Segmenting multiple overlapping Nuclei in H&E Stained Breast Cancer Histopathology Images based on an improved watershed

Pengfei shen[1,2],Wenjian Qin[1],Jie Yang[2],Wanming Hu[3],Shifu Chen[1],Ling Li[1],TieXiang Wen[1],Jia Gu[1]

[1]*Shenzhen Institutes of Advanced Techology,Chinese Academy of Sciences,Guangdong,china*
[2] *Key Laboratory of Fiber Optic Sensing Techology and Information Processing,Ministry of Education,Wuhan University of Techology*
[3]*Department of pathology,Sun Yat-sen University Cancer Center-Guangdong,China*

Abstract

In histopathology images, there often exists several Nuclei overlapped with each other which causes difficulty to automatic nuclei segmentation. As we all know, watershed algorithm has been widely employed in image segmentation. But the limitation of watershed segmentation is sensitive to noise and can lead to serious over-segmentation. In this paper, we present an improved watershed transformation that incorporates opening-closing reconstruction and the distance transform with chamfer algorithm after color deconvolution, and H-minima . Unlike the classical watershed segmentation algorithm our improved method is able to resolve over-segmentation. The experiment results demonstrate our method successfully segment out each nuclei on breast cancer histology images, effectively address over-segmentation existed in traditional watershed segmentation , and preserve the original edges of each nuclei in the image completely.

1 Introduction

Segmentation nuclei from histopathology image is an essential image processing task for many scientific and clinical applications due to the fundamentally important role of nuclei in cellular processes and diseases. Watershed is widely utilized for the segmentation of nuclei image in recent years based on mathematical morphology[1],on account of the fact that it has draw great attention for its fast computing and high accuracy in locating the weak edges of adjacent regions. But classical watershed segmentation can lead to serious over-segmentation. Many researchers have proposed various methods continually for improving, such as a pre-processing step of calculating distance transform for the binary image before watershed transformation[2], or agglomerating regions after segmentation following watershed transformation[3]. Nevertheless, owing to the influence of person or device in practical process of image making, the useful information of adjacent districts is not obviously distributed. In the rapidly emerging field of digital, the ability to deal with nuclei overlap is highly critical in different diagnostic and prognostic application. Currently, the diagnosis of diseases such as breast cancer is done manually by visual analysis of tissue samples. There appears many researcher presenting many other approaches to handle the overlap such as a Hybrid Active Contour Model[4], or C-V level set method[5], or spatial graphs built using nuclei as vertices[6].These means need user intervention, or the effect is not very good when the tissue's color is similar to the nuclei's.

In this paper we present a method for segmentation employing watershed algorithm based on morphological opening-closing reconstruction and a new distance transformation[7] after color deconvolution. The application of this method for segmentation cell image overcomes over-segmentation effectively and produces a satisfactory segmentation result apparently superior to the approach of using watershed simply based on the traditional transformation.

2 Materials and Methods

2.1 Color Deconvolution

The first step is separation of the H&E stains with the color Deconvolution technique suggest in[6],which is a special case of true spectral separation techniques which work with multispectral cameras[6].The technique uses the fact that the image information process in bright field microscopy can be modelled by the Lambert-Beer law. Given that the images are captured by the three detection channels(R,G,B) with known optical densities and the stained-specific absorption coefficients can be experimentally determined from single stain images, the concentrations of the two stains can be determined for each pixel location. These in turn can be used to obtain single stain image. Since the nuclei are stained with hematoxylin-eosin , the greyscales version of the hematoxylin single stain image is used in all subsequent processing. The Fig.4(b) shows the effect of color deconvolution.

2.2 Morphological operations

As shown in Fig.4(b) ,we can see the nuclei and the background have obvious difference, so we use the OSTU method to extract the nuclei . The effect of OSTU is shown in Fig.4(c).The new separated hematoxylin image still contains spurious structures within the nuclei. These present obstacles

for the segmentation and can be filtered out with a series of operations based on morphological binary image reconstruction[5]. Opening by reconstruction removes unconnected bright objects that are smaller than the structuring element(SE). Similarly, closing by reconstruction removes unconnected dark objects smaller than the SE. Applying these two operators in sequence. After application of these two operations the main contours of the nuclei often have an irregular shape and protrusions emanating from the edges hampering the segmentation result. To address this problem, additional morphological closing with a small SE is applied. This simplifies the shape of the object, eliminates small protrusions, disconnects "loosely" connected objects and does not significantly affect the location of the main contours. The SE for this operation is chosen to be a disk with half the radius of the one used for the opening and closing by reconstruction operators. An example of pre-processing morphological operations is show in Fig.4(d)

2.3 the Distance transformation based on chamfer

Fig.1(a) shows the binary image of two sticky cell. The aim of using distance transform is to transform the image position information between pixels to a different gray information[8]. Fig.1(c) shows the gray image of the distance transform.

