Spatial Statistics for Segmenting Histological Structures in H&E Stained Tissue Images

Luong Nguyen, Akif Burak Tosun, Jeffrey L. Fine,
Adrian V. Lee, D. Lansing Taylor, and S. Chakra Chennubhotla*

***Abstract*— Segmenting a broad class of histological structures in transmitted light and/or fluorescence-based images is a prerequisite for determining the pathological basis of cancer, elucidating spatial interactions between histological structures in tumor microenvironments (e.g., tumor infiltrating lymphocytes), facilitating precision medicine studies with deep molecular profiling, and providing an exploratory tool for pathologists. This paper focuses on segmenting histological structures in hematoxylin- and eosin-stained images of breast tissues, e.g., invasive carcinoma, carcinoma *in situ*, atypical and normal ducts, adipose tissue, and lymphocytes. We propose two graph-theoretic segmentation methods based on local spatial color and nuclei neighborhood statistics. For benchmarking, we curated a data set of 232 high-power field breast tissue images together with expertly annotated ground truth. To accurately model the preference for histological structures (ducts, vessels, tumor nets, adipose, etc.) over the remaining connective tissue and non-tissue areas in ground truth annotations, we propose a new region-based score for evaluating segmentation algorithms. We demonstrate the improvement of our proposed methods over the state-of-the-art algorithms in both region- and boundary-based performance measures.**

***Index Terms*— Histopathological image analysis, image segmentation, evaluation metrics, graph partitioning, image statistics.**

Manuscript received January 4, 2017; accepted March 7, 2017. Date of publication March 16, 2017; date of current version June 29, 2017. This work was supported in part by the PA Department of Health SAP (DLT) under Grant 4100054875, in part by the NIH Cancer Center-Chemical Biology Facility (DLT) under Grant P30CA047904, in part by the Breast Cancer Research Foundation (AVL), in part by the internal Pitt-GE Global Research (AVL, DLT) under Grant 824, in part by NHGRI through BD2K initiative (SCC) under Grant U54HG008540, in part by the UPMC Center for Commercial Applications of Healthcare Data (DLT, SCC) under Grant 711077, and in part by NIH/NCI (SCC, DLT) under Grant U01CA20482601. *Asterisk indicates corresponding author.*

L. Nguyen and A. B. Tosun are with the Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA 15213 USA (e-mail: lun5@pitt.edu; tosun@pitt.edu).

J. L. Fine is with the Department of Pathology, University of Pittsburgh, Pittsburgh, PA 15213 USA, and also with the Magee-Womens Hospital of UPMC, Pittsburgh, PA 15213 USA (e-mail: finejl@upmc.edu).

A. V. Lee is with the Department of Pharmacology and Chemical Biology, University of Pittsburgh, Pittsburgh, PA 15213 USA, and also with the Women's Cancer Research Center, University of Pittsburgh Cancer Institute and Magee-Womens Research Institute, Pittsburgh, PA 15213 USA (e-mail: leeav@upmc.edu).

D. L. Taylor is the Director of the University of Pittsburgh Drug Discovery Institute and Allegheny Foundation Professor of Computational & Systems Biology, University of Pittsburgh, PA 15213 USA (email: dltaylor@pitt.edu).

*S. C. Chennubhotla is with the Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA 15213 USA (e-mail: chakracs@pitt.edu).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the author.

Digital Object Identifier 10.1109/TMI.2017.2681519

I. Introduction

HISTOLOGICAL structure determination helps elucidate spatial tumor biology and inform the pathological basis of cancer. For example, in the hematoxylin and eosin (H&E) stained tissue image shown in Fig. 1 (left), normal ducts (top left corner) have two layers of nuclei, epithelial (inner) and myoepithelial (outer), surrounding a cavity (lumen). The structure of the normal ducts is disturbed when ducts develop into carcinoma in situ (outlined in green), in which the epithelial nuclei proliferate in close proximity and fill the lumen cavity. The structure is further perturbed as the proliferating cells become invasive carcinoma (outlined in blue), destroying the duct confinement, freely infiltrating into the breast stroma, and heading toward a blood vessel (outlined in teal), indicating an increased risk of metastasis. Host response to invasive carcinoma can be seen in Fig. 1 (right), where the tumor nest is infiltrated with lymphocyte nuclei (small, dark purple). For other histological structures such as adipose tissue, the nuclei are small and found on one side of the cell wall surrounding large lipid droplets (white blobs). Accurate segmentation of histological structures can thus help build a spatial interaction map to serve as an exploratory tool for pathologists [1]. Segmentation can also facilitate precision medicine studies which perform microdissection for deep molecular profiling [2].

Histological structure segmentation is very challenging because structures such as normal ducts and carcinoma in situ have well-defined boundaries, but many others, invasive carcinoma and stroma for example, do not. Structural morphologies also vary significantly depending on tissue origins (e.g., breast vs lung), tissue preparation and staining practices. Historically, biomedical image analysis literature has focused on segmenting nuclei, since nuclei are building blocks for all higher level tissue structures [3], [4] (e.g. a duct is a hollow structure lined by rounded epithelial cells surrounding a lumen). More recent methods have expanded to segmenting specific histological structures, such as the glands in prostate and breast tissue images, ductal carcinoma in situ in breast tissue images [5], with approaches based on nuclei-lumen association [6], region growth [7], region-based active contour in combination with Markov Random Field [8], deep learning [9], and graph-based techniques [5].

Other approaches involve engineering disease- and organ-specific extractors [10], [11] to facilitate analysis of publicly available datasets, such as MITOS (mitotic figures) and GlaS (glands) [12]. For example, a typical gland segmentation strategy may involve first identifying lumen and then searching for the surrounding epithelial layer of cells [13]. However, this

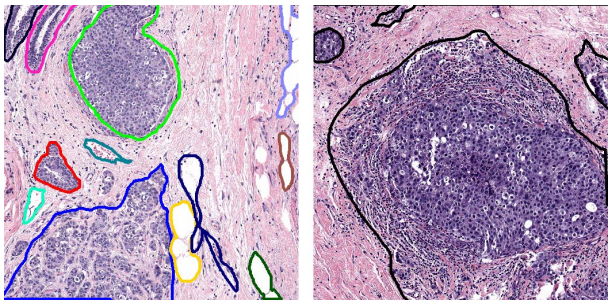

Fig. 1. **Broad class of histological structures found in breast tissue images.** These include: (Left) two normal ducts outlined in magenta and black; ductal carcinoma in situ outlined in green; a large nest of invasive carcinoma outlined in blue; adipose tissue (fat cells) outlined in different colors; (Right) a large tumor nest with infiltrating lymphocytes outlined.

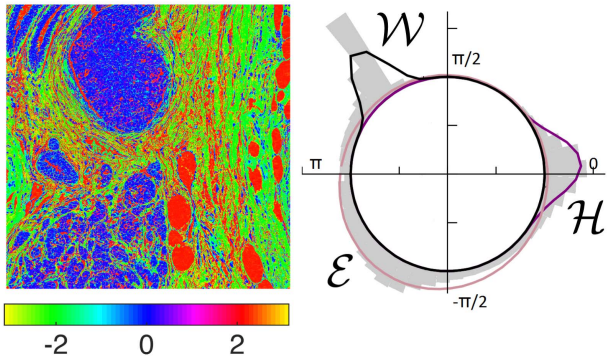

Fig. 2. **H&E optimized color space.** The H&E image from Fig. 1 (left) is transformed into H&E-hue space and shown as a heatmap (left) and as an angular histogram (right). Hematoxylin stained pixels map into a distribution of angular values centered around 0 radians and are shown in dark blue in the heatmap. Eosin stained pixels map into a distribution of angular values widely spread around -1.7 radians and are shown in yellow-green colors in the heatmap. White pixels map into a distribution of angular values narrowly spread around 2.24 radians and are shown in red in the heatmap. We model the hue values as a mixture of three univariate von Mises distributions ($\mathcal{H}$, $\mathcal{E}$, and $\mathcal{W}$, see text) and a uniform distribution ($\mathcal{O}$, not shown) to account for other pixels (black tissue folds, red blood cells).

strategy is unlikely to work in the case of breast carcinoma in situ, where the duct lumens may be completely filled by tumor cells (structure outlined in green in Fig. 1, left). Generative models have been proposed for foreground-background segmentation for H&E images with one dominant boundary [14]. Our goal in this paper is to go beyond segmenting nuclei and glands by designing broadly applicable methods for extracting a large class of histological structures.

