Automatic Classification for Pathological Prostate Images Based on Fractal Analysis

Po-Whei Huang* and Cheng-Hsiung Lee

***Abstract*—Accurate grading for prostatic carcinoma in pathological images is important to prognosis and treatment planning. Since human grading is always time-consuming and subjective, this paper presents a computer-aided system to automatically grade pathological images according to Gleason grading system which is the most widespread method for histological grading of prostate tissues. We proposed two feature extraction methods based on fractal dimension to analyze variations of intensity and texture complexity in regions of interest. Each image can be classified into an appropriate grade by using Bayesian, k-NN, and support vector machine (SVM) classifiers, respectively. Leave-one-out and k-fold cross-validation procedures were used to estimate the correct classification rates (CCR). Experimental results show that 91.2%, 93.7%, and 93.7% CCR can be achieved by Bayesian, k-NN, and SVM classifiers, respectively, for a set of 205 pathological prostate images. If our fractal-based feature set is optimized by the sequential floating forward selection method, the CCR can be promoted up to 94.6%, 94.2%, and 94.6%, respectively, using each of the above three classifiers. Experimental results also show that our feature set is better than the feature sets extracted from multiwavelets, Gabor filters, and gray-level co-occurrence matrix methods because it has a much smaller size and still keeps the most powerful discriminating capability in grading prostate images.**

***Index Terms*—Classification, fractal dimension, Gleason grading, prostatic carcinoma, prostate image.**

I. Introduction

PROSTATE carcinoma becomes the most common cancer in men over the last few years. In the U.S., prostate cancer is the most frequently diagnosed cancer and ranks second among cancer deaths in men [1]. In the U.K., more than 32 000 men were diagnosed to have the disease of prostate cancer and approximately 10 000 died from that disease [2]. Biopsy of the prostate, usually stained by Hematoxylin and Eosin (H&E) technique, is a key step for confirming the diagnosis of malignancy and guiding treatment [3]. By viewing the microscopic images of biopsy specimens, pathologists can determine the histological grades. The most widespread method for histological grading of prostate is the Gleason grading system [4].

Manuscript received August 07, 2008; revised December 02, 2008. First published January 19, 2009; current version published June 24, 2009. This research was supported in part by NSC 96-2628-E-005-084-MY3 and in part by the Ministry of Education, Taiwan. *Asterisk indicates corresponding author.*

*P. W. Huang is with the Department of Computer Science and Engineering, National Chung Hsing University, Taichung 40227, Taiwan (e-mail: powhei.huang@msa.hinet.net).

Cheng-Hsiung Lee is with the Department of Computer Science and Engineering, National Chung Hsing University, Taichung 40227, Taiwan.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMI.2009.2012704

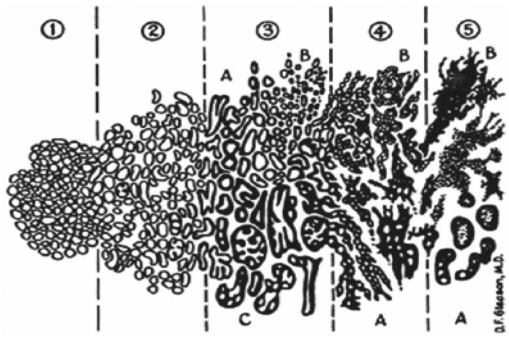

Fig. 1. Gleason grading diagram.

The Gleason grading system, developed by Gleason during the 1970s, is the most powerful tool for determining the long-term prognosis of prostate cancer cases. In this grading system, the prostate cancer can be classified into five tumor grades represented by a number ranging from 1 to 5 with 5 being the worst grade possible [5]. A classic Gleason grading diagram containing the five basic tissue patterns associated with the five tumor grades is shown in Fig. 1. This diagram shows the continuum of deteriorating cancer cell architecture and the four dividing lines along this continuum to identify patients with significantly different prognosis. Gleason grading is based upon the degree of loss of the normal glandular tissue architecture [5]. The degree of tumor resemblance to normal gland architecture is called differentiation. Fig. 2 shows four pathological images of prostatic carcinoma from well differentiated (grade 2) to very poorly differentiated (grade 5) in our image set. As reported in [6] and [7], the use of texture analysis for prostatic lesions is very essential to the identification of tissue composition in prostatic neoplasia. From Figs. 1 and 2, we can also see that the texture of prostate tissue plays an important role in Gleason grading for prostate cancer.

Although pathologists know how aggressive the cancer is likely to be and how quickly it may spread from the result of a Gleason grade, human visual grading is time-consuming and very subjective due to inter- and intra-observer variations. At present, most diagnosis of cancer is still done by visual examination of radiological images, microscopy of biopsy specimens, direct observation of tissues, and so on. These views are typically interpreted in a qualitative manner by clinicians trained to classify abnormal features such as structural irregularities or high indices of mitosis. A more quantitative and reproducible approach for analyzing images is highly desired. Therefore, how to develop a more objective computer-aided technique for automatically and correctly grading prostatic carcinoma is the goal of this research study.

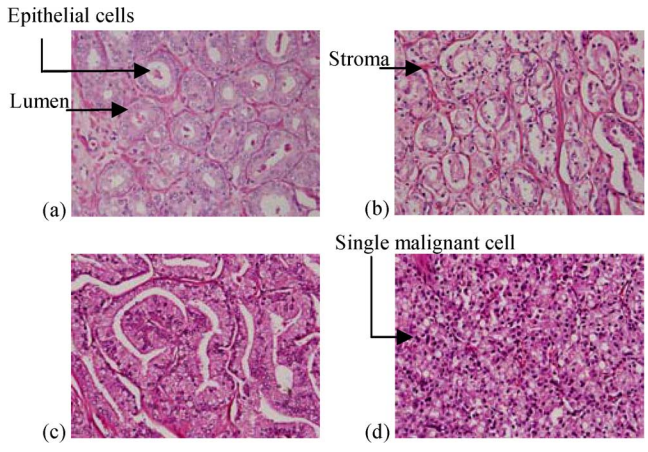

Fig. 2. Prostate images of different cancer grades. (a) Gleason grade 2. (b) Gleason grade 3. (c) Gleason grade 4. (d) Gleason grade 5.

To achieve this goal, we propose an automated system for grading prostatic carcinoma using novel texture analysis techniques to analyze the histological prostate images. For grading prostate cancer accurately, the texture feature must be powerful enough to differentiate the five basic tissue patterns as shown in Fig. 1. Many well-known techniques for texture analysis such as extracting useful texture features from gray-level co-occurrence matrix (GLCM), Gabor filters, and multiwavelet transforms have been exploited. Applications of using these techniques for medical image analysis have been reported in the literature [7]–[11].

A different approach recently received more attention is the use of fractal dimension (FD) to describe the texture feature. This approach has been found to be very effective in texture analysis and segmentation of images [12]. Recent studies have also shown that fractal geometry is useful for describing the pathological architecture of tumors and yielding insights into the mechanisms of tumor growth [13]. Cancer is often characterized as a chaotic, poorly regulated growth. By focusing on the irregularity of tumor growth, fractal geometry is well suited to quantify those morphological characteristics that pathologists have long used in a qualitative sense to describe malignancies [13]. As a consequence, it becomes quite natural and very reasonable to consider fractal dimension as a morphometric measure of irregular structures typical of tumor growth. The use of fractal analysis in pathology can be found in the literature [14]–[20].

Since Gleason grading is mainly based on texture properties of images, the concept of fractal dimension is applied in this paper for analyzing the texture of prostate tissue. In physical phenomena, the growth of cancer shows the features of fractal. Thus, the fractal theory can provide clinically useful information for discriminating pathological tissue from healthy tissue [13].

Based on the above observations and discussions, we proposed two fractal dimension texture features that can be extracted through *differential box-counting* (DBC) method [21] and our *entropy-based fractal dimension estimation* method and then combine them together as a FD-based feature set to analyze pathological images of prostatic carcinoma. To evaluate the effectiveness of our FD-based feature set, we applied both leave-one-out (LOO) and k-fold cross-validation procedures [22] to a set of 205 pathological prostate images and tested against these samples using Bayesian, k-nearest neighbor (k-NN), and support vector machine (SVM) classifiers, respectively, to estimate the correct classification rates (CCR). As compared to other features derived from multi-wavelet transforms, Gabor filters, and GLCM, our experimental results show that the FD-based feature set proposed in this paper has better discriminating capability in terms of higher CCR no matter which classifier is used. To further optimize the above mentioned feature sets by reducing their dimensionalities, we applied the sequential floating forward selection (SFFS) method [23] for selecting an optimal set of features. The experimental results show that our FD-based feature set has a smaller dimensionality than other feature sets derived from multiwavelets, Gabor filters, and GLCM while having better performance in terms of CCR. In particular, the CCR of our approach is 94.6% using Bayesian classifier, 94.2% using k-NN classifier, and 94.6% using SVM classifier.

This paper is organized as follows. Related research work in recent years is described in Section II. Our feature extraction methods are described in Section III. The classification and feature selection techniques, as well as the experimental results are presented in Sections IV and V, respectively. Finally, concluding remarks are given in the last section.

II. Related Work

A. Fractal Geometry

The term fractal comes from the Latin word "fractus" which means irregular fragments. The concept of fractal geometry was originally introduced by mathematician Mandelbrot [24]. According to Mandelbrot's concept, many natural objects exhibit the fractal property of self-similarity. Based on this concept, a fractal can be viewed as an object made of sub-objects similar to the whole, exactly or statistically, in some way. The property of self-similarity occurs over an infinite range of scales in pure mathematical fractal structures such as Koch's curve, Cantor set, or Sierpinski triangle, but occurs over a finite range in many natural objects such as coastlines and snowflakes. The complexity of such forms can not be described by the classical Euclidean geometry, largely because it is limited to considering structures in terms of their topological dimension which is always an integer. Mandelbrot successfully calculated the coastline of Britain by plotting the measured perimeter at different scales on a log-log graph to come-up with a straight line and then to take the slope of that line as the fractal dimension [25]. He found the fractal dimension of the coastline was 1.25. Pentland [26] presented the evidence that most natural surfaces are spatially isotropic fractals and the intensity images of these surfaces are also fractals. Since then, fractal analysis has been successfully applied to the field of digital image processing [12], [27] and applications of medical image analysis [13]–[20]. In particular, fractals can be used as morphometric tools to approximate cancer for diagnostic and prognostic purposes and a family of fractal-producing mathematical models known as statistical

growth processes can mimic tumor growth as pointed out by Baish and Jain [13].

