A Boosting Cascade for Automated Detection of Prostate Cancer from Digitized Histology

Scott Doyle[1], Anant Madabhushi[1], Michael Feldman[2], and John Tomaszeweski[2]

[1] Dept. of Biomedical Engineering, Rutgers Univ., Piscataway, NJ 08854, USA
[2] Dept. of Surgical Pathology, Univ. of Pennsylvania, Philadelphia, PA 19104, USA

Abstract. Current diagnosis of prostatic adenocarcinoma is done by manual analysis of biopsy tissue samples for tumor presence. However, the recent advent of whole slide digital scanners has made histopathological tissue specimens amenable to computer-aided diagnosis (CAD). In this paper, we present a CAD system to assist pathologists by automatically detecting prostate cancer from digitized images of prostate histological specimens. Automated diagnosis on very large high resolution images is done via a multi-resolution scheme similar to the manner in which a pathologist isolates regions of interest on a glass slide. Nearly 600 image texture features are extracted and used to perform pixel-wise Bayesian classification at each image scale to obtain corresponding likelihood scenes. Starting at the lowest scale, we apply the AdaBoost algorithm to combine the most discriminating features, and we analyze only pixels with a high combined probability of malignancy at subsequent higher scales. The system was evaluated on 22 studies by comparing the CAD result to a pathologist's manual segmentation of cancer (which served as ground truth) and found to have an overall accuracy of 88%. Our results show that (1) CAD detection sensitivity remains consistently high across image scales while CAD specificity increases with higher scales, (2) the method is robust to choice of training samples, and (3) the multi-scale cascaded approach results in significant savings in computational time.

1 Introduction

There will be an estimated 234,000 new cases of prostate cancer in the US in 2006, and approximately 27,000 men will die on account of it (Source: *American Cancer Society*). Trans-rectal ultrasound (TRUS) guided biopsy of the prostate followed by histological analysis under a microscope is currently the gold standard for prostate cancer diagnosis [1]. Up to twenty biopsy samples may be taken from a single TRUS procedure, making manual inspection time-consuming and labor-intensive. Computer-aided diagnosis (CAD), the use of computers to assist clinical diagnosis, has been traditionally applied to radiological images. Madabhushi, et al. [2] presented a powerful CAD system to automatically detect prostatic adenocarcinoma from high-resolution prostate MRI studies. The recent advent of high resolution whole slide digital scanners, however, has made histopathology amenable to CAD as well.

In the context of prostate histology, CAD methods have been proposed which utilize image features such as color, texture, and wavelets [3], textural second-order statistical [4], and morphometric attributes [5] to characterize and detect cancer. However, in these studies the image analysis operations are applied at arbitrarily chosen image scales. This is contrary to the multi-scale approach employed by pathologists who obtain most of the information needed for a definitive diagnosis at the coarser image scales with the finer or higher scales usually serving to confirm their diagnoses. An effective CAD system to automatically detect prostatic adenocarcinoma should therefore incorporate the spirit of this hierarchical, multi-scale paradigm. In [6] Viola and Jones proposed a computationally efficient "Boosting Cascade" in which the AdaBoost classification algorithm [7] was used to quickly classify image regions using a small number of image features. The process is repeated using an increasingly larger number of image features and an increasing classification threshold at each iteration.
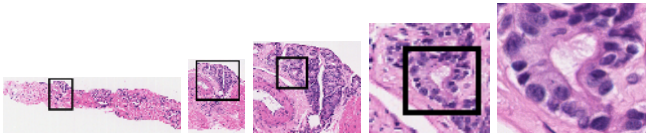
Fig. 1. A multi-scale representation of digitized human prostate histopathology

In this work, we propose a fully automated CAD system to extract and then combine multiple texture features within a Boosting Cascade framework [6] to detect prostatic adenocarcinoma from digitized histology. Pyramidal decomposition [8] is first applied to reduce the image into its constituent scales (Figure 1). At each image scale, we extract nearly 600 texture features at every image pixel. A likelihood scene corresponding to each texture feature is generated, in which the intensity at every pixel represents its probability of malignancy. A Boosting Cascade scheme is used to efficiently and accurately combine the different likelihood scenes at each image scale. Only pixels identified as adenocarcinoma with a pre-determined confidence level at each specific scale are analyzed further at the subsequent higher image scales. The novelty of our work lies in the following:

