

Nuclei Segmentation from Breast Cancer Histopathology Images

Amresh Vijay Nikam

PhD scholar, Shresh Gyan Vihar University, Jaipur

Dr. Arpita Gopal

Director, Sinhgad Institute of Business Administration and Research, Pune

Abstract: Biopsy is one of the available techniques for the garneted conformation of breast cancer. Histology image analysis using computer-aided diagnosis systems has become increasingly important during the last years. One reason is the need to alleviate the heavy workload of medical experts. Segmentation of cells and nuclei is an important first step towards automatic analysis of digitized microscopy images. In this paper, we introduce a general purpose framework which is able to segmentation of cell nuclei from microscopic images based on region growing segmentation.

Keywords: Scale-space analysis, region growing segmentation

1. INTRODUCTION

Extraction and classification of cells in histology images is an important problem that has many applications in practical medicine. The availability of high resolution multispectral multimodal imaging of tissue biopsies provides a new opportunity to develop improved tissue segmentation algorithms for developing computer-aided diagnostic classification of histological images in a clinical setting. The problem of image segmentation is not a new one. Many papers have been written on the subject but the difficulty lies in finding the appropriate algorithm for a specific application.

Image segmentation extracts objects of interest from the background; these objects are the focus for further cancer identification and classification. Early segmentation methods (which are still used) include thresholding, edge detection, and region growing [1] [2] [3]. Edge detection applies spatial filters (e.g. Canny and Sobel filters) to analyze neighboring pixel intensity or gradient differences to determine the border among objects and background. More advanced algorithms have been proposed for better performance, including active contours [4] [5] [6] [7] and clustering-based techniques [8] [9]. For clustering-based methods, when training samples are available, supervised algorithms can be applied to build the classifier, which include artificial neural network, boosting approaches [10], support vector machine (SVM), decision trees, and Bayesian model-based approaches such as Markov random field (MRF), hidden Markov model (HMM), and conditional random field (CRF). Without a set of labeled samples, unsupervised techniques, such as K-means, fuzzy c-means, ISODATA clustering, self-organizing map [11], and adaptive resonance theory [12], have to be used to group image points to different objects.

In this research we have create a framework for segmentation of cell nuclei from histology images. For primary segmentation we have used the region growing

segmentation method. Region growing is a useful, simple and sophisticated method for image segmentation. Many algorithms have been proposed based on region growing [13] which uses one or several uniformity criteria to measure the characteristics of pixel and its neighborhood. There are several criteria for the uniformity depending on the image properties which need to be emphasized. These include gray level uniformity, statistical and/or structure feature uniformity, and texture feature uniformity. The basic idea of region growing is that neighboring pixels with the same or similar uniformity values are clustered into the same region. The paper is organized as follows: section 2 describes the proposed framework in detail, section 3 evaluation, section 4 results and section 5 conclusions.

2. PROPOSED FRAMEWORK

The following figure 1 show the overall process flow of proposed framework. The proposed framework divided into two parts, seed detection and seed region growing. Seeded region growing (SRG) [12] is robust, rapid and free of tuning parameters. These characteristics allow implementation of a very good algorithm which could be applied to large variety of images. SRG is also very attractive for semantic image segmentation by involving the high-level knowledge of image components in the seed selection procedure. However, the SRG algorithm also suffers from the problems of automatic seed generation and pixel sorting orders for labeling [13].

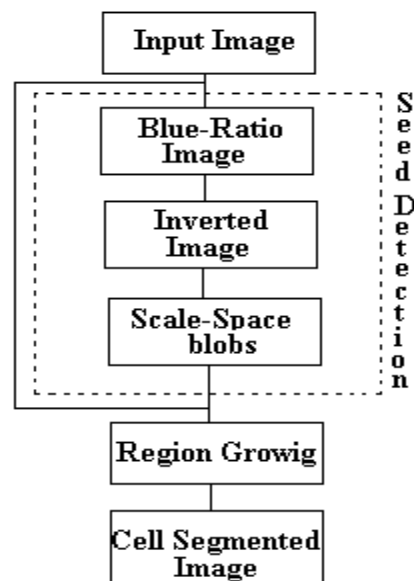


Figure1: Proposed framework

To segment the cells from histology images, first we want to find the points which are inside the cells. Once we get, that points are used as seed for the SRG algorithm.

