33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Detection of Prostate Cancer on Histopathology using Color Fractals and Probabilistic Pairwise Markov Models

Elaine Yu, James P. Monaco, John Tomaszewski, Natalie Shih, Michael Feldman, and Anant Madabhushi

Abstract— **In this paper we present a system for detecting regions of carcinoma of the prostate (CaP) in H&E stained radical prostatectomy specimens using the color fractal dimension. Color textural information is known to be a valuable characteristic to distinguish CaP from benign tissue. In addition to color information, we know that cancer tends to form contiguous regions. Our system leverages the color staining information of histology as well as spatial dependencies. The color and textural information is first captured using color fractal dimension. To incorporate spatial dependencies, we combine the probability map constructed via color fractal dimension with a novel Markov prior called the Probabilistic Pairwise Markov Model (PPMM). To demonstrate the capability of this CaP detection system, we applied the algorithm to 27 radical prostatectomy specimens from 10 patients. A per pixel evaluation was conducted with ground truth provided by an expert pathologist using only the color fractal feature first, yielding an area under the receiver operator characteristic curve (AUC) curve of 0.790. In conjunction with a Markov prior, the resultant color fractal dimension + Markov random field (MRF) classifier yielded an AUC of 0.831.**

I. Introduction

The current screening protocol for CaP includes a digital rectal exam and prostate-specific antigen test. If these tests reveal abnormalities, a transrectal ultrasound guided biopsy is required to excise tissue for histological examination. If this histological analysis reveals the presence of cancer, a surgeon may perform a radical prostatectomy (RP), excising the entire gland. Following prostatectomy, the prostate is sliced into histological sections (typically whole or quarter sections). These histological sections (HS) are stained with hemotoxylin and eosin (H&E). The hematoxylin colors the nucleic material blue, while eosin stains the cytoplasmic structures and extracellular matrix red. The color and texture of these stains provide valuable information to distinguish cancer tissue from benign tissue; there exists higher variation in color in cancerous regions due to the dense packing of nulcei.

Though the grading of HSs is currently performed by clinicians, there are several advantages that have motivated the development of Computer-Aided Diagnosis (CAD) systems [1][2] for detection of prostate cancer on histological specimens. These advantages include: (1) CAD allows for analysis of the vast amount of the data present in HSs, a time-consuming task currently performed by pathologists, (2) extraction of quantitative and reproducible features can help reduce inter-expert variability in prostate cancer grading, and (3) CAD allows for speedy and more accurate diagnosis which can potentially influence the making of treatment related decisions.

Fractal dimension (FD) is a method widely used to capture intensity and texture information. In the past, the computation of FD of images has been defined for only binary and gray scale images. Recently, a color version has been proposed by Ivanovici [3] that captures information from the three different color channels.

The use of color fractal dimension (CFD) on HS enables the valuable color information to be exploited. Due to the known difference in color intensity and texture between cancer and benign regions on HS, the FD value obtained for pixels in the cancer region are different from pixels in the benign regions. Furthermore, we know that a cancer pixel is more likely to be adjacent to another cancer pixel. Therefore, instead of evaluating each pixel in isolation, we can consider the spatial dependencies among pixels to obtain a more accurate CaP detection result. This is done using a Markov Random Field (MRF). We use a novel type of Markov prior called the Probabilistic Pairwise Markov Model (PPMM) [1].

The CaP detection system in this paper performs classification of pixels using CFD. We incorporated new constraints in the CFD algorithm to calculate CFD on a per pixel basis. Having computed the feature on a per pixel basis, we determine the per pixel likelihood of cancer given the corresponding feature value. Contextual information is then incorporated with a Markov prior to yield a final classification of cancer or benign at each pixel.

The remainder of this paper is organized as follows: in Section II, we introduce previous applications of FD to medical image analysis and also describe the specific novel contributions of our methodology. In Section III, we provide an overview of the CFD algorithm, as well as our extension to this proposed algorithm. In Section IV, we introduce our comprehensive CaP detection system. This is followed by quantitative and qualitative evaluation of our integrated CFD and MRF classifier in Section V. Concluding remarks are presented in Section VII.

