

Prostate cancer diagnosis using quantitative phase imaging and machine learning

Tan H. Nguyen^{a,b}, Shamira Sridharan^a, Virgilia Marcias^c, Andre K. Balla^c, Minh N. Do^b,
Gabriel Popescu^{*a}

^aQuantitative Phase Imaging Laboratory, Beckman Institute for Advanced Science and Technology,
Dept. of Electrical and Computer Engineering, Univ. of Illinois at Urbana-Champaign, IL USA
61801.

^bComputational Imaging group, Coordinated Science Laboratory, Dept. of Electrical and Computer
Engineering, Univ. of Illinois at Urbana-Champaign, IL USA 61801.

^cDept. of Pathology, Univ. of Chicago, IL USA 60637.

ABSTRACT

We report, for the first time, the use of Quantitative Phase Imaging (QPI) images to perform automatic prostate cancer diagnosis. A machine learning algorithm is implemented to learn textural behaviors of prostate samples imaged under QPI and produce labeled maps of different regions for testing biopsies (e.g. gland, stroma, lumen etc.). From these maps, morphological and textural features are calculated to predict outcomes of the testing samples. Current performance is reported on a dataset of more than 300 cores of various diagnosis results.

Keywords: automatic diagnosis, Quantitative Phase Imaging, texton analysis, prostate cancer

1. INTRODUCTION

Prostate cancer is the second most common type of cancer that causes deaths to men in the US [1]. In 2010, 196,038 men were diagnosed with prostate cancer, and 28,560 men eventually died from it. During prostate cancer diagnosis, if the patient fails under the Digital Rectal Examination and Prostate-Specific Antigen (PSA) screening, a urologist or radiologist will take a cylindrical biopsies of prostate tissue through the rectum of a patient with hollow needles. These samples are stained. Then, its glandular architecture is evaluated under microscopy inspection and a Gleason score is assigned by a pathologist based on the Gleason grading system [2]. This score is a sum of two partial scores corresponding to the most and second most predominant grades in the tissue. Each partial score is a grade ranging from 1 through 5, measuring the degree of malignancy and cancer aggressiveness. Benign tissues usually have grade 1 or 2. These grades are characterized by large, well separated glands, large and branchy lumen and thick boundaries. Grade 3 has smaller glands, thinner boundary, smaller and round lumen. In grade 4, the glands merge together and create a mass of glands containing multiple lumens. Grade 5 indicates that poorly differentiated tissue and low chance of survival. The Gleason grading is an important factor to prostate diagnosis since it has strong correlation with the extent of the disease and patient survival rate. It also helps the pathologist determine most suitable treatments to the patients.

Several efforts have been spent to produce a reliable Gleason grading based on the histological H&E images. Such efforts can be divided to two categories: classification-based and segmentation-based. Methods in the first group utilize various H&E-based features to produce Gleason scores without needing a label map. These features include textures from H&E images [3], multi-spectra images [4]. Methods in the second group produce a Gleason score in two stages. In the first stage, label maps of different parts of the biopsies are produced from the H&E images. Then, morphological features are extracted from these maps. Then, subsequent classifiers are deployed to produce the final Gleason scores. Naik et al. [5] build statistical models of the likelihood for the class of a pixel given its color and location in the training set. Although the authors reported good separation accuracies of 86.35% between Grade 3 and benign, 95.19% between Grade 3 and Grade 4, 92% between Grade 4 and benign, the information of stroma and lumen is not fully deployed. In [6], Nguyen et al. use the (L, a, b) color space and various constraints on the relative arrangements of tissue regions sections to refine the segmentation map.

One common point between these methods is the use of H&E images in the diagnosis. The need for staining comes from the fact that many biopsies are known to be “phase object” i.e. they are nearly transparent under bright-field inspection. Exogenous factors such as H&E molecular tags must be introduced to enhance the contrast. Unfortunately, this introduction sacrifices intrinsic properties of the specimens. Beside long sample preparation time, variation in imaging condition hinders translation of research findings into medical clinics. In contrast, QPI modalities [7-13] render high contrast of these phase objects without staining and measure the refractive index of imaging sample non-invasively. This refractive index information is intrinsic and have shown high potential for medical diagnosis. In [14], the authors reported, for the first time, the use of QPI to classify cancerous areas vs. benign areas in 11 prostate biopsies. In their research, mean and median of the phase distribution in a sliding windows are used as the feature. Although the number of biopsies is relatively small, this pioneer work has deeply motivated us in building a fully automatic diagnosis system for prostate cancer diagnosis.

Our paper follows the second approach to produce automatic diagnosis results. Here, a high throughput Spatial Light Interference Microscopy (SLIM) [9] system is used to image biopsies with frame rate up to 12.5 frames/s. These biopsies come from a dataset of more than 300 cores from the Department of Pathology at University of Chicago. A Random Forest classifier [15] is trained to segment the SLIM images into different regions (e.g. glands, stroma and lumen) with accuracy of at least 90%, based on the texture of the tissue. Given all features computed from segmented glands, a Bag-of-Word (BoW) model [16] is used to extract the feature of each cores from all the glands inside the core. Using the core features, we show a binary classification problem for Gleason grade 3 vs. Gleason grade 4 cores using the Support Vector Machine (SVM) classifiers.

