

Gland Segmentation and Computerized Gleason Grading of Prostate Histology by Integrating Low-, High-level and Domain Specific Information

Shivang Naik¹, Scott Doyle¹, Michael Feldman², John Tomaszewski², Anant Madabhushi¹

¹Rutgers, The State University of New Jersey, Piscataway, NJ, 08854

²The University of Pennsylvania, Philadelphia, PA, 19104

Abstract—In this paper we present a method of automatically detecting and segmenting glands in digitized images of prostate histology and to use features derived from gland morphology to distinguish between intermediate Gleason grades. Gleason grading is a method of describing prostate cancer malignancy on a numerical scale from grade 1 (early stage cancer) through grade 5 (highly infiltrative cancer). Studies have shown that gland morphology plays a significant role in discriminating Gleason grades. We present a method of automated detection and segmentation of prostate gland regions. A Bayesian classifier is used to detect candidate gland regions by utilizing low-level image features to find the lumen, epithelial cell cytoplasm, and epithelial nuclei of the tissue. False positive regions identified as glands are eliminated via use of domain-specific knowledge constraints. Following candidate region detection via low-level and empirical domain information, the lumen area is used to initialize a level-set curve, which is evolved to lie at the interior boundary of the nuclei surrounding the gland structure. Features are calculated from the boundaries that characterize the morphology of the lumen and the gland regions, including area overlap ratio, distance ratio, standard deviation and variance of distance, perimeter ratio, compactness, smoothness, and area. The feature space is reduced using a manifold learning scheme (Graph Embedding) that is used to embed objects that are adjacent to each other in the high dimensional feature space into a lower dimensional embedding space. Objects embedded in this low dimensional embedding space are then classified via a support vector machine (SVM) classifier as belonging to Gleason grade 3, grade 4 cancer, or benign epithelium. We evaluate the efficacy of the automated segmentation algorithm by comparing the classification accuracy obtained using the automated segmentation scheme to the accuracy obtained via a user assisted segmentation scheme. Using the automated scheme, the system achieves accuracies of 86.35% when distinguishing Gleason grade 3 from benign epithelium, 92.90% distinguishing grade 4 from benign epithelium, and 95.19% distinguishing between Gleason grades 3 and 4. The manual scheme returns accuracies of 95.14%, 95.35%, and 80.76% for the respective classification tasks, indicating that the automated segmentation algorithm and the manual scheme are comparable in terms of achieving the overall objective of grade classification.

I. INTRODUCTION

Currently, diagnosis of prostate cancer is done by manual visual analysis of prostate tissue samples that have been obtained from a patient via biopsy. If a cancerous region is found in a tissue sample, the Gleason grading scheme [1] is used to assign a numerical grade to the tissue to characterize the degree of malignancy. Gleason grades range between 1

(relatively benign tissue) and 5 (highly aggressive, malignant tissue) based on qualitative tissue characteristics such as the arrangement of nuclei and the morphology of gland structures in the image. Changes in the level of malignancy of the tissue alter the appearance of the tissue: the glands in the cancerous region become small, regular, and more tightly packed as cancer progresses from benign to highly malignant [2]. The Gleason grade is used to assist in directing treatment of the patient. Typical prostate tissue images are shown in Figure 1 corresponding to benign epithelium (Fig. 1 (a)), Gleason grade 3 (Fig. 1 (b)), and Gleason grade 4 tissues (Fig. 1 (c)). Each image contains one or more gland regions.

While the Gleason system has been in use for many decades, its dependence on qualitative features leads to a number of issues with standardization and variability in diagnosis. Intermediate Gleason grades (3 or 4) are often under- or over-graded, leading to inter-observer variability up to 48% of the time [3]. Our group has been developing schemes for automated grading of prostate histology to reduce the inter- and intra-observer variability. In [4], we presented a novel set of features for prostate cancer detection and grading on prostate histopathology based on architectural, textural, and morphological features calculated from the image. Glandular features, in particular the size and roundness factor have been shown useful in distinguishing cancerous versus benign tissue [5]. In [6], we investigated the ability of morphology-based features obtained from manually delineated gland margins to classify the tissue and found that these features were capable of achieving approximately 73% accuracy when distinguishing between Gleason grade 3 and grade 4. However, to develop a fully automated feature extraction algorithm, the glands must be detected and segmented without user interaction.

The majority of histological image segmentation work has been done on the detection of nuclei, as they are clearly marked on histology when using a variety of staining techniques. Often, image thresholding is used to segment the nuclei, such as in the algorithm developed by Korde, et al. for bladder and skin tissue [7]. Other algorithms have been proposed using more complex techniques, such as an active contour scheme for pap-stained cervical cell images by Bamford and Lovell [8] and a fuzzy logic engine proposed by Begelman, et al. for prostate tissue that uses both color and shape-based constraints [9]. However, these studies focus only on finding individual nuclei. Segmentation of multiple structures on prostate histology has been done by Gao, et al. using color

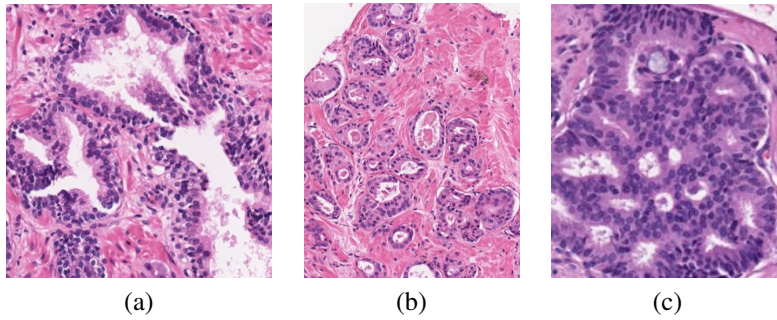


Fig. 1. Examples of tissue from (a) benign epithelium, (b) Gleason grade 3, and (c) Gleason grade 4.

