HHS Public Access

Author manuscript

Proc IEEE Int Symp Biomed Imaging. Author manuscript; available in PMC 2017 April 06.

Published in final edited form as:

Proc IEEE Int Symp Biomed Imaging. 2009 ; 2009: 795–798. doi:10.1109/ISBI.2009.5193169.

AUTOMATED CELL COUNTING AND CLUSTER SEGMENTATION USING CONCAVITY DETECTION AND ELLIPSE FITTING TECHNIQUES

Sonal Kothari[1], **Qaiser Chaudry**[1], and **May D Wang**[1,2,*]

[1]Electrical and Computer Engineering, Georgia Institute of Technology

[2]Biomedical Engineering, Georgia Institute of Technology and Emory University

Abstract

This paper presents a novel, fast and semi-automatic method for accurate cell cluster segmentation and cell counting of digital tissue image samples. In pathological conditions, complex cell clusters are a prominent feature in tissue samples. Segmentation of these clusters is a major challenge for development of an accurate cell counting methodology. We address the issue of cluster segmentation by following a three step process. The first step involves pre-processing required to obtain the appropriate nuclei cluster boundary image from the RGB tissue samples. The second step involves concavity detection at the edge of a cluster to find the points of overlap between two nuclei. The third step involves segmentation at these concavities by using an ellipse-fitting technique. Once the clusters are segmented, individual nuclei are counted to give the cell count. The method was tested on four different types of cancerous tissue samples and shows promising results with a low percentage error, high true positive rate and low false discovery rate.

Introduction

Pathologists often depend on parameters such as the number, shape and size of cells in a tissue sample to make important diagnostic decisions. In healthy conditions, nuclei in cells are mostly distinct and parameters can be determined by direct image segmentation methods such as region-based methods, histogram-based methods and edge detection based methods. However, in pathological conditions, individual cells come close together and nuclei form dense clusters. Figure 2(a) shows dark elliptical nuclei touching and overlapping in a 2-D tissue sample. Therefore, accuracy of cell-counting, cell shape and size determination depends on the segmentation of these dense clusters.

Previous work addresses segmentation of simple-clusters and touching cells by extending and improving image-segmentation methods [1, 2]. Few authors have developed algorithms that address cluster segmentation specifically [3, 4, 5, 6]. All these methods addressing cluster segmentation either could segment only simple clusters [1, 2] or give good results only for circular cells [1, 2, 3] or resulting cell shape is not a good model for the original cell shape [4, 5, 6] or have very complex algorithm [3, 4, 6]. However, this paper presents an

Corresponding Author: May D. Wang, Ph.D., maywang@bme.gatech.edu, (404) 385-1790, 313 Ferst Dr. Atlanta, GA 30332.

edge-based image segmentation method, shown in the flow-diagram (figure 1(a)), that is simple to implement and can segment complex clusters with reasonable accuracy. The method involves detection of concavities on cell cluster edges and segmentation at these concavities by ellipse fitting. The elliptical model used is a good approximation to the original cell shape. Recently, Wang and Song [7] and Bai et al. [8] introduced the concept of cluster segmentation using concavity detection. In this paper, we present a novel method for notch detection using cross-product (section 3), and a new technique for cluster segmentation using ellipse fitting (section 4). Also, we perform quantitative analysis of segmentation result using standard statistic parameters. Using proposed methodology, pathologists will be in better position to take diagnostics decisions.

Preprocessing

We have implemented our method for different types of tissue samples including standard photo micrographs of H&E stained biopsy tissue sections of renal cell carcinoma (RCC) and IHC stained biopsy tissue sections of head and neck (H&N) cancer. (Refer to section 7 for detail.)

Due to the nature of tissue images and the variability in the sample preparation, staining, and image acquisition process, it is imperative to pre-process these images in order to remove variations.

The preprocessing steps have been depicted in figure 1(b) and corresponding images are shown in figure 2. In stained RGB tissue images various entities in a tissue slice such as nuclei, glands, cytoplasm and red blood cells appear as different colors. The first pre-processing step involves the generation of a binary mask for cell nuclei from the RGB image using K-means clustering [9, 10], where seed points are selected by user interaction. The binary mask of a tissue sample often has clusters with holes as shown in figure 2(b). If these holes are not filled, they can be detected as false boundaries during the edge-detection process. Therefore, the next step involves filling in the holes using an algorithm based on morphological reconstruction [11] to obtain properly connected clusters as shown in figure 2(c). Very small objects in the binary mask are generally due to noise and due to misclassification during the k-means clustering. As such, the next step involves noise removal using the size threshold. Images in figure 2(c) and figure 2(d) show the mask before and after noise removal. Based on connected component analysis each object in the image is processed as an individual cluster. The boundary of each cluster is then detected based on a neighborhood of 8 pixels. Figure 2(e) shows result after edge detection. The resulting sequence of pixels that form the boundary of the cluster is then processed using smoothing techniques for better notch detection. Noise or jaggedness present on the edges of the clusters may lead to false concavity detection on the edge and consequently may be treated as a notch for segmentation. Therefore, it is necessary to make the boundary smooth and preserve true concavities. Our algorithm performs simple smoothing using a moving average low-pass filter. The Resulting image after smoothing is shown in figure 2(f).

