Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Detecting Prostatic Adenocarcinoma From Digitized Histology Using a Multi-Scale Hierarchical Classification Approach

Scott Doyle, Carlos Rodriguez,
Anant Madabhushi
Dept. of Biomedical Engineering
Rutgers University
Piscataway, NJ 08854, USA
anantm@rci.rutgers.edu

John Tomaszeweski, Michael Feldman
Dept. of Surgical Pathology
University of Pennsylvania
Philadelphia, PA 19104, USA
feldmanm@mail.med.upenn.edu

Abstract— **In this paper we present a computer-aided diagnosis (CAD) system to automatically detect prostatic adenocarcinoma from high resolution digital histopathological slides. This is especially desirable considering the large number of tissue slides that are currently analyzed manually – a laborious and time-consuming task. Our methodology is novel in that texture-based classification is performed using a hierarchical classifier within a multi-scale framework. Pyramidal decomposition is used to reduce an image into its constituent scales. The cascaded image analysis across multiple scales is similar to the manner in which pathologists analyze histopathology. Nearly 600 different image texture features at multiple orientations are extracted at every pixel at each image scale. At each image scale the classifier only analyzes those image pixels that have been determined to be tumor at the preceding lower scale. Results of quantitative evaluation on 20 patient studies indicate (1) an overall accuracy of over 90% and (2) an approximate 8-fold savings in terms of computational time. Both the AdaBoost and Decision Tree classifiers were considered and in both cases tumor detection sensitivity was found to be relatively constant across different scales. Detection specificity was however found to increase at higher scales reflecting the availability of additional discriminatory information.**

Index Terms— **Hierarchical classifier, decision trees, AdaBoost, prostate cancer, digitized histology.**

I. INTRODUCTION

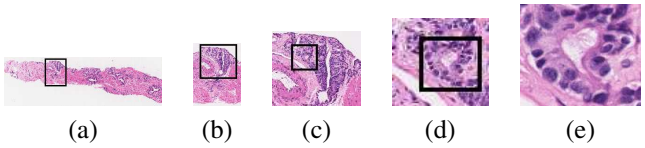
Prostate cancer is a major problem in the United States, with a predicted 234,000 cases and 27,000 deaths in 2006 according to the American Cancer Society. Patient prognosis is greatly increased if the condition is diagnosed early. The current gold standard for prostate cancer diagnosis is histological analysis of tissue samples obtained via transrectal ultrasound (TRUS) biopsy. Current TRUS protocols mandate between 12-20 biopsy samples per patient. The low accuracy of TRUS (20-25%) for elevated prostate specific antigen levels means that pathologists spend several man-hours sieving through mostly benign tissue.

The advent of digital high-resolution scanners has made available digitized histological tissue samples that are amenable to computer-aided diagnosis (CAD). CAD can relieve the pathologists' burden by discriminating obviously benign and malignant tissue so as to reduce the amount of tissue area to be analyzed by a pathologist. While histology-based CAD is relatively recent compared to radiology-based CAD, some researchers have developed CAD methods to analyze prostate histology. Previous CAD work has mostly used color, texture, and wavelet features [1], texture-based second-order features [2], or morphological attributes [3] to distinguish manually defined regions of interest on the image. The choice of scale at which to do the image analysis, however, is typically arbitrary. This ad hoc scale selection runs contrary to the multi-scale approach adopted by pathologists who usually identify suspicious regions at lower resolutions and only use the information at the higher scales (where the high level shape and architectural information is present) to confirm their suspicions (Figure 1). Figure 1 shows an image of digitized prostate histopathology at multiple scales. While low level attributes such as texture and intensity are available at the lower image scales (Figure 1 (a)-(c)) to distinguish benign from cancerous regions, higher level shape and architectural attributes of tissue become apparent only at the higher scales (Figure 1 (d), (e)). In this paper we present a multi-scale approach to detecting prostate cancer from digitized histology. Nearly 600 texture and intensity features are extracted at every image pixel and at every image scale. A hierarchical classification scheme (a variant of the cascade classifier originally proposed by Viola and Jones [4]) at each scale analyzes only those regions that were determined as suspicious in the scale immediately preceding it. Thus without compromising on the sensitivity of cancer detection, the classifier's detection specificity increases at higher scales. Our hierarchical CAD paradigm is not specific to any particular classifier and similar results are obtained with the Decision Tree [7] and AdaBoost [6] algorithms. The novel aspects of this work are in the following.

