Towards Generalized Nuclear Segmentation in Histological Images

Abhishek Vahadane and Amit Sethi

Abstract— **Computer aided diagnosis in cancer pathology (computational pathology) using histological images of biopsies is an emerging field. Segmentation of cell nuclei can be an important step in such image processing pipelines. Although seeded watershed segmentation is a simple and computationally efficient segmentation technique, it is prone to errors like over-segmentation when applied to histological images. We report specific enhancements to this technique to improve segmentation of cell nuclei in histological images. Foreground seeds were generated by fast radial symmetry transform (FRST). Otsu thresholding was used on enhanced image to estimate tentative foreground map. Background markers were computed from the tentative foreground map. False detections in the segmented output were removed by logical AND with the tentative foreground map. Using these enhancements nuclear segmentation was significantly improved on histological images (H&E stained breast and intestinal tissue images, Feulgen stained images of prostate tissues).**

Keywords: seeded watershed segmentation, histological images, segmentation of cell nuclei.

I. Introduction

Cancer is a major health problem in many parts of the world [8]. The breast, prostate, lung, and colorectal cancer are the leading types in number of estimated new cases and deaths. A key to reduce the mortality rate in cancer is early detection. Ideally, diagnosis should have no false positives nor false negatives. Biopsy, which involves extracting cells from the suspected tissue and examining them under a microscope, is one of the most accurate but invasive diagnosis technique. Computationally-assisted diagnosis based on biopsies is an emerging research field. It has the potential to relate previously unsuspected visual features to clinical diagnosis and future outcome as shown by a recent study [1].

Segmentation of objects from an image or video is an important and first step for further computational interpretation. It is a process of multi-object (nuclei) extraction from the background (cytoplasm and stroma). In histological images, the objects of interest are nuclei. Diagnosis related features like mitotic count, nucleopleomorphism etc. can be extracted from segmented nuclei. Segmentation of cell nuclei is a challenging problem due to wide size and shape variations of the nuclei [4], complex heterogeneity within nucleus and in stroma, improper staining, imaging artifacts, variation in density of the nuclei in an image and across different images, and the occluded nuclei.

Nuclear segmentation has been approached by using classical image segmentation methods such as thresholding, region based approaches (like region growing), energy minimization techniques (like snake model, active contour level sets), and the classification based segmentation [2]. A machine learning based framework was used by A. Veillard et.al [9] to find the probability map of pixels belonging to the nuclei. Linear discriminant analysis (LDA) was performed on Law's texture measure (180 dimensional feature space) for optimizing inter- to intra-scatter ratio and it was followed by an application of softmax function to map it to the probability. This method is not generalizable across different histological image types in the sense that different classifier hypothesis are required for the different training image types. Also, the labeled image data is generated by the manual nuclei annotations. Assuming a set of hypothesis for different training sets, we need to decide on a hypothesis for the test image from this set of hypothesis. In general, machine learning methods degrade in performance if sufficient training examples are not used.

After a survey on cell segmentation, E. Bengtsson proposed a comparatively robust and an upgraded seeded watershed model [2]. Here, the seeds are generated by H-maxima transform of the distance transform of the background map. However, the parameters are manually tuned. They reported that use of intensity, gradient, connectivity, and shape information in the watershed model leads to robust segmentation methods. Another work by P. Shete [7] in avoiding over-segmentation used transformation of more than one markers to one per nucleus. However, this method is very specific due to use of color thresholding and is prone to errors by removing wrong markers. An analysis of past five decades of research on cell segmentation by Erik [6] reported that cell segmentation approaches so far are combinations of discussed or basic approaches tailored to a specific application. They are biased to application at hand.

A recent work on nuclear segmentation by M.Veta et.al [10] proposed an unsupervised segmentation in which marker controlled watershed transform is seeded with new foreground marking scheme based on fast radial symmetry transform (FRST) [5]. They reported improved results than regional minima marking scheme that is popular in watershed segmentation. Here, the post-processing includes solidity rejection, small and big sized region removal for the further improvement in the segmentation accuracy. The same framework is reflected in the work by Huang and Lai [3] with some

Manuscript received July 30, 2013. This work was supported by MHRD (Govt. of India) through research funding