The concept of self-similarity can be used to estimate the fractal dimension as follows. Given a bounded set S in Euclidean n-space, S is self-similar if it is the union of N_r distinct (non-overlapping) copies of itself scaled down by a ratio r. The fractal dimension D of S is given by the relation $1 = N_r r^D$ or calculated by the following equation [12]:

$$D = \frac{\log(N_r)}{\log(1/r)}. \tag{1}$$

However, natural objects do not exhibit deterministic self-similarity because usually they are classified as random fractals, meaning that each of their smaller parts is statistically similar to the whole. Therefore, a part of an object becomes statistically identical to the original object if the original one is scaled down by a ratio r in all dimensions such that (1) is satisfied.

B. Analysis for Pathological Prostate Images

Several methods have been proposed for analyzing pathological prostate images during the last few years. These methods can be categorized into two main approaches. The first approach focused on the identification of the normal and abnormal tissue composition. Roula *et al.* [28] proposed an automated system to classify pathological images into four classes including stroma (muscular normal tissue), benign prostatic hyperplasia (BPH, a benign condition), prostatic intraepithelial neoplasia (PIN, a precursor state for cancer), and prostatic carcinoma (PCa, abnormal tissue development corresponding to cancer) for the purpose of prostate cancer diagnosis. The images are captured in multiple spectral bands. Six texture features and two structural features were extracted from the image captured in each channel. To reduce dimensionality of feature space, principal component analysis (PCA) was applied to select dominant features. These features were subsequently used by a supervised classical linear discrimination method to classify pathological images of prostate tissues. Diamond *et al.* [6] proposed a machine vision system to automatically identify tissue composition in prostatic neoplasia. In their method, texture and morphological characteristics were used to distinguish stroma, normal tissue (nonneoplastic epithelial component), or prostatic carcinoma regions in the image.

The second approach focused on automatic Gleason grading for prostatic carcinoma. Stotzka *et al.* [29] extracted statistical and structural features from the spatial distribution of epithelial nuclei over the image area. A hybrid neural network/Gaussian statistical classifier was used to distinguish moderately and poorly differentiated lesions of the prostate. The authors did not provide any algorithm for segmenting the epithelial nuclei. It seems that this stage of operation was still performed manually. Wetzel *et al.* [30] proposed a content-based image retrieval system to assist pathology diagnosis by searching for similar prostate tumor samples from the image database to evaluate the prostate tumor grade for the current sample. They used prostate tumor samples that have been graded by the Gleason system and stored in the database as an initial domain for analogical matching. Since the system made no attempt to directly reproduce pathologists' visual analysis, no quantitative result for the performance of this system was reported. Smith *et al.* [31] used power spectrum to represent the texture characteristics of tissue images and applied principal component analysis to reduce dimensionality of feature space. The nearest-neighbor (NN) classifier was used to classify the images. Their method can only discriminate grades 1, 2, and 3, but can not discriminate grades 4 and 5.

Jafari–Khouzani *et al.* [7] proposed a method for grading the pathological images of prostate biopsy samples by using energy and entropy features calculated from multiwavelet coefficients of an image. These multiwavelet features were used by the k-nearest-neighbor classifier for classification and leave-one-out procedure was applied to estimate the error rate. Their image set consists of 100 prostate images with grades from 2 to 5. Although their method had good classification performance and could discriminate various grades for prostatic carcinoma, it has the problem that a large number of features are produced in the feature extraction stage and a lot of computing time is consumed in the classification stage. For instance, if a multiwavelet transform with second level of decomposition is performed, there will be 28 submatrices to be generated. Each submatrix of multiwavelet coefficients is used to compute energy and entropy features, respectively. Thus, their method will produce 56 (28×2) multiwavelet features to represent texture information for a prostate image.

Tabesh *et al.* [32]–[34] proposed an automatic two-stage system for prostate cancer diagnosis and Gleason grading. The color, morphometric, and texture features are extracted from prostate tissue images in their system. Then, linear and quadratic Gaussian classifiers were used to classify images into tumor/nontumor classes, and further into low/high grades for cancer images. Although their method can discriminate low/high grades, it cannot discriminate grade 2 from grade 3 or grade 4 from grade 5 in Gleason grading system.

III. Feature Extraction

A. Extracting Feature by Differential Box-Counting

There exist several approaches to estimate the fractal dimension (FD) of an image. Peleg *et al.* [35] used the ε-blanket method, which is a 2-D generalization of the original approach suggested by Mandelbrot. Pentland [26] considered the image intensity surface as fractal Brownian function (fBf) and estimated FD from Fourier power spectrum of fBf. Gangepain and Roques-Carmes [36] and Keller *et al.* [37] used variations of box-counting approach to estimate FD. The differential box-counting (DBC) method proposed by Sarkar and Chaudhuri [21] is probably most commonly used because their method is computationally efficient and can cover a wide dynamic range. A complete comparison of the DBC method with the methods mentioned above was reported in [21]. Since the DBC method is superior to other methods in terms of computational efficiency and giving a better approximation for an image's intensity surface, we adopt the DBC method to calculate the intensity surface fractal features for pathological images of prostate.

In general, the images with high-grade prostatic carcinoma have sharp gray-level variations in neighboring pixels due to the scenario that dark malignant cells invade stroma. Thus, the DBC

method can better distinguish low- and high-grade prostatic carcinoma by measuring the variations of intensity in local regions. Before applying the DBC method, the color pathological images of prostatic tissues are transformed to gray-level images by getting the R channel from the RGB color space for enhancing the contrast between malignant cells and background tissues. In the above preprocessing step, the malignant cells will become darker because they are stained as blue. Other pathological objects such as stroma and lumens are stained as red or do not get stained in H&E-stained pathological images. The DBC method is described as below.

Consider an image of size $M \times M$ pixels that has been scaled down to a size $s \times s$, where $1 < s \leq M/2$ and s is an integer. Then, we can get the scale ratio $r = s/M$. Consider the image as a 3-D space such that (x, y) represents a 2-D position and the third coordinate (z) represents the gray level of an image at position (x, y). The (x, y)-space is divided into grids of size $s \times s$. Thus, there will be a column of boxes of size $s \times s \times h$ on each grid, where $\lfloor G/h \rfloor = \lfloor M/s \rfloor$ and G is the total number of gray levels in an image. Let the maximum and minimum gray levels of an image in the (i, j)th grid fall in box number k and l, respectively. The contribution of N_r in the (i, j)th grid is expressed as follows:

$$n_r(i, j) = k - l + 1. \tag{2}$$

The contribution from all grids is

$$N_r = \sum_{i,j} n_r(i, j). \tag{3}$$

N_r is counted for different scale ratio r. Then, the fractal dimension D can be estimated from the slope of line approximated by least-squares linear fitting for $\log(N_r)$ versus $\log(1/r)$ in (1).

B. Extracting Feature by Entropy-Based Fractal Dimension Estimation

To analyze the texture complexity in pathological images for different Gleason grades of prostate carcinoma, this paper further proposes an entropy-based method for estimating the fractal dimension of an image. Based on our observation, the DBC method only captures the information about intensity difference which is necessary but not sufficient enough to differentiate all patterns of different Gleason grades. Our entropy-based fractal dimension estimation (EBFDE) method can further capture the information about randomness of pixels. Although the FD-features extracted by DBC method and our EBFDE method are not totally independent of each other, they are partially complementary. This conjecture is verified by our experiment to be presented in Section V, which shows that accuracy of classification can be promoted when we combine these two types of features together.

The EBFDE method is described as follows. First, a 2-D image is partitioned into several grids of size $s \times s$. Then, we compute the entropy for the (i, j)th grid using the following equation:

$$e_r(i, j) = -\sum_{k=0}^{G-1} p_k \log_2(p_k). \tag{4}$$

In this equation, index k is taken over all grayscales in the (i, j)th grid of an image, p_k is the probability of gray-level k occurring in the (i, j)th grid of an image, and G is the total number of gray levels. The contribution from the (i, j)th grid is $e_r(i, j)^2$. So the total contribution from all grids is

$$E_r = \sum_{i,j} e_r(i, j)^2. \tag{5}$$

Again, by applying (1), the fractal dimension D of an image can be estimated using least-squares linear fitting for $\log(E_r)$ versus $\log(1/r)$.

C. Combination of Two Fractal Dimension Texture Features

Voss [38] have showed that different textures may have the same fractal dimension. Therefore, using one single fractal dimension texture feature to discriminate different textures is inadequate. Moreover, natural objects usually exhibit the property of random fractals; in other words, they may not be self-similar over all scales. Based on this concept, we assume that various self-similarity properties in a prostatic carcinoma (PCa) image may be reflected in different individual ranges of scales. Remember that the scaled down ratio is $r = s/M$, where s^2 is the grid size and M^2 is the image size. Since $M = 384$ for prostate images, we choose $s = 2, 4, 8, 16, 32, 64$, and 128 to include all feasible grid sizes, from 2×2 (the smallest one) to 128×128 (one ninth of the whole image). Therefore, the range of scales (r) is $\{1/192, 1/96, 1/48, 1/24, 1/12, 1/6, 1/3\}$, which is subsequently divided into three subranges: the subrange of small scales $\{1/192, 1/96, 1/48\}$, the subrange of medium scales $\{1/48, 1/24, 1/12\}$, and the subrange of large scales $\{1/12, 1/6, 1/3\}$. Here, we allow a small portion of overlapping between two neighboring sub-ranges because there is no clear cut between two subranges reflecting different self-similarity properties. We choose three scales in each subrange because this is the minimum requirement for using the technique of least square linear fit. Since we do not exclude the possibility that the same self-similarity property is reflected in all scales, we also use all of the seven scales to estimate the fractal dimension of an image. As a result, four fractal dimension texture features can be obtained by DBC method and another four fractal dimension texture features can be obtained by our EBFDE method. Then, we combine these eight features to become a feature set $\{f_{D1}, f_{D2}, f_{D3}, f_{D4}, f_{E1}, f_{E2}, f_{E3}, f_{E4}\}$ as follows.

- f_{D1} is the FD calculated from grids of size $s^2 (s = 2, 4, 8)$ using DBC method.
- f_{D2} is the FD calculated from grids of size $s^2 (s = 8, 16, 32)$ using DBC method.
- f_{D3} is the FD calculated from grids of size $s^2 (s = 32, 64, 128)$ using DBC method.
- f_{D4} is the FD calculated from grids of size $s^2 (s = 2, 4, 8, 16, 32, 64, 128)$ using DBC method.
- f_{E1} is the FD calculated from the grids of size $s^2 (s = 2, 4, 8)$ using EBFDE method.
- f_{E2} is the FD calculated from grids of size $s^2 (s = 8, 16, 32)$ using EBFDE method.
- f_{E3} is the FD calculated from grids of size $s^2 (s = 32, 64, 128)$ using EBFDE method.

