Cardinal Multiridgelet-based Prostate Cancer Histological Image Classification for Gleason Grading

Hong-Jun Yoon1, Ching-Chung Li1, Christhunesa Cristudass2, Robert Veltri2, Jonathan I. Epstein2,3,4, and Zhen Zhang3,4
1Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, PA 15661
2Department of Urology, 3Department of Pathology, and 4Department of Oncology
Johns Hopkins Medical Institutions, Baltimore, MD 21224

Abstract

Computer-aided Gleason grading of prostate cancer tissue images has been in rapid development during the past decade. Automated classifiers using features derived from multiwavelet transform, fractal dimension and other measurements, and using texture forests have shown considerable successes. This paper presents our study on application of cardinal multiridgelet transform (CMRT) to prostate cancer images to extract texture features in the transform domain. CMRT can provide cardinality, approximate translation invariance and rotation invariance simultaneously. With 32 images of Gleason grade 3 and grade 4 as a training set and using texture features extracted therefrom, a support vector machine with Gaussian kernel has been trained to classify grade 3 and grade 4. The leave-one-out cross-validation showed its accuracy of 93.75% and AUC of 0.9651. 10 test images of grade 4 showed 100% accuracy.

1. Introduction

Prostate cancer is one of the most frequently diagnosed cancers and ranks the second among the cancer deaths of men in the United States [1]. One of the most reliable detection methods of prostate cancer is the examination of prostate histological specimens under a microscope by pathologists. The histological grading of prostate cancer tissue is assigned according to the Gleason grading system [2; 3]. It is based on microscopic tumor patterns assessed by pathologists while interpreting the biopsy specimen. The Gleason grading consists of five basic tissue patterns that reflect the degree of loss of normal glandular structure caused by the cancer. The grade, ranging from 1 to 5, increases with the increasing level of malignancy. In essence, the Gleason grade characterizes the degree of resemblance of a tissue under examination to the normal tissue. Grade 1 designates a well-differentiated tissue having the highest degree of resemblance to the normal tissue, and grade 5 designates a very poorly differentiated tissue showing the drastic departure from the normal tissue pattern.

Because human visual grading is very time-consuming and also subject to inter- and intra-observer variations, the development of machine vision techniques in aiding pathologists to analyze prostate tissue images and detect cancer in different stages has been in steady progress during the past decade. Applications of image texture analysis to computer-aided Gleason grading of prostate cancer tissue images have been reported since 2003 [4-19]. Wavelet and multiwavelet transforms, fractal analysis, and texture forest / random tree have been utilized for texture feature extraction and classification in studies of the automated Gleason grading [6-8; 10; 11; 14; 15].

As the malignancy of the prostate cancer is manifested by the loss of the normal glandular architecture (i.e., shape, size and differentiation of the glands) [2], it is appealing to apply the cardinal multiridgelet transform that has the excellent directional selectivity to quantitatively represent the glandular architecture effectively and reproducibly. We thus propose to explore the application of the developed cardinal multiridgelet transform to extract tissue texture features for use in a Gaussian-kernel support vector machine to aid the Gleason grading.

The biopsy Gleason score given to urologists is the sum of the primary grade (representing the majority of tumor cells) and the secondary grade (assigned to the minority of the tumor cells), and is a number ranging from 2 to 10, but today in the prostate specific antigen screening era the range tends to be 5-10 [20]. The higher the Gleason score, the more aggressive the cancer is likely to act and the worse the patient’s prognosis! would be [21]. The pathologist assigns a ‘score’ on a scale of Gleason grade 2 to 5. The Gleason grading system (tissue and cellular changes indicative of cancer) and tumor stage (pathologic extent of
Proper grading is a key to predict the patient’s prognosis and to provide adequate prescription for treatment [21]. The objective of our study is to explore the use of the cardinal multiridgelet transform to aid the classification between Gleason grade 3 and grade 4. Two sample images are in Figure 1.

