

# Expectation–Maximization-Driven Geodesic Active Contour With Overlap Resolution (EMaGACOR): Application to Lymphocyte Segmentation on Breast Cancer Histopathology

Hussain Fatakdawala, Jun Xu, Ajay Basavanahally, Gyan Bhanot, Shridar Ganesan, Michael Feldman, John E. Tomaszewski, and Anant Madabhushi

**Abstract**—The presence of lymphocytic infiltration (LI) has been correlated with nodal metastasis and tumor recurrence in HER2+ breast cancer (BC). The ability to automatically detect and quantify extent of LI on histopathology imagery could potentially result in the development of an image based prognostic tool for human epidermal growth factor receptor-2 (HER2+) BC patients. Lymphocyte segmentation in hematoxylin and eosin (H&E) stained BC histopathology images is complicated by the similarity in appearance between lymphocyte nuclei and other structures (e.g., cancer nuclei) in the image. Additional challenges include biological variability, histological artifacts, and high prevalence of overlapping objects. Although active contours are widely employed in image segmentation, they are limited in their ability to segment overlapping objects and are sensitive to initialization. In this paper, we present a new segmentation scheme, expectation–maximization (EM) driven geodesic active contour with overlap resolution (EMaGACOR), which we apply to automatically detecting and segmenting lymphocytes on HER2+ BC histopathology images. EMaGACOR utilizes the expectation–maximization algorithm for automatically initializing a geodesic active contour (GAC) and includes a novel scheme based on heuristic splitting of contours via identification of high concavity points for resolving overlapping structures. EMaGACOR was evaluated on a total of 100 HER2+ breast biopsy histology images and was found to have a detection sensitivity of over 86% and a positive predictive value of over 64%. By comparison, the EMaGAC model (without overlap resolution) and GAC model yielded corresponding detection sensitivities of 42%

and 19%, respectively. Furthermore, EMaGACOR was able to correctly resolve over 90% of overlaps between intersecting lymphocytes. Hausdorff distance (HD) and mean absolute distance (MAD) for EMaGACOR were found to be 2.1 and 0.9 pixels, respectively, and significantly better compared to the corresponding performance of the EMaGAC and GAC models. EMaGACOR is an efficient, robust, reproducible, and accurate segmentation technique that could potentially be applied to other biomedical image analysis problems.

**Index Terms**—Breast cancer (BC), detection, expectation–maximization (EM), geodesic active contour (GAC), histopathology, lymphocytes, segmentation.

## I. INTRODUCTION

**B**REAST cancer (BC) is the most common type of cancer in women in the U.S. with an estimated life-time 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].

The visual detection of lymphocytes in BC histopathology images is complicated by similar appearing structures in the image such as cancer nuclei. Lymphocyte nuclei and cancer nuclei are often confused with one another during manual segmentation, which may adversely affect a human expert's ability to precisely determine the extent of LI. Hence, a clinician's ability to predict survival and disease outcome may be affected by inter- and intraobserver variability. Consequently, there exists a clear need for accurate automated detection of lymphocyte nuclei in BC histopathology images. However, an automated lymphocyte detection algorithm has to be able to deal with the variability in digital slide appearance due to inconsistencies in histological staining, and poor image quality with tissue samples due to slide digitization and tissue fixation. Furthermore, LI may be characterized by a high density of lymphocytes, which could

Manuscript received September 26, 2009; revised December 1, 2009; accepted December 29, 2009. Date of publication February 17, 2010; date of current version June 16, 2010. This work was supported by the Wallace H. Coulter Foundation, New Jersey Commission on Cancer Research, National Cancer Institute under Grant R01CA136535-01, Grant ARRA-NCI-3, Grant R21CA127186-1, Grant R21CA127186, Grant R03CA128081-01, and Grant R03CA143991-01, The Cancer Institute of New Jersey, Life Science Commercialization Award from Rutgers University, and the Aresty Undergraduate Research Grant. Asterisk indicates corresponding author.

H. Fatakdawala, J. Xu, A. Basavanahally, and A. Madabhushi are with the Department of Biomedical Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ 08854 USA (e-mail: hussainf@eden.rutgers.edu; junxu@rci.rutgers.edu; anantm@rci.rutgers.edu).

G. Bhanot and S. Ganesan are with The Cancer Institute of New Jersey, New Brunswick, NJ 08903 USA (e-mail: gyanbhanot@gmail.com; ganesash@umdnj.edu).

M. Feldman and J. E. Tomaszewski are with the Hospital of the University of Pennsylvania, Anatomic and Surgical Pathology Informatics, Philadelphia, PA 19104 USA (e-mail: feldmanm@mail.med.upenn.edu; john.tomaszewski@uphs.upenn.edu).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TBME.2010.2041232

cause significant overlap among lymphocyte nuclei and other structures in BC images. An accurate LI detection and segmentation algorithm would need to be able to identify overlapping lymphocytes as unique independent objects.

In this paper, we present a new segmentation scheme for automatic segmentation and detection of lymphocytes on BC histopathology images by avoiding the need for training datasets that are difficult to define for highly variable data and improving active contour segmentation via a specific initialization using the expectation–maximization (EM) algorithm and a novel contour-splitting scheme to resolve overlap between objects. The proposed scheme is generalizable and can be readily applied to other biomedical image segmentation problems.

## II. PREVIOUS WORK

Segmentation of structures in breast histopathology images has been attempted using fuzzy c-means clustering [3] and adaptive thresholding [4]. Thresholding tends to work only on uniform images and does not produce consistent results if there is considerable variability across image sets. In [5], a watershed algorithm was used to provide segmentation of BC cells based on the “topology” of the image derived from its gradient. Watershed algorithms, however, tend to oversegment the image and results can be highly sensitive to noise. Like thresholding-based methods, they are also usually sensitive to variations in staining. In addition, they require the existence of weak boundaries/edges to segment-overlapping objects. Manual detection and semiautomated segmentation routines were employed in [6] to first detect and segment nuclei and glands and use these results to subsequently distinguish low and high grades of BC using textural and nuclear architectural features. In [7], a region-growing method [8] was employed to obtain gland segmentation on digitized prostate whole mount histopathology images.

Active contours, which are widely used in image segmentation are deformable curves that can be used to delineate structures in an image using gradient (edge) information and global region-based information. The curve is initialized either using an implicit formulation or as a parametric representation and is allowed to deform toward object boundaries. A stopping function is designed to impede the movement of the curve beyond potential boundaries. The curve is deformed until it converges to the final segmentation. The geodesic active contour (GAC) model was introduced by Caselles *et al.* [9] and it allowed for computation of geodesics, or minimal distance curves, along with classical active contour model [10]. Significant advances in active contour models have been made since then [11], [12] in relation to improving segmentation and implementation. Numerous variants of active contour models have been recently proposed based on how the information from the image is used, the contour initialization scheme, contour representation, and convergence [13]–[17]. A major limitation of several active contour models, however, continues to be their inability to resolve multiple overlapping objects that are often segmented as a single object, and their sensitivity to initialization. In [13], a GAC model was provided a specific initialization using results from mathematical morphological operations, however, the

issue of overlapping/touching objects was not addressed. In [14], a novel active contour model was proposed based on a level-set formulation but the issue of a single contour enclosing multiple touching objects was not discussed. Li *et al.* [15] proposed the use of a watershed transform to obtain an initial segmentation and employed it to derive the final result using a region-based active contour model. Hafiane *et al.* [18] presented a model to detect nuclei on prostate histopathology using active contours initialized via results from fuzzy c-means clustering. The nuclei were then detected based on an iterative voting scheme using oriented kernels. More recently, probabilistic models have been employed to drive segmentation techniques [19], [20]. A Bayesian classifier in conjunction with template matching was used to segment nuclei in prostate and BC histopathology [19] and to distinguish different levels of LI in HER2+ BC [20]. A similar probabilistic model was used to improve segmentation by active contours in [21]. The ability to manually train these models is constrained by the availability of expert annotations of the objects of interest. Datasets for training are difficult to define due to variability across images. Furthermore, such models may not be generalizable and have limited application due to the manual training step.

Segmentation of overlapping structures in different types of images has been previously attempted. Clocksin [22] presented a method to segment overlapping nuclei in fluorescence *in situ* hybridization images by first extracting seed points via mean-shift clustering. These seed points were then employed to initialize a contour model that includes a repulsive term to avoid intersection of two contours. This method however, relies heavily on the ability of the mean-shift clustering algorithm in providing fairly accurate detection results. Nandy *et al.* [23] presented a novel method to extract boundaries of 4'-6-diamidino-2-phenylindole-stained nuclei in microscopy images by first identifying seed points and utilizing dynamic programming. Seed points are geometrically located via an intelligent scheme using automatic thresholding of the gradient image. However, the method assumes that the nuclear structures are fairly distinct from the background. Cloppet and Boucher [24] present a scheme for segmentation of overlapping nuclei in immunofluorescence images by providing a specific set of markers to the watershed algorithm. The markers are defined as splits between overlapping structures. He and Liao [25] segmented overlapping elliptically shaped nuclei by extracting contours via minimization of an energy function that measures homogeneity of regional intensity and includes a statistical shape model constraint. Fang and Chan [26] proposed a similar scheme where a statistical shape model is incorporated into the GAC model to segment occluded objects. These methods, however, require manual training to construct the shape model, are sensitive to initialization, and are potentially limited in their ability to segment multiple overlapping objects.

In this paper, we present a new segmentation scheme, EM-driven GAC with overlap resolution (EMaGACOR), which we apply to the problem of segmenting and detecting lymphocytes in BC histopathology images (see Fig. 1). Preliminary results of this work were presented in [27].

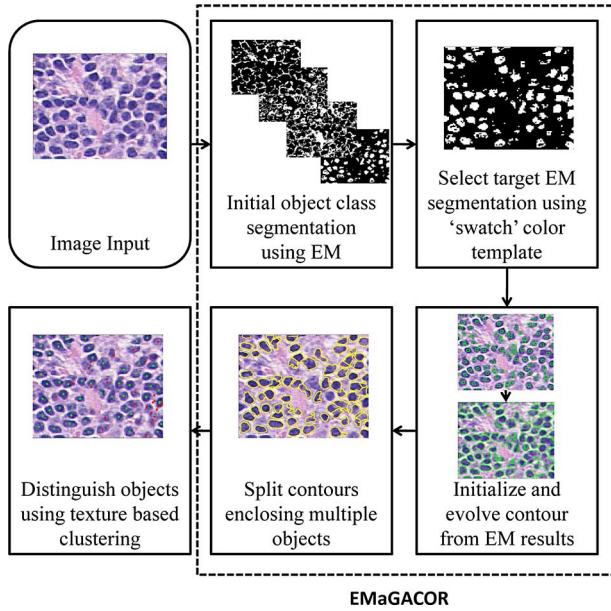


Fig. 1. Flowchart depicting our model for segmenting and detecting lymphocytes from HER2+ BC histopathology images.

