WAVELET-BASED ANALYSIS OF HYPERSPECTRAL DATA FOR DETECTING SPECTRAL FEATURES

Pai-Hui HSU[*], Yi-Hsing TSENG [**]
National Cheng-Kung University, China-Taipei
Department of Surveying Engineering,
*p6885101@sparc1.cc.ncku.edu.tw
**Tseng@mail.ncku.edu.tw

Inter-technical IC-21

KEY WORDS: Hyperspectral Data, Spectral Feature Extraction, Wavelet Transform, Modulus Maxima, Remote Sensing.

ABSTRACT

The purpose of feature extraction is to abstract substantial information from the original data input and filtering out redundant information. In this paper we transfer the hyperspectral data from the original-feature space to a scale-space plane by using a wavelet transform to extract significant spectral features. The wavelet transform can focus on localized signal structures with a zooming procedure. The absorption bands are thus detected with the wavelet transform modulus maxima, and the Lipschitz exponents, are estimated at each singularities point of the spectral curve from the decay of the wavelet transform amplitude. The local frequency variances provide some useful information about the oscillations of the hyperspectral curve for each pixel. Different type of materials can be distinguished on the basis of the differences in the local frequency variation. The new method generates more meaningful features and is more stable than other known methods for spectral feature extraction.

1 INTRODUCTION

Multispectral sensors have been widely used to observe Earth surface since the 1960's. However, traditional sensors can only collect spectral data less than 20 bands due to the limitation of sensor technology. In recent years, spectral image sensors have been improved so as to collect spectral data in several hundred bands, which are called hyperspectral image scanners. For example, the AVIRIS scanners developed by NASA JPL provide 224 contiguous spectral channels. Theoretically, using hyperspectral images should increase our abilities in classifying land use/cover types. However, the data classification approach that has been successfully applied to multispectral data in the past is not as effective for hyperspectral data as well.

As the dimensionality of the feature space increases subject to the number of bands, the number of training samples needed for image classification has to increase too. If training samples are insufficient for the need, which is quite common for the case of using hyperspectral data, parameter estimation becomes inaccurate. The classification accuracy first grows and then declines as the number of spectral bands increases, which is often referred to as the Hughes phenomenon (Hughes, 1968), as shown in figure 1.

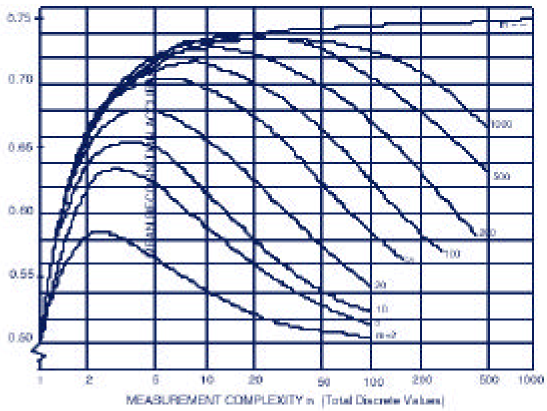


Figure 1. mean recognition accuracy vs. measurement complexity for the finite training cases (Hughes, 1968)

One of the approaches to improve the classification performance is to extract important features from the original hyperspectral data. The goal of employing feature extraction is to remove the redundant information substantially without sacrificing significant information. Some proposed feature extraction methods are based on stochastic theory (Hsu, 1999), such as the principal component analysis (Schowengerdt, 1997), discriminant analysis feature extraction (Tadjudin and Landgrebe, 1998), and decision boundary feature extraction (Lee and

Landgrebe, 1993).

Due to the high spectral resolution, it becomes possible to analyze the absorption characteristics of hyperspectral data. Symbolic descriptions of features can be used as quantitative indices of absorption bands to distinguish different materials. Some methods related to the physical characteristics are performed in the scale-space plane, such as the algorithms of spectral derivative ratio (Philpot , 1991) and spectral finger-prints (Piech and Piech, 1987).

In this paper, we attempt to transform the spectral data from the original feature space to a scale-space plane by using a wavelet transform. The wavelet transform can focus on localized signal structures with a zooming procedure. The local frequency characteristics such as the Lipschitz exponents provide some useful information about the oscillation of the spectral curve for each pixel. Different type of materials can be distinguished on the basis of the differences in the local frequency variation. The method we proposed in this study is called modulus maxima feature extraction (MMFE) as the reason that the features are extracted according to the wavelet transform modulus maxima. The new method generates more meaningful features and is more stable than other known methods for spectral feature extraction.

