

# Feature reduction of hyperspectral imagery using hybrid wavelet-principal component analysis

## Sinthop Kaewpijit

The George Washington University  
Department of Electrical and Computer  
Engineering  
Washington, D.C. 20052

## Jacqueline Le Moigne

NASA Goddard Space Flight Center  
Applied Information Science Branch  
Code 935  
Greenbelt, Maryland 20771

## Tarek El-Ghazawi

The George Washington University  
Department of Electrical and Computer  
Science  
Washington, D.C. 20052  
E-mail: tarek@seas.gwu.edu

**Abstract.** Hyperspectral imagery can provide very valuable information on land cover classes. However, it also presents many challenges in data analysis and interpretation as a result of the large amounts of data collected. For example, conventional methods for land use and land cover classifications may not be directly applicable. Such conventional methods typically require a preprocessing step to transform high dimensional data to a lower dimension, mostly by eliminating data redundancy. For decades, principal component analysis (PCA) has been widely used to decorrelate spectral bands for reducing dimensionality. It is a useful technique if the spectral class structure of the transformed data is distributed along the first few axes. Otherwise, the transformed data may be similar to the original data. In such cases, we have shown in an earlier work that the wavelet decomposition technique is a better approach. Wavelet decomposition can reduce hyperspectral data in the spectral domain for each pixel. By carefully combining PCA and wavelet techniques, we engender a new method that benefits from the strength of both techniques. The intent of the hybrid method is to provide a tradeoff between the accuracy and speed, as compared with PCA and wavelet methods. The effectiveness of this method is demonstrated by using hyperspectral data from the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) hyperspectral instrument. The experimental results show that, for high reduction rates, the hybrid method is superior to pure PCA and to pure wavelet-based techniques. © 2004 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.1637907]

Subject terms: dimension reduction; principal component analysis; wavelet decomposition; maximum likelihood.

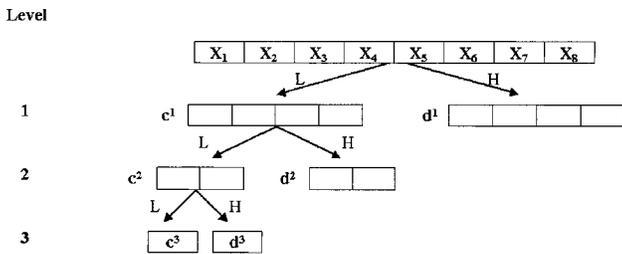
Paper 020545 received Dec. 17, 2002; revised manuscript received Jul. 21, 2003; accepted for publication Sep. 9, 2003.

## 1 Introduction

Hyperspectral imaging spectrometer data provide a wealth of information, which can be used to address a variety of Earth remote sensing applications. The rapid increase of the number of spectral channels for hyperspectral data creates a need for reducing data volume to tractable levels. The dimensionality of hyperspectral data can be reduced by applying a linear transformation, such as principal component analysis (PCA), and retaining only the significant components for further processing.<sup>1,2</sup> The object of PCA is to find a lower dimensional representation that accounts for the variance of the features. Although PCA is sufficient for reducing data volume, the process is time consuming and does not emphasize spectral signature, which is the fundamental concept of hyperspectral imagery for characterizing objects on the Earth's surface. Unlike PCA, wavelet decomposition focuses on reducing each individual spectral pixel in the spectral domain. Each reduced spectral pixel preserves the peaks and valleys of the original spectrum in a smaller representation. On the other hand, such a technique as PCA seeks to form linear combinations of the bands based on the global covariance matrix; whereas, the wavelet seeks merely a smaller subset of the original bands

based on a moving average of pixel vector values. Thus, the wavelet yields decomposed spectra that are smoother than the original spectra.<sup>3</sup> But, because of the lack of relationship among neighborhood pixels in the spatial domain, and because of the redundancy of wavelet coefficients, we are investigating a hybrid technique—a combination of wavelet and PCA—to achieve dimension reduction of hyperspectral data. Experimental results were conducted using the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) data. The results show that overall classification accuracy for the hybrid technique is superior to the other two techniques.

The remainder of this work is organized as follows. Section 2 briefly describes the PCA technique, and the multi-resolution wavelet decomposition as a new technique for hyperspectral reduction. Section 3 discusses the hybrid wavelet-PCA technique, including analytical efficiency. Section 4 presents the experiments that were conducted, including the hyperspectral datasets that were used. Section 5 shows and discusses experimental results and the impact of each method on classification accuracy, as well as computational efficiency. Section 6 concludes with brief remarks.



**Fig. 1** The fast discrete wavelet transform, where  $c^j$  represent the smoothed coefficients and  $d^j$  represent the detail coefficients;  $L$  is the low-pass filter and  $H$  is the high-pass filter.

## 2 Overview of Reduction Techniques

### 2.1 Principal Component Analysis

PCA, also referred to as the Hotelling transform or the Karhunen-Loeve transform, is a widely used dimension reduction technique in data analysis. To perform standard PCA, there are generally phases of computations.<sup>1</sup> These include assembling of the covariance matrix of the image to be transformed, determining the eigenvalues and corresponding eigenvectors of the covariance matrix, and forming the components. The first few principal components (PCs) contain the most information/variance, and are generally linear combinations of information from several spectral classes. The remaining principal components contain much less information, usually less than 1% of the data variance.

Let  $x$  be a pixel vector in the hyperspectral vector space, then PCA is a linear transformation  $G$ , such that

$$y = Gx, \quad (1)$$

with the constraint that the covariance matrix in the  $y$  space is diagonal. Moreover,  $G$  will be recognized as the transpose, provided that  $G$  is an orthogonal matrix, i.e.,  $G^{-1} = G^t$ .

### 2.2 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. Using the Mallat algorithm,<sup>3</sup> discrete wavelet transforms (DWTs) can be computed very quickly. 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.<sup>4</sup> This characteristic is related to the intrinsic property of wavelet transforms of preserving high- and low-frequency features during the signal decomposition, therefore preserving the peaks and valleys found in typical spectra. One of the most localized filters, DAUB4,<sup>5</sup> which has only four coefficients, has been used.

Figure 1 shows the principle of the Mallat algorithm: two 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, thereby halving its overall length. This is done recursively by reapplying the same procedure to the result of the

$L$  filter, as shown in Fig. 1. In the original vector  $x$  of length  $N = 2^J$  from Fig. 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.

