like a conventional morphological closing, is increasing, extensive, and idempotent. Thus the arithmetic difference $\Psi_{anclo}(f, B) - f$ is non-negative and yields the removed isolated dark spots. These serve as candidate cell nuclei markers. This idea is illustrated in Fig. 1 for a single cervical cell and an annular structuring element. The inner radius and outer radius dimensions of the annular structuring element provide control over the size and relative isolation of the nuclei that can be detected.

The squamous epithelium is made up of three principal layers [11]: the basal cell layer (immature), the intermediate cell layers, and the superficial cell layers (most mature). The nuclei of superficial cells are pyknotic and considerably smaller than intermediate and parabasal cells with a nuclear

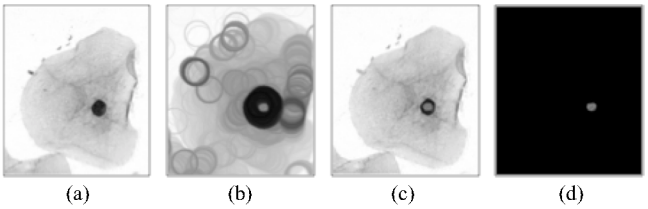

Figure 1. Extracting an inner marker for a free-lying cell nucleus. (a) Original image. (b) Gray-scale erosion with an annular structuring element. (c) Pixel-wise maximum of *(a)* and *(b)*. (d) Arithmetic difference between *(c)* and *(a)*.

diameter of about 4 μm [11]. In a normal Pap smear usually only the upper few layers of the squamous epithelium are removed and so the immature cells near the base of the epithelium are not sampled [13].

In order to detect a nucleus, an annular structuring element with an inner radius larger than that of the nucleus is needed. The nuclei of normal intermediate and parabasal cells measure approximately 8 μm in diameter and may enlarge up to 15 μm in the case of malignant or rare benign changes [11]. Hence, to detect all nuclei within this range, a set of independent annular closings with structuring elements with a range of inner diameters is needed. This is then the basis for the more sophisticated inner marker extraction algorithm presented in Table II.

Steps 2-3: Marker-controlled watershed segmentation of the detected nucleus-like objects

Marker-controlled watershed segmentation [14] is used to segment the boundaries of the nuclei-like objects located by the inner markers. First a watershed segmentation of the original image f with respect to the inner markers is performed to obtain the outer marker. Next the original image f is median filtered, its magnitude of gradient is computed, and the result is Gaussian filtered. A watershed segmentation of this filtered gradient with respect to the union of the inner markers and the outer marker yields the desired segmentation of the nuclei-like objects. This idea is illustrated in Fig. 2.

TABLE II
PROPOSED NUCLEI INNER MARKER EXTRACTION ALGORITHM

Inputs: Gray-scale image (f), and parameters λ_0, α, r_1 and r_2
Output: Binary mask (X_m) containing an inner marker for each free-lying nucleus-like object detected.

1. *Let B_0 be a disk structuring element of radius λ_0*
2. **for** $\lambda = r_1$ ***to*** r_2 **do**
3. *Let B_{an} be an annular structuring element with inner and outer radii of λ and $\lambda + \alpha$ respectively*
4. $g = \Psi_{anclo}(f, B_{an}) - f$
5. $X_1 = g > 0$
6. $X_2 = (X_1 \ominus B_0) \oplus B_0$
7. $X = X \cup X_2$
8. **end for**
9. X_m = set of centroids of the connected components in X

Note: $\oplus$ and $\ominus$ denote dilation and erosion respectively.