This paper first reviews some feature extraction methods that have been developed to speed up the process and increase the precision of classification. After that, the basic theory of wavelet transform is described for preparing the description of MMFE methods. Finally, three kinds of 220-band AVIRIS data are analyzed to illustrate our discovery and test to show the efficiency of the new feature extraction methods.

2 SPECTREAL FEATURE EXTRACTION

The high spectral resolution of hyperspectral data provide more information about the radiance of different materials, but the large amount of data in the high dimensional data space also create opportunities to extract meaningful spectral features. Generally speaking, features are any extractable measurement used. They may be numerical, symbolic, or both values. There are many approaches to extract important spectral features to reduce the dimensionality of hyperspectral data. In general, current known methods can be divided into two categories: the stochastic-based methods and the physical-related methods.

2.1 Stochastic-Based Feature Extraction

In the approaches of stochastic-based feature extraction, features are defined as band combinations that provide significant contributions to the statistical classification. The stochastic-Based methods allow for statistical treatment of multivariate data such as correlation analysis and regression, and finally produce quantitative results. The common method used for remote sensing application is the principal components transform (PCT). This method can remove the correlation between the spectral bands and concentrate the most variances in the first few components. Another method for generating a transformed set of feature axes in which class separation is optimized is called canonical analysis (Richard, 1986) or discriminant analysis feature extraction (DAFE). This approach uses the ratio of a between-class covariance matrix to within-class covariance matrix as a criterion function. Thus a transformation matrix is determined to maximize the ratio, that is, the separability of classes will be maximum after transformation. Besides, Lee and Landgrebe (1993) showed that discriminantly information features and redundant information features can be extracted from the decision boundary itself. The approach is called decision boundary feature extraction (DBFE).

Although stochastic-based methods can be tailored to a wide range of classification problems, but there are some restrictions on the application of hyperspectral image For examples, the computation time of PCT is large for hyperspectral images; the number of training samples is usually not enough to prevent singularity or yield a good covariance estimate in DBFE.

2.2 Physical-Related Feature Extraction

The molecular absorption bands of water and carbon dioxide cause deep absorption features that complete block transmission of radiation. These spectral regions were avoided for traditional remote sensing of the earth's surface (Schowengerdt, 1997). However, hyperspectral data give laboratory-like curves with spectral resolution sufficient to describe key absorption features of many materials. The ability for spectral analysis has also renewed interest in extracting physical-related spectral features in contrast to stochastic approaches.

One of the earliest physical-related feature extraction specifications for hyperspectral data was the calculation of image "residual" spectra for mineral detection and identification (Schowengerdt, 1997). This method emphasizes the absorption bands of different minerals relative to an average signature without absorption features.

A symbolic description called finger-print of the absorption bands for hyperspectral data was developed by Piech and Piech. The finger-print is a representation based on a scale space filtering of the hyperspectral data. A scale space image is a set of progressively smoothed versions of the original spectral curve. As the smoothing scale increase, features of the curve disappear until only a dominant spectral shape remains. The net result of scale space analysis of a hyperspectral data curve is a sequence of triplets. Each triplet describes a spectral feature and contains a measure of important directly related to the area contained within the spectral feature and the left and right inflection points of the spectral feature.

Another method proposed by Philpot (1991) to reduce the effects of atmospheric scattering and absorption on spectral signatures is the derivatives analysis. The derivatives are estimated using finite divided difference approximation algorithm with a finite band separation (Tsai and Philpot, 1997), $\Delta l = l_{i+1} - l_i$. The derivatives not only emphasize subtle spectral details, but also minimize illumination and atmospheric effects. Thus, derivatives are well-suited to extracting spectral features relating to specified target properties.

In this paper, a wavelet transform is applied to extract physical-related features. The wavelet transform can focus on localized signal structures with a zooming procedure. The local frequency characteristics such as the Lipschitz exponents provide useful information about the oscillation of the spectral curve for each pixel. In next section, we first briefly introduce the basic theory of wavelet transform and then explain the theory of MMFE method.