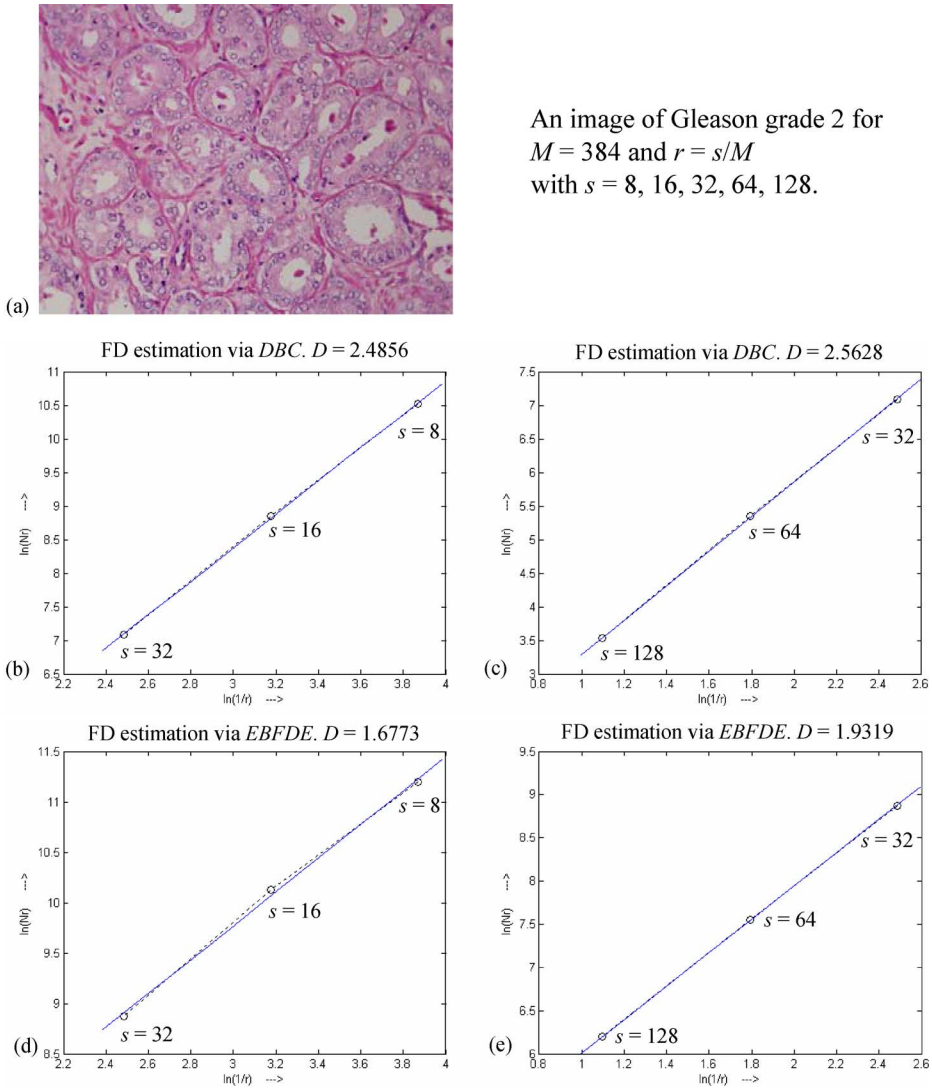

Fig. 3. (a) Original image. (b) Fractal dimension estimated by DBC method on grids of size $s = 8, 16, 32$. (c) Fractal dimension estimated by DBC method on grids of size $s = 32, 64, 128$. (d) Fractal dimension estimated by EBFDE method on grids of size $s = 8, 16, 32$. (e) Fractal dimension estimated by EBFDE method on grids of size $s = 32, 64, 128$.

- f_{E4} is the FD calculated from grids of size $s^2 (s = 2, 4, 8, 16, 32, 64, 128)$ using EBFDE method.

D. Example

Here, we use an example to demonstrate how the fractal-based features can be obtained from a pathological prostate image. Given an image of size 512×384 pixels, as shown in Fig. 3(a), the scale ratio r will be 1/48, 1/24, 1/12, 1/6, and 1/3 when the image is partitioned into grids with $s = 8, 16, 32, 64$, and 128, respectively. Then, we apply (2) to each grid of the image and sum up the results obtained from all grids as N_r. The three dots from left to right in Fig. 3(b) represent the $\log(N_r)$ versus $\log(1/r)$ plot for grids with $s = 32, 16$, and 8, respectively. The slope of the least-squares fitting line for these three dots is 2.4856, which is defined as f_{D2}. Similarly, Fig. 3(c) is the least-squares fitting line for the plot created by the grids with $s = 128, 64$, and 32. Therefore, the slope of this line is $f_{D3} = 2.5628$. Now we apply (4) to each grid of the image and sum up the squares of the results obtained from all grids as E_r. The least-squares fitting line for the plot created by the grids with $s = 32, 16$, and 8 is shown in Fig. 3(d). The least-squares fitting line for the plot created by the grids with $s = 128, 64$, and 32 is shown in Fig. 3(e). Thus, $f_{E2} = 1.6773$ which is the slope of the line in Fig. 3(d), and $f_{E3} = 1.9319$ which is the slope of the line in Fig. 3(e). All other fractal dimension texture features can be calculated in a similar way.

IV. Classification and Feature Selection

In this section, we first present the three classifiers, Bayesian, k-NN, and SVM used to test against the feature sets described in the previous section for classification of pathological images of prostate carcinoma. Then, we present the leave-one-out and k-fold cross-validation procedures adopted in this paper to estimate accuracy of classification. Finally, we describe the feature selection method used in this paper for choosing optimal subsets of features.

A. Classification Methods

The first classification technique used in this paper for automatic Gleason grading is Bayesian classifier. According to the Bayesian decision rule of parametric classifier, an observation is classified to a specific class that has the highest posteriori probability among all other classes. Suppose we know the prior probability $P(c_i)$ and conditional probability density $p(\mathbf{x}|c_i)$ of class c_i, the Bayes formula in (6) shows how to use the observed data $\mathbf{x}$ and prior probability $P(c_i)$ to obtain *a posteriori* probability $P(c_i|\mathbf{x})$ [39]

$$P(c_i|\mathbf{x}) = \frac{p(\mathbf{x}|c_i)P(c_i)}{\sum_{i=1}^{C} p(\mathbf{x}|c_i)P(c_i)} \tag{6}$$

where $\mathbf{x}$ is the feature vector extracted from a prostate image by using specific texture analysis method in this paper, C is the total number of classes, and c_i represent the class of prostatic carcinoma of different grade. The decision function can be defined as

$$g_i(\mathbf{x}) = \ln(p(\mathbf{x}|c_i)) + \ln(P(c_i)) \tag{7}$$

where ln denotes natural logarithm. This paper assumed that $\mathbf{x}$ has a multivariate normal distribution with mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$. Based on the above Gaussian assumption, we can easily obtain the following decision function [22], [39]:

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - \frac{1}{2}\ln|\boldsymbol{\Sigma}_i| + \ln P(c_i) \tag{8}$$

where $|\boldsymbol{\Sigma}_i|$ and $\boldsymbol{\Sigma}_i^{-1}$ are the determinant and inverse of the covariance matrix for class c_i, respectively. If $g_i(\mathbf{x}) > g_j(\mathbf{x})$ for all $i \neq j$, then feature vector $\mathbf{x}$ will be classified to class c_i with $i, j \in \{1, 2, \ldots, C\}$.

The second classifier used in this paper for automatic Gleason grading is k-NN which is well-known among all nonparametric classifiers. The k-nearest-neighbor decision rule classifies an observation by assigning it the label which is most frequently represented among the k nearest neighbors. A decision is made by examining all the labels on the k nearest neighbors and taking a vote. The operation of a k-NN classifier can be summarized by the following basic steps [40].

1) Compute the distances between the new sample and all previous samples already classified into clusters.
2) Sort the distances in increasing order and select k samples with the smallest distance values.
3) Apply the voting principle: a new sample is added (classified) to the largest cluster out of k selected samples.

The third classification technique used in this paper for grading carcinoma prostate images is the SVM method [41], [42]. Compared with traditional classification methods which minimize the empirical training error, the goal of SVM is to minimize the upper bound of the generalization error by finding the largest margin between the separating hyperplane and the data. The theory of nonlinear SVM is briefly described as follows.

Consider a training set of N samples in binary classification. Each sample is denoted by a tuple $(\mathbf{x}_i, y_i)$, where $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{id})^T$ corresponds to the feature vector for the ith sample $(i = 1, 2, \ldots, N)$ in d-dimensional space and $y_i \in \{-1, 1\}$ denotes its two-class label. Any point $\mathbf{x}$ on the hyperplane must satisfy the decision boundary $\mathbf{w} \cdot \mathbf{x} + b = 0$, where parameter $\mathbf{w}$ is normal to the hyperplane. In practicality, a nonlinear SVM is more widely used for the general case due to its nonlinear mechanism that can effectively classify data which are nonseparable by a linear SVM. A nonlinear SVM can be formulated by the following optimization problem:

$$\min_{\mathbf{w},b,\xi} \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{N}\xi_i$$
$$\text{subject to } y_i(\mathbf{w}\cdot\Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \xi_i \geq 0, \quad i = 1, 2, \ldots, N \tag{9}$$

where notation $||\cdot||$ represents the norm of a vector. In the above objective function, a training data $\mathbf{x}_i$ is mapped to a higher dimensional space by a kernel function Φ, and penalty C is a user-specified parameter. By minimizing (1/2) $||\mathbf{w}||^2$, we can get the maximum margin between the separating hyperplane and the data. To reduce the number of training errors in linearly non-separable case, the penalty term $C\sum_{i=1}^{N}\xi_i$ consists of a number of positive-valued slack variables ξ_i which can be used to construct a soft margin hyperplane. Then, a test sample $\mathbf{z}$ can be classified according to the following equation:

$$f(\mathbf{z}) = \text{sign}(\mathbf{w}\cdot\Phi(\mathbf{z}) + b) = \text{sign}\left(\sum_{i=1}^{N}\alpha_i y_i \Phi(\mathbf{x}_i)\cdot\Phi(\mathbf{z}) + b\right). \tag{10}$$

The test sample $\mathbf{z}$ is classified to positive class if $f(\mathbf{z}) = +1$, and is classified to negative class if $f(\mathbf{z}) = -1$. In the above decision boundary equation, the parameters α_i are Lagrange multipliers which can be obtained by using quadratic programming [43].