EMaGACOR is able to overcome the drawbacks associated with probabilistic segmentation schemes, namely, the need for annotated training data and defining representative datasets. We attempt to avoid these issues by using an EM algorithm to initialize a GAC. The EM algorithm and Gaussian mixture models have been widely used in image segmentation problems [28]–[33]. Raja *et al.* [28] presented the use of a color mixture model to estimate the probability densities of object foreground and scene background colors for segmentation of multi-colored objects. Shamsi *et al.* [29] used a modified EM algorithm for facial skin segmentation. Zhang *et al.* [30] presented a segmentation model that incorporates spatial information into a mixture model by using the hidden Markov Random Field model fitted using the EM algorithm. Diplaros *et al.* [32] presented a graphical model and a modified EM algorithm for image segmentation. Additionally, mixture models and the EM algorithm have also been employed in content-based image retrieval problems [34], [35]. In this paper, we use the magnetostatic active contour (MAC) [17] as the GAC model. The EM algorithm effectively replaces the Bayesian framework and hence eliminates the need for representative datasets for training and reduces the effect of dataset variability on segmentation results. Initialization of the active contour using EM allows the model to focus on relevant objects of interest. The EM algorithm provides an initial segmentation in the form of a likelihood scene for each of the different object classes (e.g., stroma, lymphocyte nuclei and cancer nuclei, as shown in Fig. 2).

A color swatch (a set of representative pixels that characterize the object of interest) is then defined to automatically identify the target EM-derived likelihood scene. The centroids of the objects detected in this scene are used as seed points to initialize the active contour. It is also worth noting that the probability estimates derived from the EM algorithm may be used to formulate an energy term in the form of a negative logarithm [36]

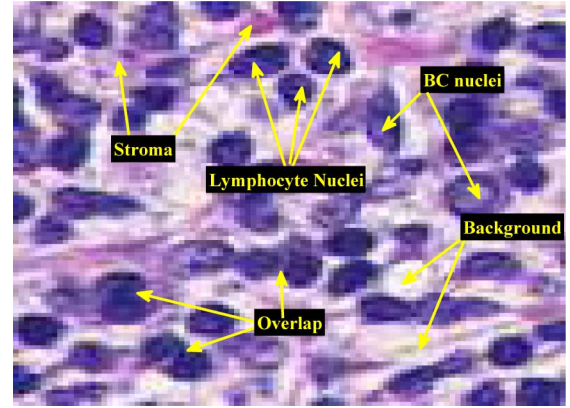


Fig. 2. Example of a HER2+ BC histopathology image showing lymphocyte nuclei, BC nuclei, stroma, and the background. Note the overlap between adjacent nuclei and the similarity in appearance between cancer and lymphocyte nuclei.

which may then be minimized during the curve evolution. However, the region dependency of the parameters of the energy term would be difficult to address in the context of segmenting an image with multiple regions [37]. Shape gradients are commonly employed [38] but may not be appropriate when the problem involves multiple overlapping objects against a noisy background.

Overlapping/touching objects are characteristic of our problem and we further process the contour result by splitting the contour between high concavity points. While Chetverikov [39] and Yang *et al.* [40] have previously described concavity detection algorithms, our methodology involves the use of directional image vectors to allow continuous concavity detection on closed contours. Additionally, a novel contour-splitting scheme is devised where the concavity points are connected by an *edge-path* algorithm that defines paths through relevant edge points within the contour while simultaneously ensuring an optimal split. The algorithm incrementally defines a path by including single edge points at a time to ensure that the split represents an edge or a potential overlap boundary. An intelligent heuristic rule based on object size is used to determine the need for a split. This enables us to distinguish between multiple objects identified as a single entity by the active contour. A texture based clustering step is applied to the output of the EMaGACOR result to distinguish lymphocyte nuclei from other structures in the image. Fig. 1 illustrates the various components of our algorithm for automated detection and segmentation of lymphocytes on HER2+ BC histopathology.

In summary, our segmentation method (EMaGACOR) is more ideally suited to the task of automatically detecting and segmenting lymphocytes in BC histopathology images by: 1) avoiding the need for training datasets that are difficult to define owing to variability across images; 2) enhancing the performance of active contours by providing a specific initialization via the EM algorithm; and 3) resolving the issue of contours enclosing multiple overlapping objects by splitting contours in favor of obtaining better lymphocyte nuclei segmentation. By altering the size heuristic and specifying a different color swatch,



TABLE I  
DESCRIPTION OF NOTATION

Symbol	Description	Symbol	Description
$\mathcal{C}$	2D image scene	$C$	2D Cartesian grid of pixels $c = (x, y)$
$f$	Function that assigns intensity value to pixel $c$	$K$	Number of classes/Gaussian mixtures
$D_k$	One of $K$ classes	$p_k$	Prior probability of $c$ belonging to $D_k$
$P(D_k f(c))$	Probability of $c$ belonging to $D_k$ given $f(c)$	$\mu_k$	Mean of $k^{th}$ Gaussian component
$\mathcal{L}_j$	Likelihood scene obtained from $P(D_k f(c))$	$\mathcal{L}_j^B$	Binarized scene obtained from $\mathcal{L}_j$
$\Sigma_k$	Covariance matrix of $k^{th}$ Gaussian component	$\mathcal{N}(f(c) \mu_k, \Sigma_k)$	$D$ dimensional Gaussian distribution function
$L$	Log-likelihood estimation of $\mathcal{C}$	$S$	Set of closed sub-contours $s$
$\epsilon$	Threshold for EM convergence	$c_w$	Point on contour $s$
$\phi$	Level set function	$\mathbf{f}$	Speed function for evolution of $\phi$
$\mathcal{C}_G$	Gray-scale scene from $\mathcal{C}$	$F$	Magnetostatic Force field matrix
$\alpha$	Real constant	$g$	Stopping function
$\nabla(\cdot)$	2D gradient along $X$ and $Y$ axes	$O_\ell$	Objects detected in $\mathcal{L}_j^B$
$q_\ell$	Centroids of $O_\ell$	$r$	Radius of initial contour centered at $q_\ell$
$\rho$	Threshold for contour evolution	$i$	Iteration number in EM algorithm
$t$	Iteration number in contour evolution	$\mathcal{A}(s)$	Area of $s$
$\tau_A$	Ideal area of lymphocyte nuclei	$\theta(c_w)$	Angle between pair of vectors
$\theta_{\max}$	Threshold to determine concavity point	$V_s$	Set of concavity points $c_m$ on $s$
$E$	Set of edge points $c_u$ enclosed by $s$	$Q_{ab}$	Path connecting pair of concavity points $\{c_a, c_b\}$
$\Gamma$	Ratio of areas of split contours $s_1$ and $s_2$	$\psi$	Threshold for determining split
$\sigma$	Standard deviation in intensity	$\kappa$	Average of intensities
$O_\zeta$	Object enclosed by $s$	$\mathbf{F}(O_\zeta)$	6D attribute vector
$\mathcal{I}$	Vector of sample intensities for object template	$\mathcal{G}$	Manually defined closed boundary

the EMaGACOR model is generalizable and may be employed in segmenting other objects (e.g., cancer nuclei) in the image. Hence, the scheme is readily extensible to other related biomedical image segmentation problems.

The rest of the paper is organized as follows. Section III describes the overall methodology of our segmentation model. In Section IV, we describe the experiments performed and evaluation methods employed to illustrate the improvements in lymphocyte detection and segmentation using our method. Quantitative and qualitative results are shown in Section V. Concluding remarks are presented in Section VI.

### III. METHODS

#### A. Data Description, Notation, and Ground Truth

Hematoxylin and eosin (H&E) stained BC biopsy cores were scanned into a computer using a high resolution whole slide scanner at 40x optical magnification at The Cancer Institute of New Jersey (CINJ). A total of 100 HER2+ BC images (from nine patients) exhibiting various levels of LI were used in our analysis. The images were down sampled by a factor of 2 and saved as  $200 \times 200$  pixels digital images. The ground truth for spatial presence of LI was obtained via manual detection and segmentation performed by a BC oncologist from CINJ. The ground truth for LI detection evaluation was obtained in the form of highlighted pixels representing the approximate centers of each of the lymphocytes in all 100 images. Note that, since

the 100 images comprised over 4000 individual lymphocytes, and on account of the effort involved in manual segmentation, only 100 lymphocytes randomly chosen from the set of 100 images were delineated by the expert to allow the evaluation of the segmentation performance of the model. The detection performance of the model, however, was evaluated on all lymphocytes across all 100 images. The H&E-stained histopathology images comprise of four main structures or entities, namely: 1) BC nuclei; 2) lymphocyte nuclei; 3) stroma; and 4) background, as illustrated in Fig. 2. Note the extent of overlap between objects and the similarity between lymphocyte nuclei and BC nuclei. Lymphocyte nuclei tend to be stained deeper than BC nuclei and are often smaller in size.

An image is defined as  $\mathcal{C} = (C, f)$ , where  $C$  is a 2-D grid representing pixels  $c \in C$ , with  $c = (x, y)$  representing the Cartesian coordinates of a pixel and  $f$  is a function that assigns intensity values to  $c \in C$ . Hence, for an RGB image  $\mathcal{C}$ ,  $f$  assigns a three element vector to  $c$  composed of red, green, and blue values. The GAC model is evolved on a gray-scale scene  $\mathcal{C}_G = (C, f_G)$ , where  $f_G$  assigns pixel  $c \in C$ , the single gray-scale intensity. A list of frequently appearing symbols is summarized in Table I.