• The method is fully automated and involves extraction of nearly 600 texture features at multiple scales and orientations to discriminate between benign and malignant tissue regions.
• The use of a multi-scale classification framework to accurately and efficiently analyze very large digitized specimens (> 2 GB). Hence, only those pixels determined as adenocarcinoma with high probability at a given scale are considered for further analysis at the subsequent higher scales.

The rest of this paper is organized as follows. In Section 2 we describe the methodology and in Section 3 we present our results. Concluding remarks and future directions are presented in Section 4.

2   Methodology

2.1   Data Description and System Overview

Human prostate tissue samples cut into 6 μm slices are scanned into the computer at $40\times$ optical magnification. Typical image file sizes were between 1-2 GB. We represent each digitized 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 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 and $\mathcal{C}^j = (C^j, f^j)$ for $j \in \{1, 2, \cdots, n\}$ is the representation of $\mathcal{C}$ at scale j, for $1 \leq j \leq n$. Hence, $\mathcal{C}^1$ represents the image at the coarsest scale and $\mathcal{C}^n$ at the finest scale. Our methodology is outlined in the flowchart in Figure 2. Digital scenes are acquired from a whole slide digital scan-

Fig. 2. Outline of our methodology

ner and are decomposed into n constituent scales using Burt's pyramidal scheme [8]. At each scale, feature extraction is performed to create a series of likelihood scenes using Bayes Theorem [9]. An expert pathologist manually segmented cancer regions from $\mathcal{S}(\mathcal{C})$ for each of 22 images. During the training stage (off-line) probability density functions (pdf's) for cancer for each of the texture features are generated using the cancer masks determined by the expert. Following feature extraction, Bayesian classification via the feature pdf's is used to generate cancer likelihood scenes for each feature. At each scale j the various likelihood scenes are combined via the AdaBoost algorithm [7]. Only regions determined as cancer at scale j with a pre-specified confidence level are considered for analysis at scale $j + 1$.
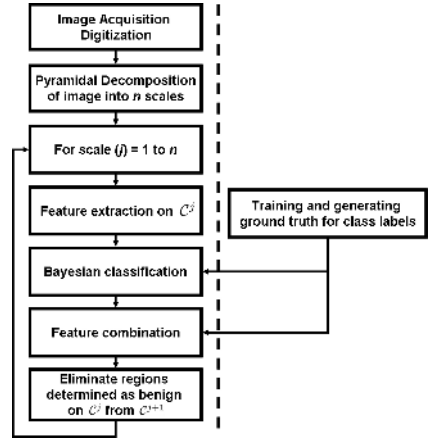
2.2   Feature Extraction

Each image $\mathcal{C}$ is first converted from the RGB color space to the HSI space. We obtain 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 feature Φ_γ at scale j and at pixel c. The choice of features was motivated by the textural appearance of prostatic adenocarcinoma at the 3 scales ($\mathcal{C}^1$, $\mathcal{C}^2$, $\mathcal{C}^3$) considered for analysis. A total of 594 texture features from the following three classes of texture operators were extracted.

First-Order Statistics: A total of 117 first-order statistical features from each image corresponding to average, median, standard deviation, difference, derivatives along the X, Y, and Z axes, 3 Kirsch filter features, and 3 Sobel

filter features were extracted at three different pixel neighborhood sizes (3×3, 5×5, 15×15).

Co-occurrence Features: A total of 117 Haralick features [10] corresponding to angular second moment, contrast, correlation, variance, inverse difference moment, entropy, sum average, sum variance, sum entropy, difference variance, difference entropy, and two measurements of correlation for three different pixel neighborhoods (3×3, 5×5, 7×7) were extracted.