In the following, Section 2 shows some background, Section 3 describes CMRT briefly, Section 4 discusses texture feature extraction in the transform domain and training of a kernel support vector machine for classifying Gleason grade 3 and grade 4, Section 5 presents our experimental result, and Section 6 gives a summary and conclusion.

2. Background

2.1. Radon Transform

The Radon transform definition used in many science and engineering fields can be found in [24]. If a bivariate function \( f(x,y) \) has no preferred orientation, its Radon transform is described as the integral over the line \( \rho = x \cos \theta + y \sin \theta \), written as

\[
R_{f(x,y)}(\rho, \theta) = \int \int f(x,y) \delta(x \cos \theta + y \sin \theta - \rho) dx dy
\]

where \( \delta(\cdot) \) is the Dirac delta function, \( \rho \in \mathbb{R} \) is the perpendicular offset of the line with respect to the origin \( (0,0) \), and \( \theta \in [0, \pi) \) is the angle of a radial line over which the integral projection is formed. That is, \( R_{f(x,y)}(\rho, \theta) \) is the integral of the function \( f(x,y) \) over the line \( \rho = x \cos \theta + y \sin \theta \). The Radon transform converts each of the line components into individual peaks positioned in corresponding to the line parameters, thus the investigation on line singularities may turn to an investigation on the local peaks.

2.2. Ridgelet Transform

For a given smooth univariate wavelet function \( \psi : \mathbb{R} \rightarrow \mathbb{R} \) define the bivariate ridgelet \( \psi_{a,b,\theta} : \mathbb{R}^2 \rightarrow \mathbb{R} \) by

\[
\psi_{a,b,\theta}(x,y) = \frac{1}{\sqrt{a}} \psi \left( \frac{x \cos \theta + y \sin \theta - b}{a} \right)
\]

for each scale \( a > 0 \), radial position \( b \in \mathbb{R} \), and orientation \( \theta \in [0, 2\pi) \). Therefore, a ridgelet is constant along lines \( x \cos \theta + y \sin \theta = \rho \). Transverse to these ridges, it is a single wavelet. The ridgelet transform of a given integrable bivariate function \( f(x,y) \) is defined by

---

**Figure 1:** Two samples of H&E stained cancerous prostate tissue images, (a) Gleason grade 3 and (b) Gleason grade 4.
The ridgelet transform can be interpreted as the 1-D wavelet transform to the slices of the Radon transform of $f(x, y)$. Note that

$$
RT_{f(x,y)}(a,b,\theta)\triangleq \left\langle f(x,y), \psi_{a,b}^{\theta}(x,y) \right\rangle
= \int \int f(x,y) \psi_{a,b}^{\theta}(x,y) \, dx \, dy
$$

The ridgelet transform is precisely the application of 1-D wavelet transform to each slice of the Radon transform where angular variable $\theta$ is held constant and $\rho$ varies as indicated in Equation (5).

### 2.3. Cardinal Multiwavelets

An orthogonal scaling function $\phi(t)$ satisfies the following condition

$$
\phi(n) = \delta(n) = \begin{cases} 
1, & n = 0 \\
0, & n = \pm 1, \pm 2, \ldots
\end{cases}
$$

is called a cardinal orthogonal scaling function (COSF). In this case the standard sampling theorem

$$
f(t) = \sum_n f\left(\frac{n}{2^j}\right) \phi\left(2^j t - n\right), \forall f(t) \in V_j(\phi)
$$

holds for every COSFs [25].