#### B. EM-Based Segmentation of Object Classes

The EM algorithm is used to determine the probability of each pixel  $c$  belonging to one of  $K$  classes,  $D_k$ ,  $k \in \{1, 2, \dots, K\}$ ,

in an image scene. The EM algorithm attempts to identify the individual Gaussian distribution from a mixture of  $K$  normal class densities. For the application considered in this paper, we set  $K = 4$ , where  $D_k \in \{\text{lymphocyte nuclei, stroma, cancer nuclei, background}\}$ . The EM algorithm computes the posterior class conditional probability  $P(D_k|f(c))$  of each pixel  $c$  belonging to class  $D_k$ , given the prior class conditional probability  $p(f(c)|D_k)$ . The algorithm is run iteratively and is comprised of two steps: expectation (E-step) and maximization (M-step). The E-step calculates  $P(D_k|f(c))$  based on the current parameters of Gaussian mixture model, while the M-step recalculates or updates the model parameters at each iteration  $i$ ,  $\gamma_k^i = \{\mu_k^i, \Sigma_k^i, p_k^i\}$  where  $\mu_k^i$  and  $\Sigma_k^i$  are the mean and covariance of each Gaussian component, respectively, and  $p_k^i = p^i(f(c)|D_k)$ , also referred to as the mixture coefficients in the Gaussian mixture model. After convergence, the EM algorithm will assign each pixel  $c$  a  $1 \times K$  probability vector whose elements are the respective  $P(D_k|f(c))$  values. The algorithmic details of the EM algorithm are summarized shortly [41]. *Parameter initialization:* The initial parameters  $\gamma_k^0 = \{\mu_k^0, \Sigma_k^0, p_k^0\}$  are obtained from the  $K$ -means clustering [42] result. The  $K$ -means algorithm is employed to initially group all the pixels in the image into  $K$  clusters. The mean and covariance of each of these clusters are used as the initial estimates. The initial value of the mixture coefficients/prior probabilities  $p_k^0$  are set as the fraction of the number of pixels allocated to the corresponding  $K$ -means clusters. *E-step:* Calculate the posterior probabilities using the current parameters  $\gamma_k^i$

$$P(D_k|f(c)) = \frac{p_k^i \mathcal{N}(f(c)|\mu_k^i, \Sigma_k^i)}{\sum_{j=1}^K p_j^i \mathcal{N}(f(c)|\mu_j^i, \Sigma_j^i)}$$

for all  $c \in C$ , where  $j, k \in \{1, \dots, K\}$ , and

$$\begin{aligned} \mathcal{N}(f(c)|\mu_k^i, \Sigma_k^i) \\ = (2\pi)^{-D/2} |\Sigma_k^i|^{-1/2} \exp\left\{-\frac{1}{2}(f(c) - \mu_k^i)^T \Sigma_k^i^{-1} (f(c) - \mu_k^i)\right\} \end{aligned}$$

is a  $D$ -dimensional Gaussian distribution. For an RGB image,  $D$  is set to 3.

*M-step:* The mean, covariance, and prior probability of each class,  $\gamma_k^i$  are updated by the posterior probabilities obtained in E-step and are computed as follows:

$$\begin{aligned} \mu_k^{i+1} &= \frac{1}{n_k} \sum_{c \in C} P(D_k|f(c)) f(c) \\ \Sigma_k^{i+1} &= \frac{1}{n_k} \sum_{c \in C} P(D_k|f(c)) (f(c) - \mu_k^{i+1})(f(c) - \mu_k^{i+1})^T \\ p_k^{i+1} &= \frac{n_k}{|C|} \end{aligned}$$

where  $n_k = \sum_{c \in C} P(D_k|f(c))$  and  $|C|$  is the cardinality of  $C$ .

*Convergence evaluation:* Convergence is evaluated by calculating the Euclidean distance of log likelihood between current  $(i+1)$  and preceding iterations  $(i)$ . Assuming that the Gaussian distribution for every pixel  $c \in C$  is independent of one another,

based on Gaussian mixture model, the log-likelihood function of the model with parameters obtained in M-step can be computed as follows:

$$\begin{aligned} L^{i+1}(C|\mu, \Sigma, p) &= \ln \left\{ \sum_{k=1}^K p_k^{i+1} \mathcal{N}(f(c \in C)|\mu_k^{i+1}, \Sigma_k^{i+1}) \right\} \\ &= \ln \left\{ \sum_{k=1}^K p_k^{i+1} \left[ \prod_{c \in C} \mathcal{N}(f(c)|\mu_k^{i+1}, \Sigma_k^{i+1}) \right] \right\} \\ &= \sum_{c \in C} \sum_{k=1}^K \ln \{ p_k^{i+1} \mathcal{N}(f(c)|\mu_k^{i+1}, \Sigma_k^{i+1}) \} \end{aligned}$$

where  $f(c \in C)$  is the intensity associated with all  $c \in C$  and  $L^{i+1}$  is the log-likelihood estimation of  $C$  with respect to the Gaussian mixture model. The convergence criterion can be expressed via the following inequality:

$$\left\| \frac{L^{i+1} - L^i}{L^i} \right\| \leq \epsilon$$

where  $\epsilon$  is an empirically determined threshold and  $\|\cdot\|$  is the  $L_2$  norm. In our experiments  $\epsilon$  was set to  $10^{-5}$ . If the convergence criterion is not attained, the algorithm returns to the E-step. Otherwise, the algorithm will return a probability matrix obtained from last E-step, which is then used to group every pixel  $c \in C$  into  $K$  different classes. Following the convergence of the EM algorithm, we obtain  $K$  likelihood scenes  $\mathcal{L}_j = (C, \ell_j)$ ,  $j \in \{1, \dots, K\}$ , where  $\ell_j(c) = P(D_k|f(c))$  for each  $c \in C$ . For each  $\mathcal{L}_j$ , we obtain the corresponding binarized scenes  $\mathcal{L}_j^B = (C, \ell_j^B)$  where

$$\ell_j^B(c) = \begin{cases} 1, & \text{if } \ell_j = \max_k [P(D_k|f(c))] \\ 0, & \text{otherwise.} \end{cases}$$

The binarized scene  $\mathcal{L}_j^B$  represents the EM based segmentation for objects belonging to class  $D_k \in \{\text{lymphocyte nuclei, stroma, cancer nuclei, background}\}$ . The appropriate scene for lymphocyte nuclei is used to initialize the active contour model. In order to do this, we specify a swatch (color template) for the object class of interest. The swatch ( $\mathcal{I}$ ) is defined as  $\mathcal{I} = \{f(c_1), \dots, f(c_\beta)\}$  representing sample intensities of the object of interest, where  $\{c_1, \dots, c_\beta\} \in C$  are pixels obtained from the object of interest on a representative image. The objects of interest in our problem are the lymphocyte nuclei. The individual binary scenes  $\mathcal{L}_j^B$ ,  $j \in \{1, \dots, K\}$  are then overlaid on the original color scene  $C$  and the mean color vector of all nonzero pixels is computed. The target scene  $\mathcal{L}^B$  is chosen as the one whose mean intensity is most similar to the average intensity of the color swatch  $\mathcal{I}$ . A sample image with its corresponding binarized scenes representing the different object classes is shown in Fig. 3. Although the EM algorithm is often unable to distinguish between BC nuclei and lymphocyte nuclei and suffers from fragmented segmentation of a single object, it can provide a specific initialization to an active contour, as described shortly.

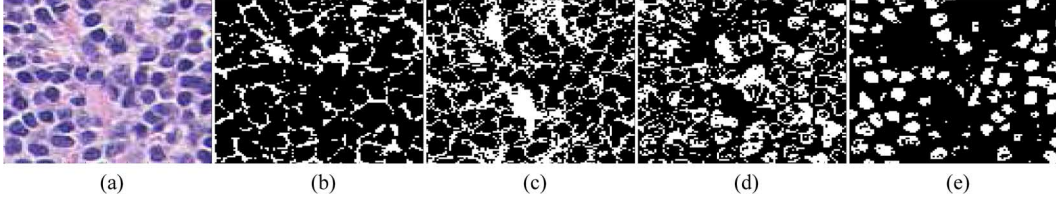


Fig. 3. (a) Original HER2+ BC histopathology image with corresponding (b)–(e) class binarized scenes  $\mathcal{L}_j^B$  for  $j \in \{1, \dots, 4\}$  obtained via EM. The binary image corresponding to lymphocytes is shown in (e).

### C. EM-Driven GAC Model (EMaGAC)

An active contour model is used in our boundary segmentation problem where a set of contours  $S$  is evolved using level-set method [43] to find target object boundaries.  $S$  is a set of closed subcontours where the cardinality of  $S$  depends upon the initialization and the final evolved result. Each closed subcontour  $s \in S$  is composed of an ordered set of  $M$  points such that  $s = \{c_w | w \in \{1, \dots, M\}\}$ , where each point  $c_w \in C$  is connected to only two adjacent points  $c_{w-1}$  and  $c_{w+1}$  with  $c_{M+1} = c_1$  and  $c_0 = c_M$  to form a closed loop. All  $s \in S$  are evolved simultaneously in time  $t$  over a 2-D gray-scale scene  $\mathcal{C}_G$ .  $s$  is represented as a zero level set  $s = \{c_w | \phi(t, c_w) = 0\}$  of a level-set function  $\phi$  that is defined as zero at the boundary, +1 in the interior of the boundary, and -1 outside the boundary. In general, the evolution of the level-set function  $\phi$  is accomplished by an iterative solution to the following partial differential equation:

$$\frac{\partial \phi}{\partial t} + \mathbf{f} |\nabla \phi| = 0$$

where  $\mathbf{f}$  is the evolving speed function that drives the subcontour  $s$  toward the desired object boundary.  $\mathbf{f}$  is unique to an active contour model and its design depends on how the information in the image is used. In this paper, a specific initialization from the EM results is provided to the MAC model [17], which is briefly discussed below. The advantage of employing the MAC model is that it utilizes a bidirectional force that allows the contour to grow or shrink toward object boundaries.

1) *MAC Model*: The MAC model implements a bidirectional force field  $F$  generated from a hypothetical magnetostatic interaction between the set of contours  $S$  and the object boundary. The force field  $F$  drives the contour toward the boundary. Both the boundary and the contour are treated as current carrying loops and the magnetic field from the boundary is computed using the well known Biôt–Savart law. This magnetic field determines the force  $F$  acting on the contour, where  $F$  is defined over the entire scene  $\mathcal{C}_G$ . The level-set implementation of the contour as proposed in [17] takes the form

$$\frac{\partial \phi}{\partial t} = \alpha g(\mathcal{C}_G) \nabla \cdot \left( \frac{\nabla \phi}{|\nabla \phi|} \right) |\nabla \phi| - (1 - \alpha) F(\mathcal{C}_G) \cdot \nabla \phi$$

where  $\alpha$  is a real constant,  $g(\mathcal{C}_G) = 1/(1 + |\nabla \mathcal{C}_G|)$ , and  $\nabla(\cdot)$  represents the 2-D image gradient  $(\partial(\cdot)/\partial X, \partial(\cdot)/\partial Y)$  along the  $X$  and  $Y$  axes. We refer the reader to [17] for additional details on the MAC model.

2) *Contour Initialization Scheme*: For a given image  $\mathcal{C}$ , the binary image corresponding to lymphocyte nuclei,  $\mathcal{L}^B$ , is iden-

---

#### Algorithm : EMaGAC

**Input** :  $\mathcal{L}^B, \mathbf{f}, r, \rho, \Delta t, \mathcal{I}$ .

**Output** : Final evolved contour.

#### BEGIN

Select target binary scene  $\mathcal{L}^B$  using color swatch template  $\mathcal{I}$ ,  
 Determine all objects  $O_\varrho$  by connected component labeling,  
 Obtain corresponding centroids  $q_\varrho = \frac{1}{|O_\varrho|} \sum_{c \in O_\varrho} c$ ,

Define initial contour  $\phi_0 = \phi(0, c)$  as circle of radius  $r$   
 centered at each  $q_\varrho$ ,

while  $\|\phi^{t+1} - \phi^t\| > \rho$ ,

Evolve contour,  $\phi^{t+1} = \phi^t + [\mathbf{f}(\nabla \phi^t)](\Delta t)$ ,

end

#### END

---

Fig. 4. EMaGAC algorithm

tified via the color swatch matching scheme (as described in Section III-B). For all objects  $O_\varrho$ , where  $\varrho \in \{1, \dots, \Omega\}$ , detected in  $\mathcal{L}^B$  via connected component labeling [44], centroids  $q_\varrho = \frac{1}{|O_\varrho|} \sum_{c \in O_\varrho} c$  are computed that serve as seeds points for initializing the GAC [see Fig. 8(a)]. The initial contour  $\phi_0 = \phi(0, c)$ , is defined as a circle of radius  $r$  centered at each  $q_\varrho$  [see Fig. 8(b)].  $r$  is empirically chosen to be approximately half the radius of the target object. In our problem, the radius of a typical lymphocyte nuclei was empirically determined to be 4 pixels. Hence, we set  $r$  to 2 pixels. The contour is then evolved till the differences in spatial location of the contours of the current iteration ( $\phi^t$ ) to the next ( $\phi^{t+1}$ ) are below an empirically determined threshold  $\rho$ . We found that setting  $\rho = 10^{-6}$  allowed the model to converge to a stable solution. The algorithm for initializing the contour model is illustrated in Fig. 4.

