

# Spectral Data Reduction via Wavelet Decomposition

S. Kaewpijit <sup>a</sup>, J. Le Moigne <sup>b</sup>, and T. El-Ghazawi <sup>c</sup>

<sup>a</sup> School of Computational Sciences, George Mason University  
4400 University Drive, Fairfax, Virginia 22030 USA.

<sup>b</sup> NASA Goddard Space Flight Center  
Applied Information Science Branch, Code 935  
Greenbelt, Maryland 20771 USA

<sup>c</sup> Department of Electrical and Computer Engineering,  
The George Washington University  
Washington DC, 20052 USA

## ABSTRACT

The greatest advantage gained from hyperspectral imagery is that narrow spectral features can be used to give more information about materials than was previously possible with broad-band multispectral imagery. For many applications, the new larger data volumes from such hyperspectral sensors, however, present a challenge for traditional processing techniques. For example, the actual identification of each ground surface pixel by its corresponding reflecting spectral signature is still one of the most difficult challenges in the exploitation of this advanced technology, because of the immense volume of data collected. Therefore, conventional classification methods require a preprocessing step of dimension reduction to conquer the so-called "curse of dimensionality."<sup>1</sup> Spectral data reduction using wavelet decomposition could be useful, as it does not only reduce the data volume, but also preserves the distinctions between spectral signatures. This characteristic is related to the intrinsic property of wavelet transforms that preserves high- and low-frequency features during the signal decomposition, therefore preserving peaks and valleys found in typical spectra. When comparing to the most widespread dimension reduction technique, the Principal Component Analysis (PCA), and looking at the same level of compression rate, we show that Wavelet Reduction yields better classification accuracy, for hyperspectral data processed with a conventional supervised classification such as a maximum likelihood method.

**Keywords:** Remote Sensing, Dimension Reduction, Wavelet Decomposition, and Maximum Likelihood.

## 1. INTRODUCTION

Hyperspectral imagery provides richer information than traditional multispectral imagery. Once the diagnostic spectral signatures are extracted from the hyperspectral data, the difficulty of actually identifying the material reflecting these spectral signatures prevents the full potential of hyperspectral technology from being realized. Furthermore, it is clear that more effective data processing techniques are needed to deal with hyperspectral cubes. Because it is necessary to have a minimum ratio of training pixels to the number of spectral bands in order to ensure a reliable estimate of class statistics [1], dimension reduction has become a significant part of the hyperspectral image interpretation. Dimension reduction is the transformation that brings data from a high order dimension to a low order dimension, thus conquering the curse of dimensionality [2]. Dimension reduction is becoming an even more important issue due to the fact that the first few space-borne hyperspectral sensors are currently in orbit, producing great amounts of data.

In remote sensing, one of the most widely used dimension reduction techniques is the Principal Component Analysis. Principal Component Analysis is a popular technique for eliminating redundancy in the data. PCA is based on decorrelation and obtains redundancy reduction by discarding low variance components, but this rotational transform is

---

<sup>1</sup> The curse of dimensionality refers to the fact that the sample size needed to estimate a function of several variables to a given degree of accuracy grows exponentially with the number of variables.

time-consuming because of its global nature. Moreover, 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, we are proposing a new dimension reduction method based on wavelet decomposition.

The principle of our method is to apply a discrete wavelet transform to hyperspectral data in the spectral domain and at each pixel. This not only reduces the data volume, but it also preserves the distinctions between spectral signatures. This characteristic is related to the intrinsic property of wavelet transforms of preserving high- and low-frequency features during the signal decomposition, therefore preserving peaks and valleys found in typical spectra.

The paper is organized as follows: Section 2 provides an overview of the wavelet decomposition that we will be using for dimension reduction. Section 3 discusses the computational complexity of our wavelet-based dimension reduction method compared to Principal Component Analysis. Section 4 presents the experimental results, including the hyperspectral test data and classification accuracies for different conventional classification methods.