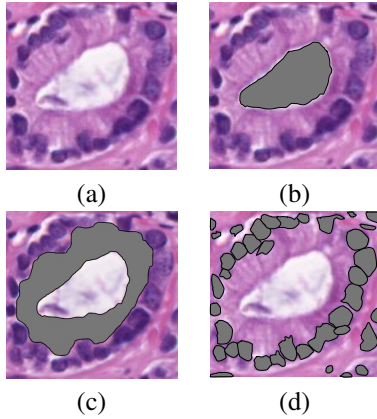


Fig. 2. Examples of the regions of interest of a gland structure (a). Shown outlined in black are the (b) lumen area, (c) epithelial cytoplasm, and (d) epithelial nuclei.

histogram thresholding to enhance regions of cytoplasm and nuclei in an tissue to aid in manual cancer diagnosis [10].

To the best of our knowledge, there have been no studies on the automated segmentation of whole gland structures for prostate histopathology. The difficulty arises in the varied appearance of gland structures across different classes of images. Figure 2 illustrates the three regions of interest that make up a gland (Fig. 2 (a)): (i) a lumen region surrounded by epithelial cells (Fig. 2 (b)), (ii) the cytoplasm of the cells bordering the lumen (Fig. 2 (c)), and (iii) the nuclei of the surrounding cells (Fig. 2 (d)). These structures undergo a variety of changes in cancer progression, as noted by the Gleason scheme. As the Gleason grade increases, the glands start to fuse, become regular, tightly packed, and smaller in size. However, despite these changes, all gland regions share certain characteristics that can be used to identify the Gleason grade of a tissue region. (1) Color values identify structures of interest: lumen regions appear white, cytoplasm appears pink, and nuclei appear purple. In addition, while benign, healthy tissue has large and irregular lumen regions, higher grade cancers have small, narrow lumens. (2) Each gland has different structures arranged in a sequential manner: the lumen area are surrounded by epithelial cell cytoplasm, with a ring of nuclei defining the outer boundary of the gland region.

In this paper, we implement an algorithm to automatically segment these gland regions. These segmented regions are used to extract morphological features from the image, which

are used in a classification algorithm to distinguish between images of Gleason grade 3, Gleason grade 4, and benign epithelium. Our segmentation algorithm is evaluated by comparing the classification accuracies obtained using features from the automatic segmentation scheme and from manually extracted boundaries. The segmentation of gland structures is only a means of obtaining features automatically with the goal of image classification, so evaluation of the algorithm should be in terms of classification accuracy as opposed to other metrics for segmentation evaluation such as accuracy or overlap measures.

In Section 2 we provide an overview of the system. In Section 3 we describe the gland detection algorithm, and in Section 4 we explain our high-level model of segmentation. In Section 5, grade classification is described and the results are shown in Section 6. In Section 7, we list our concluding remarks.

II. SYSTEM OVERVIEW

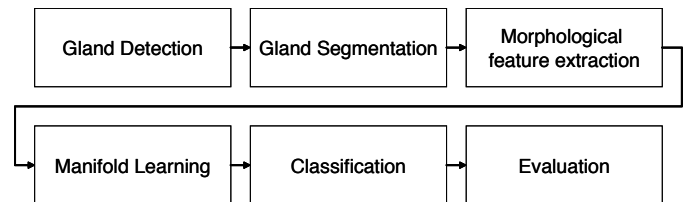


Fig. 3. Overview of our methodology. (1) Potential lumen objects are detected using pixel color values. (2) Gland regions are detected and segmented using a probabilistic framework. (3) Morphological features are calculated from the segmented areas. (4) Manifold learning is performed to obtain a lower dimensional embedding space. (5) Classification is performed using a support vector machine. (6) Evaluation is performed by comparing classification accuracies between the automated and manual schemes.

Hematoxylin and eosin stained prostate tissue samples are digitally imaged at 40x optical magnification by pathologists at the Department of Surgical Pathology, University of Pennsylvania, using a high-resolution whole slide scanner. The images are manually graded by an expert pathologist as belonging to Gleason grade 3, grade 4, or benign epithelium. Our data set consists of a total of 16 images of Gleason grade 3, 11 images of Gleason grade 4, and 17 images of benign epithelial tissue. We denote a tissue region by a digital image $\mathcal{C} = (C, f)$ where C is a 2D grid of image pixels $c \in C$ and f is a function that

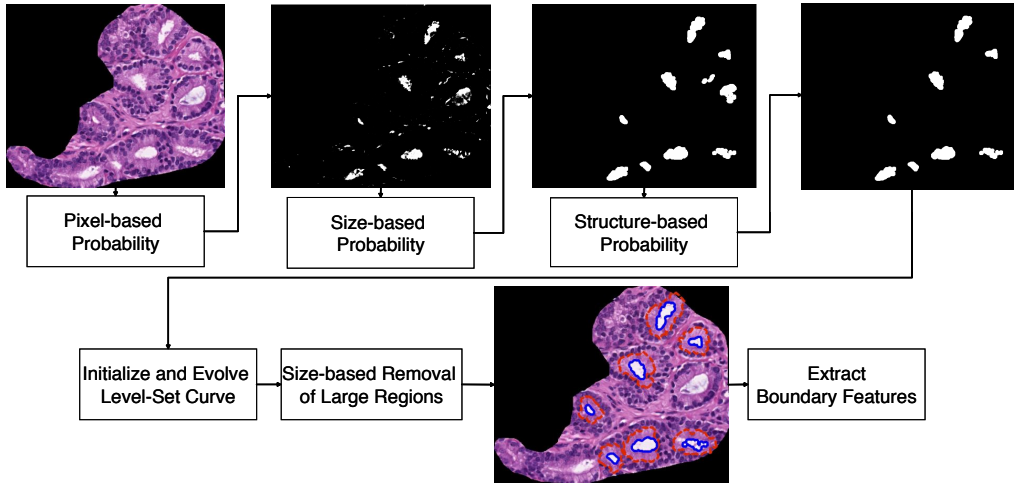


Fig. 4. Flowchart outlining our automated gland detection and segmentation scheme (top half of the flowchart in Figure 3). (1) Color values are used to detect candidate lumen objects. The pixel-based probability that an object is a lumen region is calculated. (2) Area distributions give a size-based probability to individual areas. (3) Structure-based probabilities are calculated using domain knowledge. (4) Lumen objects are used to initialize a level-set curve, which evolves to the detected nuclei. (5) A final size-based check eliminates remaining non-gland regions. (6) Final boundaries are used to calculate morphological feature values.

assigns a pixel value (representing the red, blue, and green channels of the RGB space as well as the hue, saturation, and intensity channels of the HSV space) to c .