3 WAVELET-BASED FEATURE EXTRACTION

3.1 Wavelet Transform

The continuous wavelet transform which decomposes signals over dilated and translated wavelets was first introduced by Morlet and Grossmann (1984). A function $y \in \mathbf{L}^2(\mathbb{R})$ is named a wavelet if and only if

$$C_y = \int_0^{+\infty} \frac{|\hat{y}(w)|^2}{w} dw < \infty \tag{1}$$

This condition implies that the wavelet has a zero average:

$$\int_{-\infty}^{+\infty} y(x)dx = 0 \tag{2}$$

It is normalized by setting $\|y\| = 1$, and centered in the neighborhood of $x = 0$. A family of wavelets $y_{u,s}(x) = \frac{1}{\sqrt{s}} y\left(\frac{x-u}{s}\right)$ are obtained by introducing a scale factor s and a translation factor u. The wavelet transform of a function $f \in \mathbf{L}^2(\mathbb{R})$ is defined by

$$Wf(u,s) = \langle f, y_{u,s} \rangle = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{s}} y^*(\frac{x-u}{s})dx \tag{3}$$

The wavelet transform $Wf(u,s)$ is a function of the scale and the spatial position x. It measures the variation of f in the neighborhood of u, whose size is proportional to s. When the scale s varies from its maximum to zero, the decay of the wavelet coefficients characterizes the regularity of f in the neighborhood of u. This is the key idea to detect the absorption band position from reflectance spectra.

One can prove (Grossmann and Morlet, 1984) that the wavelet transform is invertible and $f(t)$ is recovered with

$$f(t) = \frac{1}{C_y} \int_0^{+\infty} \int_{-\infty}^{+\infty} Wf(u,s) \frac{1}{\sqrt{s}} y(\frac{x-u}{s}) du \frac{ds}{s^2} \tag{4}$$

3.2 Lipschitz Regularity

Singularities and irregular structures often carry essential information in a signal. To characterize singular structure of a signal, it is necessary to quantify local regularities precisely. Lipschitz exponents provide not only uniform regularity measurements over time intervals, but also pointwise Lipschitz regularity at any point v. If f has a singularity at v, which means that it is not differentiable at v, then the Lipschitz exponent at v characterizes this singular behavior. (Mallat, 1997)

If the wavelet has a compact support, the value of $Wf(u,s)$ depends upon the value of f(x) in the neighborhood of u, of size proportional to the scale s. At fine scales, it provides a localized information on $f(x)$. Measuring this asymptotic decay is equivalent to zooming into signal structures with a varies from its maximum to zero.

The relationship between the decay of the wavelet transform amplitude across scales and the pointwise Lipschitz regularity of the signal was described by Jaffard (1991). He proved a necessary and sufficient condition on the wavelet transform for estimating the Lipschitz regularity of f at a point v. Suppose that the wavelet ψ has n vanishing moments and n derivatives having a fast decay. If $f \in \mathbf{L}^2(\mathbb{R})$ is Lipschitz $\alpha \le n$ at v, then there exists $A > 0$ such that

$$\forall (u,s) \in \mathbb{R}\times\mathbb{R}^+ ,\ \left|Wf(u,s)\right| \le As^{\alpha+\frac{1}{2}}\left(1+\left|\frac{u-v}{s}\right|^{\alpha}\right) \tag{5}$$

Conversely, if $\alpha \le n$ is not an integer and there exists A and $\alpha' \le \alpha$ such that

$$\forall (u,s) \in \mathbb{R}\times\mathbb{R}^+ ,\ \left|Wf(u,s)\right| \le As^{\alpha+\frac{1}{2}}\left(1+\left|\frac{u-v}{s}\right|^{\alpha'}\right) \tag{6}$$

then f is Lipschitz α at v.

In order to simplify the above necessary and the sufficient condition, we assume that ψ has a compact support equal to $[-C,C]$. The cone of influence of v in the scale-space plane is the set of points (u, s) such that v is included in the support of $\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi(\frac{t-u}{s})$. Since the support of $\psi(\frac{t-u}{s})$ is equal to $[u-Cs, u+Cs]$, the cone of influence of v is defined by $|u-v| \le Cs$. It is illustrated in Figure 2. If u is in the cone of influence of v then $Wf(u,s)$ depends on the value of f in the neighborhood of v. Since $|u-v| \le Cs$, the conditions (5,6) can be written as

$$\left|Wf(u,s)\right| \le A' s^{\alpha+\frac{1}{2}} \tag{7}$$

which is identical to the uniform Lipschitz condition given by Mallat (1997).