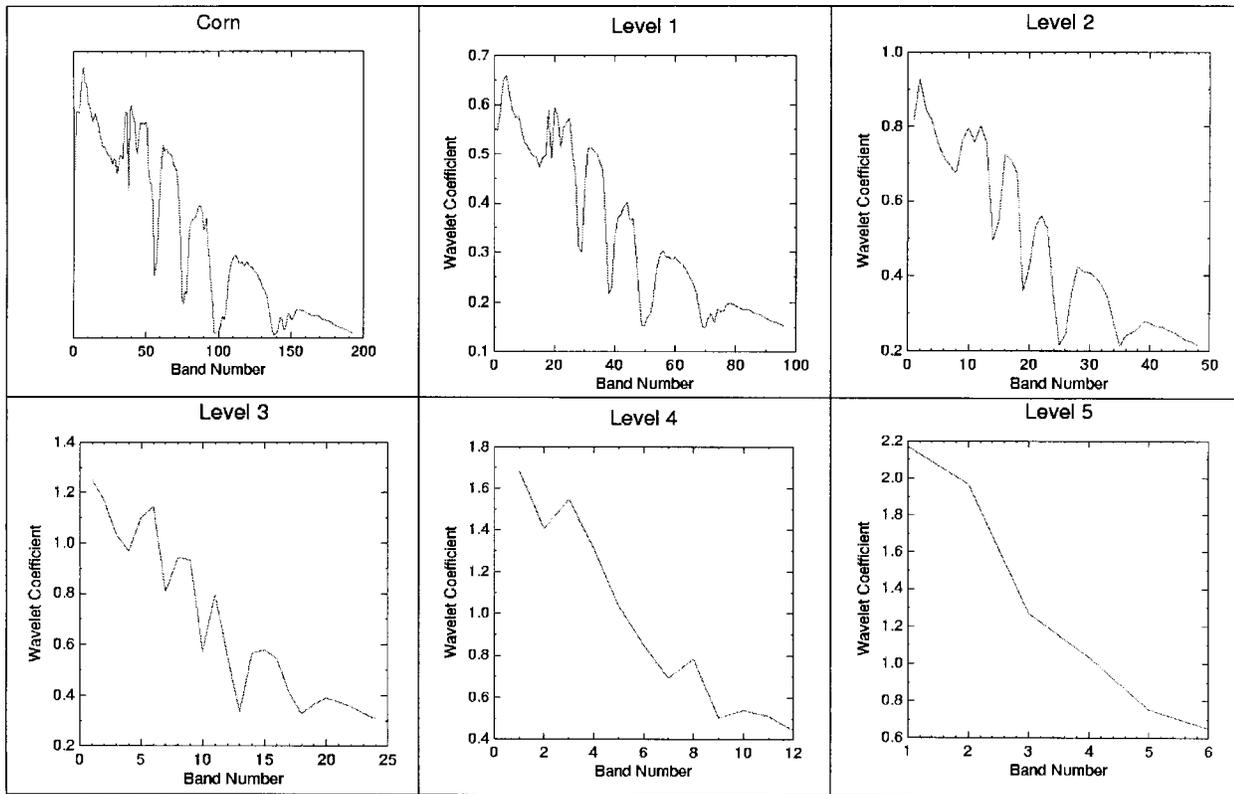
Examples of the decomposed spectral signature of corn for different wavelet decomposition levels are shown in Fig. 2. The wavelet technique reduces the effects of high-frequency details by retaining only spectra obtained from the low-pass filter. One of the important issues with wavelet reduction is to determine how many levels of decomposition can be applied while still yielding good classification accuracy. Applying the inverse DWT to the coefficient approximation<sup>3,6</sup> at the lower level by inserting a vector of zeros in place of the detail coefficients vector, we can get the reconstructed spectral data of its real approximation at the next higher level. This process can be recursively applied so that it yields a reconstructed signal of length equal to the original spectral signature. In our earlier work, our automatic wavelet spectral reduction algorithm<sup>4</sup> was developed by measuring the similarity<sup>7,8</sup> between the original spectral signature and the reconstructed spectral approximation using correlation (Corr). Then, the global histogram of this similarity measurement was computed as a measure of the desired probability distribution for all pixels in the scene. Based on a user's specified threshold (Th), the global decomposed level was determined as the lowest level that preserves the best the information content of each pixel. Details of this wavelet-based dimension reduction can be found in Ref. 4.

## 3 Hybrid Wavelet-PCA Reduction

### 3.1 Description of the Hybrid Algorithm

Similar to the methods presented in Sec. 2, our hybrid transformation is a preprocessing technique, which removes high-frequency components and reduces band-to-band correlation, especially in the case of noisy data. This hybrid transform combines both wavelet and PCA techniques. The transformation first performs an initial reduction using a wavelet decomposition, where the original hyperspectral data is compressed into a reduced-compact form. Then PCA is applied, thus significantly reducing the computational load compared to the conventional PCA technique. Usually, the wavelet method filters and subsamples each original spectrum into a reduced set of wavelet coefficients. In our algorithm, PCA decorrelates the band-to-band spectral information contained in the wavelet coefficients, and therefore yields a new smaller dataset in an uncorrelated coordinate system. The two advantages of the hybrid technique are: 1. it takes into account local spatial information among neighborhood class pixels, property that the wavelet is missing for classification purposes; and 2. it removes spectral correlation among wavelet coefficient bands.

Due to tremendous hyperspectral data dimensionalities, it is difficult to perceive much from statistical values computed from hyperspectral data, i.e., mean vector, covariance matrix.<sup>9</sup> To show that these values are preserved by wavelet decomposition, we use the visualization proposed by Kim and Swain.<sup>9</sup> In this representation, the correlation matrix (and therefore the covariance matrix) is shown by convert-

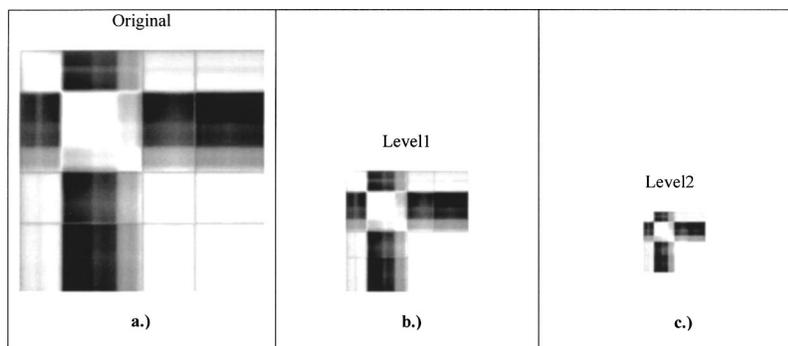


**Fig. 2** An example of Corn spectral signature and different levels of wavelet decomposition for the low-pass component.

ing absolute values of the interchannel correlations to gray values between 0 and 255. Tone is then proportional to the absolute value of the correlation (white=1 or -1; black=0).<sup>8</sup> Figure 3 shows the global statistical correlation coefficient image of the IndianPines'92 scene (described in Sec. 4) for the original data and two different levels of wavelet decomposition. This representation shows that correlation coefficient images in Figs. 3(b) and 3(c) look similar to the image produced from the original data [as shown in Fig. 3(a)], but they are smaller than that of the original data. It is also crucial to note that low correlations exist away from the diagonal, while the diagonal blocks show high correlations. Furthermore, all values along the diago-

nal are 1, since they measure the correlation of the data from each band with itself.