An overview of our system is shown in Figure 3. A detailed flowchart of the automated detection and segmentation algorithm is shown in Figure 4. The automated detection and segmentation scheme comprises the following steps:

- 1) Detection of potential lumen areas is done using a Bayesian classifier using low-level pixel values and size constraints to eliminate detected regions too small or too large to be a true lumen region (i.e. noise resulting from insufficient staining, or large gaps in the tissue caused by the mounting process). Structure constraints are applied by examining the region surrounding the candidate lumen regions to identify the presence of epithelial cell cytoplasm. If no cytoplasm is found, then the candidate region is eliminated. (Section III)
- 2) A level-set curve [11] is automatically initialized at the detected lumen boundary and are evolved until they reach a region with a high likelihood of belonging to nuclei. A final size constraint is applied to the resulting level-set curve to eliminate regions that are too large to be a gland. (Section IV)
- 3) The resulting evolved curves and the boundaries of the candidate lumen regions are used to calculate a set of morphological features from the gland regions. (Section V)
- 4) Manifold learning [12] is employed to reduce the dimensionality of the extracted feature set to calculate object adjacencies in the high-dimensional feature space.
- 5) A support vector machine algorithm is then used to classify each image as belonging to benign epithelium, Gleason grade 3 tissue, or Gleason grade 4 tissue.
- 6) To evaluate the efficacy of our segmentation algorithm, we examine the differences in classification accuracy

between the automatically generated feature set and a second set of morphological features that are calculated from a manually-initialized level-set boundary.

III. DETECTION OF GLANDS

A. Detection of tissue structures via pixel-wise classification

A gland comprises (Fig. 2) three main structures: lumen, cytoplasm, and nuclei. The structures are arranged in a specific fashion (lumen is surrounded by cytoplasm, which is surrounded by a ring of nuclei). Therefore, our first step in detecting glands in an image is to detect these structures. To detect the pixels that correspond to lumen, cytoplasm, and nuclei in the images, we employ a Bayesian classifier trained on the image pixel values. A training set \mathcal{T} of pixels representing each of the three classes is manually selected from the data to act as the ground truth. We use approximately 600 manually denoted pixels from each class for training. The color values $f(c)$ of pixels $c \in \mathcal{T}$ are used to generate probability density functions $p(c, f(c)|\omega_v)$, where ω_v represents the pixel class with $v \in \{L, N, S\}$ indicating the lumen, nucleus, and cytoplasm tissue classes, respectively. For each image $\mathcal{C} = (C, f)$, Bayes Theorem is used to obtain a pixel-wise likelihood for each pixel $c \in C$, where $P(\omega_v|c, f(c))$ is the probability that c belongs to class ω_v given image pixel value $f(c)$. Using Bayes Theorem [13], the posterior conditional probability that c belongs to ω_v is given as,

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

where $p(c, f(c)|\omega_v)$ is the *a priori* conditional probability obtained during training via the probability density function, and $P(\omega_v)$ is the prior probabilities of occurrence for each class (assumed as non-informative priors). These pixel-wise likelihoods generate likelihood scenes (Fig. 5), where the intensity in the likelihood image is the probability of pixel

