WAVELE-BASED REDUCTION OF HYPERSPECTRAL IMAGERY

B.Salehi[a,*], M.J.Valadan Zoej[b]
Department of Remote Sensing, Faculty of Geodesy and Geomatics Engineering, K.N.Toosi University of Technology, Tehran Iran
[a] (salehi_bahram@yahoo.com), [b] (Valadanzouj@kntu.ac.ir)

PS WG VII/1

KEY WORDS: remote sensing, hyper spectral, analysis, classification, reconstruction, spectral

ABSTRACT:

New sensor technology has made it possible to gather multispectral images in hundreds and potentially thousands of spectral bands, this tremendous increase in spectral resolution should provide a wealth of detailed information, but the techniques used to analyze lower dimensional data often perform poorly on high dimensional data. Therefore, it is necessary to investigate the problem and to explore effective approaches to hyperspectral data analysis. Studies indicate that the key problem is to need very large number of labeled samples. It has been found that the conventional approaches can be retained if a preprocessing stage is established.
Dimension reduction is a preprocessing stage that brings data from a high order dimension to a low order dimension. Some stochastic-based techniques are used for dimension reduction such as Principal Component Analysis (PCA), Discriminant Analysis Feature Extraction (DAFE) and Decision Boundary Feature Extraction (DBFE).But these techniques have some restrictions. For example PCA is computationally expensive and does not eliminate anomalies that can be seen at one arbitrary band; the number of training samples is usually not enough to prevent singularity or yield a good covariance estimate in DBFE.
Spectral data reduction using Automatic Wavelet Decomposition could be useful because it preserves the distinction among spectral signatures. It is also computed in automatic fashion and can filter data anomalies. This is due to the intrinsic properties of Wavelet Transform that preserve high and low frequency feature therefore preserving peaks and valleys found in typical spectra. Compared to PCA, for the same level of data reduction this paper shows that automatic wavelet reduction yields better or comparable classification accuracy.

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 (Hsu, Pai-Hui, 2000). 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 (Hsieh and Landgrebe, *1998*).
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. Fukunaga (1989) proved that the required number of training samples is linearly related to the dimensionality for a linear classifier and to the square of dimensionality for a quadratic classifier. It has been estimated that as the number of dimensions increases the training samples size need to increases exponentially in order to have an effective estimate of the multivariate densities needed to perform a non-parametric classification. 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.

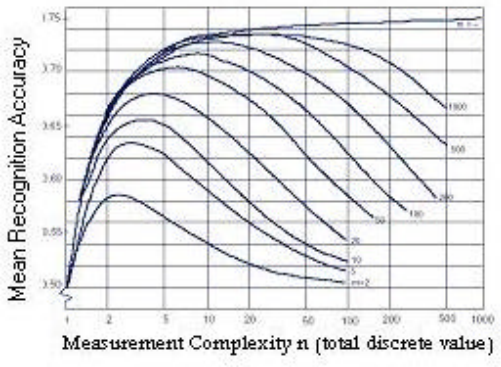

Figure1. Mean recognition accuracy vs. measurement of complexity for the finite training cases (Houghes, 1968)

One of the approaches to improve the classification performance is to reduce dimensionality via a preprocessing method, which takes into consideration high dimensional spaces properties. Dimension reduction is the transformation that brings data from a high order dimension to a low order dimension, similar to lossy compression method, dimension reduction reduced the size of the data but unlike compression, dimension reduction is applicant-driven (Kaewpijit *et al*, 2003)