Integer shifts of scaling function $\phi(t)$ and the corresponding wavelet function $\psi(t)$ serve as basis functions in the multiresolution analysis of a finite energy signal. Discrete wavelet transform decomposes a signal at a certain scale level into the scaling component (low frequency) and wavelet component (high frequency) at the next coarser scale level (in dyadic scale) through scaling and wavelet filters, respectively, and the scaling component can be similarly decomposed into the coarser scale level. Then the scaling coefficients and wavelet coefficients together give a multi-resolution representation of the original signal based on which characteristic features may be extracted to describe the signal. When vector scaling function and vector wavelet are used, it becomes a multiwavelet analysis with matrix filters.

$$
H(n) = \begin{bmatrix} h_n(2n) & h_n(2n+1) \\
h_n(2n) & h_n(2n+1) \end{bmatrix}
$$

and

$$
G(n) = \begin{bmatrix} g_n(2n) & g_n(2n+1) \\
g_n(2n) & g_n(2n+1) \end{bmatrix}
$$

Selesnick [26] proposed the cardinal multiwavelet transform that can achieve cardinality, orthogonality, compact support, and higher balancing order given by

$$
\begin{align*}
&h_n(a) = \frac{1}{\sqrt{2}}(a,0,0,1,c,0,d,0,e,0,0,0,0,0) \\
h_n(b) = \frac{1}{\sqrt{2}}(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0) \\
g_n(a) = \frac{1}{\sqrt{2}}(-a,0,-b,-1,-c,0,-d,0,-e,0,0,0,0,0,0) \\
g_n(b) = \frac{1}{\sqrt{2}}(0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
\end{align*}
$$


Let $\phi_0(t)$ and $\phi_1(t)$ be the cardinal orthogonal multiscaling functions (COMSFs) with multiplicity equal to 2. In this case

$$
f(t) = \sum_n \left( f(n) \phi_0(t) + f(n+\frac{1}{2}) \phi_1(t) \right), \forall f(t) \in V_j(\phi)
$$

holds, where $\phi_0(n/2) = \delta(n)$ and $\phi_1(n/2) = \delta(n-1)$.

### 3. Methods

Ridgelet transform was developed to efficiently represent objects with highly anisotropic elements such as line segments in various orientations, while the wavelets are non-geometrical and do not exploit the regularity of the edges in an image. Multiridgelet transform is the extension of the ridgelet transform where multiwavelet transform can be incorporated in the Radon domain. Radon transform extracts piecewise line segments in an edge-dominated image. It maps a line singularity in the image into a point singularity in an oriented profile upon which multiwavelet transform is used in decomposition to capture the point singularities. As discussed below, the cardinal orthogonal multiridgelet transform has approximate translation invariant and rotation invariant properties in addition to cardinality, it may provide a better representation for image textures of the glandular architecture.
3.1. Multiridgelet Transform

For a chosen multiwavelet (for multiplicity=2) \( \psi = [\psi_0(t), \psi_1(t)]^T \), we define a bivariate multiridgelet \( \psi_{a,b,\theta}(x,y) \) by

\[
\psi_{a,b,\theta}(x,y) = a^{-1/2} \psi\left(\frac{x \cos \theta + y \sin \theta - b}{a}\right)
\]

(9)

for each scale \( a > 0 \), each radial position \( b \), and each orientation \( \theta \). A multiridgelet is a 2 \( \times \) 1 vector with components having constant values along a line \( x \cos \theta + y \sin \theta = \rho \) and giving ridges at various locations of \( b \). Traverse to these ridges, it is a multiwavelet. The multiridgelet transform of an integrable bivariate function \( f(x,y) \) is defined by

\[
MRT_{f(x,y)}(a,b,\theta) = \left\langle f(x,y), \psi_{a,b,\theta}(x,y) \right\rangle
= \int \int f(x,y) \psi_{a,b,\theta}(x,y) dx dy
\]

(10)

The multiridgelet transform can be interpreted as the 1-D multiwavelet transform to the slices of the Radon transform of \( f(x,y) \). Therefore the multiridgelet transform can be computed by first applying the Radon transform \( R_f(\rho,\theta) \) and then applying a 1-D multiwavelet transform to the slices of Radon transform.

\[
MRT_f(a,b,\theta) = \int \int \psi_{a,b}(t) R_f(\rho,\theta) d\rho
\]

(11)

To compute Radon transform, the discretization method based on slant stacks can be utilized.