In comparison to these characteristics of the wavelet reduction method shown in Fig. 3, our hybrid technique preserves the usual characteristics of the conventional PCA, such that the higher order components with low variance can be discarded without significant loss of information content. Additionally, the original hyperspectral imagery can be reconstructed from the reduced representation using an inverse principal component transform and an inverse discrete wavelet transform, although with some loss of information. Moreover, for visualization purposes, a color composite image can be formed after the hybrid transform



**Fig. 3** Global correlation coefficient image of IndianPines'92 for different levels of wavelet decomposition.

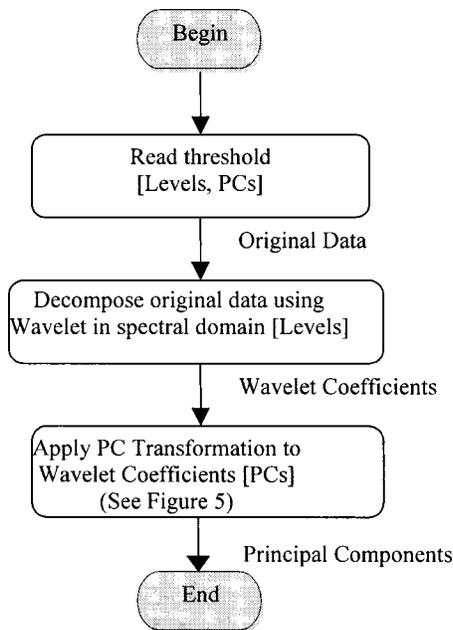


Fig. 4 A flow diagram of the hybrid wavelet-PCA reduction.

in the same manner as in conventional PCA. Since pure wavelet reduction represents the data, yet with smaller dimensionality, it is difficult to form a three-color composite for image display purposes, because any three arbitrary wavelet bands use only a small percentage of the total data variance. By applying PCA after the wavelet decomposition, the most informative three features (the first three PCs) obtained from the hybrid transform can be used. Similarly to the wavelet reduction technique,<sup>4</sup> the hybrid technique can also be applied to handle a situation in which the number of training samples is too limited to permit the use of all available features.

Figure 4 illustrates the flow diagram of the hybrid reduction technique. The hybrid algorithm gets the input from the user for the number of levels of decomposition [Level] and the number of principal component [PCs] chosen. First, the algorithm decomposes the original hyperspectral imagery in the spectral domain for each individual pixel to produce multiresolution wavelet-compressed spectra for the chosen level. Then PCA is applied to the wavelet coefficients to produce the chosen number of components.

The principle of our method is to apply a discrete wavelet transform to hyperspectral data in the spectral domain and at each pixel. As the wavelet transform includes both convolution and decimation, convolution helps retain interesting spectral features, while decimation helps reduce the data. This not only reduces the data volume, but it also preserves the distinctions between spectral signatures.<sup>4</sup>

### 3.2 Computational Complexity

Multiresolution wavelet decomposition with a Mallat algorithm is very fast for reducing hyperspectral dimensionality because of its pyramidal model and because we chose to only apply the low-pass filter. Let  $N$  be the length of the original spectral signature that is equivalent to the number of bands, and let the length of the original low-pass filter ( $F$ ) be  $L$ . For a filter of length  $L$ , wavelet decomposition

requires in the order of  $L$  operations per invocation. After the first invocation of the low-pass filter ( $F$ ), we obtain half the number of bands  $N/2$ , then we apply the low-pass filter again. Thus, each level processes half the number of pixels than the previous level. Since  $L$  is fixed for any particular wavelet filter, the wavelet-based reduction method yields the order of  $O(N)$  computations per pixel, where  $N$  is the number of bands.<sup>5</sup> The inverse discrete wavelet transform (IDWT) can, likewise, be computed in  $O(N)$  computations per pixel to reconstruct the real approximation to the original stage for the similarity measurement (correlation). The time complexity of the correlation metric (computed between the original spectral signature and its reconstructed approximation) is  $O(N)$  operations per pixel. Therefore, for the wavelet reduction algorithm described in Ref. 4 the order of complexity is  $O(MN)$ , where  $M$  is the number of pixels per scene.

By comparison with the wavelet method, PCA is computationally expensive. The time complexity of the different phases of PCA for an  $M$  pixel image of  $N$  spectral bands is as follows: 1. find mean vector  $O(MN)$ ; 2. assemble covariance matrix  $O(MN^2)$ ; 3. use eigen analysis to generate the transformation matrix performed as weighting coefficient  $O(N^3)$  for the standard eigen problem; and 4. perform pixel-by-pixel linear transformation-forming components  $O(RMN)$ , 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 ( $R \leq N$ ). The overall complexity of PCA is  $O(MN^2 + N^3)$ .<sup>10</sup>

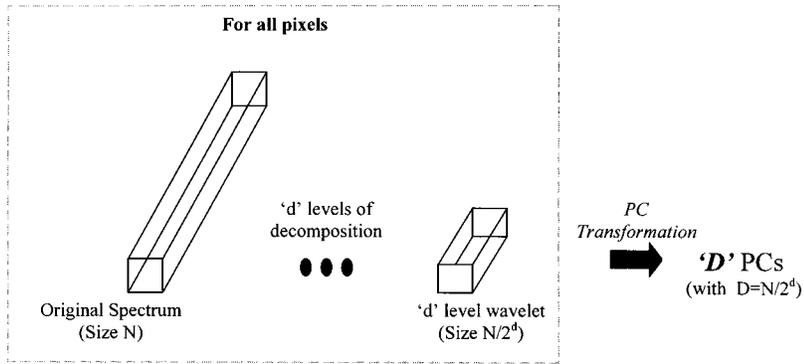
The great advantage of the hybrid technique compared to conventional PCA is to save computation time. Time spent on each computational phase is reduced, especially in forming components. The computational complexity of the hybrid method combines both complexities of the wavelet and of the PCA methods. Let  $D$  be the number of decomposed features from the wavelet that is equivalent to the new number of features (bands) for further processing of PCA,  $D = N/2^d$ , where  $d$  = number of levels of wavelet decomposition. Therefore, the overall complexity of the hybrid transform is  $O(MN + MD^2 + D^3)$ . The computational complexity of PCA will be decreased tremendously depending on the compression rate obtained from the wavelet  $D/N$ .<sup>10</sup> (see Fig. 5).