2. METHODS

2.1 Principle of SLIM imaging

Assuming a phase object with sample transmission $T(\mathbf{r})$ is imaged using QPI where \mathbf{r} denotes the transverse coordinate in the sample plane. The phase quantity of interest in QPI, $\phi(\mathbf{r}) = \arg[T(\mathbf{r})]$, is given by $\phi(\mathbf{r}) = \langle k \rangle h(\mathbf{r}) \Delta n(\mathbf{r})$, where $\Delta n(\mathbf{r}), h(\mathbf{r})$ is the difference in refractive index and the thickness of the sample. $\langle k \rangle$ is the mean wavenumber of the illumination. For coherent illumination, assuming the incident field has unit amplitude, the total field at the output port of the microscope relates to the sample’s transmission as $U_t(\mathbf{r}) = T(\mathbf{r})$. To measure $\phi(\mathbf{r})$, SLIM decomposes U_t into a DC component U_o and an AC component U_s such that $U_t(\mathbf{r}) = U_o + U_s(\mathbf{r}) = U_o \{1 + \beta(\mathbf{r}) \exp[i\Delta\phi(\mathbf{r})]\}$ where $U_o = \langle U_t(\mathbf{r}) \rangle_r$, obtained by spatial filtering the total field U_t . The quantities $\beta(\mathbf{r}) = |U_s(\mathbf{r})|/|U_o|$, $\Delta\phi(\mathbf{r}) = \arg[U_s(\mathbf{r})] - \arg U_o$, denote relative amplitude and phase difference between the two fields. If both quantities are known, the phase of interest can be computed as

$$\phi(\mathbf{r}) = \tan^{-1} \left(\frac{\beta(\mathbf{r}) \sin[\Delta\phi(\mathbf{r})]}{1 + \beta(\mathbf{r}) \cos[\Delta\phi(\mathbf{r})]} \right). \quad (1)$$

Figure 1 shows the SLIM optical setup. Here, the system is designed as an add-on module to a commercial microscope. First, the total field U_t is relayed to the output port of the scope. Then, it is spatially decomposed at the back focal plane of L_1 into U_o and U_s . The DC field, U_o , has the Fourier spectrum matches the condenser phase annulus while that of AC field occupies the rest of the focal plane. An SLM, placed at this plane, generates modulating phase rings (see Figure 1(b)) matching the spatial spectrum U_o and shifting its wavefront by $\varphi_n = n\pi/2$ with $n = 0, 1, 2, 3$. A camera placed at the back focal plane of lens L_2 captures the interference intensity between U_s and the phase-shifted U_o , namely:

$$I(\mathbf{r}; \varphi_n) = |e^{i\varphi_n} U_o + U_s(\mathbf{r})|^2 = |U_o|^2 \left[1 + |\beta(\mathbf{r})|^2 + 2\beta(\mathbf{r}) \cos(\Delta\phi - \varphi_n) \right], \quad (2)$$

where φ_n is the external phase modulation.