Wavelet Features: The phase and orientation values of the result of applying a family of 360 Gabor filters were obtained at every image pixel [2]. A Gabor wavelet is a Gaussian function modulated by a sinusoid [11]. The modulating function G for the family of 2D Gabor filters is given as:

$$G(x, y, \theta, \kappa) = e^{-\frac{1}{2}((\frac{x'}{\sigma_x})^2 + (\frac{y'}{\sigma_y})^2)} \cos(2\pi\kappa x'), \tag{1}$$

where $x' = x\cos(\theta) + y\sin(\theta)$, $y' = y\cos(\theta) + x\sin(\theta)$, κ is the filter scale factor, θ is the filter phase, σ_x and σ_y are the standard deviations along the X, Y axes, and x and y are the 2D Cartesian coordinates of each image pixel. We convolved the Gabor kernel with the image at 3 pixel neighborhood sizes (3×3, 5×5, 15×15) using five different scale parameter values $\kappa \in \{0, 1, \cdots, 4\}$ and eight orientation parameter values ($\theta = \frac{\epsilon \cdot \pi}{8}$ where $\epsilon \in \{0, 1, \cdots, 7\}$). In Figures 3 ((b)-(f)) are shown some representative feature images for the digitized histopathological image in Figure 3 (a).
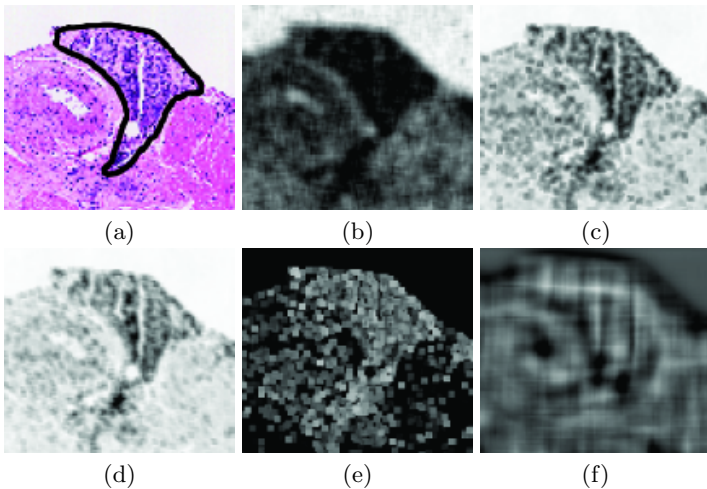
Fig. 3. (a) Original digitized prostate histopathological image with the manual segmentation of cancer overlaid (black contour), and 5 feature scenes generated from (a) and corresponding to (b) correlation (7×7), (c) sum variance (3×3), (d) Gabor filter ($\theta = \frac{5 \cdot \pi}{8}$, $\kappa = 2$, 3×3), (e) difference (3×3), and (f) standard deviation (15×15)

2.3   Training and Determining Cancer Ground Truth

Ground truth for the cancer class was generated by an expert pathologist who manually traced cancer regions on the digitized images at each image scale. The set of pixels marked by the pathologist as ground truth are denoted $E(\mathcal{C}^j)$ at scale j. The feature values g_γ^j of pixels $c^j \in E(\mathcal{C}^j)$ are used to generate pdf's $p_\gamma^j(c^j, g_\gamma^j | \omega_T)$ at each scale for the cancer class (ω_T), for each texture feature Φ_γ. In this study, we use 3 images to generate the pdf's. Figure 4 shows pdf's for 3 different texture features for the cancer and non-cancer classes at the lowest image scale $(j = 1)$.

