High-Throughput Biomarker Segmentation on Ovarian Cancer Tissue Microarrays via Hierarchical Normalized Cuts

Andrew Janowczyk, Sharat Chandran, Rajendra Singh, Dimitra Sasaroli, George Coukos, Michael D. Feldman, and Anant Madabhushi*

***Abstract*—We present a system for accurately quantifying the presence and extent of stain on account of a vascular biomarker on tissue microarrays. We demonstrate our flexible, robust, accurate, and high-throughput minimally supervised segmentation algorithm, termed hierarchical normalized cuts (HNCuts) for the specific problem of quantifying extent of vascular staining on ovarian cancer tissue microarrays. The high-throughput aspect of HNCut is driven by the use of a hierarchically represented data structure that allows us to merge two powerful image segmentation algorithms—a frequency weighted mean shift and the normalized cuts algorithm. HNCuts rapidly traverses a hierarchical pyramid, generated from the input image at various color resolutions, enabling the rapid analysis of large images (e.g., a 1500 × 1500 sized image under 6 s on a standard 2.8-GHz desktop PC). HNCut is easily generalizable to other problem domains and only requires specification of a few representative pixels (swatch) from the object of interest in order to segment the target class. Across ten runs, the HNCut algorithm was found to have average true positive, false positive, and false negative rates (on a per pixel basis) of 82%, 34%, and 18%, in terms of overlap, when evaluated with respect to a pathologist annotated ground truth of the target region of interest. By comparison, a popular supervised classifier (probabilistic boosting trees) was only able to marginally improve on the true positive and false negative rates (84% and 14%) at the expense of a higher false positive rate (73%), with an additional computation time of 62% compared to HNCut. We also compared our scheme against a *k*-means clustering approach, which both the HNCut and PBT schemes were able to outperform. Our success in accurately quantifying the extent of vascular stain on ovarian cancer TMAs suggests that HNCut could be a very powerful tool in digital pathology and bioinformatics applications where it could be used to facilitate computer-assisted prognostic predictions of disease outcome.**

***Index Terms*—Biomarker discovery, hierarchical normalized cuts (HNCuts), high-throughput, mean shift, multivariate histology, normalized cuts, segmentation, tissue microarray (TMA).**

I. Introduction

With the advent of whole slide digital scanners, histological data have now become amenable to digital and quantitative image analysis [1], [2], [48], [49], [52], [53], [56]–[58]. Additionally, with tissue microarray (TMA) technology, it is now possible to simultaneously stain several hundred tissue sections (tissue cylinders) for the presence of various biomarkers [47]. In the digital (uncompressed) form, these TMAs can be several gigabytes in size with image dimensions of up to 100 000 × 100 000 pixels. Some researchers are currently looking to increase the number of cylinders that can be accommodated on a single TMA to over 10 000 [3]. Since manual analysis of such large amounts of data is clearly not tractable, high-throughput, reproducible, and accurate computerized image analysis methods are required for quantification of the presence and extent of different biomarkers on TMAs [36]–[38], [41], [46], [50], [54].

It is estimated[1] that 21 990 women will be diagnosed with and 15 460 women will die of cancer of the ovary (OCa) in 2011. The five-year survival rates of these women are highly correlated with the early detection of OCa. Recent work [4] suggests that specific tumor vascular biomarkers (TVMs), identifiable on OCa TMAs [46], could have prognostic significance, which would enable not only predicting the aggressiveness of the disease, but could also help in tailoring a personalized treatment regime for the patient.

The manual quantification of extent of biomarker staining is, however, a laborious, time consuming, and error-prone affair. Consequently, there is a real need for high-throughput quantitative image analysis algorithms which can automatically and efficiently estimate biomarker extent on very large pathology slides in a few seconds [40], [44], [45]. For illustrative purposes, consider having ten patient studies each with a TMA of 500 cylinders. Overall, there are 5000 1500 × 1500 images to analyze. An expert clinician could expect to invest 5 min per image, thus resulting in over 400 h to analyze all of the data.

Manuscript received August 1, 2011; revised October 25, 2011; accepted November 15, 2011. Date of publication December 13, 2011; date of current version April 20, 2012. This work was supported by the National Cancer Institute under Grant R01CA136535-01, Grant R01CA140772, and Grant R03CA143991-01, The Cancer Institute of New Jersey, the Life Science Commercialization Award, Ovarian Cancer SPORE Grant P50 CA083638, and First Solutions Limited. *Asterisk indicates corresponding author.*

A. Janowczyk is with the Indian Institute of Technology Bombay, Powai, Mumbai, Maharashtra 400076, India, and also with the Department of Biomedical Engineering, Rutgers University, New Brunswick, NJ 08901 USA (e-mail: andrew@cse.iitb.ac.in).

S. Chandran is with the Indian Institute of Technology Bombay, Powai, Mumbai, Maharashtra 400076, India (e-mail: sharat@cse.iitb.ac.in).

R. Singh is with Quest Diagnostics, Inc, Madison, NJ 07940 USA (e-mail: rajendra.s.singh@questdiagnostics.com).

D. Sasaroli and G. Coukos are with Ovarian Cancer Research Center, University of Pennsylvania, Philadelphia, PA 19104-3339 USA (e-mail: biodemi@yahoo.com; gcks@mail.med.upenn.edu).

M. D. Feldman is with the Department of Pathology and Lab Medicine, University of Pennsylvania, Philadelphia, PA 19104-3339 USA (e-mail: feldmanm@mail.med.upenn.edu).

*A. Madabhushi is with the Department of Biomedical Engineering, Rutgers University, New Brunswick, NJ 08901 USA (e-mail: anantm@rci.rutgers.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.2011.2179546

[1]Cancer Facts and Figures, American Cancer Society, Atlanta, GA, 2011.

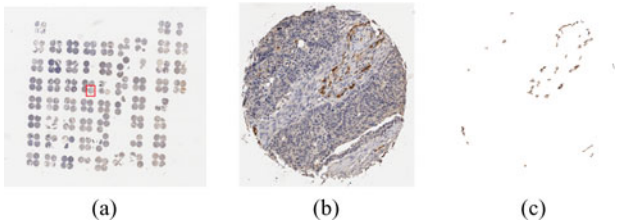

Fig. 1. (a) TMA. (b) Representative magnified tissue cylinder drawn from (a) with the extracted (c) stained TVM. A typical TMA could contain over 500 individual cylinders, making the biomarker detection via traditional image analysis algorithms a challenge.

When we consider our proposed algorithm, using a single laptop, we could complete the same work in about 10 h. Using a standard eight-core machine, the analysis could be completed in just over an hour.

Accurate identification of the extent and intensity of the stain could provide a quantitative and reproducible prognostic metric that could help predict risk of disease recurrence and patient survival. In Fig. 1, the region of interest is the reactive chemically stained dark brown region, corresponding to the TVM ESM-1 [4]. While the extent and intensity of the ESM-1 stained region may have prognostic significance, it is currently impractical in terms of both time and effort for an expert pathologist to perform this segmentation manually.

Most previous computerized image analysis algorithms for TMAs have involved thresholding-based schemes [5]–[7]. These methods are known to be highly sensitive to even slight changes in color and illumination [55]. Clustering-based approaches, including k-means [42], have also been investigated for the analysis of TMAs. However, k-means is a nondeterministic algorithm and is highly sensitive to the initial choice of cluster centers [8]. Active contour schemes [9], while suitable for cell and nuclear segmentation in digital pathology, are not ideally suited to the problem of pixel-level classification. Additionally, they are typically infeasible for problems where hundreds of objects need to be concurrently segmented on very large images [10]. In [11], it was shown that using hierarchical normalized cuts (HNCuts) as a preprocessing step for an active contour approach drastically improved computation time of their active contour approach by providing a significantly better initial estimate to the region.

While supervised learning methods such as probabilistic boosting trees (PBT) [12], [13] have become popular for image classification and segmentation, these methods are constrained by the difficulty [14] in obtaining ground-truth segmentations from experts for classifier training of the object of interest. Manual annotation of the data, apart from being time consuming and laborious, can also be expensive if only a medical practitioner is capable of providing accurate annotations. Additionally, if the target of interest changes, considerable effort might be required to generate new annotations and retrain the classifier [43].

Normalized Cuts (NCuts) [15] is among the final mature descendants from a series of graph-cutting techniques ranging from max cut to min cut [16]–[19]. It is a popular scheme in spite of its main drawbacks: 1) the large number of calculations needed for determining the affinity matrix and 2) the time-consuming eigenvalue computation. For large images, the computation and overhead of these border on the infeasible [15]. Consequently, a significant amount of research has focused on avoiding their direct calculations [20], [21].

The mean-shift algorithm (MS) [22], [39] has been employed and modified in [23] as an unsupervised technique for mode discovery instead of k-means [51]. The MS algorithm attempts to identify the cluster mean within a predefined bandwidth. By using a steepest gradient approach, a fast convergence to the set of true means of the statistical data can be found [24]. The improved fast Gauss transform (IFGT) implementation of the MS algorithm [25] allowed computation times for large images to become reasonable. For the rest of this paper, we will make no distinction between IFGT-MS and MS.

There are two major differences between this study and a previous, preliminary conference version of this paper [26]. The first difference is that this study uses a frequency weighted mean shift (FWMS) which significantly improves the computation time over that in [26]. The second is an extension of the original work by the inclusion of a number of additional experiments to rigorously and quantitatively evaluate our scheme on a much larger data cohort compared to what was initially presented in [26]. The strength of HNCut is in that it combines a powerful unsupervised clustering technique (mean shift [22]) with an equally powerful graph partitioning scheme (NCuts [15]). By performing clustering and partitioning in the color space (as opposed to pixel-level classification), the HNCut algorithm is highly efficient and precise. For large images, such as TMAs where there are often many fewer unique colors than pixels, performing the analysis in the color as opposed to the spatial domain could result in significant improvements in computational processing time. HNCut only requires specifying a few representative pixels from the target class and, unlike more traditional supervised classification algorithms, does not require more detailed target object annotation. More importantly, the HNCut algorithm is more flexible compared to supervised schemes in its ability to segment different object classes. The combination of both the high-throughput efficiency and flexibility of HNCut makes it ideally suited to applications requiring high-throughput analysis, such as quantifying the expression of biomarkers on TMAs. In this paper, we demonstrate the specific application of HNCut to a problem of automated quantification of stain extent associated with a vascular marker on OCa TMAs (see Fig. 1).