The computation of $\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$, i.e., the dot product between pairs of vectors in the transformed space, is quite cumbersome and may suffer from the curse of dimensionality problem [43]. Since a kernel function K can be expressed as $K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)\cdot\Phi(\mathbf{x}_j)$ according to the Mercer's theorem [43], the decision function of a nonlinear SVM can be written as

$$f(\mathbf{z}) = \text{sign}\left(\sum_{i=1}^{N}\alpha_i y_i K(\mathbf{x}_i, \mathbf{z}) + b\right). \tag{11}$$

In this study, we train all data sets with the radial basis function kernel $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma||\mathbf{x}_i - \mathbf{x}_j||^2/2\sigma^2), \gamma > 0$. Also notice that the above description of SVM is only for binary classification. A number of methods have been proposed to extend binary SVM to solve multiclass problems [44], [45]. They essentially separate L mutually exclusive classes by solving many two-class problems and combining their predictions in various ways. Two well-known extensions are one-against-one method [45], [46] and one-against-rest method [44], [47]. The one-against-one method uses $L(L-1)/2$ binary SVM classifiers for L classes, and each of them provides a partial decision for classifying a test sample. During the test, each of the $L(L-1)/2$ classifiers votes for one class. The

winning class is the one with the largest number of accumulated votes. In case two classes have the same number of votes, we simply choose one randomly. The one-against-rest method compares a given class with the rest of all other classes by constructing L hyperplanes where each hyperplane separates one class from the other classes. Thus, it generates L decision functions so that an observed data can be mapped to a class with the largest decision function value. It has been shown that the one-against-one method is more suitable for practical use than the one-against-rest method [44]. In this study, we use the one-against-one multiclass classification method based on LIBSVM [45], [48] by combining all pair-wise comparisons of binary SVM classifiers.

B. Estimation for Accuracy of Classification

In order to estimate the effectiveness of FD-based feature set proposed in this paper, we adopt the definition of correct classification rate (CCR) provided by Lee *et al.* [18]

$$\mathrm{CCR} = \sum_{i=1}^{C} P(c_i)\frac{n_i}{N_i} \tag{12}$$

where n_i is the number of samples correctly classified to the ith class via the classifier, N_i is the total number of samples in the ith class, $P(c_i)$ is the prior probability that an observed data falls in class c_i.

Given a small set of samples, appropriate strategies for learning and testing become very critical to avoid over-fitting. Leave-one-out (LOO) and k-fold cross-validation are two popular error estimation procedures to reduce bias in machine learning and testing when sample size is small [22]. The procedure of LOO method is to take one out of n observations and use the remaining $n-1$ observations as the training set for deriving the parameters of the classifier. The classifier is then used to classify the removed observation. This process is repeated for all n observations to obtain the estimation for accuracy of classification. As to k-fold cross-validation method, the entire sample set is randomly partitioned into k disjoint subsets of equal size, where n is the total number of samples in the entire set. Then, $k-1$ subsets are used to train the classifier and the remaining subset is used to test for accuracy estimation. This process is repeated for all distinct choices of k subsets and the average of correct classification rates is calculated. Notice that k-fold cross-validation is reduced to LOO if $k = n$. Many examples of using all of their samples to evaluate the classifiers based on LOO or k-fold cross-validation procedures can be found in [7], [28], [32], [49]. In this paper, both methods are used to estimate accuracy of classification for the sake of completeness.

C. Feature Selection

Feature selection (FS) is a problem of deciding an optimal subset of features based on some selecting algorithm. Not only the cost of recognition can be lowered down by reducing the number of features, but in some cases the performance of classification can be promoted by the optimal subset of features [50]. The sequential floating forward selection (SFFS) method is very effective in selecting an optimal subset of features [23]. It performs a number of backward steps after applying each forward step as long as the resulting subsets are better than previous feature sets of the same size. There will be no backward step if the performance can not be improved. The SFFS method performs almost as well as the branch-and-bound algorithm for optimal feature selection while demanding lower cost of computation [51].

In this study, we adopt the SFFS method for feature selection. First, we use the five-fold cross-validation procedure to estimate the CCR for the candidate feature subsets selected by the SFFS method at each stage for each of the Bayesian, k-NN, and SVM classifiers. Then, we apply both leave-one-out and five-fold cross-validation procedures to evaluate the performance of the selected feature set using each of the above three classifiers. Notice that, in applying the five-fold cross-validation procedure, the five groups of data used in feature selection are different from the five groups of data used in training and testing by random reassignment. Since we use the entire data set for feature selection, the test results by using the selected features are likely optimistically biased.

V. Experimental Results and Analysis

A. Image Acquisition

Prostatic tissue was embedded in paraffin cubes after chemical processing and then cut into very thin sections with thickness of 3–5 μm. These sections were placed on glass slides and stained with colored dyes using Hematoxylin-and-Eosin technique. The pathological images of prostatic cancer were acquired by a set of equipments including a high-quality optical microscope, a high resolution CCD camera, and an image acquisition computer system. All images were captured under the same illumination condition. There were 205 pathological images with resolution 512×384 pixels captured by the above procedure. We use the following two strategies to avoid inter- and intra-observer variations that may cause possible ambiguities in classification: 1) images were commonly analyzed by a group of experienced pathologists in Taichung Veterans General Hospital of Taiwan and classified into four classes in advance as "gold standard" for later comparison; 2) the pattern of the cancer observed in each sample must be greater than 60% of the total pattern seen in order to assign a primary Gleason grade to that sample.

Since Grade-1 patterns are very rare, Grade-1 and Grade-2 patterns are regarded as the same class. As a result, our image set was divided into four classes: 50 images in Class-1 (Grade-1 and Grade-2), 72 images in Class-2 (Grade-3), 31 images in Class-3 (Grade-4), and 52 images in Class-4 (Grade-5).

B. Feature Sets Used for Comparison

In this research, we use the feature sets derived from multiwavelets, Gabor filters, and GLCM methods to compare with our FD-based feature set and demonstrate the superiority of our approach over other methods.

Ten sets of multiwavelet features were proposed by Jafari–Khouzani and Soltanian–Zadeh [7] for grading pathological prostate images. Among them, multiwavelet-SA4,

multiwavelet-CL, and multiwavelet-GHM have the best performance. The details of defining multiwavelet-SA4, -CL, and -GHM can be found in [52]–[54]. In multiwavelet methods [7], a two-level multiwavelet transform of an image is performed to generate 28 subband images. Since two features, energy and entropy, are extracted from each subband image for classification, there will be a total number of 56 features in a feature set extracted from multiwavelet-SA4, multiwavelet-CL, or multiwavelet-GHM method.

In Gabor-filter method, a Gabor filter can be viewed as a sinusoidal plane of particular frequency and orientation modulated by a Gaussian envelope. It is a promising method for texture feature extraction in existing multi-channel filtering approaches [55], [56]. Roughly speaking, a 2-D Gabor filter acts like a local band-pass filter with certain optimal joint localization properties in the spatial domain and in the spatial frequency domain. An image is filtered with a set of Gabor filters of different preferred orientations and spatial frequencies to generate filtered images from which texture features can be extracted. In our experimental system, we implemented a bank of Gabor filters using the following five radial frequencies $\sqrt{2}/2^6$, $\sqrt{2}/2^5$, $\sqrt{2}/2^4$, $\sqrt{2}/2^3$, $\sqrt{2}/2^2$ and four orientations $0°$, $45°$, 90^0, and $135°$. How to choose appropriate radial frequencies for a bank of Gabor filters can be found in [55]. In our case, a set of 20 filtered images will be generated by Gabor filters method. We extract three features *energy*, *entropy*, and *magnitude* from each of the 20 filtered images. As a consequence, three sets of features called Gabor-Energy, Gabor-Entropy, and Gabor-Magnitude are formed with each one containing 20 features. We can combine these three sets of features to form a feature set of dimension 60 called Gabor-Combination.

In GLCM method, five statistical texture feature sets (*energy, entropy, contrast, correlation*, and *homogeneity*) are extracted from co-occurrence matrixes based on a particular scalar distance and four orientations $0°$, $45°$, 90^0, and $135°$. To determine an appropriate scalar distance for better capturing a specific feature, we estimate the CCR of that feature using ten distances (from 1 to 10 pixels) by Bayesian classifier and choose the distance which generates the highest CCR. In our experiment, the best distance is three pixels to capture *Energy* feature, four pixels to capture *Entropy* and *Contrast* features, one pixel to capture *Correlation* feature, and eight pixels to capture *Homogeneity* feature. Once we obtain the best distance which allows us to achieve the highest CCR for a specific feature, we can use that distance to generate four co-occurrence matrixes with each matrix corresponding to an orientation. Therefore, five feature sets are generated with each one containing four features. The five feature sets are called GLCM-Contrast, GLCM-Correlation, GLCM-Energy, GLCM-Homogeneity, and GLCM-Entropy, respectively. Like the Gabor-filter method, we can combine the above five feature sets together to form a feature set of dimension 20 called GLCM-Combination.

C. Performance of FD-Based Feature Set

Let us look at the performance of the FD-based feature set proposed in this paper in terms of CCR. For convenience of explanation, let $\boldsymbol{f_D} = \{f_{D1}, f_{D2}, f_{D3}, f_{D4}\}$, $\boldsymbol{f_E} = \{f_{E1}, f_{E2}, f_{E3}, f_{E4}\}$, and $\boldsymbol{f_D} + \boldsymbol{f_E} = \{f_{D1}, f_{D2}, f_{D3}, f_{D4}, f_{E1}, f_{E2}, f_{E3}, f_{E4}\}$. Table I shows the performance of feature sets $\boldsymbol{f_D}$, $\boldsymbol{f_E}$, and $\boldsymbol{f_D} + \boldsymbol{f_E}$ using Bayesian, k-NN, and SVM classifiers evaluated by leave-one-out method. Hereafter, we use notation $x \pm y\%$ to represent a 95% confidence interval for CCR, where x and y are the sample mean and 1.96 sample standard error, respectively. Moreover, we compute the p-value for the null hypothesis $H_0 : P_0 \leq P_1$, where P_0 is the CCR of using method A and P_1 is the CCR of using method B. If this p-value is less than 0.05, we can reject the above null hypothesis at the 5% significance level. In other words, we may conclude that the CCR of using method A is significantly higher than the CCR of using method B at the 5% significance level. The p-values are listed inside the brackets in tables from Table I–VIII.