## 4 Experimental Approach

### 4.1 Hyperspectral Data Cube

The two hyperspectral datasets used for the experiments are as follows

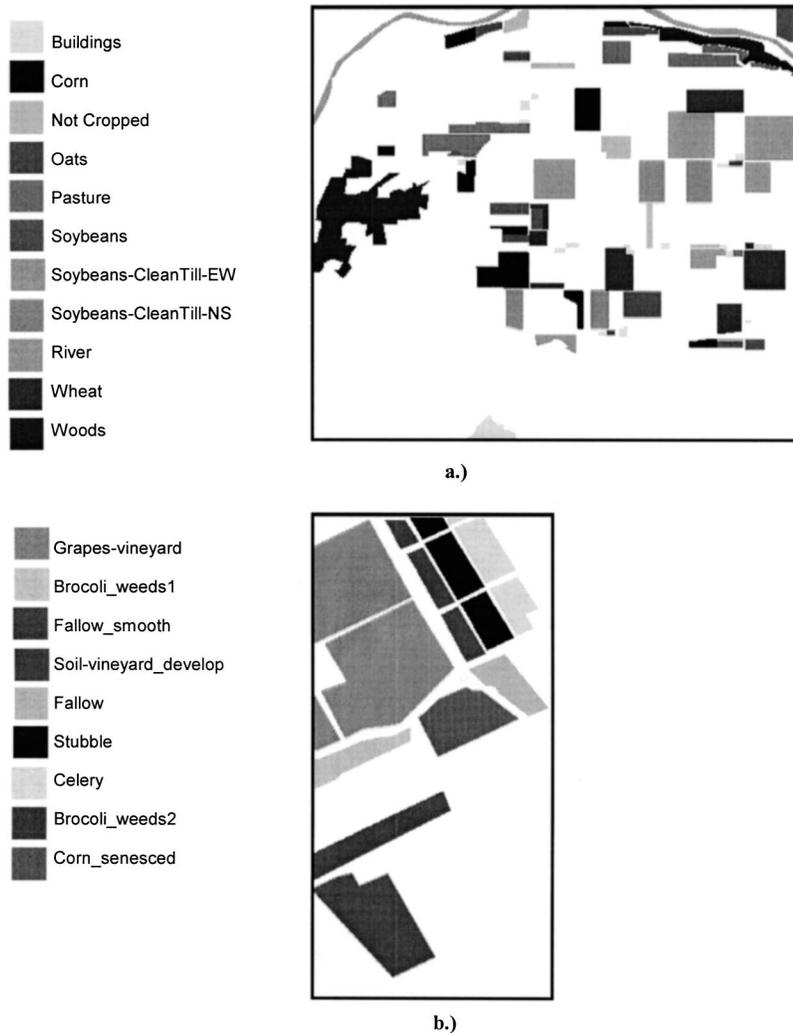
*IndianPines'92*. This AVIRIS farmland scene was acquired on 12 June, 1992 in the northern part of Indiana at high altitude with a ground pixel size of  $17 \times 17$  m. AVIRIS acquires images of very narrow, contiguous spectral bands throughout the visible, near-IR, and mid-IR portions of the spectrum (0.4 to  $2.5 \mu\text{m}$ ), in 224 bands at about 10-nm-wide intervals. We have used a dataset that consists of  $400 \times 400$  pixels by 192 bands. The 192 bands were selected by discarding significant water absorption bands, especially at about 1.4 and  $1.9 \mu\text{m}$ , and spectral overlaps bands resulting from the use of four individual spectrometers in the



**Fig. 5** Scheme of hybrid wavelet-PCA reduction.

AVIRIS instrument.<sup>11</sup> The ground truth is shown in Fig. 6(a). Original noisy bands, especially water absorption bands, will lead to some principal components with high variance that may be misleading in the classification process. The 192 bands used are therefore contiguous and equally spaced in the spectrum, except for the two bands out of the 192 that were discarded due to water absorption.

For this scene, the ground truth covers 20% of the full  $400 \times 400$  scene and is divided among 11 classes. The 11 classes are Buildings, Corn, Not Cropped, Oats, Pasture, Soybeans, Soybeans CleanTill EW rows, Soybeans CleanTill NS rows, River, Wheat, and Woods, ranging from 5926 to 1032 ground truth pixels per class. CleanTill refer to the amount of very little residue from the previous year's crop



**Fig. 6** The reference data: (a) IndianPines'92 and (b) Salinas'98.

**Table 1** Number of training and testing pixels for the IndianPines'92 scene.

Class name	Training data number of pixels	Testing data number of pixels
Buildings	496	745
Com	1415	2123
Not cropped	528	793
Oats	413	619
Pasture	1314	1970
Soybeans	828	1241
Soybeans-C-EW	2370	3556
Soybeans-C-NS	1076	1613
River	610	915
Wheat	1688	2533
Woods	2364	3545

that is on the surface of the ground. A random training sample of 40% of the pixels was chosen from the known ground truth from each class. A maximum-likelihood classifier was applied to the remaining 60% of the known ground pixels for coverage of those classes.<sup>2,12</sup> The number of training and testing pixels for each class are given in Table 1.

**Salinas'98.** This AVIRIS dataset was acquired on 9 October, 1998, south of the city of Greenfield in the Salinas Valley in California. It was taken at low altitude with a pixel size of 3.7 m. We are interested in a scene that has vegetables, bare soils, and vineyard fields. This scene consists of  $217 \times 512$  pixels by 192 bands of radiance data, discarding significant water absorption and instrument's overlap bands. Its ground truth is shown in Fig. 6(b). We selected nine classes for testing as follows: Grapes-vineyard, Broccoli-weed1, Fallow\_smooth, Soil-vineyard\_develop, Fallow, Stubble, Celery, Broccoli\_weed2, and Corn\_senesced, ranging from 18539 to 2010 ground truth pixels per class. The random training samples were chosen as 30% of the pixels for each class from the known ground truth. The trained classifier was applied to the remaining 70% of that coverage.<sup>12</sup> The number of training pixels and testing pixels for each class are given in Table 2.

**Table 2** Number of training and testing pixels for the Salinas'98 scene.

Class name	Training data number of pixels	Testing data number of pixels
Grapes-vineyard	5562	12977
Broccoli_weeds1	603	1407
Fallow_smooth	803	1875
Soil-vineyard_develop	1861	4342
Fallow	593	1383
Stubble	1188	2771
Celery	1074	2505
Broccoli_weeds2	1118	2608
Com_senesced	983	2295

Since training samples were selected randomly, the procedure of selecting samples was repeated three times for each of the three scenes,<sup>2</sup> and the median classification accuracy resulting from the experiments is reported.

## 4.2 Supervised Classification

We have experimentally validated the hybrid technique by using remotely sensed image test suites from two hyperspectral scenes as mentioned earlier, and the environment for visualizing images (ENVI) as a tool for classification assessment. The experiments performed in this work use the maximum likelihood classifier (ML),<sup>13,14</sup> which is probably the most common supervised classification method used with remote sensing data. The effectiveness of the ML depends on a reasonably accurate estimation of the mean vector and the covariance matrix for each spectral class. We used the same level of compression as the basis of comparison among the three techniques: wavelet, conventional PCA, and hybrid methods. For example, the second level of decomposition (decimated by 4 from 192 bands of original data) is analogous to 48 principal components (PCs), the third level to 24 PCs, and so on. For the hybrid technique, a given level of wavelet decomposition is computed, then PCA is applied at the same compression rate as the wavelet and conventional PCA, i.e., Level 1+48 PCs. The effectiveness of the reduction is demonstrated by determining the classification accuracy when a supervised classification such as ML is used on the reduced data. The supervised classification algorithm is trained on labeled data, so it can identify the class to which a pixel or a region belongs, and thus provide a high-level characterization of the data.<sup>15,16</sup> In practice, it has often been observed that if the number of training pixels is small, the addition of more dimensions leads to a worse performance in the testing pixels (often known as "curse of dimensionality"). It is estimated that a minimum of  $10N$  pixels per training class is necessary, with as many as  $100N$  desired for reliable estimates, where  $N$  is the number of spectral bands.<sup>17</sup> Therefore, we investigate all three techniques with a minimum of  $10N$  pixels (per training class) in both hyperspectral datasets. The details about the number of training and testing pixels are shown in Tables 1 and 2.