3.2. Cardinal Multiridgelet Transform (CMRT)

We consider the multiridgelet incorporating the order 2 cardinal balanced multiwavelets (with multiplicity=2) [17]. There are two wavelets \( \psi_0(t) \) and \( \psi_1(t) \), and two corresponding scaling functions \( \phi_0(t) \) and \( \phi_1(t) \) for representing a sequence of Radon transform coefficients \( R_f(\rho,\theta) \) for each \( \theta \).

The decomposition output gives two channels (channel \( i=0, 1 \) of transform coefficients, each channel has a low-pass subband (L1 or L2) and a high-pass subband (H1 or H2) for each orientation angle \( \theta \) from \( 0^\circ \) to \( 179^\circ \), as illustrated in Figure 2, which shows the flow of the multiridgelet decomposition.

![Figure 2: The multiridgelet (multiplicity=2) decomposition of one scale level gives two-channel output, each channel has a low-pass (L1 or L2) and a high-pass (H1 or H2) subbands.](image)

The cardinal multiwavelets may be considered as a nearly approximate Hilbert transform pair. This is supported by the computation result that Fourier transforms of the two cardinal scaling filters, \( \{h_m(n)\} \) and \( \{h_n\} \), have almost the same magnitude response in their pass-band and the differences between their group delays and between their phase delays in the pass-band are nearly one half sample; the similar result also holds for two cardinal wavelet filters. This leads to the notion that the modulus of two channels of CMRT coefficients, i.e., the square root of the sum of squares of the pair of corresponding transform coefficients, may possess approximate translation invariant and rotation invariant properties that are advantageous for texture feature extraction.

Suppose that a square object \( f(x,y) \) is translated from the center coordinates by \( \zeta \) pixels along the \( y \) axis

\[
f(x,y) = f(x,y - \zeta)
\]

(12)

as shown in Figure 3. Its Radon transform is given by

\[
R_{f_{x,y}}(\rho,\theta) = \int \int f(x,y) \delta(x \cos \theta + (y + \zeta) \sin \theta - \rho) dx dy
= \int \int f(x,y) \delta(x \cos \theta + y \sin \theta - (\rho - \zeta \sin \theta)) dx dy
\]

(13)
which implies that the translation of the object is reflected as a sinusoidal shift in the $\rho$ variable in the Radon transform.

![Figure 3](image1.png)

Figure 3: (a) object aligned at center, (b) Radon coefficients of (a), (c) object translated, (d) Radon coefficients of (c), (e) high frequency coefficients of ridgelet transform of (a), (f) high frequency coefficients of ridgelet transform of (c), (g) magnitude of high frequency coefficients of CMRT of (a), and (h) magnitude of high frequency coefficients of CMRT of (c).

Note that the single ridgelet transform is translation sensitive, the high frequency coefficients of the single ridgelet coefficients of Figure 3(a) and Figure 3(c) are shown in Figure 3(e), and Figure 3(f), respectively. They showed highly translation dependent. On the other hand, the modulus of high frequency coefficients of the CMRT of the original and the translated object given in Figure 3(g) and Figure 3(h), respectively, show that the magnitude peaks are not changed by the shift of an object. This supports our assumption that the CMRT is almost translation invariant. The rotation of an off-center object produces a shift of the Radon transform in $\theta$ variable as well as a sinusoidal shift in $\rho$ variable, hence, the CMRT is approximately rotation invariant except a shift in $\theta$ variable.