In Table I, feature set $\boldsymbol{f_D}$ has a CCR of $90.2 \pm 4.1\%$ using Bayesian classifier, $92.2 \pm 3.7\%$ using k-NN classifier for $k = 3$, and $93.2 \pm 3.5\%$ using SVM classifier. The feature set $\boldsymbol{f_E}$ has a CCR of $86.3 \pm 4.7\%$ using Bayesian classifier, $88.8 \pm 4.3\%$ using k-NN classifier for $k = 5$, and $90.2 \pm 4.1\%$ using SVM classifier. The combined feature set $\boldsymbol{f_D} + \boldsymbol{f_E}$ has a CCR of $91.2 \pm 3.9\%$ using Bayesian classifier, $93.2 \pm 3.5\%$ using k-NN classifier for $k = 1$, and $93.7 \pm 3.3\%$ using SVM classifier. The experimental result indicates that both feature sets $\boldsymbol{f_D}$ and $\boldsymbol{f_E}$ are powerful in discriminating pathological images of prostatic carcinoma. Feature sets $\boldsymbol{f_D}$ and $\boldsymbol{f_E}$ may be complementary to each other because the CCR has a chance to become higher when we combine these two sets together. However, there are

TABLE I
COMPARISONS OF CCR FOR THE FEATURE SETS PROPOSED IN THIS PAPER USING BAYESIAN, k-NN, AND SVM CLASSIFIERS EVALUATED BY LEAVE-ONE-OUT METHOD

Features	$\boldsymbol{f_D}$		$\boldsymbol{f_E}$		$\boldsymbol{f_D}+\boldsymbol{f_E}$	
CCR by Bayesian (%)	90.2 ± 4.1 [0.364]		86.3 ± 4.7 [0.058]		91.2 ± 3.9	
CCR by k-NN (%)	k=1	89.8 ± 4.2 [0.108]	k=1	86.3 ± 4.7 [0.010]	k=1	93.2 ± 3.5
	k=3	92.2 ± 3.7 [0.500]	k=3	87.8 ± 4.5 [0.068]	k=3	92.2 ± 3.7
	k=5	90.7 ± 4.0 [0.360]	k=5	88.8 ± 4.3 [0.161]	k=5	91.7 ± 3.8
CCR by SVM (%)	93.2 ± 3.5 [0.419]		90.2 ± 4.1 [0.096]		93.7 ± 3.3	

Note: The number in the bracket [] is a p-value of the null hypothesis H_0: $P_0 \leq P_1$. P_0 is the CCR of using $\boldsymbol{f_D}+\boldsymbol{f_E}$ and P_1 is the CCR of using $\boldsymbol{f_D}$ (or $\boldsymbol{f_E}$).

TABLE II
COMPARISONS OF CCR FOR THE FEATURE SETS PROPOSED IN THIS PAPER USING BAYESIAN, k-NN, AND SVM CLASSIFIERS EVALUATED BY FIVE-FOLD CROSS-VALIDATION METHOD

Features	$\boldsymbol{f_D}$		$\boldsymbol{f_E}$		$\boldsymbol{f_D}+\boldsymbol{f_E}$	
CCR by Bayesian (%)	90.2 ± 4.1 [0.364]		86.8 ± 4.6 [0.077]		91.2 ± 3.9	
CCR by k-NN (%)	k=1	89.8 ± 4.2 [0.075]	k=1	86.3 ± 4.7 [0.006]	k=1	93.7 ± 3.3
	k=3	91.2 ± 3.9 [0.288]	k=3	87.8 ± 4.5 [0.047]	k=3	92.7 ± 3.6
	k=5	89.8 ± 4.2 [0.108]	k=5	89.8 ± 4.2 [0.108]	k=5	93.2 ± 3.5
CCR by SVM (%)	90.2 ± 4.1 [0.182]		88.3 ± 4.4 [0.064]		92.7 ± 3.6	

Note: The number in the bracket [] is a p-value of the null hypothesis H_0: $P_0 \leq P_1$. P_0 is the CCR of using $\boldsymbol{f_D}+\boldsymbol{f_E}$ and P_1 is the CCR of using $\boldsymbol{f_D}$ (or $\boldsymbol{f_E}$).

TABLE III
COMPARISONS OF CCR FOR VARIOUS FEATURE SETS USING BAYESIAN CLASSIFIER EVALUATED BY LEAVE-ONE-OUT AND FIVE-FOLD CROSS-VALIDATION PROCEDURES

Feature Sets	CCR evaluated by leave-one-out (%)	CCR evaluated by 5-fold cross-validation (%)
Multiwavelet - SA4 (*56*)	---	---
Multiwavelet - CL (*56*)	---	---
Multiwavelet - GHM (*56*)	---	---
Gabor - Energy (*20*)	82.0 ± 5.3 [0.002]	75.7 ± 5.9 [0.000]
Gabor - Entropy (*20*)	67.3 ± 6.4 [0.000]	64.9 ± 6.5 [0.000]
Gabor - Magnitude (*20*)	83.9 ± 5.0 [0.012]	79.0 ± 5.6 [0.000]
Gabor - Combination (*60*)	---	---
GLCM - Contrast (*4*)	82.9 ± 5.2 [0.006]	80.5 ± 5.4 [0.001]
GLCM - Correlation (*4*)	68.8 ± 6.3 [0.000]	68.0 ± 6.4 [0.000]
GLCM - Energy (*4*)	66.8 ± 6.5 [0.000]	66.2 ± 6.5 [0.000]
GLCM - Homogeneity (*4*)	73.2 ± 6.1 [0.000]	72.4 ± 6.1 [0.000]
GLCM - Entropy (*4*)	85.9 ± 4.8 [0.045]	84.1 ± 5.0 [0.014]
GLCM - Combination (*20*)	84.4 ± 5.0 [0.017]	82.2 ± 5.2 [0.003]
$\boldsymbol{f_D} + \boldsymbol{f_E}$ (*8*)	91.2 ± 3.9	91.2 ± 3.9

Note 1: The notation '---' represents that the covariance matrix ($\sum$) of a specific feature set is singular, so its inverse doesn't exist and there is no way to calculate CCR by Bayesian classifier.

Note 2: The number in the bracket [] is a p-value of the null hypothesis H_0: $P_0 \leq P_1$.
P_0 is the CCR of using our feature set (the last row); P_1 is the CCR of using other feature set.

TABLE IV
COMPARISONS OF CCR FOR VARIOUS FEATURE SETS USING k-NN CLASSIFIER EVALUATED BY LEAVE-ONE-OUT AND FIVE-FOLD CROSS-VALIDATION PROCEDURES

Feature Sets	CCR evaluated by leave-one-out (%)			CCR evaluated by 5-fold cross-validation (%)		
	$k=1$	$k=3$	$k=5$	$k=1$	$k=3$	$k=5$
Multiwavelet - SA4 (*56*)	90.2 ± 4.1 [0.135]	92.2 ± 3.7 [0.500]	89.8 ± 4.2 [0.253]	89.3 ± 4.2 [0.054]	90.7 ± 4.0 [0.231]	89.8 ± 4.2 [0.108]
Multiwavelet - CL (*56*)	92.2 ± 3.7 [0.349]	92.7 ± 3.6 [0.576]	90.2 ± 4.1 [0.298]	91.2 ± 3.9 [0.169]	91.2 ± 3.9 [0.288]	90.2 ± 4.1 [0.135]
Multiwavelet - GHM (*56*)	90.2 ± 4.1 [0.135]	87.3 ± 4.6 [0.050]	88.3 ± 4.4 [0.125]	88.3 ± 4.4 [0.028]	86.3 ± 4.7 [0.017]	86.3 ± 4.7 [0.010]
Gabor - Energy (*20*)	88.8 ± 4.3 [0.059]	90.7 ± 4.0 [0.294]	87.3 ± 4.6 [0.073]	86.8 ± 4.6 [0.009]	89.3 ± 4.2 [0.114]	87.8 ± 4.5 [0.031]
Gabor - Entropy (*20*)	76.1 ± 5.8 [0.000]	76.6 ± 5.8 [0.000]	74.6 ± 6.0 [0.000]	75.6 ± 5.9 [0.000]	74.6 ± 6.0 [0.000]	75.6 ± 5.9 [0.000]
Gabor - Magnitude (*20*)	86.8 ± 4.6 [0.015]	89.8 ± 4.2 [0.198]	88.3 ± 4.4 [0.125]	88.8 ± 4.3 [0.039]	88.8 ± 4.3 [0.086]	87.8 ± 4.5 [0.031]
Gabor - Combination (*60*)	92.2 ± 3.7 [0.349]	91.7 ± 3.8 [0.426]	91.7 ± 3.8 [0.500]	91.2 ± 3.9 [0.169]	91.7 ± 3.8 [0.353]	90.7 ± 4.0 [0.176]
GLCM - Contrast (*4*)	76.6 ± 5.8 [0.000]	74.6 ± 6.0 [0.000]	77.1 ± 5.8 [0.000]	74.2 ± 6.0 [0.000]	76.6 ± 5.8 [0.000]	78.1 ± 5.7 [0.000]
GLCM - Correlation (4)	63.4 ± 6.6 [0.000]	63.4 ± 6.6 [0.000]	62.9 ± 6.6 [0.000]	62.0 ± 6.6 [0.000]	61.5 ± 6.7 [0.000]	61.5 ± 6.7 [0.000]
GLCM - Energy (*4*)	61.0 ± 6.7 [0.000]	65.4 ± 6.5 [0.000]	64.9 ± 6.5 [0.000]	63.9 ± 6.6 [0.000]	63.9 ± 6.6 [0.000]	64.4 ± 6.6 [0.000]
GLCM - Homogeneity (*4*)	75.1 ± 5.9 [0.000]	76.1 ± 5.8 [0.000]	78.1 ± 5.7 [0.000]	71.2 ± 6.2 [0.000]	72.7 ± 6.1 [0.000]	74.6 ± 6.0 [0.000]
GLCM - Entropy (*4*)	80.0 ± 5.5 [0.000]	79.0 ± 5.6 [0.000]	79.0 ± 5.6 [0.000]	76.6 ± 5.8 [0.000]	74.6 ± 6.0 [0.000]	73.2 ± 6.1 [0.000]
GLCM - Combination (*20*)	83.4 ± 5.1 [0.001]	82.4 ± 5.2 [0.001]	83.4 ± 5.1 [0.005]	85.4 ± 4.8 [0.003]	82.9 ± 5.2 [0.001]	83.4 ± 5.1 [0.001]
$\boldsymbol{f_D} + \boldsymbol{f_E}$ (*8*)	93.2 ± 3.5	92.2 ± 3.7	91.7 ± 3.8	93.7 ± 3.3	92.7 ± 3.6	93.2 ± 3.5

Note: The number in the bracket [] is a p-value of the null hypothesis H_0: $P_0 \leq P_1$.
P_0 is the CCR of using our feature set (the last row); P_1 is the CCR of using other feature set.

some redundancies between $\boldsymbol{f_D}$ and $\boldsymbol{f_E}$ because the improvement in CCR is not significant for $\boldsymbol{f_D} + \boldsymbol{f_E}$. We will discuss the issue of feature selection (or feature optimization) in Section V-E.