1) Nearly 600 texture features at multiple orientations are extracted to build signature vectors to distinguish adenocarcinoma from benign stromal epithelium,
2) A multi-resolution approach is used wherein feature extraction and feature classification are performed at each image scale, which is similar to the manner in which a pathologist analyzes tissue slides, and the
3) Use of a hierarchical classifier (with the AdaBoost [6] and Decision Tree [7] algorithms) to analyze specific regions at each image scale determined as tumor on the immediate preceding scale significantly helps reduce execution time while simultaneously not compromising on accuracy.

1-4244-0033-3/06/$20.00 ©2006 IEEE.

The rest of this paper is organized as follows. In Section II we describe our methodology. In Section III we present our main results (qualitative and quantitative). Our conclusions are presented in Section IV.



(a) (b) (c) (d) (e)

Fig. 1. Digitized histological image at multiple scales. The higher scales ((d), (e)) yield incrementally more discriminatory shape and architectural information.

II. METHODS

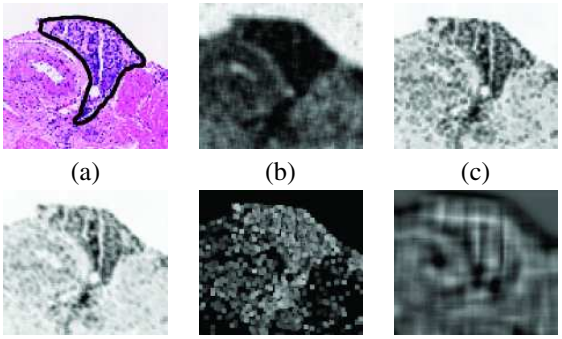
Our methodology comprises the following steps. Following slide digitization, the image is decomposed into its constituent scales and at every image scale nearly 600 texture features are extracted. A hierarchical classifier is then trained using pixels at each scale that have been manually labeled as "tumor" or "non-tumor" by an expert pathologist. In this work, we consider the AdaBoost [6] and Decision Tree classifiers [7] in the multi-scale paradigm to classify each image pixel as either benign or malignant. Only pixels determined to be cancer at the preceding scale are analyzed at each subsequent higher scale. Details on the individual modules are described below.

A. Image Acquisition and Decomposition

A set of 20 H&E stained prostate tissue samples were scanned using a high resolution glass tissue slide scanner at a magnification of 40x optical resolution at the Hospital at the University of Pennsylvania, Department of Surgical Pathology. The images were subsequently saved as TIFF images. We represent each scanned tissue slide image by a pair $\mathcal{C} = (C, f)$, where C is a 2D grid of image pixels c and f is the intensity at each pixel $c \in C$. The digital image $\mathcal{C}$ is decomposed into its constituent scales by using Burt's pyramidal scheme [5]. The set of image scales for $\mathcal{C}$ is denoted as $\mathcal{S}(\mathcal{C}) = \{\mathcal{C}^1, \mathcal{C}^2, \cdots, \mathcal{C}^n\}$, where n is the total number of image scales. The motivation behind the multi-scale framework is derived from the approach employed by pathologists to analyze tissue samples. In addition, the large size of the original uncompressed TIFF images (>1 GB) mandates analyzing the images at the lower scales first. In this study, we only consider the lowest 3 scales ($\mathcal{C}^1$, $\mathcal{C}^2$, $\mathcal{C}^3$) for each image scene $\mathcal{C}$.

B. Feature Extraction

Each image $\mathcal{C}$ is first converted from the RGB space to the HSI space. We extract a set of K feature scenes $\mathcal{F}_\gamma^j = (C^j, g_\gamma^j)$ for $\gamma \in \{1, 2, \cdots, K\}$ from each $\mathcal{C}^j \in \mathcal{S}(\mathcal{C})$ where for any $c^j \in C^j$, $g_\gamma^j(c^j)$ is the value of texture feature Φ_γ at scale j and at pixel c. The choice of features was motivated by the textural appearance of prostatic adenocarcinoma at $\mathcal{C}^1$, $\mathcal{C}^2$, and $\mathcal{C}^3$. Since only the lower 3 scales were considered for analysis, morphological and architectural features (visible only at higher scales) were not used. The extracted features included 13 statistical features (average, median, standard deviation, difference, Sobel and Kirsch filters, and derivatives in the horizontal, vertical, and diagonal directions), 13 co-occurrence features (angular second moment, contrast, correlation, variance, entropy, inverse difference moment, sum average, variance, and entropy, difference variance, and difference entropy), and a bank of 40 Gabor features at five different scales and eight different orientations. In Figure 2 (b)-(f) are shown 5 feature scenes corresponding to the original histological image in Figure 2 (a).