The active contour provides a segmentation that focuses on lymphocyte nuclei and cancer nuclei [see Fig. 8(c)]. Note that various contours contain two or more overlapping/touching objects that cannot be resolved by the EMaGAC model alone. An additional step to explicitly split object contours is required (overlap resolution), resulting in the EMaGACOR model.

### D. Resolving Overlap—EMaGACOR Model

In addition to providing a specific initialization to GAC via EM, the EMaGACOR model aims at improving segmentation by providing overlap resolution where contours enclosing multiple objects are split using a size heuristic. The spirit behind the overlap resolution scheme in EMaGACOR is as follows. In the first step, we identify contours enclosing multiple objects and determine high concavity points on the contour. An *edge-path*

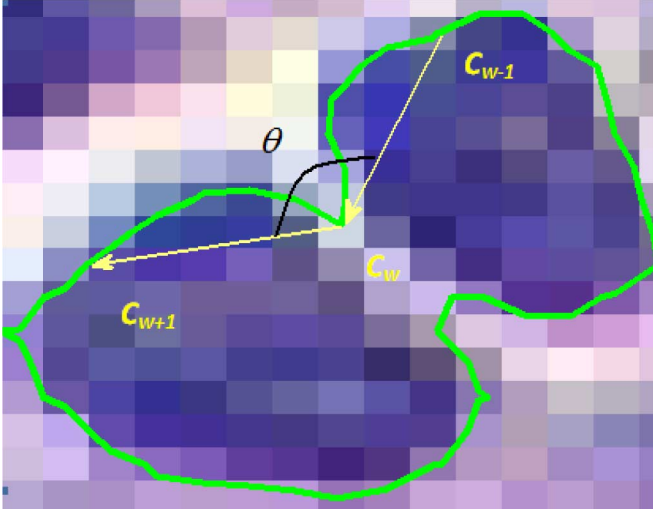


Fig. 5. Concavity detection: Three consecutive points on  $s$  ( $c_{w-1}$ ,  $c_w$  and  $c_{w+1}$ ) are used to define two vectors (shown with arrows). The angle  $\theta$  between them is a measure of concavity/convexity of the point  $c_w$ . Concavity points can be distinguished from convex points by computing the cross product between the vectors where a positive cross product would indicate a concavity point if moving in a counter-clockwise direction on  $s$ .

algorithm is then used to define potential paths between the previously detected concavity points using edge information within the contour. Based on a size heuristic scheme, an optimal path is selected to split the original contour and isolate individual objects.

1) *Concavity Detection*: High concavity points are characteristic of contours that enclose multiple objects and represent junctions where object intersection occurs (see Fig. 5). The area  $\mathcal{A}(s)$  of the closed subcontour  $s$  is compared to predetermined area of an ideal lymphocyte nucleus  $\tau_A$ . For our experiments  $\tau_A$  was set to 35. Hence, a subcontour is eligible for a split if  $\mathcal{A}(s) > \tau_A$ . Since  $c = (x, y)$ , the difference between any two points  $c_w$  and  $c_{w-1}$  will represent a vector in 2-D. Concavity points are detected by computing the angle between vectors defined by three consecutive points  $(c_{w-1}, c_w, c_{w+1}) \in s$ . The degree of concavity/convexity is proportional to the angle  $\theta(c_w)$ , as shown in Fig. 5.  $\theta(c_w)$  can be computed from the following dot-product relation:

$$\theta(c_w) = \pi - \arccos \left( \frac{(c_w - c_{w-1}) \cdot (c_{w+1} - c_w)}{\|c_w - c_{w-1}\| \|c_{w+1} - c_w\|} \right). \quad (1)$$

Concavity points can be distinguished from convexity points by computing the cross product of the vectors  $(c_w - c_{w-1})$  and  $(c_{w+1} - c_w)$ , where a concavity point would yield a positive cross product if the point were moving in a counterclockwise direction on  $s$  (see Fig. 5). The value of  $\theta(c_w)$  for an eligible concavity point  $c_w$  is constrained to be less than an empirically determined value  $\theta_{\max}$ . The value of  $\theta_{\max}$  serves as a threshold for detecting meaningful concavity points and in our case it was found that  $\theta_{\max} = \frac{8}{9}\pi$  yielded optimal results. Note that numerous false concavity points may also be detected due to noisy boundaries on the contour. However, these concavity points will not be used to split the contour as they will not satisfy the size

---

**Algorithm:** ConcavityDetection

**Input :**  $s = \{c_w | w \in \{1, \dots, M\}\}$ ,  $\theta_{\max}$ ,  $\tau_A$ .

**Output :** concavity points  $V_s$ .

**BEGIN**

for all  $s \in S$ ,

compute  $\mathcal{A}(s)$ ,

if  $\mathcal{A}(s) > \tau_A$ ,

for all points  $c_w \in s$ ,

compute vectors  $(c_w - c_{w-1})$  and  $(c_{w+1} - c_w)$ ,

compute  $\theta(c_w)$  between vectors,

if  $\theta(c_w) \leq \theta_{\max}$  & cross product  $\geq 0$ , save  $c_w \rightarrow V_s$ ,

end

end

end

**END**

---

Fig. 6. Concavity detection algorithm.

heuristic condition (3) explained shortly (see Section III-D2). The concavity detection algorithm is summarized in Fig. 6.

2) *Edge-Path Algorithm*: The *edge-path* algorithm aims at splitting large contours using edge information within the contour and employing a size heuristic. Initially, a path between a pair of concavity points through edge points enclosed within  $s$  is defined. From all the possible paths between various concavity points, the path that satisfies a split yielding a subcontour whose size is close to that of an ideal lymphocyte nuclei ( $\tau_A$ ) is favored. Numerous paths may satisfy the size heuristic, however, the algorithm will select the shortest path to define the initial split. This ensures that the contour is not abruptly split between distant concavity points.

Let  $V_s$  be the set of  $N$  concavity points detected on  $s$  such that  $V_s = \{c_m | m \in \{1, \dots, N\}\}$  and  $N \leq M$ . Let  $E$  represent the set of  $H$  edge points enclosed by  $s$  such that  $E = \{c_u | u \in \{1, \dots, H\}\}$ . In our implementation, the set of edge points is derived using a Sobel filter on scene  $\mathcal{C}_G$ . For a given pair of concavity points  $\{c_a, c_b\} \in V_s$  and  $a \neq b$ ,  $a, b \in \{1, \dots, N\}$ , the path  $Q_{ab}$  between them is defined through a total  $h$  ordered edge points in  $E$ , such that  $Q_{ab} = \{c_a, c_1, \dots, c_h, c_b\}$  is an ordered set with each of its points connected to only to two of its adjacent points and satisfies the following condition:

$$\|c_a - c_b\| \geq \|c_1 - c_b\| \geq \dots \geq \|c_h - c_b\| \geq 0 \quad (2)$$

and there does not exist  $c_{h+1} \in Q_{ab}$  such that  $\|c_h - c_b\| \geq \|c_{h+1} - c_b\|$ . In addition,  $c_{h-1}$  is the closest point to  $c_h$  that satisfies the above condition (2). Hence, the path  $Q_{ab}$  is defined in an incremental manner by first finding the closest edge point to  $c_a$  and ensuring that (2) is satisfied. The step is repeated until  $c_a$  connects  $c_b$ . In this manner, the most relevant edge between any two given concavity points  $\{c_a, c_b\}$  is included while simultaneously guaranteeing a path between them. Fig. 7 illustrates a hypothetical example of numerous paths defined to split a contour enclosing three objects. The paths are defined between each pair of four concavity points labeled by roman numerals. Once a path  $Q_{ab}$  is defined, it splits  $s$  into two subcontours— $s_1$  and  $s_2$  (not shown in figure). The area of these contours  $\mathcal{A}(s_1)$  and  $\mathcal{A}(s_2)$  are computed and compared to the predetermined area