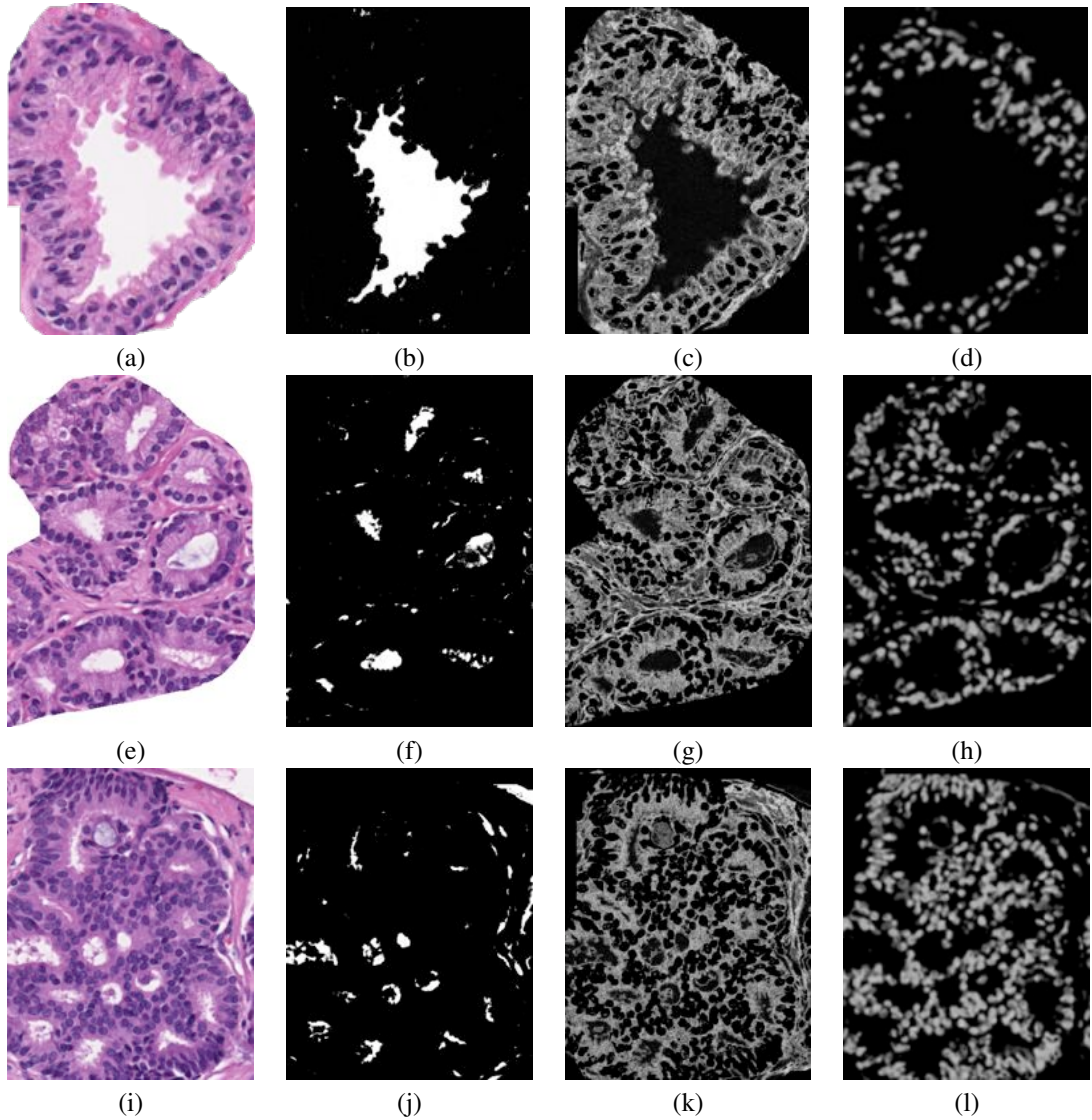


Fig. 5. Examples of likelihood scenes for images of benign epithelium (a)-(d), Gleason grade 3 (e)-(h), and Gleason grade 4 (i)-(l) corresponding to the lumen area, ((b), (f), (j)), the epithelial cytoplasm, ((c), (g), (k)), and the epithelial nuclei ((d), (h), (l)). Higher intensity values correspond to higher likelihood of the pixel belonging to the structure of interest.

c belonging to class ω_v . Shown in Figure 5 are examples of the likelihood scenes for an image of benign epithelium (Fig. 5 (a)), grade 3 (Fig. 5 (e)), and grade 4 (Fig. 5 (i)), corresponding to lumen likelihood (Fig. 5 (b),(f),(j)), cytoplasm likelihood (Fig. 5 (c),(g),(k)), and nuclear likelihood (Fig. 5 (d),(h),(l)). An empirically determined threshold is applied to the likelihood images to obtain the pixels that belong to the class.

B. Identifying candidate gland lumen objects

We define a set of pixels as an object O if for all $c \in O$, $P(\omega_L|c, f(c)) > \tau$ where τ is an empirically determined threshold. This thresholding of the lumen likelihood scene is done to remove background noise. Additionally, for any two pixels $b, c \in O$ a pre-determined neighborhood criterion must be satisfied. We are interested in the probability that an object O belongs to the class of true lumen objects, ω_L . We calculate the pixel-based probability of an object O

being a true lumen based on its pixel likelihood values as $P_B = \frac{1}{|O|} \sum_{c \in O} P(\omega_L|c, f(c))$.

C. Incorporating gland size constraint

After identifying candidate lumen areas using low-level pixel values, we utilize *a priori* knowledge of gland sizes to identify and remove potential noise. During training, the areas of the lumen and interior gland regions are manually calculated and size histograms are generated for each of the three classes. The presumed underlying Gaussian distribution in the histogram is estimated by curve fitting. Our experiments show that the lumen areas are smaller than the gland areas, and that the Gleason grade 4 distributions show areas that are smaller and more uniform compared with benign epithelium, which has a wide range of large gland and lumen areas. This tendency is reflected in the Gleason paradigm [1], where smaller, uniform glands are indicative of higher grade.

The lumen area distribution constitutes the size-based probability density function for object O in each image. The conditional probability of an object O belonging to the lumen class based on its area is given as $\mathbf{P}_S = P(O \hookrightarrow \omega_L | \mathbf{A})$, where \mathbf{A} is the area of object O .

D. Incorporating structure-based probability

Due to the structure of glands in the tissue, we can eliminate non-gland regions by ensuring that the structure of the gland is as shown in Figure 2; that is, a lumen region is immediately surrounded by epithelial cytoplasm, and that the cytoplasm is bordered by a ring of nuclei. If these conditions are met, then the objects are considered true gland regions, otherwise they are removed. The likelihood scene generated for cytoplasm is used to check for the presence of the epithelial cytoplasm surrounding the detected lumen regions. The neighborhood of O is obtained by performing morphological dilation on O to obtain a dilated object \hat{O} . The neighborhood η of O is defined as $\hat{O} - O$. For every pixel $c \in \eta$, $P(c \in \eta | \omega_S, f(c))$ is calculated. The average likelihood, \mathbf{P}_G , of O being a true gland is $\mathbf{P}_G = \frac{1}{|\eta|} \sum P(c \in \eta | \omega_S, f(c))$.

We assume that each of the above object-wise probabilities are independent, so that the joint probability that the object O identified as lumen actually belongs to a gland is given as the product of the independent probabilities \mathbf{P}_B , \mathbf{P}_S , and \mathbf{P}_G :

$$P(O \hookrightarrow \omega_L | \mathbf{P}_B, \mathbf{P}_S, \mathbf{P}_G) = \prod_{\alpha \in \{B, S, G\}} \mathbf{P}_\alpha, \quad (2)$$

where $\hookrightarrow \omega_L$ denotes membership in the lumen class ω_L .