The rest of this paper is organized as follows. In Section II, we describe previous related work and contributions of our work. In Section III, we provide the methodological description of HNCut, and in Section IV, we demonstrate its application in segmenting the stain extent of a TVM from OCa TMAs. Qualitative and quantitative evaluations of the segmentation results for the TVM on OCa TMAs are presented in Section IV, followed by a discussion of the results in Section V. Finally, concluding remarks and future directions are presented in Section VI.

II. Previous Work and Contributions

The attempt to merge NCuts and mean shift is not new [27]. To overcome the computational issues associated with NCut, a

novel approach of combining both the MS and NCut algorithms was presented in [27]. Clustering the image by running the MS algorithm to convergence produced class assignments for the pixels. By taking the average intensity value of the regions obtained via the MS clustering step and using them as the vertices in the NCut algorithm, a significant speed improvement was obtained.

It was later noticed in [28] that when points of similar values are within an ϵ neighborhood of each other, their contribution to the overall system can be merged, providing an efficiency improvement by reducing the number of computations needed per iteration. We use this to extend the MS work in [27] in a hierarchical fashion which is more pertinent and amenable to problems in digital pathology and biomedical imaging. This allows us to perform the same detection or segmentation task in half the time (under 0.5 s for HNCut as compared to the reported 1.78 s for [28]).

While there are similarities between our work and [27] and [28], there are also significant differences. The proposed algorithm is specifically designed for rapid extraction of pixels of interest in a minimally supervised manner, as opposed to unsupervised clustering which is insensitive to the user's domain knowledge as the aforementioned approaches take. Thus, we first manually identify the desired target class based on individual representative colors (referred to as a swatch) selected from the target class by a user. This swatch, which can be changed based on the desired target class or domain, lends HNCut significant flexibility and ease of use.

Second, to our knowledge, this is the first attempt at combining an FWMS with a partitioning algorithm that accomplishes the same task as MS but does it significantly faster. The FWMS exploits the fact that as each iteration of MS completes, more points converge. We demonstrate in this paper how the convergence of our novel FWMS scheme allows us to perform clustering 15 times faster than the traditional MS algorithm [27]. We can see from the run times presented in [27] that for a 240×160 pixel image, the running time is 2.18 s. By working directly in the color space and using an FWMS, we can perform a similar operation in 6 s on an image 58 times larger.

The work presented in this paper represents important methodological and clinical contributions summarized as follows.

1) A new minimally supervised hierarchical segmentation approach that combines a FWMS and normalized cuts (HNCuts) for pixel-level detection and classification. HNCut is able to segment very large images rapidly.
2) HNCut is largely insensitive to choice of parameter value and is able to discriminate between regions with similar color values. The parameters for NCuts are automatically computed, and the parameters for the FWMS are automatically adjusted based on the variance of the output.
3) Layman initialization of the system is possible, obviating the need for detailed ground-truth annotation from an expert that is required for more sophisticated supervised classifiers.
4) This study represents the first attempt, to our knowledge, to accurately quantify a vascular marker on OCa TMAs with the ultimate objective of creating a quantitative image-based metric for OCa prognosis and survival.

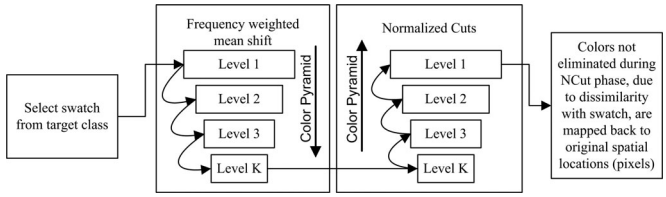

Fig. 2. Flow chart of the HNCut process. Proceeding left to right, the user selects the domain swatch, followed by the FWMS of the image. This results in the original image being decomposed into multiple levels of color resolution, which is then followed by the application of NCut at each of the color resolutions generated. At each pyramid level, colors not deemed to be part of the swatch are eliminated. Following the application of NCut on the color pyramid (from the lowest to the highest color resolution), the color values that have not been eliminated are mapped back to the spatial domain via their original pixel locations, and the final segmentation is obtained.

III. Description of HNCut

A. Overview

Fig. 2 presents a high-level overview of the four stages associated with the HNCut algorithm. Each of these stages are discussed in detail in the following sections. We present an overview here to guide the reader through the various stages.

We start by requiring the user to select a few sample pixels from the target class from an image. We use these pixels to guide the subsequent pixel classification process across all images in the same domain.

Next, we employ the MS algorithm on the color values in the image to form a hierarchical data structure (represented by the levels in the color pyramid in the second box in Fig. 2). Intuitively, the FWMS algorithm allows for identification of color values which are within some specified tolerance of each other and assigns them to the same mode. Employing the NCuts operation only on the unique values at each level of the pyramid, as opposed to all possible color values, allows for a factorization resulting in significantly fewer computations. An illustration of the application of the scheme to an OCa TMA, for detecting a TVM, is illustrated in Fig. 3. We then compute the weight for each unique mode, which reflects the actual frequency of the number of pixels associated with it.

Using this pyramid, we can drastically reduce the large segmentation problem in the color space to a set of much smaller graph partitioning problems (the third box from the left in Fig. 2), which we show can be solved far more efficiently by NCut. By starting at the bottom of the pyramid, we partition the unique values (typically on the order of ten values) into two sets such that all of the values selected by the user in the first step are assigned to the first partition. Subsequently, we eliminate the second partition and map the colors in the first partition to an immediately higher color resolution level in the pyramid. This process continues until the entire pyramid is traversed. The last step involves mapping the color values not eliminated back into the spatial domain.

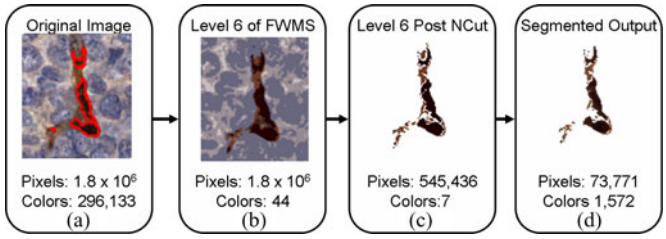

Fig. 3. (a) Original image with desired TVM stain enclosed in red. (b) Image at the bottom of the color pyramid during FWMS. (c) Image at the bottom of the color pyramid following application of NCuts. (d) Final segmentation results obtained by mapping colors not eliminated by HNCut spatially onto the original image. Note that between (a) and (b), a significant reduction in color resolution occurs, which allows NCuts to be performed on an image with several orders of magnitude fewer colors compared to the original image (a). NCut is then applied at progressively higher color resolutions, while at each pyramid level, colors not deemed to be part of the swatch are eliminated. The colors retained at the highest resolution are then spatially mapped onto the corresponding pixels to yield the final segmentation.

The hierarchical set of operations described previously makes for an extremely efficient and accurate algorithm; thus, applying the NCut at the lowest levels of the pyramid is relatively simple to do and encourages a more sophisticated definition of pixel affinity. While in this paper only chromatic information was leveraged, the method is easily and efficiently extensible to incorporate additional image features (e.g., texture).

Fig. 3 displays an image from our dataset undergoing the HNCut procedure, with the intent of quantification of the vascular marker stain (brown color). The numbers shown in the boxes in Fig. 3 represent the reduced number of colors and pixels generated by the HNCut scheme at different levels of the pyramid within a single cylinder (1500 × 1500 pixels, 300 000 colors) from a TMA.

B. Notation

An image scene is defined as $\mathcal{C} = (C, \mathbf{f})$ where C is a 2-D Cartesian grid of N pixels, $c \in C$, where $c = (x, y)$. $\mathbf{f}$ is a color intensity function, where $\mathbf{f} \in \mathbb{R}^3$.

We define as $\mathbf{F}_1 \in \mathbb{R}^3$ the vector of colors associated with all pixels $c \in C$ at the full color resolution (top of the color pyramid). The elements of $\mathbf{F}_1$, namely $f_{1,i}$, are derived such that for pixel c_i, $f_{1,i} = \mathbf{f}(c_i)$ and $f_{1,i} \in \mathbb{R}^3$. A list of commonly used notation and symbols in this paper is summarized in Table I.

C. Integrating Domain Knowledge to Guide NCuts

A user via manual selection defines a color swatch $S_1 = \{f_{1,\alpha_\tau} | \alpha_\tau, \tau \in \{1, \ldots, N\}\}$, where α_τ is an index value to the original color vector. Note that S_1 is easily obtained by annotating (manually) a few pixels from the object of interest on a representative image and may be easily changed based on the application. As we will describe in further detail later, S_1 is only used to identify which color partition [A or B from (8)] to retain during NCut. It is important to note that since S_1 is a reference to a subset of the color values in the original image, it will undergo all of the MS and NCut operations presented below. Note that S_1 is the swatch originally defined by the user at the full resolution, $k = 1$.

TABLE I
DESCRIPTION OF NOTATION AND SYMBOLS COMMONLY EMPLOYED IN THIS PAPER

Symbol	Description
$\mathcal{C}$	2D Image Scene
C	2D Cartesian grid of pixels, $c = (x, y)$
G	Gaussian function with bandwidth of size σ_{MS}
S_1	User selected swatch
k	Levels in color pyramid $k \in \{1, \ldots, \mathrm{K}\}$
N	Number of pixels, $\lvert C \rvert$
$\mathbf{F}_k$	Color vector at level k
$\mathbf{G}$	Connected graph $\mathbf{G}$ of edges E and vertices V
$\hat{\mathbf{F}}_k$	$\mathbf{F}_k$ with all duplicate values removed
$\mathbf{w}_k$	Frequency counts of $\hat{\mathbf{F}}_k$ in $\mathbf{F}_k$
M_k	$\lvert \hat{\mathbf{F}}_k \rvert$
$\mathbf{\Psi}$	Symmetric matrix comprised of $\psi(c_i, c_j)$
$\psi(c_i, c_j)$	Affinity measure between c_i, c_j
$\mathbf{D}$	Diagonal matrix with $\mathbf{D}(i,i) = \sum_j \psi(c_i, c_j)$
$R^{a,\varsigma}$	HNCut segmentation of target class
$R^{b,z}$	Ground Truth segmentation of target class

D. FWMS for Reducing the Number of Colors for NCut

1) Theory: The MS algorithm is used to detect modes in data using a density gradient estimation. By solving for when the density gradient is zero and the Hessian is negative semidefinite, we can identify local maxima. For a more detailed explanation of the algorithm, we refer the reader to [24].