*(a)*the original binary image   *(b)* the NOT operator image

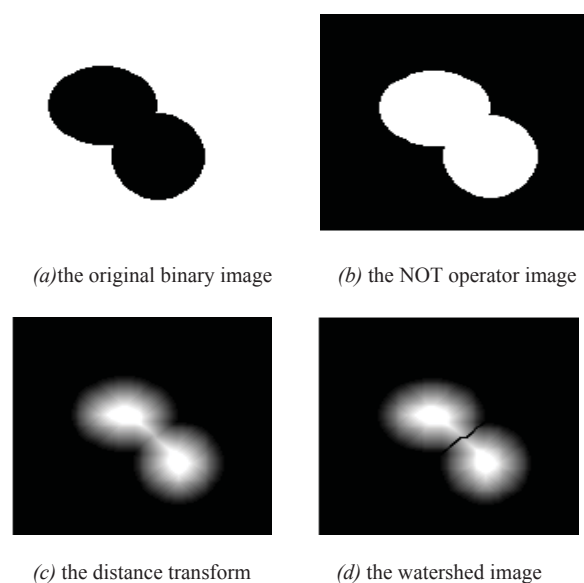*(c)* the distance transform   *(d)* the watershed image

Fig .1 the process of the distance transform

In theory, to calculate the shortest distance of a pixel to the background pixel, It is necessary to do global *operations* of the image. This will result in too much time consuming. In practice, researchers often take the approximate of the Euclidean distance, such as city block distance and chessboard distance, chamfer distance into consideration. In all these distance methods, chamfer distance algorithm is simple and fast to calculate the distance ,whose result approaches the real Euclidian geometric distance. The principle is to start from the neighbouring pixels, and every time just to calculate the minimum distance of local neighbouring pixels, according to the theory that the global distance is the superposition of local distance in proportion .

This paper utilizes the chamfer 3/4 chamfer distance transform[4], where 3 represents the distance of the two adjacent pixels connected by straight line ,and 4 represents the distance of the two adjacent pixels connected by oblique line. Chamfer 3/4 transform achieve the distance transform via a serial scan algorithm. Setting feature pixel's value is 0, the non-feature pixels with pixel values 255. The structure in Fig.2 (a) scan from left to right and from top to the end to get a minimum value, then the structure of Fig. 2 (b) scan right to left and from bottom to top to get a minimum value. As follows show the structure we use.

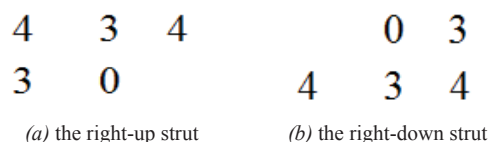*(a)* the right-up strut   *(b)* the right-down strut

Fig. 2 the strut algorithm

in the chamfer distance transform, the gray pixels whose value are 0 stay 0 after transformed, otherwise, the gray pixels whose value aren't 0 have the value that corresponding change according to the distance from the 0 pixel.Fig.4(e) shows the effect of the chamfer distance transform.

In the binary adhesion cells image, the ROI(region of interest) is the information of location of each pixel , regardless of the background. Therefore, it is wise to do the NOT operation, whose effect is shown in Fig.1(b),which makes the value of background pixels 0, the value of the pixels inside cell nonzero, which makes the gray value of background will not change, whereas the value of pixels inside cell will vary on the distance from the boundary. as shown, scanning the image using the aforementioned method, the distance transform image will appear. The information of location of a image will transform to the different gray information. On account of the fact the pixels of the edge have smaller greyscale value, they obvious distinguish from the other pixels in the image. There is no need to mark all point of the object's edge, but the point of edge where one nucleus stick to another, which is the mission of the watershed algorithm. Fig.1(d) shown the result of watershed.Fig.4(e) is the effect of distance transform.

2.4 Watershed

Watershed algorithm, which is based on mathematical morphology and is aiming at separation an image into homogeneous principle, is first proposed by S.Beucher an L.Vincent and is developed rapidly in image segmentation field in recent years[1].An image can, indeed, be observed as a relief in which we associate an altitude to the gray level of each pixel. The Watershed is defined as the peak forming the limit between two basins. It operators on the gradient image where contours are enhanced appearing as local maxima in the image. After resorting the peak identified as contours, watershed based segmentation algorithm finds out the homogenous regions as the inner basins region.