Contributions: Our paper focuses on segmenting histological structures in H&E stained images of breast tissues (Fig. 1), in which hematoxylin stains nuclei to bluish-purple colors, and eosin stains cytoplasm and the stroma matrix to red-pink colors. We hypothesize that spatial image statistics present discriminative fingerprints for segmenting a broad class of histological structures. To test this, we propose two graph-theoretic segmentation methods, each of which relies on characterizing local spatial statistics. In the first method (II-B), we measure pairwise pixel color statistics in an H&E optimized color space built to enhance the separation between hematoxylin and eosin stains. We expect the first method to be successful in segmenting structures with well-defined boundaries (e.g., adipose tissues, blood vessels). The second method (II-C) is designed to segment large amorphous histological structures (e.g., tumor nests), where we rely on the spatial statistics of inter-nuclei distances.

To benchmark our proposed H&E image segmentation algorithms, we curated a dataset of 232 breast tissue images of size 2K $\times$ 2K pixels extracted from whole slide images (WSIs) scanned with Aperio XT ® at 20$\times$ objective magnification (0.5μm/pixel) resolution (Leica Biosystems Inc., Buffalo Grove IL, USA). The segmentation ground truth was marked by two pathology trainees and corrected by a practicing breast pathologist. Since we focus on segmenting a broad class of histological structures, we chose to compare the performance of our methods to other state-of-the-art algorithms adapted from the field of natural images to the domain of H&E. Many of these methods have rich feature vocabularies (based on color, textures, etc.). In comparison, our methods use only one feature that captures spatial statistics for segmentation although more features can be integrated. We also compared our methods to publicly available H&E segmentation algorithms, one of which is gland-specific, and hence, provides baseline results for benchmarking. Finally, we present a new region based metric for evaluating the segmentation algorithms.

II. Methodology

The two segmentation methods proposed have few common elements, namely the H&E optimized color representation (II-A1) and appearance normalization (II-A2). The segmentation methods differ in the way they capture image statistics and embed them into graph partitioning strategies, and hence they will be described separately in Sections II-B and II-C.

A. H&E Color Preprocessing

1) H&E Optimized Color Space: We begin with the observation that when the standard opponent-color (with red-green, yellow-blue as opponent color axes) hue-saturation-brightness (HSV) transformation is applied to RGB images from H&E [3], [15], hematoxylin and eosin stain colors are *restricted* to blue-red quadrant of the color wheel. Our goal is to enhance the separation between the color appearances of the two stains so that the downstream spatial analysis pipeline is more robust (Fig. 2). It is worth noting that the nuclei of the various cell types stain with different intensities of hematoxylin. An additional goal of this color representation is to homogenize the variations within the hematoxylin and eosin stain colors and thus, facilitate the extraction of histological structures as in Fig. 1, which extend beyond subcellular components (e.g. nuclei).

For this, we optimize the construction of a color space to maximally separate the hematoxylin and eosin appearances. Specifically, we asked an expert to select a bag of dominantly stained hematoxylin and eosin pixels. In the future, this process can be automated by calculating stain vectors for hematolyxin and eosin over a population of H&E slides and using these vectors to derive representative pixels [15], [16].

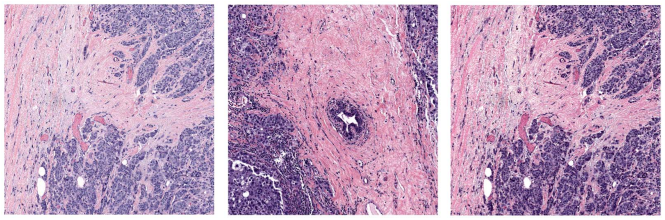

Fig. 3. **Color appearance normalization of H&E images.** A source image (left) is normalized (right) to have the same color appearance as the target image (center).

After representative pixel selection, we perform singular value decomposition on this data to obtain an orthogonal projection matrix of size 3×3. We offer a specific interpretation to the projected coordinates, similar to the HSV space. In particular, the projection onto the first singular vector (enforced to have non-negative values) yields an H&E-brightness value b. The two remaining projected coordinates c_2 and c_3 form a complex plane where we define H&E-saturation $s = \sqrt{c_2^2 + c_3^2}$ and H&E-hue, an angular variable $\theta = \tan^{-1}(c_2 + ic_3)$. From this construction, we expect the hue values of hematoxylin and eosin stained pixels to be maximally separated in the complex color plane. For illustration, the angular difference in the mean hue values of pixels stained with the two stains in Fig. 1 is 1.7 radians (Fig. 2). This spread is more than the value of ≈ 0.4 radians in the standard HSV color space. We observe that while the fibroblast nuclei appear darker than the epithilial nuclei in the RGB space (Fig. 1), this difference is less pronounced in the H&E-hue space, as indicated by the narrow spread of hematoxylin stained pixels' angular values around 0 radian (Fig. 2, right).

Hue value is unstable when the saturation is low. This is true for pixels mapped to the origin of the complex plane ($c_2, c_3 \approx 0$). In the standard HSV representation, *all* white pixels will have low saturation values and hence unstable hue angles [15]. Note that white pixels can form a significant portion of an H&E image because of adipose tissue, lumen, tissue tears, and shrinkages. In our color representation, by learning the rotation matrix from expert-selected representative pixels, we are able to give white pixels higher saturation values and more stable hue angles. However, there will be a population of pixels with low saturation values (< 0.005) that map to the origin of the complex plane. We empirically estimated this population to be around 0.3% for the H&E images of size $2K \times 2K$ that we used here. On closer inspection, these pixels lie around the boundaries of histological structures such as ducts or tumor nests. However, we do not hard code this empirical estimation but instead account for this minority population in a color spatial statistics model explicitly as described in the next section.

2) H&E Color Appearance Normalization: Any inconsistencies in sectioning, staining, and imaging result in variation in color appearance of H&E images. Previous normalization methods have utilized stain vector estimation methods [17] such as non negative matrix factorization [15]. We found these methods ineffective because the color distributions for some images in our dataset are skewed toward predominantly one stain, either hematoxylin or eosin.

We hypothesize that the color appearances of two images are similar if their color statistics match. However, matching the statistics of the whole pixel population of the source and target images as done in [18] can result in unintended artifacts. For example, if the source image has mostly stroma (eosin stained) and the target image has mostly nuclei clumps (hematoxylin stained), then matching the statistics of all pixels will turn many eosin stained pixels in the source image to bluish-purple and mistakenly change the cellular component identity of those pixels from stroma to nuclei. To address this issue, we first identify four classes of pixels: $\mathcal{H}$ (nuclei), $\mathcal{E}$ (stroma, cytoplasm), $\mathcal{W}$ (fat, shrinkage), and $\mathcal{O}$ (others: pixels with low saturation values, pixels from red blood cells and tissue folds).

To identify the four classes, we convert H&E images into H&E-hue, H&E-saturation, and H&E-brightness channels introduced in II-A1. H&E hue space is angular and given the separation between hematoxylin-stained, eosin-stained, and white pixels in this space, we model the hue values with a mixture of univariate von Mises distributions. Univariate von Mises distribution for angular statistics is the equivalent counterpart of the univariate normal distribution for linear statistics. The von Mises distribution is characterized by two parameters, a mean $-\pi < \mu \leq \pi$ and a concentration parameter $\kappa > 0$, and is given by: $f(x) = \{2\pi I_0(\kappa)\}^{-1} \exp \kappa \cos(x - \mu)$, where $I_0(\kappa)$ is the modified Bessel function of the first kind with order 0 [19]. A mixture of K univariate von Mises distributions is given by $\sum_{k=1}^{K} m_k f(x|\mu_k, \kappa_k)$, where m_k's are the prior probabilities and μ_k's, κ_k's are the means and concentration parameters. To explicitly account for pixels with low saturation values and unstable hue angles, as well as pixels arising from black tissue folds and red blood cells ($\mathcal{O}$ class), we add a uniform angular distribution as an additional mixture component. The parameters of the mixture model, including the prior probability of the uniform distribution are found using an expectation-maximization (EM) algorithm [19]. In practice, we found that the prior probability of the uniform distribution converges to a value in the range (0.001, 0.01), depending on the structural content of the H&E images.