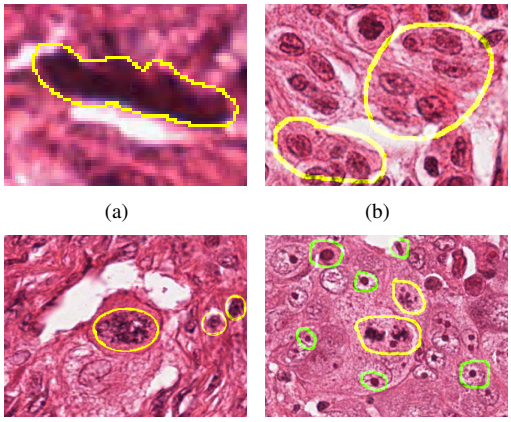
Abhishek Vahadane, Research Scholar, Department of Electronics & Electrical Engineering, Indian Institute of Technology Guwahati, India 781039 vahadane@iitg.ernet.in

Amit Sethi, Assistant Professor, Department of Electronics & Electrical Engineering, Indian Institute of Technology Guwahati, India 781039 amitsethi@iitg.ernet.in

978-1-4799-3163-7/13/$31.00 ©2013 IEEE

difference in the marking scheme. It uses a regional minima based foreground markers. This marking scheme works well with a simple tissue structure but creates problems with more complex tissue appearance and with the densely packed nuclei.

Thus, most of literature on nuclear segmentation fits the model to a specific application. Hence, there is a need of a powerful generalizable segmentation algorithm. A generalizable algorithm is such that it is applicable for a wide range of applications (different types of microscopic/histological stained images), while achieving high sensitivity and specificity. Over-segmentation is a central problem in watershed segmentation. For solving this problem and move to generic segmentation, our technique incorporates the following: 1. Gaussian smoothing for high frequency noise removal and blurred nuclei segmentation 2. Proposed background markers based on image information to reduce over-segmentation 3. Tentative foreground map for removing false detections.



(a) (b) (c) (d)

Fig. 1. Segmentation challenges: (a) Large sized nucleus, (b) Very blurred nuclei, (c) and (d) Rough nuclei boundaries, size and shape variations

II. Image Database and Challenges

We tested the proposed algorithm on three histological image data sets: 50 H&E (hematoxylin & eosin) stained breast (MITOS dataset), 65 H&E stained gastrointestinal, and 4 Feulgen stained prostate histological images. See acknowledgment. In H&E stain, the hematoxylin binds the nuclear DNA to render blue color and eosin binds the cytoplasm, other structures to give pink color. In case of Feulgen stained images, nuclei are perceived to be blue color structures while the background (cytoplasm) is white. The staining process increases the contrast between the nuclei and background (cytoplasm and other structures), thus helping the visual inspection.

Some of the challenges mentioned in the second paragraph of the introduction section can be seen in Fig.1. There are appearance variations across different regions in whole slide image and across different patient slides. Hence, it is necessary to use local adaptive enhancement techniques like local histogram equalization, adaptive thresholding etc. Occluded nuclei pose the problem of under-segmentation.

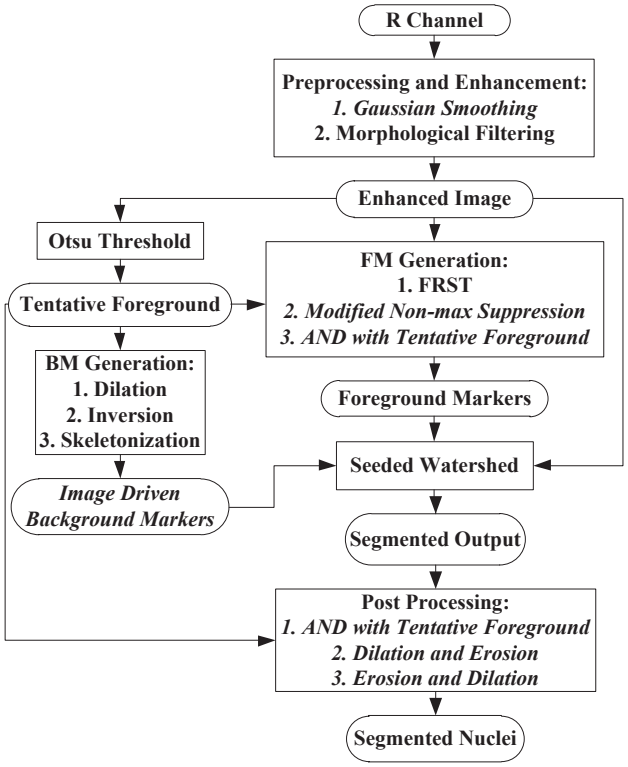


Fig. 2. Flow diagram for the improved seeded watershed segmentation. Italic texts are our improvements to seeded watershed segmentation.

