Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Image Analysis for Neuroblastoma Classification: Segmentation of Cell Nuclei

Metin N. Gurcan, *Senior Member, IEEE*, Tony Pan, Hiro Shimada, and Joel Saltz

***Abstract*— Neuroblastoma is a childhood cancer of the nervous system. Current prognostic classification of this disease partly relies on morphological characteristics of the cells from H&E-stained images. In this work, an automated cell nuclei segmentation method is developed. This method employs morphological *top-hat by reconstruction* algorithm coupled with hysteresis thresholding to both detect and segment the cell nuclei. Accuracy of the automated cell nuclei segmentation algorithm is measured by comparing its outputs to manual segmentation. The average segmentation accuracy is 90.24±5.14%.**

I. Introduction

Neuroblastoma, a pediatric cancer of the peripheral nervous system, comprises one of the most common tumors in infants and children [1]. The diagnosis, prognosis and treatment planning depend on histopathological classification of the tumor; the current classification system was developed by Shimada [2]. The Shimada classification system uses morphological information such as presence and absence of Schwannian cell development, the relative count of tumor cells in mitosis and karyorrhexis [3]. Because the type and state of the cells are very important in this classification system, accurate automatic segmentation of cells is useful. As the initial part of a system to aid pathologists in their classification system, a cell nuclei segmentation algorithm is developed. Once the cell nuclei are segmented, they can be used as the first step for the segmentation of cytoplasm; for understanding the signs of mitosis and karyorrhexis as well as for differentiation of different cell types.

Cell nuclei segmentation is a significant image analysis problem for pathology. While developing an image analysis system for the automated characterization of prostate cancer, a simple cell nuclei segmentation solution is suggested in [4]. In that work, the RGB image is converted to the hue-saturation-intensity (HIS) color space. The authors heuristically found that the H, I and S values are within certain a range of values for the nuclei. For another automated prostate cancer classification system, the nuclei are segmented in the RGB domain by examining the ratio of intensity values of red (R) and blue (B) channels as well the length to width ratio of the detected objects [5]. A cell nuclei segmentation algorithm was also developed for the automated evaluation of Her-2/neu status in breast tissue [6]. In that algorithm, only the blue channel information is used, and the initial detection and segmentation of nuclei is accomplished using Otsu's automated thresholding scheme. The thresholded image contains mainly the nuclei with some holes due to the irregularity of the intensity values. The holes are filled and clustered cell nuclei are separated from each other using the watershed segmentation algorithm.

This work is part of an automated image analysis system to assist pathologists in their histopathological classification of tumor samples. The histological sections are stained with haematoxylin and eosin (H&E) stain. Haematoxylin colors structures containing nucleic acid such as cell nuclei to a blue-purple hue, and eosin colors intra-cellular and extra-cellular protein to a bright pink hue; the red-blood cells are stained intensely red. The nuclei segmentation method relies on an adaptive thresholding method based on the morphological operations instead of absolute value thresholds or thresholds that are determined using the global image characteristics. Using an adaptive scheme has the advantage that it is more robust to variations in the staining process, which can occur if the histological sections are stained at different institutions.

Section II introduces the morphological operations that are used to detect the cell nuclei regions as well as the segmentation step which consists of hysteresis thresholding as well as post-processing operations. Experimental results are presented in Section III.

II. Nuclei Segmentation Method

A. Morphological segmentation algorithm

Figure 1 gives an overall description of the algorithm. The first step of the algorithm is the color plane decomposition. Because the H&E stained nuclei has the highest contrast in the Red (R) image plane as compared to the blue (B) and green (G) image planes, this plane is extracted and subsequent operations are carried out on the 8-bit R plane image.

This work was supported in part by NIH grant P20 EB000591.

M. N. Gurcan, T. Pan, J. Saltz are with Biomedical Informatics Department, The Ohio State University, Columbus, OH 43210 USA (phone: 614-292-1084; fax: 614-688-6600; e-mail: gurcan.1@ osu.edu).

H. Shimada is with Children's Hospital Los Angeles, Los Angeles, CA 90027 USA. (e-mail: hshimada@chla.usc.edu)

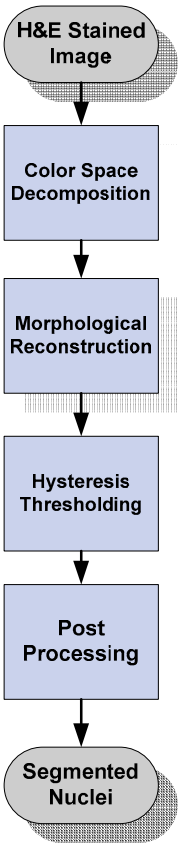

Fig.1. Flowchart of the overall nuclei segmentation algorithm.