Table II shows the performance of feature sets $\boldsymbol{f_D}$, $\boldsymbol{f_E}$, and $\boldsymbol{f_D} + \boldsymbol{f_E}$ using Bayesian, k-NN, and SVM classifiers evaluated by five-fold cross-validation method. Feature set $\boldsymbol{f_D}$ has an $90.2 \pm 4.1\%$ CCR using Bayesian classifier, $91.2 \pm 3.9\%$

TABLE V
COMPARISONS OF CCR FOR VARIOUS FEATURE SETS USING SVM CLASSIFIER EVALUATED BY LEAVE-ONE-OUT AND FIVE-FOLD CROSS-VALIDATION PROCEDURES

Feature Sets	CCR evaluated by leave-one-out (%)	CCR evaluated by 5-fold cross-validation (%)
Multiwavelet - SA4 (*56*)	93.2 ± 3.5 [0.419]	91.2 ± 3.9 [0.288]
Multiwavelet - CL (*56*)	93.2 ± 3.5 [0.419]	89.8 ± 4.2 [0.149]
Multiwavelet - GHM (*56*)	91.2 ± 3.9 [0.169]	92.2 ± 3.7 [0.424]
Gabor - Energy (*20*)	91.7 ± 3.8 [0.218]	89.3 ± 4.2 [0.114]
Gabor - Entropy (*20*)	80.5 ± 5.4 [0.000]	77.6 ± 5.7 [0.000]
Gabor - Magnitude (*20*)	92.7 ± 3.6 [0.344]	90.7 ± 4.0 [0.231]
Gabor - Combination (*60*)	90.7 ± 4.0 [0.128]	91.7 ± 3.8 [0.353]
GLCM - Contrast (*4*)	79.5 ± 5.5 [0.000]	75.1 ± 5.9 [0.000]
GLCM - Correlation (*4*)	78.5 ± 5.6 [0.000]	74.2 ± 6.0 [0.000]
GLCM - Energy (*4*)	76.1 ± 5.8 [0.000]	60.0 ± 6.7 [0.000]
GLCM - Homogeneity (*4*)	81.5 ± 5.3 [0.000]	73.7 ± 6.0 [0.000]
GLCM - Entropy (*4*)	88.8 ± 4.3 [0.039]	81.0 ± 5.4 [0.000]
GLCM - Combination (*20*)	89.3 ± 4.2 [0.055]	88.3 ± 4.4 [0.064]
$\boldsymbol{f_D}+\boldsymbol{f_E}$ (*8*)	93.7 ± 3.3	92.7 ± 3.6

Note: The number in the bracket [] is a p-value of the null hypothesis H_0: $P_0 \leq P_1$.
P_0 is the CCR of using our feature set (the last row); P_1 is the CCR of using other feature set.

TABLE VI
COMPARISONS OF CCR FOR VARIOUS FEATURE SETS WITH SFFS FEATURE SELECTION USING BAYESIAN CLASSIFIER EVALUATED BY LEAVE-ONE-OUT AND 5-FOLD CROSS-VALIDATION PROCEDURES

Feature Sets	CCR evaluated by leave-one-out (%)	CCR evaluated by 5-fold cross-validation (%)	# of Features Selected
Multiwavelet - SA4	93.7 ± 3.3 [0.500]	89.3 ± 4.2 [0.024]	11
Multiwavelet - CL	92.7 ± 3.6 [0.344]	89.8 ± 4.2 [0.034]	11
Multiwavelet - GHM	90.2 ± 4.1 [0.096]	88.8 ± 4.3 [0.016]	11
Gabor - Energy	89.3 ± 4.2 [0.055]	86.8 ± 4.6 [0.003]	11
Gabor - Entropy	80.0 ± 5.5 [0.000]	75.6 ± 5.9 [0.000]	11
Gabor - Magnitude	91.7 ± 3.8 [0.218]	88.3 ± 4.4 [0.011]	6
Gabor - Combination	91.7 ± 3.8 [0.218]	90.7 ± 4.0 [0.065]	7
GLCM - Combination	88.8 ± 4.3 [0.039]	87.3 ± 4.6 [0.005]	9
$\boldsymbol{f_D}+\boldsymbol{f_E}=\{f_{D1}, f_{D2}, f_{D3}, f_{D4}, f_{E4}\}$	93.7 ± 3.3	94.6 ± 3.1	5

Note: The number in the bracket [] is a p-value of the null hypothesis H_0: $P_0 \leq P_1$.
P_0 is the CCR of using our feature set (the last row); P_1 is the CCR of using other feature set.

CCR using k-NN classifier for $k = 3$, and $90.2 \pm 4.1\%$ CCR using SVM classifier. The feature set $\boldsymbol{f_E}$ has an $86.8 \pm 4.6\%$ CCR using Bayesian classifier, $89.8 \pm 4.2\%$ CCR using k-NN classifier for $k = 5$, and $88.3 \pm 4.4\%$ CCR using SVM classifier. The combined feature set $\boldsymbol{f_D} + \boldsymbol{f_E}$ has an $91.2 \pm 3.9\%$ CCR using Bayesian classifier, $93.7 \pm 3.3\%$ CCR using k-NN classifier for $k = 1$, and $92.7 \pm 3.6\%$ CCR using SVM classifier. The difference did not achieve statistical significance although a trend was observed from Tables I and II that the CCR could be improved if $\boldsymbol{f_D} + \boldsymbol{f_E}$ is used as the feature set instead of $\boldsymbol{f_D}$ or $\boldsymbol{f_E}$ alone.

D. Comparison of CCR Using Classifiers Without Feature Selection

Here, we want to evaluate the performance of other feature sets as described in Section V-B and compare with the performance of our feature set $\boldsymbol{f_D} + \boldsymbol{f_E}$. Table III shows the CCR for various feature sets using Bayesian classifier. If the performance is evaluated by leave-one-out method, $\boldsymbol{f_D} + \boldsymbol{f_E}$ has the highest CCR $(91.2 \pm 3.9\%)$; GLCM-Entropy has the second highest CCR $(85.9 \pm 4.8\%)$; and GLCM-Combination has the third highest CCR $(84.4 \pm 5.0\%)$. If the performance is evaluated by five-fold cross-validation method, the first three feature sets having the highest CCR are still $\boldsymbol{f_D} + \boldsymbol{f_E}(91.2 \pm 3.9\%)$, GLCM-Entropy $(84.1 \pm 5.0\%)$, and GLCM-Combination $(82.2 \pm 5.2\%)$. Table III shows that the FD-based feature set proposed in this paper is better than all other feature sets extracted from Gabor-filter and GLCM methods in terms of CCR at the 5% significance level. In this table, we did not provide any CCR information for Multiwavelet-SA4, Multiwavelet-CL, Multiwavelet-GHM, and Gabor-Combination because the covariance matrixes of these feature sets are singular and their inverse matrixes do not exist. Consequently, there is no way to apply the decision function stated in (8).

Table IV shows the classification performance of various feature sets using k-NN classifier by leave-one-out and five-fold cross-validation procedures. The feature set $\boldsymbol{f_D} + \boldsymbol{f_E}$ still has the highest CCR as compared to all feature sets extracted from Gabor-filters and GLCM methods. Such a result is consistent with the one listed in Table III. However, the difference in CCR among the feature sets Multiwavelet-SA4, Multiwavelet-CL, Gabor-Combination, and $\boldsymbol{f_D} + \boldsymbol{f_E}$ is unlikely to be statistically significant at the 5% significance level.

Table V shows the classification performance of various feature sets using SVM classifier. Overall speaking, the correct classification rates for all feature sets using SVM classifier

TABLE VII
COMPARISONS OF CCR FOR VARIOUS FEATURE SETS WITH SFFS FEATURE SELECTION USING k-NN CLASSIFIER EVALUATED BY LEAVE-ONE-OUT AND FIVE-FOLD CROSS-VALIDATION PROCEDURES

Feature Sets	CCR evaluated by leave-one-out (%)			CCR evaluated by 5-fold cross-validation (%)			# of Features Selected
	$k=1$	$k=3$	$k=5$	$k=1$	$k=3$	$k=5$	
Multiwavelet - SA4	91.2 ± 3.9 [0.169]	94.2 ± 3.2 [0.500]	92.2 ± 3.7 [0.276]	90.2 ± 4.1 [0.065]	91.2 ± 3.9 [0.169]	90.7 ± 4.0 [0.176]	11
Multiwavelet - CL	88.8 ± 4.3 [0.039]	93.7 ± 3.3 [0.416]	92.7 ± 3.6 [0.344]	88.3 ± 4.4 [0.017]	90.2 ± 4.1 [0.096]	91.7 ± 3.8 [0.283]	12
Multiwavelet - GHM	88.3 ± 4.4 [0.028]	92.2 ± 3.7 [0.210]	90.2 ± 4.1 [0.096]	87.3 ± 4.6 [0.008]	89.3 ± 4.2 [0.055]	86.3 ± 4.7 [0.010]	6
Gabor - Energy	87.8 ± 4.5 [0.019]	90.7 ± 4.0 [0.089]	90.7 ± 4.0 [0.128]	85.9 ± 4.8 [0.002]	87.8 ± 4.5 [0.019]	85.9 ± 4.8 [0.008]	9
Gabor - Entropy	74.2 ± 6.0 [0.000]	79.5 ± 5.5 [0.000]	79.0 ± 5.6 [0.000]	71.7 ± 6.2 [0.000]	77.6 ± 5.7 [0.000]	76.1 ± 5.8 [0.000]	8
Gabor - Magnitude	88.8 ± 4.3 [0.039]	91.2 ± 3.9 [0.121]	93.7 ± 3.3 [0.500]	86.8 ± 4.6 [0.005]	90.7 ± 4.0 [0.128]	89.3 ± 4.2 [0.081]	10
Gabor - Combination	91.2 ± 3.9 [0.169]	93.2 ± 3.5 [0.338]	92.2 ± 3.7 [0.276]	89.8 ± 4.2 [0.050]	90.2 ± 4.1 [0.096]	89.3 ± 4.2 [0.081]	10
GLCM - Combination	82.9 ± 5.2 [0.000]	85.9 ± 4.8 [0.002]	87.3 ± 4.6 [0.013]	81.5 ± 5.3 [0.000]	86.3 ± 4.7 [0.006]	84.9 ± 4.9 [0.003]	4
$\boldsymbol{f_D}+\boldsymbol{f_E}=\{f_{D1}, f_{D4}, f_{E4}\}$	93.7 ± 3.3	94.2 ± 3.2	93.7 ± 3.3	94.2 ± 3.2	93.7 ± 3.3	93.2 ± 3.5	3

Note: The number in the bracket [] is a p-value of the null hypothesis H_0: $P_0 \leq P_1$.
P_0 is the CCR of using our feature set (the last row); P_1 is the CCR of using other feature set.