We start with the fixed point iteration update $\forall j \in \{1, \ldots, N\}$ in MS (described in [24]) as

$$f_{k+1,j} \leftarrow \frac{\sum_{i=1}^{N} f_{k,i} G(f_{k,j} - f_{k,i})}{\sum_{i=1}^{N} G(f_{k,j} - f_{k,i})} \tag{1}$$

where G is a Gaussian function with a bandwidth parameter σ_{MS}, which is used to compute the kernel density estimate at data point c_j, $G(f_{k,j} - f_{k,i}) = \exp(-\frac{\|f_{k,j} - \hat{f}_{k,i}\|_2}{\sigma^2_{MS}})$, with $\| \cdot \|_2$ representing the $L2$ norm. $k \in \{1, \ldots, K\}$ represents various levels of color resolution produced at each iteration. The overall computation time for (1) is $O(N^2)$. By employing the IFGT [25], we can reduce the computation complexity to $O(N)$ with minimal precision loss.

It becomes possible to exploit the fact that after each iteration of the MS many of the data points, in our case color values, converge. If we consider what that convergence means mathematically, essentially two points c_{β_1}, c_{β_2}, where $\beta_1, \beta_2 \in \{1, \ldots, N\}$ meet the requirement that $|f_{k,\beta_1} - f_{k,\beta_2}| \leq \epsilon$ where ϵ is a predefined tolerance value. We can thus rewrite the numerator of (1), which is

$$f_{k,\beta_1} G(f_{k,j} - f_{k,\beta_1}) + f_{k,\beta_2} G(f_{k,j} - f_{k,\beta_2}) + \sum_{i=1, i \neq \beta_1, \beta_2}^{N} f_{k,i} G(f_{k,j} - f_{k,i}) \tag{2}$$

in the form

$$2 f_{k,\beta_1} G(f_{k,j} - f_{k,\beta_1}) + \sum_{i=1, i \neq \beta_1, \beta_2}^{N} f_{k,i} G(f_{k,j} - f_{k,i}) \tag{3}$$

thereby avoiding the explicit calculation of $G(f_{k,j} - f_{k,\beta_2})$, where $j, \beta_1, \beta_2 \in \{1, \ldots, N\}, k \in \{1, \ldots, K\}$. This results in one less computation for the Gaussian, which is by far the most expensive operation in the entire MS clustering process. The formulation in (3) results in a significant computational efficiency improvement. The computational savings apply to the denominator as well, as it follows the same reduction.

As a result, we may rewrite the update presented in (1) as a multistep update. Initially, we determine the unique values in $\mathbf{F}_k$ under the constraint that any color values $|f_{k,i} - f_{k,j}| \leq \epsilon$ are considered equivalent. Thus, from $\mathbf{F}_k = \{f_{k,1}, f_{k,2}, \ldots, f_{k,|\mathbf{F}_k|}\}$, we can construct the vector $\hat{\mathbf{F}}_k$, where $\hat{\mathbf{F}}_k \subset \mathbf{F}_k$ and $\hat{\mathbf{F}}_k$ is a set of only unique values in $\mathbf{F}_k$, with $|\hat{\mathbf{F}}_k| = M_k$. A weight vector $\mathbf{w}_k = \{w_{k,1}, \ldots, w_{k,M_k}\}$ is then computed for $\hat{\mathbf{F}}_k$ as

$$w_{k,j} = \sum_{i=1, f_{k,i} = \hat{f}_{k,j}}^{|\mathbf{F}_k|} w_{k-1,i} \tag{4}$$

where $j \in \{1, \ldots, M_k\}$. Equation (4) is summing the weights from the previous level into the new unique values that resulted from the next iteration of mean shifting. As a result, $w_{k,j}$ contains a count of the number of original pixels that have migrated to $F_{k,j}$ through mean shifting. When $k = 1$, we define $\mathbf{w}_0$ as a vector of length N, filled with ones, representing that each color value has equal weighting. Now, the number of points in the system that have converged to some intensity (color) value $\hat{f}_{k,j}$ is represented by $w_{k,j}$. It is important to note the following definition of M_k where

$$|\mathbf{w}_k| = |\hat{\mathbf{F}}_k| = |\mathbf{F}_{k+1}| = M_k \tag{5}$$

and

$$\sum_{i=1}^{M_k} w_{k,i} = N \tag{6}$$

which leads us to the update of (1):

$$f_{k+1,j} \leftarrow \frac{\sum_{i=1}^{M_k} w_{k,i} \hat{f}_{k,i} G(\hat{f}_{k,j} - \hat{f}_{k,i})}{\sum_{i=1}^{M_k} w_{k,i} G(\hat{f}_{k,j} - \hat{f}_{k,i})} \tag{7}$$

for $j \in \{1, \ldots, M_k\}$.

A illustration of the steps described in (3)–(7) is presented in Fig. 4. The images depict a standard probability density function (pdf in red) computed from the Gaussian contributions (in blue) from the 1-D data points (red circles). From Fig. 4(a), we can see that colors f_{β_1} and f_{β_2} will converge in the next iteration of the MS. We exploit the fact that once f_{β_1} and f_{β_2} converge, it becomes possible to factor out f_{β_2} from the system, and move its contribution into f_{β_1}, without altering the distribution [see Fig. 4(b)].

We call this new approach the FWMS. The FWMS helps us to produce a pyramidal scene representation $\mathcal{C}_k = (C, \mathbf{F}_k)$, where $k \in \{1, \ldots, K\}$ represents K levels of the color pyramid. Note that $M_1 \geq M_2 \geq \cdots \geq M_K$, indicating level 1 has the most colors and M_K the least. In other words, FWMS results in a series of scenes $\mathcal{C}_k$, all mutually aligned, but with a smaller

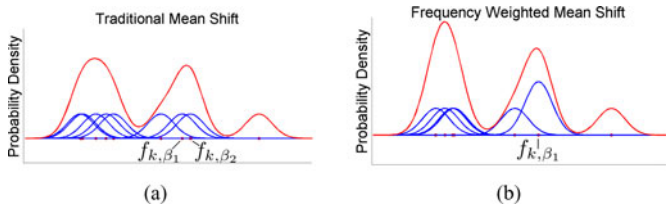

Fig. 4. Visual representation of the pdf illustrating the difference between the (a) traditional MS and the (b) FWMS. The red circles on the x-axis are the given values in a 1-D system, the blue arcs are the associated Gaussian contributions, while the red line above represents the summation of all of the contributions, i.e., the pdf. In (b), when points f_{β_1} and f_{β_2} converge, f_{β_2} is removed from the system, and its contribution is moved into f_{β_1} as a multiplication, avoiding an additional expensive step in the computation of the Gaussian pdf.

Algorithm 1 Frequency Weighted Mean Shift to Generate Color Pyramid

Input: $\mathbf{F}_1$ of $\mathcal{C}_1$
Output: $\hat{\mathbf{F}}_1, \hat{\mathbf{F}}_2, \ldots, \hat{\mathbf{F}}_K$
1: $k = 1$
2: **while** not converged **do**
3: Compute the unique values of $\mathbf{F}_k$ and store them in $\hat{\mathbf{F}}_k$
4: Compute frequency of colors in $\hat{\mathbf{F}}_k$ as they appear in $\mathbf{F}_k$ using Eq 4, store in $\mathbf{w}_k$
5: Generate $\mathbf{F}_{k+1}$ using Eq 7
6: $k = k + 1$
7: **end while**
8: **return** $\hat{\mathbf{F}}_1, \hat{\mathbf{F}}_2, \ldots, \hat{\mathbf{F}}_K$

number of colors in $\{\mathcal{C}_K, \mathcal{C}_{K-1}, \ldots\}$ compared to $\{\mathcal{C}_1, \mathcal{C}_2, \ldots\}$, which allows for NCut to be tractable.

2) Algorithm: The convergence requirement stated in line two of Algorithm 1 may be specified via three possible criteria. The first is the maximum number of iterations, a number specified by the user. The second more common approach is to stop the algorithm when the difference between any two iterations falls below a predefined threshold (i.e., the amplitude of the migrations associated with each point reduces significantly). Finally, convergence can be reached when the number of elements in $\hat{\mathbf{F}}$ becomes small enough that additional clustering provides no efficiency benefit as the overhead in the NCut starts to outweigh the computation time. This process was illustrated in Fig. 3 as the sequence of steps going from (a) to (b). It may be seen from Fig. 3 that the overall color resolution is significantly reduced as the algorithm proceeds from level 1 to level K. In this example, the original image containing about 300 000 unique color values was reduced to 44 unique values. This significantly smaller set of values makes the NCut step tractable since we operate directly in the color space.

E. NCuts on FWMS Reduced Color Space

1) Theory: NCuts [15] is a graph partitioning method, used to separate data into disjoint sets. For our problem, the hierarchical pyramid created by FWMS at various levels of color resolution $(\hat{\mathbf{F}}_1, \hat{\mathbf{F}}_2, \ldots, \hat{\mathbf{F}}_K)$ serves as the initial input to the NCut algorithm. The NCut takes a connected graph $\mathbf{G} = (E, V)$, with vertices (V) and edges (E) and partitions the vertices into disjoint groups. By setting V equal to the set of color values $\hat{\mathbf{F}}_K$,

and having the edges represent the similarity (or affinity) between the color values, we can separate the vertices into groups of similar color values. The NCut is defined as the process by which the removal of edges leads to two disjointed partitions A and B such that the variance of values (in our case colors) in A and B are minimized and the difference in average value (intensity of colors) between A and B is maximized. We present the high-level formulation as described in [15]:

$$\mathrm{NCut}(A,B) = \frac{\mathrm{cut}(A,B)}{\mathrm{assoc}(A,V)} + \frac{\mathrm{cut}(A,B)}{\mathrm{assoc}(B,V)} \tag{8}$$

where cut describes the affinity between the sets, encouraging higher dissimilarity between sets, and assoc describes the affinity between a set and the whole system, encouraging sets of significant size. The ψ function is used to define the affinity between two points. Our ψ function is defined as

$$\psi(\hat{f}_{k,i}, \hat{f}_{k,j}) = \exp\left(-\frac{\|\hat{f}_{k,i} - \hat{f}_{k,j}\|_2^2}{\sigma_{\mathrm{Ncut}}}\right) \tag{9}$$

with σ_{Ncut} as a bandwidth parameter. It is worth noting that in the traditional NCut paper [15], their affinity calculation took into account both a spatial and color component. For even small images, this made the affinity matrix intractable. As a result, the ψ function had a spatial constraint introduced such that (9) is set to zero if the associated pixels are farther away than a user-specified distance. This constraint forced the affinity matrix $\mathbf{\Psi}$ to typically be sparse, making its storage and subsequent operations applied to it less burdensome. Nevertheless, for large images, the affinity matrix is still too large (in spite of the spatial constraints), and as such we choose to operate solely in a significantly reduced color space, without the imposition of spatial constraints. In Fig. 3, we can see at the bottom of the hierarchical pyramid for a color image with original dimensions of 1200×1200, we would have an affinity matrix of only 7×7, and at the highest level a size of 1572×1572.