Concavity or Notch Detection

After preprocessing, the next step involves detection of concavities or notches. A concavity is the point on the cluster edge where two individual cells overlap. Therefore, the concavities can be found using angle (6) between adjacent normals on the edge of the cluster as suggested by previous work [8]. In this method, we divide the edge of the cluster into fixed length segments, and plot a normal at the middle point of every segment as shown in figure 3. If (hi gives the slope of normal at middle point of segment i with respect to positive x-axis, then 6 for any segment i is given by:

$$\theta = \left\{ \begin{array}{ll} |\Phi_i - \Phi_{i-1}|, & \text{if } |\Phi_i - \Phi_{i-1}| < \pi \\ \pi - |\Phi_i - \Phi_{i-1}|, & \text{else} \end{array} \right\} \quad (1)$$

Edges have a depression around a concavity and a sudden change in surface orientation. Hence, 6 has maxima at concavities as shown in figure 4(a). As illustrated in the graph using a high threshold (dotted line) only major concavities are discovered. These concavities are sharp concavities and in order to discover relatively smooth concavities the threshold needs to be decreased (solid line) and with this decrease some false detections start appearing at points with sufficient angle change, such as the ones at the edge of the individual elliptical cell with high eccentricity.

To avoid these false detections with a decrease in threshold, we exploit the fact that the desired concavities are located at the edge where the surface is concave (when viewed from inside the cell). As such, any detection at locations where the surface is convex can be rejected. The process involves splitting the cluster edge into segments of equal length. Vectors are generated for tangents at every segment. The cross product of each pair of adjacent tangential vectors is calculated while moving in clockwise direction along the cluster edge. The magnitude of the cross product depends on the angle between the vectors and its sign depends on the direction in which the first vector moves towards the second vector.

$$\mathbf{a} \times \mathbf{b} = |\mathbf{a}||\mathbf{b}| \sin \theta \hat{\mathbf{n}} \quad (2)$$

Where, a and b are first (dotted line) and second (solid line) tangential vectors; is the angle between the vectors, and n is the unit vector perpendicular to a and b in the direction given by right hand rule.

If cluster is in X-Y plane, Z component of the cross product represents magnitude and direction of cross product. As shown in figure 5 at convex surface, the direction of movement is clockwise and n is negative z-direction. While at concave surface, the direction of movement is anti-clockwise and consequently n is positive z-direction. High positive z component represents notch while negative value represent convex surface. Comparing figure 4(a) and figure 4(b), it can be observed that false detections in figure 4(b) are no longer affecting the concavity detection which can now be done at relatively lower threshold.

Segmentation

After detecting the concavities, we segment the cluster into cells. The segmentation algorithm assumes that cells have approximately elliptical shape with different eccentricities, suggested by a various authors. Figure 6 shows the flow chart for the segmentation of clusters.

4.1. Cell-size computation

First, the average cell size is computed based on user interaction. The user selects a few samples of single cells. We then compute the average cell size from samples and use it to set a threshold for identifying cell clusters.

4.2. Cluster identification

This step differentiates between cluster, single cell and noise. Depending on the standard cell size, two thresholds are set. The first threshold decides if any region is large enough to be treated as a cluster, and the second threshold decides if it is small enough to be treated as noise. The detected clusters undergo further segmentation while single cells are passed to ellipse fitting algorithm. Due to these thresholds the methodology is robust in segmenting cells in a tissue image with cell size within a range from the average cell-size.

4.3. Notch pairing based on distance threshold

In this step, we compare the distance between all the notches. Any two notches that are closer than a particular threshold, which depends on average cell size, form a pair and the cluster is split at these two notches; preference is given to the notches that are closer. Cluster splitting based on distance splitting is continued iteratively until the point is reached where no further segmentation is possible. During this process each cluster segments into either sub-clusters or sub-clusters and individual cells. The sub-clusters generated during this step generally have circular shape and cannot be further split using the threshold criteria. Therefore, these sub-clusters are passed to next step for segmentation based on centroid.

4.4. Notch pairing using centroid

After distance-based segmentation, if there are any clusters left, they are segmented by using centroid connection. In this step, starting with the notch with highest z-component (section 3), notches are connected through the centroid to split the cluster into cells. Any segmentation step is possible only if it results in regions with size larger than a minimum cell size threshold.