To normalize color appearances, we match the statistics of the source and target images. The statistics of a distribution can be characterized by an infinite set of moments. However, for analytical convenience, we compute moments only up to the fourth order (mean, standard deviation, skewness, kurtosis). In each channel, we match the moments of pixels from $\mathcal{H}$ and $\mathcal{E}$ classes from the source image to the target image [20]. We do not modify the statistics for pixels in the $\mathcal{W}$ and $\mathcal{O}$ classes because the H&E-optimized color space is not suitable for these classes and hence will introduce artifacts (e.g. white pixels can turn into gray). After normalizing the statistics in the H&E optimized color space, we convert the resulting pixel values into RGB space to obtain the normalized image, using the inverse of the rotation matrix derived in II-A1 (Fig. 3).

Normalized images will serve as inputs to two different segmentation strategies based on (1) spatial color statistics

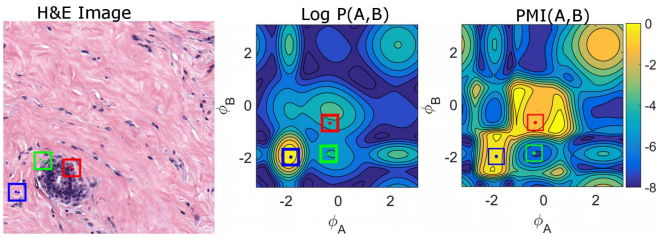

Fig. 4. **Pointwise mutual information (PMI)**. An example H&E image where PMI might be more informative than joint probability in detecting object boundaries (see text for more details). H&E image (left) shows three randomly sampled pixel pairs (red, green, and blue) and their mapping into joint probability space $\log P(A, B)$ (middle) and PMI space (right).

and (2) inter-nuclei distance distributions, as detailed in the following subsections.

B. Spatial Color Statistics Based Segmentation

1) Motivation: Fig. 4 shows a normal breast tissue with a large area of eosin stained connective tissue surrounding a small duct (lower left). The nuclei in the duct are stained with hematoxylin, while their cytoplasm exhibits a mixture of both stains, since the hematoxylin from the nuclei can leak into the cytoplasm. Statistically speaking, if we stand on any of these nuclei (red box, Fig. 4), we expect to be surrounded by hematoxylin stained pixels denoting the nuclei and mixed hematoxylin-eosin stained pixels denoting the cytoplasm. In a given neighborhood of each cell in the duct, we should find other cells exhibiting similar properties. On the other hand, if we stand on a fibroblast nucleus, which is found usually scattered in the connective tissue (blue box, Fig. 4), we will find mostly eosin stained pixels in its neighborhood. With the assumption that the spatial color statistics within a structure, such as ducts (red box, Fig. 4), is stronger than across its boundaries (green box, Fig. 4), we should be able to segment ducts while ignoring the fibroblast cells scattered among the connective tissue.

2) Modeling Spatial Color Statistics: As described in subsection II-A2, using a mixture of univariate von Mises, we can separate the image pixels into four classes, but this is insufficient to delineate histological structures, such as clusters of ducts, because such structures contain pixels from all classes (Fig. 1). In order to segment these structures, we assume that the spatial color statistics within a structure such as ducts is higher than across its boundaries and we show how to model this spatial color statistics using a mixture of bivariate von Mises distributions.

Since the H&E-hue is a angular variable, the joint distribution $P(A, B)$ of hue values from two neighboring pixels lies on a torus. We will model this joint density as a mixture of bivariate von Mises distributions. Let the values of pixel A and B in H&E-hue space be ϕ and ψ, respectively. The bivariate distribution of two angular variables, $-\pi < \phi \leq \pi$ and $-\pi < \psi \leq \pi$ is:

$$f_c(\phi, \psi) = C_c \exp\{\kappa_1 \cos(\phi - \mu) + \kappa_2 \cos(\psi - \nu) - \kappa_3 \cos(\phi - \mu - \psi + \nu)\}, \quad (1)$$

where μ, ν are the means and $\kappa_1, \kappa_2 > 0$ are the concentrations of ϕ, ψ, respectively. κ_3 is the correlation coefficient and C_c is the normalizing constant. The full bivariate von Mises model has 8 parameters, but we use a reduced 5-parameter cosine model with positive interaction (Eq. 1), as suggested in [19].

The marginal density is: $f_c(\psi) = C_c 2\pi I_0(\kappa_{13})(\psi) \exp\{\kappa_2 \cos(\psi - \nu)\}$. The value of κ_3 decides whether the distribution is unimodal or bimodal. In particular, the joint density is unimodal if $\kappa_3 < \kappa_1\kappa_2/(\kappa_1 + \kappa_2)$ and it is bimodal if $\kappa_3 > \kappa_1\kappa_2/(\kappa_1 + \kappa_2)$ when $\kappa_1 > \kappa_3 > 0$ and $\kappa_2 > \kappa_3 > 0$.

When we consider the values of neighboring pixels of the H&E image in the H&E-hue space, there are at most six possibilities for the probability masses on the torus: $\mathcal{H} - \mathcal{H}$, $\mathcal{E} - \mathcal{E}$, $\mathcal{W} - \mathcal{W}$, and the three different pairwise interactions. To model this joint distribution, we use a mixture of six unimodal bivariate von Mises distributions. We also include a uniform distribution to account for the interactions involving pixels in the $\mathcal{O}$ class (as in subsection II-A2).

A mixture model of K bivariate von Mises distributions can be parameterized by: $f_M(\phi, \psi) = \sum_{i=1}^{K} m_i f_i(\phi, \psi | \mu_i, \nu_i, \kappa_{1i}, \kappa_{2i}, \kappa_{3i})$, where m_j is the mixing coefficient ($\sum_{j=1}^{K} m_j = 1$), f_i is a cosine density with parameters $(\mu_i, \nu_i, \kappa_{1i}, \kappa_{2i}, \kappa_{3i})$ (Eq. 1). The initial values of μ_i's, ν_i's, κ_{1i}'s, and κ_{2i}'s are generated from the mixture of univariate von Mises for all the pixels in the image. The concentration parameters κ_{1i}, κ_{2i}'s and the correlation parameters κ_{3i}'s satisfy the unimodality conditions for f_i's. We constrain κ_{3i}'s to have values between -1 and 1 to avoid distortion to the elliptical patterns (observed in sampled data) [19]. Together with the above constraints, the parameters of the mixture are estimated by an EM algorithm [21].

Since there are at most seven components of the mixture model as reasoned above, we do not undertake an explicit model selection step for the mixture model. If the H&E image lacks any one of the four basic pixel classes, ($\mathcal{H}$, $\mathcal{E}$, $\mathcal{W}$, $\mathcal{O}$), the EM algorithm will drive the mixing proportions of all components involving that class to a value near zero.

3) Mutual Information in H&E-Hue Space: Consider modeling the statistical dependencies between hue angles of neighboring pixels in H&E optimized color space. If we use the joint probabilities as a measure of statistical association, we may find that the pixel pair stained with eosin in the connective tissue (CT) has a higher probability than a pixel pair stained with hematoxylin inside a duct or a hematoxylin-eosin pixel pair across the CT-duct boundary. However, because of the overabundance of eosin in some H&E images (Fig. 4, left), the combination of hematoxylin-eosin pixel pairs across the CT-duct boundary may have an equivalent or even higher probability than a pixel pair stained with hematoxylin inside the duct. As shown in Fig. 4 (middle), the pair stained with eosin (in blue square) has the highest joint probability and the pair stained with hematoxylin (in red square) have similar joint probability to the hematoxylin-eosin pair (in green square). In other words, the joint probability might not be sufficient to detect correct boundaries. As suggested by [22], this can be improved by the use of mutual information (MI) to correct for relative abundance.

To compute MI, a number of pixel pairs (i, j), with features $\vec{f}_j$ and $\vec{f}_j$ (e.g., using H&E-hue angles, $\vec{f}_i = \phi_i, \vec{f}_j = \psi_j$) and distance $d_{i,j}$, are selected randomly from the image. The value $d_{i,j}$ is defined as $d_{i,j} = 2 + 2 \times |r|$ where $r \sim \mathcal{N}(0, \sigma)$ (see III-C for the selection of σ). The joint probability of features of pixels i and j at a distance $d_{i,j}$ is denoted by $p(\vec{f}_i, \vec{f}_j; d_{i,j})$. The overall joint probability is defined as: $P(\vec{f}_i, \vec{f}_j) = \frac{1}{Z} \sum_{d_{i,j}=d_0}^{\infty} w(d) p(\vec{f}_i, \vec{f}_j; d_{i,j})$, where w is a Gaussian weighting function which decays to zero as $d_{i,j}$ increases and Z is a normalization constant.