4. Texture Feature Extraction and Classifier Training

4.1. Feature Extraction

Let us consider the application of the above-discussed cardinal multiridgelet transform to the prostate histological images, and let the decomposition be carried to 3 scale levels. In the tissue image classification, it is important to localize directional components specifying structural textures of gland units; but the gland units may lie in various orientations, so the actual orientation is not relevant and, therefore, the CMRT coefficients are not partitioned in direction but pooled together all directions in each subband for feature extraction. On the other hand, more detailed decomposition in frequency is desirable to distinguish the degree of disruption of the gland units; hence, multiwavelet packets are utilized, i.e., the high frequency component at each scale is also decomposed into two parts. The three-scale packet decomposition is taken in each direction of the Radon coefficients. Then, there exist 8 packets for each of the two CMRT channels. Packet index is frequency ordered from 0 to 7, 0 indicates the packet of the lowest frequency band, and 7 for the highest frequency band. Figure 4 shows the CMRT packet decomposition and index allocation of both channels.

![Figure 4](image2.png)

Figure 4: CMRT packet 3-scale decomposition.
The modulus of the 2-channel CMRT packet coefficients is given by

\[ \lvert W_{CMRP} \rvert = \sqrt{\lvert W_{CMRP, 0} \rvert^2 + \lvert W_{CMRP, 1} \rvert^2} \]

(14)

For each packet, variance \( \alpha_p^2 = \alpha^2 \lvert W_{CMRP, 0} \rvert \) and entropy \( s_p = s \lvert W_{CMRP, 0} \rvert \) are computed. In this case, the maximum possible number of features is sixteen, among which more discriminatory features may be selected for use in a pattern classifier.

4.2. Non-linear Support Vector Machine

A preliminary study on feature sets from the tissue images showed that they are not linearly separable. So a non-linear classifier must be considered. Let the 2-class feature vectors \( \{x_i, i=1,2,\ldots,N\} \), where \( N \) is the number of training samples, be mapped into a higher dimensional space under a non-linear mapping \( \phi: \mathbb{R} \rightarrow \mathbb{R}^m \) such that \( \{y_i = \phi(x_i), i=1,2,\ldots,N\} \) become linearly separable, and let

\[ g(y) = \mathbf{w}^T \cdot y + b = \mathbf{w}^T \cdot \phi(x) + b \]

be a discriminant function with \( g(y) > 0 \) for \( x \) belonging to the class of Gleason grade 4 and \( g(y) < 0 \) for \( x \) belonging to the class of Gleason grade 3.

![Figure 5: The non-linear support vector machine.](image)

To form a support vector machine (SVM) is to find the weight vector \( \mathbf{w} \) and threshold weight \( b \) so that the hyperplane \( g(y) = \mathbf{w}^T \cdot y + b = 0 \) in the \( y \)-space will have the maximum margin of separation for the given training data. Let \( z_i \) designate the class index of the \( i \)th training pattern \( y_i \) (or \( x_i \)), \( z_i = 1 \) for patterns of Gleason grade 4, and \( z_i = -1 \) for patterns of Gleason grade 3. Then consider

\[ z_i (w^T \cdot y + b) \geq 1, \; (i = 1,2,\ldots,N), \]

and \( \frac{1}{2} \| \mathbf{w} \|^2 \) is to be minimized. This results in a non-linear SVM classifier as shown in Figure 5, where the non-linear discriminant function is given by

\[ g(y) = \sum_{i=1}^{N} \alpha_i z_i \phi^T(x_i) \phi(x) + b = \sum_{i=1}^{N} \alpha_i z_i K(x,x_i) + b \]

\( K(x,x_i) \) denotes a non-linear kernel

\[ K(x,x_i) = \phi(x) \phi(x_i) = \phi^T(x) \phi(x_i) \]

\( N_s \) is the number of support vectors \( (N_s < N) \), support vectors are the training patterns that lie on the canonical surfaces \( \phi^T(x) \phi(x) + b = \pm 1 \) corresponding to \( \alpha_i > 0 \), and \( \alpha_i \) are the associated Lagrange multipliers. In this study, the Gaussian Radial Basis Function was used as the kernel function,

\[ K(x,x_i) = \exp \left( -\lvert \frac{x - x_i}{2\sigma^2} \rvert^2 \right). \]