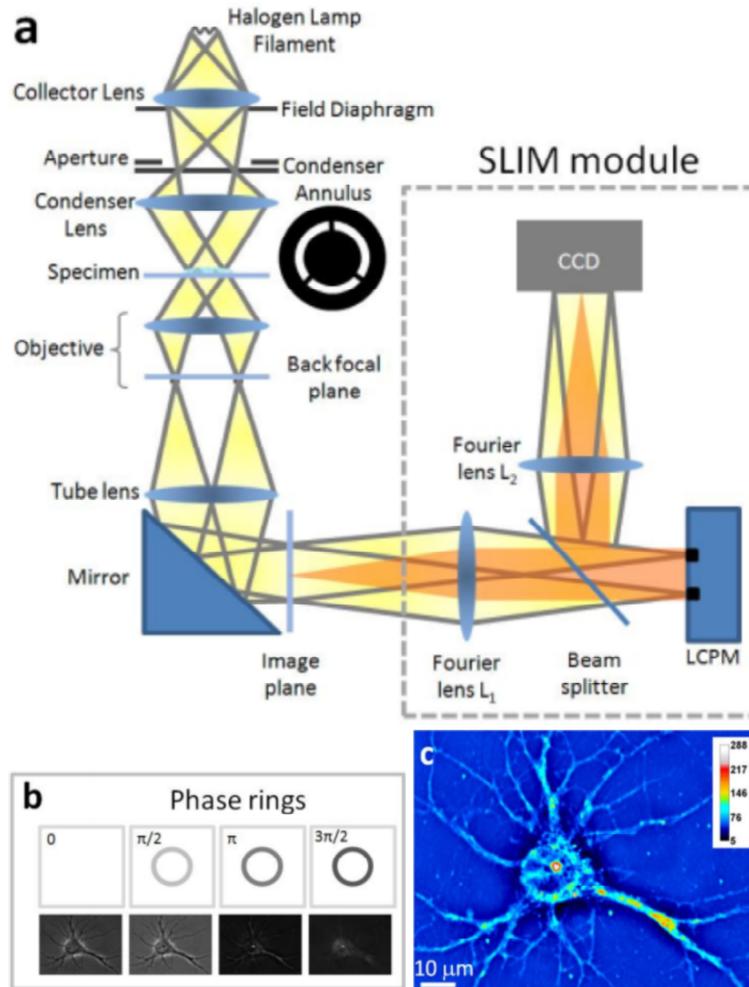


Figure 1. a) Optical setup of SLIM [8], b) Four different intensity images corresponding to 4 different modulation of the phase mask, c) Synthesize phase map

During the processing stage, four different images $I(\mathbf{r};0)$, $I(\mathbf{r};\pm\pi/2)$, $I(\mathbf{r};\pi)$ are combined to solve for $\beta(\mathbf{r}), \Delta\phi(\mathbf{r})$ using Eq. (2) and further plugged into Eq. (1) to compute $\phi(\mathbf{r})$. More details can be found in [9]. Figure 2 shows an example of H&E and SLIM images of the same core. Most of the details of the samples are observables in both images. However, the contrast of stromal area in the SLIM image are significantly better compared to that of the H&E.

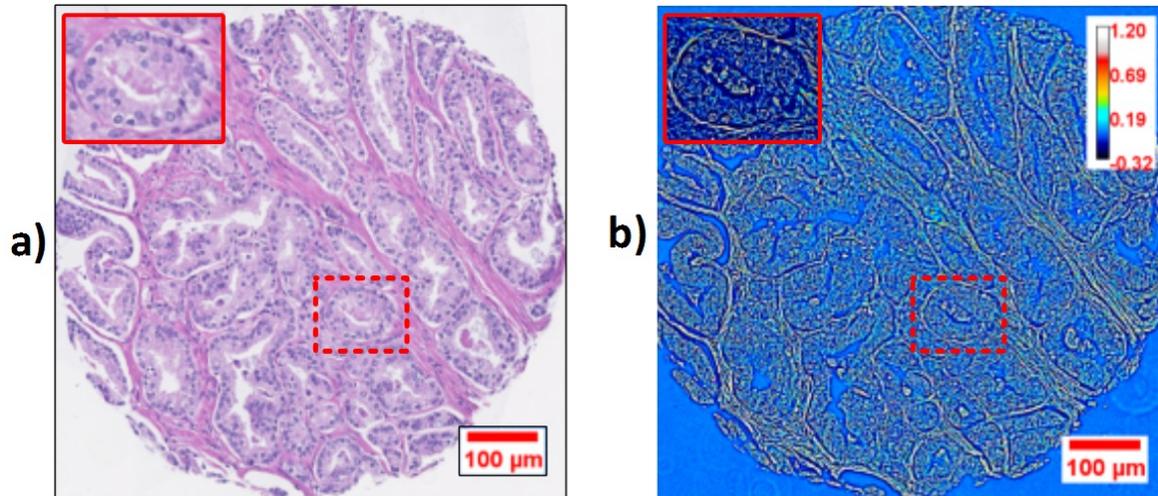


Figure 2. a) H&E image of a core in our database, b) Corresponding SLIM image

2.2 Segmentation of SLIM images

In this section, we show how to obtain the label map from the SLIM image using machine learn. The segmentation maps are produced by classifying each pixel in the phase image with a Random Forest classifier given feature vector of each pixel. Here, we use the texton-based classification framework that has been widely used in computer vision for material classification problem e.g. [17]. The framework consists of two steps. First, the textons are extracted. Second, feature vector of each pixel is calculated given extracted textons. Figure 3 (a) shows how our textons are retrieved from a set of input SLIM images. Each input SLIM images is convolved with a filter bank, here, a modified version of the Leung-Malik filter bank proposed in [18]. The bank consists of 80 directional filters and 10 symmetric filters oriented at 8 different directions and 5 scales. Each filter has dimensions of 49 x 49 pixels. For each scale and direction, we generate one first-order and one second-order derivative of Gaussian kernels with elongation factor of 3. Ten symmetric filters include 5 Gaussian at 5 Laplacian of Gaussian (LOG) kernels at 5 different scales. To capture rotational invariant features, for each scale, we keep the maximum responses over 8 directions. Filter responses obtained following this method are also known as MR-8 [17]. To train the Random Forest classifier, about 4 million filter responses or 0.18% of the total number of filter responses, are randomly sampled over 240 cores. Then, these sampled filter responses are clustered into $k=50$ clusters, whose centers are chosen as textons (see Figure 3 (a)). Found texton are used to compute the feature vector of each pixel in the next step.

