A Derivative-Aided Hyperspectral Image Analysis System for Land-Cover Classification

Fuan Tsai and William D. Philpot, *Member, IEEE*

***Abstract*—The large number of spectral bands in hyperspectral data seriously complicates their use for classification. Selection of a useful subset of bands or derived features (spectral ratios, differences, derivatives) is always desirable, strongly affects the accuracy of the classification, and is often a practical necessity to keep the processing speed and memory requirements under control. This paper examines one possible procedure for selecting spectral derivatives to improve supervised classification of hyperspectral images. The procedure is designed to identify derivative features that are more effective at separating target classes and then add them to a base subset of features for classification. The goal is to create the smallest set of features that will result in the best classification result. A key issue in this process is the interplay of the number of features and the size of the training data sets since classification accuracy declines if the dimensionality of the feature space is too large relative to the number of training samples.**

***Index Terms*—Computer-aided data analysis, hyperspectral image analysis, spectral derivative.**

I. Background

HYPERSPECTRAL remote sensing data provide researchers the opportunity to pursue complex analysis that might be difficult to carry out using traditional multispectral data. These data include more spectral details and are more adaptable to specific applications. However, using hyperspectral information also introduces new challenges. One of the challenges is simply dealing with the great increase in data volume and the corresponding increase in processing time. A more substantive challenge is to make effective use of the new information available in these data. Both issues are critical when using the more traditional image analysis methods that were designed for use with single spectral or multispectral data and are inefficient when applied to hyperspectral data analysis. Some researchers have developed new methods that are better adapted to address the high dimensionality of the data such as neural networks [1], spectral angle mapping [2], and graphically interactive approaches [3]. Another important approach has been the use of spectral libraries as reference data—an approach inherited from spectroscopic methods in analytical chemistry [4]–[6]. In this procedure, the user supplies a library of reference spectra—usually reflectance measurements collected in the laboratory—to be used as a look-up table to select training pixels from the images [7]. However, data collected in the laboratory are usually generated under well-controlled conditions with carefully prepared samples and can be a poor match to field reflectance [8].

Another common tactic is to continue to use traditional multispectral classification procedures, adapting them to use hyperspectral data more effectively. The typical approach is to select a few suitable bands or simple spectral features to optimize multispectral algorithms, thereby reducing the hyperspectral data set to a multivariate data set tailored to the specific application (e.g., [6], [7], [9], [10]). The main drawback with simply selecting a small set of discrete bands is that there is a real danger that important spectral information will be ignored.

Several tools have been developed specifically to address the efficient selection of spectral features from hyperspectral data sets. Among these are band moments [11], orthogonal subspace projection [12], band prioritization and band decorrelation [13], and the minimum noise fraction transform [14], [15]. Although very different in approach, a major purpose of these methods is to reduce the dimensionality of the spectral domain with minimal loss of useful information. Even here, though, there is some danger of loss of important spectral information, particularly if the spectral details exhibit a small variance relative to other features in the data set.

If hyperspectral data are more than just a large number of bands from which one may choose an optimal subset, the unique information will be in the relative change of adjacent bands, i.e., in the shape of the spectra. Since spectral derivatives characterize spectral shape, they are likely candidates for capturing the spectral details that would be lost by other methods. Derivative analysis is a common method in laboratory spectroscopy [16], [17] and is also suitable for remote sensing hyperspectral analysis. Unlike other spectral analysis methods, derivative analysis does not necessarily require independent reference spectra nor does it depend on the magnitude of the signal. Derivative computation involves only the change in dependent variables relative to independent variables (wavelength). Thus, derivatives describe the shape of the spectral curve and are sensitive to changes in shape, not magnitude.

Relatively few researchers have pursued the derivative approach for hyperspectral remote sensing data analysis [18]–[24]. Although their algorithms have attained certain success, there are still limitations. These algorithms were either designed for specific applications (and may still require pre-existing spectral libraries) or only use a particular order of derivative. In a previous study [25], the authors developed smoothing and derivative procedures for spectral analysis to help identify subtle features from complex hyperspectral data sets. Their procedures were designed to treat laboratory or field collected hyperspectral data as spectrally continuous data. Their algorithms also allow

Manuscript received November 14, 2001; revised June 12, 2001.

F. Tsai is with the Los Alamos National Laboratory, Los Alamos, NM 87545 USA.

W. D. Philpot is with Cornell University, Ithaca, NY 14853 USA.

Publisher Item Identifier S 0196-2892(02)00967-1.

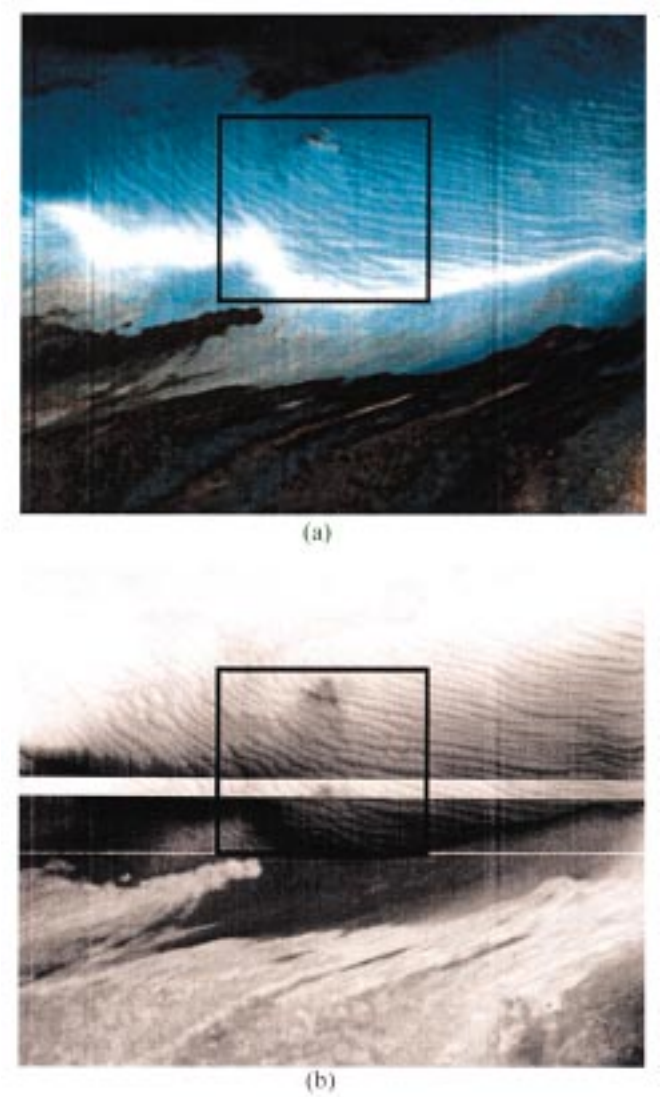

Fig. 1. Derivative feature enhancement of a PHILLS image: (a) original false-color image (B: 539.2 nm, G: 580.5 nm, R: 621.8 nm) and (b) second derivative image (center: 580.5 nm, $\mathrm{bw} = \mathrm{bs} = 41$ nm).

flexible choice of bandwidth and sampling intervals for computing derivative spectra in order to adapt to spectral features. The software has also been successfully applied to an investigation of vegetation reflectance and fluorescence spectra [26].