## 5 Experimental Results

### 5.1 Impact on Classification Accuracy

In this work, the combination of wavelet and PCA transforms is done by first applying the wavelet to the original data, and then performing PCA on the wavelet coefficients for all levels of decomposition. Then, the comparison among the three techniques is made at the same compression rate. From Figs. 7 and 8, we can see that the hybrid technique generally outperforms conventional PCA in classification accuracy when compared at the same compression rate. The pure wavelet technique gives the maximum accuracy for higher data dimensionality (48 bands) for both datasets. It is important to note that the higher information content of PCA for the hybrid technique is not a guarantee for higher classification accuracy, because PCA is performed after the wavelet transform. As we mentioned earlier, the wavelet decomposition keeps only low-frequency features and discards information on high-frequency fea-

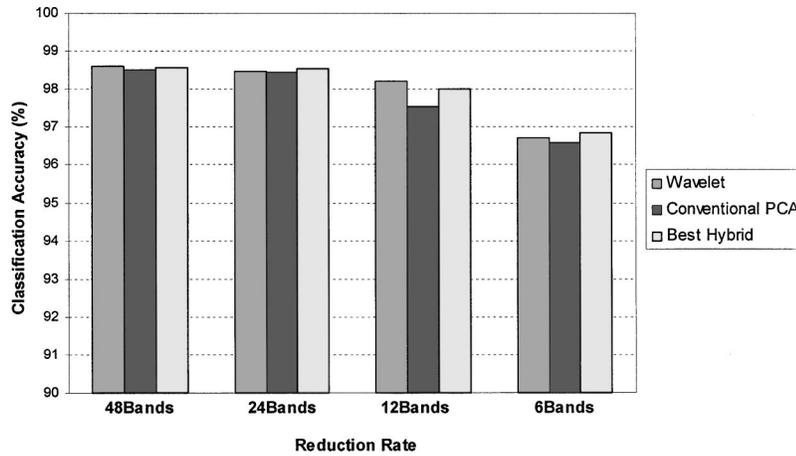


Fig. 7 Classification accuracy comparison of Salinas'98.

tures. Therefore, the information content (i.e., data variance percentage) obtained from the hybrid transform is produced from only low-frequency features of wavelet coefficients. Tables 3 and 4 show the information content of the original data computed from conventional PCA; however, in the case of the hybrid technique, the information content represents the total data variation of the wavelet coefficients that are produced from the wavelet by discarding high-frequency features.

Another explanation of the increased classification accuracy of hybrid transform data compared to PCA data is that the hybrid transform keeps a larger proportion of information content for the same rate of data reduction. For example, 48 PCs of 192 bands from conventional PCA is one fourth of the input data, as compared to 48 PCs of 96 wavelet coefficient bands from the hybrid being one half of the input data (wavelet coefficients). The results show, as expected, that classification accuracy increases with the number of PCs. In these experiments, we chose a maximum of 48 PCs, after which the classifier (ML) performance degrades because of the discrepancy between the number of tested pixels versus the number of trained pixels.<sup>1,4</sup>

By “best hybrid,” we mean the combined choice of the number of levels of decomposition and the number of principal components that yield the maximum accuracy for that compression rate. Experimental results show that the maximum classification accuracy is usually obtained by using the best hybrid with all reduction rates except for 48 bands. At 48 bands, the wavelet seems to provide better classification accuracies than the hybrid, because the wavelet preserves the distinctions between spectral signatures, and because the nature of the classifiers, which are mostly pixel-based techniques,<sup>17,18</sup> is better suited for wavelets, which are pixel-based transformations. With this compression rate (48 bands) and 11 land-cover classes of the IndianPines scene, using the ML classification, the wavelet reduction technique yields a classification rate of 80.5424% compared to 79.5248% with the hybrid transform and 78.8124% with conventional PCA, as shown in Fig. 9 and Table 3. The same trend is obtained from the Salinas scene; 98.6071 for the wavelet, 98.5822 for the hybrid, and 98.5169 for conventional PCA as shown in Fig. 7 and Table 4.

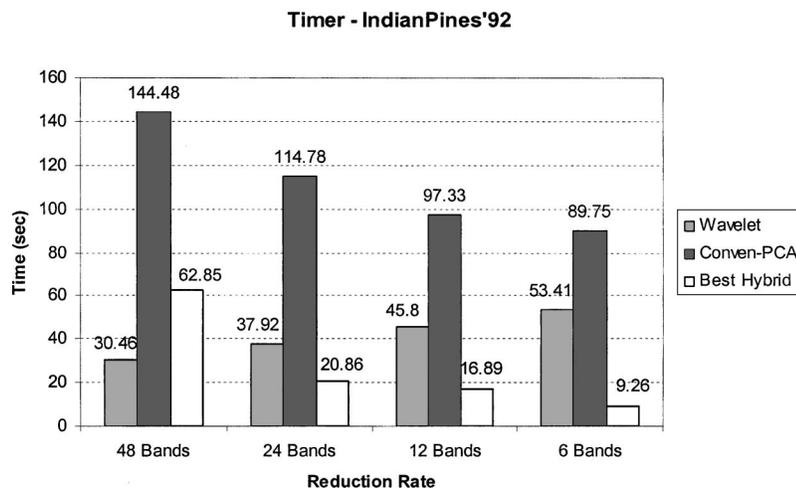


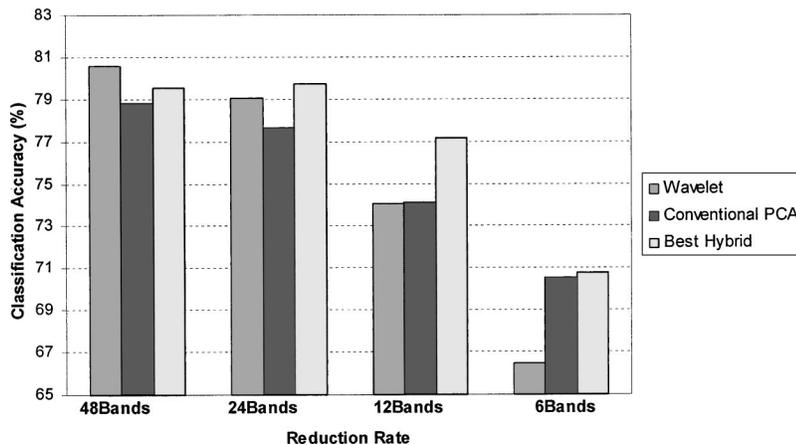
Fig. 8 Timing comparison of the wavelet, PCA, and the hybrid.

**Table 3** Information content and classification accuracy of the IndianPines'92 scene. Level 1+PCA means performing reduction for one level of wavelet decomposition and then applying PCA to its wavelet coefficients. CA means classification accuracy. IC means information content.

