

# Wavelet Spectral Analysis for Automatic Reduction of Hyperspectral Imagery

Bahram Salehi  
[Salehi\\_Bahram@yahoo.com](mailto:Salehi_Bahram@yahoo.com)

M.J.Valadan Zouj  
[Valadanzouj@kntu.ac.ir](mailto:Valadanzouj@kntu.ac.ir)

Faculty of Geodesy and Geomatics Engineering, K.N.Toosi University of Technology, Tehran Iran

Fax:0098218779476

**Bahram salehi** received the B.S degree in Geomatic engineering from Tehran University Faculty of Geodesy and Geomatics ,Iran ,in 2001.Now he is student of M.S. in Remote Sensing engineering at K.N.Toosi University of Technology, Iran

**Dr.M.J.ValadanZouj** is the head of Faculty of Geodesy and Geomatics ,K.N.Toosi University of Technology

**Keywords:** remote sensing, hyper spectral, analysis, classification, reconstruction, spectral

## Abstract:

With the number of channels in the hundreds instead of in the tens hyperspectral imagery possesses much richer spectral information than multispectral imagery .The increased dimensionality of such hyperspectral data provides a challenge to the current technique for analyzing data.

Conventional classification methods may not be used without dimension reduction preprocessing. Supervised classification techniques use labelled samples in order to train the classifier. Usually the number of such samples is limited and as the number of bands available increases this limitation becomes severe and can become dominates over the projected added value of having the additional bands available.

So because of these problems dimension reduction has become a significant part of hyperspectral image interpretation. Some techniques are used for dimension reduction such as Principle Component Analysis (PCA) and Minimum Noise Fraction (MNF) but these techniques are not useful for dimension reduction. For example PCA is computationally expensive and dose not eliminate anomalies that can be seen at one arbitrary band. MNF has also the similar problems.

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 has more separate classes and yields better or comparable classification accuracy.

## 1. Introduction

The recent development of more sophisticated remote sensing systems enable the measurement of radiation in many more spectral interval than possible previously. An example of this technology is the AVIRIS system, which collects image data in 220 bands.

The increased dimensionality of such hyperspectral data provides a challenge to the current techniques for analyzing such data. Supervised classification techniques use labeled samples in order to terrain the classifier. 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. This suggests the need for reducing the 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 is applicant-driven.

A number of techniques have been developed to reduce dimensionality. One of these techniques is Principle Component Analysis (PCA) .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-decompositopn 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 for several reasons.

1) PCA is optimal when the background associated with signal sources is Gaussian white noise. This is often not the case in hyperspectral images, where the clutter includes contributions from interface sources such as natural background signatures as well as structured (nonrandom) noise such as stripping, when the Gaussian assumption dose not hold, the background clutter can become indistinguishable from signal source.

2) The objects of interests are often small relative to the size of the scene, and therefore contribute a small amount to the overall variance. PCA often fails to capture the variability associated with small objects unless their spectra are nearly orthogonal to the background spectra.

An alternative to PCA that can alleviate some these problems is Minimum Noise Fraction (MNF) transform. This transform, also known as noise-adjusted principle components, was designed to produce orthogonal component images that are ordered by image quality as measured by the SNR, rather than the data variance.

The MNF transform is equivalent to sequence of two orthogonal transformations where the first rotates the data such that the noise covariance matrix is diagonalized, thus “whitening” the noise, followed by a standard PCA transform. PCA performance is improved because the noise effects on signal source are minimized by the whitening process, however the MNF transform still depend on “bulk” image properties, so it is not generally sensitive to small objects. For these reasons it is proposed a new dimension reduction method based on wavelet decomposition .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 dose 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 subbands 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 classification accuracy.

This paper is organized as follows. Section 2 provides an overview of the automatic multiresolution 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

Multiresolution 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

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

This condition implies that the wavelet has a zero average

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

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

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

Where “a” represents the scaling factor and “b” the translation factor.

Wavelet transforms may be either discrete or continuous. In this paper only discrete orthonormal based of wavelet are considered. For dyadic DWT the scale variables are power of 2 and the shift variables are non overlapping and discrete.