The point of using derivatives is to capture important spectral details that might be otherwise lost. As an illustration of the capacity of spectral derivatives to isolate spectral information, consider Fig. 1, which shows a false-color image displaying an area of sand shoals collected using the portable hyperspectral imager for low-light spectroscopy (PHILLS) sensor (1 January 1999, Lee Stocking Island, Bahamas) and a second-order derivative image [see Fig. 1(b)] of the same scene. Both images were created using exactly the same set of spectral bands. Inside the area marked by the bold rectangle, a triangular feature appears clearly in the derivative image but is difficult to distinguish from the noise in the color image. (The feature has recently been identified as an exposed patch of hard ground (limestone rock) bounded by a sandy bottom [27].) It is also interesting to note that seagrass, which appears as the black feature on the top portion of the original image, is a very low contrast feature in the derivative, making the triangular feature more easily distinguishable in the derivative image.

The example in Fig. 1 demonstrates the value of derivatives for extracting subtle information from a complex data set. On the other hand, including derivatives initially complicates the problem. For an N-band data set, as many as $N - 2M$ more potential features are added when an Mth order derivative spectrum is computed. If different bandwidths and sampling intervals are allowed this further expands the number of derivative features to choose from, and it is crucial to select only the most effective derivatives. Without a robust, methodical procedure for identifying and selecting the useful derivatives, it will be difficult to use derivatives in practice.

The remainder of this paper will be focused on the development and implementation of a strategy to use spectral derivatives to strengthen the classification of a vegetation land-cover study. Emphasis was placed on developing a system to help identify derivative spectra that can be added into the image in order to improve the separability among classes, especially among categories that are difficult to distinguish using standard classification methods.

II. Hyperspectral Classification and Derivatives

Supervised classification is one of the most frequently used procedures for quantitative analysis of remote sensing data. There are many multivariate algorithms designed for this task. However, because these algorithms were primarily designed to handle low-dimensional, multispectral data, problems can arise from applying them to high-dimensional hyperspectral data. Lee and Landgrebe [19] described the limitations inherent in using first-order classifiers and recommended second-order statistics for data with high dimensionality. They demonstrated that, as data dimension increases, the differences of class covariances become increasingly more important than class means. This suggests that a variance-based algorithm such as the maximum likelihood (ML) classifier can be effective when applied to hyperspectral image data.

The high dimensionality of hyperspectral data is a particularly serious concern when using a ML classifier. Not only does it significantly increase the computational load, there is evidence that the classification accuracy will actually decrease if the number of features (spectral bands) becomes too large [28, ch. 3, pp. 142–152]. The most obvious and direct method of reducing data dimensionality is to select only a few suitable bands for classification. This approach has been employed in a variety of applications [9], [18], [29], but it may also overlook subtle, but useful information in the original data. Besides the methods mentioned above [11]–[15], another commonly adopted feature reduction solution is spectral data transformation [30, ch. 10, pp. 239]. Among the transformation algorithms, principal component analysis (PCA) and canonical analysis are the most frequently used in remote sensing, [30, ch. 10, pp. 247]. The first few principal component or canonical axes will usually cover most of the data variations. As a result, if the low-variance, uncorrected variation is primarily random noise, data can be represented in several transformed bands without losing too much information [31].

There is another advantage of feature reduction. In theory, for an N-band data set, the minimum number of training pixels required for each class is $N+1$ to prevent the covariance matrix from becoming singular. In practice, it appears that $2N$ to $3.5N$ are needed (depending on the distances between class means) in order to obtain a misclassification rate within 5% of the optimal, and $10N$ to $20N$ are necessary for a within 1% misclassification rate of the optimal [32]. Derde and Massart [33] indicated that for a classification with at least ten variables the rule is generally correct, but it underestimates the sample size required for classifications with a restricted number of variables. Some researchers have suggested that $10N$ or even $100N$ training pixels are often necessary in order to reach a reliable classification result [28, pp. 151]. Unfortunately, remote sensing researchers are customarily forced to work with insufficient training data. Different techniques have been developed to overcome this problem [34]–[38]. However, these techniques may be either unable to preserve all the information, or require too much computational effort and may not work on all cases. In fact, the problem of inadequate training data can be solved by a simpler alternative, i.e., by cutting down the data dimensions, the required amount of training data is also reduced. Thus, spectral transformation (e.g., PCA) also indirectly helps minimize the requirement of training data size.

Nonetheless, PCA is not without flaws. For example, useful low-variance information will fall into later PCs and a large number of PCs may be required to include these features. Jia and Richards [39] presented a multistage segmented PC transformation for classifying hyperspectral images. However, some researchers argued that PCA may be optimal for representing data but not classification [40] and other feature extraction methods such as decision boundaries approach [41] and projection pursuit [42] may be more appropriate. However, given its relative simplicity and efficiency in representing large-variance features, PCA may still be useful if combined with algorithms that can capture low-variance, largely uncorrelated features. The task of capturing low variance, largely uncorrelated information again suggests the use of spectral derivatives [25]. Derivative analysis is independent to the feature reduction process. No matter what feature reduction method is used, it can always be combined with derivative analysis to construct a feature set for better classification. The task of derivative analysis is to identify the specific derivatives that will characterize the desired information. Adding the derivatives as features in the classification process will then optimize the classification and minimize the number of features required to achieve acceptable classification results.

III. Computational Procedure

To determine which features are helpful to classification, a standard must be established to measure the separability among classes. A simple and direct measurement is divergence [28, ch. 3, pp. 166–170], [30, ch. 10, pp. 240–244], but there are drawbacks to divergence as described in [28] and [30]. A better approach is to use the Jeffries-Matusita (JM) distance that computes the average distance between the density functions of two classes based on the Bhattacharrya distance [43], [44].

The JM distance between class-i and class-j, J_{ij}, is defined as

$$J_{ij} = \int_x \left(\sqrt{p\left(x|\omega\right)} - \sqrt{p\left(x|\omega\right)} \right) dx \tag{1}$$

where $p\left(x|\omega\right)$ is the class-conditional probability that a sample (pixel) will have the properties (spectral response); $\mathbf{x}$, give that the sample belongs to class ω and where $\mathbf{x}$ is a vector. For normally distributed class populations, (1) becomes

$$J_{ij} = 2\left(1 - e^{-\alpha}\right) \tag{2}$$

where

$$\alpha = \frac{1}{8}(\boldsymbol{m}_i - \boldsymbol{m}_j)^t \left\{ \frac{\Sigma_i + \Sigma_j}{2} \right\}^{-1} (\boldsymbol{m}_i - \boldsymbol{m}_j) + \frac{1}{2} \ln \left\{ \frac{\left| \frac{\Sigma_i + \Sigma_j}{2} \right|}{|\Sigma_i|^{1/2} |\Sigma_j|^{1/2}} \right\} \tag{3}$$

and Σ_i is the covariance matrix for the ith class.

