

July 2013

## Automated Nuclei Segmentation of Breast Cancer Histopathology

Vipin Bondre

*Deptt. of Electronics Engineering, Yeshwantrao Chavan College of Engineering, Nagpur,India,*  
Vipin.bondre@gmail.com

Amoli Belsare

*Deptt. of Electronics Engineering, Yeshwantrao Chavan College of Engineering, Nagpur,India,*  
amolibel@rediffmail.com

Follow this and additional works at: <https://www.interscience.in/ijcct>

---

### Recommended Citation

Bondre, Vipin and Belsare, Amoli (2013) "Automated Nuclei Segmentation of Breast Cancer Histopathology," *International Journal of Computer and Communication Technology*. Vol. 4 : Iss. 3 , Article 4.

Available at: <https://www.interscience.in/ijcct/vol4/iss3/4>

This Article is brought to you for free and open access by Interscience Research Network. It has been accepted for inclusion in International Journal of Computer and Communication Technology by an authorized editor of Interscience Research Network. For more information, please contact [sritampatnaik@gmail.com](mailto:sritampatnaik@gmail.com).



# Automated Nuclei Segmentation of Breast Cancer Histopathology



Vipin Bondre & Amoli Belsare

Deptt. of Electronics Engineering, Yeshwantrao Chavan College of Engineering, Nagpur, India

E-mail : [Vipin.bondre@gmail.com](mailto:Vipin.bondre@gmail.com), [amolibel@rediffmail.com](mailto:amolibel@rediffmail.com)

**Abstract** - Automated detection and segmentation of cell nuclei is an essential step in breast cancer histopathology, so that there is improved accuracy, speed, level of automation and adaptability to new application. The goal of this paper is to develop efficient and accurate algorithms for detecting and segmenting cell nuclei in 2-D histological images. In this paper we will implement the utility of our nuclear segmentation algorithm in accurate extraction of nuclear features for automated grading of (a) breast cancer, and (b) distinguishing between cancerous and benign breast histology specimens. In order to address the issue the scheme integrates image information across three different scales: (1) low level information based on pixel values, (2) high-level information based on relationships between pixels for object detection, and (3) domain-specific information based on relationships between histological structures. Low-level information is utilized by a Bayesian Classifier to generate likelihood that each pixel belongs to an object of interest. High-level information is extracted in two ways: (i) by a level-set algorithm, where a contour is evolved in the likelihood scenes generated by the Bayesian classifier to identify object boundaries, and (ii) by a template matching algorithm, where shape models are used to identify glands and nuclei from the low-level likelihood scenes. Structural constraints are imposed via domain specific knowledge in order to verify whether the detected objects do indeed belong to structures of interest. The efficiency of our segmentation algorithm is evaluated by comparing breast cancer grading and benign vs. cancer discrimination accuracies with corresponding accuracies obtained via manual detection and segmentation of glands and nuclei.

**Keywords** - Image Histology, Breast cancer, Segmentation, Detection, Grading.

## I. INTRODUCTION

Breast cancer (BC) is the most common type of cancer in women with an estimated lifetime incidence of 1 in 8 in 2008 [1]. Definitive diagnosis Of BC is performed by a pathologist via examination of tissue histopathology typically obtained via a needle biopsy. Pathologists identify specific visual patterns and distinctive phenotypic changes that are characteristic of BC. Certain kinds of phenotypic changes in tissue pathology, such as the presence of lymphocytic infiltration (LI), may be related to patient survival and outcome [2]. Specifically, LI has been correlated with nodal metastasis and recurrence in tumors expressing the human epidermal growth factor receptor-2 (HER2) protein. Precise quantification of the extent of LI on BC histopathology imagery could be significant in predicting outcome and prescribing appropriate therapy [2]

Automated segmentation of cell nuclei is now a well-studied topic for which a large number of algorithms have been described in the literature [2]–[18], and newer methods continue to be investigated. The main challenges in segmenting nuclei in histological, especially pathological tissue specimens result from the fact that the specimen is a 2-D section of a 3-D tissue sample. The 2-D sectioning can result in partially imaged nuclei, sectioning of nuclei at odd angles, and damage due to the sectioning process. Furthermore, sections have finite thickness resulting in overlapping or partially

superposed cells and nuclei in planar images. The end result of these limitations is a set of image objects that differ considerably from the ideal of round blob-like shapes.