III. Proposed Algorithm

A flow diagram for modified seeded watershed algorithm is shown in Fig.2. It shows improved performance with respect to over-segmentation, maximal nuclei segmentation, smooth boundary preservation, and false detections. This improved performance is due to inclusion of following stages: Gaussian smoothing for enhancement with its implicit advantages, an image driven background markers, and Otsu thresholding based tentative foreground estimation. In this section we describe the enhancements to the seeded watershed algorithm and their rationale in more detail. Typical parameter values used are given in Table I.

A. Channel selection and pre-processing

In both H&E and Feulgen stained images, R channel was used because it shows significant contrast between nuclei and background. The preprocessing was for the high frequency noise removal using Gaussian smoothing. It enhances the segmentation of diverse and poorly stained nuclei. The parameter (σ) of Gaussian kernel was kept low in order not to miss small sized nuclei.

B. Image enhancement

Morphological filtering (opening followed closing by reconstruction) was introduced as an image enhancement step to further smoothen the high intensity background (mainly, cytoplasm) and the low intensity foreground (mainly, nuclei), while preserving the structural or edge information. A high

degree of background smoothing is required on the MITOS dataset, so the opening operation was performed with a large sized structuring element (SE). Therefore, morphological filtering parameters for MITOS dataset are chosen to be (30,5) as shown in table 1. As the segmentation is a function of the gradient magnitude of the image, the enhancement parameters affect the performance of the segmentation. It is found that best enhancement parameters are different for differently stained histological images.

TABLE I
TYPICAL PARAMETER VALUES USED

Processing stage	Parameters	Value in pixels
Gaussian smoothing	Variance, σ	3
	Window size	10
Morphological filtering	Opening by reconstruction disk SE size	30
	Closing by reconstruction disk SE size	5
FRST	Set of radii	5,7,9,11,13,15
	Strictness of symmetry, α	2.0 (unitless)
Non-max suppression	Window size	30
	Looping window size	40
BM from FM	Dilation disk size	20
Image driven BM	Dilation disk size	7

C. Foreground and background marking

Assuming radial symmetry around the nuclei we detected foreground markers as local maximas of FRST of the enhanced image using a sparse set of radii (5,7,9,11,13,15) that were manually selected to be close to the size of the nuclei. Spurious multiple maximas were removed using non-maximum suppression parameterized by a square window of size 30 that accommodated most of the nuclei. To remove multiple maximas from asymmetrical and large size nuclei (like Fig.1(a)), a loop was run for cascaded non-maximal suppression with square mask of size 40, which converged when there was no change in the current and previous marker image.

The background markers (BMs) can be easily generated from foreground markers (FMs) obtained by applying FRST. However, the nuclei in histological images are neither truly circular in shape nor symmetrical. Even after repeated non-maximal suppression, some nuclei may have more than one marker. Thus, if we use those FMs to estimate BMs, there is possibility of artifacts such as over-segmentation, an incomplete nuclei detection, and false negatives. We added an image driven background marking scheme to avoid the above mentioned artifacts. This scheme employ estimated foreground map to compute BMs. The foreground map was a binary image obtained after Otsu threshold based segmentation of the enhanced image. The skeleton of dilated and inverted foreground map results in the true BMs. Here, dilation is done to avoid BMs on nuclei.

D. Segmentation and post-processing

Watershed segmentation was run using FMs and BMs as local gradient magnitude minima to speed up the algorithm. Obtained watershed lines represented boundaries of the nuclei. In post-processing, false detected regions were removed by use of tentative foreground map. The regions, separated by less than two pixels were connected by use of morphological dilation-erosion with small sized SE. Also, small segmented regions (less than disk SE of radius equal to three pixels) were removed.

IV. RESULTS AND DISCUSSIONS

The table I includes list of the parameters with their typical value used for the experiment. It was found that best setting for morphological filtering parameters varied by type of image (different average nuclear size). Gaussian smoothing is essential pre-processing stage depicted in Fig.3. It shows that blurred (poorly stained) nuclei are segmented if we incorporate smoothing before enhancement.

