Published in final edited form as:

Methods Inf Med. 2012 ; 51(3): . doi:10.3414/ME11-02-0015.

A Fast, Automatic Segmentation Algorithm for Locating and Delineating Touching Cell Boundaries in Imaged Histopathology

Xin Qi^{1,2}, Fuyong Xing³, David J. Foran^{1,2}, and Lin Yang^{2,3}

¹Dept. of Pathology and Laboratory Medicine, UMDNJ-Robert Wood Johnson Medical School

²Center for Biomedical Imaging & Informatics, The Cancer Institute of New Jersey, UMDNJ-Robert Wood Johnson Medical School

³Division of Biomedical Informatics, Department of Biostatistics, University of Kentucky, Lexington, KY, 40506

Summary

Background—Automated analysis of imaged histopathology specimens could potentially provide support for improved reliability in detection and classification in a range of investigative and clinical cancer applications. Automated segmentation of cells in the digitized tissue microarray (TMA) is often the prerequisite for quantitative analysis. However overlapping cells usually bring significant challenges for traditional segmentation algorithms.

Objectives—In this paper, we propose a novel, automatic algorithm to separate overlapping cells in stained histology specimens acquired using bright-field RGB imaging.

Methods—It starts by systematically identifying salient regions of interest throughout the image based upon their underlying visual content. The segmentation algorithm subsequently performs a quick, voting based seed detection. Finally, the contour of each cell is obtained using a repulsive level set deformable model using the seeds generated in the previous step. We compared the experimental results with the most current literature, and the pixel wise accuracy between human experts' annotation and those generated using the automatic segmentation algorithm.

Results—The method is tested with 100 image patches which contain more than 1000 overlapping cells. The overall precision and recall of the developed algorithm is 90% and 78%, respectively. We also implement the algorithm on GPU. The parallel implementation is 22 times faster than its C/C++ sequential implementation.

Conclusion—The proposed overlapping cell segmentation algorithm can accurately detect the center of each overlapping cell and effectively separate each of the overlapping cells. GPU is proven to be an efficient parallel platform for overlapping cell segmentation.

Keywords

Parallel Computing; Segmentation; Seed detection; Pathology images

1. Introduction

Breast cancer is a malignant neoplasm that can affect both women and men. It is the leading cancer in both white and African American women. It is the second most common cause of cancer death in white, black, Asian/Pacific Islander and American Indian/Alaska Native

women. It is estimated that approximately 209,060 new cases of invasive breast cancer were reported among women in the US during 2010. There is over a 98% five-year relative survival rate when localized breast cancer is caught before it spreads to other parts of the body(1).

In spite of the increase in the incidence of the disease, the death rates of breast cancer continue to decline. This decrease is believed to be the result of earlier detection through screening and analysis as well as improved treatment. Some of the common methods screening of breast cancer include examination by a physician, self-performed routine examinations and routine mammograms. When suspicious lesions are detected by these methods, a fine needle aspirate or biopsy can be performed and the obtained tissue is examined. The extracted tissue is mounted on glass slides and examined by a surgical pathologist to make decision of benign or cancer. Diaminobenzidine (DAB) and hematoxylin are standard staining methods used to breast cancer specimen.

Tissue Microarray (TMA) technology, e.g.(2), provides a platform to perform comparative cancer studies involving evaluation of protein expression. TMAs provide an efficient method for preserving tissue while facilitating high-throughput of multiple tissue samples in parallel. Digital microscopy has become an extremely valuable tool for visualizing and archiving pathology specimens(3).

Accurate segmentation of the cells in digitized histopathology specimens is usually the first step in automatic image analysis. An unsupervised clustering which utilizes both color and texture features to segment hematoxylin and eosin (H&E) stained prostate cancer was recently proposed in(4). A computationally efficient approach that uses color and differential invariants (how to use differential invariants? So this paper is working with Lie Group? You can double check) to assign class posterior probabilities to separate epithelial nuclei, stroma and background regions in breast microarray was reported in(5). Tissue segmentation base on color or multispectral imaging through the combination of spatial clustering and multiphase vector level set active contours was proposed for delineating histological prostate cancer tissues in(6).