5. Experimental Results

Prostate cancer tissue images utilized in this study is the Tissue MicroArray (TMA) 471 from the Pathology/Urology Department of the Johns Hopkins University (Dr. Jonathan I. Epstein). In this study, 16 of Gleason grade 3 cases and 16 of Gleason grade 4 were utilized for classifier training. To extract features from each sample image, we took an array of 768x768 pixels from the center part of the image, which is sufficient to contain the prostate cancer cells architectural information, and then subdivided it into 9 smaller blocks (patches), each of 256x256 pixels. Each feature extracted from each of the 9 patches is averaged to give a feature vector for that sample image. Due to the limited number of samples in our dataset, the number of features used should be limited.

Only three features out of 16 features mentioned in Section 4.1 were selected: variance features from multiridgelet transform packets 1 and 2, and entropy features from packet 6. The RBF SVM classifier was trained with the Gaussian kernel parameter \( \sigma = 0.15 \). The successfully trained SVM has 17 support vectors. The SVM and Kernel Methods Matlab toolbox was used for training our non-linear SVM classifier [27].
For comparison purpose, similar experiments were performed with feature extraction based upon the curvelet approach. The Curvelab Toolbox software is used for the curvelet transform [28]. Three-scale decomposition was taken. For each scale, variance and entropy features were calculated from curvelet coefficients for each orientation, and then averaged over all direction for each feature. Thus, the maximum number of features for curvelet method was 8. A curvelet-based SVM classifier was also studied for comparison where two variance features from lower frequency scales and two entropy features from two higher frequency scales were selected.

The leave-one-out (LOO) cross validation was used to estimate the generalization capability of the trained classifiers. We have 10 grade 4 samples in the dataset and used them as a test set. Even though it was a single-class testing, nevertheless it added some information to the LOO CV result. For both classifiers, LOO CV results and 10 test results are given in Table 1; as shown, the performance of our CMRT-based classifier was remarkable with the LOO CV accuracy of 93.75% and the test accuracy of 100%. The AUC measure was 0.9651 with a 95% confidence interval of (0.9090, 1.0000) which is well above the AUC value 0.8316 of curvelet-based classifier. ROC curves shown in Figure 6 and AUC values were determined using DBM-MRMC software with PROPROC (The “proper” binormal model) fit [29; 30].

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>LOO</th>
<th>AUC</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMRT</td>
<td>100%</td>
<td>93.75%</td>
<td>0.9651</td>
<td>100%</td>
</tr>
<tr>
<td>Curvelet</td>
<td>100%</td>
<td>81.25%</td>
<td>0.8316</td>
<td>70%</td>
</tr>
</tbody>
</table>

Figure 6: ROC curves of LOO CV of classifiers.

6. Conclusion

We have applied the cardinal multiridgelet transform to prostate cancer histological image texture feature extraction and trained a Gaussian kernel support vector machine for classification of Gleason grade 3 and grade 4. Among other prostate cancer tissue image classification researches, two recent works [10][16] were brought up for the comparison purpose. Since the dataset of those studies were not the same as ours, we are unable to perform one-to-one comparison between the accuracy results, however, this gave us an insight how well the proposed method performs. Next, a validation study must be conducted with a different tissue microarray to determine the accuracy of our classifier. Our result appears to be highly competitive. When a large dataset of images are available, we hope to evaluate our proposed method again and extend to classification of the interface between grade 3 and grade 4 that will be significant in computer-aided clinical decision-making. Ultimately a computer-assisted imaging program must be designed to allow pathologists and other scientists to evaluate this automated approach.

7. References

6. K. Jafari-Khouzani, and H. Soltanian-Zadeh, “Multiwavelet grading of pathological images of