Once the texton set have been specified, a vector quantization step is applied to associate each pixel in a SLIM image to a “closest” texton retrieved. Here, “closest” means smallest Euclidean distance. The output of this step is an indexing map specifying the best explained texton for each pixel (see Figure 3 (b)). Then, a Gaussian weighted histogram of texton indices is computed and used as the feature vector for each pixel.

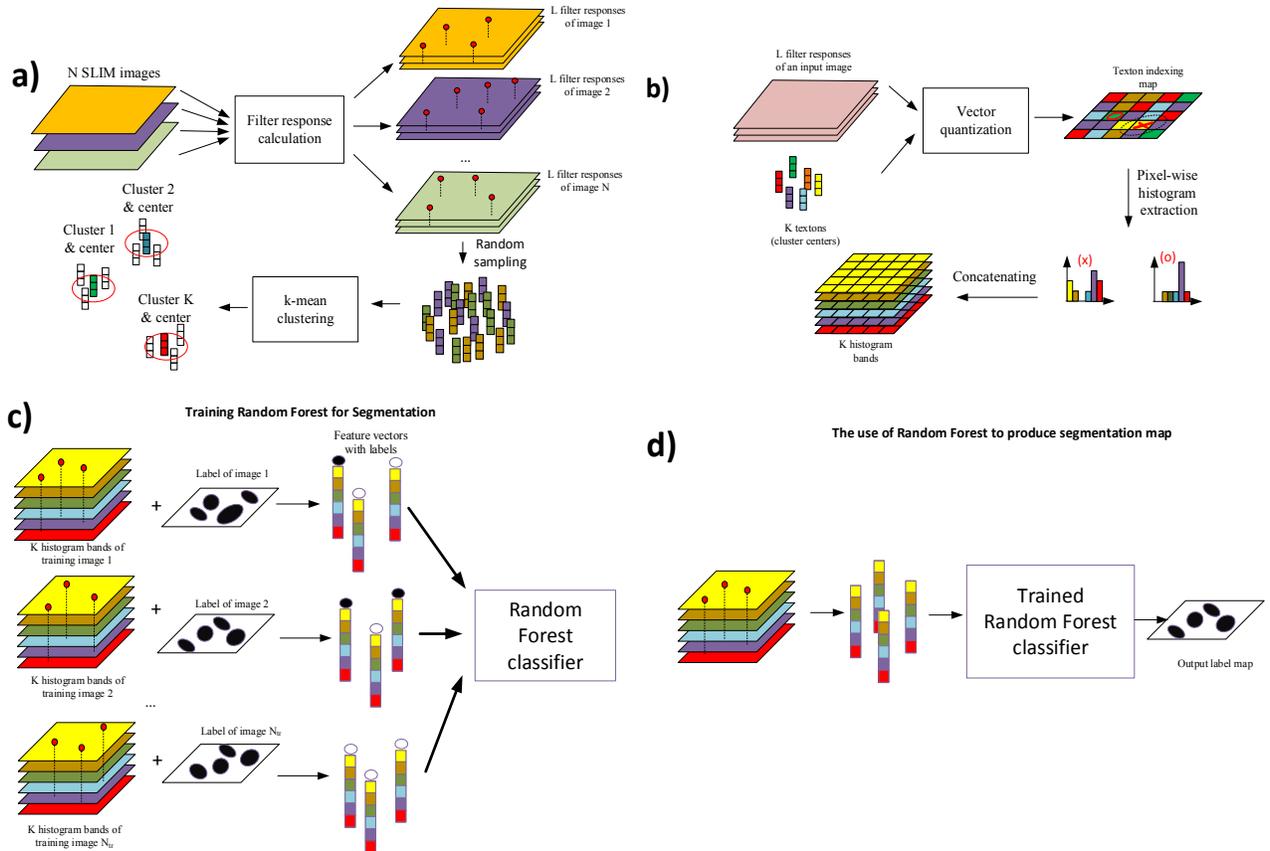


Figure 3. (a) Texton extraction from prostatic SLIM images. (b) Feature extraction: a normalized histogram of texton indices in the neighborhood of each pixel is used as its feature vector. (c) Training the Random Forest for pixel classification: pixel feature vectors with class labels randomly sampled at several locations are used to train the Random Forest classifier. (d) Label map prediction: The feature vector each pixel in the image is fed into the classifier to produce a value from 0 to 1 of pixel stroma likelihood. The label map is obtained by thresholding the stroma likelihood map.

