Detection of Prostate Cancer from Whole-Mount Histology Images Using Markov Random Fields

James P. Monaco¹, John E. Tomaszewski², Michael D. Feldman², Mehdi Moradi³, Parvin Mousavi³, Alexander Boag⁴, Chris Davidson⁴, Purang Abolmaesumi³, Anant Madabhushi¹

¹Department of Biomedical Engineering, Rutgers University, USA.
²Department of Surgical Pathology, University of Pennsylvania, USA.
³School of Biomedical Engineering, Queen’s University, Canada.
⁴Department of Computing, Queen’s University, Canada.

Abstract—Annually in the US 186,000 men are diagnosed with prostate cancer (CaP) and over 43,000 die from it. The analysis of whole-mount histological sections (WMHSs) is needed to help determine treatment following prostatectomy and to create the “ground truths” of CaP spatial extent required to evaluate other diagnostic modalities (eg, magnetic resonance imaging). Computer aided diagnosis (CAD) of WMHSs could increase analysis throughput and offer a means for identifying image based biomarkers capable of distinguishing, for example, CaP progressors from non-progressors. In this paper we introduce a novel methodology that allows the MRF to be modeled directly from training data. Our CAD system identifies cancerous regions with a sensitivity of 0.8670 and a specificity of 0.9524.

I. INTRODUCTION

Annually in the US 186,000 people are diagnosed with prostate cancer (CaP) and over 43,000 die from it. The examination of histological specimens remains the definitive test for diagnosing CaP. Though the majority of such analysis is performed on core biopsies, the consideration of whole-mount histological sections (WMHSs) is also important. Following prostatectomy, the staging and grading of WMHSs help determine prognosis and treatment. Additionally, the spatial extent of CaP as established by the analysis of WMHSs can be registered to other modalities (eg. magnetic resonance imaging), providing a “ground truth” for evaluation. The development of computer aided diagnosis (CAD) algorithms for WMHSs is also significant: 1) CAD offers a viable means for analyzing the vast amount of the data present in WMHSs, a time-consuming task currently performed by pathologists, 2) the consistent, quantified features and results inherent to CAD systems can be used to refine our own understanding of prostate histology, thereby helping doctors improve performance and reduce variability in grading and detection, and 3) the data mining of quantified morphometric features may provide means for biomarker discovery, enabling for example, the discrimination of CaP progressors from non-progressors.

With respect to prostate histology Begelman [1] considered nuclei segmentation for hematoxylin and eosin (H&E) stained prostate tissue samples. In [2] Naik used features derived from the segmentation of nuclei and glands to determine Gleason grade in core biopsy samples. To aid in manual cancer diagnosis Gao [3] applied histogram thresholding to enhance the appearance of cytoplasm and nuclei. In this paper we introduce the first CAD system for detecting CaP in WMHSs. This system is specifically designed to operate at low-resolution (0.01 mm² per pixel) and will eventually constitute the initial stage of a hierarchical analysis algorithm, identifying areas for which a higher-resolution examination is necessary. As substantiated in our previous approach [4] for prostate biopsy specimens, a hierarchical methodology provides an effective means for dealing with high density data (prostate WMHSs have 500 times the amount of data compared to a four view mammogram). Even at low resolutions, gland size and morphology are noticeably different in cancerous and benign regions [5]. In fact, we will demonstrate that gland size alone is a sufficient feature for yielding an accurate and efficient algorithm. Additionally, we leverage the fact that cancerous glands tend to be proximate to other cancerous glands. This information is modeled using Markov random fields (MRFs). However, unlike most MRF strategies which rely on heuristic formulations, we introduce a novel methodology that allows the MRF to be modeled directly from training data. Our CAD system identifies cancerous regions with a sensitivity of 0.8670 and a specificity of 0.9524.

This work was made possible due to grants from the Wallace H. Coulter foundation, New Jersey Commission on Cancer Research, the National Cancer Institute (R21CA127186-01, R03CA128081-01), the Society for Imaging and Informatics on Medicine, and the Life Science Commercialization Award.