## 2. MULTI-RESOLUTION WAVELET DECOMPOSITION

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. Using Mallat algorithm [3], discrete wavelet transforms (DWTs) can be computed very fast. In this paper, we will only consider discrete wavelets, particularly those expressed as orthonormal bases. The discrete wavelet transform is a fast, linear operation that operates on a data vector whose length is an integer power of two, transforming it into a numerically different vector of the same length [3][4]. The output represents wavelet coefficients at different positions and scales.

Level

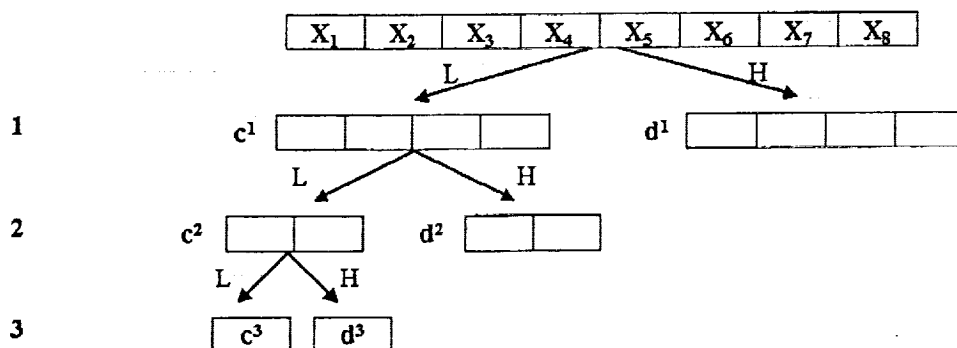


Figure 1. The fast DWT

Figure 1 shows the principle of Mallat decomposition: an orthonormal basis of wavelets is defined by a scaling function and its corresponding conjugate filter  $L$ . The two associated filters, the low-pass filter ( $L$ ) and its corresponding high-pass filter ( $H$ ) are applied to the signal, followed by dyadic decimation removing every other element of the signal and thereby halving its overall length. It operates recursively by then reapplying the same procedure to the result of the  $L$  filter, as shown in Figure 1. In the original vector  $x$  of length  $N = 2^J$  from Figure 1, the  $j$ th iteration produces the smoothed coefficients at scale  $j$ ;  $c_j = L^j x$  for  $j=1, \dots, J$ . This application of the low-pass filter ( $L$ ) causes  $c^j$  to be an increasingly smoother version of the original vector.

In this paper, such a 1-D discrete wavelet transform will be used for reducing hyperspectral data in the spectral domain at each pixel individually. This transform will decompose the hyperspectral signature of each pixel into a set of composite bands which are linear, weighted combinations of the original spectral bands.

One of the filters from the families of filters defined by Daubechies [6] has been applied to this work. Daubechies's families are compactly supported in the time-domain and have good frequency domain decay. The time-domain compact support enables the employment of fast recursive filter-bank algorithms using filters with a finite number of coefficients. Another advantage is that it is possible to control the smoothness of the analyzing wavelets. One of the simplest and most localized filter, often called DAUB4, has been used [5][6]. This filter has only four coefficients. DAUB4 has the peculiar property that its derivative exists almost everywhere. The result at each point represents a moving average of four points. Figure 2 shows an example of the actual signature of one class (Corn-notill) for 192 bands of the Indian Pines'92 AVIRIS dataset, and different levels of wavelet decomposition of this spectral signature. When the number of bands is reduced, the structure of the spectral signature becomes smoother than the structure of the original signature, and an important issue is to determine how many levels of decomposition can be applied while still yielding good classification accuracy.