Using normalized histogram of texton indices as feature vector, we train a Random Forest classifier [15] to discriminate two different kinds of textures: epithelial gland and connecting stroma. Lumen pixels are temporary excluded from the diagnosis since their phase values are very close to that of the background. These pixels will be semantically refine after classifying the others. The Random Forest classifier is a modification of bagging (or bootstrap aggregating) method that reduces the prediction variance of a single decision tree by creating an ensemble of T *de-correlated* decision trees, trained on a bootstrapped sample of the original training dataset using random sampling with replacement and “feature bagging”. The classifier has shown its priority over other methods in several classification problems e.g. [19]. The training of this classifier is explained in Fig. 3 (c). After classifier training, a query pixel can be classified using its feature vector by traversing from the root node to the leaf node in each tree. At each node before termination, a branching decision is made based on the current value of the feature vector. Each leaf node in the random forest is associated to a class probability obtained through training, telling percentage of training samples for each class reaching the node. In prediction (see Fig. 3 (d)), the forest produces class probability $p(c|L)$ by averaging the prediction over the leaf nodes $L = (l_1, \dots, l_T)$ as:

$$p(c|L) = \frac{1}{T} \sum_{t=1}^T p(c|l_t). \quad (3)$$

2.3 Automatic diagnosis of SLIM images

After the segmentation map for each core has been computed, all glands are identified. To obtain diagnosis result for the whole core, we compute six different features for each gland inside the core. These features consist of ratio between the area of gland to the maximum glandular area in the core, gland eccentricity, gland solidity, stroma invasion score, ratio between perimeter of the gland to that of the circumscribing ellipse and that of the average stroma likelihood inside the gland. The first four features are morphological and have demonstrated good classification performance in other automatic segmentation with H&E images, see [20]. Meanwhile, our experimental results show strong correlation between the average stromal likelihood feature with the consensus Gleason grades. To aggregate the glandular features and characterize the cores, we use the Bag of Word (BoW) model developed in [21]. This model combines the features of several glands inside a single core to describe the whole core. The model is constructed as shown in Fig. 4.

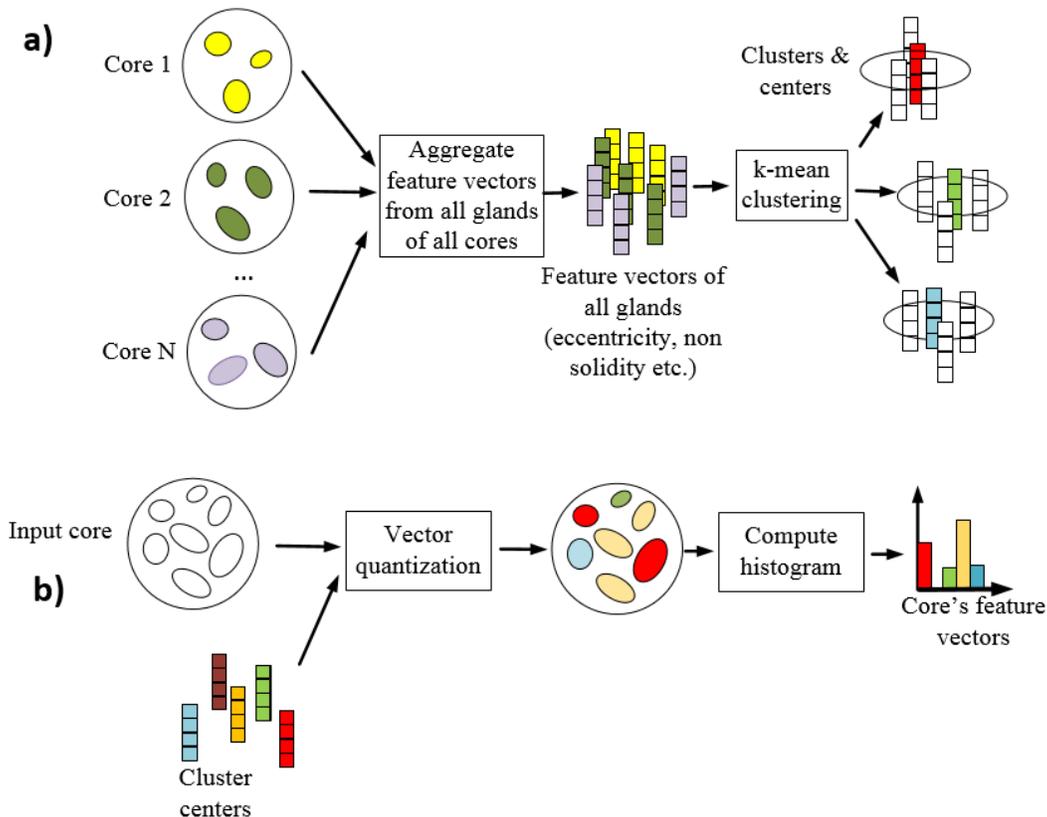


Figure 4. (a) Bag-of-word model for cores: The feature vectors of all glands in the cores are aggregated and clustered into different groups using k-mean clustering. The cluster centers are used as the “words”. (b) Feature extraction for each core: Each gland in the core is associated with a closest word. Then, a normalized histogram of word count is used to characterize the core.