These segmentation methods produce good results on regions exhibiting little or no cell overlap. However, they usually fail to accurately segment overlapping cells. The watershed is one of the most commonly used segmentation methods to handle overlapping objects. However, it often suffers from over segmentation. Some algorithms, such as a rule-based approach(7), were proposed to address this problem by merging over-segmented regions. Voronoi diagram had been applied to handle the overlapping regions in(8). However it is difficult to derive a generalized rule to merge the over-segmented regions across different image sets. Wen et. al.(9) reported their study on decomposing clumps of nuclei using high-level geometric constraints derived from maximum curvatures. This approach proved to be very effective in separating overlapping nuclei, however for some overlapping objects, the common connecting region may not exhibit local maximal curvatures. Kothari et. al.(10) proposed a semi-automatic method for overlapping cell segmentation, in which they applied concavity detection at the edge of a cluster to find the points of overlap between two nuclei. An ellipse-fitting technique was then applied to segment the concavities between two nuclei of overlap region. However, the ellipse can't accommodate the shape of some cells, especially for cancerous cells. Yang et. al.(11) proposed an approach to address overlapping cell segmentation using concave vertex graph. Graph based method was also proposed in (12) to segment overlapping stem cells within fluorescence microscopy images. Another overlapping object segmentation using semantic concept is presented in (13). These methods usually require that the images exhibit a high contrast at the edge, which is often not the case especially in the cancer regions.

2. Automatic Region of Interest (ROI) Detection

In our algorithm, the first step is to find the region of interest (ROI). Recently we developed a computer aided navigation algorithm for traversing large whole-slide virtual microscope datasets.

The algorithm employs a novel feature measurement called square annual histogram (SAH), which is both rotation and scale invariant, and is able to capture the spatial configuration of image features composing a given image region or patch. Using square annual histogram, it can discriminate among image patches which may exhibit similar intensity distributions, but differs in terms of spatial distributions. One example is shown in Figure 1 where two images show similar intensity distributions in RGB channels, but have completely different spatial configurations of the pixels, and this difference is clearly exploited by the newly proposed square annual histogram. The histograms of central bin of those two images are different while the histograms of the whole images are similar.

The χ^2 distance is utilized as similarity measurement. In our algorithm, we considered image candidates with 90% overlap among patches, and mean-shift clustering was performed to keep only 10% ranked patches.

3. Overlapping Cell Segmentation

Given the detected region of interest, cells within these patches are further processed using the touching cell segmentation algorithm. In this paper, we propose a novel algorithm to separate overlapping cells within breast TMA using standard RGB images. Hematoxylin stained breast TMA specimens were acquired using a 40× objective lens to show the cellular structure within the breast tissue. The overlapping cell segmentation algorithm for microscopic breast TMA image was composed of two main modules. The first module locates the center of each cell using a novel mean shift and voting based seed detection method, to determine the initial position for the second module. The second module is the overlapping cell segmentation component, which extracts each overlapping cell's contour using a level set function with a repulsion force to penalize object overlap. In order to speed up the whole segmentation procedure, the algorithm has been parallelized using a graphic processing unit (GPU).

3.1 Seed detection

Since the number and the location of cells are not known *a priori*, it is difficult to segment cells from microscopic images with traditional methods especially when the cells overlap with one another. The center of cell is considered as a basic perceptual cue that supports the accurate separation of overlapping cells. Parvin et. al.(14) proposed an iterative voting method to detect the center of objects, which provides excellent results in detecting the centers in the overlapping cells. In our algorithm, we define $I(x,y)$ as the original image, the image gradient $\nabla I(x,y)$ and its magnitude $\|\nabla I(x,y)\|$ are subsequently calculated. For each pixel (x, y) the voting direction $\alpha(x,y)$ is defined as its negative gradient direction

$$-\frac{\nabla I(x,y)}{\|\nabla I(x,y)\|} = (\cos(\theta(x,y)), -\sin(\theta(x,y)))$$