One property that most wavelet systems satisfy is the multiresolution analysis (MRA) property. In this paper Mallat 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 it is 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 Fig.1. 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 must localized Daubechies filter, called DAVB4 has been used. This filter has only four coefficients.

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 Fig.2. As seen from the Fig.2 as the number of wavelet decomposition levels increases, the structure of the spectral signature become smoother than the structure of original signature.

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 downsampling the wavelet reconstruction involve upsampling and filtering. The upsampling process lengthens decomposed spectral data by inserting zeros as highpass component between each element. The inverse wavelet transform is given by:

$$f(t) = \sum_{k=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} w(k,i) 2^{\frac{-k}{2}} \Psi(2^{-k} t - i)$$

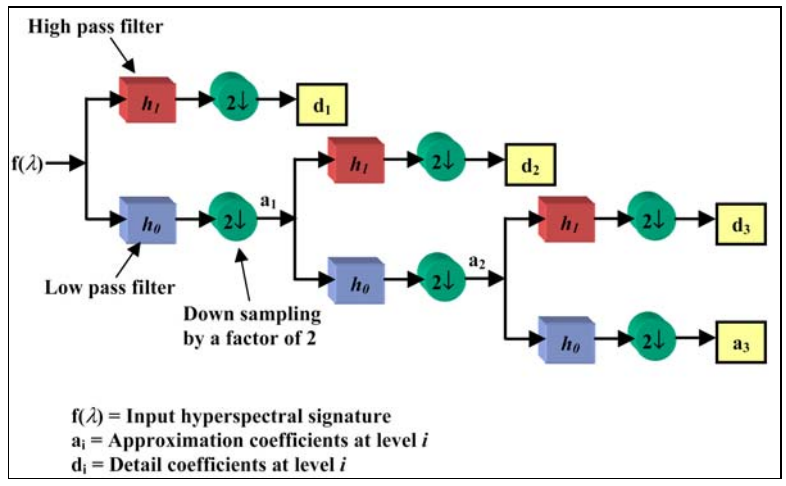


Fig.1. A dyadic filter tree implementation for a level-3 DWT

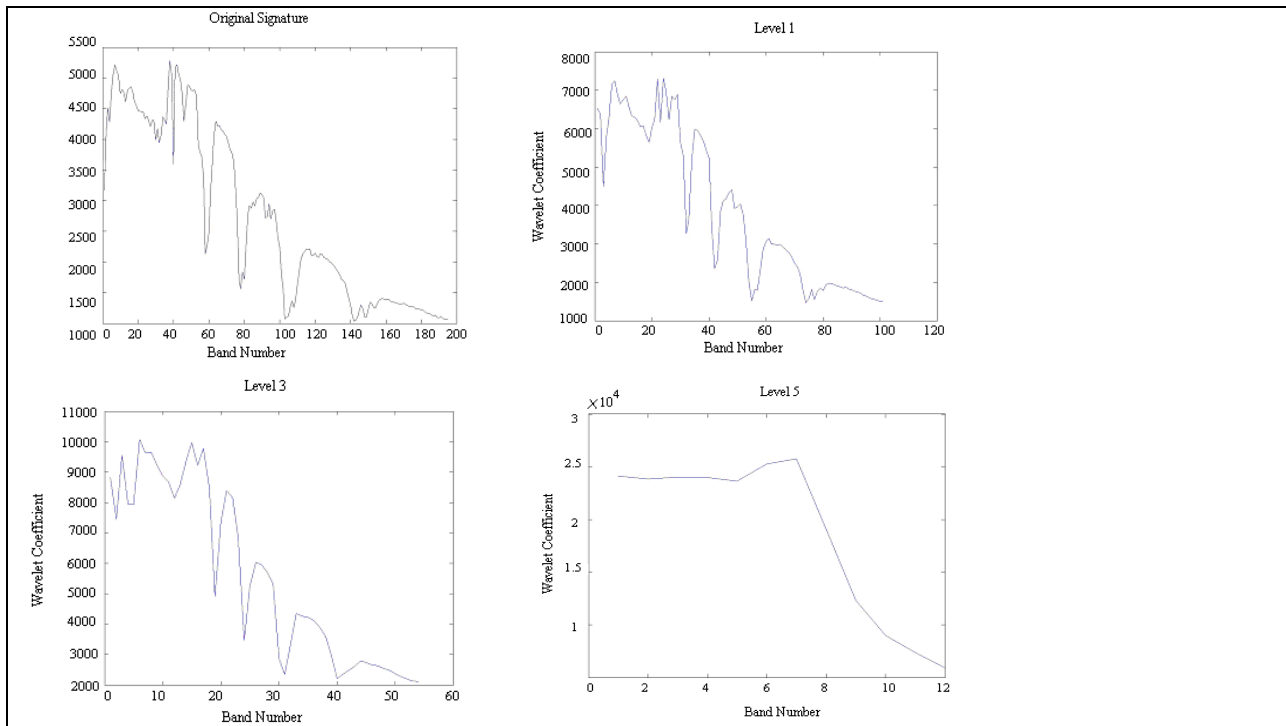


Fig.2. Example of a pixel spectral signature and different levels of wavelet decomposition for the lowpass component

### 3. Wavelet-Based Dimension Reduction

#### A. General Description of Algorithm

Wavelet-Based reduction can be effectively applied to hyperspectral imagery. Performance of wavelet reduction can be better for larger dimensions. 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  $N/2^L$

#### B. 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

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

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 reminder 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.

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

#### 4. Wavelet-Based Reduction and Classification Accuracy

We have experimentally validated our Wavelet Based dimension reduction by using remotely sensed image tests from a hyperspectral scene. Using the ENvironment for Visualizing Images (ENVI) as a tool for classification accuracy to assess the accuracy of our Wavelet Based method and PCA we calculated the error (confusion) matrix of several classification methods for the same level of compression between the Wavelet and PCA.

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. The details about the number of training pixels are shown in Table2.

Now it is the turn of introducing four statistical supervised classification methods that we used in this work.

##### *Maximum Likelihood (ML)*

This method assumes that the statistical for each class in each band are normally distributed and computes all of the probability of classes for each pixel and assign that pixel to the class with the highest probability value.

