

BIOMEDICAL IMAGE SEGMENTATION

Segmentation of Primary Breast Tumor Nuclei in Histological Images

Submitted by:

Sai Chetan Kumar Gudepu

2013

Segmentation of Primary Breast Tumor Nuclei in Histological Images

Sai Chetan Kumar Gudepu, Mark Zarella, David E. Breen, Fernando Garcia

Advanced Pathology Imaging Laboratory, Drexel University College of Medicine and

Drexel Geometric Biomedical Computing Group, Department of Computer Science, Drexel University

1. INTRODUCTION

Breast cancer (BCa) is a heterogeneous and diverse disease. They are sub-classified into four major subtypes – luminal A, luminal B, Her2-overexpressing and basal-like [1, 2]. It has been seen that the phenotypic variability between BCa occurs within, but not across these major subtypes. These subtypes not only have distinct behaviors but also differ in responses to therapy which highlights the importance of identifying each subtype of BCa for appropriate therapeutic decisions [3]. In order to determine pathologic staging, pathologists routinely evaluate various features like Regional lymph node metastasis status and histologic grade. Such histological analysis though useful and cost-effective, largely depends on the experience of the pathologist. To achieve a better reproducibility and reduced dependence on the pathologist, there is a need to develop a system to objectively predict tumor subtype which was previously possible only through expensive molecular testing and immunohistochemistry (IHC).

2. PURPOSE

In order to establish an Image analysis paradigm to generate predictions of tumor sub-type objectively, a reliable method to segment nuclei to analyze their properties individually has to be developed.

3. METHODS

The region of interest extracted from the IHC images has been pre-processed using MATLAB v.2012a before the segmentation algorithm was applied. Contrast of the images had been enhanced by mapping the intensity values such that 1% of the data was saturated at low and high intensities of the grayscale image. A binary mask of this image had been obtained using the classic threshold selection method by Otsu [4]. This was then followed by eroding the nuclei using a ‘disk’, filling the holes and then dilating them using the same structural element to preserve the size of the nuclei. The result of these three steps in the binary mask can be seen by comparing Fig.1 and 2.

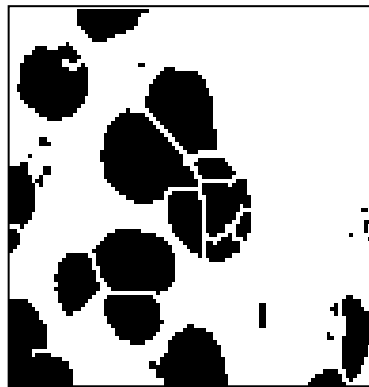


Fig.1: Before Preprocessing

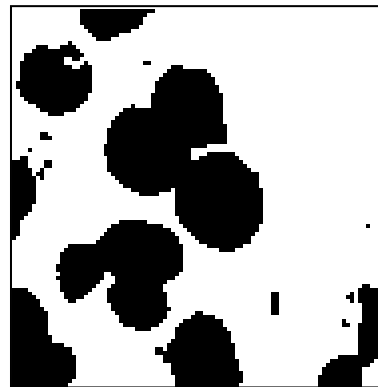


Fig.2: After Preprocessing

The resulting binary image was then passed through a two step Segmentation algorithm; first of which is the widely used Watershed transform. A Euclidian distance map (EDM) is first created which replaces each foreground pixel (nucleus) in the binary image with a gray value equal to that pixel's distance from the nearest background pixel. Starting from the peak/local maxima of the EDM which is referred to as an Ultimate eroded point (UEP), watershed algorithm dilates each of the local maxima until the edge of the nucleus is reached, or the edge of the region of another growing UEP is reached. The result of this watershed algorithm is shown in Fig.3(b).