First, feature vectors of all glands from all cores are collected. Here, from all 302 cores, 4227 gland feature vectors are extracted. Each a feature vector of dimension of 6. Then, k-mean clustering is used to partition these 4227 vector into $k=15$ different clusters. The centers of these clusters will be used as the “words” in the BoW model (see Fig. 4 (a)) while the normalized histogram of word count will be use as descriptor for the whole core (see Fig. 4 (b)).

Next, we test our BoW model in classifying cancer cases of Gleason grade 3 vs. Gleason grade 4. Here, 94 grade 3 and 32 grade 4 cancer cores were used. Since grade 2 diagnosis is not given on cores, we gather all cores with the biopsy diagnosis of grade 2 and grade 3 in majority to the grade 3 group. Meanwhile, 32 Grade 4 cancer cores come from those with biopsy diagnosis of grade 4 for the majority pattern. As a side note, here, Gleason score of biopsy evaluation is used for the ground truth instead of core evaluation.

3. RESULTS

Our code is implemented in MATLAB with the MexOpenCV plugin for Random Forest obtained from <http://vision.is.tohoku.ac.jp/~kyamagu/software/mexopencv/>. The training dataset for pixel classification consists of 4.92 million histograms of texton indices with two possible labels (“gland” or “stroma”). These histograms of texton come from 246 cores with manual labeling. For each core, 20,000 histograms of texton indices in stroma and gland regions area are randomly sampled. Note that, the whole dataset consist of 302 cores with more than 9 million histograms each. Therefore, our training set is approximately 0.18% of the total dataset. A Random Forest of 50 trees are trained. Each tree in our forests is trained on 11 randomly selected features (out of 50, the number of textons). The training are performed recursively and terminated at leaf nodes when the maximum depth $D=25$ is reached or less than 20 samples is available. Total training time is approximately 5-6 hours. Some typical segmentation results are shown in Fig. 5. Here, four different cores are displayed. The first three cores have cancer with Gleason grading from grade 3 to grade 5. The last core has High Grade Prostatic Intraepithelial Neoplasia (HGPIN), a condition in which the core is not cancer but its glandular expressions are similar to that of high-grade carcinoma. It can be seen in grade 3, the glands are well-defined. Therefore, automatic segmentation results match very well with manually ones. In grade 4, although glands lose their shapes due to fusion and distortion, the algorithm is robust enough to produce good segmentation. Moving to grade 5, some glands are not fully segmented due to heavy change in the glandular formation of epithelial cells. Therefore, only small portions of the glands or sheets of epithelial cells are detected. These remarks are also reflected in the plot of ROC curves for pixel classification shown in **Figure 6 (a)**.

Next, based on segmented maps, we build the BoW model to compute feature vector for each core in the dataset following the procedure in Section 2.2. Then, a Support Vector Machine (SVM) classifier is trained to differentiate Gleason grade 3 and 4. To handle the imbalance of the dataset between these two classes, we use the Synthetic Minority Oversampling Technique [22] to oversample 32 samples of Grade 4 with 5 nearest neighbors into 92 samples before training. After oversampling with SMOTE, a Principle Component Analysis (PCA) step is performed. Based on the explained variance of the PCA, we keep 8 principal eigenvectors and project the feature vectors of all cores into the subspace spanned by these PCA vectors after subtracting their mean. Using 4-fold stratified cross-validation on the new dataset and the standard SVM classifier with Radial Basis Function (RBF) kernel, the average accuracy of correctly classifying Grade 3 cores as Grade 3 is 68.08% and that of classifying Grade 4 cores as Grade 4 is 71.27%. The Receiver Operating Curves (ROCs) for correctly classifying the Grade 4 cores for each fold and the average curve is shown in Fig. 6 (b). The average ROC curve has an Area Under the Curve (AUC) of 0.79.

