

Unsupervised Hyperspectral Image Analysis with Projection Pursuit

Agustin Ifarraguerri, *Member, IEEE*, and Chein-I Chang, *Senior Member, IEEE*

Abstract—Principal components analysis (PCA) is effective at compressing information in multivariate data sets by computing orthogonal projections that maximize the amount of data variance. Unfortunately, information content in hyperspectral images does not always coincide with such projections. We propose an application of projection pursuit (PP), which seeks to find a set of projections that are “interesting,” in the sense that they deviate from the Gaussian distribution assumption. Once these projections are obtained, they can be used for image compression, segmentation, or enhancement for visual analysis. To find these projections, a two-step iterative process is followed where we first search for a projection that maximizes a projection index based on the information divergence of the projection’s estimated probability distribution from the Gaussian distribution and then reduce the rank by projecting the data onto the subspace orthogonal to the previous projections. To calculate each projection, we use a simplified approach to maximizing the projection index, which does not require an optimization algorithm. It searches for a solution by obtaining a set of candidate projections from the data and choosing the one with the highest projection index. The effectiveness of this method is demonstrated through simulated examples as well as data from the hyperspectral digital imagery collection experiment (HYDICE) and the spatially enhanced broadband array spectrograph system (SEBASS).

Index Terms—HYDICE, hyperspectral image, principal components analysis, projection pursuit (PP), SEBASS.

I. INTRODUCTION

HYPERSPECTRAL imaging sensors are capable of generating unprecedented volumes of radiometric data. The airborne visible/infrared imaging spectrometer (AVIRIS), for example, routinely produces image cubes with as many as 512×614 pixels and 224 spectral bands. If these data are to be used for practical purposes, they must be reduced to allow analysis by users as well as by computers. Data reduction is an operation that enhances the information content of the data in some way. A common data reduction method in situations where there is no *a priori* knowledge about a scene is principal components analysis (PCA). PCA is effective at compressing information in multivariate data sets by computing orthogonal projections that maximize the amount of data variance. It is typically performed through the eigen-decomposition of the spectral covariance matrix of an image cube. The information

can then be presented in the form of component images, which are projections of the image cube on to the eigenvectors. The component images corresponding to the larger eigenvalues are presumed to preserve the majority of the information about the scene. Unfortunately, information content in hyperspectral images does not always coincide with such projections for several reasons.

- 1) PCA is optimal when the background associated with signal sources (the clutter) is Gaussian white noise. This is often not the case in hyperspectral images, where the clutter includes contributions from interference sources such as natural background signatures [1], [2] as well as structured (nonrandom) noise such as striping. When the Gaussian assumption does not hold, the background clutter can become indistinguishable from the signal sources [2].
- 2) The objects of interest are often small relative to the size of the scene, and therefore contribute a small amount to the overall variance. PCA often fails to capture the variability associated with small objects unless their spectra are nearly orthogonal to the background spectra.

An alternative to PCA that can alleviate some of these problems is the minimum noise fraction (MNF) transform [2]–[4]. This transform, also known as noise-adjusted principal components, was designed to produce orthogonal component images that are ordered by image quality as measured by the SNR, rather than by the data variance. The MNF transform is equivalent to a sequence of two orthogonal transformations, where the first rotates the data such that the noise covariance matrix is diagonalized, thus “whitening” the noise, followed by a standard PCA transform. PCA performance is improved because the noise effects on signal sources are minimized by the whitening process. However, the MNF transform still depends on “bulk” image properties, so it is not generally sensitive to small objects.

This paper expands on a previous publication by the authors [5] where an application of projection pursuit (PP) to hyperspectral data reduction and analysis was discussed. PP is an exploratory data analysis technique used to visualize high-dimensional data in a low-dimensional space [6]–[8]. In PP, we look for a set of linear projections that are “interesting,” in the sense that they deviate from the Gaussian distribution assumption [9]. If we let \mathbf{y} be a column vector representing a data sample, then we can compute a scalar “projection score” x by

$$x = \mathbf{w}^T \mathbf{y} \quad (1)$$

where \mathbf{w} is the projection vector. A projection is the collection of the projection scores produced by all the data vectors. This ap-

Manuscript received January 4, 1999; revised September 15, 1999.