This work was made possible by grants by the Walter H. Coulter Foundation, National Cancer Institute (Grant Nos. R01CA136535, R01CA140772, and R03CA143991), Department of Defense Prostate Cancer Research Program, The Cancer Institute of New Jersey and the Society for Imaging Informatics in Medicine.

E. Yu, J. Monaco, and A. Madabhushi are with the Department of Biomedical Engineering, Rutgers University, USA. anantm@rci.rutgers.edu

J. Tomaszewski, N. Shih, and M. Feldman are with Department of Surgical Pathology, University of Pennsylvania, USA.

II. Previous Work and Novel Contribution

The idea of fractal geometry was introduced by Mandelbrot [4] to describe self-similar sets called fractals. Fractal-based analysis has been widely used for digital image analysis, including medical images and histology. Esgiar [5] has used fractal analysis in the detection of colonic cancer. In addition, Di leva [6] utilized fractal-based analysis to assess the angioarchitectural heterogeneity in human glioblastoma multiforme. Azemin [7] has also developed a methodology for fractal analysis of the retinal vasculature. However, FD algorithms employed in these studies are only defined for gray scale images.

In this paper we present a novel extension of the FD algorithm to account for color images and extend an existing CFD scheme (initially proposed by Ivanovici [3]). The specific novel contributions of this paper are as follows,

1) Extension of the CFD scheme in [3] to allow for CFD calculations on a per pixel basis with extension to hyper-rectangles instead of only hyper-cubes.
2) Leveraging the CFD features in conjunction with spatial constraints imposed via a Markov prior to build a classifier for detection of prostate cancer on histological sections.

III. Color Fractal Dimension

A. Color Fractal Dimension Algorithm by Ivanovici [3]

The CFD algorithm originally proposed by Ivanovici [3] extracts a single fractal dimension D for a given image. This algorithm considers a color image as a set S of points in 5-D space, where the five dimensions are the two spatial dimensions (x, y) and the three color dimensions (r, g, b). The arrangement of the points is characterized by the probability matrix $P(m, L)$, which is the probability of having m points included into a hyper-cube (or box) of size L centered at an arbitrary point $s \in S$. It is known that the fractal dimension D can be found using the following equation [3]:

$$N(L) = \sum_{m=1}^{N} \frac{1}{m} P(m, L) \propto L^{-D}, \tag{1}$$

where $N(L)$ is the number of pixels included in a box of size L.

Calculating the probability matrix $P(m, L)$ requires first counting at each point $s \in S$, the number of points in S that lie within the hyper-cube defined by:

$$|F^k - F^s| = \max(|F_i^k - F_i^s|) < L \in \forall i \in \{1, 2, 3, 4, 5\}, \tag{2}$$

where F_i^k is the value i^{th} dimension of point k in the 5-D space (x, y, r, g, b) and F^s is the center of the hyper-cube. Once the counts are determined for all L and $s \in S$, the matrix $P(m, L)$ can be found by enforcing the following normalization on the box counts: $\sum_{m=1}^{N} P(m, L) = 1, \forall L$.

To find the CFD, we do not directly solve Equation 1 for D. Instead, the typical method is to plot the negative of the logarithm of L against the logarithm of its corresponding $N(L)$ as shown if Figure 1. The slope of this regression line is D.

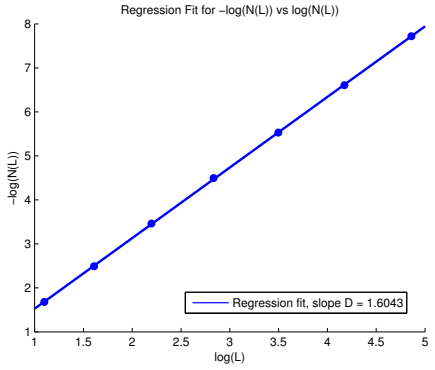


Fig. 1. Regression fit for negative logarithm of $N(L)$ against the logarithm of the corresponding box size L. The slope of this line is D, the color fractal dimension.

B. Extension of the CFD Algorithm

The CFD algorithm proposed by Ivanovici [3] calculates a single CFD value D for the entire image. In order to tailor algorithm for detection of CaP, we calculate the CFD on a per pixel basis. To do so, we calculate the CFD $D(s)$ of a $L_x \times L_y$ sub-image centered at the current pixel s. Instead of constructing a hyper-cube about the center pixel F^s, we construct a hyper-rectangle which has different bounding sizes L_x, L_y, L_R, L_G, and L_B for each color channel and spatial dimension. This modification is made to tailor the analysis to prostate histology. Through analysis of the red, blue and green channels in histology separately, the most suitable bounding size of the hyper-rectangle can be found for each channel.