IV. HIGH-LEVEL MODEL: SEGMENTATION

Once the possible gland lumen are found, boundary segmentation is performed using level-sets. A boundary \mathcal{B} evolving in time t and in the 2D space defined by the grid of pixels C is represented by the zero level set $\mathcal{B} = \{(x, y) | \phi(t, x, y) = 0\}$ of a level set function ϕ , where x and y are 2D Cartesian coordinates of $c \in C$. The evolution of ϕ is then described by a level-set formulation adopted from [11]:

$$\frac{\partial \phi}{\partial t} + F |\nabla \phi| = 0 \quad (3)$$

where the function F defines the speed of the evolution. The curve evolution is driven by the nuclei likelihood image. The initial contour $\phi_0 = \phi(0, x, y)$ is initialized automatically using the detected lumen area from the candidate gland regions. The curve is evolved outward from the detected lumen regions in the combined nuclei likelihood image to avoid noise and allow smoother evolution relative to the original image. The intensities of the nuclei likelihood image forms the stopping gradient. The algorithm is run until the difference in the contours of one iteration to the next is below an empirically determined threshold. During training, size distributions similar to those used to calculate object likelihood \mathbf{P}_S are created using the final contours. These nuclear boundary based distributions are used to remove regions that are too large to be true glands. Finally, the lumen and nuclear boundaries extracted from true gland regions are passed on to the next step for feature extraction.

V. GLEASON GRADE CLASSIFICATION

The boundaries of the detected lumen areas and the segmented interior nuclei boundary through level-set algorithm are used to extract a set of features that quantify the morphology of the gland. We define a boundary, \mathcal{B} , as a set of pixels lying on the edge of a contiguous region of an image.

A. Feature Extraction

- 1) *Area overlap ratio*: The area enclosed by \mathcal{B} divided by the area of the smallest circle enclosing \mathcal{B} .
- 2) *Distance ratio*: Ratio of average to maximum distance from the centroid of \mathcal{B} to the points lying on \mathcal{B} .
- 3) *Standard deviation*: The distances from the centroid of \mathcal{B} to all the points lying on \mathcal{B} .
- 4) *Variance*: The distances from the centroid of \mathcal{B} to all the points lying on \mathcal{B} . Standard deviation and variance are normalized by dividing all the distances by the maximum distance for individual \mathcal{B} .
- 5) *Perimeter ratio*: The ratio of estimated length of \mathcal{B} to the true length of \mathcal{B} . Estimated length is computed using linear interpolation between 5-10 points (depending on the number of points on \mathcal{B}) sampled at equal intervals from \mathcal{B} , while true length is computed using all the points lying on \mathcal{B} .
- 6) *Compactness*: The true length of \mathcal{B} squared divided by the area enclosed by \mathcal{B} [14].
- 7) *Smoothness*: For points c_{i-1}, c_i , and c_{i+1} on \mathcal{B} , where point c_{i-1} is immediately adjacent and counter-clockwise to point c_i and point c_{i+1} is immediately adjacent and clockwise from point c_i , $S_{c_i} = |d(c_i, c_g)(d(c_{i-1}, c_g) + d(c_{i+1}, c_g))|/2|$, where c_g is the centroid of the area enclosed by the boundary \mathcal{B} and $d(c_g, c_i)$ is the Euclidean distance between c_g and c_i . Smoothness is then defined as $\sum_i S_{c_i \in \mathcal{B}}$.
- 8) *Area*: enclosed within \mathcal{B} .

All of the above features are calculated for the lumen boundary as well as the interior nuclei boundary. For an image C , these 16 features constitute the feature vector \mathbf{F} .

B. Manifold Learning

Manifold learning is a method of reducing a data set from \mathbf{M} to \mathbf{N} dimensions, where $\mathbf{N} < \mathbf{M}$ while preserving inter- and intra-class relationships between the data. This is done to project the data into a low-dimensional feature space in such a way to preserve high dimensional object adjacency. Many manifold learning algorithms have been constructed over the years to deal with different types of data. In this work we employ the Graph Embedding algorithm using normalized cuts [15]. Previously, Graph Embedding has been used to improve classification between cancer and non-cancer regions of an image [12]. Graph Embedding constructs a confusion matrix \mathcal{Y} describing the similarity between any two images \mathcal{C}_p and \mathcal{C}_q with feature vectors \mathbf{F}_p and \mathbf{F}_q , respectively, where $p, q \in \{1, 2, \dots, k\}$ and k is the total number of images in the data set

$$\mathcal{Y}(p, q) = e^{-\|\mathbf{F}_p - \mathbf{F}_q\|} \in \mathbf{R}^{k \times k}. \quad (4)$$

The embedding vector \mathcal{X} is obtained from the maximization of the function:

$$\mathcal{E}_{\mathcal{Y}}(\mathcal{X}) = 2\gamma \frac{\mathcal{X}^T(D - \mathcal{Y})\mathcal{X}}{\mathcal{X}^T D \mathcal{X}}, \quad (5)$$

where $D(p, p) = \sum_q \mathcal{Y}(p, q)$ and $\gamma = |I| - 1$. The \mathbf{N} -dimensional embedding space is defined by the eigenvectors corresponding to the smallest \mathbf{N} eigenvalues of $(D - \mathcal{Y})\mathcal{X} = \lambda D \mathcal{X}$. The value of \mathbf{N} was optimized by obtaining classification accuracies for $\mathbf{N} \in \{1, 2, \dots, 10\}$ and selecting the \mathbf{N} that provided the highest accuracy for each classification task. For image \mathcal{C} , the feature vector \mathbf{F} given as input to the Graph Embedding algorithm produces an \mathbf{N} -dimensional eigenvector $\hat{\mathbf{F}} = [\hat{e}_j(a) | j \in \{1, 2, \dots, \mathbf{N}\}]$, where $\hat{e}_1(a)$ is the principal eigenvalue associated with a .