2) Algorithm: The main steps comprising the HNCut technique are shown in Algorithm 2. We begin by applying NCut on the lowest image resolution generated in the previous section, by setting $k = K$, $V_k = \{\hat{f}_{k,1}, \hat{f}_{k,2}, \ldots, \hat{f}_{k,M_k}\}$, i.e., the set of unique color values present at level K from FWMS.

Step 1: We apply NCut to partition the scene into two disjoint color sets A and B, where $A, B \subset V_k$. To perform this partition, we compute the affinity matrix $\mathbf{\Psi}_K \in \mathbb{R}^{M_k \times M_k}$ using (9) for all $i, j \in \{1, \ldots, |V_k|\}$. σ_{NCut} is a scaling parameter set to some initial value.

Step 2: As a result of the partitioning, we need to identify if either A or B uniquely contains all colors in S_k. Hence, if $S_k \subseteq A$ and $S_k \cap B = \emptyset$, then eliminate all colors in B by setting $V_k = A$. If $S_k \subseteq B$ and $S_k \cap A = \emptyset$, similarly eliminate A by setting $V_k = B$. However, if S_k is not uniquely contained in either A or B, we increase σ_{NCut} and proceed back to Step 1. We keep incrementing σ_{NCut} until S_k is uniquely contained within either of A or B, and set V_k to that partition.

Algorithm 2 NCuts on FWMS Reduced Color Space

Input: $\hat{\mathbf{F}}_1, \hat{\mathbf{F}}_2, \ldots, \hat{\mathbf{F}}_K$, S_1
Output: $\mathbf{V}_1$ is returned, which contains all retained color values

```
k = K
V_k = F̂_k
Using Equation 9 build Ψ_k from V_k
while k ≠ 1 do
    σ_NCut = intial σ value
    while σ_NCut < σ_max do
        Solve for A, B by using Eq. 8
        if S_k is not uniquely contained in A or B then
            Increase σ_NCut by a factor of 10
        else
            V_k = { A, if S_k ⊆ A
                  { B, if S_k ⊆ B
        end if
        Using Equation 9 re-construct Ψ_k from V_k
    end while
    k = k − 1
    V_k = f̂_{k,i}, ∀i where f_{k+1,i} ∈ V_{k+1}
    Using Equation 9 re-construct Ψ_k from V_k
end while
return V_1
```

Step 3: Begin the process again with the new V_k until no further partitioning of the color space at level k is possible, that is, until S_k cannot be contained uniquely within a single color partition for any value of $\sigma_{\mathrm{NCut}} < \sigma_{\max}$.

Step 4: Using this process, we sequentially climb the hierarchical data structure $\hat{\mathbf{F}}_k$ where $k \in \{1, \ldots, K\}$. Thus, we migrate to the next higher image resolution, level $k-1$ and set V_{k-1} to V_k, i.e., the set of colors retained at resolution level k, and repeat the process again. We return to Step 1 until $k = 1$.

Step 5: At level 1, V_1 contains a subset of values from $\hat{\mathbf{F}}_1$, which are considered to be the chromatic values of the region of interest. Thus, the final image is computed by retaining all pixels $j \in \{1, \ldots, N\}$ such that $f_{1,j} \in V_1$, and eliminating the others.

IV. Experimental Design and Results

A. Dataset Description and Ground-Truth Annotation

Our image database comprises a total of seven digitized TMAs of OCa, in turn comprising a total of over 500 tissue cylinders from 100 patients, from which 130 were randomly selected for performing quantitative evaluation (qualitative evaluation was done on all 500). Only 130 of them were submitted to our pathologist for annotation due to the laborious nature of the work, which further motivates the utility and clinical motivation for HNCut.

The TMAs were obtained by sampling OCa tissue and were stained for the presence of the TVM ESM-1, resulting in vascular regions with the antibody to ESM-1 staining brown. The digitized version of the TMAs were obtained by scanning the slides at 40× resolution on a whole slide digital scanner, but subsequently these were down-sampled and stored at 20×

magnification. This resulted in over 500 digital images of individual cylinders, each of which were approximately 1500 × 1500 pixels in dimension. An expert pathologist annotated the precise spatial extent of the TVM on all of the 130 tissue cylinders considered for the test. Care was taken by the pathologist to include even those instances where only a few isolated pixels were picked up by the TVM.

B. Implementation

All experiments were run on a 2.8-GHz Linux machine running MATLAB 2008b with 32-GB of RAM. The setup of HNCut was as follows. All experiments were performed after converting the RGB input images to the HSV colorspace, though the algorithm is extensible to scalar valued images (such as grayscale images) as well. The FWMS was performed using $\sigma_{\mathrm{MS}} = 0.05$. NCut was subsequently performed using the Silverman function [29] to determine the value for the initial σ_{NCut}, which was then incremented by a factor of 10 as prescribed in step 9 in Algorithm 2. The IFGT's clustering variable, as suggested by Yang *et al*. [25], was set to the square root of the number of data points. When the number of remaining clusters fell below this value, it was reset to the square root of the number of remaining clusters.

The procedure that we used to enforce the ϵ distance requirement in (7) was implemented as follows. Since the human visual system is unable to easily discriminate between subtle variations of the same color, we can set ϵ to a relatively large value. The easiest way to apply this ϵ requirement in an algorithmic form is to simply choose the desired precision level (such as 10, 0, 0.01 or 0.001, depending on the format of the data) and then simply round the value to the right of that place. Since our data are stored using double precision in the range [0, 1], we have used the thousandths decimal place. The subsequent procedure of locating unique values and computing their frequencies is as simple as generating a histogram of the data values with each unique value occupying its own bin. This is a significant benefit, as the production of histograms is not only well studied but easily transformable into a parallel computing problem [30].

C. Evaluation Description

A total of four experiments were conducted to evaluate the accuracy, efficiency, and reproducibility of the HNCut algorithm, specifically in terms of its ability at 1) identifying pixels whose colors are within the swatch and also in terms of 2) identifying contiguous vascular regions annotated by the pathologist. It was felt that both pixel-level and region-level statistics were required to comprehensively and reliably evaluate HNCut performance.

1) Region-Level Metric: We define $R^{a,\varsigma}$ as the regions identified by HNCut and $R^{b,z}$ as the corresponding expert annotated regions, with $z \in \{1, \ldots, Z\}$ and $\varsigma \in \{1, \ldots, \urcorner\}$. If for any $R^{b,z}$, $\frac{|R^{b,z} \cap R^{a,\varsigma}|}{|R^{b,z}|} > 0.3$, then $R^{a,\varsigma}$ is identified as a true positive (TP). If for any $R^{a,\varsigma}$ there is no $R^{b,z}$ for which this condition is satisfied, then $R^{a,\varsigma}$ is identified as a false positive (FP). If there is a $R^{b,z}$ for which no $R^{a,\varsigma}$ can be found that satisfies the aforementioned condition, $R^{b,z}$ is deemed to be a false negative (FN). The 0.3 threshold was experimentally determined based on interactions with our expert pathologist. The complex nature of the stain shapes necessitated a lower threshold.

2) Pixel-Level Metric: Pixel-level statistics are defined using the basis of P^a and P^b, a collection of all pixels in the segmented result ($\cup_{\varsigma=1}^{\urcorner} R^{a,\varsigma}$) and the ground truth ($\cup_{z=1}^{Z} R^{b,z}$), respectively. From there, we can define the TP rate ($\frac{|P^a \cap P^b|}{|P^b|}$), positive predictive rate ($\frac{|P^b|}{|P^b|+|P^a-(P^a \cap P^b)|}$), FN rate ($\frac{|P^b-(P^a \cap P^b)|}{|P^b|}$), and true negative rate ($\frac{|C-(P^a \cup P^b)|}{|C-P^b|}$). In all cases, the $|\circ|$ notation defines the cardinality of the set.

D. Comparative Strategies

PBT was implemented as described in [12] using suggested default values for both of PBT's variables θ and ϵ (0.45 and 0.4, respectively). PBT iteratively generates a hierarchical tree structure in the training stage where each node of the tree is converted into an Adaboost classifier [31] constituting seven weak classifiers. During testing, the conditional probability of the sample belonging to the target class is calculated at each node based on the learned hierarchical tree. The discriminative model is obtained at the top of the tree by combining the probabilities associated with probability propagation of the sample at various nodes. Unlike other commonly used classifiers, such as AdaBoost [31] and decision trees [32], which provide a hard binary classification, PBT generates a posterior conditional probability value $p(1|c), p(-1|c) \in [0, 1]$, for each sample c as belonging to one of the two classes. The feature vector was created by taking a 3×3 window around every $c \in C$, across all three color channels in HSV space, resulting in a 27 dimensional vector. 1000 random positive (stained) samples and 1000 random negative (unstained and spuriously stained) samples were selected from 25 randomly selected images, resulting in a total training vector of size $27 \times 50\,000$. Training and testing was done via 50 runs of cross validation. This consisted of randomly selecting 25 images and training the classifier as described earlier, followed by testing on the other 105 images. The probabilities returned by the PBT were subjected to thresholds at 92% and 97% (represented via the first two columns in Fig. 6). The choice of thresholds was determined as follows. During each run of the randomized cross validation, a receiver operating characteristic curve (representing the trade off between sensitivity and specificity) was generated and the threshold was set at the determined operating point. This value was found to range between 92% and 97%.