In the interest of obtaining a fully automated grading scheme, it is imperative to be able to first automatically identify and segment histological structures. Nuclei segmentation has been attempted using basic fuzzy c-means clustering [4] and adaptive thresholding [5].

Other thresholding algorithms were investigated by Korde, et al. [6] for bladder and skin cell nuclear segmentation. However, thresholding leads to poor results when there is large variability in the histology staining. Other algorithms have been proposed using more complex techniques, such as an active contour scheme for pap stained cervical cell images [7]. These techniques lead to successful results only when nuclei are non-overlapping.

In this work, we present an automated nuclei segmentation scheme for breast histopathology which utilizes a combination of low-level, high-level, and domain-specific information. A Bayesian classifier is used to generate likelihood scenes of structures of interest in the image based on image intensity and textural information. These scenes are combined with domain knowledge regarding arrangement of histological structures through which structural constraints are

imposed. A level set algorithm and template matching scheme are then used for nuclear segmentation. Nuclear centroids are used to extract architectural features via three different graph algorithms including Voronoi, Delaunay triangulation, and minimum spanning tree (MST). Our segmentation algorithm is employed for two different applications: (a) discriminating cancer from non-cancer in breast histology images, and (b) distinguishing Bloom-Richardson low grade (5, 6) breast cancer from high grade (7, 8) breast cancer. Evaluation of the algorithm is done by comparing the classification accuracies obtained for breast cancer grading and benign versus cancer discrimination with corresponding classification accuracies obtained via manual detection and segmentation of nuclei.

The rest of this paper is organized as follows. Our segmentation algorithm is described in Section 2, and the evaluation results of the algorithm in two different applications (breast cancer grading and benign versus breast cancer discrimination) are presented in Section 3. Concluding remarks presented in Section 4.

## II. INTEGRATED LOW-HIGH LEVEL AND CONTEXTUAL SEGMENTATION MODEL

### System Overview

A flowchart describing our segmentation algorithm and its applications in breast cancer detection and grading is shown in Figure 1.

#### Low-Level Information

We denote a tissue region by a digital image  $C = (C, f)$  where  $C$  is a 2D grid of image pixels  $c \in C$  and  $f$  is a function that assigns a pixel value (representing the red, blue, and green channels of the RGB space) to  $c$ . A training set  $S_v$  of pixels representing a structure of interest  $v$  (lumen ( $L$ ), cytoplasm ( $S$ ), nuclei ( $N$ ), etc.) is obtained. The color values  $f(c)$  of pixels  $c \in S_v$  are used to generate probability density functions  $p(c, f(c)|v)$ , where  $v$  represents the image class. For each image  $C$ , Bayes Theorem is used to obtain a pixel-wise likelihood for each pixel  $c \in C$ , where  $P(v|c, f(c))$  is the probability that  $c$  belongs to class  $v$  given image pixel value  $f(c)$  and obtained as,

$$P(v|c, f(c)) = \frac{P(v)p(c, f(c)|v)}{\sum_{u \in \{L, N, S\}} P(u)p(c, f(c)|u)} \quad (1)$$

where  $v \in \{L, N, S\}$ ,  $p(c, f(c)|v)$  is the *a priori* conditional probability obtained during training for class  $v$ , and  $P(v)$  is the prior probability of occurrence for each class  $v$  (assumed as non informative priors). These pixel-wise likelihoods generate a likelihood scene  $L_v$ , where the intensity in the likelihood image is the probability of pixel  $c$  belonging to class  $v$ .