As can be seen in (3), determinants of covariance matrices are required for the computation of JM distances. The calculation of matrix determinants, especially for matrices with high dimensionality, is likely to exceed machine precision boundaries and cause a numerical overflow or underflow in a computer system. To address this issue, the determinants in (3) can be calculated in their ten-based logarithmic scales. Therefore, (3) is rewritten as

$$\alpha = \frac{1}{8}(\boldsymbol{m}_j - \boldsymbol{m}_j)^t \left\{ \frac{\Sigma_i + \Sigma_j}{2} \right\}^{-1} (\boldsymbol{m}_j - \boldsymbol{m}_j) + \frac{1}{2}\delta \tag{4a}$$

with

$$\delta = \ln \left[\frac{10^{\lg(d_{ij})}}{\sqrt{10^{\lg(|\Sigma_i|)}} \sqrt{10^{\lg(|\Sigma_j|)}}} \right], \tag{4b}$$

$$d_{ij} = \left| \frac{\Sigma_i + \Sigma_j}{2} \right| \tag{4c}$$

which leads to

$$\delta = \ln(10) \left[\lg(d_{ij}) - \frac{1}{2} \lg\left(|\Sigma_i|\right) - \frac{1}{2} \lg\left(|\Sigma_j|\right) \right]. \tag{5}$$

As mentioned previously, derivatives can be used to identify features helpful to classification, i.e., resulting in larger JM distances. In this study, spectral derivatives were estimated using a finite approximation algorithm. For the first-order derivative of a spectrum, $s(\lambda)$, the estimation is

$$\frac{\partial s}{\partial \lambda} \approx \frac{s(\lambda_j) - s(\lambda_i)}{\Delta\lambda} \tag{6}$$

where $\Delta\lambda$ is the separation between adjacent bands, i.e., $\Delta\lambda = \lambda_j - \lambda_i$, and $\lambda_j > \lambda_i$. Similarly, the second derivative is

$$\frac{\partial^2 s}{\partial \lambda^2} \approx \frac{s(\lambda_k) - 2s(\lambda_j) + s(\lambda_i)}{(\Delta\lambda)^2} \tag{7}$$

where $\Delta\lambda = \lambda_k - \lambda_j - \lambda_i$; $\lambda_k > \lambda_j > \lambda_i$. Accordingly, higher orders of derivatives can be obtained using iteration from the

first derivative. Therefore, the nth derivative can be represented as

$$\left.\frac{\partial^n s}{\partial \lambda^n}\right|_j = \frac{\partial}{\partial \lambda}\left(\frac{\partial^{n-1} s}{\partial \lambda^{n-1}}\right) \approx \frac{s(\lambda_{i+n}) - \cdots + s(\lambda_i)}{(\Delta\lambda)^n} = \frac{\sum_i^{i+n} C_k s(\lambda_i)}{(\Delta\lambda)^n} \quad (8)$$

where $j = (2i+n)/2$, if $(2i+n)$ is even, or $j = (2i+n+1)/2$, if $(2i+n)$ is odd, C_k represents the weighting coefficients. In this study, the computer implementation of this procedure calculates individual $\Delta\lambda$ according to the real sampling intervals between two bands over the wavelength range. In addition, prior to the derivative calculation, the spectra can be smoothed. This is often necessary in order to minimize the noise in computing the derivative [45], [46]. The effects caused by different smoothing methods and parameters [47] are beyond the scope of this study and are not discussed here.

Using the above procedure, one can compute derivatives of the image, and identify the derivative features that generate larger JM distances among target classes. These features can then be included in the classification operation to improve the classification result. It is important to remember that no new information is created by using derivatives. The purpose is to identify the helpful spectral features that may be too subtle to be captured by other methods, so that the data dimension of the classification image can be kept as low as possible and still achieve accurate classification results.

IV. Materials and Analysis Procedure

The study site for this research is Jasper Ridge Biological Preserve (JRBP) located in the foothills of eastern Santa Cruz Mountains region near Palo Alto, CA, as shown in Fig. 2. The preserve consists of a rich population of diverse vegetation communities and includes a eutrophic lake and an intermittent stream. Elevations within the area range from 66 to 209 m above the sea level. A digital vegetation map obtained from the preserve administration was used as the primary reference data in the research. The map was created based on several aerial photographs supplemented by *in situ* ground-truth data. The aerial photographs were collected in the spring of 1995. These data were used (as the reference) to select training pixels from the study image and later used as a ground-truth image for comparing with classification results to assess the classification accuracies. In order to be compatible with the study image described later, the vector format vegetation map was converted to raster format (see Fig. 3) with 20-m spatial resolution.

The primary study imagery is an AVIRIS hyperspectral image (Flight number: 950 523C, Run 05, Scene 02) collected on 23 May 95. The image was acquired in the same season of the same year as the data used to create the vegetation map. Therefore, there should be minimum deviations due to temporal changes of vegetation and other subsequent developments on site. The 224-band AVIRIS image has 512 lines of 614 samples covering the ground from 37°25′39″N to 37°22′31″N and from 122°12′2″W to 122°16′47″W. The digital number (DN) data of the image had been radiometrically calibrated and the dark

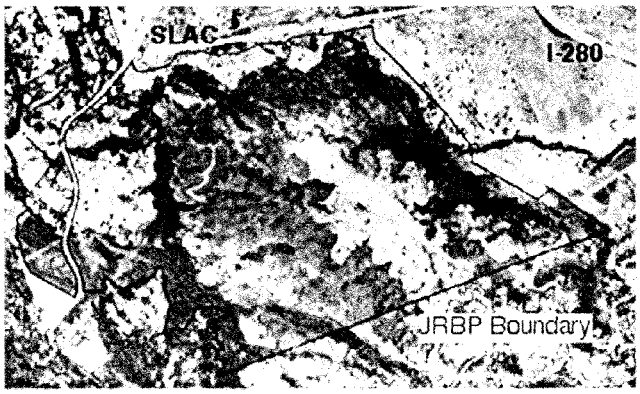

Fig. 2. Jasper Ridge Biological Preserve (JRBP) and vicinity (reproduced from a USGS DOQ).

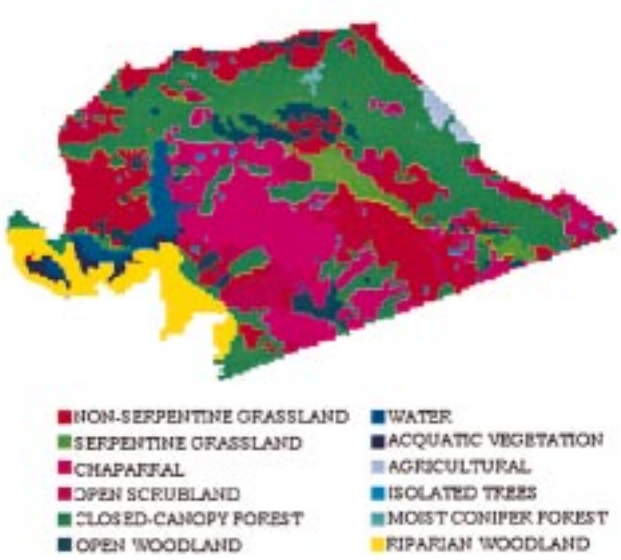

Fig. 3. Vegetation map of Jasper Ridge Biological Preserve (JRBP).

current spikes were removed, but the random noise was not filtered.

The first step of processing the AVIRIS image was to extract a subimage with a series of continuous wavelength bands but excluding wavelengths where data were too noisy because of instrument limitations or water absorption in the atmosphere. A range of wavelength bands (from 431.7 nm, up to before the water absorption region in the infrared, 1342.5 nm) was extracted from the original image in this step. In addition, because the AVIRIS instrument cannot cover the whole designated spectral range with one single sensor array, it uses several arrays with overlapping wavelength regions to ensure there are no gaps in the spectrum. One of the overlaps occurs within the extracted spectral range of the first process. Consequently, redundant bands within this overlap (band-33, 663.3 nm, to band-36, 692.02 nm; where band-32 is 687.0 nm and band-37 is 701.59 nm) were also removed from the extracted image, resulting in a 97-band image.