E. Experiment 1: Comparison of HNCut to PBT and k-Means

1) Design: We compared the detection performance of HNCut with k-means and PBT. A standard k-means algorithm [33] was performed using ten clusters. Since k-means is not deterministic and is notoriously sensitive to the choice of cluster centers, offline experiments were performed to identify initial cluster centers (cluster centers being identified both within and outside of the target object of interest), which were qualitatively determined as being optimal.

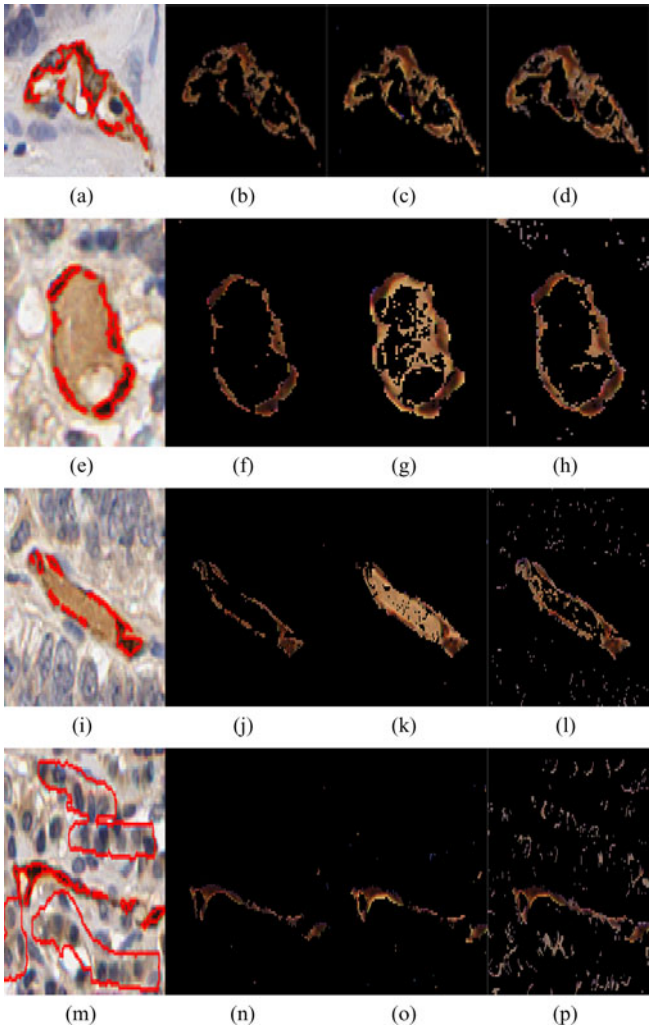

Fig. 5. (a), (e), (i), (m) Ground-truth annotations of the vascular stained areas on four different cylinders. Corresponding segmentation results from HNCut [(b), (f), (j), (n)] for $\sigma_{\mathrm{MS}} = 0.05$, PBT [(c), (g), (k), (o)] at the 97% threshold, and k-means [(d), (h), (l), (p)] using ten clusters. It can be seen that k-means always overestimates the stain extent, resulting in a large number of FPs. While PBTs perform better compared to k-means, (g) and (k) show how the PBT can occasionally retain spuriously stained pixels. On the other hand, HNCut's results closely resemble the ground truth. Note, however, that none of the algorithms are able to correctly identify the faintly stained regions in the upper portion of (m), since the stain there is barely discernible.

A subset of qualitative segmentation results are presented in Fig. 5. The first column represents the original stained TVM OCa image cropped to an area of interest, with the boundary of the ground truth highlighted by the pathologist labeled in red. The first row illustrates a case where all of the algorithms performed comparatively. The second and third rows illustrate instances where the HNCut algorithm performs better compared to PBT and k-means, both of which yield several FPs. The final row is used to illustrate a scenario where FNs occur for all three methods. The middle region for the image in Fig. 5(m) is correctly segmented in all algorithms, while the three other regions are incorrectly rejected. This specific image is a very difficult case where the stain in those regions is only barely visible to an expert. k-means results in the largest number of positives compared to the two other methods; a consequence of k-means requiring all pixels to be assigned to a cluster.

Fig. 6 quantitatively illustrates the mean and variance of the region-level metric for the different configurations across ten runs. The red line indicates the mean value across all ten runs, the blue box marks the positions where 25% of the ten values on either side of the mean are encapsulated, and the black line extends to where 75% of the values that are on either side of the mean are contained. Thus, the closer the blue and black markers are to the red mean line, the more consistent the algorithm was able to perform. HNCut provides a similar mean for FNs, while still providing a similar percentage for TPs. The FP rate for HNCut versus PBT reveals that HNCut on average yields better performance, with a much smaller variance. The threshold of 92% for the PBT encourages few FNs at the cost of many FPs.

Interestingly, randomly generating the training set for the PBT from the ground truths provided by the expert seems to lead to a larger variance in the FP metric. This can be as a result of human error in performing the ground-truth annotation, or in the selection of pixels that are not truly representative of the rest of the desired class.

It is also worth noting that k-means does quite poorly. There is no variance associated with the algorithm since we determined the optimal centers offline, thus removing the nondeterministic aspect of the scheme. Figs. 5 and 7 reveal the reason for the large number of FPs associated with k-means since it tends to retain many spuriously stained pixels as being part of the target class.

2) Pixel-Level Performance Measure: Table II quantitatively illustrates the mean and variance of the pixel-level performance measure for the different setups across 10 runs. HNCut's mean TP rate (59%) places it in between the two PBT setups (63%, 51%), while still outperforming (99%) the TP rate associated with the PBT on the two trials (98%, 98.3%). HNCut was intermediate in performance to the two runs of PBT in terms of positive predictive (36% versus 35% and 46%) and FN rates (40% versus 36% and 47%). HNCut, thus, appears to provide a good balance between precision and recall.

F. Experiment 2: Reproducibility of HNCut With Respect to Swatch and Parameter Sensitivity

The results produced by HNCut are dependent upon the selection of the swatch and the size of the σ_{MS} bandwidth parameter. Clearly, if there is a great deal of heterogeneity within the target class and the choice of swatch is not representative of the target class, the quality of the segmentation will be suboptimal. Consequently, the user has the choice of either 1) sampling additional values corresponding to the target class, or 2) repeating the segmentation with HNCut a few times with different swatches until the desired target class segmentation is obtained. Note that both tuning procedures are only made possible by the superior computational efficiency of HNCut.

1) Swatch Selection: Fig. 8 shows qualitative results reflecting the sensitivity of the segmentation as a function of the choice of the swatch. A small patch was randomly selected from the desired class by a nonexpert user. The resulting segmentation

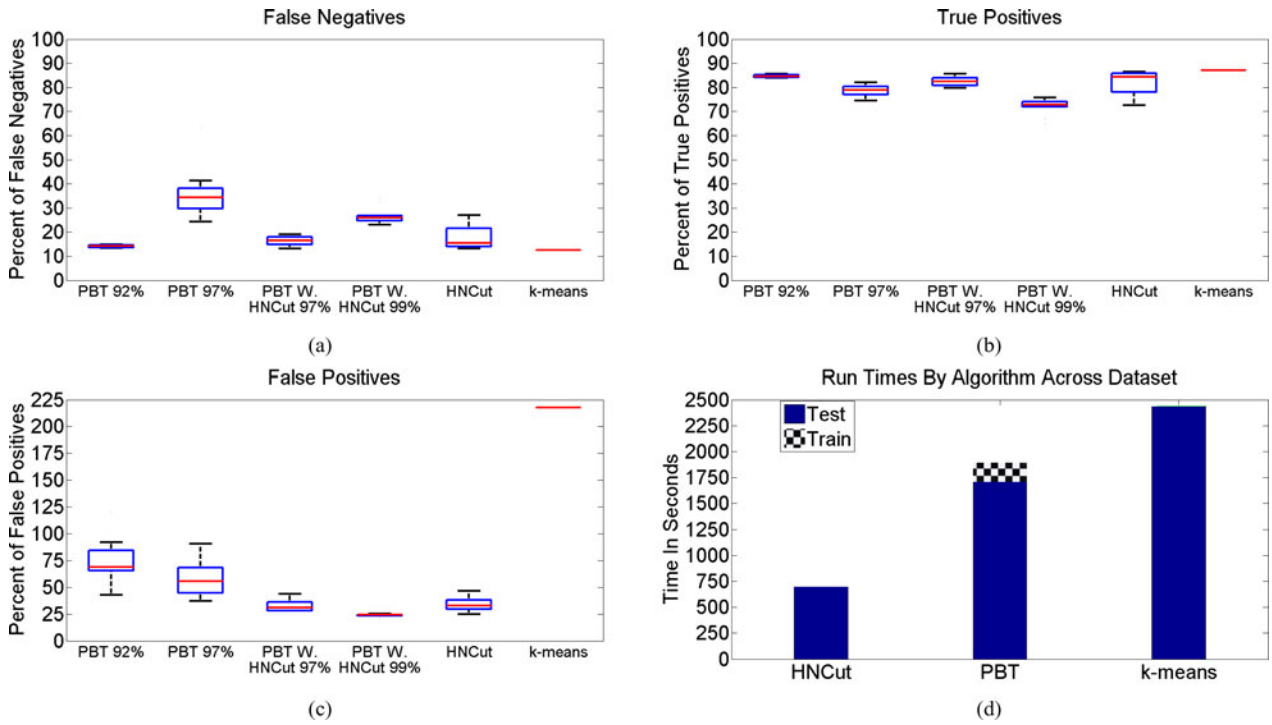

Fig. 6. Mean and variance of the region-based performance measure for (a) FNs, (b) TPs, and (c) FPs over ten runs for the PBT classifier (92% and 97% threshold), PBT classifier trained using HNCut (97% and 99% threshold), HNCut and k-means over 130 images. In the case of PBT, this involved ten runs using different training and testing sets, while for HNCut, we selected ten different swatches. The red line indicates the mean value, the blue box represents the 25th percentile of values on either side of the mean, and the black line encompasses the 75th percentile of values on either side of the mean. The plots in (a)–(c) reveal that HNcut outperforms k-means and PBT (92% and 97% thresholds) and performs comparably to a PBT trained with HNCut(97% and 99% thresholds) in terms of FPs, FNs, and TPs. (d) reveals that HNCut significantly outperforms both the PBT and k-means algorithms in terms of execution time.