The pointwise mutual information (PMI) is calculated from the joint probability $P(\vec{f}_i, \vec{f}_j)$ modeled by a mixture of bivariate von Mises distribution and the marginal probabilities $P(\vec{f}_i)$ and $P(\vec{f}_j)$ modeled by a mixture of univariate von Mises distributions. In particular, $\text{PMI}_\rho(\vec{f}_i, \vec{f}_j) = \log \frac{P(\vec{f}_i, \vec{f}_j)^\rho}{P(\vec{f}_i) P(\vec{f}_j)}$, where ρ is a free parameter that can be selected to optimize the segmentation on a training set. In this paper, we scanned through multiple values of ρ to find the best setting for our dataset (see III-C). The benefit of PMI as a measure of spatial statistics is shown in Fig. 4. Although the joint density of the pixel pair stained with hematoxylin inside the duct (red square) is similar to that of the hematoxylin-eosin pixel pair across the boundary (green square) (middle), the PMI of the former is markedly higher than that of the latter (right).

4) Graph-Based Spectral Segmentation: We pose the segmentation problem in graph-theoretic terms, where each pixel in the image is a node in the graph and nearby pixels are connected with weights (affinities) denoting the likelihood of grouping two pixels into the same histological structure. We use spectral methods to partition the graph into meaningful components. The success of spectral methods depends largely on the choice of an affinity function. We denote the affinity matrix by W with elements $w_{i,j}$ defined as: $w_{i,j} = e^{\text{PMI}_\rho(\vec{f}_i, \vec{f}_j)}$. To demonstrate the effectiveness of PMI-based affinity function, we pick in Fig. 5 (left) three patches (outlined in red, green, and blue) and compute the affinities between the center pixel in each patch and its neighbors using only H&E-hue as the feature. The heatmaps show affinities in which hotter colors indicate greater affinity values. In the red patch, the pixel stained with hematoxylin in the center has high affinity with its neighboring hematoxylin stained pixels. In the green patch, the center eosin stained pixel is more similar to the eosin stained pixels on the left and less similar to the hematoxylin stained pixels on the right. In the blue patch, the center hematoxylin stained pixel has high affinity to the hematoxylin stained pixels on the right but low affinity to the white pixels on the left. These heatmaps indicate that H&E-hue-based affinity function is a meaningful measure for H&E images. We further establish the effectiveness of H&E-hue-based affinity function by comparing results from our segmentation method to other spectral methods which use richer feature vocabularies (see section III).

We use H&E-hue-based affinity measures as input to a standard spectral graph partitioning method that has been the state-of-the-art for segmenting natural images (see [23] and references therein). From the affinity matrix W, we find eigenpairs $(\vec{v}, \lambda)$ of the generalized system: $(D - W)\vec{v} = \lambda D\vec{v}$

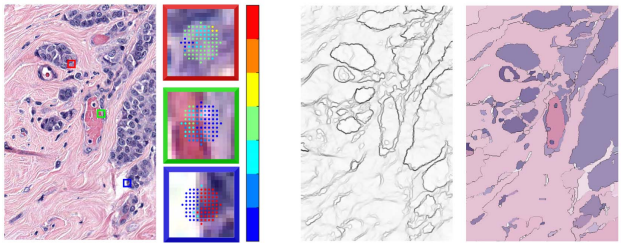

Fig. 5. **Spatial color statistics based segmentation method.** Pairwise pixel affinities are measured from PMI (left) and given as input to standard spectral graph segmentation methods (right). Note that the segmentation method sucessfully groups nuclei into tumor nests instead of separating them into individual nuclei because of the homogenizing characteristic of H&E-hue color space and the effectiveness of the affinity function. The method also successfully segments the blood vessel in the center, but absorbs the blood vessel in the top center into the stroma.

where D is a diagonal matrix with $d_{i,i} = \sum_{j \neq i} w_{i,j}$. Dominant eigenvector maps (small eigenvalues) indicate boundary locations of potential histological structures. As is well known, no single eigenvector will be capable of capturing all possible boundaries in complex images. Hence, the usual practice is to calculate an edge strength map from oriented spatial derivatives of a large number of dominant eigenvectors. A postproccessing step is used to eliminate spurious boundary pixels [23]. In Fig. 5 (right), we show the overall edge strength map (maximum over all orientations of the edge strength map) and the resulting segmentation after postprocessing.

C. Nuclei Neighborhood Statistics Based Segmentation

1) Motivation: Local spatial statistics vary between the various histological structures in breast tissues. We observe this phenomenon in Fig. 1 (left). The spatial arrangement of epithelial nuclei in normal ducts (top left corner) is disturbed as ducts develop into carcinoma in situ (outlined in green) and further damaged as carcinoma becomes invasive (outlined in blue), destroying the duct confinement and infiltrating through breast stroma. As a first approximation, we chose to characterize this spatial arrangement of the epithelial nuclei by the physical distances between pairs of nuclei and consequently developed the method `interNucDist`.

2) Characterizing Nuclei Neighborhoods With Superpixels: Nuclei segmentation in histopathological images is an extensively researched problem [24]–[27]. However, the close proximity of epithelial cells and the prevalence of mitotic figures (dividing cells) in breast cancer make it difficult to accurately detect nuclear boundaries. To avoid this issue, we identify *putative nuclei* locations in the form of *superpixels* in the hematoxylin stained tissue regions ($\mathcal{H}$) and characterize neighborhood statistics using superpixels derived from eosin stained ($\mathcal{E}$) and white ($\mathcal{W}$) tissue regions. Tissue regions in the $\mathcal{O}$ class are ignored due to their small population and being diagnostically not relevant.

In order to generate superpixels from H&E images (Fig. 6, top left), the pixel colors are first normalized as in II-A2. Then, we calculate posterior probability values using the EM algorithm for the three classes $\mathcal{H}$, $\mathcal{E}$ and $\mathcal{W}$

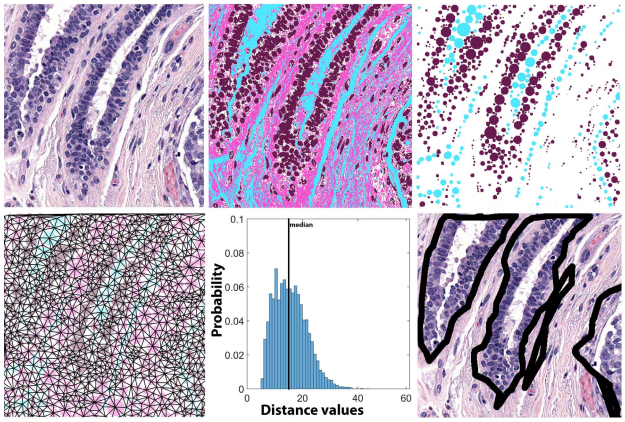

Fig. 6. **Nuclei neighborhood statistics based segmentation.** (TOP) H&E image (left) is decomposed into four separate classes: $\mathcal{H}$, $\mathcal{E}$, $\mathcal{W}$, and $\mathcal{O}$ (not further processed) (middle). Each class is modeled with a set of superpixels here in purple ($\mathcal{H}$), pink ($\mathcal{E}$), and cyan ($\mathcal{W}$) (right). (BOTTOM) Delaunay triangulation provides superpixel neighborhoods (left). Distribution of distances between superpixels of the same class is used for greedy graph partitioning (middle). Superpixels in $\mathcal{H}$ and $\mathcal{W}$ classes are segmented and merged into two dominant ducts (right).

(Fig. 6, top middle) as described in II-A2 and threshold the probabilities to generate ownership masks assigning each pixel to a unique class. Finally, we follow the algorithm proposed in [11] to fit circular shaped superpixels to the ownership masks (Fig. 6, top right). The circular shapes are used for analytical convenience and to serve as a representation for putative nuclei and their neighborhoods.

To characterize nuclei neighborhoods, we combine superpixels from all three classes, $\mathcal{H}$, $\mathcal{E}$ and $\mathcal{W}$, and perform a Delaunay triangulation using their center coordinates (Fig. 6, bottom left) as in [28]. The Delaunay triangulation preserves physical distances and allows us to build and partition graphs separately in each class. Additionally, the graph generated by the triangulation helps avoid the mistake of connecting a fibroblast nucleus with an epithelial nucleus when they are separated by a large area of stroma.