Corresponding authors: James Monaco, email: jpmonaco@rci.rutgers.edu, Anant Madabhushi email: anantm@rci.rutgers.edu


**II. METHODOLOGY**

**A. Gland Segmentation**

In the luminance channel of histological images glands appear as regions of contiguous, high intensity pixels circumscribed by sharp, pronounced boundaries. To segment these regions we adopt a routine first used for segmenting breast microcalcifications [6]. We briefly outline this algorithm. First define the following: 1) **current region** (CR) is the set of pixels representing the segmented region in the current step of the algorithm, 2) **current boundary** (CB) is the set of pixels that neighbor CR in an 8-connected sense, but are not in CR, and 3) **internal boundary** (IB) is the subset of pixels in CR that neighbor CB. These definitions are illustrated in Figure 1. The growing procedure begins by initializing CR to a seed pixel assumed to lie within the gland. At each iteration CR expands by aggregating the pixel in CB with the greatest intensity. CR and CB are updated, and the process continues. The algorithm terminates when the $L_{\infty}$ norm from the seed to the next aggregated pixel exceeds a predetermined threshold. That is, the $L_{\infty}$ norm establishes a square bounding box about the seed; the growing procedure terminates when the algorithm attempts to add a pixel outside this box. During each iteration the algorithm measures the boundary strength which is defined as the average intensity of the pixels in CB minus the average intensity of the pixels in CB. After the growing procedure terminates, the region with the greatest boundary strength is selected. Seed pixels are established by finding peaks in the image after Gaussian smoothing. Since gland sizes can vary greatly, we smooth at multiple scales, each of which is defined by the sigma $\sigma_g \in \{0.2, 0.1, 0.05, 0.025\}$ mm of a Gaussian kernel.\(^1\) The length $l$ of each side of the bounding box used for terminating the segmentation step is tied to the scale: $l = 12\sigma_g$. The final segmented regions may overlap. In this event the region with the highest boundary measure is retained. Figures 2(a) and 2(g) are H&E stained WMHSs with black ink marks providing a rough truth (RT) of CaP extent. Gland segmentation results are shown in Figures 2(b) and 2(h). Figures 2(c) and 2(i) provide magnified views of the regions of interest in Figures 2(b) and 2(h). The centroids of glands whose probability of malignancy exceeds $\rho = 0.15$ are marked with green dots in Figures 2(d) and 2(j). This labeling is refined by the MRF iteration, producing the gland centroids shown in Figures 2(e) and 2(k). Figures 2(f) and 2(l) show the aggregation of cancerous glands into regions (green) along with a high-fidelity truth (HFT) of CaP extent (yellow).

**B. Feature Extraction, Modeling, and Bayesian Classification**

Gland area is used to discriminate benign from malignant glands. Since we employ a Bayesian framework, we require estimates of the conditional probability density functions (pdfs) of gland area for both malignant $\omega_m$ and benign $\omega_b$ glands. Using the equivalent square root of gland area (SRGA), the pdfs $f(y|\omega_m)$ and $f(y|\omega_b)$ can be accurately modeled with a weighted sum of gamma distributions:

$$f(y; \theta, k, \lambda) = \lambda g^{k-1} \frac{e^{-y/\theta_1}}{\theta_1^k \Gamma(k_1)} + (1-\lambda) g^{k-1} \frac{e^{-y/\theta_2}}{\theta_2^k \Gamma(k_2)},$$

where $y > 0$ is the SRGA, $\lambda \in [0, 1]$ is the mixing parameter, $k_1, k_2 > 0$ are the shape parameters, $\theta_1, \theta_2 > 0$ are the scale parameters, and $\Gamma$ is the Gamma function. Note, we use $f$ to indicate a continuous pdf and $p$ to denote a discrete probability mass function (pmf). A Bayesian classifier uses these pdfs to calculate the probability of malignancy for each gland. Those glands whose probabilities exceed the predetermined threshold $\rho$ are labeled malignant; the remainder are classified as benign (Figures 2(d) and 2(j)).

**C. Improved Classification Using Markov Random Fields**

In addition to glandular features such as area, a highly indicative trait of cancerous glands is their proximity to other cancerous glands. This can be modeled using MRFs.

1) **Formulation of Gland Proximity as a MRF**: Let $S = \{s_1, s_2, \ldots, s_N\}$ represent a set of $N$ unique sites corresponding to the $N$ segmented glands. Let each site $s \in S$ have an associated random variable $X_s \in \{\omega_m, \omega_b\}$ indicating its state as either malignant or benign. To refer collectively to the states of all glands we have $X = \{X_s : s \in S\}$. Each state $X_s$ is unknown; we only observe an instance of the random variable $Y_s \in \mathbb{R}^D$ representing the $D$ dimensional feature vector associated with gland $s$. Though our algorithm is extensible to any number of features, currently $D = 1$ with $Y_s$ being the SRGA. To collectively refer to the entire scene of feature vectors we have $Y = \{Y_s : s \in S\}$.

\(^1\)The growing procedure operates on the original image.

\(^2\)Figures 2(d), 2(i), 2(e), 2(k), 2(f) and 2(l) will be further explained in later sections.

---

Fig. 1. Example of current segmented region (CR), internal boundary (IB), and current boundary (CB) during a step of the region growing algorithm.

Fig. 2. Examples of gland segmentation results and labels, with labeled centroids shown as green dots.