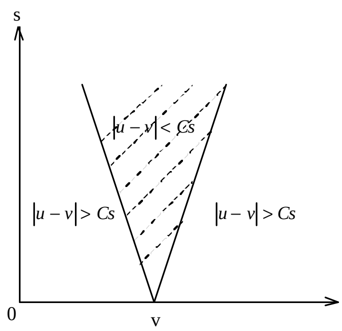


Figure 2. The corn influence in the scale-space.

3.3 Modulus Maxima of Wavelet Transform

A wavelet ψ with n vanishing moments can be written as the n^{th} order derivative of a function θ (Mallat, 1997), that is $\psi = (-1)^n \theta^{(n)}$, thus the resulting wavelet transform is a multiscale differential operator:

$$Wf(u,s) = s^n \frac{d^n}{du^n}(f * \bar{\theta}_s)(u). \tag{8}$$

Suppose the convolution $f * \bar{\theta}_s(u)$ averages $f(x)$ over a domain proportional to s. Let $\psi_1 = -\theta'$ and $\psi_2 = \theta''$ be the two wavelet defined by (8). The wavelet transforms, $W^1 f(u,s)$ and $W^2 f(u,s)$, are respectively to the first and second derivative of $f(x)$ smoothed by $f * \bar{\theta}_s(u)$. For a fixed scale, the local maxima of $W^1 f(u,s)$ and the zero-crossings of $W^2 f(u,s)$ will correspond to the inflection points of $f * \bar{\theta}_s(u)$ (figure 3). For all scales, the local maxima points of $W^1 f(u,s)$ can be connected as a set of maxima line in the scale-space plane (u, s). Similarly, the zero-crossings of $W^2 f(u,s)$ define a set of smooth curves that often look like finger-prints.

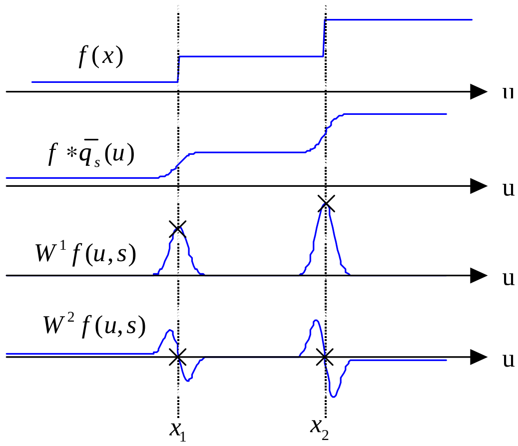


Figure 3. Local Maxima of $W^1 f(u,s)$ and the zero-crossing of $W^2 f(u,s)$ are the inflection point of function $f * \bar{\theta}_s(u)$.

By detecting the positions of local maxima or zero-crossings from coarse to fine scale, we can obtain the positions of the singularity of a signal. The two methods are very similar, but the local maxima approach has several important advantages (Mallat, 1992). The smoothing function q can be viewed as the impulse response of a low-pass filter. An important example often used in signal processing is the Gaussian function. In this study, the wavelet is defined as the first derivative of the Gaussian function, and the modulus maxima is used to detect positions of the absorption bands of a hyperspectral signal.

Although singularities can be detected by finding the abscissa where the wavelet modulus maxima converge at fine scale, the Lipschitz regularity however is not characterized yet. It is not sufficient to trace the wavelet modulus maxima across scales to detect singularities. We need more information about the decay of $|Wf(u,s)|$ along the maxima line. For simplification, we assume that all modulus maxima converge to v are included in a cone. In the other words, f does not have oscillations that accelerate in the neighborhood of v. The potential singularity at v is necessarily isolated. From (7), function f is uniformly Lipschitz a in the neighborhood of v if and only if there exist $A > 0$ such that each modulus maximum (u, s) in the cone satisfies

$$|Wf(u,s)| \le As^{a+\frac{1}{2}}, \tag{8}$$

which is equivalent to

$$\log_2|Wf(u,s)| \le \log_2 A + \left(a + \tfrac{1}{2}\right)\log_2 s \tag{9}$$

The Lipschitz regularity at v is thus the maximum slope of $\log_2|Wf(u,s)|$ as a function of $\log_2 s$ along the maxima lines converging to v.

