MARKER-CONTROLLED WATERSHED SEGMENTATION OF NUCLEI IN H&E STAINED BREAST CANCER BIOPSY IMAGES

M. Veta[1], A. Huisman[2], M.A. Viergever[1], P.J. van Diest[2], J.P.W. Pluim[1]

[1]Image Sciences Institute, University Medical Center Utrecht, P.O. Box 85500, 3508 GA Utrecht, The Netherlands
[2]Department of Pathology, University Medical Center Utrecht, P.O. Box 85500, 3508 GA Utrecht, The Netherlands

ABSTRACT

In this paper we present an unsupervised automatic method for segmentation of nuclei in H&E stained breast cancer biopsy images. Colour deconvolution and morphological operations are used to preprocess the images in order to remove irrelevant structures. Candidate nuclei locations, obtained with the fast radial symmetry transform, act as markers for a marker-controlled watershed segmentation. Watershed regions that are unlikely to represent nuclei are removed in the postprocessing stage. The proposed algorithm is evaluated on a number of images that are representative of the diversity in pathology in this type of tissue. The method shows good performance in terms of the number of segmented nuclei and segmentation accuracy.

Index Terms— Marker-controlled watershed, nuclei segmentation, breast cancer, histology images

1. INTRODUCTION

Assessing breast cancer prognosis relies largely on the Bloom-Richardson grading system. It is based on semi-quantitative scoring of the degree of tubule formation, nuclear pleomorphism, and mitotic rate [1]. Although prognostically strong, the system is traditionally done morphologically by simple eyeballing through the microscope, which has less than optimal reproducibility [2]. Automatic image analysis therefore may be useful here. Cell nuclei segmentation is an important first step towards automatic analysis of digitized histological slides of breast cancer biopsies. A successful nuclei segmentation algorithm can be used to automate nuclear pleomorphism scoring. Segmentation of nuclei can also be used, in a bottom-up manner, to locate the tumour regions within the slide or to assess the degree of tubule formation. The main challenges in achieving successful nuclei segmentation arise from the diversity and complexity of tissue appearance. Imperfect preparation, staining inhomogeneity and scanning artefacts also contribute to the intricacy of the problem.

In recent years, research in automatic analysis of histological slides has significantly increased. This interest notably includes automatic nuclei segmentation in breast cancer histopathology images. In Naik et al. [3] nuclei are segmented with template matching on an image produced by modelling low-level intensity information. Fatakdawala et al. [4] present a lymphocyte segmentation algorithm in which results from expectation-maximization clustering are used to initialize level-set active contour segmentation. Size heuristics and concavity detection are used to split contours that encompass multiple objects, thus avoiding undersegmentation. Cosatto et al. [5] detect candidate nucleus locations in breast cancer biopsy images using the Hough transform and evolve an active contour around each point. Malformed outlines are rejected using a support vector machine classifier with features describing the shape, texture and fit of the boundary to the underlying image. These segmentations are, in turn, used to distinguish between benign and malignant regions within the slides.

Watershed segmentation is a method particularly suited for nuclei segmentation. The results of the classical watershed segmentation can be significantly improved by modifying the segmentation function (topographical relief) to contain regional minima only at specific locations that mark the objects of interest and the background. These markers can be obtained in a variety of ways and the process is usually application-dependent. It is important to have one-to-one correspondence of the foreground markers and the objects of interest. Producing two markers for a single object leads to oversegmentation and failure to mark an object leads to undersegmentation.

In Huang et al. [6] a marker-controlled watershed method for segmentation of nuclei in hepatocellular carcinoma biopsy images is presented. The regional minima of a preprocessed image are used as foreground markers and the gradient magnitude as a segmentation function. The skeleton by influence zones (SKIZ) of the foreground markers is used to mark the background. This method works well only if there is considerable separation between the nuclei. However, in hematoxylin and eosin (H&E) stained breast cancer biopsy images the nuclei are often overlapping or clustered closely together. Accordingly, many of them might not belong to a regional minimum or multiple objects might belong to a single regional minimum, which leads to bad segmentation results.