where θ is the angel of the gradient direction with respect to x axis. The voting area $A(x,y;r_{min}, r_{max}, \Delta)$ of each pixel is defined by a cone shape with its vertex located at (x,y) . x, y). A cone-shape voting area is selected for two reasons. First the center of the cell is relatively far away from the boundary of the cell, thus more voting points are located within

the region closest to the center when compared to the region closest to the edge of the cell. Second, a cone shape voting area will greatly reduce the time complexity of the algorithm for calculating fewer voting points in total. We define a 2D Gaussian kernel $g(x, y)$ with its mean μ located at the center of the voting area. The center of the Gaussian kernel

$$g(x, y, \mu_x, \mu_y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x-\mu_x)^2 + (y-\mu_y)^2}{2\sigma^2}\right)$$

where

$$\mu_x = x + \frac{(r_{max} + r_{min}) \cos\theta}{2} \text{ and } \mu_y = y - \frac{(r_{max} + r_{min}) \sin\theta}{2}$$

The kernel is specifically defined in a way to reinforce the voting towards the center of the object. A voting image $V(x, y; r_{min}, r_{max}, \Delta)$ is defined, which has the same dimensions as the original image $I(x, y)$. For each pixel (x, y) , we update the voting image as

$$V(x, y; r_{min}, r_{max}, \Delta) = V(x, y; r_{min}, r_{max}, \Delta) + \Sigma_{(u,v)} \|\nabla I(x, y)\| g(x, y; \sigma, A)$$

The centers of objects are determined by running mean shift on the sum of the voting image after thresholding. The detailed algorithm is listed in Algorithm 1.

Although our work was initially motivated by (14), the method we have developed is different: For each point (x, y) with high gradient, we define a convolution Gaussian kernel located at the center of the voting area instead of (x, y) . This is a critical step which explains why our algorithm performs better. As the center of the object is usually relatively far away from the boundary, the shifted Gaussian kernel we designed can reinforce the voting towards the center of the object. Meanwhile, instead of using iterative voting, we proposed to calculate the center of overlapping objects by running mean shift on the final set of voting images. This step dramatically reduces the computational time as we avoid many iteration steps.

Algorithm 1: Mean shift based seed voting

1. Initialize the parameters: d = estimated average diameter of cells within the image, $r_{min} = 0.5d$, $r_{max} = 1.5d$, $\Delta = 30$.
 2. Calculate the Gaussian blurred gradient image and the orientation of the gradient at each pixel (x, y) . Record the set of (x, y) with large gradient magnitude as S .
 3. For each point $(x, y) \in S$, calculate the voting image $V(x, y)$.
 4. for $R = 0.3, 0.4, \dots, 0.9$ do
 5. Record all the points (x, y) in the voting image with voting number larger than $\max(V(x, y)) \times R$.
 6. end for
 7. Sum all the voting image and run mean shift to generate the final list of the seeds. The bandwidth of the mean shift is set as $1/3$ of the estimated average diameter of cells within the image.
-

3.2 Parallelization of the Seed Detection on the Graphic Processing Unit (GPU)

In our seed detection module, the voting algorithm is based on the voting pixels. Each voting pixel $((x, y) \in S)$ has a cone-shape voting area $A(x, y; r_{min}, r_{max}, \Delta)$. Each pixel in the final voting image V is updated by equation 2.

By investigating the execution time for each step of our proposed seed detection algorithm for analyzing a typical 2D pathology image (1392×1040), we found that the computational bottleneck is the voting step. Our proposed seed voting algorithm is pixel based method which has an advantage of easy parallelization due to data independence. By fully utilizing this property, we chose eight blocks for a GPU. Within each block, we create 128 threads. Each voting pixel was assigned to one thread to calculate its corresponding voting image. In total 1024 threads were created simultaneously.