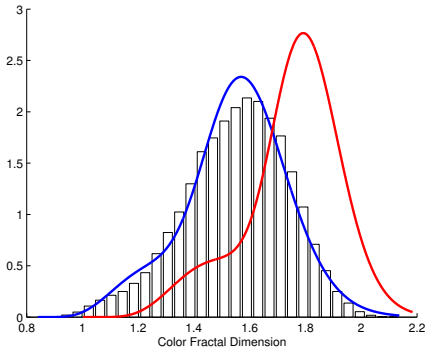


Fig. 2. Gamma fit of cancer pixel and benign pixel CFD. The blue line represents the CFD gamma fit for benign pixels, and the red line represents the CFD gamma fit for cancer pixels.

IV. Experimental Design

A. Algorithm Overview

Our CaP detection system comprises the following steps:

- Step 1: Calculate color fractal dimension on a per pixel basis
- Step 2: Bayesian Classification on a per pixel basis
- Step 3: Incorporate spatial constraints via Markov Random Field (MRF)

B. Bayesian Classification

A mixture of gamma distributions are used to model the probability density functions (PDF) for the CFD of cancer and benign pixels:

$$g(y;\theta,\mathbf{k},\alpha)=\alpha y^{k_1-1}\frac{e^{-y/\theta_1}}{\theta_1^{k_1}\Gamma(k_1)}+(1-\alpha)\,y^{k_2-1}\frac{e^{-y/\theta_2}}{\theta_2^{k_2}\Gamma(k_2)}, \tag{3}$$

where $y>0$ is the CFD, $\alpha\in[0,1]$ is the mixing parameter, $k_1,k_2>0$ are the shape parameters, $\theta_1,\theta_2>0$ are the scale parameters, and Γ is the Gamma function [8]. The PDFs generated for the cancer and benign classes are shown in Figure 2. The probability of each pixel being cancer given its CFD feature value is obtained via Bayes theorem.

C. Incorporating Spatial Dependencies with Markov Random Field (MRF)

Since cancer pixels are more likely to be next to cancer pixels, incorporating contextual spatial information is essential for accurate CaP detection. We include this contextual information by combining the conditional probabilities described in Equation 3 with a novel type of Markov prior called a probabilistic pairwise Markov model (PPMM) [1]. To determine the optimal labeling (classification), we perform maximum *a posteriori* estimation via iterated conditional modes (ICM). ICM (like all other MAP estimation schemes for MRFs) does not yield probability values, but instead produces a hard classification for each pixel. To vary the operating point (i.e. sensitivity/specificity) of the system we vary an internal threshold $T\in\{0,0.5,.1,\ldots,1\}$ as described in [9]. This yields a total of 21 different operating points.

D. Dataset Description

The dataset consists of 27 H&E stained histological sections from radical prostatectomies obtained from 10 patients at the University of Pennsylvania. All slides were digitized at 40x magnification (0.25 μm per pixel) with the following range of dimension values: $2000\leq x\leq 3000$ and $2000\leq y\leq 3000$. An expert pathologist delineated the cancerous regions as ground truth for classifier training and evaluation.

E. Implementation and Classifier Training

The following experiment was conducted to evaluate the CaP area detection capability of the algorithm. The CFD is calculated for all pixels across the 27 histological sections. Seven different values were used for the (x,y) size of the hyper-rectangle: $L_x,L_y\in\{3,5,9,17,33,65,129\}$. The red ($L_R$), green ($L_G$), and blue ($L_B$) channel bounding size were $13,23$ and 19, respectively.

Leave-one-out cross validation was used for classifier training and evaluation. First, the PDFs of CFD feature values for cancer and benign pixels in the training set were obtained. With these distributions, the probability of each pixel being cancer and benign is calculated for all the pixels in the slide. This procedure was repeated until all slides were classified. After classifying with Bayesian probability created via CFD information alone, we proceeded to incorporate spatial information with a Markov Prior to obtain a final classification of cancer or benign for each pixel.