4.5. Ellipse fitting

We picked up the ellipse fitting method proposed by Fitzgibbon et al. [12] which is reported to have better accuracy than other standard methods. The edge pixels of cells obtained in the segmented mask after step 4.4 are compared with the edges of the cluster as shown in figure 2(f). Common edges pixels are then used as data for ellipse fitting algorithm. The final result of the segmentation of the input image in figure 7(a) is shown in figure 7(b); black lines mark the cell boundaries.

Results and Discussions

In order to test the robustness of our algorithm, we selected H&E stained tissue images from three subtypes of renal cell carcinoma (RCC) – papillary (PA), chromophobe (CH) and clear cell (CC) and IHC stained head and neck (H&N) cancer tissue images, thereby introducing morphological structure variations (RCC subtypes) and stain color variations (H&E and IHC). Quantitative analysis of these four different types (PA,CH, CC, H&N) of tissue images, numbered 1–4 respectively, is shown in table 1,. Estimated number (EN) was calculated by the algorithm and false positive (FP) and False negative (FN) were estimated by comparison with manual segmentation results. The results have been analyzed based on three standard statistics parameters – True positive rate, false discovery rate and percentage error. High true positive rate, low false discovery rate and low error rate illustrate the usefulness of the method for cell-counting of various tissue samples. The method is simple to implement and can generate results in real-time, this is highly suitable for clinical applications. The method is semi-automatic and requires user interaction only in seed selection (section 2) and cell size calculation (section 4.1).

As compared to previous methods of segmentation using concavities [6, 7, 8], our method will generate better results due to higher accuracy in concavity detection. The method may generate errors at places when two cells overlap in such a fashion that concavities are very smooth or absent. Also the results are dependent on how good color segmentation is performed during the k-means clustering process for generation of the binary mask. The future work includes efforts to improve the color-segmentation and the cluster segmentation to further enhance the efficiency of the method. Also method is being tested for larger dataset of about 100 images to evaluate the robustness of the method.

Acknowledgments

We thank Dr. Mitch Parry for his valuable comments and suggestions. This research has been supported by grants from National Institutes of Health (Bioengineering Research Partnership R01CA108468, P20GM072069, CCNE U54CA119338), Georgia Cancer Coalition, Hewlett Packard, and Microsoft Research. RCC subtypes data was provided by Dr. Andrew N Young, Emory University. H&N cancer data was provided by Dr. Georgia Z. Chen, Emory University, Atlanta, GA 30332 USA.

References

1. Ruberto CD, et al. Segmentation of Blood Images using Morphological Operators. ICPR. 2000; 3:397–400.
2. Refai, H., et al. ICIP. Vol. 2. IEEE; 2003. Automatic count of hepatocytes in microscopic images; p. II-1101-II-1104.vol.3
3. Thévenaz P, et al. Snakuscules. IEEE Transactions on image processing. 2008; 17:585–593. [PubMed: 18390366]
4. Nilsson, B., et al. CVPR. IEEE; 2001. Segmentation of dense leukocyte clusters; p. 221-227.
5. Glory, E., Meas-Y edid, et al. Biomedical Imaging: Nano to Macro. IEEE; 2006. Automated image-based screening of cell cultures for cell therapy; p. 259-262.
6. Schmitta O, et al. Radial symmetries based decomposition of cell clusters in binary and gray level images. Pattern Recognition. 2008; 41:1905–1923.
7. Wang, W., et al. Third International Conference on Natural Computation. IEEE; 2007. Cell Cluster Image Segmentation on Form Analysis; p. 833-836.

8. Bai, X., et al. DICTA. IEEE; 2008 Dec.. Touching Cells Splitting by Using Concave Points and Ellipse Fitting; p. 271-278.
9. Chaudry, Q., et al. BioInformatics and BioEngineering. IEEE; 2008 Oct. Improving renal cell carcinoma classification by automatic region of interest selection; p. 1-6.
10. Weeks AR, et al. Color Segmentation in the HSI Color Space Using the k-means Algorithm. Proc. of the SPIE - Nonlinear Image Processing VIII. 1997:143–154.
11. Soille, P. Morphological Image Analysis: Principles and Applications. Springer-Verlag; 1999. p. 173-174.
12. Fitzgibbon, A., et al. Transactions on Pattern Analysis and Machine Intelligence. Vol. 21. IEEE; 1999 May. Direct least square fitting of ellipses; p. 476-480.