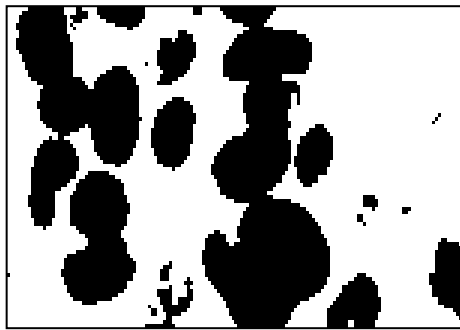


Fig.3(a): Binary mask

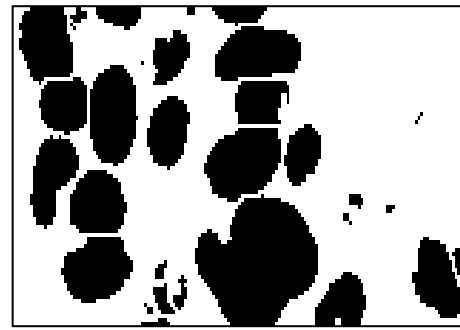


Fig.3(b): Watershed Output

As seen from the Fig.3(b), Watershed clearly works better when the nuclei don't overlap too much since there would just be one UEP in such case. To address this, the output of watershed is passed through a sequence of steps as shown in Fig.4 in order to split the clumped nuclei [5]. The details of major steps of the algorithm are presented in the following subsections.

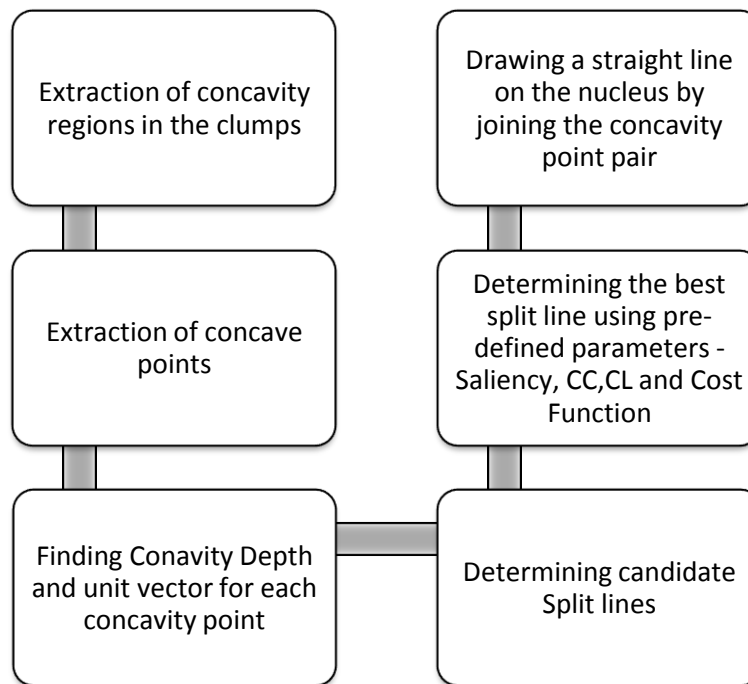


Fig.4: shows the sequence of steps in clump segmentation

Before proceeding with the second stage of the segmentation algorithm, clumps had to be differentiated with the individual nuclei. Area was chosen as a parameter for this purpose and by considering objects with areas greater than twice of standard deviation of areas of all objects, 500 pixels was determined to be appropriate to differentiate clumps from the individual nuclei.

Table 1: Determination of area threshold

Mean of Areas (M)	Standard deviation (SD)	M+2*SD
171.3295	162.7135	496.7566

3.1. Extraction of Concavity Regions

In order to find all the concavity regions present in each of the clumps, convex hull of the object (refers to individual clump) is determined and a polygon is inscribed in it. A simple logical ‘and’ to find the pixels where the inscribed polygon is 1 and the object is 0 would result in the extraction of concavity regions R1, R2 and R3 as shown in the figure 5.

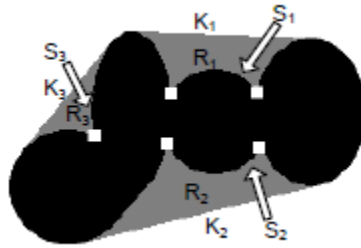


Fig. 5: shows the Concavity regions R_i , corresponding edge segments S_i and Convex hull chords K_i .