#### High-Level Information

We consider two shape based approaches for segmentation including:

(a) Level set and (b) Template matching.

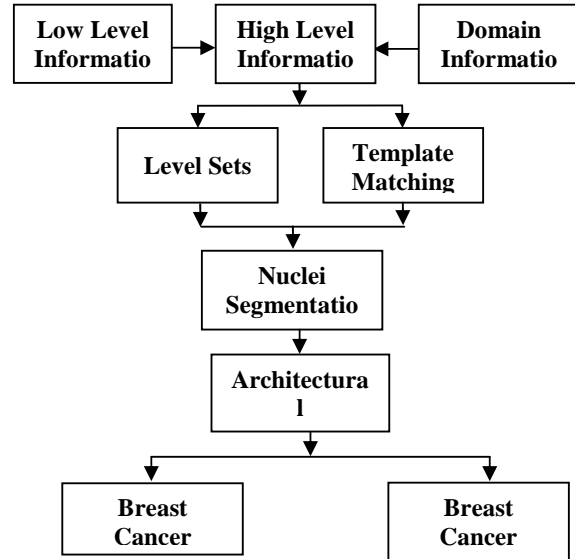


Fig. 1. Flowchart illustrating the system overview.

#### Level set Segmentation

Boundary segmentation is performed using level-sets which makes use of neighbouring pixels during evolution of the contour to find the target boundary. A boundary  $B$  evolving in time  $t$  and in the 2D space defined by the grid of pixels  $C$  is represented by the zero level set  $B = \{(x, y) | \phi(t, x, y) = 0\}$  of a level set function  $\phi$ , where  $c = (x, y)$ . The evolution of  $\phi$  is then described by a level-set formulation adopted from [8]:

$$\frac{\partial \phi}{\partial t} + F|\nabla \phi| = 0 \quad (2)$$

where the function  $F$  defines the speed of the evolution. The curve evolution is driven by the likelihood image  $L_v$ , where  $v \in \{L, N, S\}$ . The initial contour  $\phi_0 = \phi(0, x, y)$  is automatically initialized using low-level information via Bayesian classifier. The algorithm is run until the difference in the contours of one iteration to the next is below an empirically determined threshold.

#### Template Matching

Template matching [9] is done on a binary image ( $IB$ ) converted from the likelihood scene  $L_v$ . Correlation between the selected template and  $IB$  is computed at

each of the pixels  $c \in C$ . The choice of template is motivated by the size and shapes of the structure of interest. For our application to detection of nuclei; we have chosen four binary elliptical templates with different major and minor axes.

### Incorporating Domain Specific Constraints

Our segmentation algorithm exploits domain-specific information in the form of specific arrangement and

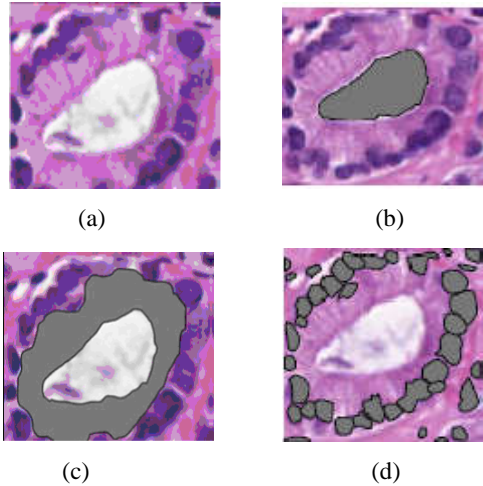


Fig. 2. Illustration of different regions of interest (a) Shown outlined in black are the (b) lumen area, (c) cytoplasm, and (d) nuclei.

relationships between histological structures. For instance, (Fig.2) comprises three main structures of interest: lumen, cytoplasm, and nuclei. The Structures are arranged in a specific fashion (lumen is surrounded by cytoplasm, which is surrounded by a ring of nuclei). We exploit the fact that a lumen area needs to be surrounded by cytoplasm, which is then surrounded by a ring of nuclei.