2.1 Seed detection and radius finding

For seed detection we used the following procedure.

2.1.1 Conversion from RGB to Blue Ratio image:

Histopathology image consists of three things background which is white like color, extra cellular which is pink like, and cytoplasm and nuclei which is blue - purple like color. To reduce the complexities for integrating LoG responses, the RGB images are transformed to accentuate the nuclear dye [14].

$$BR = \frac{100 \times B}{1 + R + G} \times \frac{256}{1 + B + R + G}$$

Where B, R and G are blue, red and green channel of RGB, respectively.

2.1.2 Invert Image

After transforming the RGB image into BR image the cell region becomes whitish color and background and extra cellular region becomes blackish color. To get the LOG response we invert the image. We used the following formula to invert the BR image.

$$\text{scaledRangeMin} + ((\text{elementToScale} - \text{rangeMin}) \times (\text{scaledRangeMax} - \text{scaledRangeMin}) / (\text{rangeMax} - \text{rangeMin}))$$

Where we used the value for scaledRangeMin = 255, scaledRangeMax = 0 and rangeMin, rangeMax is the min and max value of BR image.

2.1.3 Creation of Scale-Space

Iterative Gaussian blurring is used to generate a scale-space representation of the inverted BR image. Such a representation allows us to examine the given image using increasing aperture sizes, thereby facilitating the detection and processing of coarse to fine features under the same framework. However, looking at an image from arbitrary apertures of the same order of magnitude is often unproductive. Hence, we step along the scale axis in a logarithmic fashion, and also work within given lower and upper bounds for the scale parameter. Thus, for a given scale (t), the corresponding Gaussian kernel will be of the form:

$$\sigma = e^t$$

For each σ , we compute the size of the discrete isotropic 2D Gaussian kernel by the heuristic $\text{ceil}(3\sigma) * 2 + 1$. The inverted BR image is convolved with each Gaussian, and the results are stored.

2.1.4 Laplacian Filtering

In order to detect cell in created scale-space, we filter each stored image using a 2D discrete Laplacian filter. This is the optimal filter for detecting symmetric blob-like shapes in an image that has been convolved with a Gaussian. Due to properties of convolution, we can instead take the Laplacian of a Gaussian, and convolve that with the inverted BR image.

Therefore we can convolve the inverted BR image with a Laplacian of Gaussian filter to obtain a scale-space where high responses denote the location of blobs of a certain size. In order to determine the location and size of cells, we must find local maxima in the spatial domain as well as in the scale dimension. We define a threshold, where every pixel

with gray value larger than the threshold is considered a local maximum. Furthermore, we restrict maxima to be larger than their 8 immediate neighbors. Using this scheme, we obtain maxima for every scale.

One possible way to suffering this problem is to require local maxima to show a higher response than its immediate neighbors in scale. The problem with this method is Gaussian blurring does not respect the location of features; it moves edges. Therefore, we must consider some neighborhood as we compare different scales. It is not immediately clear what size neighborhood is optimal. Too small a neighborhood and we are left with overlapping regions. Too large a neighborhood would force the selection of a single blob when multiple small blobs exist close to each other.

To determine the accurate maxima in scale we model the blobs as circles with the center located at the local maxima with radius 1.5σ . Conceptually, we consider a list of circles of varying center location size. We iterate over all circles, computing the area of overlap between it and every other circle. If the area of overlap is larger than a threshold (0.25 is reasonable), then we keep the circle generated by a higher response in scale and discard the other. When the process is finished, we are left with maxima in space and in scale.

2.2 Region growing

The basic formulation for Region-Based Segmentation is [15]:

- $\bigcup_{i=1}^n R_i = I$
- R_i is a connected region, $i = 1, 2 \dots$
- $R_i \cap R_j = \Phi$ for all $i = 1, 2 \dots$
- $P(R_i) = TRUE$ for $i = 1, 2 \dots$
- $P(R_i \cup R_j) = FALSE$ for any adjacent region R_i and R_j

Where $P(R_i)$ represents the homogeneity criterion, which is based on the feature values established for the segmentation purpose over the region R. Also the region R_i and R_j are considered to be adjacent if a pixel belong to R_i is a neighbor of some pixel of R_j and vice versa.