3.2. Extraction of Concave points

In any concavity point based splitting method; the detection of concavity points on the boundary of objects presents the first challenge. In this algorithm, an approach using curvature analysis has been used to detect corners (concavity points) as described in [6]. Edge contours are extracted using canny edge detector and the gaps in the contours are filled. The subsequent steps in this method are discussed below.

- Circular convolution is applied to smooth the contour by using a Gaussian smoothing kernel g of $\sigma = 3$. After that, the curvature value of each pixel of the contour is computed using

$$K_i^j = \frac{\Delta x_i^j \Delta^2 y_i^j - \Delta^2 x_i^j \Delta y_i^j}{[(\Delta x_i^j)^2 + (\Delta y_i^j)^2]^{1.5}} \quad \text{for } i = 1, 2, \dots, N,$$

The first and second derivatives of x and y coordinates of each pixel are obtained by convolving with the Gaussian kernel. Here, K_i represents the value of curvature for a point (x_i, y_i) on the j th contour (object). All the local maxima of this curvature function are included in the initial list of corner points.

- Since the previously determined list would include all the local maxima, even the round nuclei would include concavity points owing to the curvature differences on the contour.

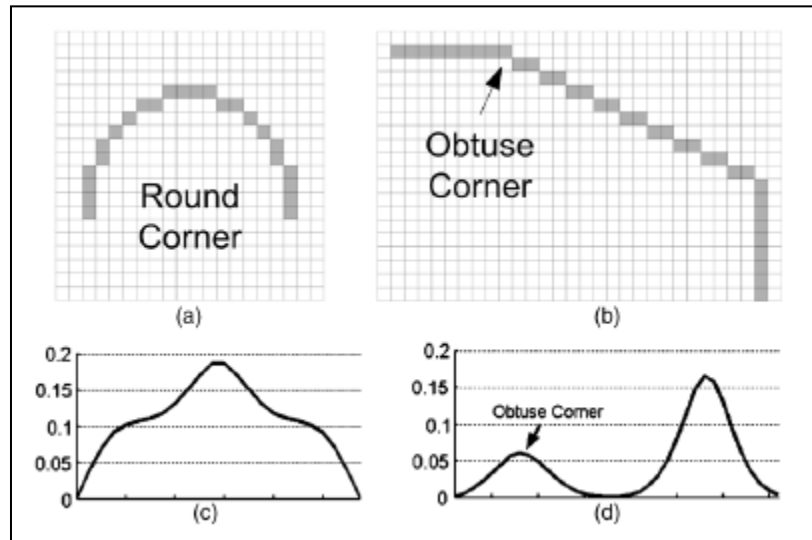


Fig.6: shows the curvature plots for Round and Obtuse corners

As seen from the curvature plots in Fig.6, although the curvature of a round corner is the largest among its neighbors, the actual difference is not significant compared to a similar plot for an obtuse corner which is our point of interest in clumps. In order to separate and remove these round corners and preserve just the obtuse corners, a Region of Support (ROS) is defined. ROS of a particular corner point is defined by the segment of the contour bounded by the pixel's two nearest curvature minima. Mean of curvature values in this ROS is determined and compared with the corner's curvature value to differentiate round corners from obtuse ones. The computed mean thus acts like a threshold and since it changes with each corner point as per the criteria discussed, it is adaptive too. All the corner points which have a curvature value more than the mean of ROS are considered true corners. This is because the curvature drops faster over the ROS for an obtuse corner than it does for a round corner and hence if the curvature value of the corner point is more than the mean, it is considered a true corner and the rest are eliminated from the list.

- The list obtained at the end of previous step would include all the potential true corners but needs a bit of fine tuning in order to avoid including minor spikes/valleys on the contour. For this purpose, for each corner point, the ROS determined is extended to include two neighboring corner points.