(a) (b) (c) (d) (e) (f)

Fig. 2. (a) Original digitized prostate histopathological image with the manual segmentation of cancer overlaid (black contour), and 5 feature scenes generated from (a) and corresponding to (b) correlation, (c) sum variance, (d) Gabor filter, (e) difference, and (f) standard deviation.

C. Training

An expert pathologist manually segmented tumor regions on the 20 patient studies in our database. The feature values $g_\gamma^j(c^j)$ for all pixels c^j in the training images at each scale $j \in \{1, 2, 3\}$ were used to train the AdaBoost [6] and the Decision Tree [7] classifiers.

D. Hierarchical Classifier

In order to exploit information at multiple scales and to reduce computational complexity we employ a hierarchical classifier which is a variant of the cascade algorithm proposed by Viola and Jones [4]. Hence, at each scale only those pixels that were classified as adenocarcinoma at the immediate preceding scale are analyzed. Below, we briefly describe the two classifiers (Decision Trees and AdaBoost) that were employed within the cascade framework.

Decision Trees A decision tree is trained using an iterative selection of individual features that are the most salient at each node in the tree [7]. A popular algorithm for generating decision trees is C4.5, proposed by Quinlan [7]. The rules generated by this algorithm are of the form "if X and Y then Z," where X and Y are the rule antecedents, and Z is the rule consequence. The C4.5 tree is trained by labeled instances corresponding to the tumor ω_T and non-tumor ω_{NT} classes. Each object label acts as a leaf in the tree, and the algorithm initially maps out a path to the object using the feature values of the training data. These paths are then pruned using a greedy elimination rule which removes paths from the tree that are not sufficiently

discriminatory. Hence for any pixel c^j at scale j, $\delta_k^j(c^j)$ represents the classification obtained for the decision tree trained with set S_k^j. For a total of A different training sets $(S_1^j, S_2^j, \cdots, S_A^j)$ we obtain A uncorrelated classifications for pixel c^j which are then combined at scale j as,

$$\Delta^j(c^j) = \sum_{k=1}^{A} \delta_k^j(c^j) \quad (1)$$

Boosting The AdaBoost algorithm [6] is used to combine a number of weak learners to generate a strong classifier. Pixels determined as cancer by a pathologist during the training stage are used to generate probability density functions (pdf's) for each of the individual texture features Φ_γ, for $1 \leq \gamma \leq K$, at each image scale j. Bayes Theorem [8] is then used to generate likelihood scenes $\mathcal{L}_\gamma^j = (\mathcal{C}^j, l_\gamma^j)$ for each Φ_γ at each j which constitute the weak learners. These are combined by the AdaBoost algorithm [6] into a strong classifier $\Pi^j = \sum_{i=1}^{I^j} \alpha_i^j l_i^j$ where for every pixel $c^j \in C^j$, $\Pi^j(c^j)$ is the combined likelihood that pixel c^j belongs to cancer class ω_T, α_i^j is the weight determined during training for feature Φ_i, and I^j is the number of iterations at scale j. We used $I^j < I^{j+1}$ since additional discriminatory information for the classifier only becomes available at higher scales.

For each of the Decision Tree and AdaBoost classifiers a binary scene $\mathcal{C}^{j,B} = (C^j, f^{j,B})$ is created where for $c^j \in C^j$, $f^{j,B}(c^j) = 1$ iff $\Pi^j(c^j)$, $\Delta^j(c^j) > \tau^j$, where τ^j is a predetermined threshold. $\mathcal{C}^{j,B}$ is resized via interpolation to obtain $\mathcal{C}^{j+1,B} = (C^{j+1}, f^{j+1,B})$. Feature extraction and classification steps are repeated at scale $j+1$, but only for those pixels $c^{j+1} \in C^{j+1}$ determined as cancer at the preceding scale (i.e. $f^{j+1,B}(c^{j+1}) = 1$).

III. RESULTS AND DISCUSSION

The CAD system was evaluated in terms of (1) efficiency, (2) segmentation accuracy and (3) reproducibility at each image scale. Efficiency was evaluated to determine the savings in computation time by using the hierarchical framework as opposed to not. Accuracy was evaluated in terms of Receiver Operating Characteristic (ROC) curves, obtained by thresholding the classifier outputs (Π^j, Δ^j) at different values and computing tumor detection sensitivity and specificity. Tumor segmentations were evaluated against cancer ground truth masks obtained manually during training. The system was also evaluated in terms of reproducibility of results for changes in training data and kind of classifier (Decision Tree or AdaBoost) employed.