### E Incorporating Domain Specific Constraints

#### 1. Low-level Information

Step 1: Bayesian classification of nuclei: Low-level image intensity information is used to generate likelihood scene  $L_N$ .

Step 2: Convert  $L_N$  to binary image:  $L_N$  is thresholded to a binary image  $I_B = (C, h)$ , where  $h(c) \in \{0, 1\}$ . Let  $B$  denote the set of pixels in the background, so that  $h(b) = 0$  for all  $b \in B$  and  $h(c) = 1$  for all  $c \in C, c \notin B$ .

#### 2. Structural Constraints

Step 3: Euclidean distance transform (EDT): EDT operation [10] is applied on  $I_B$ , which transforms into a grey level image  $DB = (C, f_B)$  where for  $c \in C, f_B(c)$  is

the Euclidean distance to the nearest  $b \in B$ . The EDT for all pixels  $c \in C$  is defined as:

$$f_B(c) = \min_{b \in B} [d(c, b)] \quad (3)$$

where  $d(c, b)$  is the Euclidean distance between pixels  $c, b \in C$ . By limiting the template matching to only those pixels  $c \in C$  for which  $f_B(c) > \lambda$ , where  $\lambda$  is some pre-defined threshold, significantly increases the computational speed and efficiency of our algorithm.

#### 3. High-level Information

Step 4: Template matching to detect nuclei: For each  $c \in DB$  for which  $f_B > \lambda$ , template matching with 4 templates  $T_1, T_2, T_3, T_4$  is done. The idea is to focus on pixels  $c \in DB$  for which  $f_B$  is high since these are the points that lie close to medial axis of object of interest. At  $c \in C$ , where  $f_B(c) > \lambda, \max_j [corr(c, T_j)]$  is computed where  $corr(c, T_j)$  is correlation obtained by placing centroid of  $T_j$  on  $c$ , for  $j \in \{1, 2, 3, 4\}$ . Thus we obtain a new scene  $\mathcal{P}^j = (C, f^j)$  where for each  $c \in C, f^j(c) = \max_j [corr(c, T_j)]$ . Nuclear centroids are determined as those for which  $f^j(c) > \delta$ .

## III. BREAST CANCER GRADING

### A. Feature Extraction

Following gland and nuclear segmentation, we calculate 8 boundary features from the interior nuclear and lumen boundaries, giving a total of 16 morphological features which quantify the size and shape of the glands [11]. The features were chosen based on features used by pathologists in detecting and grading cancer in practice. The goal of our segmentation algorithm is high classification accuracy in distinguishing cancer grades and cancer detection, so to compare our algorithm with manual segmentations; we extract each set of features twice using manual and automatic segmentation. These two feature sets were used in a support vector machine classifier to compare manual and automated segmentations in terms of classifier accuracy. Experiments using our algorithm are described in the next three sub-sections.

