Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Support Vector Classification for Pathological Prostate Images Based on Texture Features of Multi-Categories

P. W. Huang, Cheng-Hsiung Lee
Department of Computer Science and Engineering
National Chung Hsing University
Taichung 40227, Taiwan
powhei.huang@msa.hinet.net

Phen-Lan Lin
Department of Computer Science and
Information Management, Providence University
Shalu, Taichung 433, Taiwan
lan@pu.edu.tw

***Abstract*—This paper presents an automated system for grading pathological images of prostatic carcinoma based on a set of texture features extracted by multi-categories of methods including multi-wavelets, Gabor-filters, GLCM, and fractal dimensions. We apply 5-fold cross-validation procedure to a set of 205 pathological prostate images for training and testing. Experimental results show that the fractal dimension (FD) feature set can achieve 92.7% of CCR without feature selection and 94.1% of CCR with feature selection by using support vector machine classifier. If features of multi-categories are considered and optimized, the CCR can be promoted to 95.6%. The CCR drops to 92.7% if FD-based features are removed from the combined feature set. Such a result suggests that features of FD category have significant contributions and should be included for consideration if features are selected from multi-categories.**

***Keywords*—Fractal dimension, Gleason grading, prostatic carcinoma, prostate image, SVM**

I. INTRODUCTION

Over the last few years, prostate carcinoma becomes the most common cancer in men. In the US, prostate cancer is the most frequently diagnosed cancer and ranks second among cancer deaths [1]. 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 [2]. By viewing the microscopic images of biopsy specimens, pathologists can determine the histological grades. The Gleason grading system [3] is the most widespread method for histological grading of prostate carcinoma.

Although pathologists can determine the histological grades by viewing the microscopic images of biopsy specimens, the process of human visual grading is time-consuming and very subjective due to inter- and intra-observer variations. Therefore, how to develop a more objective computer-aided technique for automatically and correctly grading prostatic carcinoma is the goal of this research study.

A classic Gleason grading diagram containing the five basic tissue patterns associated with the five tumor grades is shown in Fig. 1. As reported in [4], the use of texture analysis for prostatic lesions is very essential to the identification of tissue composition in prostatic neoplasia. Figure 2 shows four pathological images of prostatic carcinoma from well differentiated (grade 2) to very poorly differentiated (grade 5) in our image set. From Fig. 1 and Fig. 2, we can also see that the texture of prostate tissue plays an important role in Gleason grading for prostate cancer.

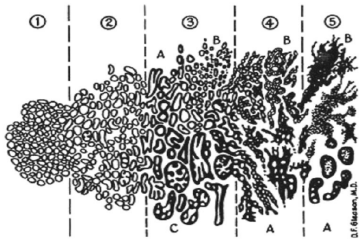

Figure 1. The Gleason grading diagram.

There are several well-known techniques for texture analysis such as extracting texture features from Gray-Level Co-occurrence Matrix (GLCM), Gabor filters, and multiwavelet transforms. The concept of fractal dimension (FD) is applied in this paper for analyzing the texture of prostate tissue. In physical phenomena, the growth of cancer shows the features of fractal. The fractal theory can provide clinically useful information for discriminating pathological tissue from healthy tissue [5].

The FD-based features can be extracted through differential box-counting (*DBC*) method [6] and entropy-based fractal dimension estimation (*EBFDE*) method to analyze pathological images of prostatic carcinoma. We combine the FD-based features with those extracted from multiwavelets, Gabor filters, and GLCM to form a set of 144 texture features of multi-categories for classification.

To evaluate the effectiveness of classification results based on these features, we use *k*-fold cross-validation procedure [7] to a set of 205 pathological prostate images and tested against these samples using Support Vector Machine (SVM) classifier to estimate the correct classification rates (CCR). For selecting an optimal set of features, we apply the Sequential Floating Forward Selection (SFFS) feature selection method [8].