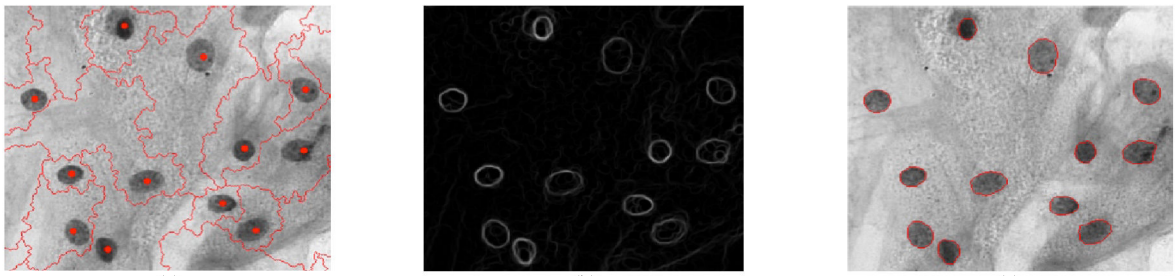

Figure 2. Segmentation of the detected nuclei-like objects. (a) Watershed segmentation of the original image with respect to the inner markers (shown as disks) yielding the outer marker. (b) Gaussian filtered gradient magnitude of the median-filtered original image in *(a)*. (c) Watershed segmentation of *(b)* with respect to the union of the inner markers and the outer marker.

Steps 4-5: Artifact rejection

Quantitative measures of the area, elliptical shape, and the texture/granularity are computed for each segmented nucleus-like object. These values are then used to reject objects that are too small or large to be intermediate cells, do not have an elliptical shape, or do not have a granular texture.

The minimum area is deemed to be the area of a circle with radius $r_{\min}$. To measure how elliptical a segmented nucleus-like object is, an ellipse is fitted to the object pixels such that it has the same normalized second central moments as the segmented object. The elliptic variance descriptor (E_{var}) [15] is then used to measure how closely the borders of the fitted ellipse agree with those of the segmented nucleus-like object. For elliptical objects E_{var} is close to 0 (see Fig. 3).

The granularity or coarseness of texture of each segmented object is computed using the coarseness feature devised by Tamura [16] (derived from a texture model corresponding to visual perception). This feature reflects the size and number of texture primitives. It is useful and robust in the sense that it does not depend directly on the exact gray-levels in the object and so has robustness to non-uniformity of illumination and staining variations (provided that these do not greatly affect the size and number of texture primitives). Fig. 3 depicts two candidate nuclei with different E_{var} and Tamura coarseness values.

III. Empirical Evaluation

The performance of the algorithm was evaluated relative to expert manual segmentation.

A. Image data

The data used in this study is a subset of 889 fields-of-view (FOVs) captured by a cytopathologist from 68 Pap smear slides. Each FOV was acquired using a CCD camera mounted on a light microscope. The images were captured with a 40× objective lens. Each FOV image is of size 1024×1344 pixels with square pixels of size 0.25μm. The gray-scale resolution is 8 bits per pixel.

Eleven slides, each containing a minimum of 100 non-superficial cervical cell nuclei, were randomly selected from among the 68 slides. For each slide three FOVs were randomly selected to yield a total of 33 FOVs.

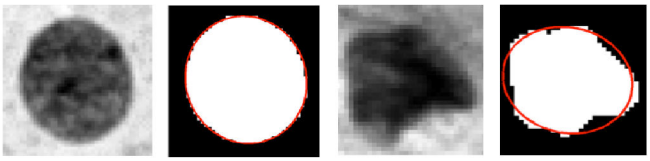

Figure 3. Assessing the shape and texture of a candidate nucleus. (a) and (b) show the result for a candidate nucleus with an elliptical shape ($E_{var} = 0.026$ and $Tamura\ coarseness = 11.17$). (c) and (d) show the result for a candidate nucleus with a non-elliptical shape ($E_{var} = 0.096$ and $Tamura\ coarseness = 10.75$).