The cell nuclei are relatively darker structures in the R color plane. First, the R image is complemented (i.e. all pixel values are subtracted from the maximum intensity value of 255) so that nuclei are relatively bright regions. The next step is morphological reconstruction of this image. Following this, morphological top-hat transform is used to detect these high intensity regions. The *top-hat by reconstruction* transform involves the subtraction of a morphologically reconstructed version of the image from the original image:

$$\varphi_{recth}(I) = I - \rho_S(I^m \mid I)$$

where $\varphi_{recth}(I)$ represents the top-hat by reconstruction operator for image *I*, $\rho_S(I^m \mid I)$ is the morphological reconstruction operator with the structuring element of S, I^m is the marker image. The reconstruction operator removes selected peaks from the image:

$$\rho_S(I^m \mid I) = \coprod_{n \geq 1} \delta_S^n(I^m \mid I)$$

where $\coprod$ denotes the supremum operation applied $n \geq 1$ times until stability of the *reconstruction by dilation* is reached by means of the dilation operator, $\delta_S^n(I^m \mid I)$. The reconstruction by dilation operator is defined as

$$\delta_S^1(I^m \mid I) = (I^m \oplus S) \wedge I$$

where $\oplus$ denotes the dilation operation, $\wedge$ is the point-wise infimum operation. The dilation is defined as

$$(I \oplus S)(x, y) = \coprod_{(\xi,\eta) \in S} I(x - \xi, y - \eta), \qquad (x, y) \in Z^2$$

where Z^2 is the two-dimensional discrete space.

Figure 2 and 3 show the part of a digitized pathological slide and complement of the R plane, respectively. From Figure 3, it is apparent that the cell nuclei have relatively higher intensities. The reconstructed image and the top-hat by reconstruction image are shown in Figures 4 and 5, respectively. The top hat transform reduces the effects of non-nuclei structures, hence increasing the signal-to-noise ratio, which is important for the subsequent hytheresis thresholding operation.

After the top-hat transform, the cell nuclei are segmented using a hysteresis thresholding method [7]. The hysteresis thresholding operation is modified for this application to remove most of the small, high-intensity structures in close proximity to the cell nuclei. To achieve this, after the initial thresholding, the nuclei candidates are labeled, and their areas are calculated. Those cell nuclei with areas less than an experimentally determined threshold were eliminated from the second thresholding stage. Subsequently, a set of post-processing operations are carried out. These operations include filling the holes in the segmented nuclei, exclusion of red-pixel values with high intensity values and morphological smoothing as well as removing nuclei cell candidates with small area from consideration. The holes within the nuclei are created because of the non-uniform intensity values within the cell nuclei. Some of the internal pixel values are lower than the lowest hysteresis threshold value, which create holes after thresholding. Such holes are identified and filled. The red pixel values, which mainly correspond to blood vessel cells, are recognized by examining the R to B and R to G plane pixel value ratios. Because red blood vessel cells have mainly red components, these ratios are relatively higher for these types of structures. For morphological smoothing, a morphological closing operation is used.

In some cases, the cell nuclei are very close to each other. When segmented, these cells form a cluster instead of being separated. As the final step of the algorithm, such clusters are divided into individual cells using a modified form of watershed transform. Figure 6 illustrates the output of the watershed segmentation [8] for the division of the nuclei

clusters into individual nuclei.. Figure 7 shows the final output of the segmentation where the boundary of each cell is shown in green; Figure 8 illustrates the segmentation from another region of interest from a stroma region.

B. Evaluation of the nuclei segmentation

The results of the manual segmentation were compared using the overlap scores:

$$OS_1 = \frac{M \cap A}{M \cup A}$$

$$OS_2 = \frac{M \cap A}{|M| + |A|} \times 2$$

Where M and A are manual and automated segmentation; $\cap$, $\cup$, $|\ |$ denote the intersection, union, and the cardinal operations (i.e. the total number of "on" pixels in the segmented area), respectively.

III. Experimental Results

IRB approval was obtained for this study. The slides were embedded in paraffin and were cut at a thickness of 5 µm according to commonly used Children's Oncology Group protocols. The tissue was fixed on a slide, and the slides were scanned using a ScanScope T2 digitizer (Aperio, San Diego, CA) at 40x magnification and then compressed using JPEG compression at approximately 1:40 compression ratio.

The structuring element *S* used in this study is a disk of radius 20. The upper hysteresis threshold was 60, and the lower hysteresis threshold was 45.