A. Efficiency

Figure 3 shows the total average computation times for feature extraction, feature calculation, and feature combination on 17 test images at 3 different scales with and without the use of the hierarchical classifier. In this case 3 randomly selected images were used for training. All computations were done on a 3 GHz Pentium IV Dell computer (2 GB RAM) using MATLAB 7.1. The average image sizes (in pixels) at the 3 different scales were 350×350, 700×700, and 1400×1400. The hierarchical classifier produced an average 4-fold and 8-fold savings in computation time at scales $j = 2$ and $j = 3$ respectively.

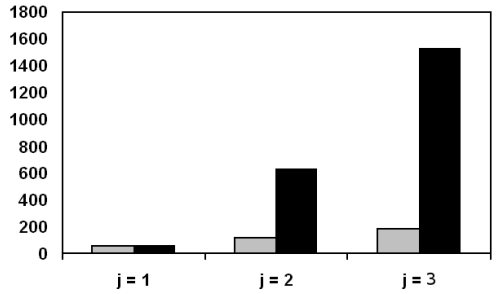


Fig. 3. Computation times (in minutes) for the CAD at each image scale ($j \in \{1, 2, 3\}$) with (gray bar) and without (black bar) using the hierarchical approach.

B. Accuracy

Figure 4 shows the combined likelihood scenes for 3 different prostate studies (Figure 4 (a), (e), (i)) at 3 different scales. Very little discriminability is observed on the classifier outputs at scale $j = 1$ between the tumor and benign classes owing to the paucity of relevant and resolvable tumor information at this scale. The classifier results at scales $j = 2$ (Figure 4 (c), (g), (k)) and $j = 3$ (Figure 4 (d), (h), (l)) show greater discriminability between benign and malignant regions (compare Figures 4 (c), (g), (k) and 4 (d), (h), (l) with the tumor masks (in black) on 4 (a), (e), (i)), with the results at scale $j = 3$ appearing to be more specific compared to the corresponding results at $j = 2$.

In Figure 5 (a) are shown the ROC curves for the hierarchical boosting classifier at scales $j \in \{1, 2, 3\}$ for a total of 51 images (17 test images at 3 scales). As can be discerned by the larger area under the ROC curve, the classifier performance at scale $j = 3$ is significantly superior compared to scales $j = 1$ and $j = 2$. Similarly classifier performance at scale $j = 2$ is superior compared to that obtained at scale $j = 1$. These results indicate that while the sensitivity of the classifier remains unchanged across the 3 scales, the specificity increases with higher scales. Since more tumor related detail and information becomes available at the higher scales, the number of false positive errors reduces.

C. Reproducibility

Figure 5 (b) shows the ROC curves obtained for a subset of the testing data that was trained using 3 different training sets containing 3, 5, and 8 images respectively. The similarity of the area under the curves indicates that the classification performance does not change appreciably with variations in the training data.