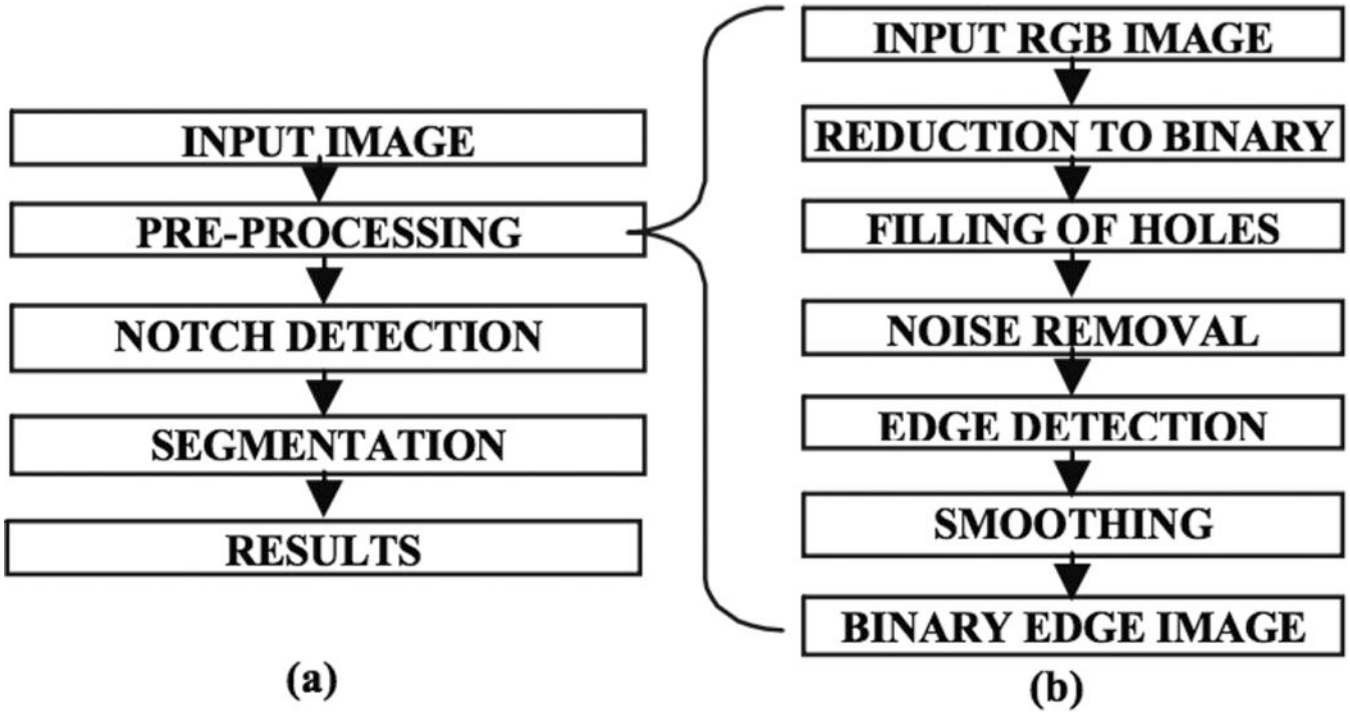

Figure 1.
a) Overall flow-diagram for the method, b) Flow diagram of pre-processing steps