Concurrent to our research, Bejnordi et al. developed a graph based method for detecting DCIS in WSI [5]. However, our methods are different in that: we use superpixels in $\mathcal{H}$, $\mathcal{E}$, $\mathcal{W}$, and $\mathcal{O}$, pixel classes that are not specific to epithelial nuclei (lymphocytes are also accounted for); our methods can segment a broad class of histological structures such as fat clusters, blood vessels, lymphocyte aggregations, tumor nests while their method focuses on ductal carcinoma in situ. In addition, our methods are unsupervised at single scale, while their method involves supervised classification at multiple scales; our methods include a rule for merging histological structures identified in different pixel classes.

3) Segmenting Superpixels Into Histological Structures: Neighborhoods derived from the Delaunay triangulation can be richly characterized [28]. However, in this paper, we chose to test a simple property, namely pairwise distances between superpixels of the same class. In particular, for each class we build a separate graph in which each superpixel is a node, and neighboring superpixels of that class are connected by an edge if their distance is under a threshold τ. We set the distance threshold τ to be at least the median of the distance distribution between neighboring superpixels (shown in Fig. 6, bottom middle). The distance threshold τ is set to maximize the performance of the algorithm on the training set (III-C).

After building the graph over superpixels, we use a greedy connected component analysis algorithm (Matlab built-in function `conncomp`) to cluster superpixels into segments. The superpixels in the $\mathcal{E}$ class are considered as background and hence, are not processed in this step.

Next, we sort the segments in $\mathcal{H}$ and $\mathcal{W}$ classes based on their tissue areas in a descending order. The segments from these classes might overlap with each other because of complex histological structures such as the ducts. We combine segments from these two classes with one simple post-processing rule: if a segment $\mathcal{W}_i$ overlaps with a segment $\mathcal{H}_j$, regardless of the overlapping area, the overlapping part is absorbed into $\mathcal{H}_j$. This rule prioritizes nuclei regions over white areas. For example, it allows us to associate the lumen area with the surrounding epithelial nuclei to form a duct. After merging segments from $\mathcal{H}$ and $\mathcal{W}$ classes, the remaining tissue regions in $\mathcal{E}$ and $\mathcal{O}$ classes are considered as background. The post-procesing rule was established to optimize the segmentation results on our dataset (Fig. 6, bottom right). Although this rule might not be universally applicable, we hope to uncover any missing rules with an expanded set of images and annotations.

III. Experiments and Results

A. Dataset

We collected 30 breast H&E WSIs and scanned them using Aperio XT ® at 20× magnification and 0.5μm pixel resolution. As a quality control measure, we eliminated WSIs with serious artifacts such as air bubbles and pen marks, resulting in a total of 23 WSIs for further analysis. We partitioned these whole slide images into tiles of size 2K × 2K and selected a subset of 232 tiles with tissue appearance ranging from normal to invasive cancer. Some images in the dataset consist mostly of histological structures with clear boundaries, such as normal ducts and carcinoma in situ (Fig. 1, left, magenta, black, and green blobs). Other images are mostly textural because of the severe disruption in tissue architecture caused by invasive cancer (Fig. 1, right). Although tiles were extracted at 20× to cover complex histological structure landscape, many methods chosen for comparison do not scale efficiently to images of size 2K × 2K such as JSEG [29], and NCuts [30]. Therefore, we downsampled the images to size 512 × 512, at 5× magnification and 2μm pixel resolution to ensure a fair comparison of methods. From the 23 WSIs, we chose 17 WSIs and 174 extracted tiles as training set for scanning parameters of all algorithms for comparisons. The remaining 6 WSIs with 58 tiles were used as a test set for evaluating their performance.

B. Ground Truth

Ground truth annotation (Fig. 1) was provided by two pathologist trainees and corrected by a practicing breast pathologist. The annotating software allows users to identify ducts, vessels, carcinoma in situ, tumor nests and other histological structures as *foreground* segments. The remaining

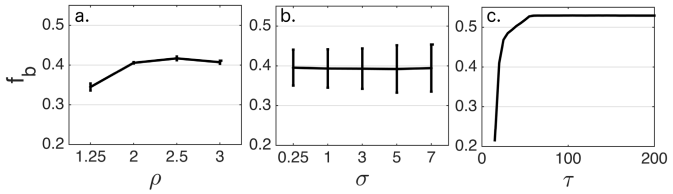

Fig. 7. **Parameter scanning.** a. `colorStats`: F_b vs. ρ. Error bars show 95% confidence interval for different σ; b. F_b vs. σ. Error bars show 95% confidence interval for different ρ's. c. `interNucDist`: F_b vs. τ.

portion of the image, namely stroma and non tissue areas, is considered as *background*. There could be more than one background segment in an image.

Interestingly, while the description of histological structures requires expert knowledge, we observed that non-expert annotators can readily segment H&E images into coherent segments without necessarily giving the segments their appropriate labels. We hypothesize that there is a transfer of knowledge from segmenting natural images (car, tree, road, etc.) to segmenting histological structures in H&E images. Therefore, we recruited two non-experts to annotate a subset of our data (79 images). The non-experts were given instructions to trace boundaries of up to 15 coherent segments per image (Fig. 7). Before starting to annotate, they were each shown five examples of expertly annotated images, but no further training was provided. To speed up the annotation process and to avoid confusions, most of the images annotated by the non-experts are well-differentiated, i.e., histological structures with clear boundaries such as normal ducts or carcinoma in situ. Our goal in collecting these annotations is to understand the performance gap between the expert, non-expert observers, and computational algorithms.

Although there are large publicly available datasets such as TCGA, high quality annotations of histological structures are not available. It is expensive and time consuming to collect such detailed annotations from pathologists. By making our data publicly available, we wish to attract attention of the biomedical imaging community and encourage other researchers and institutions to create high quality and diverse annotated histological datasets. Our data, including H&E tiles, ground truth annotations, and train/test division can be accessed from http://csb.pitt.edu/Faculty/chakra/pubs/TMI_HESegmentation.tar.gz.

C. Algorithms for Comparison

Since our focus is on segmenting a broad class of histological structures, we chose to compare the performance of our methods to other state-of-the-art algorithms adapted from the field of natural images to the domain of H&E. In addition, we also compared our methods to publicly available H&E segmentation algorithms: `GraphRLM` [28] and `GlandSeg` [6]. Although `GlandSeg` is gland-specific, it can provide baseline results for benchmarking. In summary, we evaluated the following algorithms:

Methods proposed in this paper:

- `colorStats`: our proposed color statistics based method using H&E-hue with affinity values derived from PMI. Only one feature (H&E-hue) is used to build the affinities for the spectral method.
- `inNucDist`: our proposed nuclear neighborhood statistics based method using greedy grouping of superpixels with distant thresholds derived from the data. Only one feature, distances between superpixels, is used to calculate connectivity for the graph partitioning method.
- `non-expert`: annotations on a subset of the images by two non-experts.

Methods motivated by histological data

- `GraphRLM` [28]: superpixels are calculated after identifying nuclei, stroma, and lumen pixels in RGB and neighborhoods are characterized by 16 graph run length features. Segments are produced using seed detection and region growing algorithms.
- `GlandSeg` [6]: glands are identified by first segmenting white lumens and then associating them with the surrounding nuclei.

Methods motivated by natural images

- `gPb` (gPb-OWT-UCM) [23]: current state-of-the-art natural image segmentation algorithm that uses 36 color and texton features in a spectral method.
- `crisp-bound` [22]: La*b* color and variances (six features) are used for deriving pair-wise affinities in a spectral segmentation method.
- `JSEG` [29]: color quantization followed by region growing on this map using RGB colors.
- `NCut` [30]: normalized graph cut algorithm using brightness, color, and texture features.
- `MeanShift` [31]: a local color (three features) homogenization method followed by region growing.
- `EGB` [32]: pairwise region comparison and greedy grouping of pixels using five features (RGB, xy-coordinates)
- `HED` [33]: deep learning-based boundary detection method, in which, segmentation problem is viewed as classifying pixels into either boundary or non-boundary.