A. Ifarraguerri is with the Biological and Chemical Command, U.S. Army, Edgewood Chemical Biological Center, Aberdeen Proving Ground, MD 21010 USA.

C.-I. Chang is with the Remote Sensing Signal and Image Processing Laboratory, Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD 21250 USA (e-mail: cchang@umbc.edu).

Publisher Item Identifier S 0196-2892(00)07162-X.

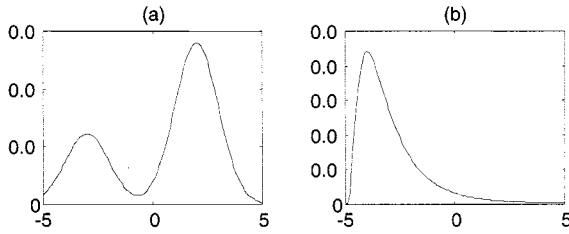


Fig. 1. Conceptual score distributions for which the projection score should attain a high value. (a) Bimodal distribution, indicating spectral classes, and (b) skewed distribution, indicating outliers.

proach is preferable to eigenanalysis of the spectral covariance since it does not depend on the white Gaussian noise assumption and is not biased toward larger objects (or spectral classes) that contribute more variance. To carry out PP analysis, we must first select an appropriate “projection index” to maximize. A projection index is a measure of the desirability of a given projection. In PCA, which can be viewed as a special case of PP, this index is the variance of the projection scores. Thus, PCA seeks projections that maximize the magnitude of signal sources (or “signal energy”). However, information related to the segmentation of different objects and background types does not generally align itself with signal magnitude. Whereas PCA and MNF provide the analyst with images ordered by a measure of signal strength, they do not necessarily represent the spectral information content of the signals of interest. We would like to find projections that cause a single pixel or a small cluster of projection scores to become “separated” from the bulk of the data points (outliers). Such a projection would serve to locate spectrally anomalous objects or low probability targets regardless of their relative size. Another desirable type of projection is one that produces multimodal projection score distributions, indicating different spectral classes. Fig. 1 shows conceptual score distributions for which our projection index should attain a high value.

The types of projections described above can be obtained by using a projection index that considers the distribution of projection scores. The projection index should attain a high value for score distributions with strong outliers or multiple modes, suggesting a spectral feature different than that of the background. An important projection index used in PP is the one proposed by Friedman [9], which is based on the mean-squared difference between the projection score distribution and the Gaussian distribution. This index gives a high value for bimodal projection score distributions, but not necessarily for those with outliers. An alternative to Friedman’s index, suggested by Huber [8] and Jones and Sibson [7], is to use information-theoretic measures such as entropy, computed on the projection score distribution. One such measure is the information divergence (or Kullback–Leibler distance) [10]. Information divergence is an information-theoretic criterion that can be used to measure the “distance” between two probability distributions. We can construct a projection index by computing the information divergence between the estimated probability distribution of a projection and the normal distribution. A high projection index indicates that its distribution is highly non-normal, and therefore presumed interesting.

The PP method presented here is designed to obtain an orthonormal transform. There is no restriction in the theory that requires the computed projections to be mutually orthogonal, but removing this requirement makes computations significantly more complex and takes away the advantages of orthogonality. To find a set of orthogonal projections, a two-step iterative process is presented. We first search for a projection that maximizes the projection index and then reduce the rank of the data matrix by projecting it onto the subspace orthogonal to all the previous projections. Once these projections are obtained, they can be used for image compression, segmentation, or enhancement for visual analysis. PP analysis of hyperspectral imagery has been reported previously by Jimenez and Landgrebe [11]. They proposed a method based on the Bhattacharyya distance as the projection index, requiring a training set with labeled samples. Our approach does not require any *a priori* knowledge about the scene. More recently [12], Chiang and Chang successfully used skewness and kurtosis as the basis of a projection index for unsupervised target detection with hyperspectral imagery.