Some proposed dimension reduction methods are based on stochastic theory such as the Principal Component Analysis (PCA), Discriminant Analysis Feature Extraction (DAFE) and Decision Boundary Feature Extraction (DBFE). These techniques are not so effective for dimension reduction of hyperspectral data for example DAFE and DBFE need to very large number of training samples for estimating the statistical properties of the hyperspectral data in the original feature space. PCA is effective at compression information in multivariate data sets by computing orthogonal projections that maximize the amount of data variance. It is typically performed through the egin-decompositon of the spectral covariance matrix of an image cube. The information can then be presented in the form of component images, which are projections of the image cube on to the eigenvectors, the component images corresponding to the large eigenvalues are presumed to preserve the majority of the information about the scene. Unfortunately information content in hyperspectral images dose not always coincide with such projections (Chang, 2000). This rotational transform is also time-consuming because of its global nature (Kaewpijit *et al*, 2003). Finally, since it is a global transformation, it does not preserve local spectral signatures and therefore might not preserve all information useful to obtain a good classification. For these reasons, some authors have proposed a dimension reduction method based on wavelet decomposition.

This paper attempts to transform the spectral data from the original feature space to a reduced feature space by using a discrete wavelet transform. The principle of this method is to apply a discrete wavelet transform to hyperspectral data in the spectral domain and at each pixel location. This does not only reduce the data, volume but it also can preserve the characteristics of the spectral of signature. This is due to intrinsic property of wavelet transforms of preserving of high and low frequency during the signal decomposition, therefore preserving peaks and valleys found in typical spectra. In addition, some of sub bands especially the low pass filter, can eliminate anomalies found in one of the bands.

Our experimental results for representative sets of hyperspectral data have confirmed that the wavelet spectral reduction as compare to PCA provides better or comparable classification accuracy while can reduce the computational requirement.

This paper is organized as follows. Section 2 provides an overview of the automatic multiresoluton wavelet analysis for dimension reduction of hyperspectral data. Section 3 discusses the automatic selection of level of decomposition. Section 4 presents results for the automatic wavelet reduction. This is accomplished by investigating the impact of the wavelet reduction on classification accuracies for different conventional classification methods and Section 5 provides our concluding remark for this work.

2. AUTOMATIC MULTIRESOLUTION WAVELET ANALYSIS

Wavelet transforms are the basis of many powerful tools that are now being used in remote sensing applications, e.g., compression, registration, fusion, and classification (Kaewpijit *et al*, 2003).Wavelet transform can provide a domain in which both time and scale information can be studied simultaneously giving a time-scale representation of the signal under observation. A wavelet transform can be obtained by projection the signal onto shifted and scaled version of a basic function. This function is known as the mother wavelet, $\Psi(t)$, A "mother wavelet" must satisfy this condition (Mathur, 2002).

$$c_\Psi = \int_{-\infty}^{\infty} \frac{|\Psi(s)|^2}{|s|} ds < \infty \qquad (1)$$

This condition implies that the wavelet has a zero average

$$\int_{-\infty}^{\infty} \Psi(x) dx = 0 \qquad (2)$$

And the shifted and scaled version of the mother wavelet forms a basis of functions. These basis functions can be represented as

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \Psi\left[\frac{t-b}{a}\right] \qquad (3)$$

where a represents the scaling factor and b the translation factor. Wavelet transforms may be either discrete or continuous. In this paper only Discrete Wavelet Transform (DWT) is considered. For dyadic DWT the scale variables are power of 2 and the shift variables are none overlapping and discrete.

One property that most wavelet systems satisfy is the multiresolution analysis (MRA) property. In this paper Mallat (1989) algorithm is utilized to compute these transforms.

Following the Mallat algorithm, two filters [the lowpass filter (L) and its corresponding highpass filter (H)] are applied to the signal, followed by dyadic decimation removing every other elements of the signal, thereby halving its overall length. This is done recursively by reapplying the same procedure to the result of the filter subbands to be an increasingly smoother version of the original vector as shown in figure2. In this paper, such 1-D discrete Wavelet transform will be used for reducing hyperspectral data in the spectral domain for each pixel individually. This transform will decompose the hyperspectral of each pixel into a set of composite bands that are linear, weighted combination of the original spectral bands. In order to control the smoothness one of the simplest and most localized Daubechies filter, called DAVB4 has been used. This filter has only four coefficients (Kaewpijit *et al*, 2003).