Our system performs very well on classifying pathological prostate images in terms of CCR. Experimental results show that the fractal dimension (FD) feature set can achieve 92.7% of CCR without feature selection and 94.1% of CCR with feature selection by support vector machine classifier. If

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

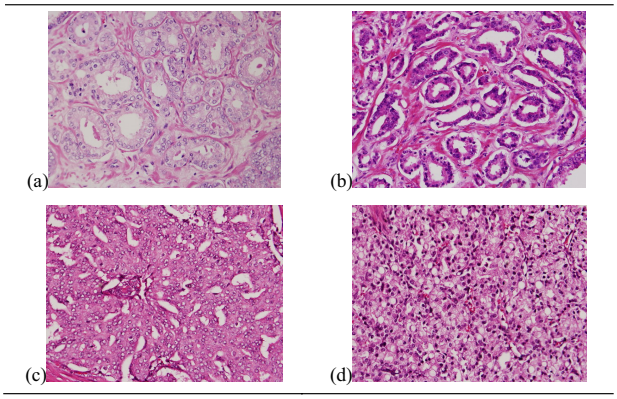

Figure 2. The prostate images of different cancer grades. (a) Gleason grade 2. (b) Gleason grade 3. (c) Gleason grade 4. (d) Gleason grade 5.

features of multi-categories are considered and optimized, the CCR can be promoted to 95.6%. The CCR drops to 92.7% if FD-based features are removed from the feature set. Such a result suggests that features of FD category have significant contributions and should be included for consideration if features are selected from multi-categories.

II. Feature Extraction

In this section, we present various well-know texture feature sets derived from multiwavelets, Gabor filters, GLCM, and FD-based methods for classifying histological prostate images.

Jafari-Khouzani et al. [4] 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. Ten sets of multiwavelet features were captured from different multiwavelet methods. Among them, multiwavelet-SA4 method has the best performance. The details of defining multiwavelet-SA4 can be found in [9]. In multiwavelet method, 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 method. We call this type of features as the Multiwavelet-category.

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 [10], [11]. 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 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^o, 45^o, 90^0, and 135^o. How to choose appropriate radial frequencies for a bank of Gabor filters can be found in [10]. 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 and call this type of features as the Gabor-category.

In GLCM method, five statistical texture feature sets (energy, entropy, contrast, correlation, and homogeneity) are extracted from co-occurrence matrices based on a particular scalar distance and four orientations 0^o, 45^o, 90^0, and 135^o. 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) and choose the distance which generates the highest CCR. In our experiment, the best distance is 3 pixels to capture Energy feature, 4 pixels to capture Entropy and Contrast features, 1 pixel to capture Correlation feature, and 8 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 matrices with each matrix corresponding to an orientation. Therefore, five feature sets are generated with each one containing four features. Like the Gabor-filter method, we combine the above five feature sets together to form a feature set of dimension 20 and call this type of features as the GLCM-category.

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]:

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

The fractal dimension texture features can be generated from a pathological image using differential box-counting (*DBC*) and entropy-based fractal dimension estimation (*EBFDE*) methods [13]. They can provide very useful information for classifying pathological prostate images into four classes in Gleason grading system. In our system, fractal dimension is estimated according to different ranges of scales for capturing the various self-similarity properties in a prostatic carcinoma image.

Before applying the *DBC* method, 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 pre-processing 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 $1 < s \le M/2$ and s is an integer. Then, we can get the scale ratio $r = s/M$. Consider the image as a three-dimensional

space such that (x, y) represents a two-dimensional 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

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

The contribution from all grids is

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

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).

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. The entropy-based fractal dimension estimation (*EBFDE*) method can further capture the information about randomness of pixels. 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:

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

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

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

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

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 sub-ranges: the sub-range of small scales {1/192, 1/96, 1/48}, the sub-range of medium scales {1/48, 1/24, 1/12}, and the sub-range 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 sub-ranges reflecting different self-similarity properties. We choose three scales in each sub-range 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 are obtained by *DBC* method and another four fractal dimension texture features are obtained by the *EBFDE* method. Then, we can combine these eight features to form a feature set denoted by $\boldsymbol{f_D}+\boldsymbol{f_E}$. We call this type of features as the FD-category. Details of extracting FD-based features from prostate images can be found in [13]. Finally, we combine the FD-based features with those extracted from multiwavelets, Gabor filters, and GLCM to form a set containing 144 texture features of multi-categories for classification.