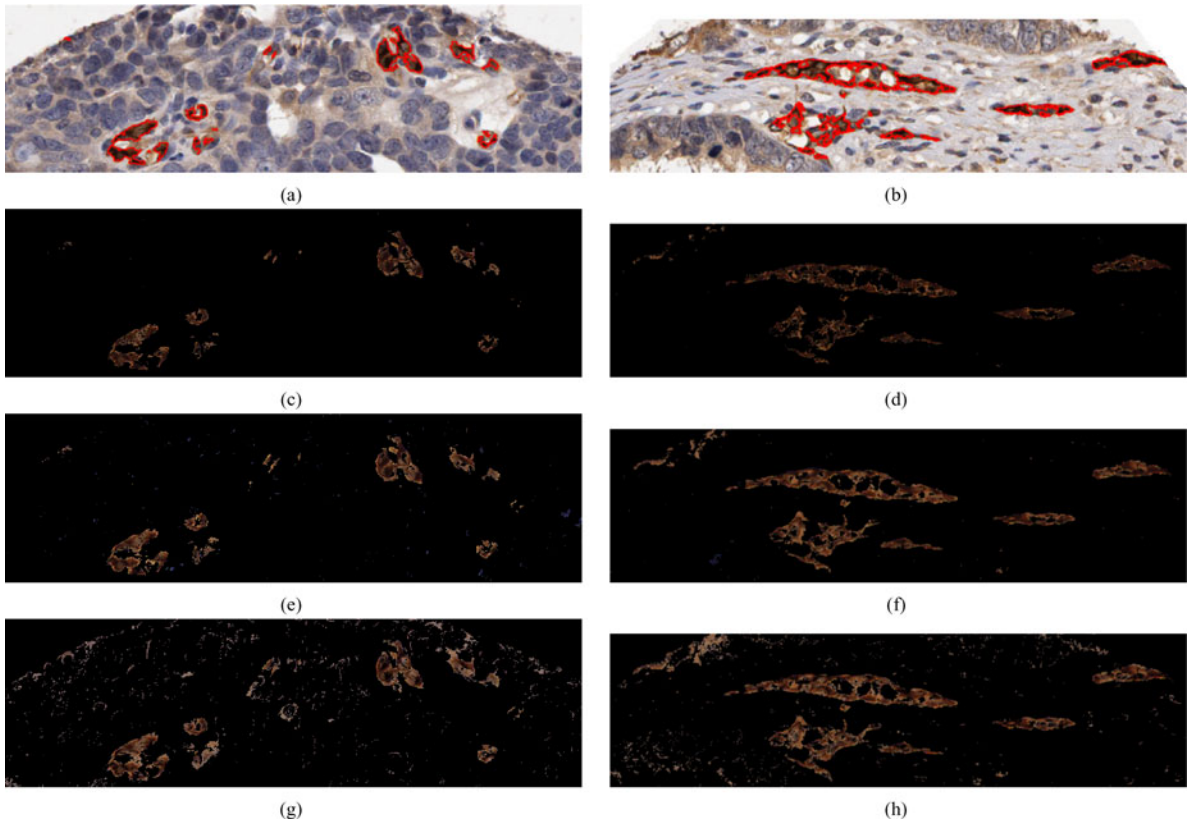

Fig. 7. Two bands across selected TMA cylinders are presented. (a), (b) original input, with the annotated ground truth in red, is presented on the top, followed by (c), (d) HNCut with $\sigma_{\mathrm{MS}} = 0.05$, (e), (f) PBT at the 97% threshold, and (g), (h) k-means using ten clusters.

TABLE II
QUANTITATIVE RESULTS PRESENTED FOR THE PIXEL-LEVEL PERFORMANCE MEASURE ACROSS ALL OF THE ALGORITHMS

	True Positive Rate	True Negative Rate	Positive Predictive Rate	False Negative Rate
HNCut	59.24%±7.36	**99.01% ± 0.56**	36.34% ±52.30	39.95% ±7.36
PBT (92%)	**62.65% ± 0.87**	98.09% ±0.16	35.23% ±14.42	35.71%±0.87
PBT (97%)	51.72%±3.26	98.33%±0.07	46.70% ±5.08	46.65%±3.26
PBT W. HNCut (97%)	58.70%±4.09	98.46% ±0.06	56.18% ±6.44	39.66% ±4.09
PBT W. HNCut (99%)	46.03% ±2.53	98.56% ±0.02	**67.60% ± 2.29**	52.34% ±2.53
k-means	71.89%	97.56%	14.34%	**27.3%**

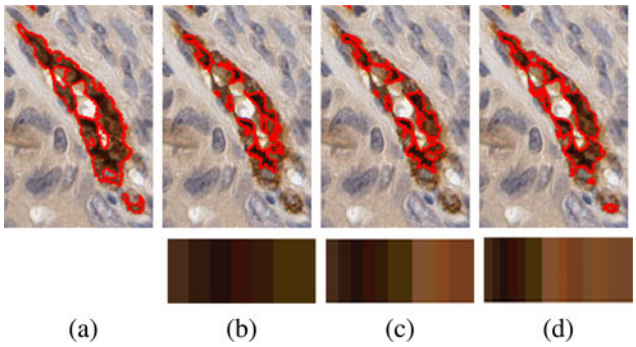

Fig. 8. (a) Ground-truth annotation of stain extent obtained from an expert pathologist. The segmentation result shown in (b) was created using a swatch comprising seven selected pixels. The next column (c) contains the same values as (b) with the addition of another five values. The final column (d) has 18 values selected from the original image. The red line encapsulates the results of the segmentation algorithm. We can see that the first set of results (b) are reasonable, but as more class representative samples are used to construct the swatch, the results improve [(c), (d)].

was overlaid using a red boundary on the original image. Subsequently, a few additional pixels were added to the swatch, and the segmentation repeated. In Fig. 8(b), we can see that when the user selects dark pixels within the swatch, the segmentation focuses on the darker aspects of the stain. When the swatch shown in Fig. 8(d) was used (a true representation of the variance in the target class), the results approached the annotations of the expert. Note that a nonexpert user could easily determine which areas of the target class were not sampled from, and include those in the swatch. This iterative process could be repeated until the nonexpert user observes the results that match the exact desired output. Once the domain swatch is selected, it can safely be used for the rest of the images in the TMA set.

2) Parameter Sensitivity: σ_{MS} is a parameter used in FWMS that is sensitive to the dataset considered. In Fig. 9, the importance of selecting the correct σ_{MS} becomes apparent. In the case where the σ_{MS} value is too large, the FWMS aggregates together pixels not contained within the swatch. As a result, they can never be pruned away as shown in (b). The highlighted blue section is dark enough in color that it becomes associated with the stain due to the large bandwidth selection. On the other hand, when the appropriate swatch representative of the desired target class is selected, almost any σ_{MS} value becomes acceptable, as shown with $\sigma_{\mathrm{MS}} = 0.01$ in Fig. 9(c). Unfortunately, in the case where a swatch that is not representative of the target class is selected, as in Figs. 9(d)–(f), the results tend to be more sensitive to the choice of value for σ_{MS}.

In our specific application, using HNCut on 500 discs, about ten of them failed to converge properly (as determined by qualitative, visual inspection), resulting in very poor segmentations. Interestingly, these ten images all had little to no stain present. By computing the variance of the color pixels in the segmented output against the domain swatch, we can assess the performance of HNCut and make relevant adjustments in an unsupervised manner. For instance, if the variance is larger then desired, adjusting σ_{MS} to a smaller value will produce new output that is more similar to the domain swatch. For all ten images considered in this experiment, the scheme for automatically adjusting σ_{MS} resulted in excellent results.

G. Experiment 3: Efficiency and Speed Considerations of HNCut

A crucially important property of HNCut is the efficiency of FWMS compared to the traditional MS. To quantitatively evaluate the computational savings in using FWMS compared to MS, the MS and FWMS procedures were executed over a total of 20 iterations and the corresponding iteration times graphed. Additionally, we compared the time it took for PBT, k-means, and HNCut to segment the 130 tissue cylinders for which quantitative evaluation was performed.

In order to clearly illustrate the high-throughput capabilities of HNCut, we compared its runtime to PBT and k-means. Fig. 6(d) illustrates a graphical representation of the results. From the onset, we can see that PBT's training time of 181 s accounts for 25% of HNCuts 643 s run time. Typically, this training time is divided amongst all of the tested samples; thus, the more samples that are tested, the cheaper it becomes to train the system. Regardless, even upon excluding the training time for PBT, HNCut still performs significantly faster. The average of 16 s per sample by PBT is easily beaten by the runtime of 6 s per sample by HNCut (for each 1500 $\times$ 1500 cylinder on the TMA). This implies that HNCut is roughly 62% faster compared to PBT.

In Table III, we can see the expected time taken to perform the classification task on different size images. On much larger images, the difference in execution time becomes even more apparent. When we compare the time needed for HNCut versus that of a human expert performing the same task, the need for a technological approach becomes apparent.

Fig. 10 shows the numerical advantages to using FWMS over MS. When the initial number of points is large, after each iteration, fewer computations need to be performed. The larger ϵ is selected, the faster FWMS will converge; on the other hand,

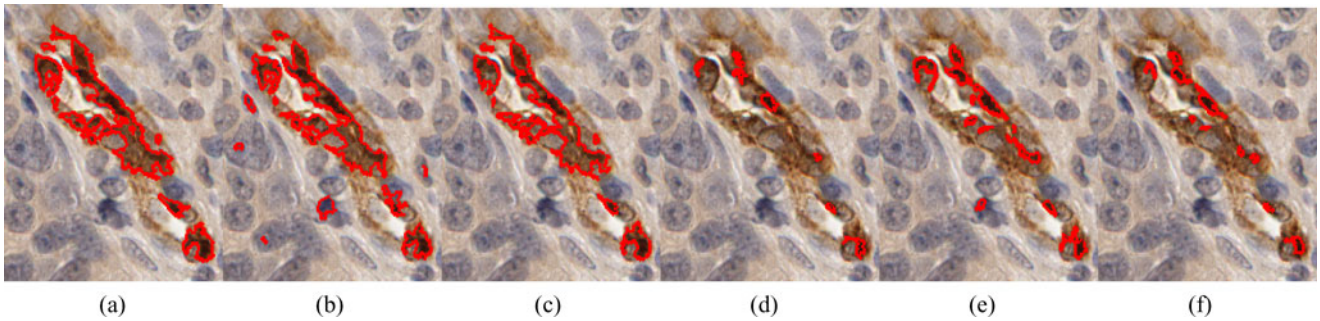

Fig. 9. (a) Ground-truth (pathologist) segmentation of stain extent. (b) and (c) Segmentation outputs for two different σ_{MS} values ($\sigma_{\mathrm{MS}} = 0.01, 0.3$). The algorithm rarely experiences unacceptable segmentations except in the case when an intentionally inappropriate value of σ_{MS} for the domain swatch is chosen. (d)–(f) are illustrated with σ_{MS} values of 0.01, 0.3, and 0.05, respectively, except that for these cases, a nonrepresentative swatch for the target class was deliberately selected.