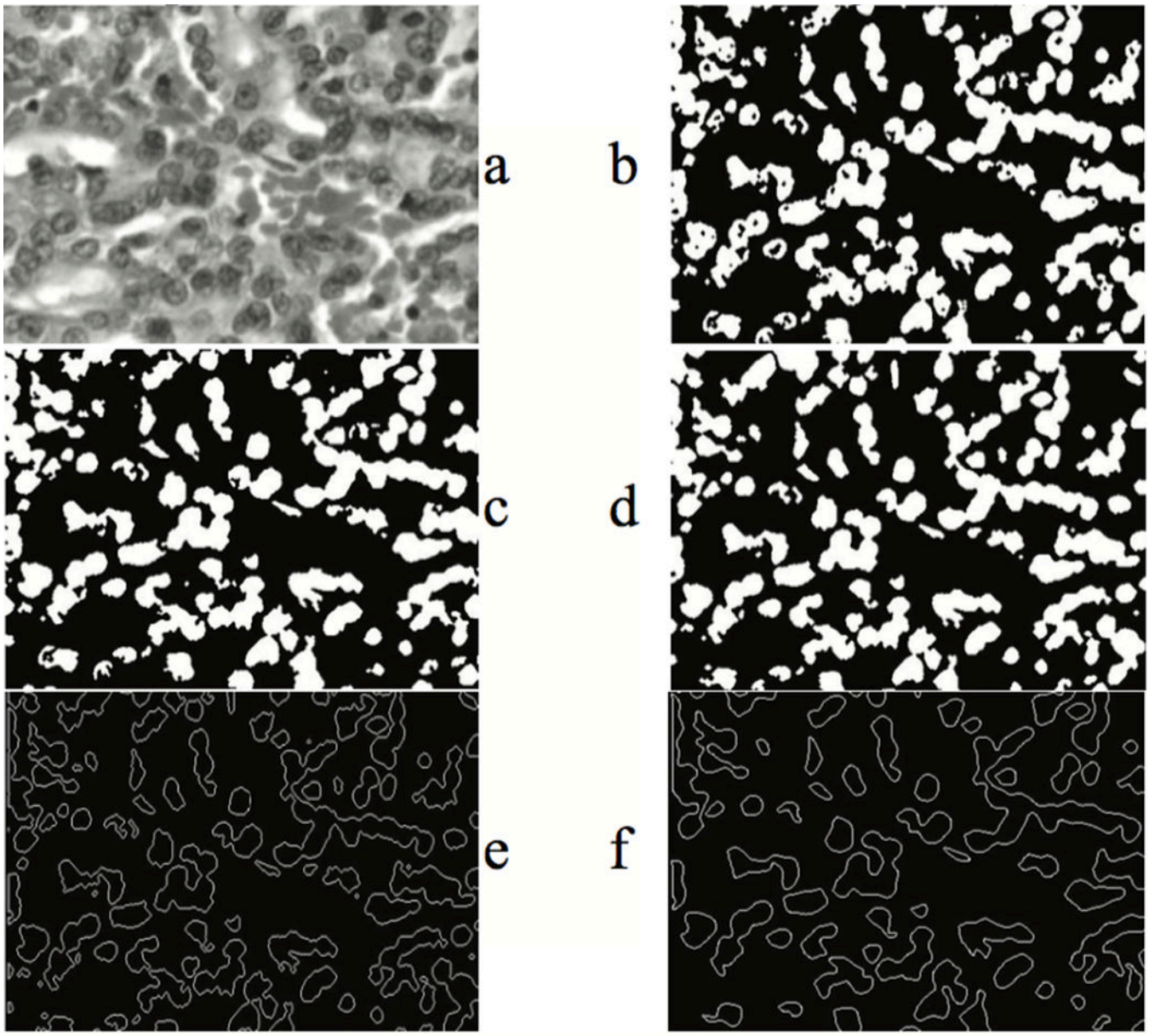

Figure 2.

Pre-processing steps implemented for papillary tissue sample. (a) Input RGB image shown in gray scale, (b) binary mask of nuclei, (c), filled binary mask, (d) mask after noise removal (e) result after edge detection, (f) result after smoothing