From the set of 23 high quality WSIs, we took 17 WSIs with 174 extracted tiles for scanning parameters and selecting the best parameter setting for each algorithm. The remaining 6 WSIs with 58 tiles were used as a test set for evaluating the performance using the best parameter settings. The parameter settings for `EGB` (14 settings), `MeanShift` (28), `NCut` (15) are taken from the SEISM package [34]. For `gPb` and `crisp-boundary`, we used the software off-the-shelf to generate ultrametric contour maps [23] and scanned through 50 different thresholds (0.01:0.02:0.99) to obtain segmentation results. For `HED`, we used pretrained model to detect boundaries for all images. Edge thinning on 50 different thresholds (0.01:0.02:0.99) was used to generate the final boundary maps. For `GraphRLM` and `JSEG`, 60 and 56 combinations of parameters were chosen in the ranges described in their respective papers.

In `colorStats`, there are two free parameters, σ and ρ. We scanned through combinations of five values of σ (0.25, 1, 3, 5, 7) and four values of ρ (1.25, 2, 2.5, 3). In `inNuctDist`, we scanned through 38 values of distance threshold τ (15 to 200 pixels with a step of 5 pixels). We select the fifteen largest segments in $\mathcal{H}$ and $\mathcal{W}$ for post-processing in `interNucDist`.

D. Evaluation Metrics

For each of the methods, we select the parameter setting that maximizes the boundary metric F_b on the training set of 174 tiles. The boundary metric F_b [34] refers to F-score for boundary accuracy. Boundary pixels from segmentation results are matched with ground truth boundaries within some distance, from which precision, recall, and F-score are calculated. The best parameter setting based on F_b is used to generate segmentation results on the test set of 58 tiles. In addition, we also calculated the Dice metric as an additional criterion for comparison [34].

Previously, Rand index was chosen as a region based evaluation metric for tissue segmentation algorithms [14]. However, Rand index is applicable to simple images with only one dominant boundary. If the images are more complicated, Rand index tends to over-reward segmentations with small unimportant islands (EGB results, Fig. 8) and fails to emphasize important histological structures such as tumor nests. More importantly, the experts emphasize partitioning the foreground (e.g. ducts, carcinoma in situ) more than the background, as in stroma, because of their diagnostic relevance. Thus, we introduce a region score F_r to more accurately reflect the ground truth preference for foreground segments.

Because the segmentation methods are unsupervised, we do not know which segments are foreground and which are background, unlike the ground truth. Our strategy is to score the background segments of the ground truth first (to account for their larger tissue areas), then score the foreground segments, and finally combine the scores in a preferential weighting scheme that emphasizes the foreground segments.

Let G_b be the sets of pixels representing background segments in the ground truth and S be the set of segments in the segmentation result. For each background segment G_{b_i}, starting with the largest one, we match it with the most overlapping segment S_j in the segmentation result and remove S_j from further consideration. For each match (G_{b_i}, S_j), we assign a region overlapping score given by: $s_{b_i} = \|G_{b_i} \cap S_j\| / \|G_{b_i} \cup S_j\|$, where $\cap$ denotes intersection of the two sets and $\cup$ denotes their union. The overall background score s_b is given by the sum of s_{b_i}, each weighted by the fractional area of background segment G_{b_i}. In the case of images that have no background, such as the ones of invasive carcinoma, the score s_b is handled by a weighting term, which will be explained shortly. This process is repeated for each foreground segment G_{f_k} in the ground truth, assigning each a score s_{f_k}. The overall foreground score s_f is given by the sum of s_{f_k}, each weighted by the fractional area of G_{f_k}. In our dataset, all images have at least one foreground segment.

We weight the importance of identifying foreground segments by a constant α, which combines foreground and background scores as: $F_r = \alpha s_f + (1 - \alpha) s_b$. If there is no background in the ground truth (e.g., invasive cancer images), we assign $\alpha = 1$, otherwise $\alpha = 0.75$. This measure can handle both over- and undersegmentation issues. If a histological structure is oversegmented, the ground truth is only matched with the biggest overlapping segment and hence penalizes the oversegmentation behavior. Vice versa, if a structure is undersegmented, dividing the overlapping area by $\|G \cup S\|$ penalizes this behavior. The values of F_b, F_r, and Dice are between 0 and 1 with higher scores implying better performance.

TABLE I
EVALUATION OF SEGMENTATION ALGORITHMS ON THE TRAINING AND TESTING SETS. THE F-SCORES REPORTED ARE FOR THE BEST PARAMETER SETTINGS LEARNED ON THE TRAINING SET AND APPLIED ON THE TESTING SET

Methods	Train			Test		
	F_b	Dice	F_r	F_b	Dice	F_r
colorStats	0.4210	0.6704	0.3712	0.3933	0.7399	0.4012
inNucDist	0.5295	0.6553	0.4991	0.5508	0.7451	0.5617
Non-expert	**0.61**	**NA**	**0.63**	**NA**	**NA**	**NA**
GraphRLM	0.2144	0.4343	0.2564	0.1785	0.4258	0.2675
GlandSeg	0.2959	0.4598	0.2098	0.2872	0.485	0.2352
gPb	0.3995	0.6592	0.3454	0.3904	0.7333	0.3814
crisp-bound	0.3371	0.6564	0.2727	0.3201	0.7426	0.3258
JSEG	0.4400	0.4667	0.4455	0.4098	0.4204	0.4178
NCut	0.3756	0.0944	0.1732	0.3538	0.0731	0.1811
MeanShift	0.3395	0.2845	0.3055	0.3277	0.2417	0.2778
EGB	0.3999	0.6886	0.3365	0.3966	0.7475	0.4165
HED	0.3367	0.6681	0.2000	0.1975	0.7312	0.2603

E. Results

Table I summarizes the evaluation scores for all segmentation algorithms considered in this paper. We conducted parameter scanning for our methods as well as publicly available methods on the training set to select the setting that maximize F_b. Fig. 7 shows the dependency plots of F_b vs. σ and F_b vs. ρ for colorStats and of F_b vs. τ for interNucDist on the training set. The best parameter settings for colorStats are $\rho = 2.5$ and $\sigma = 7$, and for interNucDist is $\tau = 105$ pixels. Additional parameter scanning results are included in the tables of the **Supplementary Materials**, available in the supplementary files /multimedia tab.

Overall, inNucDist performs the best among all methods on both the training and testing sets. The improvement by inNucDist over the second and third best methods (JSEG, EGB) is statistically significant (t-test $p < 0.05$) for F_b and F_r. There is no statistically significant difference between the top methods in terms of Dice metric.

Since the dataset has both well-differentiated and poorly-differentiated images, colorStats does not perform as well. However, compared to crisp-bound and gPb, the two spectral graph segmentation methods, colorStats has better F_b and F_r scores over both training and testing datasets. The improvement of colorStats over crisp-bound and gPb on F_r is statistically significant ($p < 0.05$). Note that both of our methods significantly outperformed the deep learning based method HED.

Fig. 8 illustrates segmentation results obtained from applying various algorithms on two representative images. Segmentation results for all images are available in the supporting documents. For display purposes, the parameter setting for each algorithm was optimized to achieve the highest F_b score over the database of images. For each image, we report both F_b and F_r scores. In Fig. 8 top, which is the same one as in Fig. 1, colorStats result looks very similar to the one annotated by non-experts. Both results miss fat boundaries