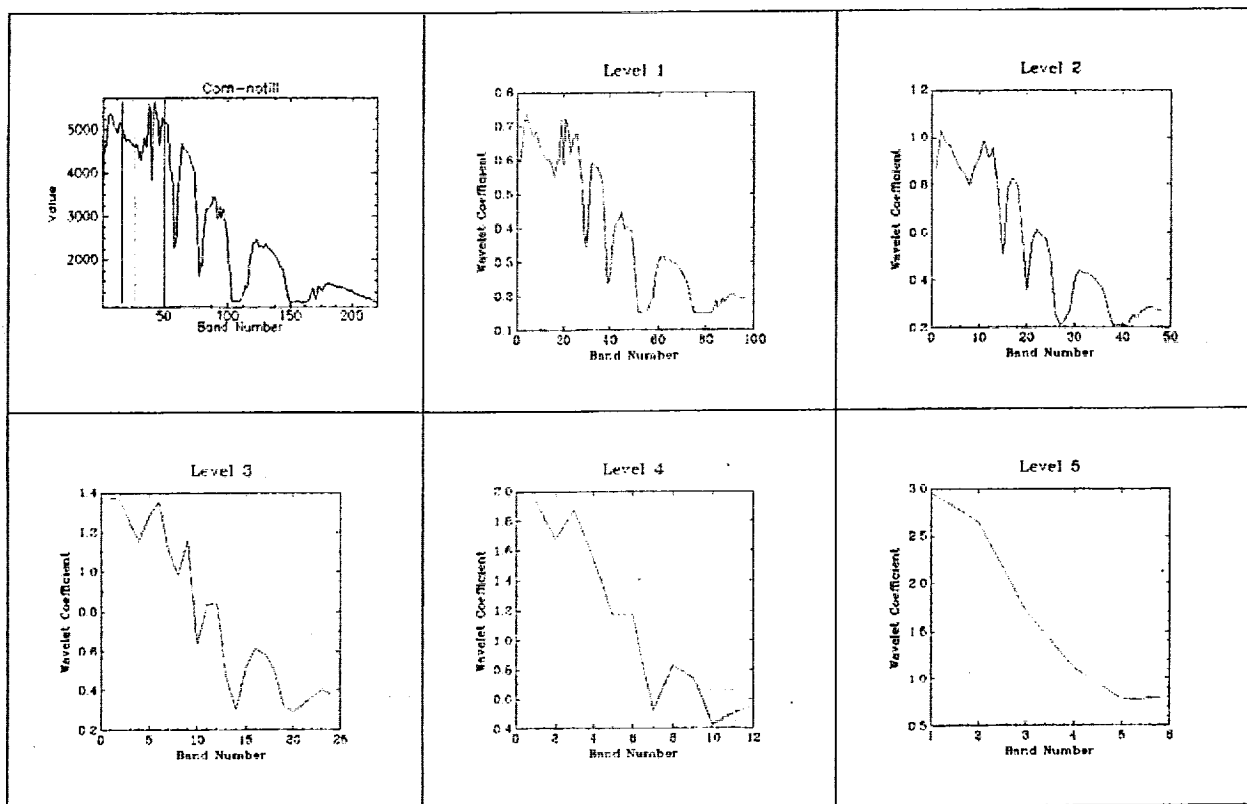


Figure 2. An example of the Corn-notill spectral signature and different levels of wavelet decomposition

### 3. REDUCTION EFFICIENCY ANALYSIS

Once the 1-D wavelet decomposition is applied to each pixel signature, the next step is to select the fewest wavelet coefficients required to give the highest accuracy for different conventional classification methods. These wavelet coefficients yield a reduced-dimensional data set that can be used as input for supervised classifications.

From a complexity point of a view, for a filter of length  $N_L$ , a wavelet decomposition requires in the order of  $N_L$  operations per invocation. After the first invocation of the low-pass filter (L) we obtain half the number of pixels, and then apply the low-pass filter again. Thus, each level processes half the number of pixels than the previous level. Since  $N_L$  is fixed for any particular wavelet filter, the wavelet-based reduction method yields the order of  $O(N)$  ( $N$  is the

number of bands) computations per pixel, which is extremely favorable [6][7]. Therefore, the whole algorithm complexity is in the order of  $O(MN)$ , where  $M$  is the number of pixels in the spatial domain.