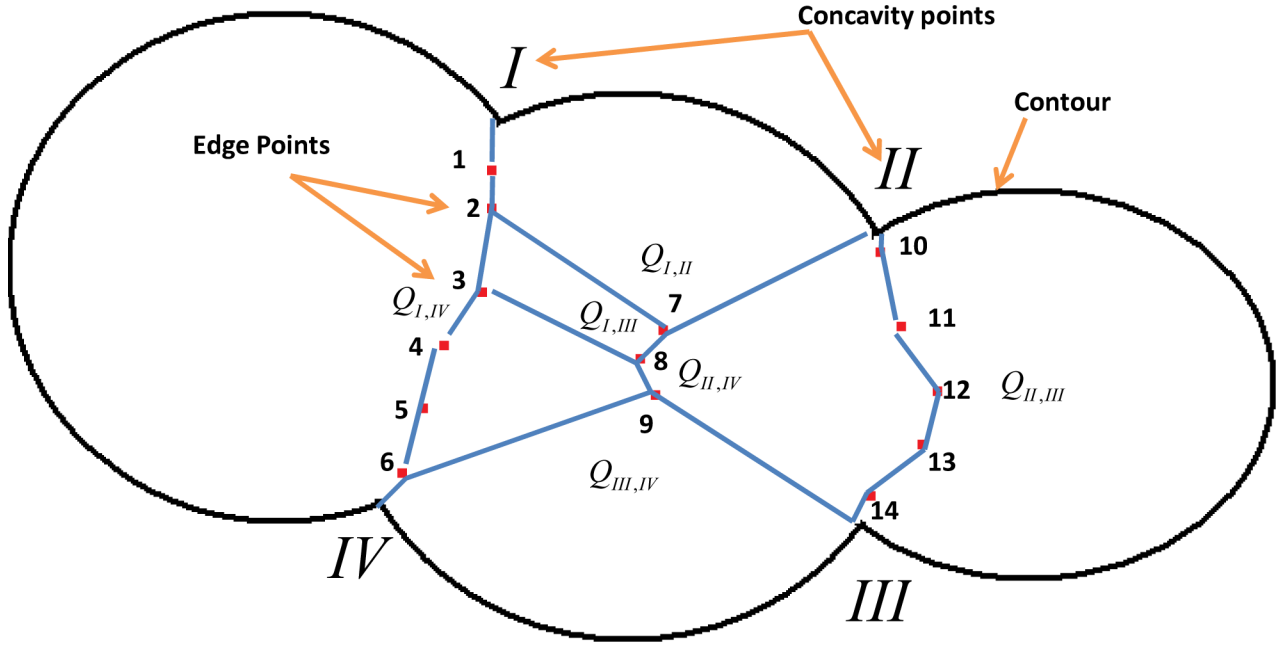


Fig. 7. Illustration of the *edge-path* algorithm. In this hypothetical example, the contour  $s$  encloses three objects. Edge points (red) included in paths (blue) are numbered. Four concavity points  $I$ ,  $II$ ,  $III$ , and  $IV$  are shown. The path between  $I$  and  $II$  is defined as  $Q_{I,II} = \{I, 1, 2, 7, II\}$ . Similarly,  $Q_{I,III} = \{I, 1, 2, 3, 8, 9, III\}$ ,  $Q_{I,IV} = \{I, 1, \dots, 6, IV\}$ ,  $Q_{II,III} = \{II, 10, \dots, 14, III\}$ ,  $Q_{II,IV} = \{II, 7, 8, 9, 6, IV\}$ , and  $Q_{III,IV} = \{III, 9, 6, IV\}$ . Of all these paths,  $Q_{I,II}$  and  $Q_{III,IV}$  do not satisfy the size heuristic. Of the remaining paths, the shortest path ( $Q_{II,III}$ ) will be used to split the contour.

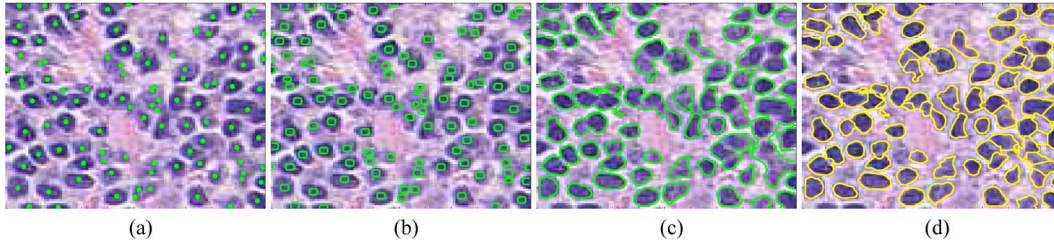


Fig. 8. (a) Object centroids  $q_o$  (green) obtained from EM derived binarized scene  $\mathcal{L}^B$  that serve as seed points for initializing the active contours  $s \in S$ . (b) Object initialization for the GAC, defined as circles centered at each  $q_o$ . (c) Contour result after evolution. (d) Improved segmentation after splitting contours enclosing multiple objects by using the *edge-path* algorithm.

of an ideal lymphocyte nucleus  $\tau_A$ . The larger ratio of the areas of the split contours  $\Gamma = (\mathcal{A}(s_1)/\mathcal{A}(s_2))$  ( $\mathcal{A}(s_1) \geq \mathcal{A}(s_2)$ ) are constrained to a threshold  $\psi$  described as follows:

$$\psi = \frac{\mathcal{A}(s) - \tau_A}{\tau_A}.$$

The path  $Q_{ab}$  that splits  $s$  is accepted if  $\Gamma$  satisfies the following conditions:

$$\begin{aligned} \Gamma &\leq \psi, & \text{if } \psi \geq 1 \\ \Gamma &\geq \psi^{-1}, & \text{if } \psi < 1. \end{aligned} \quad (3)$$

The condition in (3) favors the size heuristic so that at least one of the subcontours  $s_1$  or  $s_2$  has an area close to  $\tau_A$ . The shortest path that satisfies this condition is selected to define the initial split. The process of splitting contours is repeated until all contours have an area roughly equal to  $\tau_A$ . For the hypothetical example shown in Fig. 7, the value of threshold  $\psi$  would be computed as 2 as it encloses three objects. Hence, the area ratio  $\Gamma$  must be less than or equal to 2. Paths  $Q_{I,II}$  and

$Q_{III,IV}$  do not satisfy this condition. Of the remaining paths, the shortest path ( $Q_{II,III}$ ) will be used to split the contour into  $s_1$  (left) and  $s_2$  (right). The *edge-path* algorithm is then repeated to split the larger contour  $s_1$ . The final contour segmentation on HER2+ BC histopathology image is shown in Fig. 8(d). Detailed splitting of such contours is illustrated in Fig. 9. Note that lack of edge information does not impede the algorithm from segregating overlapping objects. The *edge-path* algorithm is described in Fig. 10.

#### E. Texture-Based Clustering to Detect Lymphocytes

$K$ -means clustering [42] of first-order statistical texture features derived from within the object interior is used to classify the object segmented by EMaGACOR as a lymphocyte or not. The first-order statistical features considered were the standard deviation  $\sigma$  and average intensity  $\kappa$  from three channels (R, G, and B) computed for the region enclosed within the contour. Thus, each candidate object  $O_\zeta$ ,  $\zeta \in \{1, \dots, T\}$ , is described by a 6-D attribute vector  $\mathbf{F}(O_\zeta)$ , comprised of  $\sigma$  and  $\kappa$



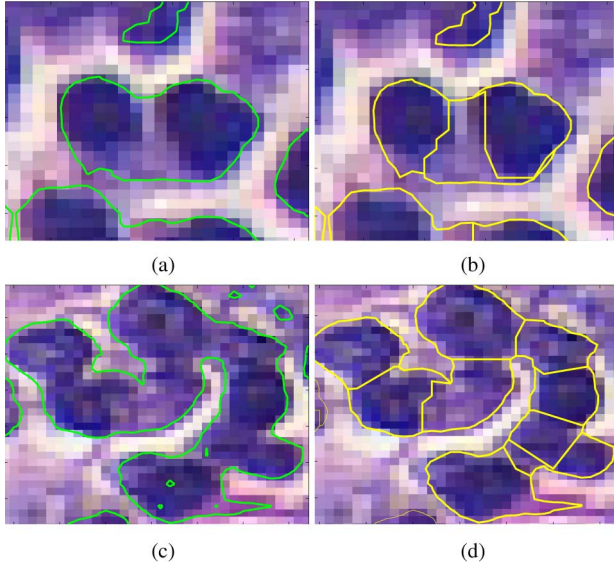


Fig. 9. (a) and (c) Example of a contour enclosing overlapping lymphocytes. Lack of edges/weak edges prevents the contour from providing accurate object segmentation. (b) and (d) Contour split by *edge-path* algorithm using size heuristic.

values within  $O_c$ . Fig. 11 illustrates the final result showing the detected lymphocyte nuclei by obtaining the centroids of the closed contours  $s$  (green dots).

#### IV. EXPERIMENTS AND PERFORMANCE MEASURES

A total of 100 images were analyzed using the EMaGACOR model. We compared our results to segmentation obtained from randomly initialized GAC to illustrate the improvement in segmentation obtained by EM initialization (EMaGAC). In addition, we compared EMaGACOR results to segmentations obtained via EMaGAC model to illustrate the importance of the overlap resolution via contour splitting. The following three models were thus quantitatively and qualitatively evaluated in terms of their detection and segmentation performance:

- 1) randomly initialized GAC model;
- 2) EM-driven GAC model (EMaGAC);
- 3) EMaGACOR.

The methods employed for quantitative evaluation are described shortly.

##### A. Evaluation of Detection Performance

The metrics used to evaluate lymphocyte detection include: 1) sensitivity (SN); 2) positive predictive value (PPV); and 3) overlap detection ratio (OR) (see Table II). The detection results from the three models (GAC, EMaGAC, and EMaGACOR) are compared to the manual detection results obtained from an expert clinician. The SN and PPV values are computed from the true-positive (TP), false-positive (FP), and false-negative (FN) values

$$SN = \frac{TP}{TP + FN} \quad (4)$$

##### Algorithm : Edge-path

**Input** : Sub-contour  $s$ , object size  $\tau_A$ ,

Edge points  $c_u \in E$ , Concavity points  $V_s$ .

**Output** : Final path  $Q_{ab}$  to split  $s$  into  $s_1$  and  $s_2$ .

**BEGIN**

if  $|V_s| > 2$ ,

set  $\mathcal{D} = \infty$ ,

compute  $\psi = \frac{A(s) - \tau_A}{\tau_A}$ ,

for each pair of concavity points  $\{c_a, c_b\} \in V_s, a \neq b$ ,

set  $c_A = c_a$ , set  $e = E$ , mark  $c_b$  as edge point,

set  $Q_{ab} = \{c_a\}$ ,

while  $c_A \neq c_b$ ,

find closest edge point  $c_u \in e$  to  $c_A$ ,

if  $\|c_A - c_b\| \geq \|c_u - c_b\|$ ,

save  $c_u \rightarrow Q_{ab}$ , set  $c_A = c_u$ ,

end

remove  $c_u$  from  $e$ ,

end

Split  $s$  into  $s_1$  and  $s_2$  using  $Q_{ab}$ ,

compute  $\Gamma = \frac{A(s_1)}{A(s_2)}$ ,

if Equation 3 is true and  $\|Q_{ab}\| < \mathcal{D}$ ,

accept split and reject any previous split,

set  $\mathcal{D} = \|Q_{ab}\|$ ,

end

end

**END**

Fig. 10. Detailed description of the *edge-path* algorithm.

TABLE II  
QUANTITATIVE EVALUATION OF DETECTION RESULTS BETWEEN EMaGACOR, EMaGAC, AND GAC MODELS OVER 100 IMAGES

	SN	PPV	OR
GAC	0.19	0.56	0.022
EMaGAC	0.42	0.63	0.14
EMaGACOR	0.86	0.64	0.91

$$PPV = \frac{TP}{TP + FP} \quad (5)$$

TP refers to the number of lymphocyte nuclei correctly identified while FP refers to the number of objects incorrectly identified as lymphocyte nuclei and FN refers to the number of lymphocyte nuclei missed by the model. The detection results are represented as centroids of the region enclosed by a closed contour. TP, FP, and FN values are obtained by comparing each centroid generated by the model to manually determined object centroids which served as our ground truth.

The overlap detection ratio (OR) is computed as follows:

$$OR = \frac{\text{Number of overlaps resolved}}{\text{Total number of overlaps}}.$$

An overlap is characterized by the existence of a common boundary between two objects which in our case may be between two or more lymphocyte nuclei, cancer nuclei, or cancer and lymphocyte nuclei.