##### *Mahalanis distance (MB)*

This algorithm is similar to ML algorithm but in MB algorithm it is assumed that the covariance matrix of all classes is equivalent and of course this algorithm is faster than ML algorithm.

##### *Minimum Distance (MD)*

In this algorithm for each training sample region the mean is calculated and also the Euclidean distance between each pixel and the means of each training sample region is calculated. The pixel is labeled to that class that has minimum distance.

##### *Parallelepiped (PP)*

The parallelepiped classifier is a very simple classifier, in this algorithm the range in all bands describes a multidimensional box or parallelepiped, if, on classification, pixel are found to lie in such a parallelepiped they are labeled as belong to that class.

#### Data Source and Study Area

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 (Fig.3). This image has 195 spectral bands about 10nm apart in the spectral region from 0.4 to  $2.45\mu\text{m}$  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.



Fig.3. Test AVIRIS data. California 1994

## Experimental results

In this work we used two software for implementation of our research.

MATLAB 6.5: This software has been used for implementation of Wavelet reduction and PCA algorithms.

ENVI 3.5: This software also was used for assessing the effect of Wavelet reduction algorithm in overall accuracy of different classification method compare to PCA. The overall classification accuracies obtained from both of dimension reduction methods are listed in Table4. As shown in Table4 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.

- 1) The nature of classification, 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.

## References

- [1] J.A. Richards, *Remote Sensing Digital Image Analysis: An Introduction*, 2<sup>nd</sup> Ed. New York: Springer-Verlarrg, 1993.
- [2] S.Kaewpijit, J.L.Moigne and T.EL-Ghazawi, "Automatic Reduction of Hyperspectral Imagery Using Wavelet Spectral Analysis," *IEEE Geosci. Remote Sensing*, Vol.41, No.4, April 2003.

[3] L.O.Jimenez and D.Landgrebe, "High dimensional Feature Reduction Via Projection Pursuit," School of Electrical and Computer Engineering Purdue University West Lafayette, April 1995.

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

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

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

[7] H.Emami, "Evaluation and Decomposition of Mixed Pixel in Remotely Sensed Images for Accuracy Improvement of Classification Results",A thesis for MS. Degree at K.N.Toosi University of Technology,Tehran,Iran,2002.