4 EXPERIMENTS

4.1 Test data

The test data are a set of hyperspectral data delivered from the airborne visible/Infrared imaging spectrometer (AVIRIS). The data has 220 spectral bands from 400nm to 2450nm with 10 nm spectral resolution. The spectral curves of three different materials are shown in Figure 4.

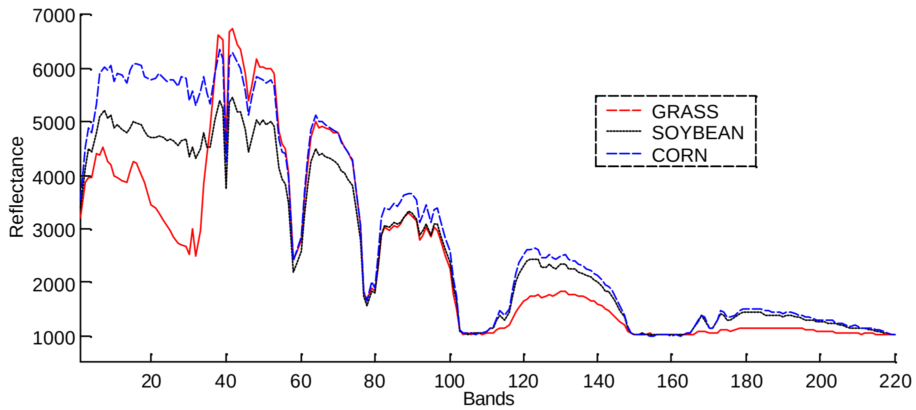
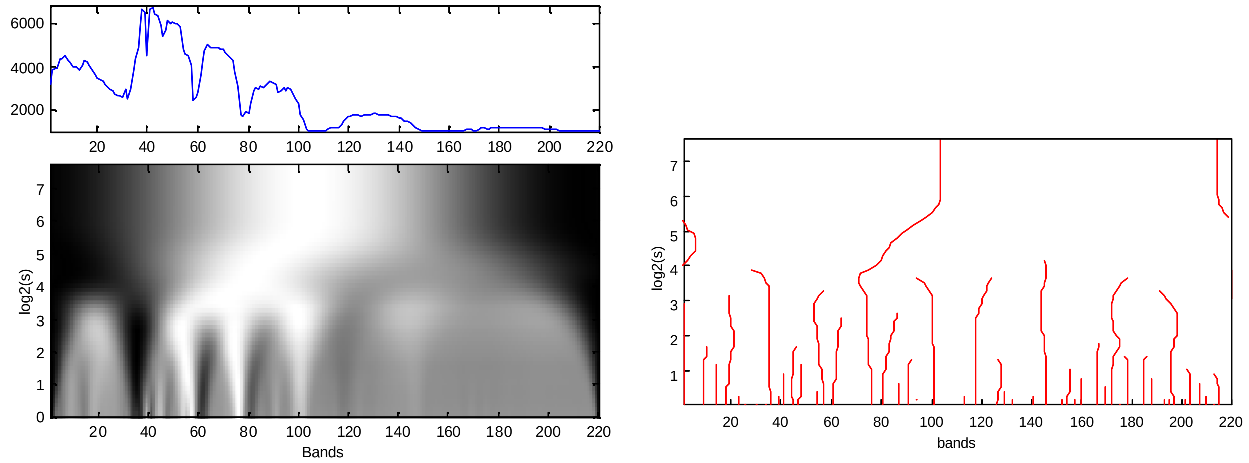


Figure 4. Three kinds of hyperspectral curves from AVIRIS images.