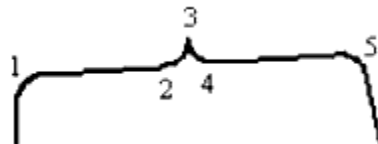


Fig.7: shows a minor spike over a contour

After determining the ROS of corner points, the angle of a corner is defined as the angle between the lines joining the corner point and the two centers of mass on both sides of the ROS, where the centre of mass is defined as the mean position of all pixels on one arm of the ROS. This definition is redefined for corners on arcs using tangents [6]. In figure 7, if the points 2 and 4 are preserved after removing round corners, the angle of point 3 would be acute and if they are removed, angle of point 3 with 1 and 5 would almost be a straight line (obtuse), and thereby implies point 3 is the result of some local variation on the curve – spike. The effect of this angle to differentiate between true and false corners was seen and the value of 180 degrees was fixed above which the corners were eliminated from the list.

3.3. Concavity Depth and Unit Vector

Concavity depth (CD) is the perpendicular distance (Euclidian distance metric has been used) between the concavity point and the convex hull chord of the concavity region to which the point belongs.

In order to determine the orientation of the concavity region, unit vector of the concavity point is obtained. A predefined distance of 7 index values was used to determine two points on either side of the concavity point. Mid-point of the line joining those points is considered to be the tail-point of the vector while the head of the vector is put on the line that connects this point and the concavity point (Fig.8).

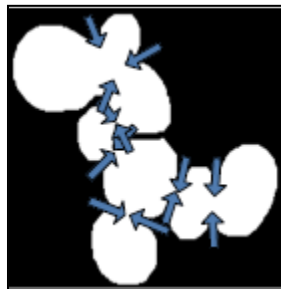


Fig.8: shows the unit vectors

3.4. Determining Candidate Split Lines

Normally, when we consider all the possible split lines, there would be a lot of redundancy and most of these split lines either are invalid (pass through the background) or intersect each other and hence it is necessary to exclude them. The function `Delaunay` in MATLAB can be used to eliminate such lines. The properties of Delaunay triangulation are such that it disallows the lines to intersect each other and that it maximizes the minimum of the interior angles of the triangles formed between the split lines to obtain a list of candidate split lines.

3.5. Determining Best Split Line

A set of features – Saliency, Concavity-concavity alignment (CC), and Concavity-line alignment (CL) are taken for all the concavity points in the list of candidate split lines. A Split line is chosen so that it is as short as possible in addition to having concave enough regions at both of its ends. This is ensured by Saliency by considering Concavity depth as well as the distance between the two concavity points.

$$SA_{i,j} = \frac{\min(CD_i, CD_j)}{0.1 \times \min(CD_i, CD_j)^2 + D_e(C_i, C_j)^2}$$

SA represents Saliency between points i and j; and D_e represents the Euclidian distance between those points.

A valid split line is characterized by a good alignment between the concavity regions. CC and CL alignment features take this into consideration which is based on the orientation of the two concavity regions obtained by the corresponding unit vectors.

CC and CL features are determined as defined in [7] and are given by:

$$CC_{i,j} = \pi - \cos^{-1}(v_i \cdot v_j),$$

where v_i and v_j are the unit vectors of the two concavity points. CL alignment describes how well the two regions are aligned with the split line and is given by:

$$CL_{i,j} = \max(\cos^{-1}(v_i \cdot u_{ij}), \cos^{-1}(v_j \cdot u_{ji}))$$

Where u_{ij} is the unit vector along the line from point i to j.

A Cost Function is defined by incorporating these features and is given by:

$$CF_{i,j} = SA_{i,j} + CC_{i,j} + 2 \times CL_{i,j}$$

The points i and j which result in minimizing the cost function are paired and the split lines are applied to the binary image to split the nuclei clumps.

3.6. Drawing Line on the Object

A straight line with value zero is drawn after obtaining the x and y coordinates of concave-point pair.