B. Ground truth

A graphical user interface (GUI) was developed to permit a user to review each FOV and to place a marker on individual nuclei. Three untrained subjects were recruited to independently review the FOVs using the GUI and to mark each free-lying nucleus. Prior to performing this task each was shown examples of intermediate cell nuclei in another FOV (not one of the 33 FOVs they had to review). Each subject was specifically instructed to mark elliptical objects, of approximately the right size, with a well-defined boundary, and with a granular texture. The set of all objects selected by at least two of the three subjects were taken to be the ground truth for free-lying intermediate cell nuclei.

Two image analysis experts then used the GUI to independently trace the boundary of each ground truth nucleus. These manual segmentations were taken to be the ground truth for the boundaries of the free-lying intermediate cell nuclei.

C. Method

The proposed segmentation algorithm was applied to the 33 FOV images. The parameters for different steps of the algorithm were selected after several experiments on a small subset of images independent of the 33 selected FOV images. The minimum and maximum values of the inner radius (r_1 and r_2) of the annular structuring elements (see Table I) were set to 22 and 33 pixels. The values were selected based on the nucleus diameter range for the intermediate cell nuclei defined by Koss [11]. The outer radius of each annulus was set to be two pixels more than the inner radius (i.e. $\alpha = 2$) to guarantee the extraction of inner markers of the adjacent free-lying nuclei. Finally the size of the disk-structuring element (λ_0) was set to 3 pixels (for noise mitigation).

Objects selected by the algorithm were compared to the ground truth nuclei obtained manually and used to compute the sensitivity and specificity of the algorithm for the detection of free-lying intermediate cell nuclei.

To evaluate the accuracy of the segmentation of each nucleus detected by the algorithm, the resulting boundary was compared to the two corresponding boundaries (one from each expert) in the ground truth. More specifically the similarity between pairs of masks was computed in terms of the Dice similarity coefficient (DSC) scores [17]. The coefficient ranges between 0 and 1. A value of 1 indicates perfect agreement and a value of 0 indicates no agreement.

D. Results

The sensitivity and specificity of the algorithm for the detection of free-lying intermediate cell nuclei is 94.71% and 85.30% respectively. Boxplots of the DSC scores for the comparison of the proposed automatic segmentation to the two manual segmentations, and for the comparison between the two manual segmentations are shown in Fig. 4.

The agreement between the algorithm and the two manual segmentations is $97.30 \pm 1.3\%$ and $96.96 \pm 1.7\%$ respectively (mean DSC $\pm$ standard deviation). The overall agreement between the two expert segmentations is $97.26 \pm 1.2\%$.

E. Discussion

The sensitivity of the algorithm to the detection of free-lying intermediate cell nuclei is high. However its specificity, whilst still quite high, could be improved. A review of false positives indicates that some of them are due to segmentation failures as the result of severe background noise and artifacts. However other apparent failures in fact represent genuine free-lying nuclei overlooked by the three recruited subjects.

The DSC scores for boundary delineation evaluation (see Fig. 4) show that nuclei boundaries obtained using the marker-controlled watershed transform are highly accurate and consistent with the two experts' visual perception of the intermediate cell nuclei boundaries.

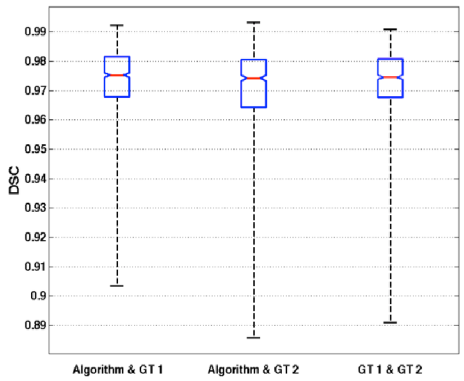

Figure 4. DSC scores for the comparison of the proposed automatic segmentation to the two manual segmentations, and for the comparison bewteen the two expert segmentations (GT1 and GT2).