3.3 Cell Segmentation

Once the seeds have been accurately detected, the overlapping cell segmentation is performed using level set with an interaction model. The interaction model includes two mechanisms: a repulsion term prevents the contours of adjacent cells from overlapping and separates the overlapping cell boundaries; a competition term which determines the membership of each pixel which is assigned to the cells producing the smallest difference. Considering an image I that has N cells, let $C_i (i = 1, \dots, N)$ denote the contours that evolve towards the boundaries. Each cell is represented by its own level set energy function. Instead of examining each contour independently, the interaction between neighboring contours was integrated into the level set energy function. The energy function E for cell segmentation which combines the repulsion and competition terms is:

$$\begin{aligned} E = & \lambda_0 \sum_{i=1}^N \int_{in(C_i)} |I - c_i|^2 dx dy \\ & + \lambda_b \sum_{i=1}^N \int_{\Omega_b} |I - c_b|^2 dx dy \\ & + \mu \sum_{i=1}^N \int_0^1 g(|\nabla I(C_i(q))|) |C'_i(q)| dq \\ & + \omega \sum_{i=1}^N \sum_{j=1, j \neq i}^N A_i \cap A_j \end{aligned}$$

where A_i denotes region of cell $\{A_i | i = 1, 2, \dots, N\}$ and Ω_b is the background which represents the region outside all the cells $out(C_1) \cap out(C_2) \cap \dots \cap out(C_N)$. The operator $in()$ and $out()$ represent the regions inside and outside of the cell, respectively. The c_i and c_b are the intensity means of the cell regions and background regions, respectively. The λ_0 , λ_b and μ are the fixed weighting parameters. Function g is chosen to be a sigmoid function

$g(x) = \left(1 + e^{\left(\frac{x-b}{\alpha}\right)}\right)^{-1}$ where α is used to control the slope the output curve and β controls the window size. By penalizing the union of the overlapped region $\{A_i | i = 1, \dots, N\}$ enclosed by contours $C_i (i = 1, \dots, N)$, the last item in Energy function E is the repulsion term which is used to represent the repulsion force between each adjacent overlapping object and the ω is the regulation parameter.

Segmentation is achieved by minimizing the energy function E . In order to express the energy function using level set, we bring in the regularized Heaviside function H_ε (15)

$$H_\varepsilon(z) = \frac{1}{2} \left(1 + \frac{2}{\pi} \tan^{-1} \left(\frac{z}{\varepsilon} \right) \right)$$

Where ε is the regulation parameter of the Heaviside function and Delta function is defined as

$$\delta_\varepsilon(z) = \frac{d}{dz} H_\varepsilon(z)$$

The energy function can be minimized iteratively by employing the gradient descent method. The evolution equation for each energy function ($\psi_i(t, x, y)$) is then obtained by deducing the associated Euler-Lagrange equation as

$$\frac{\partial \psi_i}{\partial t} = \delta(\psi_i) \left(\lambda_0 |I - c_i|^2 - \lambda_b |I - c_b|^2 \prod_{j=1, j \neq i}^N H(\psi_j) + \mu \nabla g \frac{\nabla \psi_i}{|\nabla \psi_i|} + \gamma g \operatorname{div} \left(\frac{\nabla \psi_i}{|\nabla \psi_i|} \right) + \omega \sum_{j=1, j \neq i}^M (1 - H(\psi_j)) \right)$$

After evolving the level sets, the means c_i and c_b of the cell and the background regions are iteratively updated. This method has proven to be quite effective and accurate for RNAi fluorescent cellular image segmentation in (16). Throughout the experiments, the parameters were selected as $\lambda_0 = 1$, $\lambda_b = 0.3$, $\mu = 0.5$, $\gamma = 0.2$, $\omega = 0.6$, $\epsilon = 1$, $\alpha = 1$, $\beta = 7$ empirically.