We propose a method similar to [6] that meets the ob-
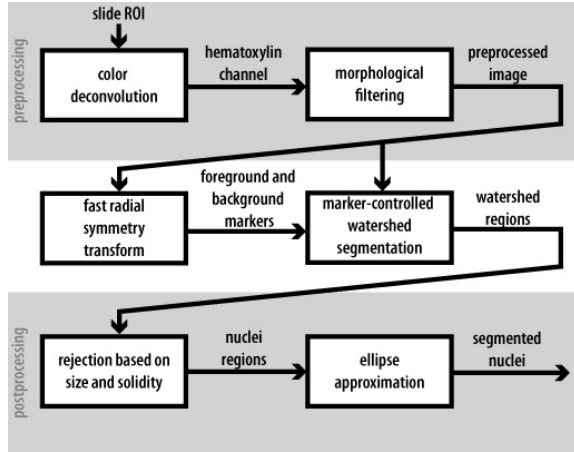
Fig. 1. Block-diagram of the proposed method

jective of nuclei segmentation in H&E stained breast cancer biopsy images by applying the fast radial symmetry transform [7] to produce markers for the watershed segmentation. A similar operator [8] was used in [9] to decompose regions of clustered nuclei obtained by fuzzy c-Means clustering and level-set segmentation in H&E stained prostate cancer biopsy images. The method is unsupervised, automatic and computationally efficient, and requires only a small number of parameters to be defined. We also present a scheme that can be used to make a trade-off between the number of segmented nuclei and the accuracy of the segmentation.

2. MATERIALS AND METHODS

This study includes 19 H&E stained slides of breast cancer biopsies. The slides were chosen by an experienced pathologist to represent the diversity in pathology in this type of tissue. The digitization was done using a ScanScope XT whole slide scanner (Aperio, Vista, CA, USA) at magnification of 40x and resolution of 0.25 μm/pixel at our pathology department [10].

From each slide, a region of interest (ROI) of size 1024x1024 pixels was chosen. These ROIs were selected from larger regions previously scored by a pathologist for nuclear pleomorphism according to the Bloom-Richardson grading system. Out of the 19 selected ROIs, six belong to score one nuclear pleomorphism region, 11 to score two region and two to score three region. All results presented in this paper are obtained on these ROIs, which represent the diversity in appearance of the tissue and the nuclei well, while still being a manageable set suitable for manual annotation for verification purposes. Figure 2a gives a section of one of the ROIs with more complex tissue architecture, and thus more challenging for segmentation.

The nuclei segmentation method presented here consists of three main steps: preprocessing, marker-controlled water-