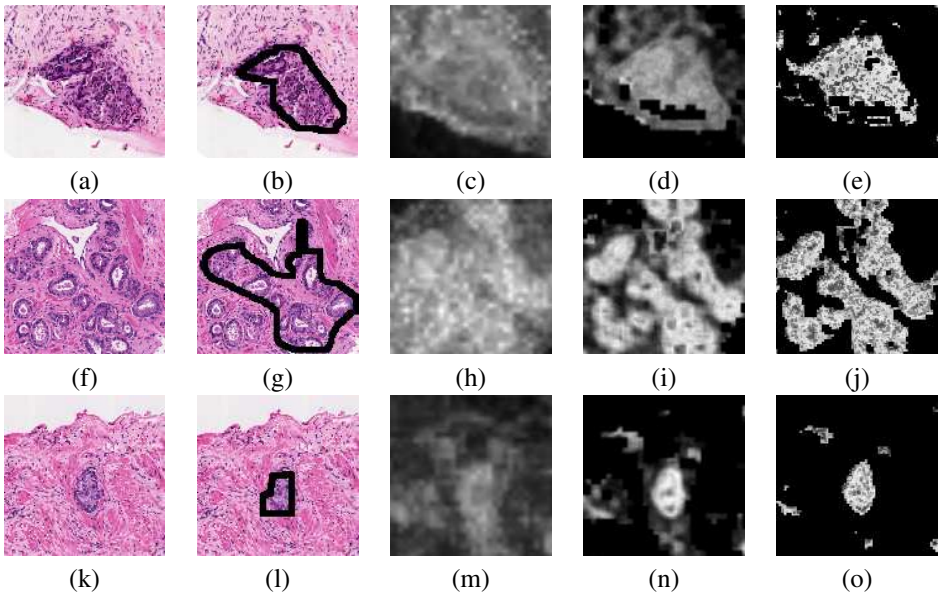


Fig. 4. (a), (f), (k) Digital histopathological prostate studies. (b), (g), (l) Tumor regions corresponding to the histopathological studies in (a), (f), and (k). Classifier outputs indicating regions with a high likelihood of being cancer at scale $j = 1$ ((c), (h), (m)), $j = 2$ ((d), (i), (n)), and $j = 3$ ((e), (j), (o)).

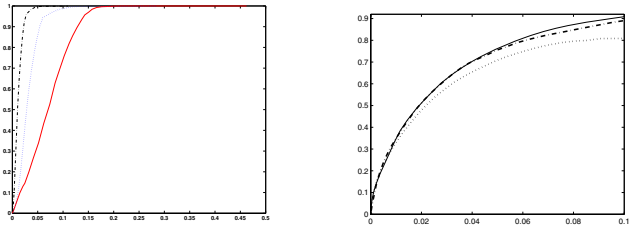


Fig. 5. (a) ROC plots of the hierarchical AdaBoost classifier at scales $j = 1$ (solid line), $j = 2$ (dotted line), and $j = 3$ (dot-dashed line). Note that while detection sensitivity remains constant across scales, specificity increases as a function of scale. (b) ROC curves for different sets of training data, using 3 (dot-dashed line), 5 (dotted line), and 8 (solid line) images. The similarity between the area under the curve indicates robustness to training.

IV. CONCLUDING REMARKS

In this paper we have presented a fully automated CAD system to detect prostatic adenocarcinoma from digitized histological images. Our methodology involves extracting nearly 600 texture features at multiple scales. A hierarchical classifier is used to efficiently and accurately distinguish image pixels between the cancer and benign classes at every scale. At each image scale only those pixels determined as cancer on the preceding image scale are analyzed, leading to a highly computationally efficient algorithm which does not compromise on accuracy. In fact on a total of 20 studies our results indicate that while the tumor detection sensitivity is relatively constant across different scales, detection specificity increases at higher scales due to introduction of additional cancer relevant information not resolvable at lower scales. Our results were confirmed independently by two classifiers – AdaBoost and Decision Trees. Future work will focus on integrating information at higher scales into the CAD framework. The additional morphological and architectural information at these scales will enable us to not only distinguish benign and malignant prostate tissue, but also different grades of prostatic adenocarcinoma.

ACKNOWLEDGMENTS

This work was supported by grants from the Wallace H. Coulter foundation (WHCF 4-29368, 4-29349) and the Office of Technology Transfer (NJCST 4-21754) at Rutgers University.

REFERENCES

[1] Wetzel, A.W., et al., Evaluation of prostate tumor grades by content based image retrieval, *Proc. of SPIE*, 1999, pp 244-252.
[2] Esgiar, A.N., et al., Microscopic image analysis for quantitative measurement and feature identification of normal and cancerous colonic mucosa, *IEEE Trans. on Inf. Tech. in Biomedicine*, 1998, pp 197-203.
[3] Tabesh, A., et al., Automated prostate cancer diagnosis and Gleason grading of tissue microarrays, *Proc. of SPIE*, 2005, pp 58-70.
[4] Viola, P., and Jones, M., Robust Real-Time Face Detection, *Intl. J. of Computer Vision*, vol. 57(2), 2004, pp 137-154.
[5] Adelson, E.H., Burt, P.J., Image data compression with the Laplacian pyramid, *Proc. of Patt. Recog. Inf. Proc.*, 1981, pp 218-223.
[6] Freund, Y., and Schapire, R., Experiments with a new boosting algorithm, *Proc. of Natural Conf. on Machine Learning*, 1996, pp 148-156.
[7] Quinlan, J.R., Bagging, Boosting, and C4.5, *13th Nat. Conf. on Artificial Intel.*, 1996, pp 725-730.
[8] Duda, R.O., and Hart, P.E., Pattern Classification and Scene Analysis, NY, Wiley, 1973.