On the other hand, the PCA computational complexity of an  $M$  pixels image of  $N$  spectral bands can be computed as follows:

- (1) Find the mean vector:  $O(MN)$ ;
- (2) Assemble the covariance matrix:  $O(MN^2)$ ;
- (3) Perform eigenanalysis, i.e. generate the transformation matrix used to compute the eigenvectors with a Jacobi method:  $O(N^3)$ .
- (4) Perform pixel-by-pixel linear transformation:  $O(RMN)$ .

Therefore, the total estimated time complexity of a PCA is  $O(MN^2 + N^3)$ , where  $M$  is the number of pixels of the image data,  $N$  is the number of bands, and  $R$  is the number of formed components ( $R \leq N$ ) [8][9].

An example of computational efficiency is shown when reducing the Indian Pines'92 AVIRIS data (192 Bands, 145x145 pixels, ~8MBytes) to the third level of wavelet decomposition compared to 24 Principal Components. The wavelet-based reduction takes only 1.907 seconds, while PCA is much more time-consuming with 20.309 seconds (including IO operations). These timing results are obtained on a Pentium III 450 MHz Linux-based Workstation. The wavelet-based reduction has therefore decreased the computational workload and is faster than PCA by about 10 times.

## 4. EXPERIMENTAL RESULTS

### 4.1. Experiment Setup and Hyperspectral Test Data

We have experimentally validated our wavelet-based dimension reduction by using remotely sensed image test suites from 2 hyperspectral scenes, and then using ENVI (the Environment for Visualizing Images) as a tool for classification assessment. ENVI is a state-of-the-art image processing system designed to provide a comprehensive analysis of satellite and aircraft remote sensing data, especially hyperspectral data [10]. It includes functions useful for multispectral and hyperspectral classification analysis that enable to select samples (points or regions) for training and testing pixels. Then, a confusion matrix is computed in order to assess the classification accuracy. We used the same level of compression as the basis comparison between the two methods. For example, the first level of decomposition (decimated by 2 from 192 bands of original data) is compared to 96 Principal Components (PCs), the second level to 48 PCs, and so on.

Supervised classification algorithms are trained on labeled data, so they are able to identify the class to which a pixel or a region belongs and thus provide a high-level characterization of the data [11]. Three conventional supervised classification methods are selected to test both PCA and our Wavelet Reduction technique.

**Maximum Likelihood (ML):** This classification method assumes that the statistics for each class in each band are normally distributed, and calculates the probability that a given pixel belongs to a specific class [1][10]. A maximum likelihood classification involving  $N$  spectral bands and  $C$  classes requires  $CPN(N+1)$  multiplications where  $P$  is the numbers of pixels in the image of interest [1].

**Minimum Distance:** The minimum distance classification uses the mean vectors of each training sample region and calculates the Euclidean distance from each unknown pixel to the mean vector for each class [1][10].

**Parallelepiped:** The parallelepiped classification uses a simple decision rule to classify multispectral data. The decision boundaries form an  $n$ -dimensional parallelepiped in the image data space. The dimensions of the parallelepiped are defined based upon a standard deviation threshold from the mean of each selected class [1][10].

Two hyperspectral data sets used in the experiment are as follows:

- 1) **Indian Pines'92:** The first dataset is a subset scene of the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) data, Indian Pines'92. This AVIRIS spectrometer has a ground pixel size of 17m x 17m, and a spectral resolution of 224 channels, covering the range from 400 nm to 2500 nm, centered at 10 nm intervals. We focus on the farmland scene taken June 12, 1992 in the northern part of Indiana, which consists of 145x145 pixels by 192 bands of radiance data as shown in Figure 3a. For this scene, the ground truth covers 49% of the

full 145x145 scene and is divided among 16 classes. In this work, we selected 9 classes from the 16 classes to test our techniques. The 9 classes are Corn-notill, Corn-min, Grass/Pasture, Grass/Trees, Hay-windowed, Soybeans-min, Soybean-clean, Woods, and Soybeans-notill ranging in size from 489 pixels to 2468 pixels. A random training sample of 20% of the pixels was chosen from the known ground truth from each class. The trained classifiers were applied to the remaining 80% of the known ground pixels in the scene [12][13].