Unlike simple threshold segmentation, watershed segmentation is an adaptive iterative threshold algorithm used for segmentation. In practical application, minima of

morphological gradient are taken as markers of flooding. Due to the noise or local irregularities of greyscale of original signal, there may be many incorrect minima in gradient image which will cause over-segmentation. Although smoothening is done to the image, the number of minima is often larger than the number of objects in original image. In this situation markers should be obtained by other pre-processing approaches instead of being set equal to the minima in gradient image. This is where the opening-closing reconstruction , distance transform and H-minima transform function. The morphological operator can reduce the noise, and distance transform based on chamfer convert the information of location to different gray information, making the value of boundaries is very low and the boundaries apart from other pixel area, and H-minima suppress undesired minima. The Fig.4(f) shows the effect of classical segmentation.Fig.4(h) shows the region of over-segmentation using classical segmentation.

2.4 H-minima transform and Minima Imposition Technique

The large number of local minima which is the main cause of over-segmentation is reduced to some extend after applying morphological filters[7]. But some undesired minima are still present in the distance image. So in the proposed we use marker extraction method which utilize H-minima transform to produce a marker image for watershed transform. The H-minima transform is a powerful mathematical tool to suppress undesired minima. Performing the H-minima transform on the distance image can effectively decrease over-segmentation. The H-minima transform is performed by
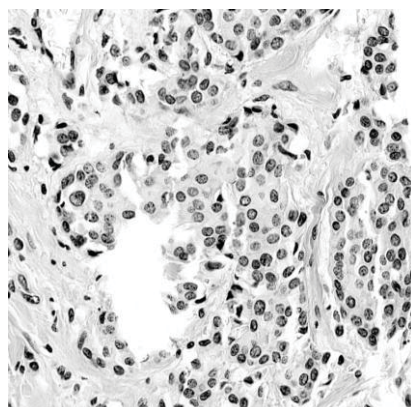
$$H_h(g) = R_g^\varepsilon(g + h) \qquad (1)$$

In Equation (1), with R the morphological greyscale reconstruction by erosion operator. This transform suppresses all minima in g whose depth is less than h.
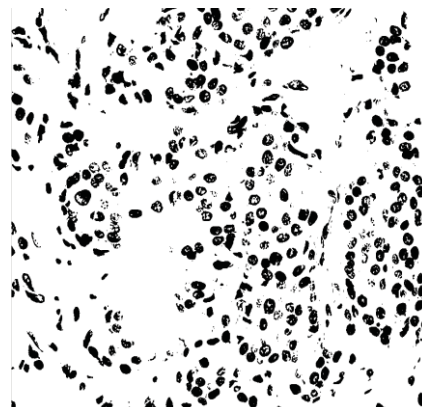
So in the proposed work we use marker extraction method which uses h-minima transform to produce a marker image for watershed transform.
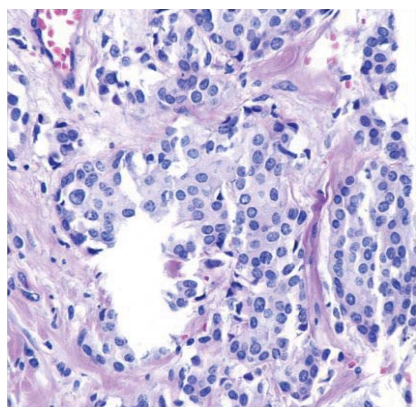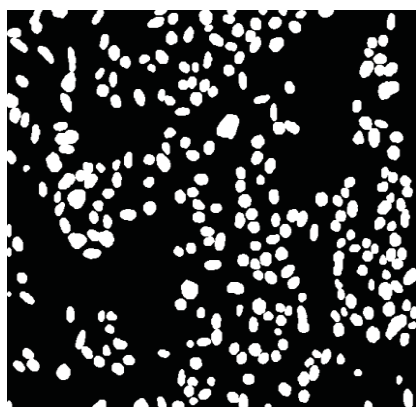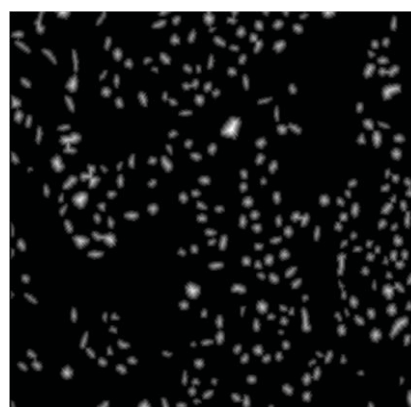