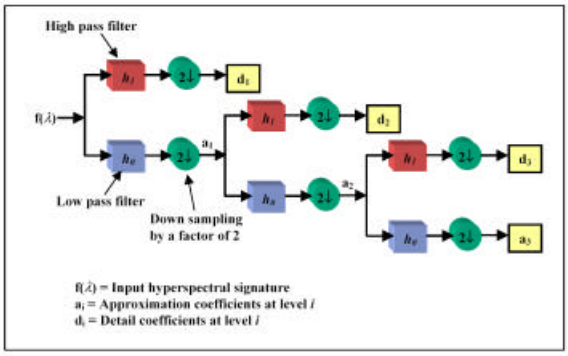

Figure2. A dyadic filter tree implementation for a level-3 DWT

An example of the actual signature of one pixel for 195 bands of the California 94 AVIRIS dataset and different level of lowpass component of wavelet decomposition of this spectral signature is shown in figure3. As s seen from this figure as the number of wavelet decomposition levels increases, the structure of the spectral signature becomes smoother than the structure of original signature.

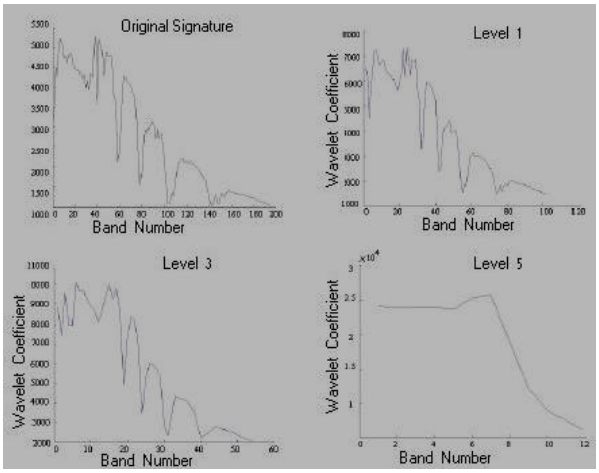

Figure3. Example of a pixel spectral signature and different levels of wavelet decomposition for the low pass component

In the algorithm of wavelet reduction we need to reconstruct the spectral signature to automatically select the number of levels of wavelet decomposition.

While wavelet decomposition involves filtering and down sampling the wavelet reconstruction involve up sampling and filtering. The up sampling process lengthens decomposed spectral data by inserting zeros as high pass component between each element.

3. WAVELET- BASED DIMENSION REDUCTION

3.1 General Description of Algorithm

Wavelet-Based reduction can be effectively applied to hyperspectral imagery. Performance of wavelet reduction can be better for larger dimensions (Kaewpijit *et al*, 2003). This property is due to very nature wavelet compression, where significant feature of the signal might be lost when the signal is under sampled. The general description of the wavelet reduction algorithm follows;

1. For each pixel in a hyperspectral scene, the 1-D signal corresponding to its spectral signature is decomposed using Daubechies wavelet.
2. For each hyperspectral pixel, approximation the original spectral is reconstructed using IDWT. The needed level of decomposition for a given pixel is the one that corresponds to producing an acceptable correlation whit the original signature.
3. Combining results from all pixels, the number of the level of decomposition (L) is automatically computed as the lowest level needed after discarding outliers.
4. Using the number of L computed in (3) the reduced output data are composed of all pixels decomposed to level L. Therefore, if the original number of bands was N the output number of bands is $N/2^L$.