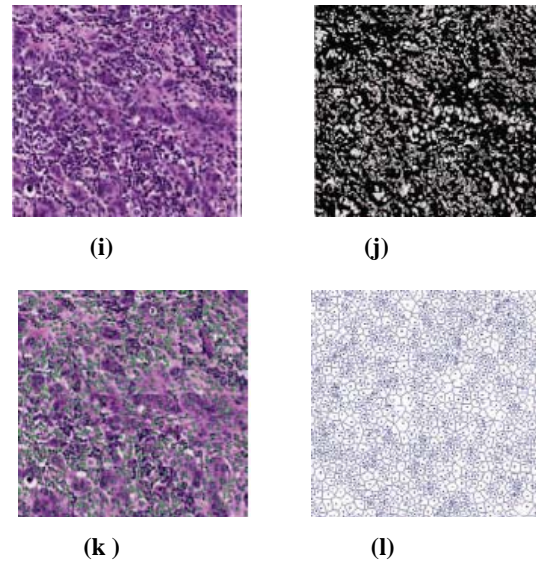
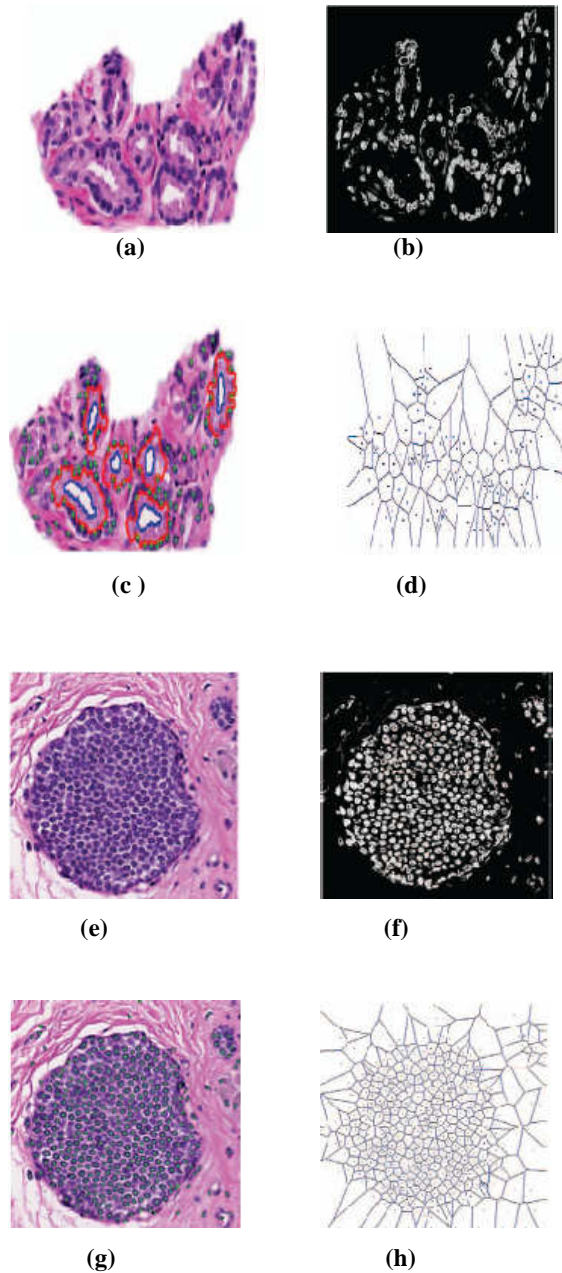
### B. Application 1: Breast Cancer Detection

We discriminate cancer from non-cancer in breast histology images using architectural features. The dataset contains a total of 18 benign images and 36 cancer images. The feature set is reduced using principal component analysis (PCA) and a SVM classifier is used for classification. We obtain an accuracy of 81.91% using automatic segmentation and 77.10% with manually segmented structures (Table1), indicating comparable performance between manual and automatic segmentation. A sample cancer image is shown in Fig. 3(f).

### Application 2: Breast Cancer Grading



We seek to distinguish 21 images of low-grade breast cancer (grades 5 and 6 on the Bloom-Richardson scale) from 9 images of high grade cancer (7 and 8). As in the prostate cancer grading example, graph-based features are used to classify high grade vs. low grade cancer. The feature set is reduced using PCA and classification is done using a SVM classifier. The classification accuracies are shown in Table 1



- Fig. 3. Tissue images corresponding to (a) Gleason grade 3 cancers, (e) Bloom-Richardson grade 5 and (i) grade 8 cancer  $L_N$  for (a), (e), (i) are shown in (b), (f), and (g). Segmented gland boundaries and nuclei centroids are shown in (c), (g), and (k). Voronoi diagrams used to generate architectural features are shown in (d), (h), and (l).