(a)   (b)   (c)
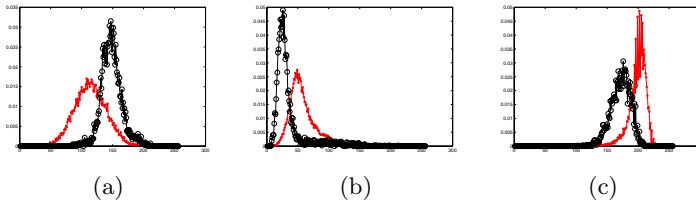
Fig. 4. Pdf's for cancer (red dots) and non-cancer regions (black circles) corresponding to (a) Gabor filter $(\theta = \frac{6*\pi}{8}, \kappa = 4, 15 \times 15)$, (b) difference entropy (3×3), and (c) correlation (3×3) at the lowest image scale $j = 1$

2.4   Feature Classification

For each scene $\mathcal{C} = (C, f)$, Bayes Theorem [9] is employed to obtain a series of likelihood scenes $\mathcal{L}_\gamma = (\mathcal{C}, l_\gamma)$, for $\gamma \in \{1, 2, \cdots, K\}$, where for each pixel $c \in C$, $l_\gamma(c)$ is the posterior conditional likelihood $P(\omega_T | c, g_\gamma)$ that c belongs to cancer class ω_T given feature value $g_\gamma(c)$. Using Bayes Theorem [9] the posterior conditional probability that c belongs to ω_T is given as

$$P(\omega_T | c, g_\gamma) = \frac{P(\omega_T) p_\gamma(c, g_\gamma | \omega_T)}{\sum_{v \in \{T, NT\}} P(\omega_v) p_\gamma(c, g_\gamma | \omega_v)} \tag{2}$$

where ω_{NT} denotes the non-cancer class, $p_\gamma(c, g_\gamma | \omega_T)$ is the *a-priori* conditional density obtained during training via the pdf for feature Φ_γ, and $P(\omega_T)$ and $P(\omega_{NT})$ are the prior probabilities of occurrence for the two classes (cancer and non-cancer), assumed as non-informative priors $(P(\omega_T) = P(\omega_{NT}) = 0.5)$.

2.5   Feature Combination and the Boosting Cascade

We employ a hierarchical version of the well-known classification ensemble scheme AdaBoost [7] to create a single, strong classifier from 594 likelihood scenes or base learners. The method comprises two steps: Training and Testing.

Training. We generate a Boosted classifier $\Pi^j = \sum_{i=1}^{I^j} \alpha_i^j l_i^j$ at each image scale j, where for every pixel $c^j \in C^j$, $\Pi^j(c^j)$ is the combined likelihood that pixel c^j belongs to class ω_T, α_i^j is the feature weight determined during training for base learner $\mathcal{L}_i$, and I^j is the number of iterations used to train the AdaBoost

algorithm. We used $I^j < I^{j+1}$ since additional discriminatory information is incorporated into the classifier only at higher scales. Three images randomly chosen from our database were used for training the Boosted classifier.

Testing. At scale j we create a combined likelihood scene $\mathcal{L}^j = (C^j, \Pi^j)$. Using $\mathcal{L}^j$, 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 $\Pi^j(c^j) > \delta^j$, where δ^j is a predetermined threshold. We then resize $\mathcal{C}^{j,B}$ to obtain $\mathcal{C}^{j+1,B} = (C^{j+1}, f^{j+1,B})$. The feature extraction and Bayesian classification steps are then repeated to obtain Boosted classifier $\mathcal{L}^{j+1}$, considering only those pixels c^j in $\mathcal{C}^{j,B}$ for which $f^{j,B}(c^j) > 0$. The Boosting Cascade algorithm is shown below.

Algorithm. $BoostingCascade()$
Input: Image pyramid $\mathcal{S}(\mathcal{C})$, ground truth for cancer $E(\mathcal{C})$,
                number of pyramidal levels n, set of predetermined thresholds δ^j
Output: Set L of binary cancer segmentations at each scale
begin
    0. *for* $j = 1$ to n *do*
    1.   Obtain combined likelihood scene $\mathcal{L}^j$ for $\mathcal{C}^j$ via AdaBoost [7];
    2.   Obtain tumor mask $\mathcal{C}^{j,B}$ by thresholding $\mathcal{L}^j$ at δ^j;
    3.   Obtain $\mathcal{C}^{j+1,B}$ by interpolating $\mathcal{C}^{j,B}$ so that $C^{j+1,B} = C^{j+1}$;
    4.   *for* each c^{j+1} in $\mathcal{C}^{j+1,B}$ *do*
    5.   *if* $f^{j+1,B}(c^{j+1}) < 1$ *then* $f^{j+1}(c^{j+1}) = 0$;
   6.   *endfor*
   7. $L[j] = \{\mathcal{L}^j\}$;
    8. *endfor*
    9. Output L;
end