3.2 Automatic Decomposed Level Selection

The correlation between the original spectral signature and the reconstructed spectral approximation is an indicator, which measures the similarity between two spectral signatures and used for selecting how many levels of decomposition can be applied while steel yielding good classification accuracy. The correlation function between the original spectral signature (x) and its reconstructed approximation (y) is shown in (Kaewpijit *et al*, 2003)

$$r(x,y) = \frac{\sum x_i y_i - (\frac{1}{N})\sum x_i \sum y_i}{\sqrt{(\sum x_i^2 - (\frac{1}{N})(\sum x_i)^2)(\sum y_i^2 - (\frac{1}{N})(\sum y_i)^2)}} \quad (4)$$

where N is the original dimension of the signal.

Table.1 shows the similarity between the original spectral signature and its reconstructed approximation of one class for the scene in our image. As seen from the table, as the number of levels of decomposition increases and the signal become more different from the original data, a proportionate decrease in correlation is observed. For each pixel in the hyperspectral scene and for each level of decomposition the correlation between original and reconstructed signal is computed. All correlation higher than the user-specified threshold contributes to the histogram for that level of decomposition. When all pixels are processed, the lowest level of wavelet decomposition needed to produce such correlation is used for the remainder of the algorithm.

Level	Correlation
1	*.9974*
2	*.9936*
3	*.9804*
4	*.9558*
5	*.9224*

Table1. Similarity Measures Between The Original Versus the Reconstructed Spectral Signature for one Class in our image.

4. WAVELET-BASED REDUCTION AND CLASSIFICATION ACCYRACY

We have experimentally validated the Wavelet Based dimension reduction by using remotely sensed image test from a hyperspectral scene, using the ENvironment for Visualizing Images (ENVI) as a tool for classification accuracy assessment. Using the wavelet-reduced data, an error (confusion) matrix of several classification methods for the same level of decomposition between the Wavelet and PCA was calculated.

Supervised classification methods are trained on labeled data. As the number of bands increases the number of training data for classification is increased too. In usual the minimum number of training data for each class is 10N, where N is the number of bands (Swain and Davis, 1978). The details about the number of training pixels are shown in Table2.

Class Name	Training data (NO.of pixels)
Wood	*1290*
Grass\paster	*467*
Soybeans-notill	*1108*
Corn	*700*
Corn-notill	*1527*
Hay_windrowed	*630*
Grass\trees	*868*
Alfalfa	*92*
Oatas	*303*
Grass\pastuer-moved	*2645*
Soybeans-clean	*710*
Corn-min	*921*

Table2. Number of training data for classification

Four statistical supervised classification methods are selected to test both PCA and the automatic wavelet reduction technique: *Maximum Likelihood (ML), Mahalanibis distance (MB), Minimum Distance (MD) and Parallelepiped (PP).*
In this work we used an image of a portion of the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) of hyperspectral data taken over an agricultural area of California, USA in 1994 (figure4). This image has 195 spectral bands about 10nm apart in the spectral region from 0.4 to 2.45µm with a spatial resolution of 20m.The test image has a pixel of 145 rows by 145 columns. And its corresponding ground truth map is involving 12 class. The number of training pixel for each class is in Table2.

Figure4. Test AVIRIS data. California 1994

The overall classification accuracies obtained from both of dimension reduction methods are listed in Table 3.
As shown in Table3 for ML algorithm the Wavelet reduction gives 95.73% overall accuracy for the first level of decomposition, while PCA only gives 95.3% .The same trend is seen for MB classification method and for all level of decomposition. The two other classification methods (MD and PP), are sometimes chose over the ML classification because of their speeds. Yet they are known to be much less accurate than the ML classification. Some authors believe that there are two main factors that make Automatic Wavelet Reduction outperform the PCA as follows (Kaewpijit *et al*, 2003).
1) The nature of classifiers, which are mostly pixel-based techniques and are thus well suited for Wavelet, which is pixel-based transformation.
2) The lowpass and some of highpass portions of the remaining information content, not includes in the firsts PCs, are still present in the Wavelet reduced data