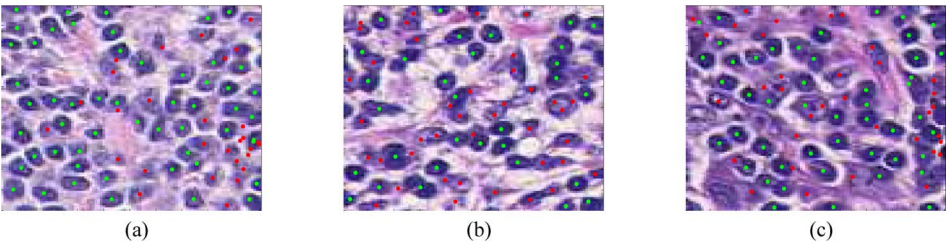


Fig. 11. Final lymphocyte detection results from EMaGACOR for three studies (a), (b), and (c) (green dots are centers of objects detected as lymphocyte nuclei). The result is obtained by clustering textural features (average and standard deviation in intensity) extracted from final contour result into two groups. Red dots are for objects identified as not being lymphocytes.

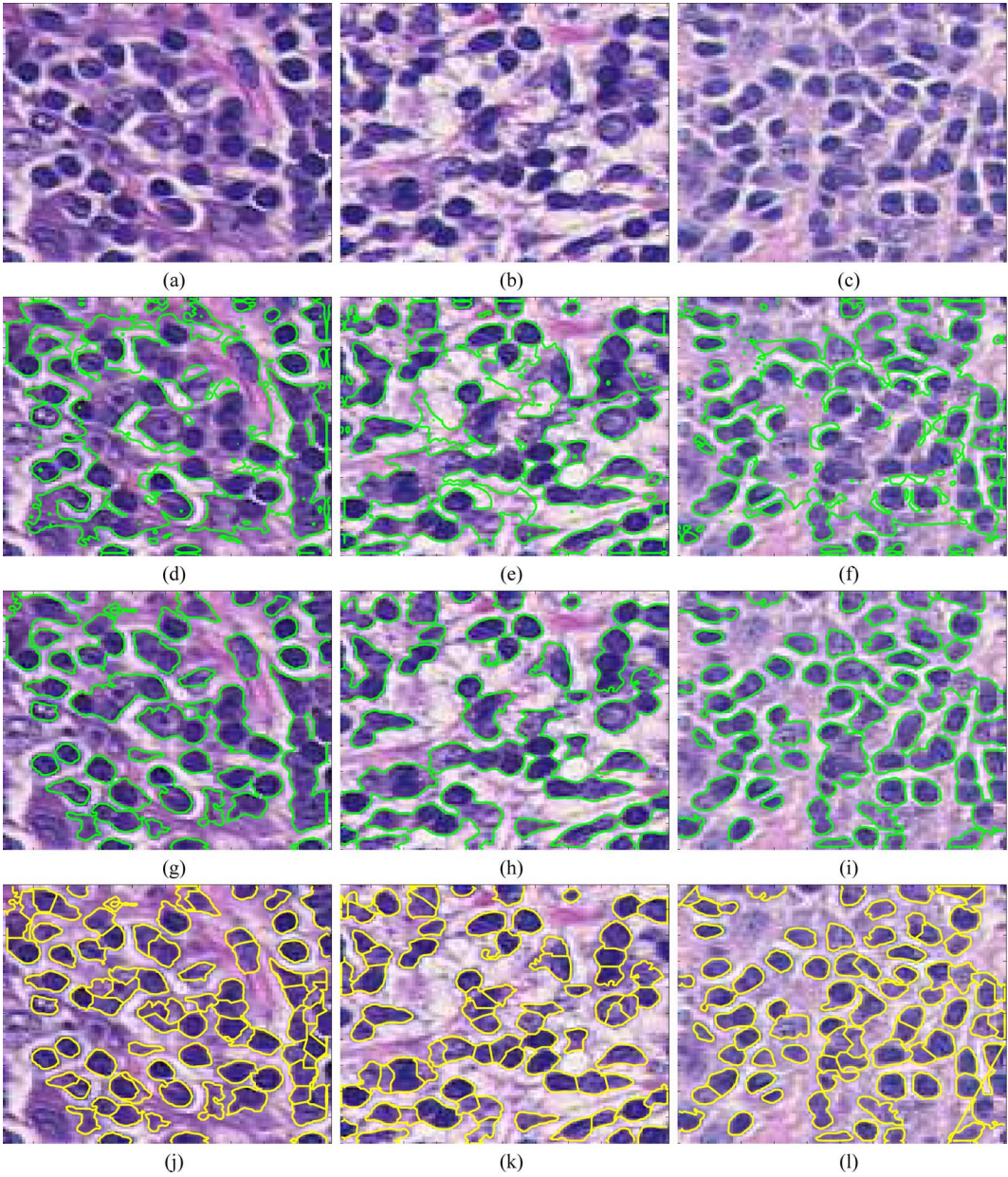


Fig. 12. (a)–(c) Original HER2+ BC histopathology image with corresponding segmentation results from (d)–(f) randomly initialized GAC, (g)–(i) GAC initialized via EM (EMaGAC), and (j)–(l) EMaGACOR that implements contour splitting using concavity detection and the *edge-path* algorithm.



TABLE III  
QUANTITATIVE EVALUATION OF SEGMENTATION RESULTS BETWEEN EMAGACOR, EMAGAC, AND GAC MODELS FOR 100 LYMPHOCYTES

	$SN_a$	$SP_a$	$PPV_a$	$OV_a$	HD	MAD
<b>GAC</b>	0.31	0.94	0.21	0.17	11.3	7.6
<b>EMaGAC</b>	0.80	0.98	0.58	0.51	3.2	1.5
<b>EMaGACOR</b>	0.80	1.0	0.86	0.72	2.1	0.9

### B. Evaluating Segmentation Performance

Segmentation results are compared to manual delineations performed by an expert oncologist (which serves as ground truth for segmentation evaluation) by computing boundary based metrics [45], namely: 1) Hausdorff distance (HD) and 2) mean absolute distance (MAD); and area overlap metrics [45] [true-positive area ( $TP_a$ ), false-positive area ( $FP_a$ ), true-negative area ( $TN_a$ ), and false-negative area ( $FN_a$ )]. The manual delineation is represented as a closed boundary  $\mathcal{G} = \{c_\chi | \chi \in \{1, \dots, \varpi\}\}$ . The HD and MAD values for each contour  $s$  are defined as

$$HD = \max_w \left[ \min_{\chi} \|c_w - c_\chi\| \right]$$

$$MAD = \frac{1}{M} \sum_{w=1}^M \left[ \min_{\chi} \|c_w - c_\chi\| \right].$$

Since manual delineation of 100 images is tedious and time consuming, the expert was asked to provide manual delineations of 100 lymphocyte nuclei randomly chosen from the set of 100 images. One random lymphocyte was picked from each of the 100 images. For each of the 100 object boundary segmentations, a corresponding value of HD and MAD were obtained. The HD and MAD results are illustrated as frequency histogram plots (see Fig. 13). HD and MAD values close to zero correspond to better segmentation. The area overlap metrics are used to compute the sensitivity  $SN_a$ , specificity  $SP_a$ , positive predictive value  $PPV_a$  and the overlap ratio  $OV_a$  of the segmentation results for each of the three models. The area overlap metrics are computed as follows:

$$TP_a = \frac{|\mathcal{A}(s) \cap \mathcal{A}(\mathcal{G})|}{|\mathcal{A}(\mathcal{G})|}$$

$$FP_a = \frac{|\mathcal{A}(s) \cup \mathcal{A}(\mathcal{G}) - \mathcal{A}(\mathcal{G})|}{|\mathcal{A}(\mathcal{G})|}$$

$$FN_a = \frac{|\mathcal{A}(s) \cup \mathcal{A}(\mathcal{G}) - \mathcal{A}(s)|}{|\mathcal{A}(\mathcal{G})|}$$

$$TN_a = \frac{|C - \mathcal{A}(\mathcal{G})|}{|\mathcal{A}(\mathcal{G})|}$$

where  $\mathcal{A}(\cdot)$  is the area of the closed boundary. The  $SN_a$  and  $PPV_a$  values are computed in a similar fashion as described in (4) and (5), respectively.  $SP_a$  and  $OV_a$  values are computed as follows:

$$SP_a = \frac{TN_a}{TN_a + FP_a}$$

$$OV_a = \frac{TP_a}{TP_a + FP_a + FN_a}.$$

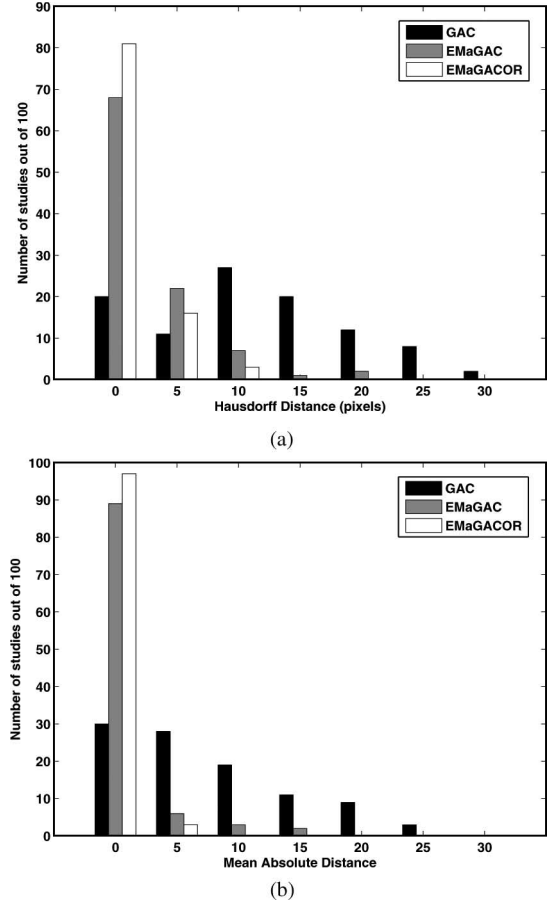


Fig. 13. Histogram for (a) HDs and (b) MADs for each of the three models. Note that HD and MAD values close to zero correspond to better segmentation. The plot reflects the number of studies (y-axis) for which the HD and MAD values were below a certain number of image pixels (x-axis).

## V. RESULTS AND DISCUSSION

### A. Qualitative Results

Qualitative results for three of the 100 different studies are illustrated in Fig. 12. These images were chosen to reflect the superiority of EMaGACOR over the EMaGAC and GAC models. The GAC model is limited in its ability to segment objects of interest [see Fig. 12(d)–(f)]. Note that the EM initialization allows the contour to focus on objects of interest and prevents the contour from enclosing numerous objects [see Fig. 12(g)–(i)]. In addition, the concavity detection scheme and the *edge-path* algorithm provide overlap resolution resulting in improved segmentation [see Fig. 12(j)–(l)].