TABLE III
RUN TIMES FOR SEGMENTATION OF IMAGES OF VARIOUS SIZES. WE CAN SEE IN ALL CASES THE HNCUT ALGORITHM PROVIDES THE BEST RUN TIMES. ALL TIMINGS WERE PERFORMED ON A 2-GHZ DUAL-CORE LAPTOP HAVING 8 GB OF RAM.

	323x323 (104,329)	646 x 646 (417,316)	1292 x 1292 (16,69,264)	2584 x 2584 (96,677,056)
Hncut	0.6s	1.2s	7.2s	25s
K-means	6.9s	30s	104s	504s
PBT	0.678s	2.9s	12.5s	46s
Ncut	43s	585s	Insufficient Memory	Insufficient Memory
Est. Manual Segmentation	30s	160s	600s	2600s

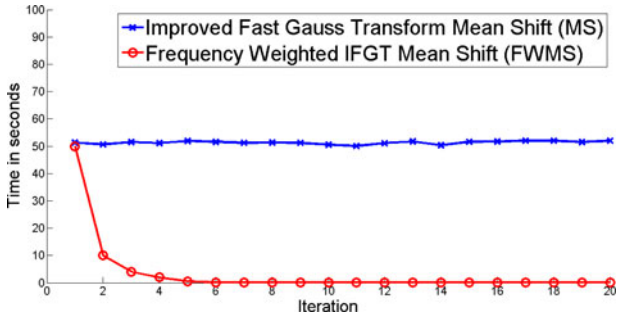

Fig. 10. Graph showing the typical time for each iteration of the MS and FWMS procedures. The original IFGT (MS) mean shift (top, in blue) has constant time for each iteration. The benefits of the FWMS algorithm (bottom, in red) become apparent within a few iterations of the clustering procedure as each additional iteration requires significantly less time as additional data points converge to the cluster mean.

when ϵ is selected to be extremely small the execution time for FWMS begins to approach that of MS.

H. Experiment 4: Using the Turing Test to Evaluate HNCut

The original Turing test [34] is a test of a machine's ability to demonstrate intelligence. A human judge is blinded and attempts to differentiate between a human and a machine using only a natural language conversation. If the judge cannot differentiate reliably between the machine and the human, the machine is said to have passed the Turing test. The question that we pose is similar: Is it possible to differentiate a supervised classifier trained with human annotated data from a supervised classifier trained with HNCut segmented data? In general, supervised methods are viewed as more dependable because they rely on training data. However, the question we pose via the Turing test is whether it is possible to use a minimally supervised method to train a supervised method and obtain on par or better results to a supervised classifier trained with manually annotated data. Toward this end, we performed ten iterations of the training/testing procedure using the HNCut output as the ground truth for training in the PBT, and compared it against the PBT output resulting from the pathologist annotated data. The choice of thresholds was determined in a similar fashion as Experiment 1, except the operating point was found to range between 97% and 99%, and thus we chose those two values.

The results presented in Fig. 6 (using the "PBT With HNCut" label) suggest that when a PBT is trained with the results from the HNCut, the results are actually superior to all other classifier configurations considered (including PBT, k-means, and HNCut), with a much smaller standard deviation. In the case of FP, the variance at the 99% threshold is almost negligible, giving a high confidence of reproducibility. As a result, the output suggests that it is possible to use HNCuts layman's initialization to produce data that is of a similar quality to the expert's laborious annotation work, minimizing user interaction. This is especially interesting because it means that the combination of the two outperforms a supervised method trained with expert data. This result suggests that supervised classifier methods can be employed for accurate quantification of biomarker extent by using HNCut to create the training set. This would be highly beneficial, avoiding the extremely expensive overhead of laboriously and manually annotating the target class. Based on these results, HNCut would appear to pass the Turing test for segmentation.

V. DISCUSSION OF SEGMENTATION ERRORS

As with any segmentation algorithm, HNCut is also subject to FP and FN errors. Below, we briefly discuss some of these errors and possible reasons for these errors.

Since the stain severity is proportional to the quantity of the biomarker, the stain will vary greatly in intensity of color across not only all cylinders but also across all stained areas themselves. This high variance is one of the reasons why thresholding and k-means type algorithms tend to do poorly. Additionally, the rim of the cylinders are often corrupted with noise which manifests

as a dark stain. The removal of these artifacts could be done by simply choosing to ignore pixels that lie on or very close to the cylinder boundary. In the situation where the disc is not well formed, either on account of tissue tearing or an absence of cells, there is the possibility for large scale pooling of FP stain within the void. Since the chromatic qualities of the FP regions are very similar to TP areas, this specific type of error is difficult to identify and eliminate.

Psammomas are calcified material within the center of a laminated whorl of elongated fibroblastic cells [14]. Unfortunately, psammomas are exactly the same in color and texture as the TPs, making it difficult for all save an expert reader to identify. In the absence of additional domain knowledge, it would be impossible for any segmentation algorithm (let alone HNCut) to distinguish these FP errors from the TPs.

VI. Concluding Remarks

In this paper, we have presented a minimally supervised segmentation scheme termed HNCuts for precise and accurate quantification of extent of vascular staining of OCa TMAs. The extent and severity of this vascular stain has been predicted to be an important prognostic marker in predicting outcome of women with OCa. The strength of HNCut is derived from the fact that it integrates the best of both an FWMS clustering and the NCut algorithm. While other schemes have been previously proposed in an attempt to combine both mean shift and NCuts, we believe that HNCut is the only approach which provides the flexibility to be able to extract different target classes based on user input. Additionally, the HNCut algorithm's hierarchical usage of the color space, a novel feature, allows it to operate faster compared to other similar approaches. By using our newly presented combination of FWMS and NCuts, and by operating in the color space, HNCut is able to handle large images efficiently. HNCut was found to be 62% faster compared to a state-of-the-art supervised classification scheme. A major advantage of HNCut, apart from its efficiency and accuracy, is that it is not encumbered by the need for precisely annotated training data. The quantitative evaluation of HNCut on 130 images corresponding to OCa TMAs (and qualitative, visual inspection on a total of 500 images), stained to identify a vascular biomarker, reveals that HNCut performs better than two popular classifier and clustering techniques—PBT and k-means. Over ten different runs of the HNCut with different swatches, it was shown that the HNCut results had lower variance compared to PBT. Additionally, when using the output from HNCut to train PBT, the results were comparable to a supervised classifier trained directly with expert annotations. Hence, HNCut its highly flexible, allowing for a lay person (nonexpert) to specify a swatch comprising a few pixels representative of the target class. Thus HNCut is ideally suited for applications in digital histopathology and biomaker discovery where the need is for an image segmentation tool to rapidly identify different types of structures or classes of interest. While HNCut had a sensitivity of over 80% in identifying the vascular regions, the question of whether this level of accuracy is acceptable for predicting disease outcome can only be answered in a clinical trial setting. In other words, only if a quantitative metric derived from the HNCut segmentation results can be correlated with patient outcome, can the HNCut results be deemed to have acceptable accuracy. Clearly, such a validation would only be possible under the auspices of a clinical trial. Since HNCut operates in the color space, and is thus highly efficient, the only limitation in the size of the image that can be analyzed by HNCut is the amount of computer memory available to read in the image data. In future work, we intend to explore the applicability of HNCut to other biomarker quantification and digital pathology problems.

References

[1] A. Madabhushi, "Digital pathology image analysis: Opportunities and challenges," *Imag. Med.*, vol. 1, pp. 7–10, 2009.

[2] G. Alexe, J. Monaco, S. Doyle, A. Basavanhally, A. Reddy, M. Seiler, S. Ganesan, G. Bhanot, and A. Madabhushi, "Towards improved cancer diagnosis and prognosis using analysis of gene expression data and computer aided imaging," *Exp. Biol. Med.*, vol. 234, pp. 860–879, 2009.

[3] H. Rui and M. J. LeBaron, "Creating tissue microarrays by cutting-edge matrix assembly," *Expert Rev. Med. Devices*, vol. 2, no. 6, pp. 673–680, Nov. 2005.

[4] R. Buckanovich, D. Sasaroli, A. O'Brien-Jenkins, J. Botbyl, R. Hammond, D. Katsaros, R. Sandaltzopoulos, L. A. Liotta, P. A. Gimotty, and G. Coukos, "Tumor vascular proteins as biomarkers in ovarian cancer," *J. Clin. Oncol.*, vol. 25, pp. 852–861, Mar. 2007.

[5] H. Vrolijk, W. Sloos, W. Mesker, P. Franken, R. Fodde, H. Morreau, and H. Tanke, "Automated acquisition of stained tissue microarrays for high-throughput evaluation of molecular targets," *J. Mol. Diagnost.*, vol. 5, no. 3, pp. 160–167, 2003.

[6] J. Wu, J. Dong, and H. Zhou, "Image quantification of high-throughput tissue microarray," in *Proc. Soc. Photo-Opt. Instrum. Eng.*, Mar. 2006, vol. 6143, pp. 509–520.

[7] A. Rabinovich, S. Krajewski, M. Krajewska, A. Shabaik, S. Hewitt, S. Belongie, J. Reed, and J. H. Price, "Framework for parsing, visualizing and scoring tissue microarray images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 10, no. 2, pp. 209–219, Apr. 2006.