- 2) **Salinas'98:** This AVIRIS dataset was acquired on October 9, 1998, South of the city of Greenfield in the Salinas Valley in California. It includes vegetables, bare soils, and vineyard fields as shown in Figure 3b, and consists of 217x512 pixels by 192 bands of radiance data. We selected 9 classes as follows: grapes-vineyard, broccoli-weed1, fallow\_smooth, soil-vineyard\_develop, fallow, stubble, celery, broccoli\_weed2, and corn\_senesced. The training samples were chosen as 5 percent of the pixels from the known ground truth over the entire scene. The trained classifiers were applied to the remaining 95 percent of the scene.

Since training samples were selected randomly, the procedure of selecting samples was repeated three times for both scenes [13].

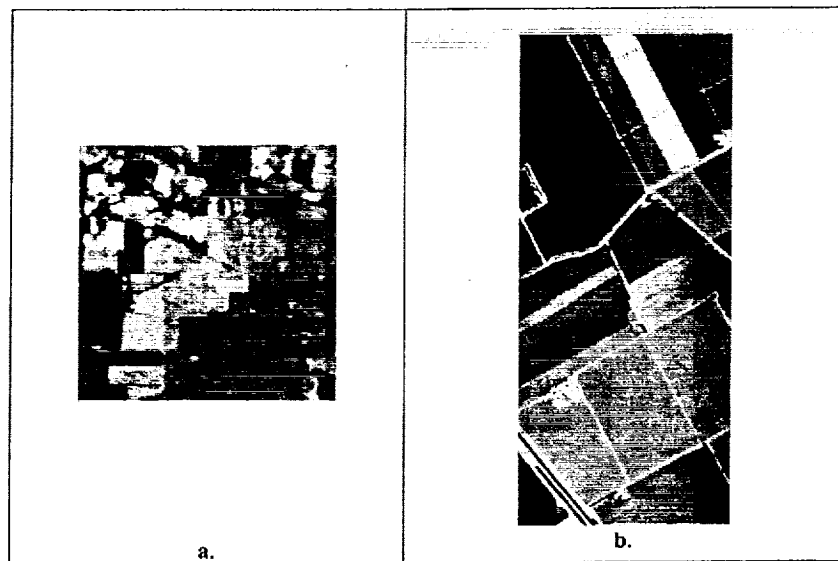


Figure 3. Test AVIRIS Data: a.) Indian Pines'92; b.) Salinas'98

#### 4.2. Experimental Results

The results show that our new dimension reduction method, the wavelet-based technique, provides a greater computational efficiency as well as a better overall classification accuracy than the widely used PCA method. The overall classification accuracies obtained from both dimension reduction methods are listed in Tables 1 and 3. For the Indian Pines'92 dataset and the ML classification, it is shown that the Wavelet Reduction gives 82.4 percent overall accuracy for the third level of decomposition, while PCA only gives 72.2 percent for 24 PCs for the ML classification. The same trend is seen for the Salinas'98 scene, in which wavelet gives 98.6 percent, while PCA gives 98.3 percent. Tables 2 and 4 show the complete results with confusion matrices of the classified wavelet coefficients at the third level of decomposition, equivalent to 24 PCs, for the testing areas of both data sets with the ML classification [14].

The two other classification methods, Minimum Distance and Parallelepiped, are sometimes chosen over the Maximum Likelihood classification because of their speed, but they are known to be much less accurate than the ML classification. Therefore, as expected, when comparing PCA to Wavelet Reduction, it can be seen that both minimum distance and parallelepiped classifiers provide significantly lower accuracy (below 50%) than maximum likelihood after dimensionality reduction. It should also be noted that, for the ML classifications of Tables 1 and 3, the classification accuracy at the first level of decomposition (or 96 PCs) is significantly lower because of Hughes phenomenon, demonstrating a loss of classifier performance with higher dimensionality.