### B. Quantitative Results

Results of quantitative evaluation of detection and segmentation performance for each of the three models are shown in Tables II and III. These results reflect the improved performance



TABLE IV

*p*-VALUES OF *t*-TEST BETWEEN EMAGACOR, EMAGAC, AND GAC MODELS FOR DETECTION AND SEGMENTATION PERFORMANCE METRICS FROM 100 STUDIES

	<i>SN</i>	<i>PPV</i>	<i>OR</i>	<i>SN<sub>a</sub></i>	<i>SP<sub>a</sub></i>	<i>PPV<sub>a</sub></i>	<i>OV<sub>a</sub></i>	<i>HD</i>	<i>MAD</i>
<b>GAC/EMaGAC</b>	$2.6 \times 10^{-34}$	$6.1 \times 10^{-3}$	$1.8 \times 10^{-14}$	$1.1 \times 10^{-16}$	$8.1 \times 10^{-3}$	$1.9 \times 10^{-15}$	$1.9 \times 10^{-17}$	$4.5 \times 10^{-16}$	$1.5 \times 10^{-13}$
<b>GAC/EMaGACOR</b>	$1.3 \times 10^{-73}$	$1.4 \times 10^{-3}$	$1.3 \times 10^{-81}$	$6.1 \times 10^{-18}$	$8.2 \times 10^{-4}$	$2.8 \times 10^{-33}$	$1.2 \times 10^{-31}$	$3.8 \times 10^{-21}$	$2.1 \times 10^{-17}$
<b>EMaGAC/EMaGACOR</b>	$1.5 \times 10^{-50}$	$6.9 \times 10^{-1}$	$1.4 \times 10^{-57}$	$9.5 \times 10^{-1}$	$1.3 \times 10^{-4}$	$7.9 \times 10^{-12}$	$1.7 \times 10^{-10}$	$4.6 \times 10^{-3}$	$1.7 \times 10^{-2}$

of EMaGACOR over EMaGAC and GAC, respectively. The SN and PPV values listed in Table II reflect the efficacy of the EMaGACOR model in detecting lymphocyte nuclei in BC images as compared to GAC and EMaGAC models. A total of 952 cases of overlapping objects were manually identified in 100 images and the EMaGACOR model was able to resolve 91% of all overlaps. In terms of segmentation performance, EMaGACOR easily outperformed both EMaGAC and GAC models, respectively, in terms of all six segmentation evaluation measures over 100 lymphocytes. One random lymphocyte was chosen from each of the 100 images to evaluate segmentation results. Fig. 13(a) and (b) shows histogram plots of the HD and MAD values for the three models over 100 lymphocytes for which segmentation evaluation was done. The HD and MAD values for the EMaGACOR model were less than 5 pixels for over 95% and 99% of the studies respectively. The statistical measures from the area overlap metrics are summarized in Table III.

### C. Test of Statistical Significance Between Models

For each of the 100 studies, performance evaluation metrics (SN, PPV,  $SN_a$ ,  $PPV_a$ ,  $SP_a$ ,  $OV_a$ , HD, and MAD) were compared for every pair of models (GAC/EMaGAC, GAC/EMaGACOR, EMaGAC/EMaGACOR) using the paired *t*-test under the null hypothesis that there is no significant difference in the values of the metrics between the EMaGACOR, EMaGAC, and GAC models. The null-hypothesis was rejected for *p*-value  $\leq 0.05$  (see Table IV). For almost every performance metric, EMaGACOR showed a statistically significant improvement compared to GAC and EMaGAC models. The difference in *PPV* values between the EMaGAC and EMaGACOR models is not statistically significant as the overlap resolution step sometimes results in false positives due to aggressive splitting. However, EMaGACOR had an overall higher PPV compared to the other models. Nonsignificant difference was also obtained between the area overlap sensitivities ( $SN_a$ ) for the EMaGAC and EMaGACOR models. This may be attributable to the fact that our sample size was limited. Note that for the evaluation of segmentation results we were able to obtain expert delineations of only 100 lymphocytes from the clinicians. A statistically significant difference may have been obtained if the evaluation were performed using a larger cohort. However, obtaining such a large set of manual annotations is a challenge, especially given the fact that clinicians would have to provide them.

## VI. CONCLUDING REMARKS

In this paper, we have presented a novel detection and segmentation scheme, EMaGACOR. Our segmentation scheme overcomes a number of issues that plague popular segmenta-

tion schemes. Specifically, our model is able to overcome the sensitivity associated with random initialization of the active contour model and is able to resolve object overlap. In addition, the scheme differs from supervised classifier detection methods that are encumbered by the need for a large number of annotated training samples. The initialization of the contour via the EM scheme allows for: 1) improved active-contour-based segmentation, and also 2) eliminates the need for supervised classifier-based approaches for lymphocyte detection, which typically require the availability of a large number of accurately and carefully annotated instances. The concavity detection scheme in conjunction with the *edge-path* algorithm effectively resolves the issue of segmenting overlapping objects.

We applied EMaGACOR for segmentation and detection of lymphocyte nuclei in HER2+ BC histopathology images. Quantifying the extent of LI in BC histopathology images is important due to the correlation of LI to nodal metastasis in BC. The scheme described in this paper could be developed into an automated image analysis routine for making prognostic predictions of outcome associated with HER2+ BC. While EMaGACOR was able to address most of the major issues in detecting lymphocyte nuclei such as similarity with other structures, high variability across images and high prevalence of overlapping/touching objects, the model, however, may fail on account of extreme noise and poor (or faint) staining of the target object. A poor EM segmentation result may in turn yield a suboptimal initialization to the active contour model. Note that our model works in the RGB space. Results (not shown) in the hue, saturation, and intensity space did not show any significant improvement. In addition, the gray-scale scene derived from the R, G, and B values worked very well with the active contour model. Experimental results have shown that our new EMaGACOR model performs significantly better, both in terms of detection and segmentation, compared to EMaGAC and GAC models over a total of 100 images.

The ability to automatically and accurately segment lymphocytes in BC histopathology images may serve to be useful in studying LI and its relation to BC prognosis [20]. The relationship between the presence of LI and good prognosis has also been observed in other pathologies such as ovarian cancer [46]. Hence, the algorithm developed in this study might also be similarly employed in prognosis prediction of other diseases. In future work, we intend to apply EMaGACOR to other biomedical image analysis problems.

## REFERENCES

- [1] A. Jemal, R. Siegel, E. Ward, Y. Hao, J. Xu, T. Murray, and M. J. Thun, "Cancer statistics, 2008," *CA Cancer J Clin*, vol. 58, pp. 71–96, 2008.
- [2] G. Alexe, G. S. Dalgin, D. Scanfeld, P. Tamayo, J. P. Mesirov, C. DeLisi, L. Harris, N. Barnard, M. Martel, A. J. Levine, S. Ganesan, and G. Bhanot, "High expression of lymphocyte-associated genes in node negative her2+