And figures as those for Application A are shown in Fig.3 (i-l) for a tissue image with a grade 8 breast cancer.

TABLE I  
SVM CLASSIFICATION ACCURACY FOR THE  
THREE APPLICATIONS USING THE  
AUTOMATICALLY AND MANUALLY  
EXTRACTED FEATURE SETS.

<i>Tissue Type</i>	<i>Task</i>	<i>Automated</i>	<i>Manual</i>
Breast	Cancer vs. Non-Cancer	81.91%	77.10%
Breast	High vs. Low Grade	80.52%	93.33%

#### IV. CONCLUDING REMARKS

In this paper, we have demonstrated an integrated nuclear segmentation and detection scheme. The strength of the model is derived from the fact that it incorporates low-, high-level knowledge, and structural constraints imposed via domain knowledge. The architectural attributes derived from the segmented nuclei and glands were used for (a) grading of prostate cancer, (b) discriminating cancer from non-cancer in breast histology, and (c) discriminating low grade from high

grade breast cancer. The corresponding classification accuracies obtained for case (a) were 95.19% for grade 3 vs. grade 4, 86.35% for grade 3 vs. benign, and 92.90% for grade 4 vs. benign; for case (b) was 81.91% and for case (c) was 80.52%. These accuracies compare favourably with the corresponding results obtained via manual segmentation. Future work will focus on evaluating our methods on a larger cohort of images.

## REFERENCES

- [1] H.J.G. Bloom and W.W. Richardson, "Histological grading and prognosis in breast cancer," *British Journal of Cancer*, vol. 11, pp. 359–377, 1957.
- [2] R. Montironi, R. Mazzuccheli, et al., "Gleason grading of prostate cancer in needle biopsies or radical prostatectomy specimens: contemporary approach, current clinical significance and sources of pathology discrepancies," *BJU Int.*, vol. 95, no. 8, pp. 1146–1152, 2005.
- [3] S. Doyle, M. Hwang, K. Shah, A. Madabhushi, et al., "Automated grading of prostate cancer using architectural and textural image features," *IEEE ISBI*, pp. 1284–1287, 2007.
- [4] L. Latson, B. Sebek, and K. Powell, "Automated cell nuclear segmentation in color images of hematoxylin and eosin-stained breast biopsy," *Analytical and Quantitative Cytology and Histology*, vol. 25, no. 6, pp. 321–331, 2003.
- [5] S. Petushi, F.U. Garcia, M.M. Haber, et al., "Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer," *BMC Medical Imaging*, vol. 6, no. 14, October 2006.
- [6] V.R. Korde, H. Bartels, et al., "Automatic segmentation of cell nuclei in bladder and skin tissue for karyometric analysis," *Biophotonics 2007: Optics in Life Science*, vol. 6633, pp. 6633V, 2007.
- [7] P. Bamford and B. Lovell, "Unsupervised cell nucleus segmentation with active contours," *Signal Processing*, vol. 71, pp. 203–213, 1998.
- [8] C. Li, C. Xu, C. Gui, and M.D. Fox, "Level set evolution without re-initialization: a new variational formulation," *IEEE CVPR*, vol. 1, pp. 430–436, 2005.
- [9] I. Sintorn, M. Homman-Loudiyi, et al., "A refined circular template matching method for classification of human cytomegalovirus capsids in TEM images," *Computer Methods and Programs in Biomedicine*, vol. 76, no. 2, pp. 95–102, 2004.
- [10] A. Datta and S. Soundaralakshmi, "Fast and scalable algorithms for the Euclidean distance transform on a linear array with a reconfigurable pipelined bus system," *Journal of Parallel and Distributed Computing*, vol. 64, pp. 360–369, 2004.
- [11] S. Naik, S. Doyle, A. Madabhushi, et al., "Gland segmentation and computerized gleason grading of prostate histology by integrating low-, high-level and domain specific information," *MIAAB Workshop*, 2007.