shed segmentation and postprocessing. A block-diagram with the individual steps of the method is given in Fig. 1. The parameters that need to be defined are the smallest and largest expected semi-axial length of the nuclei in the image, a_{min} and a_{max}, the radial strictness α and the solidity rejection threshold used in the postprocessing stage.
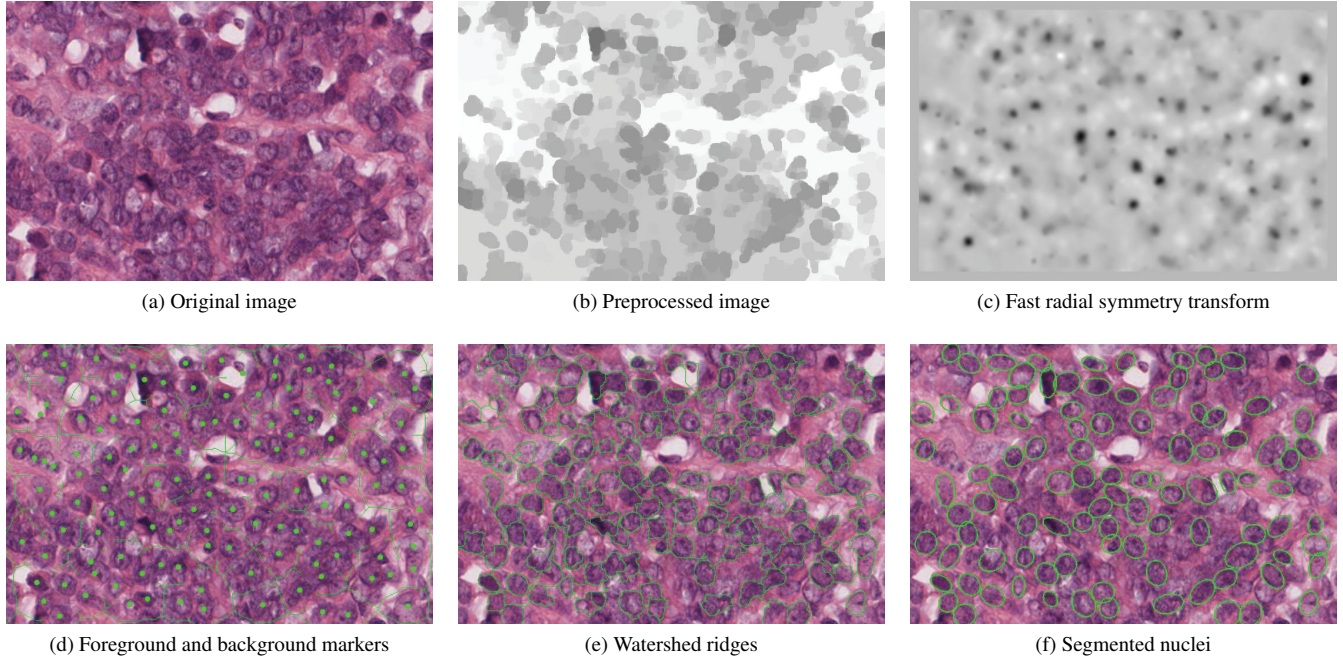
2.1. Preprocessing

The aim of the preprocessing is to remove irrelevant content while preserving the boundaries of the nuclei. This starts with separation of the hematoxylin and eosin stains with colour deconvolution [11]. Since the nuclei are dyed by hematoxlin, only this channel is used for the segmentation. The separated hematoxylin channel still contains a lot of noise and irrelevant structures. These present obstacles for the marker extraction and segmentation, and need to be filtered out from the image. This is done with a series of morphological operations. First, opening by reconstruction is applied, which is followed by closing by reconstruction. Opening and closing by reconstruction, as contrasted with morphological opening and closing, preserve the shape of the objects that are not removed. The structuring element used is a disk with radius a_{min}. In this way, only details that are smaller than the expected size of the nuclei are removed. Furthermore, to eliminate the small protrusions around the nuclei boundaries, morphological closing with a very small structuring element is performed. This set of operations successfully eliminates most of the noise and unwanted structures in the image (Fig. 2b).

2.2. Marker-controlled watershed segmentation

The fast radial symmetry transform [7] is a computationally efficient, non-iterative procedure that operates along the direction of the image gradient to infer centres of radial symmetry. Since nuclei are somewhat radially symmetric objects, this operation is suited for their localization. To produce candidate nuclei locations, we use the orientation-based version of this transform, which discards gradient magnitude information and relies only on the orientation. This can be beneficial in the case of low contrast between the nuclei and the background. The transform is computed for the preprocessed image for the range of radii $[a_{min}, a_{max}]$. The radial strictness parameter α defines how strict the radial symmetry must be for the transform to return a high interest value. Dark regions of symmetry in the input image correspond to minima in the resulting transform image (Fig. 2c) and can be easily detected with non-minima suppression.