TABLE VIII
COMPARISONS OF CCR FOR VARIOUS FEATURE SETS WITH SFFS FEATURE SELECTION USING SVM CLASSIFIER EVALUATED BY LEAVE-ONE-OUT AND FIVE-FOLD CROSS-VALIDATION PROCEDURES

Feature Sets	CCR evaluated by leave-one-out (%)	CCR evaluated by 5-fold cross-validation (%)	# of Features Selected
Multiwavelet - SA4	94.2 ± 3.2 [0.430]	93.2 ± 3.5 [0.354]	8
Multiwavelet - CL	94.2 ± 3.2 [0.430]	93.7 ± 3.3 [0.433]	11
Multiwavelet - GHM	93.7 ± 3.3 [0.349]	92.7 ± 3.6 [0.284]	8
Gabor - Energy	93.2 ± 3.5 [0.277]	92.7 ± 3.6 [0.284]	11
Gabor - Entropy	84.4 ± 5.0 [0.000]	82.9 ± 5.2 [0.000]	8
Gabor - Magnitude	93.7 ± 3.3 [0.349]	94.1 ± 3.2 [0.500]	12
Gabor - Combination	92.7 ± 3.6 [0.215]	91.2 ± 3.9 [0.130]	15
GLCM - Combination	88.3 ± 4.4 [0.011]	86.8 ± 4.6 [0.006]	6
$\boldsymbol{f_D}+\boldsymbol{f_E}=\{f_{D1}, f_{D2}, f_{D3}, f_{E1}, f_{E4}\}$	94.6 ± 3.1	94.1 ± 3.2	5

Note: The number in the bracket [] is a p-value of the null hypothesis H_0: $P_0 \leq P_1$
P_0 is the CCR of using our feature set (the last row); P_1 is the CCR of using other feature set.

are slightly higher than those using Bayesian or k-NN classifiers. Comparing with Gabor-Entropy, GLCM-Contrast, GLCM-Correlation, GLCM-Energy, GLCM-Homogeneity, and GLCM-Entropy, our FD-based feature set $\boldsymbol{f_D}+\boldsymbol{f_E}$ has a better performance in terms of CCR at the 5% significance level. However, comparing with Multiwavelet-SA4, Multiwavelet-CL, Multiwavelet-GHM, Gabor-Energy, Gabor-Magnitude, Gabor-Combination, and GLCM-Combination, the difference in CCR did not achieve statistical significance although a trend was observed. Notice that the dimensionalities of the feature sets derived from multiwavelets, Gabor filters, and GLCM-Combination are 56, 20, and 20, respectively; however, the dimensionality of our feature set $\boldsymbol{f_D}+\boldsymbol{f_E}$ is only 8.

E. Comparison of CCR Using Classifiers With Feature Selection

Feature selection is a fundamental problem in statistical pattern recognition and machine learning. In Section III, we present the method of constructing feature set $\boldsymbol{f_D}+\boldsymbol{f_E}$ based on fractal dimension calculation for discriminating pathological images of prostate carcinoma. In the previous subsection, we have demonstrated the effectiveness of $\boldsymbol{f_D}+\boldsymbol{f_E}$ and compare its performance with those feature sets extracted from multiwavelets, Gabor filters, and GLCM methods. Although we have shown that $\boldsymbol{f_D}$ and $\boldsymbol{f_E}$ are complementary to a certain degree, some redundancy may exist between them. To reduce this redundancy, or equivalently to reduce the dimensionality of feature set $\boldsymbol{f_D}+\boldsymbol{f_E}$, the SFFS feature selection method described in Section IV is applied. Similarly, all other feature sets are also optimized by the SFFS method and their classification results are presented in this subsection. Notice that feature sets GLCM-Contrast, GLCM-Correlation, GLCM-Energy, GLCM-Homogeneity, and GLCM-Entropy are removed from the list of comparison because their respective dimension is only 4. We can find their performance without feature selection from Tables III–V.

Tables VI–VIII show the classification performance of various feature sets after feature selection using Bayesian, k-NN, and SVM classifiers, respectively. We would like to highlight two important aspects based on the results shown in these three

tables: 1) a trend that feature set $\boldsymbol{f_D} + \boldsymbol{f_E}$ has the highest CCR was observed; 2) the feature set $\boldsymbol{f_D} + \boldsymbol{f_E}$ has the smallest dimension.

If we use Bayesian classifier to cooperate with the SFFS feature selection method, the feature set $\boldsymbol{f_D}+\boldsymbol{f_E}$ will be reduced to a set containing only five features $\{f_{D1}, f_{D2}, f_{D3}, f_{D4}, f_{E4}\}$. As we can see from Table VI, the correct classification rates of the optimized FD-based feature set $\boldsymbol{f_D} + \boldsymbol{f_E}$ are $93.7 \pm 3.3\%$ and $94.6 \pm 3.1\%$ using leave-one-out and five-fold cross-validation procedures, respectively. Let us look at the third column of Table VI, the feature set $\boldsymbol{f_D}+\boldsymbol{f_E} = \{f_{D1}, f_{D2}, f_{D3}, f_{D4}, f_{E4}\}$ has significantly higher CCR than all other feature sets at the 5% significance level except Gabor–Combination. Although Gabor–Combination has a very competitive performance, it has a dimension of 7 which is still longer than the length of $\boldsymbol{f_D} + \boldsymbol{f_E}$.

If we use k-NN classifier to cooperate with the SFFS feature selection method, the feature set $\boldsymbol{f_D} + \boldsymbol{f_E}$ will be reduced to a set containing only 3 features $\{f_{D1}, f_{D4}, f_{E4}\}$. Table VII shows the classification performance of various feature sets using k-NN classifier with SFFS as the feature selection method. The optimized feature set $\boldsymbol{f_D}+\boldsymbol{f_E}$ can achieve a CCR of $94.2 \pm 3.2\%$ for $k = 3$ using leave-one-out procedure and a CCR of $94.2 \pm 3.2\%$ for $k = 1$ using five-fold cross-validation procedure. The competitive feature sets are Multiwavelet-SA4, Multiwavelet-CL, and Gabor-Combination. We would like to point out that the dimensions of Multiwavelet-SA4, Multiwavelet-CL, and Gabor-Combination are 11, 12, and 10, respectively, while the dimension of optimized $\boldsymbol{f_D} + \boldsymbol{f_E}$ is only 3.

When we use SVM classifier to cooperate with the SFFS feature selection method, the feature set $\boldsymbol{f_D}+\boldsymbol{f_E}$ is reduced to a set containing five features $\{f_{D1}, f_{D2}, f_{D3}, f_{E1}, f_{E4}\}$, which is still the smallest among all other feature sets. As we can see from Table VIII, the correct classification rates of $\boldsymbol{f_D} + \boldsymbol{f_E}$ are $94.6 \pm 3.1\%$ and $94.1 \pm 3.2\%$ using leave-one-out and five-fold cross-validation procedures, respectively. The competitive feature sets are Multiwavelet-SA4, Multiwavelet-CL, Multiwavelet-GHM, Gabor-Energy, Gabor-Magnitude, and Gabor-Combination in terms of CCR. However, dimensions of Multiwavelet-SA4, Multiwavelet-CL, Multiwavelet-GHM, Gabor-Energy, Gabor-Magnitude, and Gabor-Combination are 8, 11, 8, 11, 12, and 15, respectively, while the dimension of optimized $\boldsymbol{f_D} + \boldsymbol{f_E}$ is only 5.

By summarizing all the experimental results from Table III–VIII, we may conclude that the FD-based feature set proposed in this paper has the smallest dimension while having either better or at least the same performance based on p-values as well as statistical results with a 95% confidence interval when compared to the feature sets extracted from multiwavelets, Gabor-filter, and GLCM methods. With feature selection, the correct classification rates for all feature sets using the three classifiers become slightly higher and their dimensions are reduced significantly. The optimized $\boldsymbol{f_D} + \boldsymbol{f_E}$ feature set did not exclude all the features in $\boldsymbol{f_D}$ or all the features in $\boldsymbol{f_E}$ as we can see from the last row in Tables VI–VIII. This implies that *Differential Box-Counting* method and *Entropy-Based Fractal Dimension Estimation* method used in this paper are both helpful to extract most informative FD-based features for achieving best performance in classification.

VI. Discussion and Conclusion

This paper presents an automated system for grading pathological images of prostatic carcinoma. The feature set proposed in this paper can be generated from pathological images using differential box-counting and entropy-based fractal dimension estimation techniques. The effectiveness of our proposed feature set was tested by Bayesian, k-NN, and SVM classifiers, respectively, using both of leave-one-out and k-fold cross-validation sampling procedures. A set of 205 pathological prostate images were used as the samples for classification. Experimental results demonstrated that the FD-based feature set proposed in this paper can provide very useful information for classifying pathological prostate images into four classes in Gleason grading system. As compared to other feature sets derived from multiwavelets, Gabor filters, and GLCM methods, our feature set has the highest correct classification rate and smallest dimensionality. With feature selection, our proposed feature set achieved a CCR of 93.7% using Bayesian classifier, a CCR of 94.2% using k-NN classifier, and a CCR of 94.6% using SVM classifier if leave-one-out was used as the evaluation procedure. When five-fold cross-validation was used as the evaluation procedure, a CCR of 94.6%, 94.2%, and 94.1% was achieved by Bayesian, k-NN, and SVM classifiers, respectively.

The concept of fractal geometry was originally introduced by Mandelbrot. The formula used to estimate fractal dimension $D = \log(N_r)/\log(1/r)$ is also well known. However, we would like to emphasize the following points to highlight the main contributions of this paper.

1) We successfully propose a fractal dimension feature set of very small size to grade prostate images effectively.
2) It is unobvious to calculate N_r and to define down scaled ratio r so that the property of "self-similarity" in prostate images can be reflected appropriately in a range of scales. In this paper, we successfully provide an elaborative design for defining the subranges of scales so that feasible FD texture features can be extracted from prostate images.
3) We propose a novel EBFDE method to calculate N_r based on entropy. By combining the FD features derived from entropy with the FD features derived from intensity difference, we come-up with a very powerful and concise feature set to facilitate Gleason grading for prostate images.
4) If we want to select a single category of features for grading prostate images, we demonstrate by extensive experiments that FD category has either better or at least the same performance statistically as compared to multiwavelet, Gabor, and GLCM categories. However, the feature set in FD category proposed in this paper has the smallest size. Such a result suggests that FD category will still have significant contributions and should be included for consideration if we want to select features from multicategories.

We would like to further point out that the results from our feature selection process are likely optimistically biased because we only have a small data set (205 images in total and 31 to 72 images for each of the four classes) and we use the entire data set for feature selection only once for each feature

set and each of the classifiers. Currently, we treat Grade-1 and Grade-2 in Gleason grading system as the same class because Grade-1 patterns are very rare and difficult to acquire. In our future study, we will apply our method to test the discriminating capability between Grade-1 and Grade-2 patterns when we gather enough number of images belonging to Grade-1 pattern.

ACKNOWLEDGMENT

The authors would like to thank all colleagues in the Department of Pathology of Taichung Veterans General Hospital for providing sample images and the grading results used in this research study. Special thanks to Prof. H.-P. Chan and the anonymous referees for their invaluable suggestions to make this paper more sound and complete.