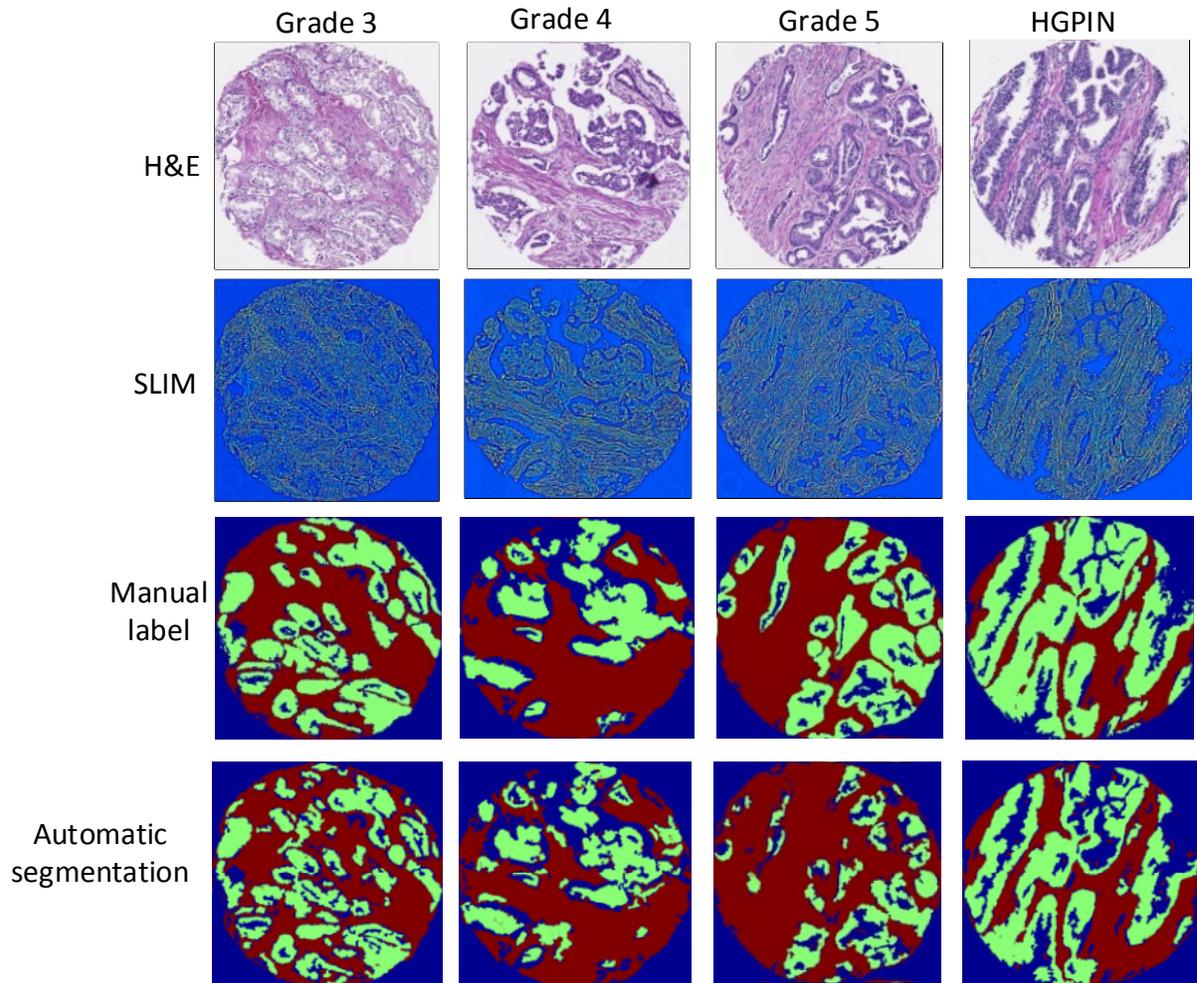


Figure 5. Segmentation results. First row: (from left to right) four H&E images of 4 cores containing Gleason patterns grade 3, 4, 5 and HGPIN. Second row: corresponding SLIM images. Third & fourth rows: manual and automatic labeled images. It be further seen that the performance of the classifier is reduced when the grade is higher. Generally, when the grade is high, the density of epithelial cell become less uniform, which makes the classifier produces under-segmented gland areas

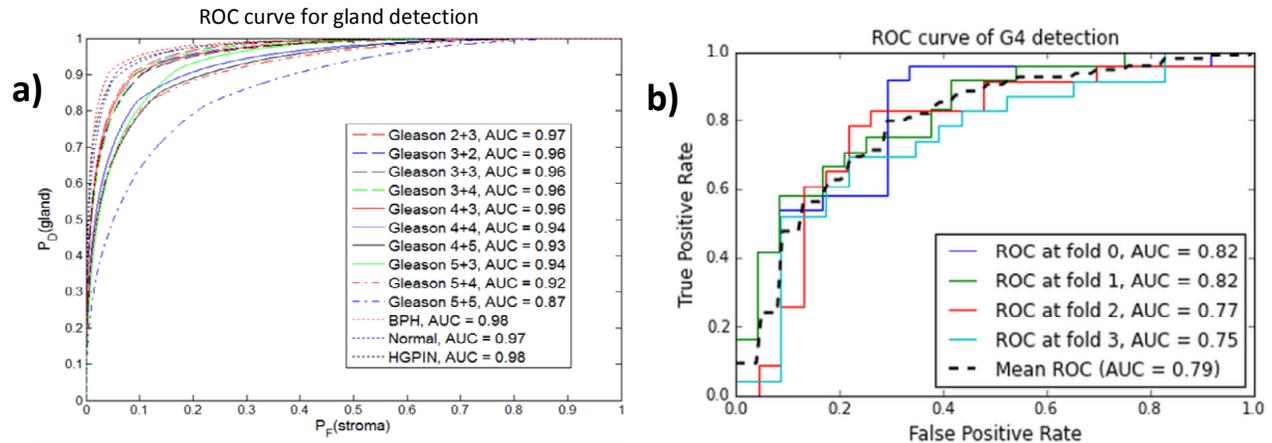


Figure 6. (a) Receiver Operating Curves for different diagnosis results in the binary classification problem between glands vs. stroma. Lumen pixels are excluded from the classification. (b) Four different ROC curves of the 4 folds in stratified 4-fold cross-validation and the average ROC curve (dashed one).