3   Results and Discussion

Figure 6 (a) shows the ROC curves for our CAD system obtained by evaluating all 22 images in our database at 3 different scales. The increase in ROC area at higher scales corresponds to an increase in specificity, further reiterating that information at higher scales is necessary to achieve a more accurate classification. Figure 6 (b) shows the ROC curves for a subset of testing images that were trained using 3 training sets comprising 3, 5, and 8 images respectively. As can be observed, the curves have a similar area, indicating that CAD is robust with respect to training. In Figure 6 (c) is a bar chart showing the comparative computational savings by using the Boosting Cascade scheme at each image scale. As might be expected, the savings are greater at the higher scales; an 8-fold savings at $j = 3$. The system was evaluated on a total of 22 separate patient studies. CAD tolerance was evaluated in terms of (i) accuracy, (ii) precision (robustness to training), and (iii) computational complexity. Accuracy was evaluated via the receiver operating characteristic (ROC) curve [2]. Figure 5 shows qualitative results for 3 images in our database. Figures 5 (a), (f), and (k) show the original
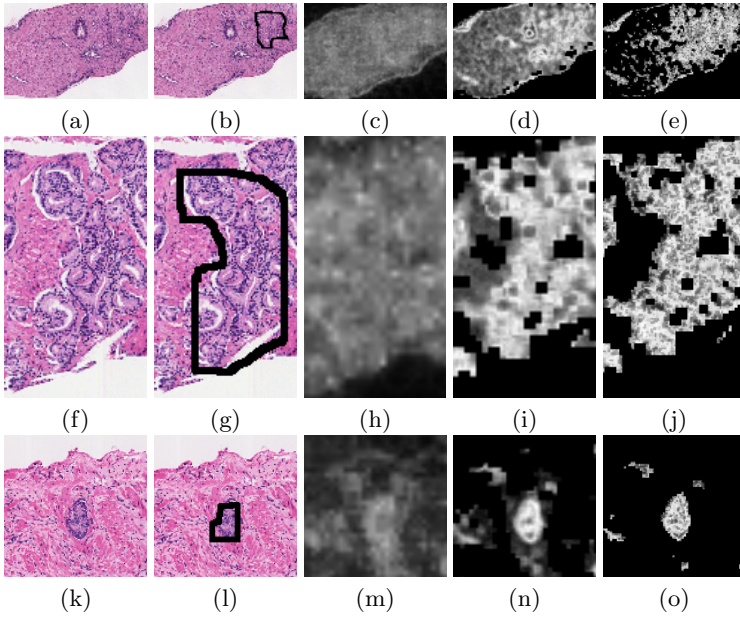
Fig. 5. (a), (f), (k) Digital histopathological prostate studies. (b), (g), (l) Tumor masks corresponding to the studies shown in (a), (f), and (k). Corresponding combined likelihood scenes at scale $j=1$ ((c), (h), (m)), $j=2$ ((d), (i), (n)), and $j=3$ ((e), (j), (o)). Note the increase in detection specificity at higher scales.
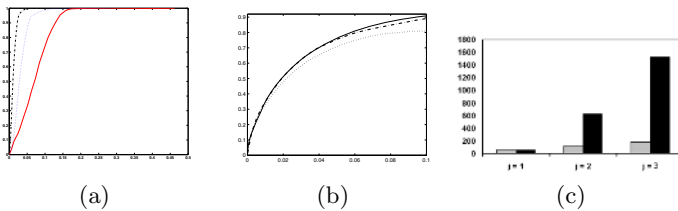
Fig. 6. (a) Average ROC curve for all 22 studies in our database at scales $j = 1$ (solid line), $j = 2$ (dotted line), and $j = 3$ (dot-dashed line). The increase in ROC area at high scales demonstrates the increase in CAD detection specificity at high image resolutions. (b) ROC curves obtained for a subset of testing images trained using 3 (dot-dashed line), 5 (dotted line), and 8 (solid line) images. The similarity of the 3 ROC curves indicates that CAD is robust to training. (c) Computation times (in minutes) for CAD at each image scale with (gray bar) and without (black bar) the Boosting Cascade.