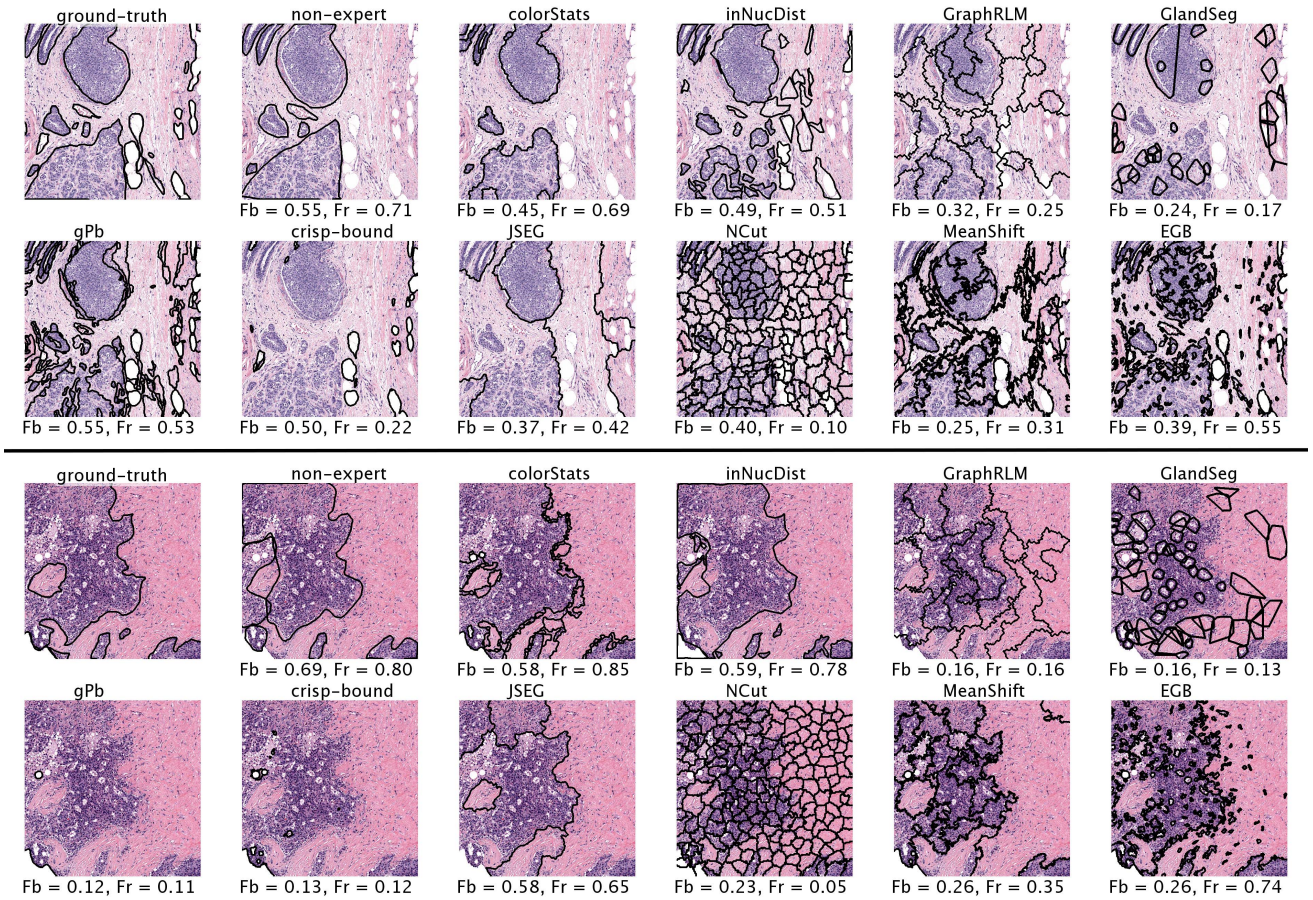

Fig. 8. Segmentation results for two representative images with boundary and region metrics displayed. Segmentation results for the entire dataset are available in the supporting documents.

which lowers F_b score, while account for the majority of the foreground segments which increases F_r score. `inNucDist` is able to detect the majority of histological structures, but picks up some islands in the stroma which are histologically indeterminate, leading to low F_r score.

`GraphRLM` and `NCut` divide the image into homogeneous regions but do not group them into histological structures (low F_r scores) or respect structural boundaries (low F_b score). `GlandSeg` misidentifies ducts and undersegment other structures, leading to both low F_b and F_r scores. `gPb` introduces some false positive edges within blobs of fat and across stroma but accurately traces the boundaries of histological structures and thus it has higher F_b score than `colorStats` and higher F_r score than `inNucDist`. `Crisp-bound` focuses on detecting boundaries between white segments and other color segments. It misses most boundaries of histological structures and undersegments foreground, resulting in low F_b and F_r scores. `JSEG` undersegments the invasive carcinoma blob and misses boundaries between the two normal ducts in the upper left corner, leading to low F_b and F_r scores. `MeanShift` tends to oversegment but respects structural boundaries. `EGB` picks up small spurious islands but mostly in the stroma background and hence it has decent F_r score. F_r scores clearly reflect the perceptual quality of segmentation results.

In Fig. 8 bottom, `colorStats` carves out long and thin islands next to the large cancer segment (low F_b), but preserves structural integrity of all foreground segments (high F_r). `inNucDist` slightly oversegments the large cancer blob and hence scores lower than `colorStats` on F_r. `gPb` and `crisp-bound` more severely under-segment the image, while white holes inside the invasive cancer patch are mistakenly identified as lumens by `GlandSeg`. `EGB` punctures holes in the largest segment that it picks up, thus scoring high in F_r but low in F_b. `JSEG` oversegments the cancer patch and has lower F_r score than `EGB`.

With our current non-optimized version of the code in Matlab and Java, `colorStats` takes roughly 40 seconds to segment an image of size 512×512 on a laptop with 16GB memory and `interNucDist` takes roughly 15-45 seconds on an image of size $2K \times 2K$. In fact, most methods take roughly the same amount of time (45 seconds-1 minute) on images of size 512×512. In addition, we have also tested our methods on both high resolution ($2K \times 2K$) and low resolution images (512×512) and observed no significant changes in performance. However, when it comes to scaling these methods to WSIs, the spectral methods are difficult to scale unless more optimized eigensolvers become available. On the other hand, we have scaled interNucDist to work on WSIs of size 20k x 20k where it takes roughly 7 minutes on a single i7

core machine with 64GB memory. With parallelization, the average running time reduces to 100 seconds per image of size $20K \times 20K$ (results not shown).

IV. Discussion

In this paper, we proposed two graph-based image segmentation methods based on local spatial color and nuclei neighborhood statistics. In `colorStats`, we analytically modeled the spatial color statistics between neighboring pixels using bivariate von Mises mixture models in the H&E optimized color space. In `inNucDist`, we segmented histological structures using nuclei neighborhood statistics in the form of pairwise distances between superpixels. Working with 232 expertly annotated breast H&E images, we demonstrated the ability of our algorithms to identify significant histological structures, and thus enable the understanding of their spatial relationships, and perhaps infer the status of the disease. Our method `inNucDist` performs better than the state-of-the-art methods and the improvement is statistically significant.

Segmentation methods adapted from the domain of natural images (`JSEG`, `EGB`, `crisp-bound`, `gPb`) performed much better than methods that have been specifically designed for H&E images (`GraphRLM`, `GlandSeg`). Notably, non-experts outperformed all the algorithms considered here, except for our method `inNucDist` on the region score. We observed that non-experts tend to annotate more segments than experts. This is reasonable since experts with more training in histopathology are able to group together segments into relevant histological structures while non-experts focus on grouping pixels into coherent segments, without the histological knowledge.

`colorStats` performed better than `gPb` and `crisp-bound`, demonstrating that our H&E optimized color space and von Mises modeling offered a more effective affinity measure for spectral segmentation. Note that our method `colorStats` uses only one feature, while `gPb` uses 36 color and texton features and `crisp-bound` uses six La*b* and variance features. We presented results using only one feature (H&E-hue), but this can be easily extended to incorporate more features. `inNucDist` uses only one feature, namely a threshold on pairwise superpixel distances, yet performs better than `GraphRLM`, which uses sixteen graph run length features.

In solving the challenging task of segmenting histological structures, we have encountered two major hurdles. First and foremost is the difficulty in obtaining the ground truth. In addition to being cost-prohibitive, curating ground truth annotations requires extensive communication with the collaborating pathologists because there are inevitable questions regarding what and how histological structures should be annotated. We recruited two pathology trainees and one practicing breast pathologist for the early proof-of-concept work in this paper. We hope that the publication of this paper will encourage a collaborative annotation effort among pathologists.

The second challenge is the use of appropriate metrics for performance evaluation. The current evaluation framework for natural image segmentation algorithm matches the algorithmic output with ground truth boundaries, but in clinical work, not every histological structure can be identified with certainty; especially in cancer images, the boundaries of the tumor, which are inherently weak, become more ambiguous that even expert pathologists find it hard to segment. When the ground truth is unidentifiable, this type of evaluation measure could severely misjudge the performance of the algorithm. It is the reason why we proposed a histologically relevant region metric, focusing on evaluating important histological structures involving clumps of nuclei and fat. Other metrics that are commonly in use in biomedical image analysis literature [12] are component-specific and are not applicable to the approaches presented here for segmenting a broad class of structures in breast tissue images.