The foreground map is obtained by placing a disk-shaped marker at each detected minimum (Fig. 2d). To produce a tentative background map, a naïve assumption is made that at the location of each interest point there is a circular nucleus with radius a_{max}. The morphological skeleton of the

(a) Original image   (b) Preprocessed image   (c) Fast radial symmetry transform

(d) Foreground and background markers   (e) Watershed ridges   (f) Segmented nuclei

Fig. 2. Example segmentation

tentative background obtained in this way is used as a background marker. The Sobel gradient magnitude of the preprocessed image, which is used as a segmentation function, is modified by imposing regional minima at locations specified by the foreground and background maps. Subsequently, watershed segmentation is performed to obtain candidate nuclei regions (Fig. 2e).

2.3. Postprocessing

In the postprocessing stage, regions produced by the background markers are discarded, as well as regions that are too small (area smaller than πa_{min}^2) or too big (area larger than πa_{max}^2). Since it is expected that the nuclei are somewhat convex objects, regions of very small solidity (ratio of the area and the convex area of the region) are discarded. The solidity of the region as a feature that describes the "quality" of the segmentation is further explored in Section 3. Finally, with the goal of obtaining more regular contours, the segmented nuclei regions are approximated by ellipses (Fig. 2f).
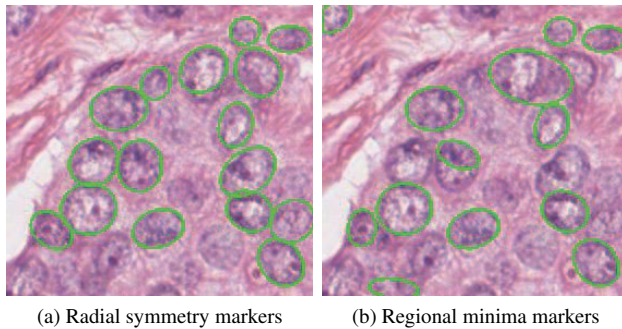
3. RESULTS AND DISCUSSION

The proposed method was used to segment nuclei in all 19 ROIs. In all cases, the same parameter set was used ($a_{min} = 12$ pixels, $a_{max} = 28$ pixels, $\alpha = 3$). The solidity rejection threshold was set at 0.8. All produced segmentations were manually labelled as correct or erroneous by one of the authors. A segmentation was labelled as erroneous if it was an

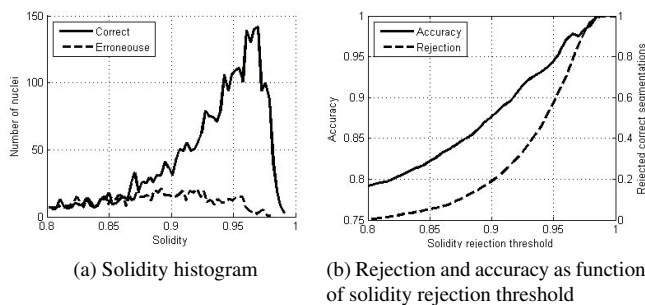**Table 1**. Accuracy and number of segmented nuclei

	Radial symmetry markers	Regional minima markers
Average accuracy per ROI	81.5% ($\pm$ 9.1%)	81.0% ($\pm$8.2%)
Overall accuracy	79.2%	79.6%
Total number of correct segments	2900	2306

apparent over- or undersegmentation or an object that does not represent a nucleus. For comparison, the same was done for watershed segmentation using regional minima markers as in [6] and identical pre- and postprocessing steps. The results are summarized in Table 1.

The average percentage of correct segmentations per ROI in both cases is similar: 81.5% for the radial symmetry markers and 81.0% for the regional minima markers. However, the radial symmetry markers produced 25.8% more segmentations compared to the regional minima markers. From the analysis of individual ROIs, it can be concluded that when there is clear separation of the nuclei, both algorithms perform similarly. The radial symmetry markers perform a lot better in the case of more complex tissue architecture, such as tubule formations and clustered nuclei (Fig. 3). In individual cases with more complex tissue architecture, like the one presented in Fig. 2, the proposed method segments up to 60% more nuclei compared to the regional minima markers.