4. Experimental Results

Hematoxylin stained breast TMA specimens were used throughout these experiments. The images were captured using 10× and 40× objective lens. Images taken under 10× objective were utilized for retrieval ROI patches. Then the detected patches with its corresponding 40× magnification images were utilized for cell segmentation. In total we utilized 100 image patches containing more than 1000 imaged cells. Figure 2 shows an example of detected ROI on a breast TMA image. The red square on the left image is the input patch; the red crosses on right top images are the centers of highly related image patches similar to query region; the green squares on right bottom images are the final retrieval result after mean-shift clustering. Figure 3 shows the seed detection results on a synthetic image with five overlapping cells. From this experiment, it can be seen that the iterative voting method tends to put the seed at the overlaid region when several cells are overlapping together. This is because the voting number calculated in (14) is the convolution results of a Gaussian centered at the boundary point (x, y) , which is the location with the higher gradient. However, the center of the object should be far away from the boundary. In Figure 4 we show the results using real data. The seed detection errors using the method in (14) are marked with yellow rectangles. In our algorithm, we utilize a Gaussian kernel defined at the center of the voting area for each voting point (x, y) calculated by thresholding the gradient image. In this way we penalize the voting towards the boundary of the object and enforce the voting towards the center of the object. This is the major reason that our method can provide better results for cells with overlapping regions.

To measure the accuracy of seed detection, an error function is defined as the pixel-wise distance between seeds manually located and seeds extracted using seed detection methods. Table I shows the quantitative results of our algorithm and (14) compared with the ground truth annotation. The 80% column in Table I represents the sorted 80% accuracy among all the seed detection results. Meanwhile a number of missing seeds and false seeds are calculated. When compared with the ground-truth annotation, the mean of the number of missing seeds and false seeds using our algorithm were 0.2 and 0.8 respectively. The mean of the number of missing and false seeds using (14) were 1 and 1.2 respectively.

In Figure 5, we showed four representative image segmentation results using of our method and compared to five other algorithms including marker-based watershed, mean-shift, isoperimetric (17), and methods presented in (14) and (18), respectively. The method we developed achieved more accurate seed detection results, which is prerequisite major contributing factor towards accurate cellular segmentation. Using the seeds detected by our method, marker-based watershed still tends to over-segment the whole cell in the image as it can't compensate for the intensity variation within the cells. Mean-shift, Isoperimetric (17), and the method reported in (18) segmentation tends to under-segment the image by merging

multiple cells into one object. The method presented in (14) tends to create false seeds in the overlapping region of cells, which leads to unreliable segmentation of cells in the overlapping regions. All of the source codes and binaries of the state-of-the-art methods were downloaded from the respective algorithm authors' websites.

In order to quantitatively measure the accuracy, precision (P) and recall (R) are calculated for our overlapping cell segmentation algorithm and compared with the ground truth annotations. P is defined as the intersection between the segmentation results and the manually annotation results divided by the segmentation results. R is defined as the intersection between the segmentation results and manually annotation results divided by the manually annotation results. In Table II, we show the detailed statistics of the quantitative segmentation results compared with the ground-truth annotation. The 80% column in Table II represents the sorted 80% accuracy among all the 234 image patches, which contains more than 2200 cells. From Table II, we achieved better segmentation result compared with other methods because of less false seeds. The mean and standard deviation of P and R for our overlapping cell segmentation algorithm is 0.90 ± 0.01 and 0.78 ± 0.03 , respectively, which indicates a good agreement between the manual and ground truth annotations. Varying the magnification of the input images will affect the initialization parameters of seed detection because our algorithm is not scale invariant. However to compensate for changes in magnification, only the average diameter of objects and bandwidth of mean shift play a significant effect for seed detection. Once they are estimated, the seed detection algorithm can achieve relatively robust results.

The parallelization code was developed using Graphic Processing Unit (GPU). GPU is a massively parallel multi-core chip that can execute thousands of concurrent threads. The GPU cores (also called stream processors) are grouped into several streaming multiprocessors. They are managed by the thread manager.

Our GPU is the NVIDIA Quadro FX5800 which has 240 cores and 30 streaming multiprocessors, each of which contains 8 GPU cores. It supports both single and double float point precision, offers 933 GFlops single precision, and has memory bandwidth with 102 GB/second. Unlike CPU threads that are heavy weight, GPU threads are light weight with little creation overhead, instant switching, and instruction and memory latency hiding. The Compute Unified Device Architecture (CUDA) environment from NVIDIA Corporate is utilized in our parallel implementations.