In Section II of this paper, we define the projection index and show how it can be computed. Section III then introduces an orthogonal transform based on our projection index. In Sections IV and V, we illustrate our method with examples, and present our conclusions in Section VI.

II. PROJECTION INDEX

The key to success in PP is the selection of an appropriate projection index. In this section, we define the information divergence index and show how it can be estimated in practice. Given two continuous probability distributions $f(x)$ and $g(x)$ of the projection score x , the relative entropy of $f(x)$ with respect to $g(x)$ is defined as

$$d(f||g) \equiv \int_{-\infty}^{\infty} g(x) \log \frac{f(x)}{g(x)} dx. \quad (2)$$

The absolute information divergence between $f(x)$ and $g(x)$ can be defined by

$$j(f, g) \equiv |d(f||g)| + |d(g||f)|. \quad (3)$$

which is symmetric, i.e., $j(f, g) = j(g, f)$, and nonnegative. The absolute information divergence is zero if $f = g$ and increases as the two distributions diverge. If we let $g(x)$ be the Gaussian distribution, then we can compute the divergence of $f(x)$ from normality. To do this, we must estimate $f(x)$ from the data. One possible approach is to use a kernel estimator such as a Parzen window [13], but the resulting integral to be solved would result in a cumbersome expression. A simpler and effective approach is to approximate the continuous distributions $f(x)$ and $g(x)$ with their discrete counterparts \mathbf{p} and \mathbf{q} respectively. The relative entropy of (2) for discrete distributions is defined as:

$$D(\mathbf{p}||\mathbf{q}) \equiv \sum_i p_i \log \frac{p_i}{q_i} \quad (4)$$

where p_i and q_i are the i th components of the \mathbf{p} and \mathbf{q} vectors respectively. Then the absolute information divergence becomes

$$J(\mathbf{p}, \mathbf{q}) \equiv D(\mathbf{p}||\mathbf{q}) + D(\mathbf{q}||\mathbf{p}). \quad (5)$$

The absolute value is not needed in the discrete case since $D(\mathbf{p}||\mathbf{q}) \geq 0$ for all discrete distributions \mathbf{p} and \mathbf{q} , though this is not necessarily true for continuous probability distribution functions [10]. We can now compute the projection index as follows.

- 1) Select the number of bins n and their width Δx for quantization of $f(x)$ and $g(x)$.
- 2) Standardize the projection scores by subtracting their mean and dividing by their standard deviation [i.e., $(x - \bar{x})/\sigma_x$].
- 3) Construct the projection score histogram \mathbf{p} from the standardized data and normalize \mathbf{p} to unity, i.e., $\sum_i p_i = 1$.
- 4) Quantize the standard Gaussian distribution by computing the integral $q_i = 1/\sqrt{2\pi} \int_{i\Delta x}^{(i+1)\Delta x} e^{-t^2/2} dt$ for $i = -n/2$ to $n/2$. This is the value that p_i would assume in the limit as the number of pixels goes to infinity if the projection were Gaussian.
- 5) Compute the projection index $J(\mathbf{p}, \mathbf{q})$ using (4) and (5).

The number of bins and their width should be chosen to cover at least the range $[-5\sigma, 5\sigma]$ in order to retain the majority of the Gaussian shape, but the number should be increased depending on the maximum and minimum values of the projection scores. The bin width should also be chosen small enough to provide an accurate estimate of the divergence, but not so small that it causes numerical instability. For the examples in this paper, we have used $\Delta x = 1$, which appears to be appropriate for sample sizes of about 1000 points. If a simulation is performed where a number of Gaussian random numbers are generated, the computed divergence of the samples from the true Gaussian will change significantly as a function of Δx , with a minimum at a location that depends on the sample size. For a sample size of 1000, the optimal Δx is close to 1 and decreases with increasing sample size. Also, the zero values in \mathbf{p} are replaced by small positive numbers to avoid numerical instability in the projection index estimate (a “small” number is defined here as one that would correspond to less than one sample occupying the bin). This has a smoothing effect on the distribution estimate.