Fig.9 (a): Clump of nuclei

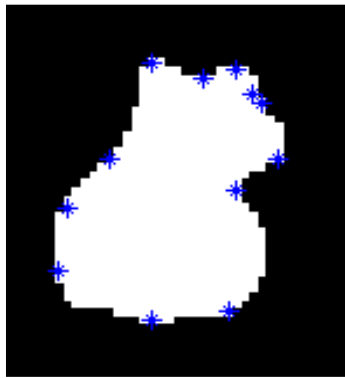


Fig.9 (b): Concave points



Fig.9 (c): Splitted Clump

4. RESULTS

4.1. Comparison with Initial Segmentation

A random sample of 5 images has been chosen and the result of the initial Watershed Segmentation is compared to the final result. Binary map of one such case is shown in the Figure.10.

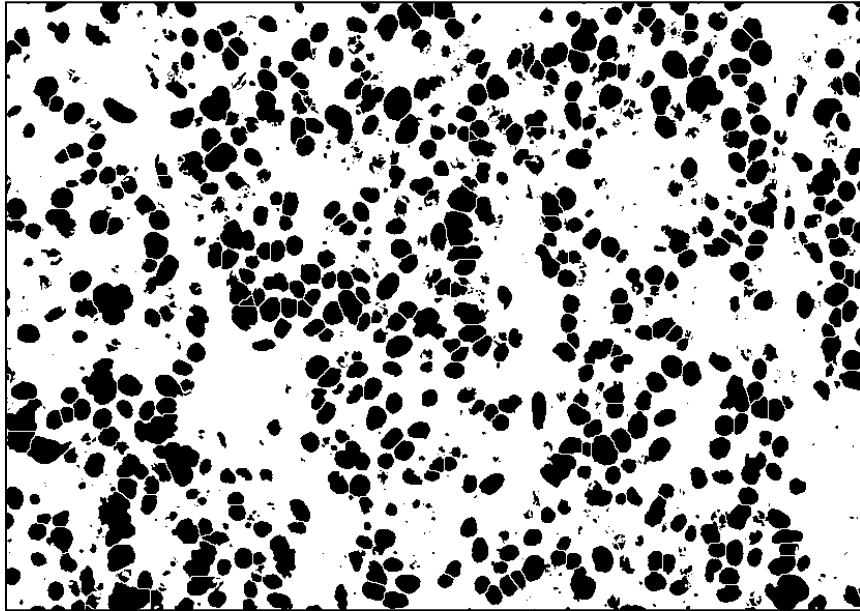


Fig. 10 (a): shows the Watershed Segmentation

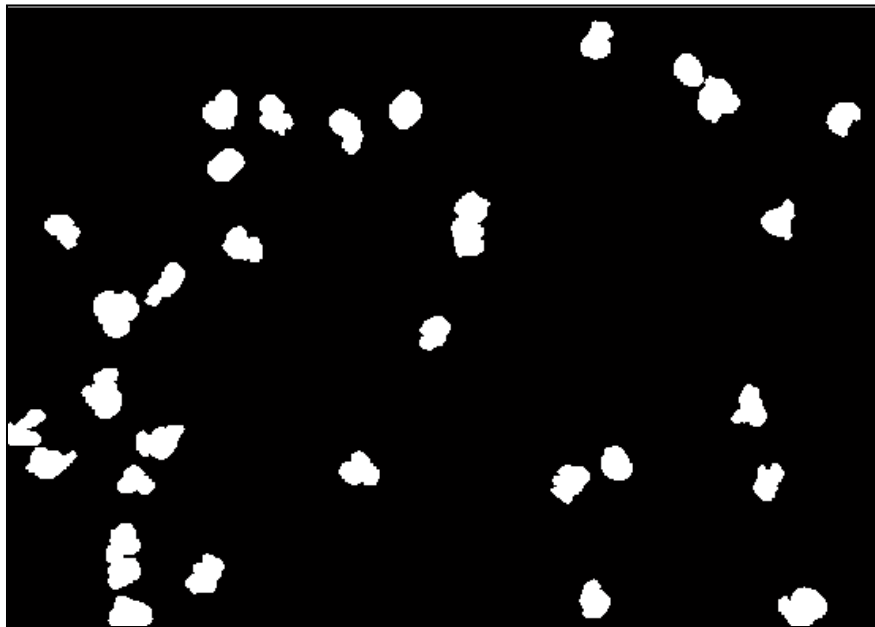


Fig. 10 (b): shows the Clump map

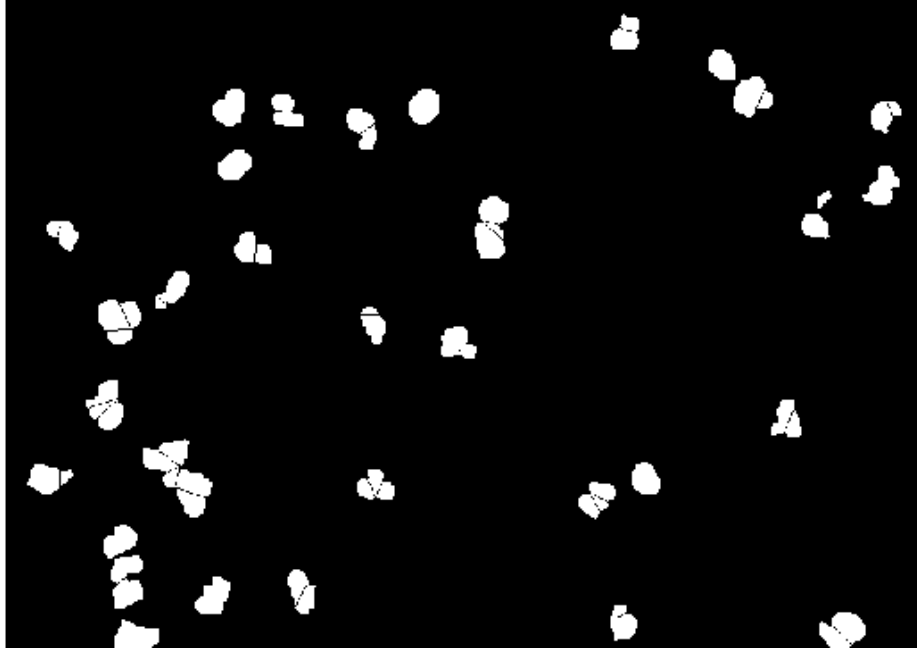


Fig. 10 (c): shows the Clump map after Splitting

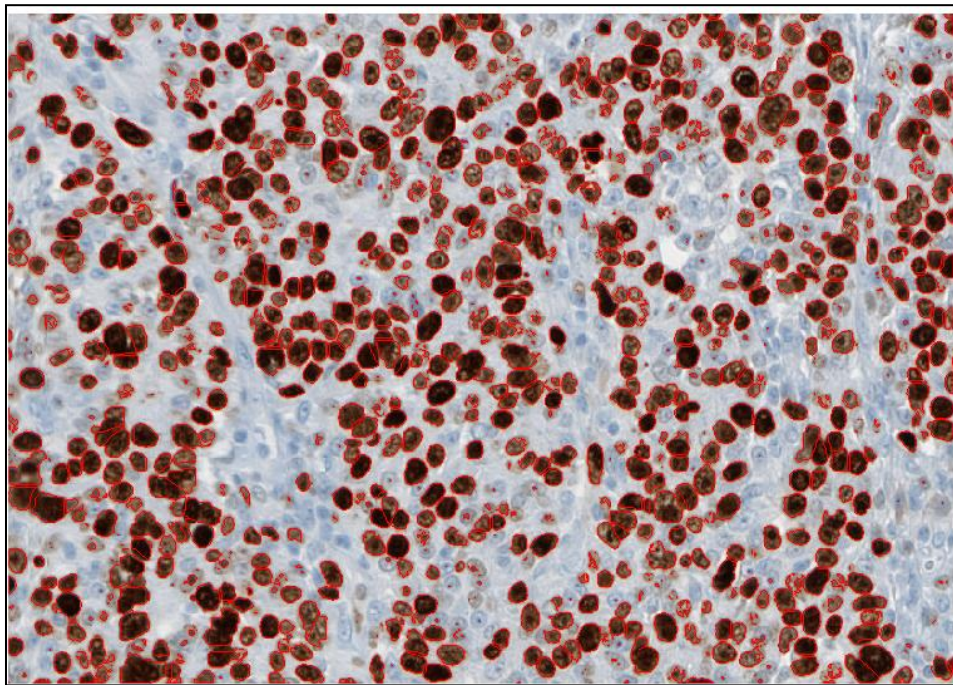


Fig. 10 (d): shows the Final Segmentation