Techniques	Extent of reduction Reduction rate			
	48 bands	24 bands	12 bands	6 bands
Wavelet	CA=80.5424	CA=79.0312	CA=74.0396	CA=66.4326
Conventional PCA	CA=78.8124	CA=77.637	CA=74.1414	CA=70.5457
	IC=99.92	IC=99.80	IC=99.67	IC=99.36
Level 1+PCA	CA=79.5248	CA=78.8226	CA=76.5939	CA=70.605
	IC=99.99	IC=99.95	IC=99.87	IC=99.66
Level 2+PCA		CA=79.6876	CA=77.1434	CA=70.6711
		IC=99.99	IC=99.95	IC=99.80
Level 3+PCA			CA=76.6244	CA=70.5236
			IC=99.99	IC=99.90
Level 4+PCA				CA=70.7831 IC=99.97

**Table 4** Information content and classification accuracy of the Salinas'98 scene. CA means classification accuracy. IC means information content.

Techniques	Extent of reduction Reduction rate			
	48 bands	24 bands	12 bands	6 bands
Wavelet	CA=98.6071	CA=98.4672	CA=98.206	CA=96.7167
Conventional PCA	CA=98.5169	CA=98.4361	CA=97.5407	CA=96.59
	IC=99.99	IC=99.99	IC=99.97	IC=99.91
Level 1+PCA	CA=98.5822	CA=98.5231	CA=97.8516	CA=96.6079
	IC=100	IC=100	IC=99.99	IC=99.93
Level 2+PCA		CA=98.4765	CA=97.8702	CA=96.5955
		IC=100	IC=99.99	IC=99.94
Level 3+PCA			CA=98.007	CA=96.8442
			IC=100	IC=99.96
Level 4+PCA				CA=96.6608 IC=99.99



**Fig. 9** Classification accuracy comparison of IndianPines'92.

**Table 5** The IndianPines test data: confusion matrix of the wavelet technique (Level 2).

Class	User's acc. (%)	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10	Class 11	Row total
Class 1	58.51	567	45	20	24	186	27	13	4	2	39	42	969
Class 2	77.87	40	1207	68	20	36	77	16	31	0	47	8	1550
Class 3	85.01	20	22	550	1	25	6	11	1	0	4	7	647
Class 4	65.35	0	7	4	445	25	7	1	2	0	190	0	681
Class 5	84.94	69	40	25	11	1528	32	2	1	0	10	81	1799
Class 6	64.45	7	248	22	1	38	805	64	14	0	47	3	1249
Class 7	75.83	0	221	29	12	0	251	3153	478	0	14	0	4158
Class 8	64.9	3	269	2	13	1	7	287	1080	0	2	0	1664
Class 9	100	0	0	0	0	0	0	0	0	913	0	0	913
Class 10	89.04	17	22	57	84	55	22	9	2	0	2177	0	2445
Class 11	95.14	22	42	16	8	76	7	0	0	0	3	3404	3578
Column total		745	2123	793	619	1970	1241	3556	1613	915	2533	3545	19,653
Producer's acc. (%)		76.11	56.85	69.36	71.89	77.56	64.87	88.67	66.96	99.78	85.95	96.02	
Overall (%)	<b>80.54</b>												

Tables 5, 6, and 7 show the complete results with the confusion matrices for the testing areas of the IndianPines'92 scene with the ML classification, while Tables 8, 9, and 10 show the complete results for the Salinas'98 scene at second levels of decomposition that are analogous to 48 PCs. In general, there is little difference in accuracy for each class obtained among these techniques for the Salinas'98 scene. However, in the IndianPines'92 scene, some improvements in accuracy from the wavelet compared to conventional PCA can be observed for the Oats and Soybeans classes. This trend can be applied also for the hybrid compared to conventional PCA for the Not Cropped, Oats, and Pasture classes.

### 5.2 Computational Efficiency

Figure 8 shows the comparison of efficiency in time among the three techniques. For the Best Hybrid (as shown in Fig. 8), 48 bands of the best hybrid are produced from wavelet level 1+48 PCs; 24 bands are produced from wavelet level 2+24 PCs; 12 bands are produced from wavelet level 2+12 PCs; and 6 bands are produced from wavelet level 4+6 PCs. This figure shows that time spent for the wavelet method is small compared to the conventional PCA and the best hybrid transform at 48 bands. Then it increases when the number of bands decreases because of the reconstruction process to compute the similarity measurement.<sup>4</sup> Con-

**Table 6** The IndianPines test data: confusion matrix of PCA (48 PCs).

Class	User's acc. (%)	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10	Class 11	Row total
Class 1	59.52	550	39	21	25	175	23	11	3	0	33	44	924
Class 2	76.4	41	1191	69	28	44	59	25	43	0	53	6	1559
Class 3	82.94	24	15	530	0	19	11	20	2	0	11	7	639
Class 4	60.54	0	8	4	405	34	5	0	2	0	211	0	669
Class 5	84.03	73	53	35	11	1531	37	1	0	0	5	76	1822
Class 6	63.42	8	242	24	8	31	742	64	15	0	36	0	1170
Class 7	73.02	3	246	28	13	0	285	3036	524	0	23	0	4158
Class 8	58.63	1	261	2	21	0	45	389	1022	0	2	0	1743
Class 9	100	0	0	0	0	0	0	0	0	913	0	0	913
Class 10	88.04	24	25	63	99	44	26	10	2	0	2157	0	2450
Class 11	94.62	21	43	17	9	92	8	0	0	2	2	3412	3606
Column total		745	2123	793	619	1970	1241	3556	1613	915	2533	3545	19,653
Producer's acc. (%)		73.83	56.1	66.83	65.43	77.72	59.79	85.38	63.36	99.78	85.16	96.25	
Overall (%)	<b>78.81</b>												

**Table 7** The IndianPines test data: confusion matrix of the hybrid technique (Level 1+48 PCs). Class 1: Buildings; Class 2: Corn; Class 3: Not Cropped; Class 4: Oats; Class 5: Pasture; Class 6: Soybeans; Class 7: Soybeans-C-EW; Class 8: Soybeans-C-NS; Class 9: River; Class 10: Wheat; and Class 11: Woods.

Class	User's acc. (%)	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10	Class 11	Row total
Class 1	61.82	549	38	21	25	143	29	15	3	2	32	31	888
Class 2	76.84	40	1198	71	23	49	77	24	36	0	35	6	1559
Class 3	83.23	19	16	541	4	27	8	11	2	0	13	9	650
Class 4	61.36	0	5	4	416	27	5	0	1	0	220	0	678
Class 5	84.72	74	45	35	12	1563	37	1	1	0	7	70	1845
Class 6	62.56	7	265	18	6	27	752	76	13	0	38	0	1202
Class 7	73.78	2	238	29	12	1	277	3056	501	0	26	0	4142
Class 8	60.83	2	256	1	22	0	33	363	1056	0	3	0	1736
Class 9	100	0	0	0	0	0	0	0	0	913	0	0	913
Class 10	89.05	28	15	58	92	47	15	10	0	0	2156	0	2421
Class 11	94.75	24	47	15	7	86	8	0	0	0	3	3429	3619
Column total		745	2123	793	619	1970	1241	3556	1613	915	2533	3545	19,653
Producer's acc. (%)		73.69	56.43	68.22	67.21	79.34	60.6	85.94	65.47	99.78	85.12	96.73	
Overall (%)	<b>79.52</b>												

ventional PCA is a time-consuming technique that is the slowest among the three techniques for all reduction rates. The time spent for PCA decreases, depending on the number of formed components. In general, the time in the hybrid technique is the highest for 48 bands and decreases for a smaller number of bands because this time is saved in the PCA portion of the algorithm.