(a) the original image
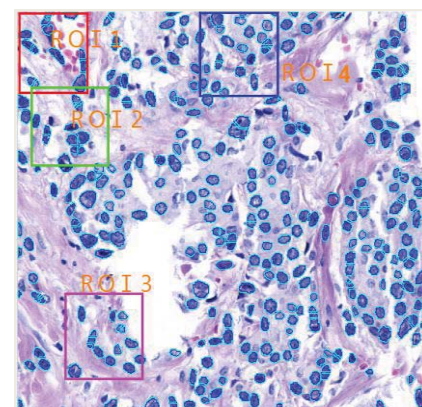

(b) the extraction image

(c) the image by OSTU
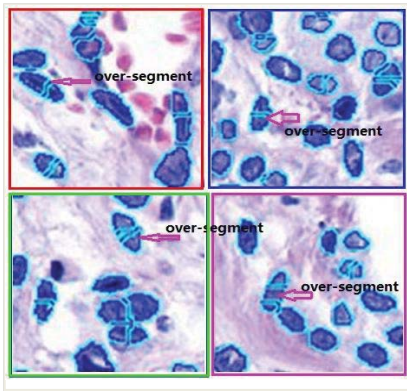

(d)the image by morphological operator
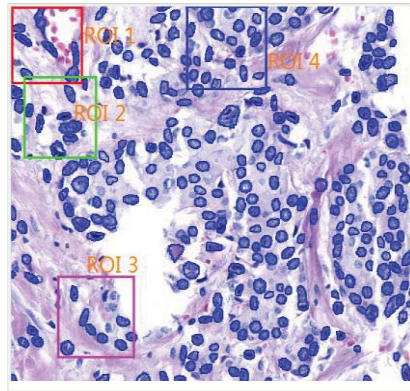

(e) fast distance transform

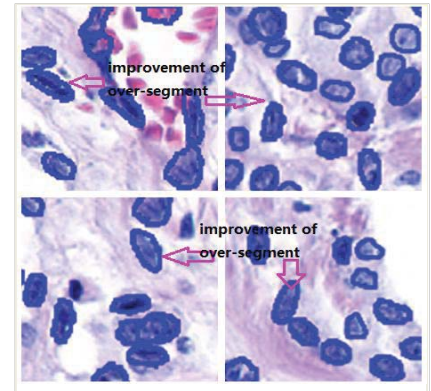(f) classical segmentation algorithm

(h) the over-segmentation region   (i) the improved watershed algorithm   (g) the improvement of over-segmentation

Fig.4 the processing and compare of the improved watershed algorithm

3   Experiment result

The use of chamfer distance transformation and H-minima transform before applying traditional watershed segmentation algorithm has achieved the objective of deducing the problem of over-segmentation when employed to histopathological image. For example, applying only the traditional watershed technique to the image in Figure 4(a) will achieve serious over-segment in Fig.4(f) and Fig.4(h)(the region of Fig.4(f)),but our current methodology is able to produce a better effect nearly no over-segment in Fig.4(i) and Fig.4(g)(the region of Fig.4(i)). Through visual inspections of the segmentation results, it can be observed that there is no visible over-segment. We have tested our method on more than 300 images . Each image has more than 400 nuclei. The rate of over-segmentation computed from those image. The Table I shows that the improved method can apparently solve the over-segment.

The rate of over-segmentation(the number of over-segment cells/the total number of cells)	The direct watershed algorithm	The improved method
	0.69	0.05

Table 1:Quality comparison of nuclei segmentation

4   Conclusion

To Figure out the over-segment problem in traditional watershed algorithm, in this paper, we develop an improved watershed algorithm, which use chamfer distance transformation to transform information of location to different gray information , and utilize H-minima transform to suppress undesired minima leading in over-segment after the first segment of color deconvolution . The experimental results have shown that our proposed process of using distance transformation and H-minima before applying watershed segmentation directly is effective.

Acknowledgements

References

[1]   XM Chen, GQ Guo, and MQ Liu. (2007) "On infrared image segmentation algorithm based on watershed," Journal of Optoelectronics. laser, vol.12, no. 10, pp. 1072-1075.

[2]   SH Liu, LB Zeng, B Liu, and Y Fang. (2007) " Separating algorithm for overlapping granule images, " Computer Engineering, vol. 28, no. 2, pp.198-199.

[3]   Sahirzeeshan Alia and Anant Madabhushib. (2010) "Segmenting multiple overlapping objects via a hybrid active contour model incorporating shape priors: Applications to digital pathology " Medical Imaging 2011: Image Processing, vol.19,no.13.pp.90-95

[4]   T. F. Chan, L. A. Vese. (2007) "Active contours without edges,"IEEE Transactions on Image Processing, 10(2): 266-277.

[5]   Vincent L (1993) Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. IEEE Trans Image Process 2: 176-201.

[6]   Ruifrok AC, Johnston DA (2008) Quantification of histochemical staining by color deconvolution. Anal Quant Cytol Histol 23: 291-299.PubMed: 11531144.

[7]   Ratael C.Gonzalez, Richard E. Woods, and Steven L.Eddins.(2004) Digital Image Processing Using Matlab. Prentice Hall, Upper Saddle River, Version 1.0, 2004.

[8]   Garini Y, Young IT, McNamara G (2006) Spectral imaging: principles and applications. Cytometry A 69: 735-747. PubMed: 16969819.