prostate images at scale $j = 3$. Figures 5 (b), (g), and (l) show the corresponding ground truth for cancer (black contour). Figures 5 (c)-(e), (h)-(j), and (m)-(o) show the combined likelihood scenes for images shown in (a), (f), and (k) at scales $j = 1$ ((c), (h), (m)), $j = 2$ ((d), (i), (n)), and $j = 3$ ((e), (j), (o)). These images show that integration of additional discriminatory information at higher scales (higher resolution) increases the CAD detection specificity.

4   Conclusions and Future Work

In this work, we have presented a novel fully automated CAD system that integrates nearly 600 texture features extracted at multiple scales and orientations into a hierarchical multi-scale framework to automatically detect adenocarcinoma from prostate histology. To the best of our knowledge this work represents the first attempt to automatically analyze histopathology across multiple scales (similar to the approach employed by pathologists) as opposed to selecting an arbitrary image scale [3]-[5]. Further, the use of a multi-scale framework allows for efficient and accurate detection of prostatic adenocarcinoma. At the higher scales, our hierarchical classification scheme resulted in an 8-fold savings in computation time. Also, while CAD detection sensitivity was consistently high across image scales, detection specificity was found to increase at higher scales. While the CAD system was trained using only 3 images, the inclusion of additional training data did not significantly change CAD accuracy, indicating robustness to training. In future work, we intend to incorporate additional morphological and shape-based features at the finer scales and to quantitatively evaluate our CAD technique on a much larger database of prostate histopathological images.

References

1. Matlaga, B., Eskew, L., and McCullough, D.: Prostate Biopsy: Indications and Technique. The Journal of Urology, **169**:1 (2003) 12–19
2. Madabhushi, A., et al.: Automated detection of prostatic adenocarcinoma from high resolution ex-vivo MRI. IEEE Trans. on Med. Imaging, **24**:12 (2005) 1611–1625
3. Wetzel, A.W., et al.: Evaluation of prostate tumor grades by content based image retrieval. Proc. of SPIE Annual Meeting **3584** (1999) 244–252
4. Esgiar, A.N., et al.: Microscopic image analysis for quantitative measurement and feature identification of normal and cancerous colonic mucosa. IEEE Trans. on Information Tech. in Biomedicine **2**:3 (1998) 197–203
5. Tabesh, A., et al.: Automated prostate cancer diagnosis and Gleason grading of tissue microarrays. Proc. of the SPIE **5747** (2005) 58–70
6. Viola, P., and Jones, M.: Rapid object detection using a boosted cascade of simple features. IEEE Conf. Comp. Vision and Pattern Recog **1** (2001) 511–518
7. Freund, Y., and Schapire, R.: Experiments with a new boosting algorithm. Proc. of the Natural Conf. on Machine Learning (1996) 148–156
8. Adelson, E.H., and Burt, P.J.: Image data compression with the Laplacian pyramid. Proc. of Pattern Recog. and Inf. Proc. Conf. (1981) 218–223
9. Duda, R.O., and Hart, P.E.: Pattern Classification and Scene Analysis. Wiley (1973)
10. Haralick, R.M., Shanmugan, K., and Dinstein, I.: Textural features for image classification. IEEE Trans. on Systems, Man, and Cybernetics **SMC-3** (1973) 610–621
11. Manjunath, B.S., and Ma, W.Y.: Texture features for browsing and retrieval of image data. IEEE Trans. Pattern Anal. Machine Intell. **2** (1996) 837–842