Fig. 3. Graph with six sites and binary states.
edges. A local neighborhood \( \eta_s \) is defined as follows: 
\[
\eta_s = \{ r : r \in S, r \neq s, \{r, s\} \in E \}.
\]

The set of all local neighborhoods establishes a neighborhood structure: 
\[
\eta = \{ \eta_s : s \in S \}.
\]
A clique is a set of the vertices of any fully connected subgraph of \( G \). The set \( \mathcal{C} \) contains all possible cliques. These concepts are best understood in the context of an example. The graph in Figure 3 has sites \( S = \{1, 2, 3, 4, 5, 6\} \) and edges \( E = \{\{1, 2\}, \{1, 4\}, \{1, 5\}, \{2, 3\}, \{2, 6\}, \{4, 5\}\}. \) The neighborhood of site 5, for example, is \( \eta_5 = \{1, 4\} \). There are six one-element cliques \( \mathcal{C}_1 = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\} \), six two-element cliques \( \mathcal{C}_2 = E \), and one three-element clique \( \mathcal{C}_3 = \{\{1, 4, 5\}\} \). The set \( \mathcal{C} \) is the union of these three
sets. The state $X_s$ of each site is either black or white, i.e. $\Lambda = \{b, w\}$. Our specific neighborhood structure is determined by the distance between gland centroids. If $m_s$ denotes the centroid of gland $s$, then $r \in \eta_s$ if $\|m_s - m_r\| < d$. Motivated by the pathology we choose $d = 0.7$mm.

To simplify notation we use $Pr \{X_r = x_r, X_s = x_s\} \equiv p(x_r, x_s)$ for indicating the probability of a specific event, where $x_r, x_s \in \{\omega_m, \omega_b\}$. If $X$ is a MRF with respect to the neighborhood structure $\eta$, then $X$ satisfies the local Markov property $p(x_r, x_s) = p(x_r | x_s)$, where $x_s$ indicates $x$ without $x_r$ and $x_{\eta_s} = \{x_s : s \in \eta_s\}$. Additionally, $X$ is a MRF with respect to $\eta$ if and only if $p(x)$ is a Gibbs distribution [7]: $p(x) = \prod_{c \in \mathcal{C}} V_c(x)$, where $V_c$ are nonzero functions that depend only on those $x_s$ for which $s \in c$. The local conditional probabilities follow directly:

$$p(x_r, x_{\eta_s}) = \frac{\prod_{c \in \mathcal{C}} V_c(x)}{\sum_{x_{\eta_s}} \prod_{c \in \mathcal{C}} V_c(x)} = \frac{p(x_r | x_s) \prod_{c \in \mathcal{C}} V_c(x)}{\sum_{x_{\eta_s}} \prod_{c \in \mathcal{C}} V_c(x)} = \frac{p(x_r | x_s) p(s)}{\sum_{s} p(s)}.$$

This distribution has the same form as the Gibbs distribution for $p(x)$, but now the product is only over those cliques $c$ that contain $s$.

2) Integration of Data Derived PMFs into the MRF: Since it is difficult to derive Gibbs distributions that model a set of training data, generic models are usually assumed. The most prevalent formulation is the Potts model which is defined as follows for two-element cliques:

$$V_{\{r,s\}}(x_r, x_s) = \begin{cases} e^{-\beta} & \text{if } x_r = x_s \text{ and } \{r,s\} \in \mathcal{C} \\ e^\beta & \text{if } x_r \neq x_s \text{ and } \{r,s\} \in \mathcal{C}. \end{cases}$$

(3)

The Potts model disregards all cliques having more or less than two elements, i.e. if $|c| \neq 2$ we have $V_c(x) = 1$, where $|\cdot|$ signifies cardinality. Such generic models are unnecessary; assuming that all $X_s$ are i.i.d. and all $X_c$ given $X_s$ are i.i.d. for every $r \in \eta_s$, we can determine the appropriate $V_c$ directly from the data. To our knowledge, the following equations represent the first proposed means for incorporating arbitrary pmfs into the MRF structure:

$$V_s(x_s) = p(x_s)^{1-|\eta_s|} \text{ for } s \in S$$

(4)

$$V_{\{r,s\}}(x_r, x_s) = p(x_r, x_s) \text{ for } \{r,s\} \in \mathcal{C}. \text{ (5)}$$

The functions $V_c$ for higher-order cliques are identically one. The validity of (4) and (5) can be seen by inserting them into (2):

$$\prod_{s \in \mathcal{C}} V_c(x) = \frac{p(x_r)^{1-|\eta_s|} \prod_{r \in \eta_s} p(x_r, x_s) \prod_{c \in \mathcal{C}} p(x_s)}{\sum_{x_{\eta_s}} \prod_{c \in \mathcal{C}} V_c(x)} = \frac{p(x_r, x_s)}{\sum_{x_{\eta_s}} \prod_{c \in \mathcal{C}} V_c(x)} = \frac{p(x_r, x_s)}{\sum_{x_{\eta_s}} \prod_{c \in \mathcal{C}} V_c(x)} = p(x_r, x_s).$$