IV. SUMMARY AND CONCLUSION

In this paper we presented a new automated algorithm for detecting and segmenting free-lying cell nuclei in bright-field microscope images of Pap smears. The empirical results show that the algorithm has a high sensitivity and specificity for the detection of free-lying nuclei, and that the algorithm is able to delineate the boundaries of these with high accuracy. This work represents the first step in the development of an automated screener for Pap smears based on the malignancy-associated change phenomenon.

REFERENCES

[1] R. M. DeMay, "Common problems in papanicolaou smear interpretation," *Archives of Pathology & Laboratory Medicine,* vol. 121, pp. 229-238, 1997.

[2] H. S. Wu, J. Gil, and J. Barba, "Optimal segmentation of cell images," *Vision, Image and Signal Processing, IEE Proceedings,* vol. 145, pp. 50-56, 1998.

[3] O. Lezoray and H. Cardot, "Cooperation of color pixel classification schemes and color watershed: a study for microscopic images," *Image Processing, IEEE Transactions on,* vol. 11, pp. 783-789, 2002.

[4] P. Bamford and B. Lovell, "Unsupervised cell nucleus segmentation with active contours," *Signal Process,* vol. 71, pp. 203–213, 1998.

[5] A. Garrido and N. P. de la Blanca, "Applying deformable templates for cell image segmentation," *Pattern Recognition,* vol. 33, pp. 821-832, 2000.

[6] M. E. Plissiti, C. Nikou, and A. Charchanti, "Accurate localization of cell nuclei in Pap smear images using gradient vector flow deformable models," *Proceedings of 3rd International Conference on Bio-inspired Signals and Systems(BIOSIGNALS 2010),* pp. 284-289, 2010.

[7] H. S. Wu, J. Barba, and J. Gil, "A parametric fitting algorithm for segmentation of cell images," *Biomedical Engineering, IEEE Transactions on,* vol. 45, pp. 400-407, 1998.

[8] C. H. Lin, Y. K. Chan, and C. C. Chen, "Detection and segmentation of cervical cell cytoplast and nucleus," *Int. J. Imaging Syst. Technol,* vol. 19, pp. 260–270, 2009.

[9] S. F. Yang-Mao, Y. K. Chan, and Y. P. Chu, "Edge Enhancement Nucleus and Cytoplast Contour Detector of Cervical Smear Images," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on,* vol. 38, pp. 353-366, 2008.

[10] L. M. Isenstein, D. J. Zahniser, and M. L. Hutchinson, "Combined malignancy associated change and contextual analysis for computerized classification of cervical cell monolayers," *Anal Cell Pathol,* vol. 9, pp. 83-93, 1995.

[11] L. G. Koss and M. R. Melamed, "Diagnostic cytology of organs," in *Koss' Diagnostic Cytology and Its Histopathologic Bases*. vol. 1, 5 ed: Lippincott Williams & Wilkins, 2006, p. 186.

[12] C. Ronse and H. J. A. M. Heijmans, "A Lattice-Theoretical Framework for Annular Filters in Morphological Image Processing," *Applicable Algebra in Engineering, Communication and Computing,* vol. 9, pp. 45-89, 1998.

[13] E. S. Cibas and B. S. Ducatman, *Cytology: Diagnostic principles and clinical correlates*, 3 ed.: Saunders, 2009.

[14] F. Meyer and S. Beucher, "Morphological segmentation," *Journal of Visual Communication and Image Representation,* vol. 1, pp. 21-46, 1990.

[15] M. Peura and J. Iivarinen, "Efficiency of Simple Shape Descriptors," *Aspects of Visual Form, Proceedings of the Third International Workshop on Visual Form,* pp. 443-451, 1997.

[16] H. Tamura, S. Mori, and T. Yamawaki, "Textural Features Corresponding to Visual Perception," *Systems, Man and Cybernetics, IEEE Transactions on,* vol. 8, pp. 460-473, 1978.

[17] L. R. Dice, "Measures of the Amount of Ecologic Association Between Species," *Ecological Society of America,* vol. 26, pp. 297-302, 1945.