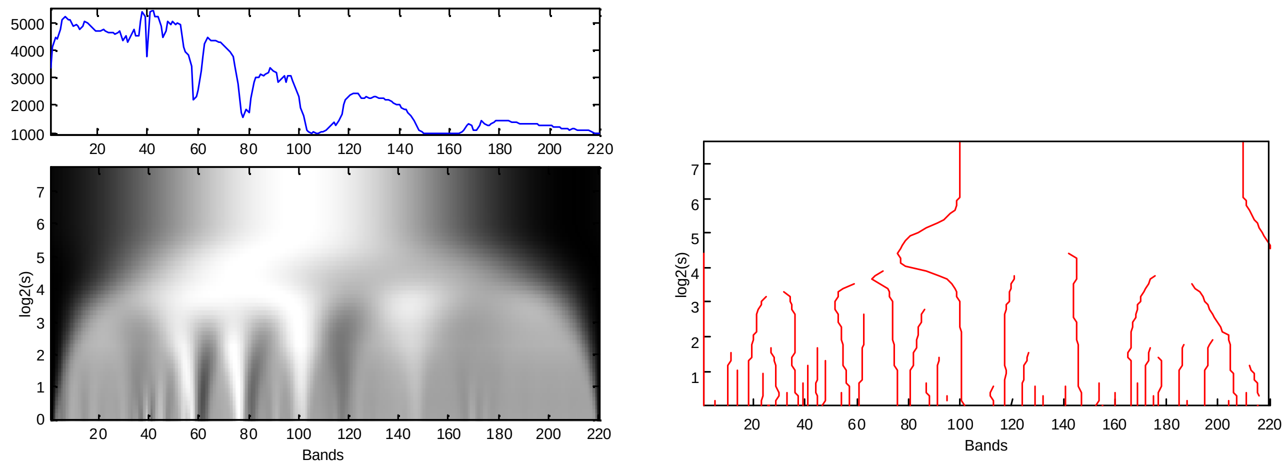
4.2 Wavelet Transform Modulus

Figure 5(a), 6(a) and 7(a) respectively show the wavelet transform $Wf(u,s)$ of the spectral curves with respect to the three materials. They were calculated with $y = -q'$, where q is a Gaussian function. The position parameter u and the $\log_2$ of scale s vary respectively along the horizontal and vertical axes. Black, gray and white points represent positive, zero and negative wavelet coefficients respectively. The absorption bands create large amplitude coefficients in their cone of influence. Figure 5(b), 6(b) and 7(b) show the results of extracting modulus maxima from $Wf(u,s)$. It could be seen that all singularities can be detected easily along the maxima line from coarse to fine scale. The maxima lines begun at large scale are corresponding to the main absorption bands of the spectral curves .



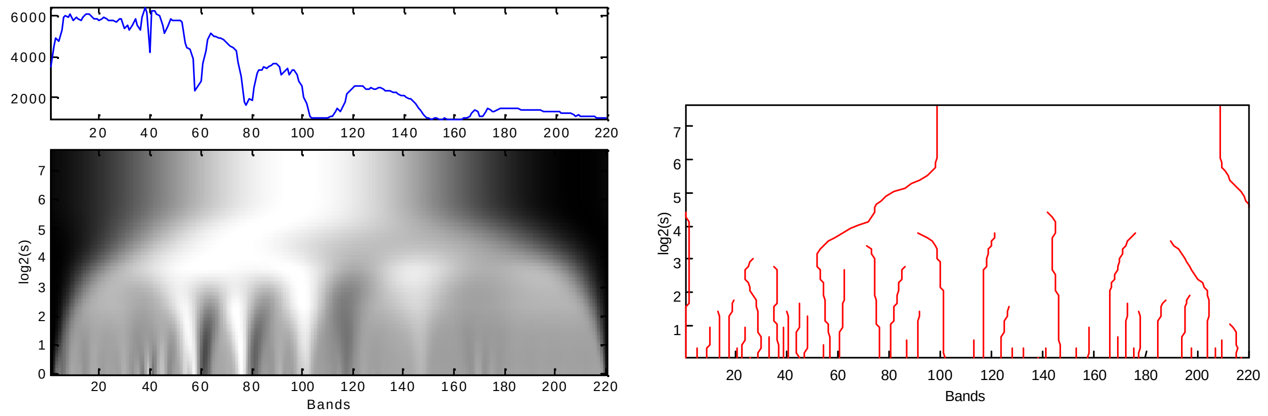
(a) (b)

Figure 5. (a) $Wf(u,s)$, wavelet transform of the grass spectrum, (b) The modulus maxima of $Wf(u,s)$



(a) (b)

Figure 6. (a) $Wf(u,s)$, wavelet transform of the soybean spectrum, (b) The modulus maxima of $Wf(u,s)$



(a) (b)

Figure 7. (a) $Wf(u,s)$, wavelet transform of corn spectrum, (b) The modulus maxima of $Wf(u,s)$

4.3 Lipschitz Exponent

Figure 8 shows the results of Lipschitz exponents calculated at each singularity point for the test data. Because the Lipschitz regular a was estimated under the assumption of the compact support wavelet y with one vanishing moment, the values of a should satisfy $0 \le a < 1$ for isolated singularity. The negative values of Lipschitz exponents indicate the corresponding points possessing high-frequency oscillations in their neighborhood.