III. Classification and Feature Selection

Compared with traditional classification methods which minimize the empirical training error, the goal of Support Vector Machine (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 non-linear 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 $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{id})^T$ corresponds to the feature vector for the ith sample $(i = 1, 2, ..., N)$ 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 non-linear SVM is more widely used for the general case due to its non-linear mechanism that can effectively classify data which are non-separable by a linear SVM. A non-linear SVM can be formulated by the following optimization problem:

$$\min_{\mathbf{w},b,\xi} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N} \xi_i$$

$$\text{Subject to} \quad y_i(\mathbf{w} \cdot \Phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, 2, ..., N, \quad (6)$$

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. In this study, we use the one-against-

one multi-class classification method based on LIBSVM [14], [15] with the RBF (Radial Basis Function) kernel in equation (7) by combining all pair-wise comparisons of binary SVM classifiers.

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \| \mathbf{x}_i - \mathbf{x}_j \|^2 / 2\sigma^2),\ \gamma > 0. \quad (7)$$

In addition, this study applies *k*-fold cross-validation method to estimate classification performance. In *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.

The correct classification rate (CCR) [16] is defined as

$$CCR = \sum_{i=1}^{C} P(c_i) \frac{n_i}{N_i}, \quad (8)$$

where n_i is the number of samples correctly classified to the *i*th class by using SVM classifier, C is the total number of classes, N_i is the total number of samples in the *i*th class, $P(c_i)$ is the prior probability that an observed data falls in class c_i.

In feature selection, we adopt the Sequential Floating Forward Selection (SFFS) method to choose optimal subsets of features. Feature selection (FS) is a problem of deciding an optimal subset of features based on some selecting algorithm. The SFFS feature selection method is very effective in selecting an optimal subset of features [8]. 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.

IV. Experimental Results

The image acquisition process is described as follows. First, prostatic tissue was embedded in paraffin cubes after chemical processing and then cut into very thin sections with thickness of 3~5μm. 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. These images were analyzed by experienced pathologists and classified into four classes in advance for later comparison. Since Grade-1 pattern is very rare, Grade-1 and Grade-2 are regarded as the same class. As a result, our image set was divided into four classes: 50 images in Class-1, 72 images in Class-2, 31 images in Class-3, and 52 images in Class-4.

Table I shows the classification results of multi-category feature sets using SVM classifier. In this table, the classification results in terms of CCR of three feature sets are compared. Multi-category-1 denotes the set of 144 features whose extraction methods have been described in Section II. Multi-category-2 is a set of 136 features formed by removing the 8 FD-based features from Multi-category-1. The $\boldsymbol{f_D}$+$\boldsymbol{f_E}$ feature set contains 8 FD-based features purely extracted from *DBC* and *EBFDE* methods [13]. Then, we use a 5-fold cross-validation procedure to train and test the above three feature sets using SVM classifier. As we can see from Table I, the CCRs of Multi-category-1 and Multi-category-2 are both 92.2% while the CCR of $\boldsymbol{f_D}$+$\boldsymbol{f_E}$ is 92.7%. It seems that feature set $\boldsymbol{f_D}$+$\boldsymbol{f_E}$ has the same discriminating capability as Multi-category-1 and Multi-category-2 have. However, the dimension of $\boldsymbol{f_D}$+$\boldsymbol{f_E}$ is only 8 while the dimensions of Multi-category-1 and Multi-category-2 are 144 and 136, respectively.

When the Sequential Floating Forward Selection method is applied to optimize the above three feature sets, the dimension of optimized Multi-category-1 is reduced to 10, the dimension of optimized Multi-category-2 is reduced to 8, and the dimension of optimized $\boldsymbol{f_D}$+$\boldsymbol{f_E}$ is reduced to 5. More specifically, optimized Multi-category-1 contains 1 FD-based feature, 3 Gabor-entropy features, 3 Gabor-energy features, and 3 Multiwavelet-SA4 features. Optimized Multi-category-2 contains 2 Gabor-entropy features, 1 Gabor-magnitude feature, and 5 Multiwavelet-SA4 features. The optimized $\boldsymbol{f_D}$+$\boldsymbol{f_E}$ feature set contains 3 FD-based features extracted by *DBC* method and another 2 FD-based features extracted by *EBFDE* method. Notice that the CCR of optimized $\boldsymbol{f_D}$+$\boldsymbol{f_E}$ feature set is 94.1% which is pretty close to the performance of optimized Multi-category-1 (95.6%) and outperforms the CCR of optimized Multi-category-2 (92.7%). This implies that the FD-based features are important to classifying pathological prostate images. If texture features from multi-categories are considered, FD-based features must be included. Otherwise, the classification performance will be degraded. The features from GLCM-category seem negligible in our classification problem domain.