Property (a) ensures that segmentation must be complete; that is, every pixel must be in a region. Second property (b) requires that points in a region must be connected in some predefined sense. Third (c) indicates that the regions must be disjoint. Property (d) ensures that the region satisfies the homogeneity criterion defined by the user. And (e) indicates that region R_i and R_j are different in the sense of predicate P . Successes of cell segmentation using region growing algorithm depends on initial seed selection and criteria used to terminate the recursive region grow process. Hence choosing appropriate criteria is the key in extracting the cell region. In general these criteria include region homogeneity, object contrast with respective background, strength of the region boundary, size, and conformity to desire texture feature like texture, size, and color.

We used criteria mainly based on region homogeneity and region aggregation using intensity values and their gradient direction and magnitude. This criterion is characterized by a cost function which exploits certain features of images around the seed [16]. These cost function are verified for their match with the specified conditions of homogeneity criteria by comparing their value to be less than 1. If there is

a match then pixel under consideration is added to the growing region otherwise excluded from consideration.

We also add one more criteria the pixel distance should not exceed the radius which calculate in seed finding procedure.

Following list show the criteria we used.

1. x and y must be within image size. Where x and y is the position of the pixel which we want to include or exclude from the region
2. Pixel at x, y position is not already examined.
3. Distance of x, y from seed pixel is not greater than $1.5 \times \text{radius}$.
4. Difference between mean of the pixels values of included region and current pixel must be less than TI.
5. Pixel at x, y color intensity must be less than TC.

We have implemented the 2D seeded region grow algorithm using queue data structure. In our implementation we considered 8-neighbours while growing the region.. Similar pseudo code for our implementation is as follows:

```

Initialize the Queue
For each seed location
Insert seed location to queue
While (queue not empty)
Delete location
If above mention criteria matches for location's
neighbour pixel
Mark location's neighbour as Region
Inset neighbour into queue

```

3. Evaluation

The automatic segmentations were compared with the manual segmentations obtained with systematic random sampling in the following way: if a manual segmentation was not intersected by an automatic segmentation with a Dice coefficient or Sørensen index of at least 0.25, it was counted as a false negative (FN). Otherwise, it was counted as a true positive (TP). The Dice coefficient was taken as a measure of quality of the segmentation. The Dice coefficient is a measure of overlap between two regions, commonly used for evaluation of segmentation techniques. It is defined as:

$$D(x, y) = 2 \frac{|x \cap y|}{|x| + |y|}$$

Where x and y are the number of species in samples x and y, respectively.

The quotient of similarity $D(x, y)$ ranges from 0 to 1. We used a cut-off value of 0.25 was to avoid unsegmented nuclei that are "touched" by a neighbouring segmentation to be counted as TP.

To estimate the positive predictive value a subset of 40 automatically segmented nuclei from each slide was randomly generated. An expert labelled all segmentations that did not correspond to epithelial nuclei, such as stroma, lymphocytes, "junk" particles etc.

For each representative region the sensitivity, positive predictive value and the median Dice coefficient were estimated. We refer to the sensitivity, positive predictive value and median Dice coefficient measures as estimates because they are based on an annotated subset of the entire population of nuclei in the images. Because of the

asymmetric left-skewed distribution, the median of the Dice coefficient is a better measure of central tendency than the mean. To test our algorithm we used the parameter values as below:

For seed and radius detection:

Scale space: 2.0 to 3.5 with 0.2 increment

Finding the maxima: 0.25 threshold

Overlap blob detection: 0.25 threshold

For region growing:

Min pixel difference (TI): 10 threshold

Colour intensity (TC): 10 threshold

4. Result

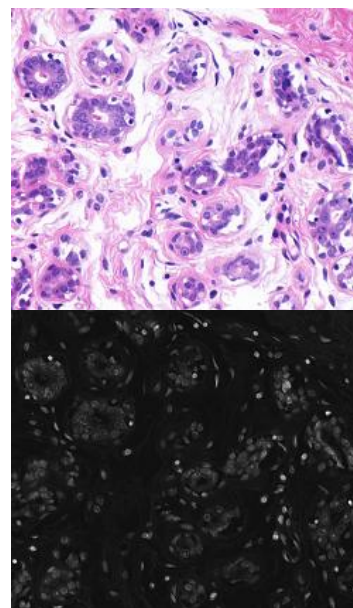
Figure 2 shows the segmentation results for a few regions from our dataset. The sensitivity, positive predictive value and median Dice coefficient for our datasets are summarized in Figure 3.

The sensitivity was estimated as the percentage of manual segmentations that were matched to an automatic segmentation. The positive predictive value was estimated as the percentage of the annotated automatic segmentations marked as corresponding to an epithelial nucleus. The mean estimated sensitivity was 0.912 (± 0.087). The mean estimated positive predictive value was 0.937 (± 0.082). The distribution of the estimated Dice coefficients had a high peak around 0.93, with the vast majority of segmentations having values larger than 0.84.

6. Conclusion

We have presented an accurate technique for automated segmentation of nuclei in images derived from digital slides of H&E stained breast cancer sections. The evaluation revealed that the proposed method has good performance in both detection and segmentation accuracy.

The obtained binary images can be used for measurements of cell characteristic for scientific and diagnostic purposes.