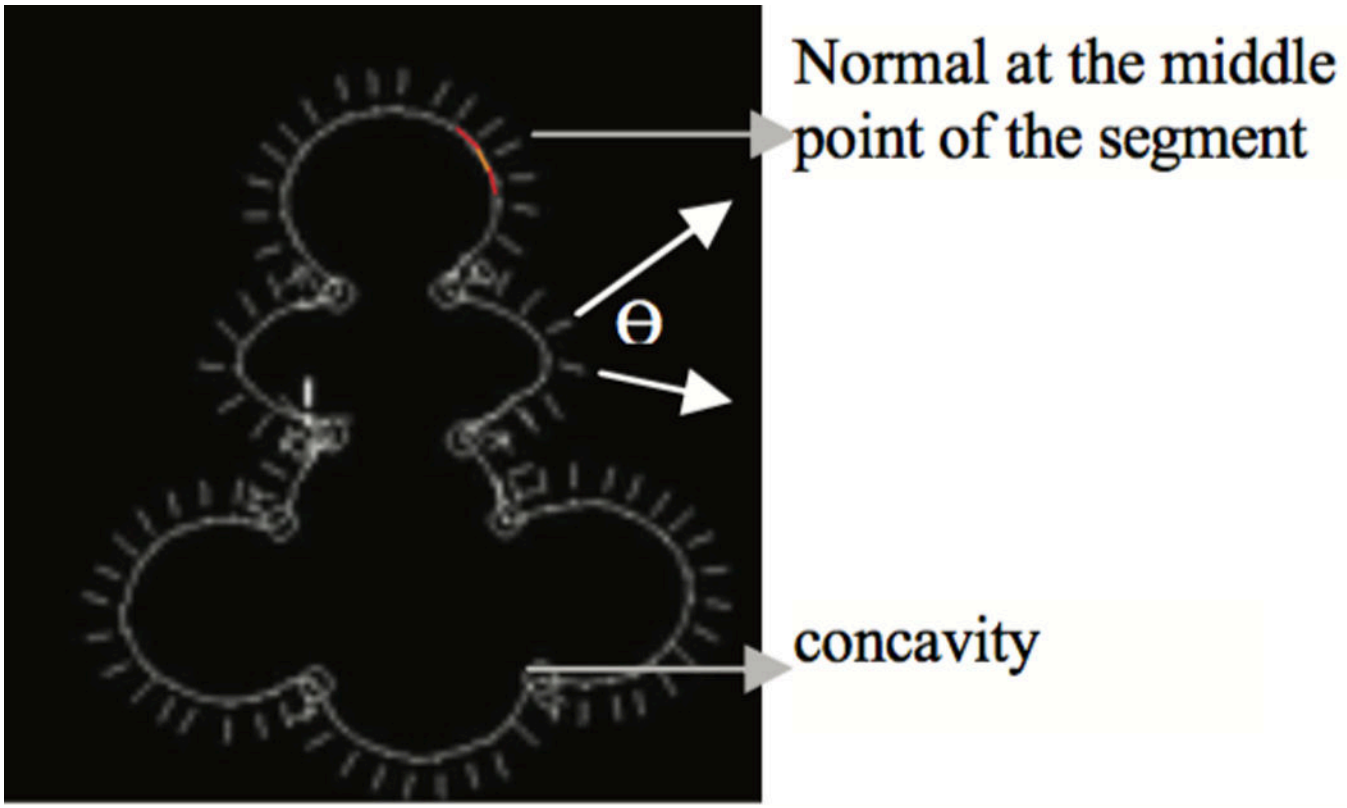

Figure 3.
A synthetic cluster illustrating the method of calculating ∈, angle between adjacent normals.

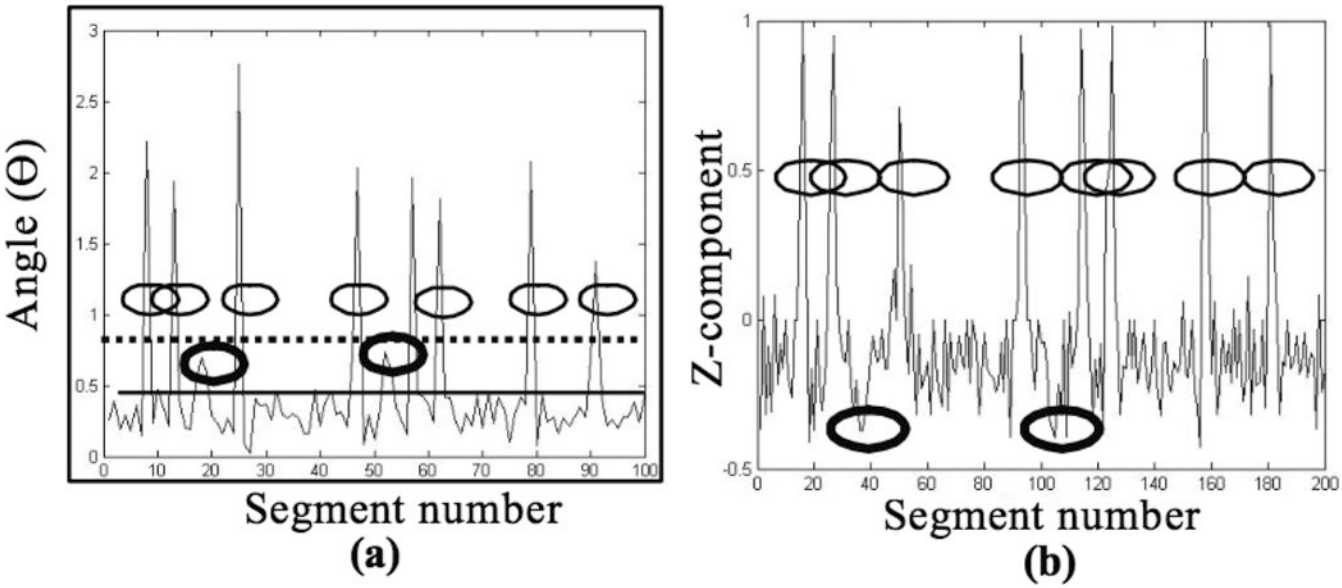

Figure 4.

a) Graph illustrating angles between adjacent normals for different segment number along the edge of cluster in figure 3. Dotted line and complete line represents high and low threshold respectively. b) Relation between z-component and segment number for the same cluster. Thin circles mark true concavities and thick circles mark the possible false concavities