The experimental results have shown significant speed up (22 times faster than the sequential implementation on CPU and thousands of times faster than the Matlab implementation). In the parallel version of the algorithm, we can complete the seed detection procedure for an image with dimensionality 1392×1040 in 197 milliseconds.

5. Conclusion

In this paper we presented a fast and accurate overlapping cell segmentation algorithm and rested in large clinical datasets. We demonstrated that square annual histogram (SAH) followed by mean-shift clustering can accurately identify the most salient image patches. In the following step, mean shift based voting algorithm can accurately detect the center of each overlapping cell. Given these estimated seeds, a level set active contour based on the interactive model can effectively separate each of the overlapping cells. The improved segmentation results are achieved by accurately estimating the seed locations and utilizing a repulsion term in the level set energy function to separate the overlapping boundaries. The method is automatic and is not limited to domain specific prior knowledge. Therefore it holds promise for other multichannel overlapping object segmentation applications with

minor effort. The GPU implementation of our algorithm can reliably process one 2D image in less than 0.2 second. The experimental results show that the GPU is an extremely efficient parallel platform for overlapping cell segmentation.

Acknowledgments

This research was funded, in part, by grants from the National Institutes of Health through contract 5R01CA156386-06 from the National Cancer Institute; and contracts 5R01LM009239-04 and 1R01LM011119-01 from the National Library of Medicine. Additional support was provided by a gift from the IBM International Foundation.

References

1. Jemal A, Siegel R, Xu J, Ward E. Cancer Statistics, 2010. *CA Cancer J Clin*. 2010
2. Rimm DL, Camp RL, Charette LA, Costa J, Olsen DA, Reiss M. Tissue microarray: a new technology for amplification of tissue resources. *Cancer Journal*. 2001; 7(1):24–31.
3. Bartels PH, Gahm T, Thompson D. Automated microscopy in diagnostic histopathology: from image processing to automated reasoning. *International Journal of Image Systems and Technology*. 1997; 8:214–223.
4. Datar M, Padfield D, Cline H. Color and texture based segmentation of molecular pathology images using hsoms. *Proceedings of IEEE International Symposium on Biomedical Imaging*. 2008:292–295.
5. Amaral T, McKenna S, Robertson K, Thompson A. Classification of breast-tissue microarray spots using colour and local invariants. *Proceedings of IEEE International Symposium on Biomedical Imaging*. 2008:999–1002.
6. Hafiane A, Bunyak F, Palaniappan K. Evaluation of level set-based histology image segmentation using geometric region criteria. *Proceedings of IEEE International Symposium on Biomedical Imaging*. 2009; MP-PA1:1–4.
7. Wahlby C, Lindblad J, Vondrus M, Bengtsson E, Björkstén L. Algorithms for cytoplasm segmentation of fluorescence labelled cells. *Anal Cell Pathol*. 2002; 24:101–111. [PubMed: 12446959]
8. Zhou X, Liu KY, Bradblad P, Perrimon N, Wang STC. Towards automated cellular image segmentation for RNAi genome-wide screening. *Proceeding of Medical Image Computing and Computer Assisted Intervention*. 2005; 3749:885–892.
9. Wen Q, Chang H, Parvin B. A delaunay triangulation approach for segmenting clumps on nuclei. *Proceedings of IEEE International Symposium on Biomedical Imaging*. 2009:9–12.
10. Kothari S, Chaudry Q, Wang WD. Automated cell counting and cluster segmentation using convexity detection and ellipse fitting techniques. *Proceedings of IEEE International Symposium on Biomedical Imaging*. 2009; 2:795–798.
11. Yang L, Tuzel O, Meer P, Foran DJ. Automatic image analysis of histopathology specimens using concave vertex graph. *International Conference on Medical Image Computing and Computer Assisted Intervention*. 2008; 5241:833–841.
12. Faustino GM, Gattass M, Rehen S, Lucena CJP. Automatic embryonic stem cells detection and counting method in fluorescence microscopy images. *Proceedings of IEEE International Symposium on Biomedical Imaging*. 2009:799–802.
13. Wittenberg T, Grobe M, Münzenmayer C, Kuziela H, Spinnler K. A semantic approach to segmentation of overlapping objects. *Methods of Information in Medicine*. 2004; 43(4):343–353. [PubMed: 15472745]
14. Parvin B, Yang Q, Han J, Chang H, Rydberg B, Barcellos-Hoff MH. Iterative voting for inference of structural saliency and characterization of subcellular events. *IEEE Transactions on Image Processing*. 2007; 16(3):615–623. [PubMed: 17357723]
15. Chan TF, Vese LA. Active contours without edges. *IEEE Transaction of Image Processing*. 2001; 10(2):266–277.