V. Results and Discussion

A. Quantitative Results

The Bayesian probabilities calculated with CFD over all pixels were thresholded to produce a receiver operator characteristic (ROC) curve with an area under the curve (AUC) of 0.790. Once the Markov prior is introduced, the AUC improves to 0.831. Both ROC curves are shown in Figure 3. As can be seen, the ROC curve after incorporation of spatial dependencies has a higher AUC than CFD classification alone, indicating that including the spatial dependencies improves detection accuracy.

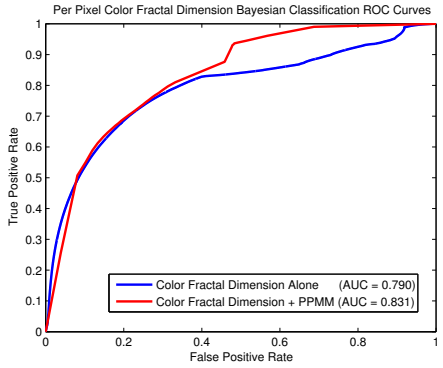


Fig. 3. ROC curves for classification of pixels with CFD alone and after adding the Markov prior, across 27 studies, using a leave one out validation strategy.

VI. Qualitative Results

To help visualize the algorithm behavior, we show the (1) Bayesian probability map obtained with CFD, (2) a cancer decision map obtained by thresholding the probability scene at the operating point on the ROC curve (best trade-off between sensitivity and specificity), and (3) a cancer decision map obtained by incorporating spatial dependencies to CFD with a Markov Prior, for two prostate histological slides. The histological sections with carcinoma region annotations and corresponding probability maps are shown in Figure 4. As can be observed, the cancer areas in the probability maps obtained via the CFD and CFD+MRF classifiers agree very well with the ground truth annotation for disease extent provided by the expert pathologist. Figures 4(d) and (h) also reveal that the spatial constraints imposed by the MRF allow for improved disease detection sensitivity and specificity compared to the CFD classifier alone.

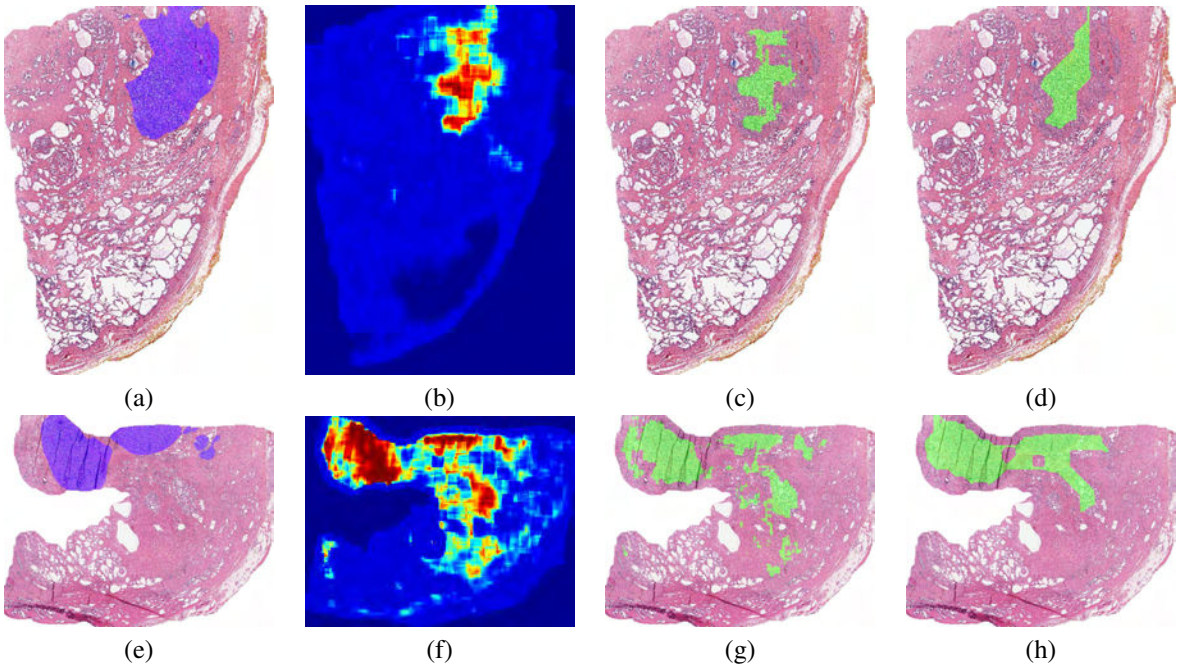
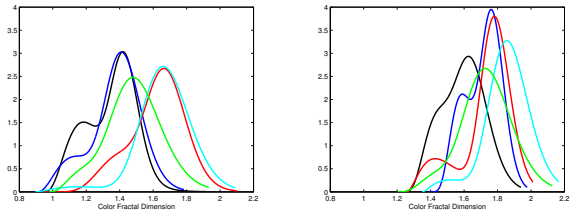