The determination of $p(x_r, x_s)$ and $p(x_r, x_s)$ from training data is straightforward. For example, consider the randomly selected two-element clique $\{r, s\}$ where both sites are malignant. The probability $V_{\{r,s\}}(\omega_m, \omega_m) = p(\omega_m, \omega_m)$ can be found by examining all permutations of two neighboring glands and determining the probability in which both are cancerous. The pmf $p(x_r, x_s)$ is the marginal mass function of $p(x_r, x_s)$. Since $p(x_r, x_s)$ is symmetric, both marginals are identical.

3) Label Estimation and Aggregation: The goal is to estimate the hidden states $X$ given the observations $Y$ using maximum a posteriori (MAP) estimation, i.e. maximizing $p(x|y)$ over $x$. Bayes laws yields $p(x|y) \propto f(y|x)p(x)$, where $\propto$ signifies proportionality. The Iterated Conditional Modes (ICM) [8] algorithm indicates that the maximization of $p(x|y)$ need not occur at all sites simultaneously; we can perform MAP estimation on each site individually by maximizing $p(x_r, x_{\eta_s}) \propto f(y_s|x_s)p(x_s, x_{\eta_s})$. After estimating each individual state $X_s$, the entire scene of states $X$ is updated. The process iterates until convergence, usually requiring only five or six iterations (Figures 2(c) and 2(k)).

Our ultimate objective is to delineate the spatial extent of the cancerous regions. Following the neighborhood structure defined in Section II-C.1, each gland centroid can be considered the center of a disk of diameter $d$. If the disks of two centroids overlap, they are considered neighbors. This leads to a simple formulation for cancerous regions; the union of all disks of diameter $d$ centered at the centroids of the malignant glands (Figures 2(f) and 2(l)).

III. RESULTS AND DISCUSSION

A. Data and CAD Training

The data consists of four H&E stained prostate WMHSs obtained from different patients. An initial pathologist used a black marker to delineate a very rough truth (RT) of CaP extent. An second pathologist performed a more detailed annotation of the digitized slices, producing a high fidelity truth (HFT). The digital images have a resolution of 0.01mm² per pixel. The approximate image dimensions are 2.1×3.2 cm, i.e. 2100×3200 pixels. The training step involves estimating the parameters for the SRGA pdfs $f(y|\omega_m)$ and $f(y|\omega_b)$ using (1) and determining the MRF pmfs in (4) and (5). The training/testing procedure uses a leave-one-out strategy.

B. Quantitative Results

We first assess the ability of the CAD system to discriminate malignant and benign glands. A gland is considered cancerous if its centroid falls within the HFT. The performance of the classification step described in Section II-B varies as the threshold $\rho$ increases from zero to one, yielding the receiver operator characteristic (ROC) curve (solid) in Figure 4. Since the resulting classification serves as the initial condition for the MRF iteration, the MRF performance also varies as a function of $\rho$, producing the dashed ROC curve in Figure 4. The operating points for the qualitative results in Figure 2 are shown by the ring and dot ($\rho = 0.15$). We next

\footnote{The quality of the gland segmentation is implicit in this performance measure.}
measure the accuracy of the final CAD regions produced at this operating point by comparing them with the HFT. The sensitivity (the ratio of the cancerous area correctly marked to the total cancerous area) and specificity (the ratio of benign area correctly marked to the total benign area) are 0.8670 and 0.9524, respectively.

Fig. 4. The solid ROC curve describes the performance (obtained over four studies) of the initial gland classification (without MRF) as the probability threshold $\rho$ varies from zero to one. The classification results at each threshold $\rho$ are passed to the MRF iteration, yielding the dashed ROC curve. The operating points for the qualitative results in Figure 2 are shown by the ring and dot ($\rho = 0.15$).

C. Qualitative Results

Qualitative results for two WMHSs were previously presented as Figure 2.

IV. CONCLUDING REMARKS

Over a cohort of four studies we have demonstrated a simple, powerful, and rapid –requiring only four to five minutes to analyze a 2100×3200 image on 2.4 Ghz Intel Core2 Duo Processors– method for the detection of CaP from low-resolution whole-mount histology specimens. Relying only on gland size and proximity, the CAD algorithm highlights the effectiveness of embedding physically meaningful features in a probabilistic framework that avoids heuristics. Additionally, we introduced a novel method for formulating data derived pmfs as Gibbs distributions, obviating the need for generic MRF models.

REFERENCES