A) Original

B) Blue ratio

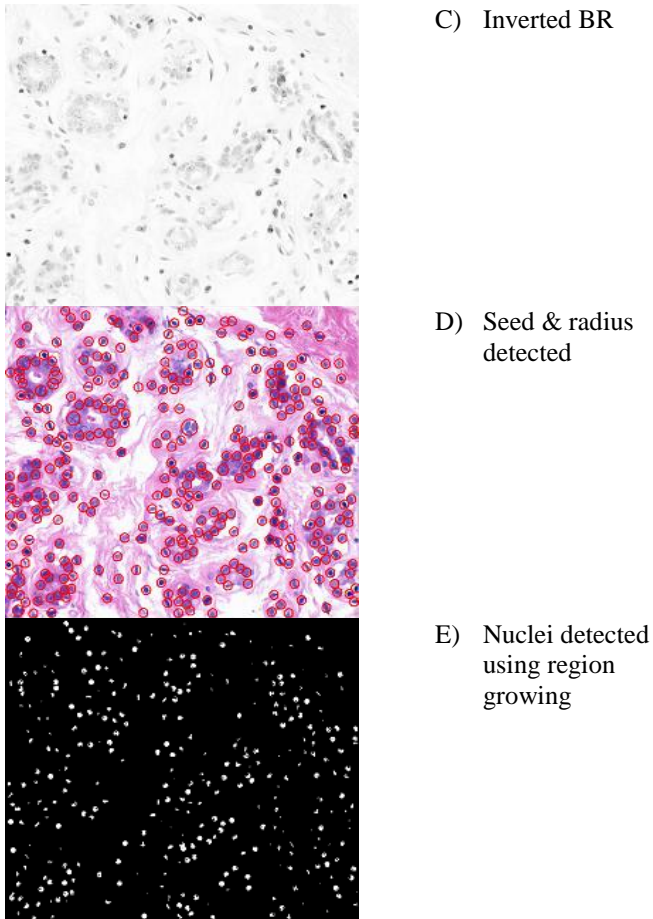


Figure 2. Segmentation of nuclei from H&E breast cancer image

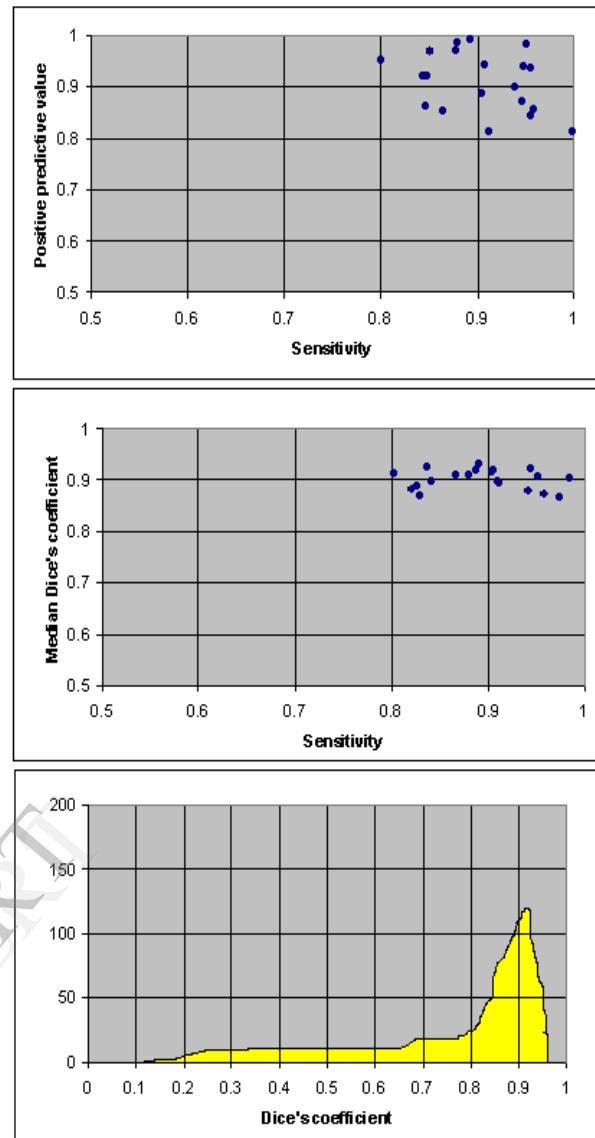


Figure 3. Performance measurement

REFERENCES

- [1] Gonzalez, R. C. & Woods, R. E. (2008). *Digital Image Processing*. Pearson Prentice Hall, 3rd Ed.
- [2] Sezgin, M. & Sankur, B. (2003). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1), 146–165.
- [3] Loukas, C. G. & Linney, A. (2004). A survey on histological image analysis-based assessment of three major biological factors influencing radiotherapy: proliferation, hypoxia and vasculature. *Computer Methods and Programs in Biomedicine*, 74(3), 183–199.
- [4] Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: active contour models. *International Journal of Computer Vision*, 1(4), 321–331.
- [5] Malladi, R., Sethian, J., & Vemuri, B. (1995). Shape modeling with front propagation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17(2), 158–171.
- [6] Caselles, V., Kimmel, R., & Sapiro, G. (1997). Geodesic active contours. *International Journal of Computer Vision*, 22(1), 61–79.
- [7] Vese, L. A. & Chan, T. F. (2002). A multiphase level set framework for image segmentation using the Mumford and Shah model. *International Journal of Computer Vision*, 50(3), 217–293.
- [8] Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern Classification*. Wiley, 2nd Ed.
- [9] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- [10] Doyle, S., Madabhushi, A., Feldman, M., & Tomaszewski, J. (2006). A boosting cascade for automated detection of prostate cancer from digitized histology. In *13th International Conference on Medical Image Computing and Computer Assisted Intervention* (pp. 504–511).
- [11] R. M. Haralic and L. G. Shapiro (1985), *Image Segmentation Techniques*, Computer Vision, Graphics and Image Processing 29, pp 100-132.
- [12] Adams, R., Bischof, L., 1994. Seeded region growing. *IEEE Trans. Pattern Anal. Machine Intell.* 16, 641–647.
- [13] Mehnert, A., Jackway, P., 1997. An improved seeded region growing algorithm. *Pattern Recognition Lett.* 18, 1065–1071.
- [14] Chang H, Loss LA, Parvin B. Nuclear segmentation in H and E sections via multi-reference graph-cut (MRGC). *International Symposium Biomedical Imaging*; 2012.
- [15] B.Chanda, D.Dutta Majumder. *Digital Image Processing and Analysis*, Prentice Hall of India, New Delhi, 2000.
- [16] Runzhen Huang Kwan-Liu Ma. *RGVis: Region Growing Based Techniques for Volume Visualization*.

IJERT