One limitation of our methods is that the normalization step can obfuscate the histological structure identities. For example, lymphocytes and mitotic nuclei typically stain darker than epithelial nuclei. After normalization, their color appearances can become more similar, thus, leading to a loss of information in their identities. However, color normalization is crucial for downstream analysis. This problem is true for any normalization methods in the literature and can be mitigated by analyzing the original color images overlayed with segmentation boundaries.

One of our end goals is to build a computer assisted diagnostic system for pathologists that parses WSIs and triages relevant histological structures for rapid diagnosis. Tumor diagnosis and classification is a difficult, labor-intensive task that requires the expertise of highly trained physicians. A computer-assisted workflow could tremendously reduce the amount of pathologist effort required to do this diagnostic work, which would be of great value to healthcare organizations. Further, tumor biology is complex and a computational approach may empower pathologists to glean additional prognostic information from images, than is currently possible. Another important goal is to enable molecular-profiling based precision medicine approaches that build spatial interaction maps combining protein, DNA/RNA biomarkers.

As future work, we can automate the color preprocessing step to be completely unsupervised using the recently proposed non-linear tissue-component separation method [35]. In addition, we can improve the performance of our segmentation algorithms, by combining multi-scale methods with higher-order statistics in H&E-hue space. While this study attempts to identify segments with distinct spatial statistics, assigning them labels (e.g., normal duct, carcinoma in situ) will fall under the challenging task of tissue recognition. We can potentially rank histological structures segmented by our algorithms from the most to least abnormal (cancer, atypia, inflammation, etc.). Our approach also raises the possibility of using spatial statistics in recognizing tissue origins. Finally, we will initiate open-source collaborative efforts among pathologists for annotating H&E images.

Acknowledgments

We would like to thank B. Campbell, D. Spagnolo, J. Duboy, M. Marx and O. Choudhary for help at various stages of this project.

REFERENCES

[1] J. L. Fine, "21st century workflow: A proposal," *J. Pathol. Inform.*, vol. 5, no. 1, p. 44, Nov. 2014.

[2] B.-R. Wei and R. M. Simpson, "Digital pathology and image analysis augment biospecimen annotation and biobank quality assurance harmonization," *Clin. Biochem.*, vol. 47, nos. 4–5, pp. 274–279, 2014.

[3] J. Monaco, J. Hipp, D. Lucas, S. Smith, U. Balis, and A. Madabhushi, "Image segmentation with implicit color standardization using spatially constrained expectation maximization: Detection of nuclei," in *Proc. MICCAI*, vol. 45. 2012, pp. 365–372.

[4] F. Liu and L. Yang, "A novel cell detection method using deep convolutional neural network and maximum-weight independent set," in *Proc. MICCAI*, 2015, pp. 349–357.

[5] B. E. Bejnordi *et al.*, "Automated detection of DCIS in whole-slide H&E stained breast histopathology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 9, pp. 2141–2150, Sep. 2016.

[6] K. Nguyen, A. Sarkar, and A. K. Jain, "Structure and context in prostatic gland segmentation and classification," in *Proc. MICCAI*, 2012, pp. 115–123.

[7] Y. Peng, Y. Jiang, L. Eisengart, M. A. Healy, F. H. Straus, and X. J. Yang, "Segmentation of prostatic glands in histology images," in *Proc. ISBI*, 2011, pp. 2091–2094.

[8] J. Xu, J. P. Monaco, and A. Madabhushi, "Markov random field driven region-based active contour model (MaRACel): Application to medical image segmentation," in *Proc. MICCAI*, 2010, pp. 197–204.

[9] J. Wang, J. D. MacKenzie, R. Ramachandran, and D. Z. Chen, "Detection of glands and villi by collaboration of domain knowledge and deep learning," in *Proc. MICCAI*, 2015, pp. 20–27.

[10] N. M. Rajpoot, H. El-Daly, E. Simmons, and N. M. Rajpoot, "HyMaP: A hybrid magnitude-phase approach to unsupervised segmentation of tumor areas in breast cancer histology images," *J. Pathol. Inform.*, vol. 4, no. 2, p. 1, 2013.

[11] A. B. Tosun, M. Kandemir, C. Sokmensuer, and C. Gunduz-Demir, "Object-oriented texture analysis for the unsupervised segmentation of biopsy images for cancer detection," *Pattern Recognit.*, vol. 42, no. 6, pp. 1104–1112, 2009.

[12] MICCAI. (2015). *Grand Challenges in Biomedical Image Analysis*. accessed on Jun. 4, 2015. [Online]. Available: https://www2.warwick.ac.uk/fac/sci/dcs/research/combi/research/bic/glascontest/

[13] A. Basavanhally *et al.*, "Incorporating domain knowledge for tubule detection in breast histopathology using O'Callaghan neighborhoods," *Proc. SPIE*, vol. 7963, p. 796310, Mar. 2011.

[14] M. T. McCann, D. G. Mixon, M. C. Fickus, C. A. Castro, J. A. Ozolek, and J. Kovačević', "Images as occlusions of textures: A framework for segmentation," *IEEE Trans. Image Process.*, vol. 23, no. 5, pp. 2033–2046, May 2014.

[15] X. Li and K. N. Plataniotis, "A complete color normalization approach to histopathology images using color cues computed from saturation-weighted statistics," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 7, pp. 1862–1873, Jul. 2015.

[16] A. Vahadane *et al.*, "Structure-preserving color normalization and sparse stain separation for histological images," *IEEE Trans. Med. Imag.*, vol. 35, no. 8, pp. 1962–1971, Aug. 2016.

[17] J. Vicory *et al.*, "Appearance normalization of histology slides," *Comput. Med. Imag. Graph.*, vol. 43, pp. 89–98, Jul. 2015.

[18] E. Reinhard *et al.*, "Color transfer between images," *IEEE Comput. Graph. Appl. Mag.*, vol. 21, no. 5, pp. 34–41, Sep./Oct. 2001.

[19] K. V. Mardia and J. Frellsen, "Statistics of bivariate von mises distributions," in *Bayesian Methods in Structural Bioinformatics*. Springer, 2012, pp. 159–178.

[20] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *Int. J. Comput. Vis.*, vol. 40, no. 1, pp. 49–70, Oct. 2000.

[21] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, vol. 382. Hoboken, NJ, USA: Wiley, 2007.

[22] P. Isola, D. Zoran, D. Krishnan, and E. H. Adelson, "Crisp boundary detection using pointwise mutual information," in *Proc. ECCV*, 2014, pp. 799–814.

[23] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.

[24] Y. Al-Kofahi, W. Lassoued, W. Lee, and B. Roysam, "Improved automatic detection and segmentation of cell nuclei in histopathology images," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 4, pp. 841–852, Apr. 2010.

[25] S. Arslan, T. Ersahin, R. Cetin-Atalay, and C. Gunduz-Demir, "Attributed relational graphs for cell nucleus segmentation in fluorescence microscopy images," *IEEE Trans. Med. Imag.*, vol. 32, no. 6, pp. 1121–1131, Jun. 2013.

[26] C. Chen, W. Wang, J. A. Ozolek, and G. K. Rohde, "A flexible and robust approach for segmenting cell nuclei from 2d microscopy images using supervised learning and template matching," *Cytometry A*, vol. 83A, no. 5, pp. 495–507, 2013.

[27] A. B. Tosun, O. Yergiyev, S. Kolouri, J. F. Silverman, and G. K. Rohde, "Detection of malignant mesothelioma using nuclear structure of mesothelial cells in effusion cytology specimens," *Cytometry A*, vol. 87, no. 4, pp. 326–333, 2015.

[28] A. B. Tosun and C. Gunduz-Demir, "Graph run-length matrices for histopathological image segmentation," *IEEE Trans. Med. Imag.*, vol. 30, no. 3, pp. 721–732, Mar. 2011.

[29] Y. Deng and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 8, pp. 800–810, Aug. 2001.

[30] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[31] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[32] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, Sep. 2004.

[33] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1395–1403.

[34] J. Pont-Tuset and F. Marques, "Measures and meta-measures for the supervised evaluation of image segmentation," in *Proc. IEEE CVPR*, Jun. 2013, pp. 2131–2138.

[35] J. S. Sarnecki *et al.*, "A robust nonlinear tissue-component discrimination method for computational pathology," *Lab. Invest.*, vol. 96, pp. 450–458, Apr. 2016.