- breast cancers correlates with lower recurrence rates," *Cancer Res.*, vol. 67, pp. 10 669–10 676, 2007.
- [3] L. Latson, B. Sebek, and K. A. Powell, "Automated cell nuclear segmentation in color images of hematoxylin and eosin-stained breast biopsy," *Anal. Quant. Cytol. Histol.*, vol. 25, pp. 321–331, 2003.
  - [4] S. Petushi, F. U. Garcia, M. M. Haber, C. Katsinis, and A. Tozeren, "Large-scale computations on histology images reveal grade differentiating parameters for breast cancer," *BMC Med. Imag.*, vol. 6, p. 14, 2006.
  - [5] S. Essafi, R. Doughri, S. M'hiri, K. B. Romdhane, and F. Ghorbel, "Segmentation and classification of breast cancer cells in histological images," in *Proc. 2nd Inf. Commun. Technol. ICTTA 2006*, vol. 1, pp. 1097–1102.
  - [6] S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski, "Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features," in *Proc. 5th IEEE Int. Symp. Biomed. Imag.: Nano Macro (ISBI 2008)*, 14–17, pp. 496–499.
  - [7] J. P. Monaco, J. E. Tomaszewski, M. D. Feldman, M. Moradi, P. Mousavi, A. Boag, C. Davidson, P. Abolmaesumi, and A. Madabhushi. (2008). "Detection of prostate cancer from whole-mount histology images using Markov random fields," in *MIAAB* [Online]. Available: <http://www.miaab.org/miaab-2008-papers/28-miaab-2008-paper-22.pdf>
  - [8] S. A. Hojjatoleslami and J. Kittler, "Region growing: A new approach," *IEEE Trans. Image Process.*, vol. 7, no. 7, pp. 1079–1084, Jul. 1998.
  - [9] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," in *Proc. 5th Int. Conf. Comput. Vis.*, 20–23 Jun., 1995, pp. 694–699.
  - [10] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. V1, no. 4, pp. 321–331, Jan. 1988.
  - [11] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 359–369, Mar. 1998.
  - [12] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.
  - [13] C. Duan, S. Bao, H. Lu, and J. Lu, "Robust automatic segmentation of cell nucleus using multi-scale space level set method," in *Proc. Med. Imag. Inf.: 2nd Int. Conf. (MIMI 2007), Beijing, China, Aug. 14–16*, pp. 80–88. Revised selected papers 2008.
  - [14] C. Li, C. Xu, C. Gui, and M. Fox, "Level set evolution without re-initialization: A new variational formulation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR 2005)*, Jun. 20–25, vol. 1, pp. 430–436.
  - [15] N. Li, M. Liu, and Y. Li, "Image segmentation algorithm using watershed transform and level set method," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP 2007)*, vol. 1, pp. I-613–I-616.
  - [16] D. Mukherjee, N. Ray, and S. Acton, "Level set analysis for leukocyte detection and tracking," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 562–572, Apr. 2004.
  - [17] X. Xie and M. Mirmehdi, "Mac: Magnetostatic active contour model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 4, pp. 632–646, Apr. 2008.
  - [18] A. Hafiane, F. Bunyak, and K. Palaniappan, "Fuzzy clustering and active contours for histopathology image segmentation and nuclei detection," in *Proc. ACIVS*, 2008, pp. 903–914.
  - [19] S. Naik, S. Doyle, S. Agner, A. Madabhushi, M. Feldman, and J. Tomaszewski, "Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology," in *Proc. 5th IEEE Int. Symp. Biomed. Imag.: Nano Macro (ISBI 2008)*, pp. 284–287.
  - [20] A. Basavanahally, S. Ganesan, S. Agner, J. Monaco, G. Bhanot, and A. Madabhushi, "Computerized image-based detection and grading of lymphocytic infiltration in her2+ breast cancer histopathology," *IEEE Trans. Biomed. Eng.*, to be published.
  - [21] C. Lee, W. Snyder, and C. Wang, "Supervised multispectral image segmentation using active contours," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA 2005)*, Apr. 18–22, pp. 4242–4247.
  - [22] W. F. Clocksin, "Automatic segmentation of overlapping nuclei with high background variation using robust estimation and flexible contour models," in *Proc. 12th Int. Conf. Image Anal. Process.*, Sep. 17–19, 2003, pp. 682–687.
  - [23] K. Nandy, P. Gudla, and S. Lockett. (2007). "Automatic segmentation of cell nuclei in 2d using dynamic programming," in *Proc. 2nd Workshop Microscopic Image Anal. Appl. Biol., Piscataway, NJ*, D. Metaxas, J. Rittscher, S. Lockett, and T. Sebastian, Eds. [Online]. Available: <http://www.miaab.org/miaab-2007-papers.html>
  - [24] F. Cloppet and A. Boucher, "Segmentation of overlapping/aggregating nuclei cells in biological images," in *Proc. 19th Int. Conf. Pattern Recognit. (ICPR 2008)*, Dec. 8–11, pp. 1–4.
  - [25] X. He and Q. Liao. (2008). "A novel shape prior based segmentation of touching or overlapping ellipse-like nuclei," in *Medical Imaging 2008: Image Processing*, J. M. Reinhardt and J. P. W. Pluim, Eds., vol. 6914, no. 1 [Online]. Available: <http://link.aip.org/link/?PSI/6914/69141T/1>
  - [26] W. Fang and K. L. Chan, "Incorporating shape prior into geodesic active contours for detecting partially occluded object," *Pattern Recognit.*, vol. 40, no. 8, pp. 2163–2172, 2007.
  - [27] H. Fatakdawala, A. Basavanahally, J. Xu, G. Bhanot, S. Ganesan, M. Feldman, J. Tomaszewski, and A. Madabhushi, "Expectation maximization driven geodesic active contour with overlap resolution (emagacor): Application to lymphocyte segmentation on breast cancer histopathology," in *Proc. IEEE Int. Symp. Bioinf. Bioeng.*, 2009, pp. 69–76.
  - [28] Y. Raja, S. J. McKenna, and S. Gong, "Segmentation and tracking using color mixture models," in *Proc. Third Asian Conf. Comput. Vision—Vol. I (ACCV 1998)*. London, U.K.: Springer-Verlag, 1998, pp. 607–614.
  - [29] M. Shamsi, R. A. Zoroofi, C. Lucas, M. S. Hasanabadi, and M. R. Alsharif, "Automatic facial skin segmentation based on em algorithm under varying illumination," *IEICE—Trans. Inf. Syst.*, vol. E91-D, no. 5, pp. 1543–1551, 2008.
  - [30] Y. Zhang, J. Brady, and S. Smith, "Segmentation of brain mr images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.
  - [31] H. Greenspan, A. Ruf, and J. Goldberger, "Constrained gaussian mixture model framework for automatic segmentation of mr brain images," *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1233–1245, Sep. 2006.
  - [32] A. Diplaros, N. Vlassis, and T. Gevers, "A spatially constrained generative model and an em algorithm for image segmentation," *IEEE Trans. Neural Netw.*, vol. 18, no. 3, pp. 798–808, May 2007.
  - [33] K. Blekas, A. Likas, N. P. Galatsanos, and I. E. Lagaris, "A spatially constrained mixture model for image segmentation," *IEEE Trans. Neural Netw.*, vol. 16, no. 2, pp. 494–498, Mar. 2005.
  - [34] H. Yuan, X.-P. Zhang, and L. Guan, "Content-based image retrieval using a gaussian mixture model in the wavelet domain," in *Visual Communications and Image Processing 2003*, T. Ebrahimi and T. Sikora, Eds. vol. 5150, no. 1Bellingham, WA: SPIE, 2003, pp. 422–429.
  - [35] M. Najjar, C. Ambroise, and J. P. Cocquerez, "Image retrieval using mixture models and em algorithm," in *Image Analysis*. vol. 2749/2003. Berlin, Germany: Springer, 2003, pp. 335–344.
  - [36] D. Cremers, M. Rousson, and R. Deriche, "A review of statistical approaches to level set segmentation: Integrating color, texture, motion and shape," *Int. J. Comput. Vis.*, vol. 72, no. 2, pp. 195–215, Apr. 2007.
  - [37] T. Brox and J. Weickert, "Level set segmentation with multiple regions," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 3213–3218, Oct. 2006.
  - [38] A. Herbulot, S. Jehan-Besson, S. Duffner, M. Barlaud, and G. Aubert, "Segmentation of vectorial image features using shape gradients and information measures," *J. Math. Imag. Vis.*, vol. 25, no. 3, pp. 365–386, 2006.
  - [39] D. Chetverikov, "A simple and efficient algorithm for detection of high curvature points in planar curves," in *Proc. Comput. Anal. Images Patterns*, 2003, pp. 746–753.
  - [40] L. Yang, O. Tuzel, P. Meer, and D. J. Foran, "Automatic image analysis of histopathology specimens using concave vertex graph," in *Proc. Med. Image Comput. Comput. Assist. Interv. Int. Conf. Med. Image Comput. Assist. Interv.*, 2008, vol. 11, pp. 833–841.
  - [41] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *J. R. Stat. Soc. Series B (Methodol.)*, vol. 39, no. 1, pp. 1–38, 1977.
  - [42] C. M. Bishop, *Neural Networks for Pattern Recognition*. London, U.K.: Oxford Univ. Press, 1995.
  - [43] J. Sethian, *Level Set Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*. Cambridge, U.K.: Cambridge Univ. Press, 1996.
  - [44] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*. Boston, MA: Addison-Wesley/Longman, 1992.
  - [45] A. Madabhushi and D. N. Metaxas, "Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions," *IEEE Trans. Med. Imag.*, vol. 22, no. 2, pp. 155–169, Feb. 2003.
  - [46] L. Zhang, J. R. Conejo-Garcia, D. Katsaros, P. A. Gimotty, M. Massobrio, G. Regnani, A. Makrigiannakis, H. Gray, K. Schlienger, M. N. Lieberman, S. C. Rubin, and G. Coukos, "Intratumoral t cells, recurrence, and survival in epithelial ovarian cancer," *N. Engl. J. Med.*, vol. 348, no. 3, pp. 203–213, 2003.



**Hussain Fatakdaawala** received the B.S. degree in biomedical engineering from Rutgers, the State University of New Jersey, Newark, in 2009.

Mr. Fatakdaawala is the recipient of the Samuel Welkowitz Award for excellence in biomedical engineering as an Undergraduate Student and the Aresty Undergraduate Research Grant. He was also a Vice President of the Biomedical Engineering Society (BMES) at Rutgers (2008 and 2009).



**Shridar Ganesan** received the A.B. degree in chemistry from Princeton University, Princeton, NJ, in 1985, and the Ph.D. degree in cell biology and the M.D. degree from Yale University, New Haven, CT, in 1993.

He trained in internal medicine at Brigham and Women's Hospital. He was a Medical Oncology Fellow at the Dana-Farber Cancer Institute. He is currently an Assistant Professor of medicine at Robert Wood Johnson Medical School, New Brunswick, NJ. His current research interests include the molecular characterization of breast cancer subtypes.



**Jun Xu** received the M.S. degree in applied mathematics from the University of Electronic Science and Technology of China, Hefei, China, in 2004, and the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2007.

He is currently a Postdoctoral Associate in the Laboratory for Computational Imaging and Bioinformatics (LCIB), Department of Biomedical Engineering, Rutgers University, Piscataway, NJ. His research interests include medical image analysis, computer vision, machine learning, and their applications to

cancer detection and diagnosis. He is an author or coauthor of more than ten papers in leading international journals and conferences.

Dr. Xu is an author or coauthor of articles in various international journals such as the IEEE TRANSACTIONS ON NEURAL NETWORKS, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-I, NEUROCOMPUTING, the IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE.



**Michael Feldman** is currently an Associate Professor in the Department of Pathology and Laboratory Medicine, Hospital of the University of Pennsylvania, Philadelphia, where he is also the Assistant Dean for the Information Technology and the Medical Director of the Pathology Informatics. His current research interests include the development, integration, and adoption of information technologies in the discipline of pathology, especially in the field of digital imaging.



**Ajay Basavanahally** received the B.S. degree in biomedical engineering from Case Western Reserve University, Cleveland, OH, in 2007.

He is currently a Graduate Assistant in the Laboratory for Computational Imaging and Bioinformatics (LCIB), Department of Biomedical Engineering, Rutgers University, Piscataway, NJ. His current research interests include the development of computer-aided diagnosis and prognosis tools for breast cancer histopathology.



**John E. Tomaszewski** is currently a Professor and an Interim Chair of Pathology and Laboratory Medicine, the University of Pennsylvania, Philadelphia. He is a nationally recognized expert in diagnostic genitourinary pathology. His current research interests include high-resolution MRI of prostate and breast cancer, computer-assisted diagnosis, and high dimensionality data fusion in the creation of a new diagnostic testing paradigms.



**Gyan Bhanot** received the Ph.D. degree in theoretical particle physics from Cornell University, Ithaca, NY, in 1979.

He is currently a Professor in the Department of Biology and Biochemistry and the Department of Physics, Rutgers University, New Brunswick, NJ. After Postdoctoral appointments at Brookhaven National Laboratory, The European Center for Nuclear Physics (CERN) in Geneva, Switzerland, IAS Princeton, and ITP Santa Barbara, he was appointed as Associate Professor in the Physics Department, Florida

State University. From 1994 to 2006, he was a Research Staff Member in the Physics Department, IBM Research, where he was involved in various projects involving high-performance computing. He is a member of the Cancer Institute of New Jersey and a long-term Visiting Member at the Simons Center for Systems Biology, Institute for Advanced Study in Princeton.

Dr. Bhanot is the recipient of the Gordon Bell Award for application performance on the IBM Blue Gene/L "the fastest computer in the world at the time."



**Anant Madabhushi** received the B.S. degree in biomedical engineering from Mumbai University, Mumbai, India, in 1998, the M.S. degree in biomedical engineering from the University of Texas, Austin, in 2000, and the Ph.D. degree in bioengineering from the University of Pennsylvania, Philadelphia, in 2004.

Since 2005, he has been an Assistant Professor in the Department of Biomedical Engineering, Rutgers University, Piscataway, NJ, where he is also the Director of the Laboratory for Computational Imaging and Bioinformatics (LCIB). He is also a member of

the Cancer Institute of New Jersey and an Adjunct Assistant Professor of radiology at the Robert Wood Johnson Medical Center, NJ. He has authored or coauthored 60 publications and book chapters in leading international journals and peer-reviewed conferences. He has several patents pending in the areas of medical image analysis, computer-aided diagnosis, machine learning, and computer vision.

Dr. Madabhushi is also the recipient of a number of awards for both research as well as teaching, including the Coulter Phase 1 and Phase 2 Early Career award (2006, 2008), the Excellence in Teaching Award (2007–2009), the Society for Imaging Informatics in Medicine (SIIM) New Investigator award (2008), and the Life Sciences Commercialization Award (2008). In addition his research work has also received grant funding from the National Cancer Institute (NIH), New Jersey Commission on Cancer Research, and the Department of Defense.