Classification Method	Reduction Method	Classification accuracy (%)				
		No. Of Component/Level of Decomposition				
		101/1	54/2	30/3	18/4	12/5
Maximum Likelihood	**Wavelet**	*95.7356*	*92.5062*	*88.9342*	*84.6901*	*81.1451*
	PCA	*95.3119*	*91.3712*	*87.9837*	*85.3308*	*82.3436*
Mahalanobis Distance	**Wavelet**	*58.9236*	*58.6642*	*58.2423*	*57.4795*	*58.4189*
	PCA	*58.1104*	*56.9035*	*56.5256*	*55.3933*	*51.7098*
Minimum Distance	**Wavelet**	*40.6104*	*40.5617*	*40.4415*	*40.6796*	*39.6672*
	PCA	*40.5239*	*40.5140*	*40.5140*	*40.4842*	*40.4148*
Parallelepiped	**Wavelet**	*27.4137*	*27.0976*	*26.8934*	*26.8224*	*25.1447*
	PCA	*20.1573*	*21.1996*	*21.8945*	*22.1529*	*20.4247*

Table3. Classification result from comparing PCA and Wavelet Reduction

5. CONCLUSION

The high spectral resolution of hyperspectral data provides the ability for diagnostic identification of different materials. In order to analyze such hyperspectral data by using the current techniques and to increase the classification performance, dimension reduction is pre-processing for removing the redundant information substantially without sacrificing significant information and of course preserving the characteristics of the spectral signature. In this paper, we have presented an efficient dimension reduction technique for hyperspectral data based on automatic Wavelet decomposition. With a high number of bands produced from hyperspectral sensors, we showed that the Wavelet Reduction method yields similar or better classification accuracy than PCA. This can be explained by the fact that Wavelet reduced data represent a spectral distribution similar to the original distribution, but in a compressed form. Keeping only the approximation after Wavelet transform is a lossy compression as the removed high frequency signal (details) may contain useful information for class separation and identification. PCA also has a similar problem when not all the components are kept. This is, however the tradeoff when compression or reduction is used.

REFRENCES

[1] Fukunaga, K., 1989, "Effect of Sample Size in Classifier Design," IEEE Pattern Analysis and Machine Intelligence, Vol. 11, No 8, pp. 873-885.

[2] Hsieh, P.F., and Landgrebe, D.A., 1998 "Classification of High Dimensional Data", School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana, USA.

[3] HSU, P.H., and TSENG, Y.H., 2000, "Wavelet Based Analysis of Hyperspectral Data for Detecting Spectral Features," International Archives of Photogrammetry and Remote Sensing .Vol. XXXIII, Supplement B7. Amsterdam.

[4] Ifarraguerri,A., and Chang, C.I.,2000 "Unsupervised Hyperspectral Image Analysis With Projection Pursuit,"IEEE Geosci.Remote Sensing, Vol.38, No. 6, November 2000.

[5] Kaewpijit,S., Moigne,J.L. and EL-Ghazawi,T.,2003 "Automatic Reduction of Hyperspectral Imagery Using Wavelet Spectral Analysis," IEEE Geosci. Remote Sensing, Vol.41, No.4.

[6] Mathur, A., 2002, "Dimensionality Reduction of Hyperspectral Signatures Detection of Invasive Species", Department of Electrical and Computer Engineering, Mississippi State University, Mississippi.

[7] Mallat, S.G., 1989 "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693.

[8] Richards, J.A., 1993 *Remote Sensing Digital Image Analysis: An Introduction*, 2nd Ed. New York: Springer-Verlag.

[9] Salehi, B., Valadan Zouj, M.J., 2004,"Studying the Analyzing Methods for Classification of Hyperspectral Imagery", Faculty of Geodesy & Geomatics Engineering, K.N.T. University of Technology, Tehran, Iran .

[10] Swain, P.H., and Davis, S.M., 1978, *Remote Sensing: The Quantitative Approach*. New York: McGraw-Hill.