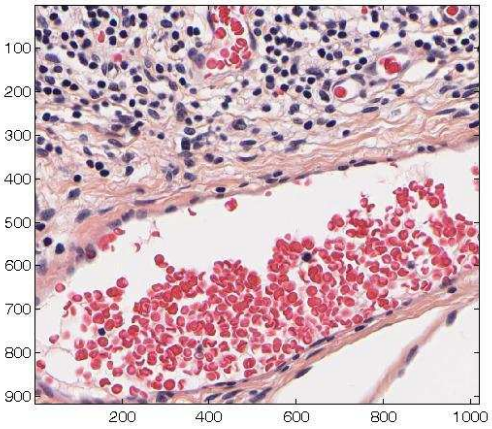

Fig. 2. Part of a pathological slide with visible cell nuclei.

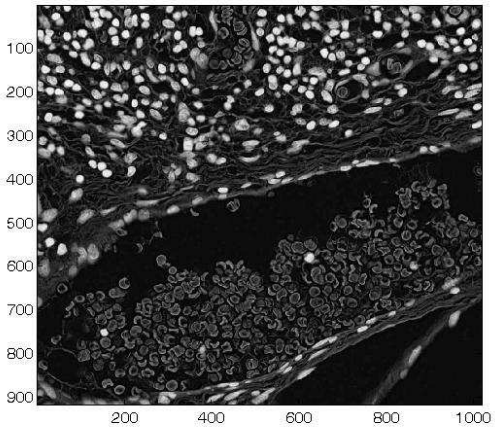

Fig.3. Complement of the R plane.

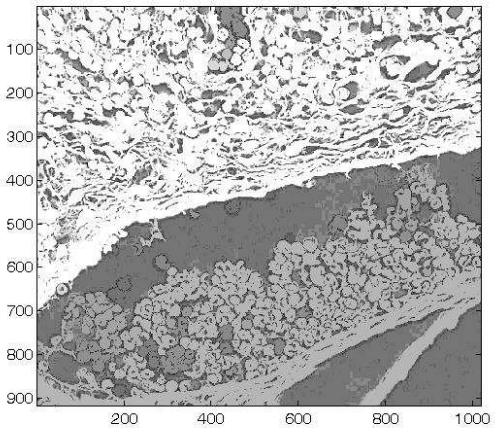

Fig 4. Output of the reconstruction filter.

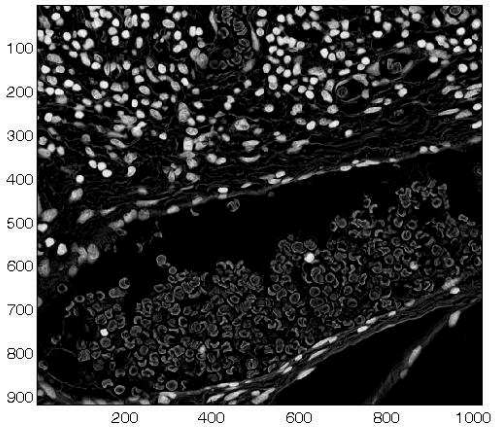

Fig.5. Output of the top-hat transform by reconstruction operation.

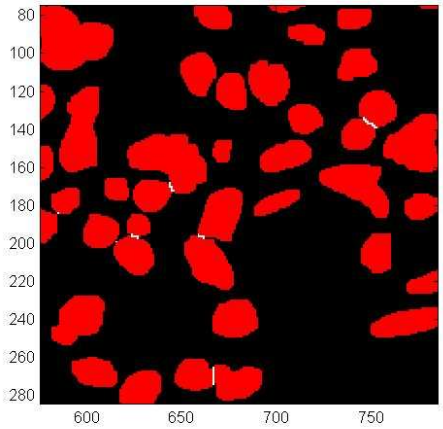

Fig.6. Output of the watershed segmentation. The final segmentation of the cells is shown in red against the black background. The regions where the watershed segmentation separates are shown in white.

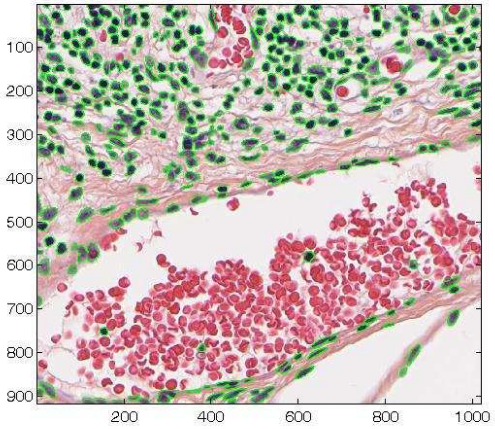

Fig.7. Output of the final segmentation of the region of interest shown in Fig. 2. The contours of the cell nuclei are in green.