4.2. Comparison with Preliminary Results

Performance of the current method, result from preliminary segmentation by David.E.Breen et al. as described in [8] has been evaluated on 1 image. A region of interest of 300x300 was chosen at two places on the image (a and b) and the ground truth was determined by manually counting the number of nuclei. As seen from Table.2, the mean percentage improvement of error is 34.57%.

Table.2 shows the Performance Evaluation

Slide. No.	Ground Truth	No. of objects returned by [8]	No. of objects returned by the current method	Error % (old)	Error % (New)	Improvement
71528_a	108	34	115	68.52	6.48	62.04
71528_b	116	80	144	31.1	24	7.1

Mean Improvement in Error % = 34.57%

The large difference in the Improvement % can be attributed to the way the ROI was chosen and due to the removal of the objects touching the boundary of the image by using a 4-neighborhood. The figures 11(a) and (b) demonstrates this improvement in performance.

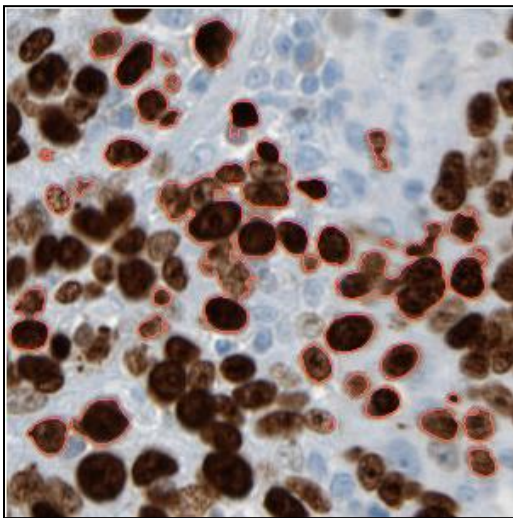


Fig.11 (a): Shows Preliminary result

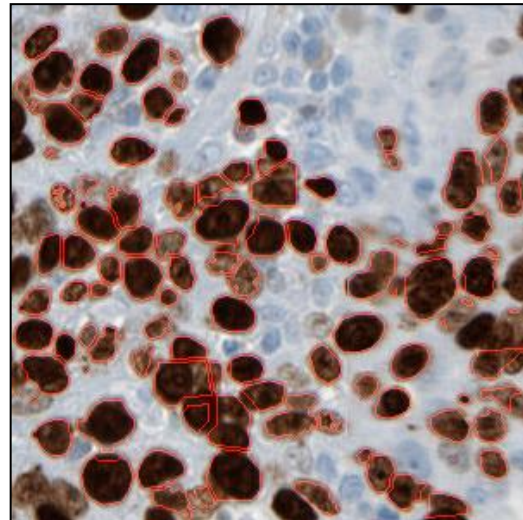


Fig.11 (b): Shows Result of Current Method

Though there is certainly an improvement over the preliminary segmentation, the current method fails in identifying the concave points at certain places as shown in Fig. 12. The possible reasons for this failure and the potential ways to obtain an improved segmentation are discussed in the subsequent sections.