Classification Method	Reduction Method	Classification Accuracy				
		No. of Component/Level of Decomposition				
		6/5	12/4	24/3	48/2	96/1
Maximum Likelihood	PCA	72.6214	73.0496	72.2066	70.186	15.3486
	Wavelet	75.1372	78.429	75.2481	81.5335	15.3486
Minimum Distance	PCA	42.3659	42.6602	42.968	43.0483	43.142
	Wavelet	41.1214	41.9912	42.2722	42.3792	42.3926
Parallelepiped	PCA	36.9865	37.2541	36.8259	35.8624	31.1521
	Wavelet	32.5037	32.7579	33.7883	33.4404	33.0925

Table 1. Classifications results after PCA vs. Wavelet Reduction using different levels of decomposition (Indian Pines'92)

Class	User's Acc.	Corn-notill	Corn-min	Grass/Pasture	Grass/Trees	Hay-windrowed	Soybeans-min	Soybeans-clean	Woods	Soybeans-notill	Row Total
Corn-notill	78.84	924	72	0	0	0	180	1	0	15	1172
Corn-min	82.18	8	442	0	0	0	74	14	0	0	538
Grass/Pasture	97.86	0	0	334	0	0	4	0	4	0	342
Grass/Trees	91.88	10	1	10	577	3	12	0	8	7	628
Hay-windrowed	99.74	0	0	0	0	388	0	1	0	0	389
Soybeans-min	73.75	151	134	32	17	0	1447	45	0	135	1982
Soybeans-clean	70.93	4	14	7	0	0	143	410	0	0	578
Woods	99.8	0	0	1	1	0	0	0	1023	0	1025
Soybeans-notill	73.42	50	4	13	2	0	134	20	0	616	839
Column Total		1147	987	387	597	391	1974	491	1035	774	7473
Producer's Acc.		80.56	88.27	84.13	96.65	98.23	73.3	83.5	98.84	79.59	
OVERALL ACCURACY	82.44										

Table 2. Confusion Matrix using the third level of wavelet decomposition and the ML classification (Indian Pines'92)

Classification Method	Reduction Method	Classification Accuracy				
		No. of Component/Level of Decomposition				
		6/5	12/4	24/3	48/2	96/1
Maximum Likelihood	PCA	97.2278	97.9587	98.3321	98.2313	40.3501
	Wavelet	97.8968	98.5979	97.9587	98.5681	40.3501
Minimum Distance	PCA	93.7844	93.8096	93.821	93.821	93.8187
	Wavelet	93.5736	93.6927	93.8086	93.821	93.821
Parallelepiped	PCA	81.7678	82.2054	80.2053	74.638	64.8277
	Wavelet	76.5488	76.4915	76.2326	74.8213	73.6689

Table 3. Classifications results after PCA vs. Wavelet Reduction using different levels of decomposition (Salinas'98)

Class	User's Acc.	grapes-vineyard	brocoli_weeds1	fallow_smooth	soil_vineyard_develop	fallow	stubble	celery	brocoli_weeds2	corn_senesced	Row Total
grapes-vineyard	98.09	17587	0	78	14	1	25	12	7	201	17829
brocoli_weeds1	100	0	1858	0	0	0	0	0	0	0	1858
fallow_smooth	99.92	0	0	2449	0	2	0	0	0	0	2451
soil_vineyard_develop	99.57	0	0	0	5741	0	0	0	0	25	5766
fallow	99.1	0	0	17	0	1889	0	0	0	0	1889
stubble	99.95	0	0	2	0	0	3738	0	0	0	3738
celery	100	0	0	0	0	0	0	3388	0	0	3388
brocoli_weeds2	98.52	0	53	0	0	0	0	0	3532	0	3585
corn_senesced	94.72	25	0	0	133	3	0	0	0	2889	3049
Column Total		17812	1909	2544	5992	1877	3781	3400	3538	3114	43848
Producer's Acc.		99.86	97.22	98.27	97.44	99.57	99.34	99.65	99.8	92.74	
OVERALL ACCURACY	98.82										