An improvement in the segmentation and the background marking by use of an image driven background marking scheme is shown in Fig.4. The FM based background marking scheme is inaccurate in marking the background and it increases false negatives unlike image driven BMs. Nuclear segmentation result by the proposed algorithm on Feulgen stained image is shown in Fig.5. The prostate tissue Feulgen stained images have uniform background (cytoplasm and other structures). Therefore the best setting of morphological filtering parameters for opening and closing are respectively 20 and 5. Quantitative segmentation results (table 2) were obtained only on 50 intestinal microscopic images due to accessibility of ground truth (expert-marked nuclei). Under-segmentation is moderately high, which shows that occluded nuclei is an unsolved problem.

TABLE II
PERFORMANCE MEASURES ON INTESTINAL MICROSCOPIC IMAGES

Recall	Precision	Oversegmentation	Undersegmentation
93.33%	90.79%	8.63%	15.49%

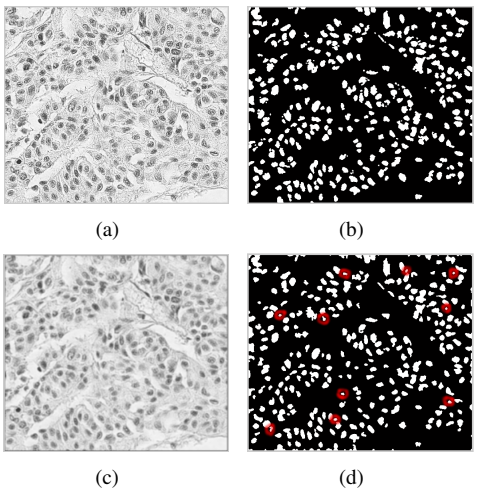


Fig. 3. Improvement in nuclear segmentation by Gaussian smoothing (a) R channel of input color image, (b) Segmentation without smoothing, (c) Smoothed R channel, and (d) Segmentation with smoothing. Blurred nuclei missed by (b), are retained in (d), as shown by annotations in (d).