V. Conclusion

This paper presents an automated system for grading pathological images of prostatic carcinoma based on a set of texture features extracted from multi-categories of methods including multi-wavelets, Gabor-filters, GLCM, and fractal dimension. Our system has very good performance on classifying pathological prostate images in terms of correct classification rate. Experimental results show that the FD-based feature set can provide very useful information for classifying pathological prostate images. A CCR of 92.7% can be achieved by a set of 8 FD-based texture features without feature selection. The CCR can be promoted to 94.1% if an optimized set of 5 FD-based features is used for classification. If multi-categories of features are considered and optimized, the CCR can be promoted to 95.6%. The CCR drops to 92.7% if FD-based features are not included in the feature set used for classification. Thus, FD-based features play an important role in prostate image classification.

TABLE I.
COMPARISONS OF THREE FEATURE SETS IN TERMS OF CCR USING SVM CLASSIFIER EVALUATED 5-FOLD CROSS-VALIDATION PROCEDURE

Feature Sets	Without feature selection		With feature selection	
	# of Features	*CCR (%)*	*# of Features*	*CCR (%)*
Multi-category-1	144	92.2	10	95.6
Multi-category-2	136	92.2	8	92.7
$f_D + f_E$	8	92.7	5	94.1

REFERENCES

[1] American Cancer Society, Cancer Facts & Figures 2007. Atlanta, GA: American Cancer Society, 2007.

[2] Y. Zhu, S. Williams, and R. Zwiggelaar, "Computer technology in detection and staging of prostate carcinoma: A review," *Medical Image Analysis*, vol. 10, pp. 178-199, 2006.

[3] D. F. Gleason, "The veteran's administration cooperative urologic research group: Histologic grading and clinical staging of prostatic carcinoma," *in Urologic Pathology: The Prostate, M. Tannenbaum Ed. Lea and Febiger, Philadephia*, PA, pp. 171-198, 1977.

[4] K. Jafari-Khouzani and H. Soltanian-Zadeh, "Multiwavelet grading of pathological images of prostate," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 6, pp. 697-704, June 2003.

[5] J. W. Baish and R. K. Jain, "Fractals and cancer," *Cancer Research*, vol. 60, pp. 3683-3688, July 2000.

[6] N. Sarkar and B. B. Chaudhuri, "An efficient differential box-counting approach to compute fractal dimension of image," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 24, no. 1, pp. 115-120, Jan 1994.

[7] K. Fukunaga, *Introduction to Statistical Pattern Recognition,* 2nd ed. Academic, New York, 1990.

[8] J. Novovicova, P. Pudil, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition* 28, pp. 1389-1398, 1995.

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

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

[11] 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 Recognition*, vol. 29, no. 5, pp. 733-742, 1996.

[12] B. B. Chaudhuri and N. Sarkar, "Texture segmentation using fractal dimension," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 1, January 1995.

[13] P. W. Huang and Cheng-Hsiung Lee, "Automatic classification for pathological prostate images based on fractal analysis," *to appear in IEEE Transactions on Medical Imaging*, 2009.

[14] 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.

[15] C.-C. Chang and C.-J. Lin, LIBSVM: a library for support vector machines, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[16] W.-Li Lee, Y.-Chang Chen, and K.-Sheng Hsieh, "Ultrasonic liver tissues classification by fractal feature vector based on M-band wavelet transform," *IEEE Transactions on Medical Imaging*, vol. 22, no. 3, pp. 382-392, March 2003.