The 97-band image was then geographically registered to the vegetation map. The registration was done using ten positions with known coordinates as the control points and

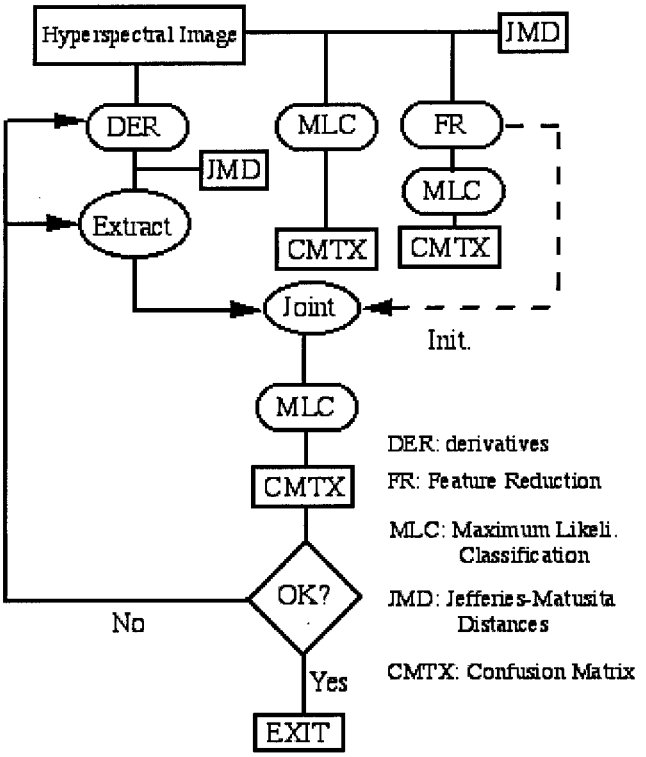

Fig. 4. Procedure for derivative-aided ML classification for hyperspectral image analysis.

using the nearest neighbor linear transformation for resampling. A subarea surrounding JRBP and its proximity was then extracted from the registered image. The final product of the pre-processing was a 195-sample by 114-line, 97-band image with 20-m sample spacing. This image was used as the initial working image for the subsequent ML classification and derivative analysis.

Before proceeding to the classification, a set of training data was randomly selected from the vegetation map. However, among the 12 ground types shown in Figs. 3, five of them were not considered valid classes because of an insufficient population of "qualified" candidates for training data. For the purposes of this study, a pixel is regarded as a qualified training data candidate if its adjacent neighbors all belong to the same class as the pixel itself. For each of the remaining classes, at least half of the qualified pixels were randomly selected as the training data for that class. These training data were applied to all subsequent supervised classification operations in the analysis procedure.

The analysis procedure for derivative-aided classification is illustrated in Fig. 4. It began with an initial ML classification of the original (97-band) working image. The confusion (contingency) matrix was also generated based on the classification result and the vegetation map. This served as the baseline classification against which subsequent classification procedures would be compared. The JM distances among classes in the image were also calculated for later use.

The next step is to construct a "base image" by a feature reduction method. In this study, the first ten PCs of the 97-band image were used as the base image. It was classified using the ML classifier, and its confusion matrix and class

TABLE I
CLASSIFICATION ACCURACIES OF THE 97-BAND IMAGE AND THE TEN-BAND PCA IMAGE

	97-band Image		10-band PCA Image	
	Producer	User	Producer	User
Non-Serpentine	79%	71%	67%	78%
Serpentine	31%	100%	67%	72%
Chaparral	68%	68%	36%	77%
Open Scrubland	19%	100%	14%	31%
Closed-Canopy Forest	77%	75%	55%	85%
Water	46%	95%	84%	85%
Riparian Woodland	61%	85%	79%	73%
Overall	69%		55%	
Kappa	0.5851		0.4547	

JM distances were also calculated. Comparing the confusion matrix of the base image to the 97-band images, some of the classes were found to have significantly lower accuracies in the ten-PC-image. The process was then focused on increasing their classification accuracies.

It was hoped that given the relatively low data dimension of the ten-PC-image, adding appropriate extra features to the data set would improve the classification result. As mentioned before, derivative analysis should be effective in extracting suitable features from the original image, because derivatives can capture subtle information from spectra. Therefore, derivative features with greater separability (large JM distances) of target classes were identified from the derivative images and were gradually appended to the base image to improve the classification. In the meantime, an equivalent number of PCs were also added into the ten-PC-image independently for comparison, but they were not essential to the analysis.

The procedure of appending derivative features was repeated until the classification result was satisfied or the classifier reached the limits (e.g., the data dimension became too high relative to the available training data). Sometimes, the appended derivative bands were highly correlated to the existing PCs or previously added derivative bands and could have resulted in a singular covariance matrix. In these cases, redundant bands were identified from the correlation matrix and removed from the image.

V. RESULTS AND DISCUSSION

For the seven classes considered in the analysis, Table I lists the producer's and user's accuracies of the ML classification (with a threshold of 0.86) from the 97-band image and the base image. There are several significant changes apparent in the producer's accuracies. Two trends are of particular interest. One is the increase of serpentine grassland, water and riparian woodlands in the ten-PC-image. The other is the decrease of non-serpentine grassland, chaparral and closed-canopy forest classification. Examination of the confusion matrices (see Table II) further reveals that, for classes with seriously decreased producer's accuracies in the PC-image, the number of unclassified pixels grows more significantly than for the other classes. This suggests that some of the useful information was excluded in the

TABLE II
(a) CONFUSION MATRICIES OF THE 97-BAND IMAGE AND (b) THE TEN-BAND PCA IMAGE (B)

	C1	C2	C3	C4	C5	C6	C7	Total
C0	134	22	203	78	262	12	62	773
C1	2056	147	107	189	351	14	17	2881
C2	0	110	0	0	0	0	0	110
C3	90	43	1535	109	276	147	44	2244
C4	0	0	0	118	0	0	0	118
C5	310	29	398	113	3127	7	190	4174
C6	0	0	6	0	1	167	1	175
C7	15	0	0	1	55	16	483	570
Total	2605	351	2249	608	4072	363	797	11045

(a)

	C1	C2	C3	C4	C5	C6	C7	Total
C0	533	59	1079	301	1213	14	100	3299
C1	1743	30	86	127	235	12	9	2242
C2	47	236	2	18	23	0	0	326
C3	31	8	799	34	135	10	24	1041
C4	86	8	33	86	62	1	2	278
C5	118	10	195	41	2234	3	29	2630
C6	0	0	44	0	9	306	3	362
C7	47	0	11	1	161	17	630	867
Total	2605	351	2249	608	4072	363	797	11045

(b)

C0: Unclassified (# of training pixels: N/A)
C1: Non-serpentine (718) C2: Serpentine (103) C3: Chaparral (828)
C4: Open Scrubland (107) C5: Closed-Canopy Forest (1560)
C6: Water (124) C7: Riparian Woodland (298)

ten-PC subset, making it difficult to correctly distinguish these classes.

On the other hand, for serpentine grassland, water, and riparian woodland, the confusion matrix from the 97-band image shows notable misclassification. The high misclassification rate indicates that there is considerable overlap among individual distribution functions in the original spectral space. It may also imply that the overlap is a result of a broadening of the distribution due to random noise or uncorrelated, lower variance features. For these classes, the distribution functions become more distinctive after spectral transformation, and the producer's accuracies are improved in the ten-PC-image. It is also possible that the classifier was more effective in the ten-image because the training data were more adequate than for the 97-band image.