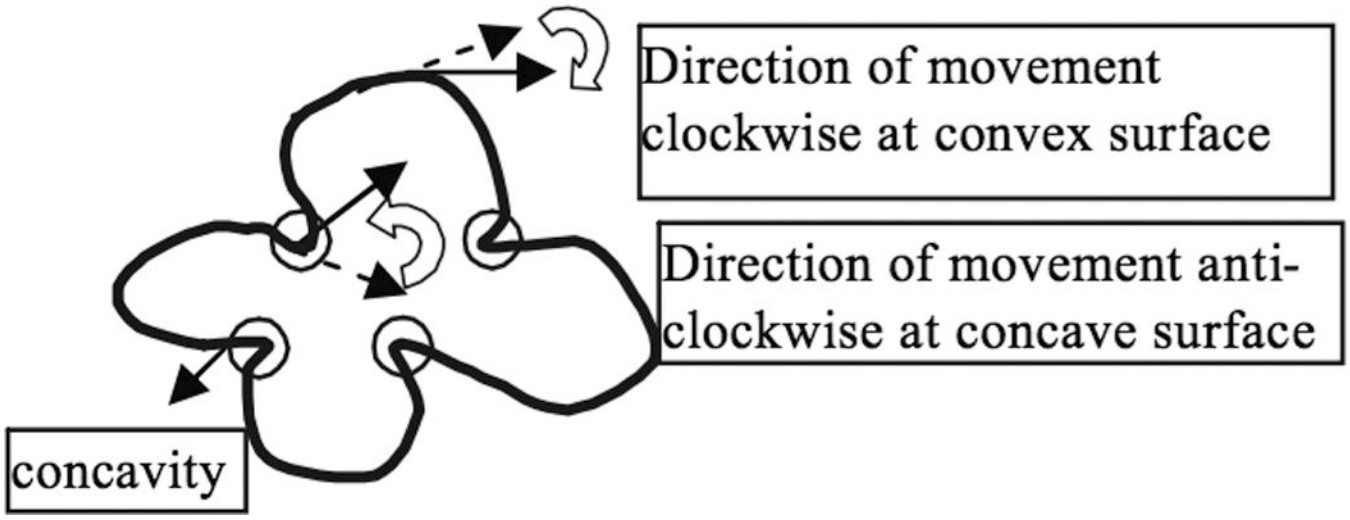

Figure 5.
Figure depicts cross product resultant direction in case of convex and concave contour locations; concavities are marked