Fig. 4. (a), (e) Histological section with annotation of cancer area; (b), (f) Bayesian probability result using CFD; (c), (g) Final labeling result after thresholding the Bayesian probability at the optimum point; (d), (h) Final labeling result after incorporation of spatial information via Markov Prior.



(a) (b)

Fig. 5. (a) Benign and (b) Cancer pixel CFD feature value gamma distributions for 5 different studies. Each color represents a separate slide.

VII. Concluding Remarks

In this paper we present a novel methodology that exploits color fractal dimensions with spatial constraints imposed via a Markov random field to accurately detect prostate cancer on histological specimens. The novel methodological contribution of this work was to extend the CFD algorithm initially proposed in [3] to allow for the CFD computation on a per pixel basis and permitting its use in the context of hyper-rectangles, as opposed to only hyper-cubes. While the overall AUC of the integrated CFD+MRF classifier is already high (0.831), we believe we could achieve even higher accuracy if the intensity non-standardness issue (on account of color and illumination variability across slides) can be addressed. The impact of this non-standardness on the CFD is shown in figure 5 which illustrates PDFs of the CFD for 5 different images. Another direction of future work is to fuse the existing classifier with the gland based cancer detection classifier presented by Monaco et al. [1].

References

[1] J. Monaco, J. Tomaszewski, M. Feldman, I. Hagemann, M. Moradi, P Mousavi, A. Boag, C. Davidson, P. Abolmaesumi, and A. Madabhushi, "High-throughput detection of prostate cancer in histological sections using probabilistic pairwise markov models," *Medical Image Analysis*, vol. 14, no. 4, pp. 617 – 629, 2010.

[2] S. Doyle, J. Tomaszewski, M. Feldman, and A. Madabhushi, "Hierarchical boosted bayesian ensemble for prostate cancer detection from digitized histopathology," *IEEE Transactions on Biomedical Engineering,*, [Epub ahead of print], 2010 (PMID: 20570758).

[3] M. Ivanovici and N. Richard, "Fractal dimension of color fractal images," *Image Processing, IEEE Transactions on*, vol. 20, no. 1, pp. 227 –235, 2011.

[4] B. B. Mandelbrot, *The fractal Geometry of Nature*, Freeman, 1982.

[5] A.N. Esgiar, R.N.G. Naguib, B.S. Sharif, M.K. Bennett, and A. Murray, "Fractal analysis in the detection of colonic cancer images," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 6, no. 1, pp. 54 –58, 2002.

[6] A. Di Ieva, F. Grizzi, C. Sherif, C. Matula, and M. Tschabitscher, "Angioarchitectural heterogeneity in human glioblastoma multiforme: A fractal-based histopathological assessment," *Microvascular Research*, vol. In Press, Corrected Proof, doi: DOI: 10.1016/j.mvr.2010.12.006.

[7] M. Z. C. Azemin, D. K. Kumar, T. Y. Wong, R. Kawasaki, P. Mitchell, and J. J. Wang, "Robust methodology for fractal analysis of the retinal vasculature," *Medical Imaging, IEEE Transactions on*, vol. 30, no. 2, pp. 243 –250, 2011.

[8] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, Inc., New York, NY, 1965.

[9] J. Monaco, S. Viswanath, and A Madabhushi, "Weighted iterated conditional modes for random fields: Application to prostate cancer detection," *Workshop on Probabalistic Models for Medical Image Analysis (PMMIA) (in conjunction with MICCAI)*, 2009.