(a) Radial symmetry markers   (b) Regional minima markers

Fig. 3. Comparison between radial symmetry and regional minima markers

(a) Solidity histogram   (b) Rejection and accuracy as function of solidity rejection threshold

Fig. 4. Solidity as a rejection criterion

One way to boost the accuracy of the segmentation is to detect more of the errors in the postprocessing stage. Based on the annotated segmentations, we have analysed a number of features for their power to discriminate between correct and erroneous nuclei regions. This analysis included a number of texture and shape features, as well as features that model the intensity profile at the region boundary. Among all examined features, the solidity of the region proved to be the most discriminative. Although rejection based on region solidity is already included in the postprocessing stage, the threshold value is set to reject only the extreme cases. Figure 4a gives the distribution of the solidity for the regions of the annotated nuclei. It can be observed that by setting a higher rejection threshold, a large portion of the erroneous segmentations will be discarded, while keeping most of the correct segmentations. A plot of the overall segmentation accuracy and the rejection rate as functions of the solidity threshold is given in Fig. 4b. By varying the solidity rejection threshold, a compromise can be made between the number of segmented nuclei and the accuracy of the segmentation.

4. CONCLUSIONS AND FUTURE WORK

In this paper we have described an unsupervised method for segmentation of nuclei in breast cancer biopsy images. The method is automatic and computationally efficient, and requires only a small number of parameters to be defined. Evaluation on a set of images with a variety of tissue appearances showed that a large number of nuclei are segmented with good accuracy. Future work will include the use of this method to automate the assessment of the degree of nuclear pleomorphism and tubule formation as part of an image analysis-based more objective Bloom-Richardson system for grading invasive breast carcinoma. The fact that this can be done on whole slide images makes it possible to integrate this application into the workflow of pathology practice.

5. REFERENCES

[1] C.W. Elston and I.O. Ellis, "Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up.," *Histopathology*, vol. 19, no. 5, pp. 403–410, 1991.

[2] P. Robbins et al., "Histological grading of breast carcinomas: a study of interobserver agreement," *Human Pathology*, vol. 26, no. 8, pp. 873–879, 1995.

[3] S. Naik et al., "Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2008, pp. 284–287.

[4] H. Fatakdawala et al., "Expectation-maximization-driven geodesic active contour with overlap resolution (EMaGACOR): Application to lymphocyte segmentation on breast cancer histopathology," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1676–1689, 2010.

[5] E. Cosatto et al., "Grading nuclear pleomorphism on histological micrographs," in *International Conference on Pattern Recognition (ICPR)*, 2008, pp. 1–4.

[6] P.W. Huang and Y.H. La, "Effective segmentation and classification for HCC biopsy images," *Pattern Recognition*, vol. 43, no. 4, pp. 1550–1563, 2010.

[7] G. Loy and A. Zelinsky, "Fast radial symmetry for detecting points of interest," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, pp. 959–973, 2003.

[8] O. Schmitt and M. Hasse, "Radial symmetries based decomposition of cell clusters in binary and gray level images," *Pattern Recognition*, vol. 41, no. 6, pp. 1905–1923, 2008.

[9] A. Hafiane et al., "Clustering initiated multiphase active contours and robust separation of nuclei groups for tissue segmentation," in *International Conference on Pattern Recognition (ICPR)*, 2008.

[10] A. Huisman et al., "Creation of a fully digital pathology slide archive by high-volume tissue slide scanning," *Human Pathology*, vol. 41, no. 5, pp. 751–757, 2010.

[11] A.C. Ruifrok and D.A. Johnston, "Quantification of histochemical staining by color deconvolution.," *Analytical and quantitative cytology and histology*, vol. 23, no. 4, pp. 291–299, 2001.