As seen in Table I, chaparral was the class that suffered the greatest degradation in producer's accuracy in the ten-PC-image. Consequently, the analysis began with the attempt to restore chaparral accuracy by supplying derivative features helpful for separating chaparral pixels from the others. The goal was to regain the unclassified pixels of chaparral and other poorly classified classes without introducing greater misclassification to other classes.

Table III lists the accuracies of the base image with four and eight extra second-order derivative features with large JM distances between chaparral and others. The derivatives were calculated at five and nine sampling intervals after smoothing with bandwidths of five and nine samples, respectively.

In the table, it is apparent that not only have chaparral accuracies increased in both images, most of other classes are also improved. This indicates that the additional derivative features have provided useful information to the classifier, improving

TABLE III
ACCURACIES OF THE BASE IMAGE (TEN PCA) WITH FOUR AND EIGHT DERIVATIVE FEATURES

	PCA+4		PCA+8	
	Producer	User	Producer	User
Non-Serpentine	70%	77%	71%	76%
Serpentine	67%	73%	67%	74%
Chaparral	43%	76%	47%	77%
Open Scrubland	18%	38%	21%	42%
Closed-Canopy Forest	59%	84%	61%	84%
Water	85%	85%	85%	85%
Riparian Woodland	81%	72%	82%	72%
Overall	59%		61%	
Kappa	0.4990		0.5178	

TABLE IV
CONFUSION MATRIX OF BASE IMAGE WITH EIGHT DERIVATIVE FEATURES

	C1	C2	C3	C4	C5	C6	C7	Total
C0	384	46	787	221	865	8	74	2385
C1	1837	35	84	140	278	16	13	2403
C2	47	236	2	18	16	0	0	319
C3	57	12	1066	50	177	9	22	1393
C4	64	7	40	129	64	0	1	305
C5	170	15	216	49	2482	6	33	2971
C6	0	0	43	0	9	307	3	362
C7	46	0	11	1	181	17	651	907
Total	2605	351	2249	608	4072	363	797	11045

the chaparral classification without any substantial negative effect on other classes. The confusion matrix (see Table IV) further verifies that the additional derivative features indeed reduce the number of unclassified pixels and assign them to the correct class (chaparral).

In Tables III and IV, the misclassification rates are generally acceptable except for nonserpentine grassland (C1), chaparral (C3), and closed-canopy forest (C5). The confusion matrices also show that a large number of chaparral pixels are misclassified as closed-canopy forest, suggesting the two classes are difficult to separate in most, if not all, of the bands in the images. Therefore, a derivative feature maximizing the discrimination between chaparral and closed-canopy forest would be likely to most improve the classification.

The process of supplying derivative features with better separability to improve classification can be performed repeatedly with different derivative parameters until the overall accuracy is satisfied or the data dimension becomes too large for the classifier to produce a reliable classification. In this test case, 50 derivative features were ultimately added to the base image. The classification result of the final 60-band image is displayed in Fig. 5 with the same color scheme used in the vegetation map (see Fig. 3).

Fig. 6 illustrates the progression of overall accuracies and Kappa values during the derivative-adding process. Accuracies and Kappa values for PC-only images with equivalent number of bands are also plotted in the same figure for comparison. The accuracies are based on the pixels that are within JRBP and can be verified from the vegetation map. As expected, the Kappa value and overall accuracy of PC-images increase gradually as

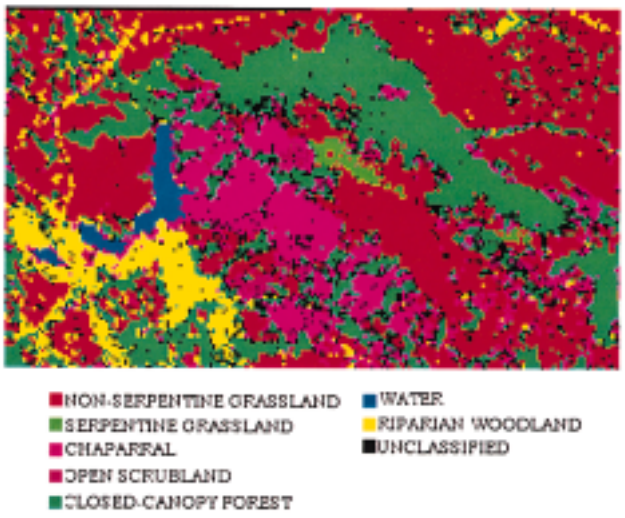

Fig. 5. Classification result of the base image with 50 derivative features.

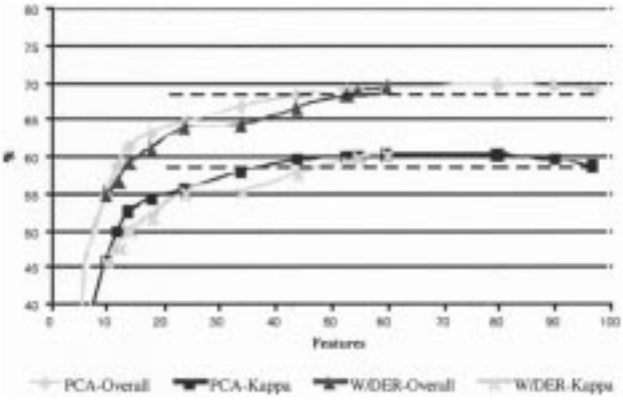

Fig. 6. Overall accuracy and kappa of PCA and derivative-added images (dashed lines represent values from the original 97-band image).

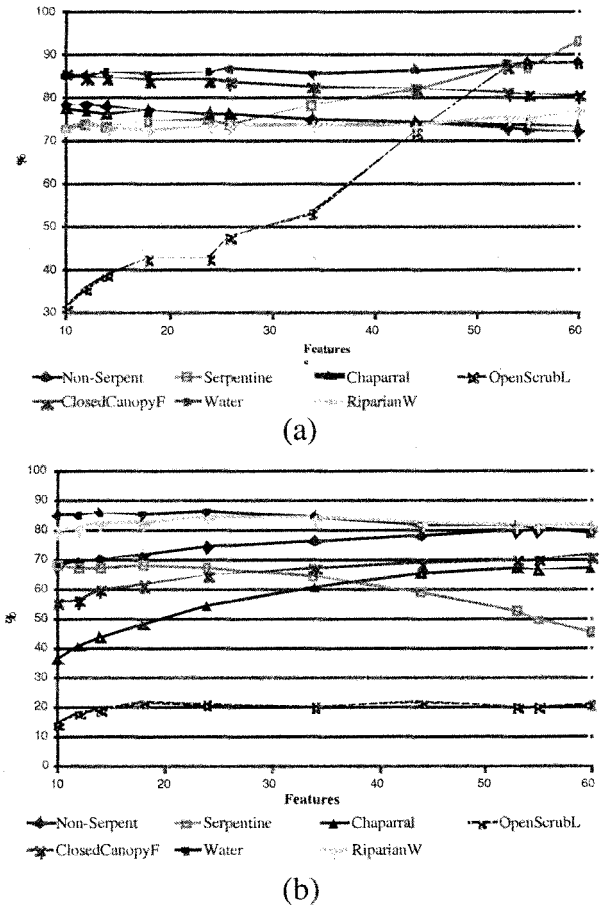

Fig. 7. (a) User's and (b) producer's accuracies of ten PCs with additional derivative features.

the number of PCs increases, eventually exceeding, and finally matching the accuracy for the original 97-band image. The improvements are initially large with the first few PCs, but the rate of increase slows significantly after more than ten PCs are included. In fact, the first ten PCs already account for 80% of the overall accuracy and 78% of the Kappa value achieved in the original 97-band image. This is the reason for using the first ten PC bands as the base image in the analysis.