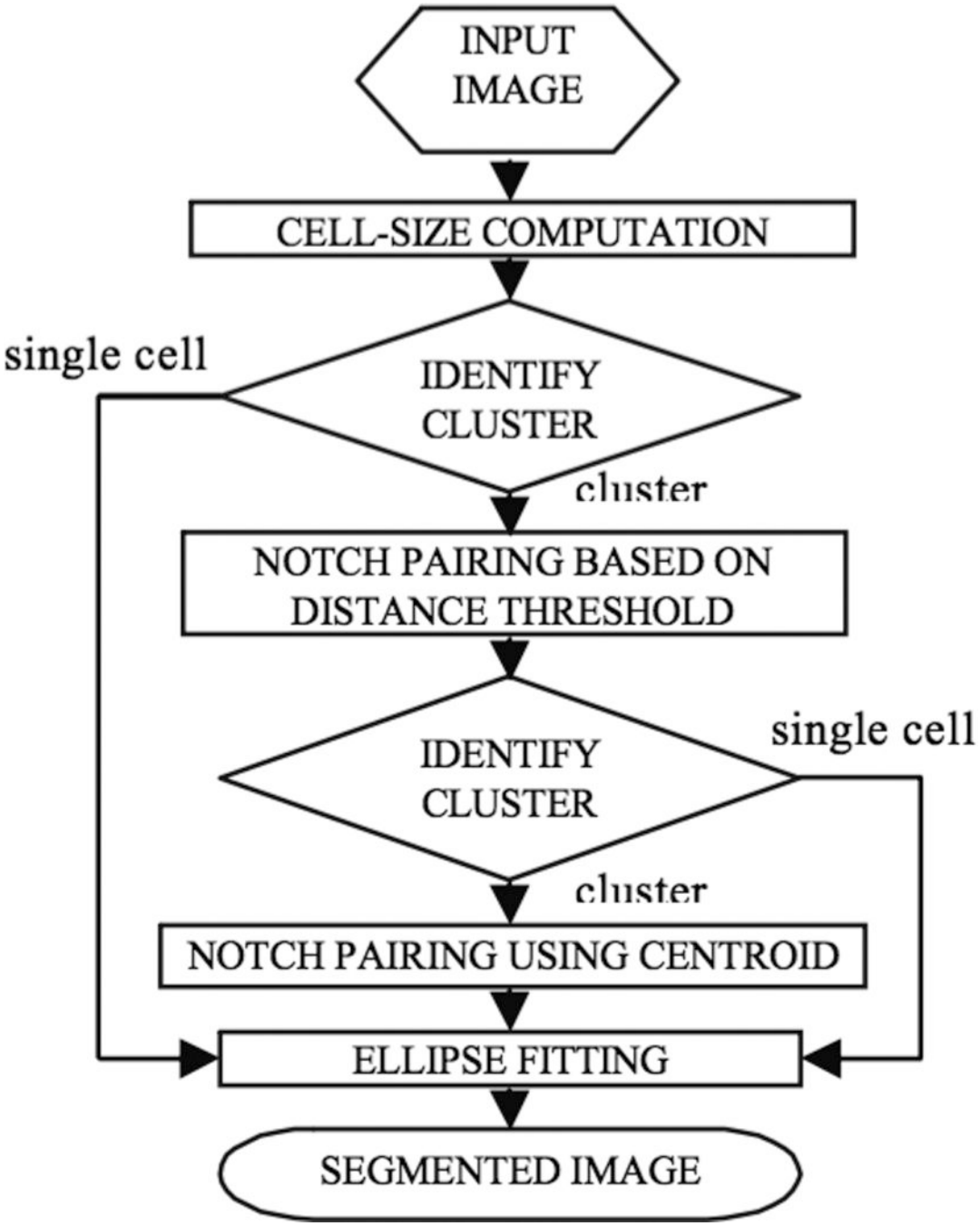

Figure 6.
Flow-chart for Segmentation

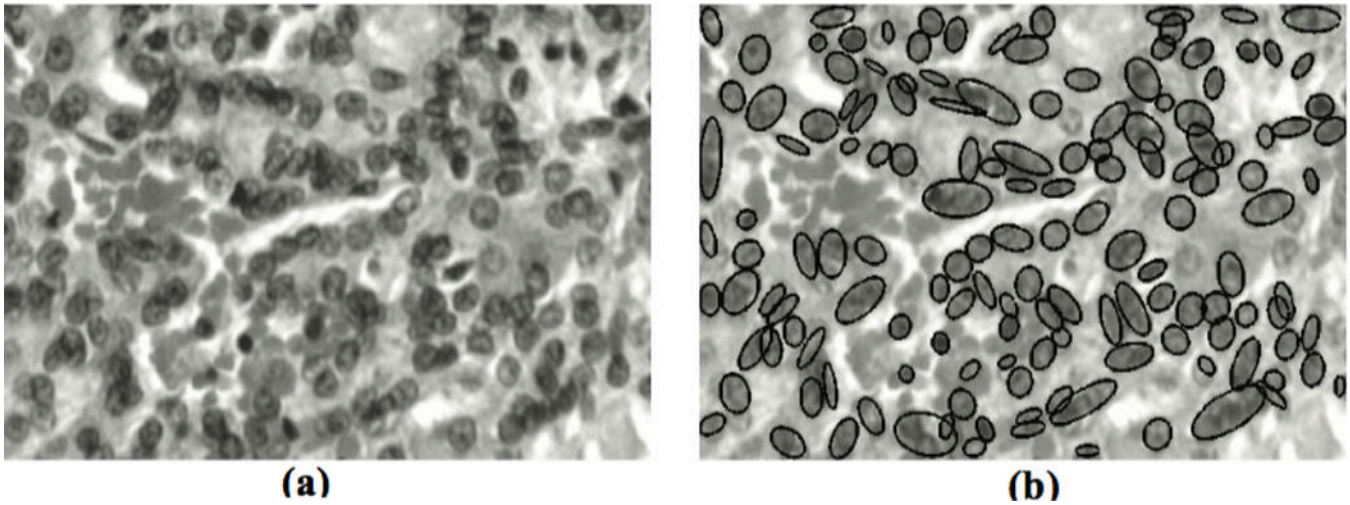

Figure 7.

a) Input papillary tissue, b) result image after segmentation of image, green line mark the cell boundaries

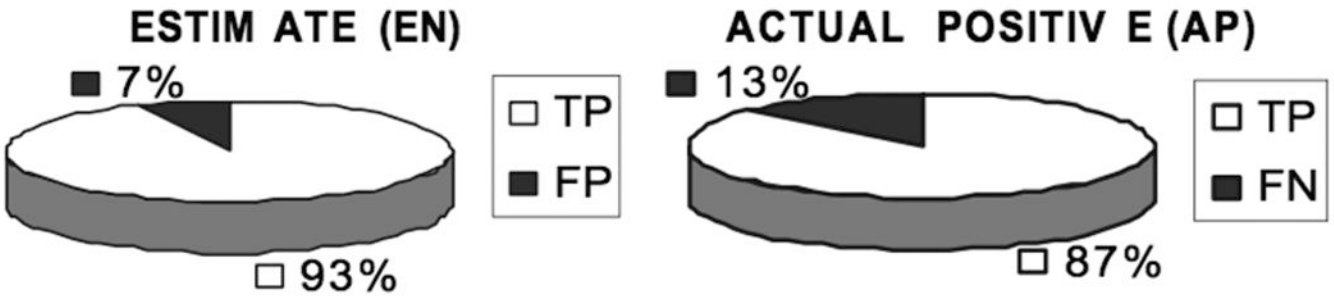

Figure 8.
Pie chart representation of estimate and actual positive for image number 1 in Table 1

Table 1

Image number	EN	FP	FN	TP	AP	TPR	FDR	E
1.	301	20	43	281	324	86.72	6.64	−7.09
2.	423	32	33	391	424	92.22	7.5	−0.23
3.	338	40	27	298	325	91.69	11.8	4.00
4.	480	60	22	420	442	95.02	12.5	8.59

Acronyms used in these tables are as follows: Estimated-Number of cells segmented by the method, EN; False positive-false detection, FP; False negative-missed detection, FN: True positive-correct detection, TP=EN − FP; Actual positive-number of cells calculated manually, AP=TP+FN; True positive rate-hit rate/sensitivity, TPR=TP/AP *100; False discovery rate, FDR=FP/EN *100; Percentage Error, E = (EN−AP)/AP *100