[8] W. Zhong, G. Altun, R. Harrison, P. C. Tai, and Y. Pan, "Improved k-means clustering algorithm for exploring local protein sequence motifs representing common structural property," *IEEE Trans. NanoBiosci.*, vol. 4, no. 3, pp. 255–265, Sep. 2005.

[9] H. Fatakdawala, J. Xu, A. Basavanhally, 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," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 7, pp. 1676–1689, Jul. 2010.

[10] L. D. Cohen and R. Kimmel, "Global minimum for active contour models: A minimal path approach," *IEEE Int. J. Comput. Vis.*, vol. 24, no. 1, pp. 57–78, Jan. 1997.

[11] J Xu, A Janowczyk, S Chandran, and A Madabhushi, "A weighted mean shift, normalized cuts initialized color gradient based geodesic active contour model: Applications to histopathology image segmentation," *Proc. SPIE*, 7623, 76230Y, 2010, Available: http://dx.doi.org/10.1117/12.845602.

[12] Zhuowen Tu, "Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2005, pp. 1589–1596.

[13] P. Tiwari, M. Rosen, G. Reed, J. Kurhanewicz, and A. Madabhushi, "Spectral embedding based probabilistic boosting tree (sceptre): Classifying high dimensional heterogeneous biomedical data," in *Proc. Med. Image Comput. Comput. Assist. Intervention*, 2009, vol. 1, pp. 844–851.

[14] G. Carneiro, B. Georgescu, S. Good, and D. Comaniciu, "Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree," *IEEE Trans. Med. Imag.*, vol. 27, no. 9, pp. 1342–1355, Sep. 2008.

[15] 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.

[16] V. Vazirani, *Approximation Algorithms*. New York: Springer-Verlag, Mar. 2004.

[17] M. Garey and S. Johnson, *Computers and Intractability; A Guide to the Theory of NP-Completeness*. New York: Freeman, 1990.

[18] Z. Wu and R. Leahy, "An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 11, pp. 1101–1113, Nov. 1993.
[19] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," in *Proc. 3rd Int. Workshop Energy Minimizat. Methods Comput. Vis. Pattern Recognit.*, 2001, pp. 359–374.
[20] S. Dhillon, Y. Guan, and B. Kulis, "Weighted graph cuts without eigenvectors a multilevel approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 1944–1957, Nov. 2007.
[21] S. Chandran, S. Hebbar, V. Mamania, and A. Sawa, "Improved cut-based foreground identification," in *Proc. Comput. Vis., Graph. Image Process., Indian Conf.*, 2004, pp. 447–454.
[22] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Trans. Inf. Theory*, vol. 21, no. 1, pp. 32–40, Jan. 1975.
[23] 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.
[24] Y. Cheng, "Mean shift, mode seeking, and clustering," *Proc. IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790–799, Aug. 1995.
[25] C. Yang, R. Duraiswami, N.A. Gumerov, and L. Davis, "Improved fast Gauss transform and efficient kernel density estimation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2003, vol. 1, pp. 664–671.
[26] A. Janowczyk, S. Chandran, R. Singh, D. Sasaroli, G. Coukos, M. D. Feldman, and A. Madabhushi, "Hierarchical normalized cuts: Unsupervised segmentation of vascular biomarkers from ovarian cancer tissue microarrays," in *Proc. Med. Image Comput. Comput. Assist. Intervention*, 2009, vol. 1, pp. 230–238.
[27] W. Tao, H. Jin, and Y. Zhang, "Color image segmentation based on mean shift and normalized cuts," *IEEE Trans. Syst., Man, Cybern.*, vol. 37, no. 5, pp. 1382–1389, Oct. 2007.
[28] X. Yuan, B. G. Hu, and R. He, "Agglomerative mean-shift clustering via query set compression," in *Proc. SIAM Int. Conf. Data Mining*, 2009, pp. 221–232.
[29] B. W. Silverman, *Density Estimation, for Statistics and Data Analysis*. London, U.K.: Chapman & Hall, 1986.
[30] A. Zomaya, Ed., *Parallel and Distributed Computing Handbook*. New York: McGraw-Hill, 1996.
[31] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *J. Comput. Syst. Sci.*, vol. 55, no. 1, pp. 119–139, 1997.
[32] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Trans. Syst. Man Cybern.*, vol. 21, no. 3, pp. 660–674, May/Jun. 1991.
[33] S. P. Lloyd, "Least square quantization in PCM," *IEEE Trans. Inf. Theory.*, vol. 28, no. 2, pp. 129–137, Mar. 1982.
[34] A. Turing, "Computing machinery and intelligence," *Mind*, vol. LIX-236, pp. 433–460, 1950.
[35] J. V. Johannessen and M. Sobrinho-Simoes, "The origin and significance of thyroid psammoma bodies," *Lab Invest*, vol. 43, pp. 287–296, 1980.
[36] W. Chen, R. Reiss, and D. Foran, "Unsupervised imaging, registration and archiving of tissue microarrays," in *Proc. Amer. Med. Inf. Assoc.*, 2002, vol. 1, pp. 136–139.
[37] D. Rosen, X. Huang, M. Deavers, A. Malpica, E. Silva, and J. Liu, "Validation of tissue microarray technology in ovarian carcinoma," *Modern Pathol.*, vol. 17, pp. 790–797, 2004.
[38] W. Chen, M. Reiss, and D.J. Foran, "A prototype for unsupervised analysis of tissue microarrays for cancer research and diagnostics," *IEEE Trans. Inf. Technol. Biomed.*, vol. 8, no. 2, pp. 89–96, Jun. 2004.
[39] U. Ozertem, D. Erdogmus, and R. Jenssen, "Mean shift spectral clustering," *Pattern Recognit.*, vol. 41, no. 6, pp. 1924–1938, 2008.
[40] O. Kallioniemi, U. Wagner, J. Kononen, and G. Sauter, "Tissue microarray technology for high-throughput molecular profiling of cancer," *Human Mol. Genet.*, vol. 10, no. 7, pp. 657–662, 2001.
[41] S. Ronald and S. Guido, "Tissue microarray (TMA) applications: implications for molecular medicine," *Expert rev. Mol. Med.*, vol. 5, no. 26, pp. 1–12, 2003.
[42] Z. Zhang, J. Zhang, and H. Xue, "Improved k-means clustering algorithm," *Image Signal Process., Congr.*, vol. 5, pp. 169–172, May 2008.
[43] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
[44] S. Doyle, M. Hwang, K. Shah, A. Madabhushi, M. Feldman, and J. Tomaszeweski, "Automated grading of prostate cancer using architectural and textural image features," in *Proc. IEEE Int. Symp. Biomed. Imag.: Nano Macro*, Apr. 2007, pp. 1284–1287.
[45] S. Doyle, A. Madabhushi, M. Feldman, and J. Tomaszeweski, "A boosting cascade for automated detection of prostate cancer from digitized histology," in *Proc. Med. Image Comput. Comput. Assist. Intervent.*, 2006, pp. 504–511.
[46] B. Karacali and A. Tözeren, "Automated detection of regions of interest for tissue microarray experiments: An image texture analysis," *BMC Med. Imag.*, vol. 7, p. 2, Mar. 2007. Available: www.biomedcentral.com/471-2342/7/2.
[47] H. Rui and M. J. LeBaron, "Creating tissue microarrays by cutting-edge matrix assembly," *Expert Rev. Med. Devices*, vol. 2, no. 6, pp. 673–680, Nov. 2005.
[48] S. Petushi, C. Katsinis, C. Coward, Z. Garcia, and A. Tözeren, "Automated identification of microstructures on histology slides," in *Proc. IEEE Int. Symp. Biomed. Imag.: Macro Nano*, 2004, vol. 1, pp. 424–427.
[49] M. Datar, D. Padfield, and H. Cline, "Color and texture based segmentation of molecular pathology images using hsoms," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2008, pp. 292–295.
[50] T. Amaral, S. J. McKenna, K. Robertson, and A. Thompson, "Classification of breast-tissue microarray spots using texton histograms," in *Proc. Med. Image Understand. Anal.*, Jul. 2008, pp. 144–148.
[51] Y. Wang, J. Yang, and N. Peng, "Unsupervised color-texture segmentation based on soft criterion with adaptive mean-shift clustering," *Pattern Recogit. Lett.*, vol. 27, no. 5, pp. 386–392, 2006.
[52] J. Monaco, J. E. Tomaszewski, M. D. Feldman, M. Moradi, P. Mousav, A. Boag, C. Davidson, P. Abolmaesumi, and A. Madabhushi, "Probabilistic pairwise markov models: Application to prostate cancer detection," *Soc. Photo-Opt. Instrum. Eng.*, vol. 7260, 2009.
[53] S. K. Nath, K. Palaniappan, and F. Bunyak, "Cell segmentation using coupled level sets and graph-vertex coloring," in *Proc. Med. Image Comput. Comput. Assist. Intervent.*, 2006, pp. 101–108.
[54] K. A. Mendhurwar, R. Kakumani, and V. Devabhaktuni, "Microarray image segmentation using chan-vese active contour model and level set method," in *Proc. IEEE Conf. Eng. Med. Biol. Soc.*, 2009, vol. 1, pp. 3629–3632.
[55] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
[56] S. Doyle, J. Tomaszewski, M. Feldman, and A. Madabhushi, "A boosted bayesian multi-resolution classifier for prostate cancer detection from digitized needle biopsies," *IEEE Trans. Biomed. Eng.*, vol. PP, no. 99, pp. 1–1, Jun. 21, 2010, doi: 10.1109/TBME.2010.2053540.
[57] J. P. Monaco, J. E. Tomaszewski, M. D. Feldman, I. Hagemann, M. Moradi, P. Mousavi, A. Boag, C. Davidson, P. Abolmaesumi, and A. Madabhushi, "High-throughput detection of prostate cancer in histological sections using probabilistic pairwise Markov models," *Med. Image Anal.*, vol. 14, no. 4, pp. 617–629, 2010.
[58] A. Basavanhally, S. Doyle, and A. Madabhushi, "Predicting classifier performance with a small training set: Applications to computer-aided diagnosis and prognosis," in *Proc. IEEE Int. Symp. Biomed. Imag.: Nano Macro*, 2010, pp. 229–232.

Authors' photographs and biographies not available at the time of publication.