The spectral curves of soybean and corn shown in figure 4 look quite similar in shape and their main absorption bands corresponding to the low-frequency components are almost the same. However, one would be easy to discover the differences when a scale factor is introduced to overlap these two curves. In figure 8 obviously shows that the high-frequency variations of the corn spectrum are larger than the soybean's in the 21^{th}, 76^{th} and 153^{th} bands. Therefore, this two different materials can be distinguished easily with the Lipschitz exponents.

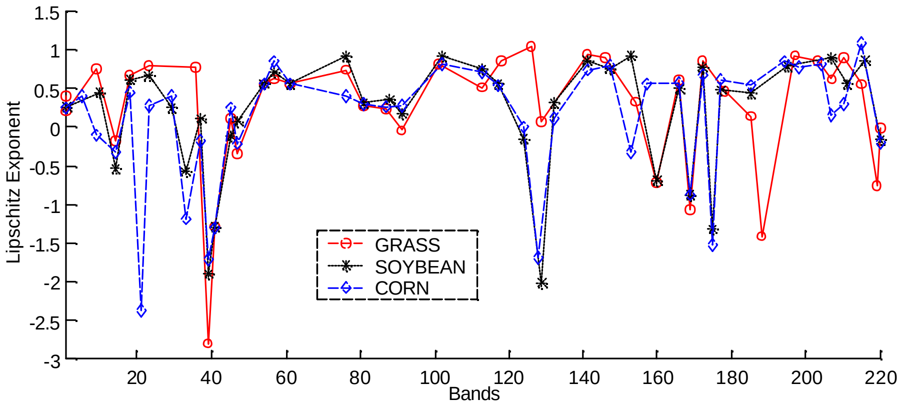


Figure 8. The Lipschitz exponents calculated at each singularity points for the test data.

4.4 Dyadic Wavelet Transform

For fast numerical computations, the detection of the wavelet transform maxima can be limited to dyadic scale $\{2^j\}_{j\in\mathbb{Z}}$. Figure 9 shows the result of the dyadic wavelet transform of corn spectrum. This dyadic wavelet transform has the same properties as a continuous wavelet transform, $Wf(u,s)$. Singularities create sequences of maxima that converge towards the corresponding location at fine scales, and the Lipschitz regularity is calculated from the maxima amplitude.

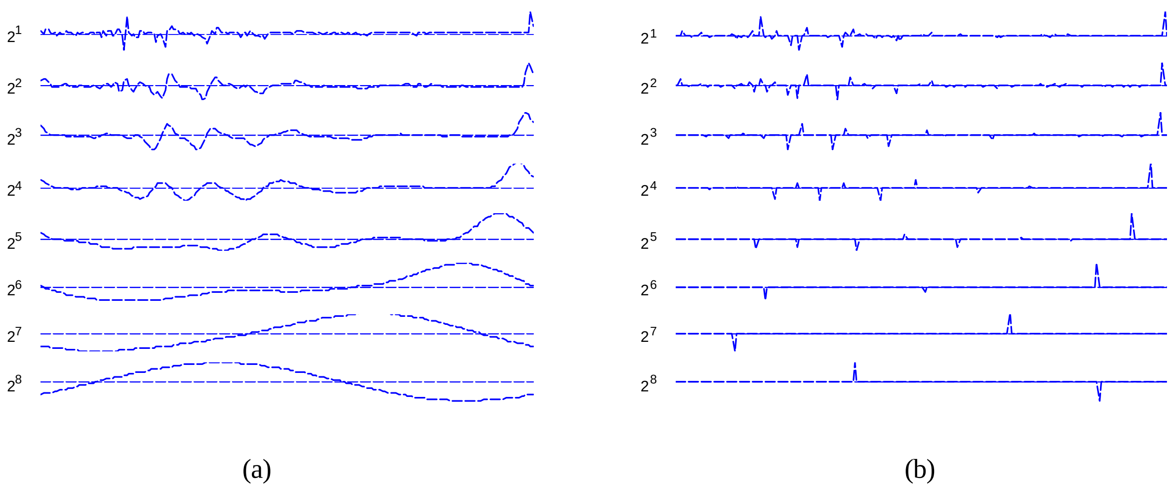


Figure 9. (a) Dyadic wavelet transform computed up to the scale 2^8,
(b) Modulus maxima of the dyadic wavelet transform.