The speed of the hybrid technique is especially important when managing massive amounts of remotely sensed hyperspectral data on board spacecraft.<sup>19,20</sup> Figure 9 shows that the classification accuracy is very close to that of the two other methods, but Fig. 8 clearly shows the advantage in the speed of the hybrid algorithm over the conventional PCA. To speed up this application even further, the computation of these techniques using parallel processing can be addressed. This was demonstrated by using a Beowulf machine developed at NASA/Goddard, the highly parallel integrated virtual environment (HIVE). In this work, we have

implemented the hybrid algorithm on a subcluster based on Gateway 2000 PCs, which has 16 nodes, each of which is a dual Pentium III Xeon (a total of 32 processors) interconnected via a Myrinet network, and with a total of 8 Gbytes of RAM and 72 Gbytes of disk.

The good data locality in the automatic wavelet technique makes it more suitable for parallel processing. This is because processors can work on different pixels independently without any need for interprocessor communication overhead. The automatic wavelet parallel algorithm equally partitions the image pixels and maps each partition to a node. Each node has its own local partition of data to perform automatic wavelet reduction. The hybrid technique does not lend itself very well to parallelism, because of its nature of global computations that requires sending and receiving data among nodes. These operations produce a lot of communication overheads.

**Table 8** The Salinas test data: confusion matrix of the wavelet technique (Level 2).

Class	User's acc. (%)	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Row total
Class 1	98.8	12,909	0	52	6	2	26	15	2	54	13,066
Class 2	100	0	1402	0	0	0	0	0	0	0	1402
Class 3	100	0	0	1800	0	0	0	0	0	0	1800
Class 4	99.45	0	0	0	4169	0	0	0	0	23	4192
Class 5	98.29	6	0	18	0	1378	0	0	0	0	1402
Class 6	99.85	0	0	4	0	0	2745	0	0	0	2749
Class 7	100	0	0	0	0	0	0	2488	0	0	2488
Class 8	99.77	0	5	0	0	0	0	1	2606	0	2612
Class 9	90.46	62	0	1	167	3	0	1	0	2218	2452
Column total		12,977	1407	1875	4342	1383	2771	2505	2608	2295	32,163
Producer's acc. (%)		99.48	99.64	96	96.02	99.64	99.06	99.32	99.92	96.64	
Overall (%)	<b>98.6</b>										

**Table 9** The Salinas test data: confusion matrix of PCA (48 PCs).

Class	User's acc. (%)	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Row total
Class 1	98.72	12,900	0	50	7	2	26	15	3	64	13,067
Class 2	100	0	1402	0	0	0	0	0	0	0	1402
Class 3	100	0	0	1801	0	0	0	0	0	0	1801
Class 4	99.47	0	0	0	4158	0	0	0	0	22	4180
Class 5	98.92	0	0	15	0	1378	0	0	0	0	1393
Class 6	99.78	0	0	6	0	0	2745	0	0	0	2751
Class 7	100	0	0	0	0	0	0	2488	0	0	2488
Class 8	99.77	0	5	0	0	0	0	1	2605	0	2611
Class 9	89.43	77	0	3	177	3	0	1	0	2209	2470
Column total		12,977	1407	1875	4342	1383	2771	2505	2608	2295	32,163
Producer's acc. (%)		99.41	99.64	96.05	95.76	99.64	99.06	99.32	99.88	96.25	
Overall (%)	<b>98.51</b>										

The parallel implementation of the automatic wavelet spectral reduction, however, has shown that the parallel algorithm is not only scalable, but has also brought the execution time of the wavelet computation to an exceptional level using cost-efficient high-performance computers.<sup>21</sup> This was shown by applying the parallel implementation to compute the automatic wavelet reduction for two levels of decomposition (48 bands) on an eight processor Beowulf. The parallel wavelet took only 7.04 s, while its sequential version took 30.46 s. A better speedup (about 6.5 times faster compared to the sequential version) of the parallel automatic wavelet can be observed for five levels of decomposition, which is the most time consuming for the wavelet. The parallel hybrid technique took 10.87 s, while its sequential hybrid took 62.85 s. The parallel conventional PCA took about 26.05 s out of 144.48 s (sequential time). These results were obtained on eight processor Beowulfs of Pentium Xeon processors. From the parallel implementation with eight processors, it was shown that the parallel wavelet generally is more scalable than the other two parallel algorithms for the worst-case scenario of the used

data. Both the parallel hybrid and the parallel PCA algorithms, however, can reduce the overall execution time.

## 6 Conclusions

We present a new hybrid wavelet-PCA reduction of hyperspectral data. Both analytical and experimental evaluations of execution time and classification accuracy are conducted. The results show that the hybrid dimension reduction technique is a very useful method for reducing dimensionality of hyperspectral data, and can produce faster results than currently used PCA techniques. The intent of the experiments is to show that the hybrid method produces better classification accuracy than a pure wavelet method and better speed than a PCA. Our experimental results for this particular dataset show that applying the wavelet first and then the PCA can provide slightly better classification accuracies than PCA for some particular cases. In general, the best results for classification accuracy are obtained when utilizing two levels of automatic wavelet decomposition. This can be explained by the fact that wavelet data

**Table 10** The Salinas test data: confusion matrix of the hybrid technique (Level 1+48 PCs). Class 1: Grapes-vineyard; Class 2: Brocoli\_weeds1; Class 3: Fallow\_smooth; Class 4: Soil-vineyard\_develop; Class 5: Fallow; Class 6: Stubble; Class 7: Celery; Class 8: Brocoli\_weeds2; and Class 9: Corn\_senesced.

Class	User's acc. (%)	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Row total
Class 1	98.81	12,903	0	50	7	4	25	13	2	54	13,058
Class 2	100	0	1402	0	0	0	0	0	0	0	1402
Class 3	100	0	0	1800	0	0	0	0	0	0	1800
Class 4	99.47	0	0	0	4164	0	0	0	0	22	4186
Class 5	98.71	1	0	17	0	1378	0	0	0	0	1396
Class 6	99.85	0	0	4	0	0	2745	0	0	0	2749
Class 7	100	0	0	0	0	0	0	2490	0	0	2490
Class 8	99.77	0	5	0	0	0	0	1	2606	0	2612
Class 9	89.84	73	0	4	171	1	1	1	0	2219	2470
Column total		12,977	1407	1875	4342	1383	2771	2505	2608	2295	32,163
Producer's acc. (%)		99.43	99.64	96	95.9	99.64	99.06	99.4	99.92	96.69	
Overall (%)	<b>98.58</b>										

represent a spectral distribution similar to the original one, but in a compressed form. It is also observed from the experiments that the automatic wavelet technique fits well the ML classification process. Although the first level of decomposition better represents the original spectral distribution, larger data dimensions cause a loss in classifier performance.