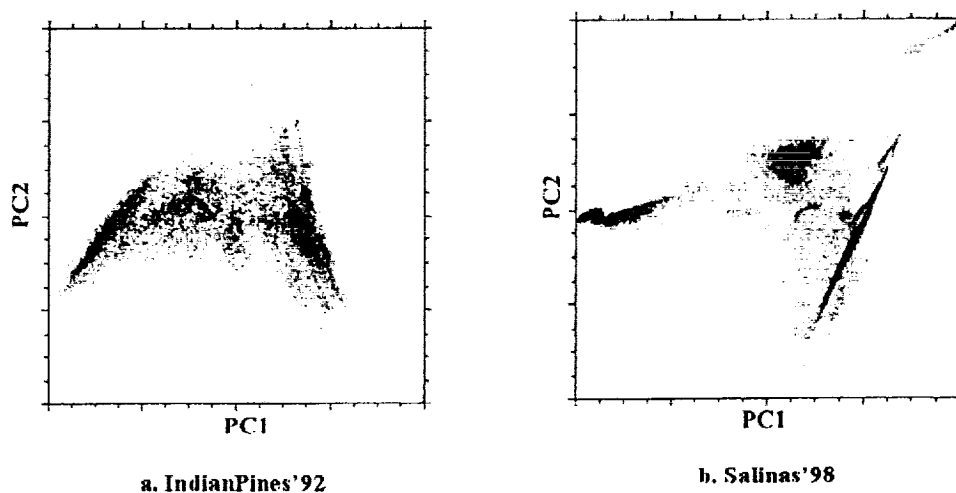
Table 4. Confusion Matrix using the third level of wavelet decomposition and the ML classification (Salinas'98)

Since the Principal Components transformation is based on a global transformation of the hyperspectral data, while Wavelet Reduction, on the other hand, is based on a pixel by pixel wavelet decomposition, this new wavelet-based method is not explicitly sensitive to the class structure in the scene. Moreover, while wavelets preserve the peaks and valleys of spectral signatures, PCA, however, provides separability of the classes, and can still be useful when classes are well distributed in the direction of the first few principal axes [1]. Figure 4 (scatter plot of PC1 and PC2 for both scenes) shows that the spectral class structure of the Salinas scene is better separated along the two first principal axes than is the Indian Pines scene [15]. This could explain why the classification accuracy obtained after PCA reduction is better for the Salinas scene than for the Indian Pines scene. It could also explain why the ML classification results obtained with the

Wavelet Reduction approach are significantly better than PCA when applied to the Indian Pines scene. In such a case, the wavelet decomposition technique seems to be a better approach.

Table 5 shows information contents (cumulative percentage of total data variation) for both the Indian Pines'92 and the Salinas'98 datasets. It is clear that 24 PCs of the Salinas'98 scene, containing a total data variation of 99.99% are sufficient to produce a good classification accuracy. On the other hand, 24 PC's of the Indian Pines'92 dataset contain only 96.60% of the total data variation, and in this case, the Wavelet Reduction method outperforms the PCA. As a summary, the two factors that explain the results are:

- (1) the nature of the classifiers, most of which are pixel-based techniques [16], might explain why wavelets, which are pixel-based transformations, work better than PCA.
- (2) the remaining information content, not included in the first PCs (such as for the Indian Pines'92 dataset), contains information that is hidden by noise but that differentiates the classes. This information, not contained in the first PCs, is still present in the wavelet reduced data.