C. Classification

Following Graph Embedding, we classify the images using a support vector machine (SVM) algorithm into (a) benign epithelial tissue, (b) Gleason grade 3 tissue, and (c) Gleason grade 4 tissue. SVMs project a set of training data \mathbf{E} representing two different classes into a high-dimensional space by means of a kernel function \mathbf{K} . The algorithm then generates a discriminating hyperplane to separate out the two classes in such a way to maximize a cost function. Testing data is then projected into the high-dimensional space via \mathbf{K} , and the test data is classified based on where it falls with respect to the hyperplane. In this study, the training and testing data is drawn from the N low-dimensional eigenvectors resulting from the Graph Embedding algorithm. The kernel function $\mathbf{K}(\cdot, \cdot)$ defines the method in which data is projected into the high-dimensional space. In this paper we use a commonly used kernel known as the radial basis function:

$$\mathbf{K}(\hat{\mathbf{F}}, \hat{\mathbf{E}}) = e^{(-\delta \|\hat{\mathbf{F}} - \hat{\mathbf{E}}\|^2)}, \quad (6)$$

where δ is a scaling factor to normalize the inputs $\hat{\mathbf{F}}$ and $\hat{\mathbf{E}}$.

There are three classification tasks for SVM classification: benign epithelium vs. Gleason grade 3; benign epithelium vs. Gleason grade 4, and Gleason grade 3 vs. Gleason grade 4. For each task, a third of each class is selected at random to train the classifier. Classification is then performed for all samples in the task using both the reduced and unreduced feature space. This process is repeated for a total of 10 trials to provide adequate cross-validation.

VI. RESULTS AND DISCUSSION

Our evaluation of the segmentation performance is based on comparing the classification accuracies from feature values obtained using: (1) the automatically initialized level-set boundaries, and (2) the manually initialized level-set boundaries. Since the segmentation of gland regions is only required to enable classification, we need to illustrate that automatically initializing the level-set curve performs approximately as well (for better or worse) as a level-set curve that has been manually initialized. Even though the automated and manual schemes may arrive at different segmentations, the classification accuracy may not change significantly. Since the goal of segmentation is to generate an accurate classification, this is the means by which the two methods (automated and

manual) are compared. We also motivate the use of manifold learning by showing classification accuracy from both the reduced and unreduced feature spaces.

Table I lists the SVM classification accuracy of each of the classification tasks. The average accuracy is shown over 10 trials using randomized cross-validation. Standard deviations are given in parentheses. Listed are each of the pairwise class comparisons: Gleason grade 3 (G3) vs. benign epithelium (BE), Gleason grade 4 (G4) vs. benign epithelium, and Gleason grade 3 vs. grade 4. Shown are the results of classification using the manually- and automatically-extracted boundary regions. In two out of three cases (G3 vs. BE and G4 vs. BE), the accuracy is higher using manual features, but in one case (G3 vs. G4) the accuracy is higher using automatically generated features. The similar performance with the manual and automated segmentation schemes implies that the automated scheme performs comparably with the manual scheme. Table I lists the accuracies and standard deviations using the unreduced feature space and the reduced feature spaced obtained using Graph Embedding. The results here indicate that in almost all classification tasks, feature reduction through manifold learning improves the classification accuracy over the unreduced feature space.

Feature space	Task	Automated	Manual
Reduced	G3 vs. G4	95.19% (0.025)	80.76% (0.016)
	G3 vs. BE	86.35% (0.016)	95.14% (0.016)
	G4 vs. BE	92.90% (0.000)	95.14% (0.017)
Unreduced	G3 vs. G4	84.46% (0.034)	77.06% (0.062)
	G3 vs. BE	87.28% (0.031)	92.11% (0.026)
	G4 vs. BE	86.43% (0.047)	91.82% (0.017)

TABLE I

AVERAGES AND STANDARD DEVIATIONS OF SVM CLASSIFICATION ACCURACY FOR EACH OF THE THREE CLASSIFICATION TASKS USING THE AUTOMATICALLY AND MANUALLY EXTRACTED FEATURE SETS, AS WELL AS THE UNREDUCED (ORIGINAL) AND REDUCED FEATURE SPACES.

Sample results from the automated segmentation algorithm are shown in Figure 6. The lumen boundaries are displayed in a solid blue contour and the interior nuclear boundaries are displayed as dashed black lines. Results are shown for sample images from the benign epithelium (Fig. 6 (a), (d)), Gleason grade 3 (Fig. 6 (b), (e)), and Gleason grade 4 (Fig. 6 (c), (e)) classes. In benign epithelial images (Figs. 6 (a) and (d)), almost perfect segmentation of the large benign glands is shown. In Gleason grade 3 images (Figs. 6 (b) and (e)) almost all gland structures are detected and segmented properly. In some cases the boundary does not evolve completely to the interior nuclear boundary, while in others the boundary passes between nuclei that are spread too far apart. In Gleason grade 4 images (Figs. 6 (c) and (e)), a number of gland-like regions are not detected due to the size or structure constraints, while others have lumen regions that are too small or occluded to be detected. These issues illustrate the difficulty of gland detection and segmentation, especially at high Gleason grades where glands can have lumen areas that are almost completely degraded [1].

A comparison between manual and automated segmentation results is given in Figure 7 for an image of Gleason grade 3