Compared to a wavelet-based reduction method, the hybrid method has slightly better classification accuracy at higher reduction rates than the wavelet technique. The computational cost of the sequential hybrid technique at the higher reduction rate is also the lowest one.

### 6.1 Future Ideas

There are several ideas that can be implemented to make this work more effective and useful. A few of these ideas are as follows.

- Currently we are using orthogonal DWT to perform dimension reduction, which requires using images that have power of 2 spectral bands. Investigating ways to remove this restriction is of great interest. One idea is to consider removing this limitation by using the biorthogonal DWT, and understanding the associated tradeoffs in accuracy, storage, and speed.
- We have used a DAUB4 filter as the mother wavelet due to its simplicity and the fact that it provides good results for our remote sensing applications. It would be of interest to conduct a systematic study that considers different types of filters that can be used for dimension reduction, along with the associated tradeoffs.

### Acknowledgments

The authors wish to acknowledge the support of the Intelligent Systems/Image Understanding Program as well as the Computational Technologies Program of the Earth Science Technology Office. They also would like to thank Tony Gualtieri and David Landgrebe for providing test hyperspectral datasets.

### References

1. J. A. Richards, *Remote Sensing Digital Image Analysis: An Introduction*, Springer-Verlag, Berlin, Heidelberg (1999).
2. D. Landgrebe, "Multispectral data analysis: A signal theory perspective," Technical Report, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 1998. Available: <http://ee.www.ecn.purdue.edu/~biehl/MultiSpec/Signal.Theory.pdf>
3. S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(7), 674–693 (1989).
4. S. Kaewpigit, J. Le Moigne, and T. El-Ghazawi, "Automatic wavelet spectral analysis for reduction of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.* **41**(4), 863–871 (2003).
5. I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA (1992).
6. S. A. Dianat and R. Rao, "Wavelet transforms: Theory and applications," Short Course Note: Proc. SPIE's 12th Annual International-Symposium on Aerosense/Defense Sensing, Simulation & Controls, 13–17 April (1998).
7. L. Bruce and J. Li, "Wavelet for computationally efficient hyperspectral derivative analysis," *IEEE Trans. Geosci. Remote Sens.* **39**(7), 1540–1546 (2001).
8. C. I. Chang, "An information theoretic-based approach to spectral variability, similarity and discriminability for hyperspectral image analysis," *IEEE Trans. Inf. Theory* **46**(5), 1927–1932 (2000).
9. H. Kim and P. H. Swain, "Evidential reasoning approach to multisource-data classification in remote sensing," *IEEE Trans. Syst. Man Cybern.* **25**(8), 1257–1265 (1995).
10. S. Kaewpigit, J. Le Moigne, and T. El-Ghazawi, "A wavelet-based PCA reduction for hyperspectral imagery," in *Proc. 2002 Intl. Geosci. Remote Sens. Symp. (IGARSS'02)*, Toronto, Canada.
11. X. Jia and J. A. Richards, "Efficient maximum likelihood classification for imaging spectrometer data sets," *IEEE Trans. Geosci. Remote Sens.* **32**(2), 274–281 (1994).
12. J. A. Gualtieri, S. R. Chettri, R. F. Cromp, and L. F. Johnson, "Support vector machine classifiers as applied to AVIRIS data," *Summaries 8th JPL Airborne Earth Sci. Workshop*, Feb. 8–11, 1999.
13. Research System Inc., *ENVI User's Guide* (1999).
14. X. Jia and J. A. Richards, "Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification," *IEEE Trans. Geosci. Remote Sens.* **37**(1), 538–542 (1999).
15. R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, John Wiley and Sons, New York (1973).
16. T. M. Lillesand and R. W. Kiefer, *Remote Sensing and Image Interpretation*, John Wiley and Sons, Inc., New York (1994).
17. P. H. Swain and S. M. Davis, *Remote Sensing: The Quantitative Approach*, McGraw-Hill, Inc., New York (1978).
18. 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," *Proc. SPIE* **3078**, 116–127 (1997).
19. T. El-Ghazawi, P. Charlemwat, and J. Le Moigne, "Wavelet-based image registration on parallel computers," *Proc. ACM/IEEE* (1997).
20. D. Ridge, D. Becker, P. Merkey, and T. Sterling, "Beowulf: Harnessing the power of parallelism in a pile-of-pCs," *Proc. IEEE Aerospace Conf.* **2**, 79–91 (1997).
21. K. Hwang and Z. Xu, *Scalable Parallel Computing*, McGraw-Hill, Boston, MA (1998).



**Sinthop Kaewpigit** received his BS degree in civil engineering from Chulachomkiao Royal Military Academy (CRMA), Thailand, in 1996, the MS degree in engineering management from the George Washington University (GWU), Washington, D.C., in 1999, and the PhD degree in computational science and informatics at George Mason University (GMU), Fairfax, Virginia, in 2002. His research interests include parallel and distributed computing on cluster of workstations, and hyperspectral Sensing. He is a member of IEEE.



**Jacqueline Le Moigne** is a senior computer scientist in the Applied Information Science Branch of the Earth and Space Data and Computing Division of the NASA Goddard Space Flight Center. She received a BS and MS in mathematics and a PhD in computer vision from the University Pierre and Marie Curie, Paris, France. Her research interests include computer vision, utilizing massively parallel computers and applied to Earth and space science problems such as robotics, land use/land cover assessment, and intelligent data management. The scientific objective of her research project is to help Earth and space scientists analyze quickly and efficiently large amounts of data. Her most recent research focuses on parallel registration and fusion of multisensor/multiscale satellite image data, for which she has been studying wavelets and their implementation on high-performance computers. Current work includes the development of an image registration toolbox, the registration of Landsat imagery, and the implementation of image processing methods on field programmable gate arrays and on Beowulf-type architectures. She was the chairman and vice-chairman of the Washington/North Virginia chapter of the IEEE Geoscience and Remote Sensing Society from 1993 to 1995. In November 1997, she was chairman of the first NASA/Goddard Image Registration Workshop, and in 1999, she was coeditor of a special issue of the journal *Pattern Recognition* on image registration. She was elected an IEEE senior member in 1996 and *IEEE TGARS* associate editor in 2001. She is also an associate editor for *Pattern Recognition*.



**Tarek El-Ghazawi** is a professor at the Department of Electrical and Computer Engineering at The George Washington University. He received his PhD degree in 1988 from New Mexico State University in electrical and computer engineering. He was previously with the George Mason University, and also taught at the Florida Institute of Technology and Johns Hopkins University. His research interests include high-performance computing and architectures, computer vision, and performance evaluations. He has more than

100 refereed publications in these areas, and his research has been frequently supported by government and industry, particularly NASA, LUCITE, DARPA and Department of Defense, the National Science foundation, and industry. He has served as an associate editor for the *International Journal on Parallel and Distributed Systems and Networking*, and a guest editor for the *IEEE Concurrency*, special track on high-performance data mining. He has served on numerous conference and technical program committees. He is a senior member of the IEEE, a member of the ACM, and a member of Phi Kappa Phi.