**Figure 4.** The scatter plot of PC1 vs. PC2 for both data sets

No. of Components	Information Contents (%)	
	IndianPines'92	Salinas'98
6	94.89	99.92
12	95.61	99.97
24	96.6	99.99
48	97.94	99.99
96	99.37	~100

**Figure 5.** The cumulative percentage of total data variation at different number of Principal Components

## 5. CONCLUSIONS

In this paper, we have presented a fast and efficient dimension reduction technique for hyperspectral data based on wavelet decomposition. Both analytical assessment of time complexity and experimental results of classification accuracy have proven that the 1-D wavelet-based dimension reduction technique is a useful method for reducing

dimensionality of hyperspectral data. On our two datasets, we showed that the Wavelet Reduction method yields similar or better classification accuracy than the 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. The Wavelet Reduction method also fits very well in the ML classification process for which it yields results superior to PCA. Furthermore, at a decomposition level similar in compression rate to the number of PCAs, Wavelet Reduction is more efficient from a computational point of view.

The best results were obtained by utilizing the third level of wavelet decomposition, and as mentioned earlier, although the first and second level decomposition better represent the original spectral distribution, larger data dimensions cause a loss in classifier performance.

Future research will involve considering a trade-off method between Wavelet Reduction and PCA, and developing a parallel hybrid dimension reduction algorithm.

### ACKNOWLEDGEMENTS

The authors wish to acknowledge the support of the ESS/HPCC Program and would like to thank Tony Gualtieri for providing test hyperspectral data.

### REFERENCES

1. J. A. Richards, *Remote Sensing Digital Image Analysis: An introduction*, Spring-Verlag, Berlin, Heidelberg, 1986.
2. D. W. Scott, "The Curse of dimensionality and Dimension Reduction," in *Multivariate Density Estimation: Theory, Practice, and Visualization*, Chapter 7, pp. 195-217, John Wiley and Sons, Inc., 1992.
3. S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Pattern Anal. and Machine Intell.*, vol. 11, no. 7, pp 674-693.
4. M. A. Cody, "The Fast Wavelet Transform," *Dr. Dobbs's Journal*, April 1992.
5. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, New York, NY, 1992.
6. I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA, 1992.
7. L.M. Bruce, J.Li, "Fast Wavelet-Based Algorithms for Multiresolutional Decomposition and Feature Extraction of Hyperspectral Signatures", *Proceedings of SPIE*, Vol. 3717, Orlando, FL., 5-6 April 1999.
8. S. Kaewpajit, J. Le Moigne, and T. El-Ghazawi. "Finding the dimensionality of hyperspectral data," to appear in *Proc. of SPIE, Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VII*, 16-20 April 2001, Orlando, FL.
9. K. I. Diamantaras, and S. Y. Kung, *Principal Component Neural Networks*, John Wiley and Sons, Inc., New York, NY, 1996.
10. Research System Inc., *ENVI User's Guide*, 1999.
11. T. M. Lillesand, and R. W. Kiefer, *Remote Sensing and Image Interpretation*, John Wiley and Sons, Inc., New York, NY, 1994.
12. D. Landgrebe, "Multispectral Data Analysis: A Signal Theory Perspective," *School of Electrical & Computer Engineering, Purdue University, West Lafayette, Technical Report*, 1998.
13. J. A. Gualtieri, S. R. Chettri, R. F. Cromp, and L. F. Johnson, "Support Vector Machine Classifiers as Applied to AVIRIS Data," *Summaries of the Eighth JPL Airborne Earth Science Workshop*, February 8-11, 1999.
14. R. G. Congalton, and K. Green, *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, Lewis Publishers, New York, NY, 1999.
15. P. H. Swain, and S. M. Davis, *Remote Sensing: The Quantitative Approach*, McGraw-Hill, Inc., New York, NY, 1978.
16. H. H. Szu, J. Le Moigne, N. S. Netanyahu, and C. C. Hsu, "Integration of Local Texture Information in the Automatic Classification of Landsat Images," *SPIE Aerosense, Wavelet Applications IV*, April 1997.