16. Yan P, Zhou X, Shah M, Wong STC. Automatic segmentation of high-throughput RNAi fluorescent cellular images. *IEEE Transactions on Information Technology in Biomedicine*. 2008; 12(1):109–117. [PubMed: 18270043]
17. Grady L, Schwartz EL. Isoperimetric graph partitioning for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2006; 28(1):469–75. [PubMed: 16526432]
18. Al-Kofahi Y, Lassoued W, Lee W, Roysam B. Improved automatic detection and segmentation of cell nuclei in histopathology images. *IEEE Transactions on Biomedical Engineering*. 2010; 57(4): 841–52. [PubMed: 19884070]

Figure 1.

Two patches exhibiting similar intensity distribution in RGB, but have completely different spatial configurations. In this figure we show SAH exhibits more discriminative power than traditional histogram by encoding the spatial configuration of the image patches.

Figure 2.

An example of the detected ROI on a breast TMA image. Red square on the representative patch is the input; the red crosses on the right top image are the centers of highly related image patches similar to query region; the green squares on right bottom image are the final retrieval result after mean-shift clustering.

Figure 3.

(a) an original synthetic image. (b) seed detection results using the iterative voting method in(14) (c) intermediate results using our mean-shift based seed detection (d) final detected seeds.

Figure 4. Seed detection results using real dataset where the errors on the overlapping part are marked with yellow rectangles. (a) original Hematoxylin stained TMA image. (b) seed detection results using the iterative voting method in (14) (c) seed detection results using our algorithm.

Figure 5.

The comparative results on four representative image patches. (a) the original image patches (b) the marker-based watershed segmentation results using our seeds (c) the mean-shift results (d) the isoperimetric segmentation results (e) the segmentation results using method presented in (14) (f) the segmentation results using the method described in (18) (g) our segmentation results

Table I

The quantification of the pixelwise seed detection errors of our algorithm and (14), both compared with the ground truth annotation of the seeds

	Pixelwise Detection Error (our method)	Pixelwise Detection Error ((14))
Mean	6.63	7.46
Variance	4.53	7.29
Median	6.81	7.75
Min	3.05	3.32
Max	10.31	12.55
80%	5.08	5.23

Table II

The segmentation accuracy compared with ground truth annotation, the watershed represents the marker-based watershed.

Watershed	Precision	Recall
Mean	0.73	0.52
Variance	0.03	0.06
Median	0.99	0.54
Min	0.07	0.03
Max	1.00	0.93
80%	1.00	0.78
Mean-shift		
Mean	0.73	0.78
Variance	0.07	0.02
Median	0.85	0.81
Min	0.17	0.26
Max	1.00	0.96
80%	0.99	0.87
Isoperimetric		
Mean	0.87	0.76
Variance	0.04	0.02
Median	0.95	0.80
Min	0.23	0.23
Max	1.00	0.95
80%	1.00	0.87
(14)		
Mean	0.84	0.64
Variance	0.04	0.02
Median	0.93	0.65
Min	0.29	0.28
Max	1.00	0.90
80%	1.00	0.75
(18)		
Mean	0.82	0.81
Variance	0.04	0.02
Median	0.90	0.87
Min	0.18	0.20
Max	1.00	0.99
80%	0.99	0.93
Our Method		
Mean	0.90	0.78

Watershed	Precision	Recall
Variance	0.02	0.01
Median	0.94	0.08
Min	0.22	0.20
Max	1.00	0.96
80%	1.00	0.85