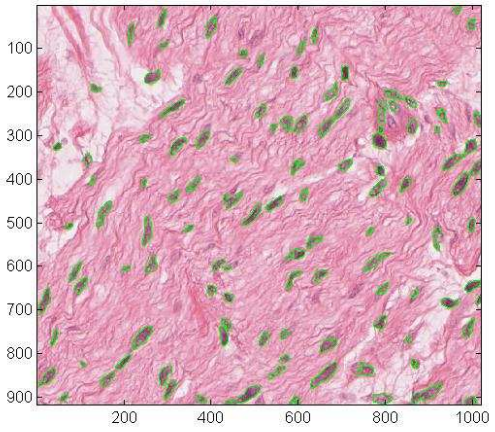

Fig.8. Output of segmentation of cells from a stromal region. The output of the cell nuclei are in green.

In order to evaluate the accuracy of the segmentation algorithm, 20 different slides of size 920x1020 from neuroblastoma slides were sampled, and the segmentation algorithm on these slides were applied. Then, nuclei were manually segmented for a randomly selected 100 cells. The manually segmented and automatically segmented nuclei boundaries were compared using two different overlap scores.

For the 100 sample nuclei, the average overlap score OS1 was 90.24% with a standard deviation of 5.14%, and OS2 was 94.79% with a standard deviation of 2.97%.

IV. Conclusion

In this work, cell nuclei detection and segmentation algorithm for neuroblastoma studies is developed. The method uses morphological top-hat by reconstruction method as well as hysteresis thresholding operation. The developed algorithm was tested using 100 randomly selected cell nuclei by comparing the automated output to the manual segmentations. The developed method uses morphological operations to minimize the variation due to different staining as well as variation in the tissue characteristics. It was observed from the visual evaluation of the final segmentation results that some of the cell nuclei that are only faintly visible within the selected slide are not detected, therefore, not segmented. It was also observed that watershed segmentation could not separate a small number of clustered cells.

In future work, observed issues will be addressed. Additionally, color and background correction will be included. Information from several color planes will be used to increase accuracy.

References

[1] L. Rees, M. Smith, J. Gurney, M. Linet, T. Tamra, J. Young, and J. Bunin, "Cancer Incidence and Survival among Children and Adolescents: United States SEER Program 1975-1995," vol. NIH Pub. No. 99-4649, 1999.

[2] H. Shimada, I. M. Ambros, L. P. Dehner, J. Hata, V. V. Joshi, B. Roald, D. O. Stram, R. B. Gerbing, J. N. Lukens, K. K. Matthay, and R. P. Castleberry, "The International Neuroblastoma Pathology Classification (the Shimada system)," *Cancer*, vol. 86, pp. 364-72, 1999.

[3] H. Shimada, I. M. Ambros, L. P. Dehner, J. Hata, V. V. Joshi, and B. Raold, "Terminology and morphologic criteria of neuroblastic tumors: Recommendation by the International Neuroblastoma Pathology Committee," *Cancer*, vol. 86, pp. 349-63, 1999.

[4] M. Gao, P. Bridgman, and S. Kumar, "Computer aided prostate cancer diagnosis using image enhancement and JPEG2000," *SPIE Medical Imaging Conference*, 2003.

[5] A. Tabesh, V. P. Kumar, H.-Y. Pang, D. Verbel, A. Kotsianti, M. Teverovskiy, and O. Saidi, "Automated prostate cancer diagnosis and Gleason grading of tissue microarrays," *SPIE Medical Imaging Conference*, 2005.

[6] F. Raimondo, M. A. Gavrielides, G. Karayannopoulou, K. Lyroudia, I. Pitas, and I. Kostopoulos, "Automated evaluation of Her-2/neu status in breast tissue from fluorescent in situ hybridization images," *IEEE Trans Image Process*, vol. 14, pp. 1288-99, 2005.

[7] P. L. Rosin and T. Ellis, "Image difference threshold strategies and shadow detection," *6th British Machine Vision Conference*, pp. 347-356, 1995.

[8] N. Malpica, C. O. de Solorzano, J. J. Vaquero, A. Santos, I. Vallcorba, J. M. Garcia-Sagredo, and F. del Pozo, "Applying watershed algorithms to the segmentation of clustered nuclei," *Cytometry*, vol. 28, pp. 289-97, 1997.