5 CONCLUSIONS

The high spectral resolution of hyperspectral data provides the ability for diagnostic identification of different materials. In order to increase the classification performance, feature extraction is a pre-processing for removing the redundant information substantially without sacrificing significant information. In this paper we transfer the hyperspectral data to the scale-space plane by using wavelet transform to extract important spectral features. The wavelet transform can focus on localized signal structures with a scaling and dilation of a wavelet. The absorption bands of spectral curves are thus detected automatically by the wavelet transform modulus maxima, and the Lipschitz exponents are estimated at each singularities point of the spectral curve from the decay of the wavelet transform amplitude. The local frequency variances provide useful information about the oscillations of the hyperspectral curve for each pixel. Different materials can be distinguished on the basis of the differences in the local frequency variation. The new method generates more meaningful features and is more stable than other known methods for spectral feature extraction.

In particular, The method proposed in this paper will be helpful for spectral analysis that reduce the multidimensional hyperspectral data to a smaller number of key features that can be both automatically processed and physically interpreted.

ACKNOWLEDGMENTS

This research project was sponsored by the National Science Council of Republic of China under the grants of NSC88-2211-E006-051.

REFERNENCES

Grossmann, A. and Morlet, J., 1984. Decomposition of Hardy functions into square integrable wavelets of constant shape, SIAM J. Math, Vol. 5, pp. 723-736.

Hsu, P. H. and Tseng, Y. H. 1999. Feature Extraction for Hyperspectral Image. Proceedings of the 20th Asian Conference on Remote Sensing, Vol. 1, pp. 405-410.

Hughes, G. F., 1968. On the Mean Accuracy of Statistical Pattern Recognizers, IEEE Transaction on Information Theory, IT-14. pp. 55-63.

Jaffard, S., 1991. Pointwise smoothness, two microlocalisation and wavelet coefficients, Publications Mathematiques, Vol. 35

Lee, C. and Landgrebe, D., 1993. Feature Extraction and Classification Algorithms for High Dimensional Data, TR-EE 93-1, Purdue University.

Mallat, S., 1997. A wavelet tour of signal processing, ACADEMIC PRESS.

Mallat, S. and Hwang, W. L., 1992. Singularity Detection and Processing with Wavelets, IEEE Transactions on Information Theory, Vol. 38, No. 2, pp. 617-643.

Piech, M. A. and Kenneth, R. P., 1987. Symbolic representation of hyperspectral data, Applied Optic, Vol. 26, No.18, pp. 4018-4026.

Piech, M. A. and Kenneth, R. P., 1989. Hyperspectral interactions: invariance and scaling, Applied Optic, Vol. 28, No. 3, pp.481-489.

Philpot, W. D., 1991. The Derivative Ratio Algorithm: Avoiding Atmospheric Effects in Remote Sensing, IEEE Transactions on Geoscience and Remote Sensing, Vol.29, No.3, pp.350-357.

Richards, J. A., 1993. Remote Sensing Digital Image Analysis: An Introduction, Springer-Verlag Berlin Heidelberg, Second Edition.

Schowengerdt, R. A.., 1997. Remote Sensing: Models and Methods for Image Processing, Academic Press.

Tadjudin, S. and Landgrebe, D., 1998. Classification of High Dimensional Data with Limited Training Samples, PhD Thesis and School of Electrical & Computer Engineering Technical Report TR-ECE 98-8., Purdue University.

Tsai, F. and Philpot, W., 1997. Derivative Analysis of Hyperspectral Data for Detecting Spectral Features, Geoscience and Remote Sensing, IGARSS '97, Remote Sensing - A Scientific Vision for Sustainable Development, 1997 IEEE International Volume: 3 , 1997 , Vol.3, pp. 1243 –1245.