4. SUMMARY & FUTURE WORK

This paper proposes a unified framework to perform automatic diagnosis of prostatic cancer directly from QPI images. The framework utilizes texture analysis to produce segmentation maps for different regions of the cores. Given the label map, all glands in each core are identified and their feature vectors are calculated. BoW model is used to compute the core feature vector from its glands. Using standard SVM classifiers, we showed that an accuracy of about 70% can be obtained in the binary classification problem for cores of Grade 3 and Grade 4 cancer. Further work on this project includes improving misclassification due to outliers (e.g. inflammatory cells, blood vessels, nerves), and incorporating more classes into automatic diagnosis.

5. ACKNOWLEDGMENTS.

This work was supported by the National Science Foundation CBET-1040461 MRI and Agilent Laboratories. For more information, visit <http://light.ece.illinois.edu/>.

REFERENCES

- [1] U. C. S. W. Group, "United States cancer statistics: 1999–2010 incidence and mortality web-based report," Atlanta, GA, (2013).
- [2] D. F. Gleason, and E. M. Tannenbaum, "The veteran's administration cooperative urologic research group: Histologic grading and clinical staging of prostatic carcinoma," *Urologic Pathology: The Prostate*, 171-198 (1977).
- [3] J. Diamond, N. H. Anderson, P. H. Bartels *et al.*, "The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia," *Human Pathology*, 35(9), 1121-1131 (2004).
- [4] R. Farjam, H. Soltanian-Zadeh, R. A. Zoroofi *et al.*, "Tree-structured grading of pathological images of prostate." 840-851.
- [5] S. Naik, S. Doyle, M. Feldman *et al.*, "Gland segmentation and computerized gleason grading of prostate histology by integrating low-, high-level and domain specific information."
- [6] K. Nguyen, B. Sabata, and A. K. Jain, "Prostate cancer grading: Gland segmentation and structural features," *Pattern Recognition Letters*, 33(7), 951-961 (2012).
- [7] G. Popescu, [Quantitative Phase Imaging of Cells and Tissues] Mcgraw-Hill, (2011).
- [8] T. H. Nguyen, and G. Popescu, "Spatial Light Interference Microscopy (SLIM) using twisted-nematic liquid-crystal modulation," *Biomedical optics express*, 4(9), 1571-1583 (2013).
- [9] Z. Wang, L. J. Millet, M. Mir *et al.*, "Spatial light interference microscopy (SLIM)," *Opt. Exp.*, 19(1016), (2011).

- [10] B. Bhaduri, C. Edwards, H. Pham *et al.*, "Diffraction phase microscopy: principles and applications in materials and life sciences," *Advances in Optics and Photonics*, 6(1), 57-119 (2014).
- [11] G. Popescu, T. Ikeda, R. R. Dasari *et al.*, "Diffraction phase microscopy for quantifying cell structure and dynamics," *Opt. Lett.*, 31(6), 775-777 (2006).
- [12] E. Cuche, F. Bevilacqua, and C. Depeursinge, "Digital holography for quantitative phase-contrast imaging," *Opt. Lett.*, 24(5), 291-293 (1999).
- [13] P. Bon, G. Maucort, B. Wattellier *et al.*, "Quadriwave lateral shearing interferometry for quantitative phase microscopy of living cells," *Opt. Express*, 17(15), 13080-13094 (2009).
- [14] Z. Wang, K. Tangella, A. Balla *et al.*, "Tissue refractive index as marker of disease," *J. Biomed. Opt.*, 16(11), (2011).
- [15] L. Breiman, "Random forests," *Machine Learning*, 45, 5-32 (2001).
- [16] R. Simon, K. Sundar, and N. Mukunda, "Twisted Gaussian Schell-model beams. I. Symmetry structure and normal-mode spectrum," *JOSA A*, 10(9), 2008-2016 (1993).
- [17] M. Varma, and A. Zisserman, [Classifying images of materials: Achieving viewpoint and illumination independence] Springer, (2002).
- [18] T. Leung, and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *International Journal of Computer Vision*, 43(1), 29-44 (2001).
- [19] B. Glocker, O. Pauly, E. Konukoglu *et al.*, [Joint classification-regression forests for spatially structured multi-object segmentation] Springer, (2012).
- [20] A. Tabesh, V. P. Kumar, H.-Y. Pang *et al.*, "Automated prostate cancer diagnosis and Gleason grading of tissue microarrays." 58-70.
- [21] L. Fei-Fei, and P. Perona, "A bayesian hierarchical model for learning natural scene categories." 2, 524-531.
- [22] N. V. Chawla, K. W. Bowyer, L. O. Hall *et al.*, "SMOTE: synthetic minority over-sampling technique," *arXiv preprint arXiv:1106.1813*, (2011).