According to the figure, the accuracy and Kappa value of the derivative-added image steadily improve as more derivative features are added. Both values had already surpassed the original 97-band image after 43 derivative images had been added. The results are nearly identical for an equivalent number of PC bands. Given that derivative analysis requires more intensive computation, it may not be a cost-effective method for the data set used for this case. Nonetheless, derivatives may be particularly useful if the data include subtle, low-variance information. PCA is known to be useful for excluding random noise (uncorrelated variance) from multivariate data. However, if the bands combinations considered to be noise and excluded by PCA are actually carrying useful information, derivative analysis may be able to detect the information and make it available for classification. The utility of a derivative feature may depend on the particular data set and classification requirements. In any case, it will be valuable to further discuss the results and phenomena observed in the process in order to better understand the effects caused by derivatives.

Fig. 7 displays the user's and producer's accuracies of individual classes during the analysis. The user's accuracy does not change much except for open scrubland and serpentine grassland, reflecting the decrease of misclassified pixels of these two classes as more derivative features are introduced into the image. The producer's accuracy of chaparral climbs steadily as more derivative features are appended, while accuracies of other classes are also increased or maintained at about the same level. This supports the process of choosing derivative features with better separability between chaparral and other classes.

Serpentine grassland is the only category showing noticeable decline in producer's accuracy. The confusion matrix of the final (60-band) image is listed in Table V for examination to determine the cause of the degeneration. According to the confusion matrix, there are 119 serpentine pixels misclassified as nonserpentine, in contrast to only 30 as shown in Table II-B. This is the primary cause for the decrease of serpentine grassland accuracy. Supplying extra derivative features with higher separability between serpentine and nonserpentine grasslands might well improve its classification. However, serpentine grassland is also one of the categories with relatively few available training pixels (due to their low populations in the preserve); when the

TABLE V
CONFUSION MATRIX OF THE FINAL (60 FEATURES) CLASSIFICATION

	C1	C2	C3	C4	C5	C6	C7	Total
C0	175	21	294	97	395	9	55	1046
C1	2065	119	116	195	364	15	12	2886
C2	5	157	2	5	3	0	0	172
C3	80	27	1490	105	281	32	29	2044
C4	6	4	1	124	7	0	0	142
C5	242	23	310	80	2876	6	55	3592
C6	0	0	32	0	7	286	1	326
C7	32	0	6	2	139	19	645	843
Total	2605	351	2249	608	4072	363	797	11051

data dimension increases, its classification may not improve as consistently as classes with abundant training pixels. This flaw is not a function of using derivatives. Rather, it is a function of using training data that are inadequate for describing high-dimensional distribution.

To further verify the relationship between data dimension and training data size, an experiment was carried out on two of the target classes being studied in this research. Taking the 60-band image as a test image, several rounds of ML classification were performed with different numbers of training pixels for serpentine grassland and closed-canopy forest. (The former represents the class with few training data, while the later is the one with ample training pixels available.) The relationship between training data size and producer's accuracy is illustrated in Fig. 8. In general, the accuracies gradually decrease as training data size shrinks. For closed-canopy forest, the accuracy remains at roughly the same level if the training data are at least five times the spectral dimension, but drops drastically when the ratio is less than two. This consequence corroborates the arguments discussed in [28], [32], and [33].

The process of adding additional features became difficult after 43 derivative bands had been appended to the base image in this study. Aside from the training data size issue, there is another consideration—it is also possible that some of the spectral bands are actually noise or closely correlated (but still within the tolerance of the classifier). Whether these bands are noise or highly correlated, removing them from the image should not significantly affect the classification. Details about this issue and possible solutions were discussed in [48]. One of the solutions proposed in [48] is particularly suitable for the derivative-aided analysis and merits further discussion here.

So far, the selected derivatives have been exclusively based on creating better separability between chaparral and any of the other classes. The risk of selecting redundant or closely correlated bands increases as more features are selected. According to Fig. 7, accuracies of other classes are also improved in addition to chaparral although the added derivatives are chosen for better "chaparral" classification. Therefore, it is possible to use derivatives intended for increasing accuracy of another class but still provide positive effects to chaparral classification. An ideal candidate in this case would be closed-canopy forest, the class to which chaparral is most likely to be misclassified according to Table V. To test this approach, after providing 24 chaparral-oriented derivative bands, the preference for derivative selection were switched to closed-canopy forest. As shown in Fig. 9, the Kappa and overall, producer's and user's accuracies are not

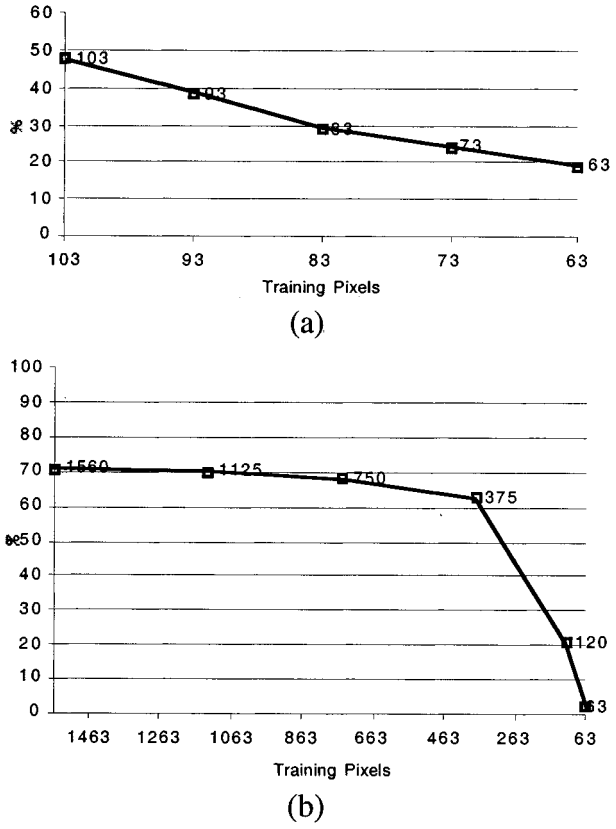

Fig. 8. Accuracies of (a) serpentine grassland and (b) closed-canopy forest versus training data size in the 60-feature image.

much different (± 3%) from the result of using chaparral-only derivatives. More importantly, the chaparral classification keeps improving even after the operation focus has switched to closed-canopy forest. This approach significantly strengthens the capability of the derivative-aided classifier to improve an individual class classification before causing a singular covariance matrix.

Among the seven categories considered for classification in the preserve, open scrubland has a particularly low accuracy in all operations. The primary reason is probably due to the nature of the class, which is actually a mix of several cover types. With the relatively low spatial resolution of AVIRIS imagery, the inhomogeneity of open scrubland pixels may result in high misclassification. The confusion matrices listed in Tables IV and V tend to support this hypothesis. Open scrubland pixels are misclassified more than twice as often as they are classified correctly in both cases.