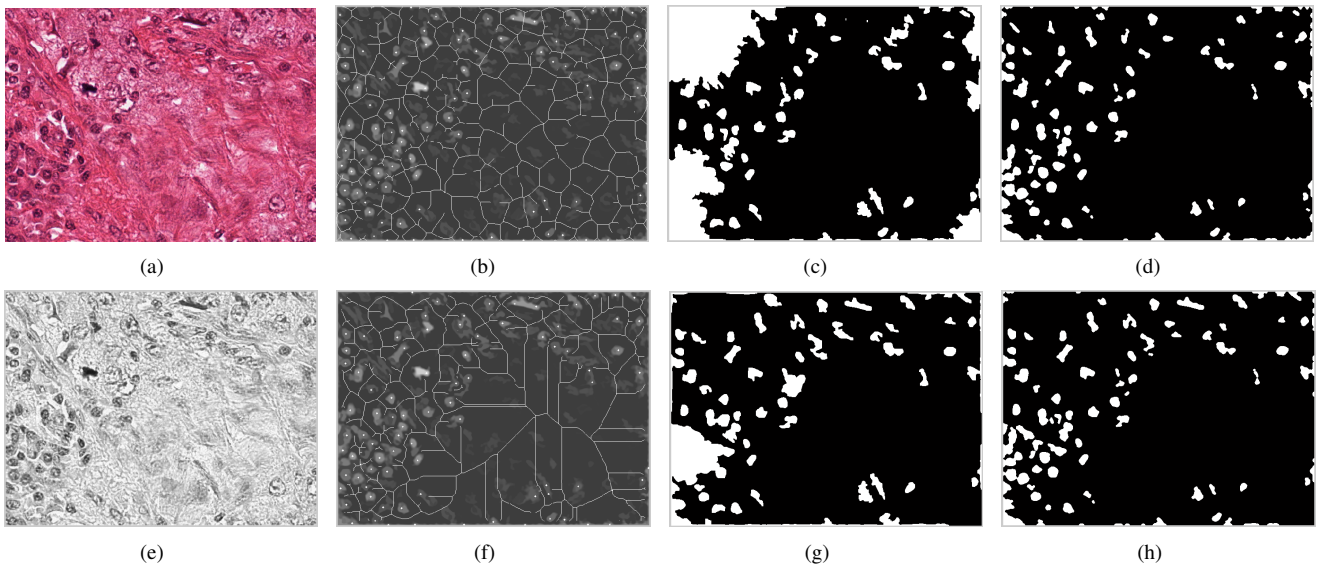


Fig. 4. Results of the proposed algorithm (a) Input color image, (e) R channel, (b, c, d)- results for background markers obtained from foreground markers, and (f, g, h)- results for proposed image driven background markers. (b, f) Enhanced image + foreground markers (dots) + background markers (lines on the background), (b) shows wrong background markers as they cross objects of interest unlike by proposed image driven background marking shown in (f). (c, g) Seeded watershed segmentation output, (g) shows less false detections, and (d, h) Processed output, between (d) & (h), (d) shows some nuclei missing unlike by proposed algorithm as shown in (h).

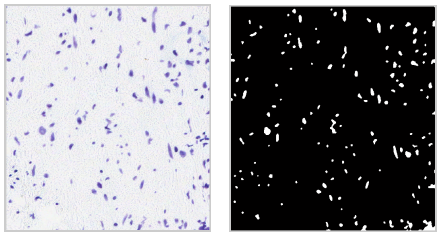


Fig. 5. Segmentation result on the Feulgen stained image (a) Input Feulgen stained image, (b) Segmented output.

V. CONCLUSIONS AND FUTURE DIRECTIONS

Seeded watershed scheme is a good choice for the multi-object segmentation, such as nuclei in histological images, due to its low computational complexity in comparison with image energy based models for segmentation. However, the watershed model is prone to over-segmentation. We improved its performance for histological images by introducing specific enhancements, and improved foreground and background marking. A tentative foreground map was used to avoid false detections.

The problem of segmenting partially occluded nuclei remains unsolved. Further improvements are required to make the proposed nuclear segmentation algorithm fully automated and more robust. For example, one can try using local or adaptive enhancement techniques and exploring image dependent marking schemes like the proposed background marking scheme.

VI. ACKNOWLEDGMENT

We thank providers of image datasets: Prof. P. Gann (University of Illinois at Chicago, USA) for Feulgen stained prostate, ICPR contest organizers (ipal.cnrs.fr/ICPR2012) for H&E stained breast, and MIP-I contest organizers (www.utee.feec.vutbr.cz/files/mikulka/challenges.html) for H&E stained gastrointestinal histological images.

REFERENCES

[1] A. H. Beck, A. R. Sangoi, S. Leung, R. J. Marinelli, T. O. Nielsen, M. J. van de Vijver, R. B. West, M. van de Rijn, and D. Koller, "Systematic analysis of breast cancer morphology uncovers stromal features associated with survival," *Sci Transl Med*, vol. 3, no. 108, pp. 108–113, 2011.

[2] E. Bengtsson, C. Wahlby, and J. Lindblad, "Robust cell image segmentation methods," *Pattern Recognition and Image Analysis C/C of Raspoznavaniye Obrazov I Analiz Izobrazhenit*, vol. 14, no. 2, pp. 157–167, 2004.

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

[4] M. Kulikova, A. Veillard, L. Roux, and D. Racoceanu, "Nuclei extraction from histopathological images using a marked point process approach," in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2012, pp. 831 428–831 428.

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

[6] E. Meijering, "Cell segmentation: 50 years down the road [life sciences]," *Signal Processing Magazine, IEEE*, vol. 29, no. 5, pp. 140–145, 2012.

[7] P. G. Shete, G. K. Kharate, and S. C. Rege, "Breast cancer cell detection using digital image processing," *International Journal of Engineering*, vol. 1, no. 9, 2012.

[8] R. Siegel, D. Naishadham, and A. Jemal, "Cancer statistics, 2012," *CA: a cancer journal for clinicians*, vol. 62, no. 1, pp. 10–29, 2012.

[9] A. Veillard, M. S. Kulikova, and D. Racoceanu, "Cell nuclei extraction from breast cancer histopathology images using color, texture, scale and shape information," *11th European Congress on Telepathology and 5th International Congress on Virtual Microscopy*, 2012.

[10] M. Veta, A. Huisman, M. A. Viergever, P. van Diest, and J. P. Pluim, "Marker-controlled watershed segmentation of nuclei in h&e stained breast cancer biopsy images," in *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium*. IEEE, 2011, pp. 618–621.