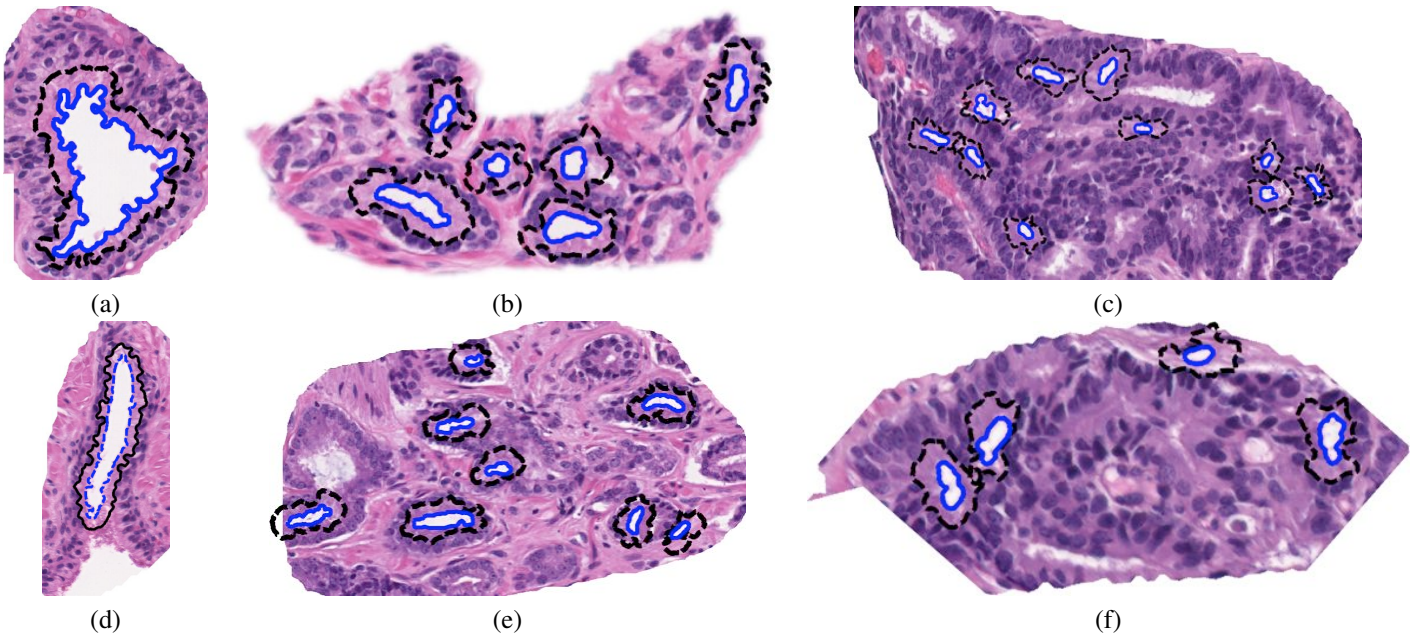


Fig. 6. Results of the automatic segmentation algorithm. The blue contours correspond to the boundaries of the lumen regions, while the black contours are the inner boundaries of the nuclei of the epithelial cells surrounding the gland. Shown are examples of tissues from benign epithelium ((a), (d)), Gleason grade 3 ((b), (e)), and Gleason grade 4 ((c), (f)).

tissue. The segmented regions obtained manually (Fig. 7 (a)) are qualitatively different from those obtained automatically (Fig. 7 (b)): certain regions are identified by one method and not the other, while some non-gland regions are identified as glands by the automated algorithm. However, the classification accuracy is still comparable between the two methods, indicating that these qualitative differences do not greatly affect the ability of the classifier to distinguish between Gleason grades.

Also shown in Figure 7 are the low-dimensional representations of the data obtained from the Graph Embedding algorithm. The points on the plot correspond to benign epithelium (red triangles), Gleason grade 3 (green circles), or Gleason grade 4 images (blue squares). Black contours denote the class clusters. Shown are the results of the automated extraction algorithm (Fig. 7 (c)) and the manual algorithm (Fig. 7 (d)). The clear grouping of different classes indicates that the Graph Embedding algorithm can successfully preserve object adjacencies in the high-dimensional features space and can map those adjacencies effectively in the low-dimensional space. Additionally, the low-dimensional manifolds from both the automated and manual schemes reflect a similar manifold structure, suggesting that the two feature extraction schemes describe the differences between the tissue types in a similar manner. Finally, the manifold structure displays a smooth transition from benign tissue on the right of the graph to high-grade cancer on the left, indicating that in terms of our extracted features the Gleason grading scheme describes a gradient of different tissue characteristics progressing from low- to high-grade cancer. With a larger data set, the manifolds shown in Fig. 7 will be more densely populated, and the low-dimensional structure of the data can be fully appreciated.

VII. CONCLUSIONS

In this work, we have demonstrated a fully automated gland detection, segmentation, and feature extraction algorithm. Our results suggest that gland morphology plays an important role in discriminating different Gleason grades of prostate cancer. Our features are capable of discriminating between intermediate Gleason grades, and our results suggest that classification accuracy obtained using our automated segmentation scheme is comparable to the accuracy of a manual extraction scheme. The main contributions of this work are:

- a fully automated gland detection, segmentation, and feature extraction algorithm;
- a novel method of segmentation evaluation by comparing classification accuracy between automatic and manually extracted segmentations;
- the use of low-, high-, and domain-level knowledge to accurately detect and segment gland regions;
- a novel set of morphological features that accurately distinguishes between intermediate Gleason grades; and
- comparable classification performance between automated and manually generated feature sets.

Our automated scheme performs comparably to the manual scheme, indicating that the segmentations obtained through this algorithm can reliably classify between different prostate tissue images at least as well as a manually denoted segmentation. This algorithm can be used to extract discriminating morphological features from large image studies, making the creation of large-scale image databases an attainable goal. In future work we aim to implement other methods of gland segmentation such as active shape models to improve the agreement between the manual and the automated schemes and to identify glands that are currently missed due to lack of sufficient lumen area or coherent structure. We are also