VI. SUMMARY AND CONCLUSION

Spectral derivatives have the potential to be useful for efficient classification of hyperspectral images. By supplying derivative features with better separability (larger JM distances) for specific classes, it may be possible to improve classification accuracy with a smaller number of features. Using producer's accuracy as the benchmark, the tests conducted in this study shows that with 43 derivative features and the first ten PCs, four of the seven targets are classified more accurately than with the original 97-band image, and one class remains at the same level. For the remaining two classes, the derivative-added

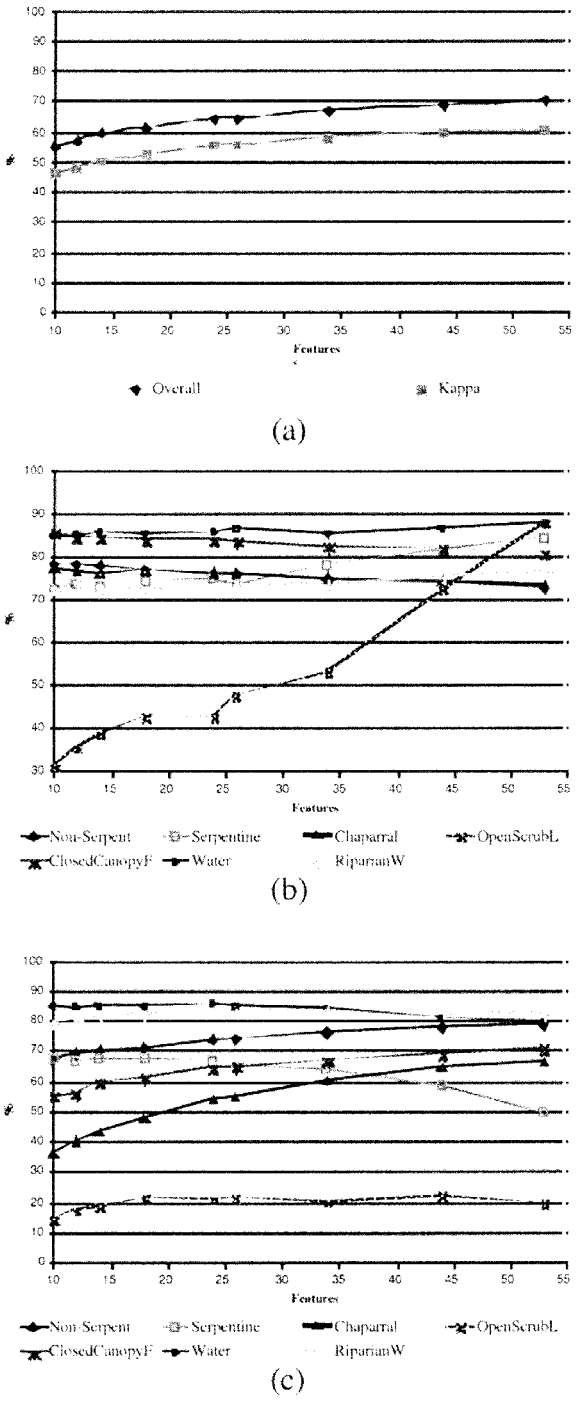

Fig. 9. (a) Kappa and overall, (b) user's, and (c) producer's accuracies of base image with derivatives for chaparral and closed-canopy forest.

image has reached at least 90% of the original accuracy. After adding 50 derivatives, the result indicates that five classes are better than in the original 97-band image and the remaining two are classified at least 92% as accurately as the original.

Regardless of the nature of the spectral features used for classification, the size of the training data set relative to the dimensionality of the feature space will limit the classification accuracy. Feature reduction algorithms such as PCA may or may not address this issue completely. Principal component transformation or other variance-maximizing procedures are good at capturing the bulk of information in relatively few features. However, they are not ideal for extracting subtle information from complex spectra. This may not be obvious in the overall classification, but is apparent when examining accuracies of individual classes. Derivative analysis is efficient in capturing subtle information and can be used in conjunction with feature reduction algorithms to construct a feature set for better classification. The task is to select a relatively small number of features that will best discriminate among the desired classes. The concern of data dimension and training data size emphasizes the need to compress useful information into the minimum number of features. A combination of variance-based and shape-based (derivative) features is likely to be an effective and efficient solution.

The described derivative-aided analysis procedure can be used as a prototype for developing a robust hyperspectral image analysis system that can handle high-dimensional images more efficiently and effectively. The design of the procedure is systematic and modularized. It is ready for automatic and computerized implementation. However, there are still places for improvements and extensions. For example, the iteration for selecting appropriate derivative features can be engineered more intelligently to reduce computational resources requirements as well as to speed up the process. In addition, a redundancy check can also be introduced into the derivative selection procedure to reduce the risk of generating a singular covariance matrix.

Although vegetation classification was the primary example used in this study, the developed derivative analysis procedure can also be used for other applications with little or no modifications. Furthermore, there is no reason that derivative analysis should be restricted in ML or other supervised classification systems. It should be valuable to explore the possibility of applying derivative analysis to unsupervised classification and other remote sensing applications.

REFERENCES

[1] J. A. Benediktsson, J. R. Sveinsson, and K. Arnason, "Classification and feature extraction of AVIRIS data," *IEEE Trans. Geosci. Remote Sensing*, vol. 33, pp. 1194–1205, May 1995.

[2] F. A. Kruse *et al.*, "The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data," *Remote Sens. Environ.*, vol. 44, pp. 145–163, 1993.

[3] H. Cetin, T. A. Warner, and D. W. Levandowski, "Data classification, visualization and enhancement using n-dimensional probability density functions (nPDF): AVIRIS, TMS, and geophysical applications," *Photogramm. Eng. Remote Sens.*, vol. 59, no. 12, pp. 1755–1764, 1993.

[4] Z. Chen, P. J. Curran, and J. D. Hansom, "Derivative reflectance spectroscopy to estimate suspended sediment concentration," *Remote Sens. Environ.*, vol. 40, no. 1, pp. 67–77, 1992.

[5] J. Peñuelas *et al.*, "Reflectance indices associated with physiological changes in nitrogen-and water-limited sunflower leaves," *Remote Sens. Environ.*, vol. 48, no. 2, pp. 135–146, 1994.

[6] B. N. Rock *et al.*, "High-spectral resolution field and laboratory optical reflectance measurements of red spruce and eastern hemlock needles and branches," *Remote Sens. Environ.*, vol. 47, pp. 176–189, 1994.

[7] J. P. Hoffbeck and D. A. Landgrebe, "Classification of remote sensing images having high-spectral resolution," *Remote Sens. Environ.*, vol. 57, pp. 119–126, 1996.

[8] J. A. Gamon, J. Peñuelas, and C. B. Field, "A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency," *Remote Sens. Environ.*, vol. 41, no. 1, pp. 35–44, 1992.

[9] M. E. Martin *et al.*, "Determining forest species composition using high-spectral resolution remote sensing data," *Remote Sens. Environ.*, vol. 65, pp. 249–254, 1998.