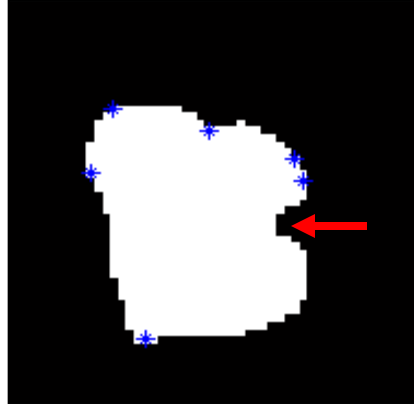


Fig.12 (a): Failure case 1

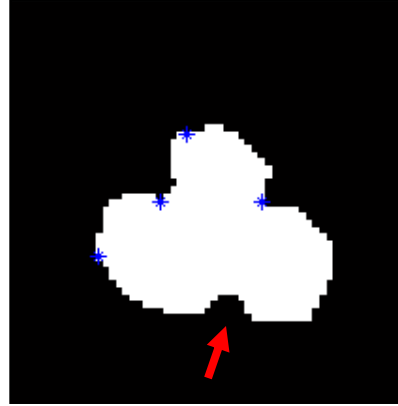


Fig.12 (a): Failure case 2

As seen from the Fig.12, the current algorithm could not detect all the useful concave points and hence the clump could not be split.

4.3. Possible Reasons for Failure

- The ROS to remove false corners is defined to include the adjacent corners and if the angle between the tangents from both ends is more than 180 degrees, the corners are eliminated. This is probably the reason for the non-identification of the more obvious corners in Fig.12 (shown in red). If the angle to differentiate true and false corners is increased to say, 340 degrees it would pick up a lot of false corners due to the local variations on the contour resulting in an in-correct splitting. The key to the success of this approach is to correctly define the angle of a corner and the extent/support over which the angle is defined.
- Edge detection may result in the loss of critical information on the contours. Since edges depend on the gradient information, it is practically not possible to obtain a well defined edge in case of microscopic images and hence to plug the discontinuities in the contour, a small structural element was used to dilate the contour. Though this does plug the discontinuities, there is always a chance of losing a spike/valley on the object in the undetected areas of the contour thereby resulting in non-detection of concave points on which the entire second stage of the algorithm is based upon. This can be possibly avoided/ improved in the second pass by some of the methods discussed in the next section.

5. FUTURE SCOPE

A list of possible ideas to explore which could result in a further improvement over the current algorithm is listed below:

- A modified watershed algorithm applied to clumps could be explored.
- Linear discriminant analysis and the related Fischer's linear discriminant can be used as a linear classifier to separate two/more nuclei inside the clump based on the intensities of overlapping nuclei.
- Maximum Entropy classifiers may be explored.
- Texture information inside the contour could be used to split the clumps.
- Shape features like eccentricity could be used.
- Morphological gradient on a gray scale image combined with a factor multiplied – Otsu can be used
- An algorithm can be developed that can make use of a relevant set of image features like Orientation/Curvature/Texture etc. and identifying discontinuities within each feature domain. This may lead to an effective and improved segmentation.

REFERENCES

- [1] Perou, C.M., et al., *Molecular portraits of human breast tumors*. Nature,2000/ **406**(6797):p.747-52
- [2] Sorlie,T.,et al., *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications*. Proc Natl Acad Sci U S A, 2001. **98**(19): p.10869-74.
- [3] Van de Vijver, M.J.,et al., *A gene expression signature as a predictor of survival in breast cancer*. N Engl J Med, 2002. **347**(25):p. 1999-2009.
- [4] M. Sezgin and B. Sankur (2004). "Survey over image thresholding techniques and quantitative performance evaluation". *Journal of Electronic Imaging* **13** (1): 146–165.
- [5] S. Kumar *et al.*, "A rule-based approach for robust clump splitting," *Pattern Recogn.*, vol. 39, no. 6, pp. 1088-1098, 2006.
- [6] X.C. He and N.H.C. Yung, "Corner detector based on global and local curvature properties", *Optical Engineering*, 47(5), pp: 057008, 2008.
- [7] An Improved Clump Splitting Method For Convex Objects. *Muhammad Farhan, Olli Yli-Harja, and Antti Niemistö*
- [8] S. Petushi, J. Zhang, A. Milutinovic, D.E. Breen, F.U. Garcia, "*Image-Based Histologic Grade Estimation Using Stochastic Geometry Analysis*," *Medical Imaging 2011: Computer-Aided Diagnosis*, SPIE Proceedings, Vol. 7963, paper 79633E, March 2011