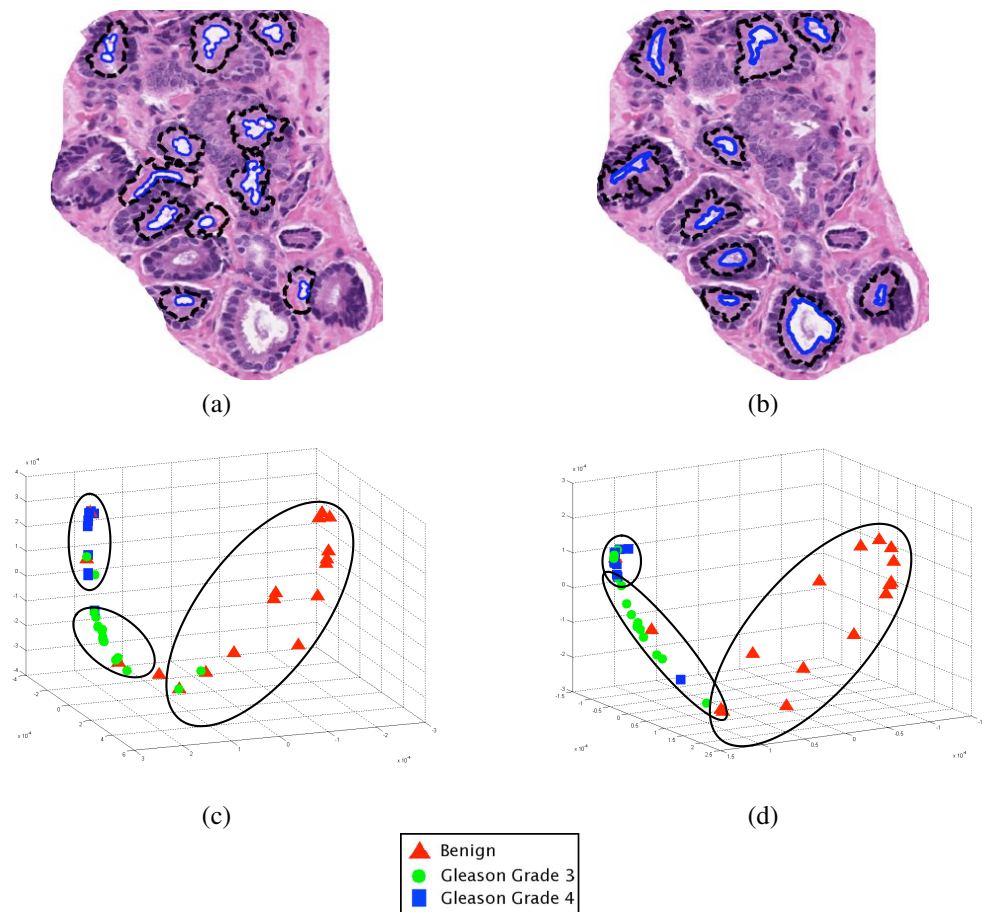


Fig. 7. Examples of (a) the results of the automated segmentation of a tissue, and (b) the same region segmented using manually initialized contours. Also shown are the results of Graph Embedding applied to the feature set obtained from (c) the automatically generated features and (d) the features obtained using a manually initialized level-set contour. Points on the graphs correspond to images of benign epithelium (red triangles), Gleason grade 3 (green circles), and Gleason grade 4 tissue (blue squares). The separation between class clusters (denoted by black ellipses) indicates that Graph Embedding can project the data to accurately show object adjacencies.

integrating the automatically detected morphological features with texture-based and architectural image features, which have also been shown [4] to discriminate between Gleason grades of prostate tissue.

VIII. ACKNOWLEDGMENTS

This work was made possible due to grants from The Coulter Foundation (WHCF4-29349, WHCF 4-29368), Busch Biomedical Award, Cancer Institute of New Jersey, New Jersey Commission on Cancer Research, National Institute of Health, and The National Cancer Institute (R21CA127186-01, R03CA128081-01).

REFERENCES

- [1] D.F. Gleason, "Classification of prostatic carcinomas," *Cancer Chemotherapy Reports*, vol. 50, no. 3, pp. 125–128, 1966.
- [2] J.I. Epstein, W.C. Allsbrook, et al., "The 2005 international society of urological pathology (isup) consensus conference on gleason grading of prostatic carcinoma," *Am. J. of Surg. Path.*, vol. 29, no. 9, pp. 1228–1242, 2005.
- [3] W. Allsbrook, K.A. Mangold, et al., "Interobserver reproducibility of gleason grading of prostatic carcinoma: General pathologist," *Hum. Path.*, vol. 32, no. 1, pp. 81–88, 2001.
- [4] S. Doyle et al., "Automated grading of prostate cancer using architectural and textural image features," *IEEE ISBI*, pp. 1284–1287, 2007.
- [5] R. Farjam et al., "An image analysis approach for automatic malignancy determination of prostate pathological images," *Cytometry Part B: Clinical Cytometry*, 2007.
- [6] S. Naik et al., "A quantitative exploration of efficacy of gland morphology in prostate cancer grading," *IEEE 33rd NEB Conference*, pp. 58–59, 2007.
- [7] V.R. Korde, H. Bartels, et al., "Automatic segmentation of cell nuclei in bladder and skin tissue for karyometric analysis," *Biophotonics 2007: Optics in Life Science*, vol. 6633, pp. 6633V, 2007.
- [8] P. Bamford and B. Lovell, "Unsupervised cell nucleus segmentation with active contours," *Signal Processing*, vol. 71, pp. 203–213, 1998.
- [9] G. Begelman et al., "Cell nuclei segmentation using fuzzy logic engine," *Int. Conf. on Image Proc.*, pp. 2937–2940, 2004.
- [10] M. Gao, P. Bridgman, and S. Kumar, "Computer aided prostate cancer diagnosis using image enhancement and jpeg2000," *Proc. of SPIE*, vol. 5203, pp. 323–334, 2003.
- [11] C. Li, C. Xu, C. Gui, and M.D. Fox, "Level set evolution without re-initialization: a new variational formulation," *IEEE CVPR*, vol. 1, pp. 430–436, 2005.
- [12] A. Madabhushi et al., "Graph embedding to improve supervised classification: Detecting prostate cancer," *Proc. of MICCAI*, vol. 3749, pp. 729–738, 2005.
- [13] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, Wiley-Interscience, second edition, 2001.
- [14] C. Demir and B. Yener, "Automated cancer diagnosis based on histopathological images: a systematic survey," *Technical Report TR-05-09, Computer Science Department at Rensselaer Polytechnic Institute*, 2005.
- [15] J. Shi and J. Malik, "Normalized cuts and image segmentation," *CVPR*, pp. 731–737, 1997.