[10] E. W. Chappelle and M. S. Kim, "Ratio analysis of reflectance spectra (RARS): An algorithm for the remote estimation of the concentrations of chlorophyll A, chlorophyll B, and carotenoids in soybean leaves," *Remote Sens. Environ.*, vol. 39, pp. 239–247, 1992.
[11] K. Staenz, "Classification of a hyperspectral agricultural dataset using band moments for reduction of the spectral dimensionality," *Can. J. Remote Sens.*, vol. 22, no. 3, pp. 248–257, 1996.
[12] J. C. Harsanyi and C. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 779–785, July 1994.
[13] C. Chang *et al.*, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, no. 6, pp. 2631–2641, 1999.
[14] A. A. Green *et al.*, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sensing*, vol. 26, pp. 65–74, Jan. 1988.
[15] J. B. Lee, A. S. Woodyatt, and M. Berman, "Enhancement of high-spectral resolution remote-sensing data by a noise-adjusted principal components transform," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, pp. 295–304, May 1990.
[16] W. L. Butler and D. W. Hopkins, "Higher derivative analysis of complex absorption spectra," *Photochem. Photobiol.*, vol. 12, pp. 439–450, 1970.
[17] G. Talsky, *Derivative Spectroscopy*. Weihein, Germany: Weinhein Verlagsgesellschaft, 1994.
[18] T. H. Demetriades-Shah, M. D. Steven, and J. A. Clark, "High-resolution derivatives spectra in remote sensing," *Remote Sens. Environ.*, vol. 33, pp. 55–64, 1990.
[19] C. Lee and D. A. Landgrebe, "Analyzing high-dimensional multispectral data," *IEEE Trans. Geosci. Remote Sensing*, vol. 31, pp. 792–800, 1993.
[20] Z. Chen, D. Elvidge, and W. T. Jansen, *Description of Derivative-Based High Spectral-Resolution (AVIRIS) Green Vegetation Index*. Bellingham, WA: SPIE, 1993.
[21] Z. Chen, D. Elvidge, and D. P. Groeneveld, "Monitoring seasonal dynamics of arid land vegetation using AVIRIS data," *Remote Sens. Environ.*, vol. 65, pp. 255–266, 1998.
[22] D. G. Goodin *et al.*, "Analysis of suspended solids in water using remotely sensed high-resolution derivative spectra," *Photogramm. Eng. Remote Sens.*, vol. 59, no. 4, pp. 505–510, 1993.
[23] H. Holden and E. LeDrew, "Spectral discrimination of healthy and nonhealthy corals based on cluster analysis, principal component analysis, and derivative spectroscopy," *Remote Sens. Environ.*, vol. 65, pp. 217–224, 1998.
[24] M. L. Adams, W. D. Philpot, and W. A. Norvell, "Yellowness index: A New measure of chlorosis in stressed vegetation," *Int. J. Remote Sens.*, vol. 20, no. 18, pp. 3663–3677, 1999.
[25] F. Tsai and W. D. Philpot, "Derivative analysis of hyperspectral data," *Remote Sens. Environ.*, vol. 66, no. 1, pp. 41–51, 1998.
[26] W. D. Philpot *et al.*, "Analysis of reflectance and fluorescence spectra for atypical features: Fluorescence in the yellow-green," *J. Plant Physiol.*, vol. 148, no. 5, pp. 567–573, 1996.
[27] R. Zimmerman, *personal communication* Moss Landing, CA, 2000.
[28] P. H. Swain and S. M. Davis, *Remote Sensing: The Quantitative Approach*. New York: McGraw-Hill, 1978.
[29] P. J. Curran *et al.*, "Reflectance spectroscopy of fresh whole leaves for the estimation of chemical concentration," *Remote Sens. Environ.*, vol. 39, pp. 153–166, 1992.
[30] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis 1998*, 3rd ed. New York: Springer-Verlag, 1998.
[31] D. F. Morrison, *Multivariate Statistical Methods*, 2nd ed. New York: McGraw-Hill, 1976.
[32] P. A. Lachenbruch, *Discriminant Analysis*. New York: Hafner, 1975.
[33] M. P. Derde and D. L. Massart, "Evaluation of the required sample size in some supervised pattern recognition techniques," *Analytica Chimica Acta*, vol. 223, no. 1, pp. 19–44, 1989.
[34] J. P. Hoffbeck and D. A. Landgrebe, "Covariance matrix estimation and classification with limited training data," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 18, no. 7, pp. 763–767, 1996.
[35] X. Jia and J. A. Richards, "Efficient maximum likelihood classification for imaging spectrometer data sets," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 274–281, 1994.
[36] C. Lee and D. A. Landgrebe, "Fast likelihood classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 29, pp. 509–517, 1991.
[37] R. E. Roger, "Sparse inverse covariance matrices and efficient maximum likelihood classification of hyperspectral data," *Int. J .Remote Sens.*, vol. 17, no. 3, pp. 589–613.
[38] B. M. Shahshahani and D. A. Landgrebe, "The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 1087–1095, May 1994.
[39] X. Jia and J. A. Richards, "Segmented principal components transformation for efficient hyperspectral remote-sensing data display and classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 538–542, Jan. 1999.
[40] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York: Academic, 1990.
[41] C. Lee and D. A. Landgrebe, "Feature extraction based on decision boundaries," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 4, pp. 388–400, 1993.
[42] L. O. Jimenez and D. A. Landgrebe, "Hyperspectral data analysis and supervised feature reduction via projection pursuit," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 2653–2667, Nov. 1999.
[43] A. Bhattacharrya, "On a measurement of divergence between two statistical populations defined by their probability density functions," *Bull. Calcutta Math.*, vol. 35, pp. 99–109, 1943.
[44] T. Kailath, "The divergence and Bhattacharrya distance measure in signal selection," *IEEE Trans. Commun. Theory*, vol. COM-15, pp. 52–60, 1967.
[45] S. Kawata and S. Minami, "Adaptive smoothing of spectroscopic data by a linear mean-square estimation," *Appl. Spectrosc.*, vol. 38, no. 1, pp. 49–58, 1984.
[46] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Anal. Chem.*, vol. 36, pp. 1627–1639, 1964.
[47] F. Tsai, "Derivative analysis of hyperspectral data," M.S., School of Civil and Environmental Engineeering, Cornell Univ., Ithaca, NY, 1996.
[48] ——, "Derivative Hyperspectral Image Analysis for Land use Classification," Ph.D., Sch. Civil and Environ. Eng.., Cornell Univ., Ithaca, New York, 2000.

Fuan Tsai received the B.S. degree in civil engineering from National Taiwan University, Taipei, Taiwan, R.O.C., in 1989, and the M.S. and Ph.D. degrees in civil and environmental engineering with a concentration in remote sensing from Cornell University, Ithaca, NY, in 1996 and 2000, respectively.

He served as a Junior Officer in the R.O.C. Marine Corps from 1989 to 1991, and was a Civil Design Engineer from 1991 to 1993. From 1996 to 1998, he was a Graduate Research Assistant in the Los Alamos National Laboratory (LANL), Los Alamos, NM, where he participated in several remote sensing and visualization projects. He is currently a Postdoctoral Research Associate at LANL. His research interests include remote sensing/hyperspectral image analysis, geographic information sciences, and scientific visualization/phenomenology.

William D. Philpot (M'88) received the B.A. degree in music from New York University in 1969, the B.S. degree in physics from the State University of New York at Stony Brook in 1973, and the M.S. and Ph.D. degrees in marine studies from the University of Delaware, Newark, in 1978 and 1981, respectively.

He served two years in the U.S. Army. In 1981, he joined the faculty of the Department of Civil and Environmental Engineering, Cornell University, Ithaca, NY, where he now teaches courses in remote sensing and digital image processing. His research interests are in the physics of optical remote sensing, spatial and spectral pattern recognition, and image processing. His current research includes the analysis of hyperspectral data for bathymetry and water quality in coastal waters and the analysis of bathymetric lidar data for water quality and bottom characteristics.