REFERENCES

[1] *Cancer Facts & Figures 2007*. Atlanta, GA: American Cancer Society, 2007.
[2] Cancer Research UK: UK Prostate Cancer Statistics 2007 [Online]. Available: http://info.cancerresearchuk.org/cancerstats/types/prostate/
[3] Y. Zhu, S. Williams, and R. Zwiggelaar, "Computer technology in detection and staging of prostate carcinoma: A review," *Med. Image Anal.*, vol. 10, pp. 178–199, 2006.
[4] D. F. Gleason, , M. Tannenbaum, Ed., "The veteran's administration cooperative urologic research group: Histologic grading and clinical staging of prostatic carcinoma," in *Urologic Pathology: The Prostate*. Philadephia, PA: Lea Febiger, 1977, pp. 171–198.
[5] G. J. O'Dowd, R. W. Veltri, M. C. Miller, and S. B. Strum, "The Gleason score: A significant biologic manifestation of prostate cancer aggressiveness on biopsy," *Prostate Cancer Res. Inst.: PCR Insights*, vol. 4, no. 1, pp. 1–5, 2001.
[6] J. Diamond, N. Anderson, P. Bartels, R. Montironi, and P. Hamilton, "The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia," *Human Pathol.*, vol. 35, no. 9, pp. 1121–1131, 2004.
[7] K. Jafari-Khouzani and H. Soltanian-Zadeh, "Multiwavelet grading of pathological images of prostate," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 6, pp. 697–704, Jun. 2003.
[8] M. A. Tahir, A. Bouridane, and F. Kurugollu, "An FPGA based coprocessor for GLCM and Haralick texture features and their application in prostate cancer classification," *Analog Integrated Circuits Signal Process.*, vol. 43, no. 2, May 2005.
[9] D.-Y. Tsai and K. Kojima, "Measurements of texture features of medical images and its application to computer-aided diagnosis in cardiomyopathy," *Measurement*, vol. 37, no. 3, pp. 284–292, Apr. 2005.
[10] R. J. Ferrari, R. M. Rangayyan, J. E. L. Desautels, and A. F. Frere, "Analysis of asymmetry in mammograms via directional filtering with Gabor wavelets," *IEEE Trans. Med. Imag.*, vol. 20, no. 9, pp. 953–964, Sep. 2001.
[11] J. V. B. Soares, J. J. G. Leandro, R. M. Cesar, H. F. Jelinek, and M. J. Cree, "Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification," *IEEE Trans. Med. Imag.*, vol. 25, no. 9, pp. 1214–1222, Sep. 2006.
[12] B. B. Chaudhuri and N. Sarkar, "Texture segmentation using fractal dimension," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 1, pp. 72–77, Jan. 1995.
[13] J. W. Baish and R. K. Jain, "Fractals and cancer," *Cancer Res.*, vol. 60, pp. 3683–3688, Jul. 2000.
[14] A. Kikuchi, S. Kozuma, K. Sakamaki, M. Saito, G. Marumo, T. Yasugi, and Y. Taketani, "Fractal tumor growth of ovarian cancer: Sonographic evaluation," *Gynecol. Oncol.*, vol. 87, no. 3, pp. 295–302, 2002.
[15] P. Paramanathan and R. Uthayakumar, "Application of fractal theory in analysis of human electroencephalographic signals," *Comput. Biol. Med.*, vol. 38, pp. 372–378, 2008.
[16] I. S. Reljin and B. D. Reljin, "Fractal geometry and multifractals in analyzing and processing medical data and images," *Arch. Oncol.*, vol. 10, no. 4, pp. 283–293, 2002.
[17] T. Chikui, K. Tokumori, K. Yoshiura, K. Oobu, S. Nakamura, and K. Nakamura, "Sonographic texture characterization of salivary gland tumors by fractal analysis," *Ultrasound Med. Biol.*, vol. 31, no. 10, pp. 1297–1304, 2005.
[18] W. L. Lee, Y. C. Chen, and K. S. Hsieh, "Ultrasonic liver tissues classification by fractal feature vector based on M-band wavelet transform," *IEEE Trans. Med. Imag.*, vol. 22, no. 3, pp. 382–392, Mar. 2003.
[19] J. M. Zook and K. M. Iftekharuddin, "Statistical analysis of fractal-based brain tumor detection algorithms," *Magn. Reson. Imag.*, vol. 23, pp. 671–678, 2005.
[20] T. Stojic, I. Reljin, and B. Reljin, "Adaptation of multifractal analysis to segmentation of microcalcifications in digital mammograms," *Physica A*, vol. 367, pp. 494–508, 2006.
[21] N. Sarkar and B. B. Chaudhuri, "An efficient differential box-counting approach to compute fractal dimension of image," *IEEE Tran. Syst., Man Cybern.*, vol. 24, no. 1, pp. 115–120, Jan. 1994.
[22] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York: Academic, 1990.
[23] J. Novovicova, P. Pudil, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit.*, vol. 28, pp. 1389–1398, 1995.
[24] B. B. Mandelbrot, *Fractal Geometry of Nature*. San Francisco, CA: Freeman, 1982.
[25] B. B. Mandelbrot, "How long is the coast of Britain? Statistical self-similarity and fractional dimension," *Science*, vol. 156, pp. 636–638, 1967.
[26] A. P. Pentland, "Fractal based description of natural scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, pp. 661–674, Nov. 1984.
[27] Y. Xia, D. Feng, and R. Zhao, "Morphology-based multifractal estimation for texture segmentation," *IEEE Trans. Image Process.*, vol. 15, no. 3, pp. 614–623, Mar. 2006.
[28] M. A. Roula, J. Diamond, A. Bouridane, P. Miller, and A. Amira, "A multispectral computer vision system for automatic grading of prostatic neoplasia," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Washington, DC, 2002, pp. 193–196.
[29] R. Stotzka, R. Männer, P. H. Bartels, and D. Thompson, "A hybrid neural and statistical classifier system for histopathologic grading of prostate lesions," *Analytical Quantitative Cytol. Histol.*, vol. 17, no. 3, pp. 204–218, 1995.
[30] A. W. Wetzel, R. Crowley, S. J. Kim, R. Dawson, L. Zheng, Y. M. Joo, Y. Yagi, J. Gilbertson, C. Gadd, D. W. Deerfield, and M. J. Becich, "Evaluation of prostate tumor grades by content-based image retrieval," in *Proc. SPIE AIPR Workshop Advances Computer-Assist. Recognit.*, Washington, DC, 1999, vol. 3584, pp. 244–252.
[31] Y. Smith, G. Zajicek, M. Werman, G. Pizov, and Y. Sherman, "Similarity measurement method for the classification of architecturally differentiated images," *Comput. Biomed. Res.*, vol. 32, no. 1, pp. 1–12, 1999.
[32] A. Tabesh, V. Kumar, H. Pang, D. Verbel, A. Kotsianti, M. Teverovskiy, and O. Saidi, "Automated prostate cancer diagnosis and Gleason grading of tissue microarrays," *Proc. SPIE Med. Imag.*, vol. 5747, pp. 58–70, Apr. 2005.
[33] A. Tabesh and M. Teverovskiy, "Tumor classification in histological images of prostate using color texture," in *Proc. Asilomar Conf. Signals, Syst., Comput. (ACSSC) 2006*, Pacific Grove, CA, Oct. 2006, pp. 841–845.
[34] A. Tabesh, M. Teverovskiy, H.-Y. Pang, V. P. Kumar, D. Verbel, A. Kotsianti, and O. Saidi, "Multifeature prostate cancer diagnosis and Gleason grading of histological images," *IEEE Trans. Med. Imag.*, vol. 26, no. 10, pp. 1366–1378, Oct. 2007.
[35] S. Peleg, J. Naor, R. Hartley, and D. Avnir, "Multiple resolution texture analysis and classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 4, pp. 518–523, Jul. 1984.
[36] J. Gangepain and C. Roques-Carmes, "Fractal approach to two dimensional and three dimensional surface roughness," *Wear*, vol. 109, pp. 119–126, 1986.
[37] J. Keller, R. Crownover, and S. Chen, "Texture description and segmentation through fractal geometry," *Comput. Vis. Graph. Image Process.*, vol. 45, pp. 150–160, 1989.
[38] R. F. Voss, , R. Pynn and A. Skjeltorp, Eds., "Random fractals: Characterization and measurement," in *Scaling Phenomena in Disordered Systems*. New York: Plenum, 1986.
[39] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 2000.
[40] M. Kantardzic, *Data Mining: Concepts, Models, Methods, and Algorithms*. Hoboken, NJ: Wiley, 2002.
[41] C. Cortes and V. Vapnik, "Support-vector network," *Mach. Learn.g*, vol. 20, pp. 273–297, 1995.
[42] V. Vapnik, *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1995.
[43] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. Reading, MA: Addison Wesley, 2006.

[44] C. W. Hsu and C. J. Lin, "A comparison of methods for multi-class support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.

[45] T. F. Wu, C. J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *J. Mach. Learn. Res.*, vol. 5, pp. 975–1005, 2004.

[46] U. Krebel, "Pairwise classification and support vector machines," in *Advances in Kernel Methods: Support Vector Learning*. Cambridge, MA: MIT Press, 1999, pp. 255–268.

[47] L. Bottou, C. Cortes, J. Denker, H. Drucker, I. Guyon, L. Jackel, Y. LeCun, U. Muller, E. Sackinger, P. Simard, and V. Vapnik, "Comparison of classifier methods: A case study in handwriting digit recognition," in *Proc. Int. Conf. Pattern Recognit.*, 1994, pp. 77–87.

[48] C.-C. Chang and C.-J. Lin, LIBSVM: A library for support vector machines 2001 [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm

[49] R. Farjam, H. Soltanian-Zadeh, R. A. Zoroofi, and K. J. Khouzani, "Tree-structured grading of pathological images of prostate," in *Proc. SPIE Med. Imag.*, San Diego, CA, 2005, vol. 5746, pp. 840–851.

[50] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, Jan. 2000.

[51] A. K. Jain and Zongker, "Feature selection: Evaluation, application, and small sample performance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 153–158, Feb. 1997.

[52] L.-X. Shen, H. H. Tan, and J. Y. Tham, "Symmetric-antisymmetric orthonormal multiwavelets and related scalar wavelets," *Appl. Computat. Harmonic Anal. (ACHA)*, vol. 8, no. 3, pp. 258–279, May 2000.

[53] C. K. Chui and J. A. Lian, "A study of orthonormal multiwavelets," *Appl. Numer. Math.*, vol. 20, pp. 273–298, 1995.

[54] J. S. Geronimo, D. P. Hardin, and P. R. Massopust, "Fractal functions and wavelet expansions based on several functions," *J. Approx. Theory*, vol. 78, no. 3, pp. 373–401, 1994.

[55] A. K. Jain and F. Farrokhnia, "Unsupervised texture segmentation using Gabor filters," *Pattern Recognit.*, vol. 24, no. 12, pp. 1167–1186, 1991.

[56] O. Pichler, A. Teuner, and B. J. Hosticha, "A comparison of texture feature extraction using adaptive Gabor filtering, pyramidal and tree structured wavelet transforms," *Pattern Recognit.*, vol. 29, no. 5, pp. 733–742, 1996.