As discussed previously, we would like to capture both multimodality and skewness with our projection index. Fig. 1 shows some interesting non-Gaussian shapes that, in our experience, tend to give informative projections with hyperspectral images. The divergence index was chosen for its ability to detect all such distribution shapes.

III. PROJECTION PURSUIT TRANSFORM

A. Preprocessing

It is a common practice in PP to prereduce the dimensionality of the data by using PCA [8]. This process is sometimes known as “sphering,” because all the components are given equal weight by scaling them with their corresponding eigenvalue. This step also serves as a first pass reduction by keeping only projections associated with significant eigenvalues, thus

discarding the noise subspace. If \mathbf{D} is the diagonal eigenvalue matrix and \mathbf{U} is the matrix whose columns contain the eigenvectors corresponding to the eigenvalues in \mathbf{D} , then a pixel vector (spectrum) \mathbf{y} is transformed by

$$\mathbf{z} = \mathbf{D}^{-1/2} \mathbf{U}^T (\mathbf{y} - \bar{\mathbf{y}}) \quad (6)$$

where $\bar{\mathbf{y}}$ is the sample mean of \mathbf{y} . We can then retain the elements of \mathbf{z} associated with the largest eigenvalues in \mathbf{D} .

B. Obtaining the Projections

Since there is no analytical expression for the projection index given in (5), obtaining a projection vector that maximizes the projection index requires a numerical optimization technique. Several different strategies are discussed in the literature [7]. In this paper, we describe a simple search strategy that, although not optimal, is intuitively straightforward and computationally manageable. The premise of our approach is that there are interesting projection vectors located in or near the data “cloud” (i.e., the multidimensional area occupied by the data points), and that they can be approximated by the pixel spectrum nearest to it in terms of the inner product (zero-order approximation). To find the best solution available in the data, we simply take each pixel spectrum as a candidate projection vector, compute the projection index and then select the one that produces the highest value. We can illustrate this point graphically by using a simple two-dimensional (2-D) example. Fig. 2 shows a scatter plot of samples that have a non-Gaussian distribution. Clearly, a projection along the horizontal dimension is more interesting than a projection along the vertical, and in fact, will give a higher projection index value. In this simple case, there is a data point along almost every direction from the origin, so we can get a very good answer by using the data points as candidate projection vectors and choosing the one that provides the largest projection index. A similar argument can be applied to a higher dimensional data space. If a high multidimensional data set is interesting, the data points will not be spread evenly, but instead will form clusters or other structures along some directions. These directions are the desired projection vectors. Therefore, even if the number of dimensions increases, the interesting projection directions of a data cloud will not be arbitrary, but will in fact still be aligned with its data points.

Starting with a matrix $\mathbf{Z} = [\mathbf{z}_1 \mathbf{z}_2, \dots, \mathbf{z}^m]$ whose columns contain the m pixel spectra vectors \mathbf{z}_i transformed via (6), we can compute an m -dimensional vector of projection scores $\mathbf{Z}^T \mathbf{z}_i$. Each single column vector \mathbf{z}_i in \mathbf{Z} acts as a candidate projection operator. The information divergence from normality for each set of projection scores can then be estimated following the algorithm outlined in the previous section. Let \mathbf{q} be the quantized standard Gaussian distribution for some Δx . The pixel that produces the highest value of the projection index $J(\mathbf{p}(\mathbf{Z}^T \mathbf{z}_i), \mathbf{q})$ is then selected as the best projection

$$i_{best} = \arg \left\langle \max_i J(\mathbf{p}(\mathbf{Z}^T \mathbf{z}_i), \mathbf{q}) \right\rangle. \quad (7)$$

with its corresponding pixel as the projection vector

$$\mathbf{w} = \frac{\mathbf{z}_{i_{best}}}{\|\mathbf{z}_{i_{best}}\|} \quad (8)$$

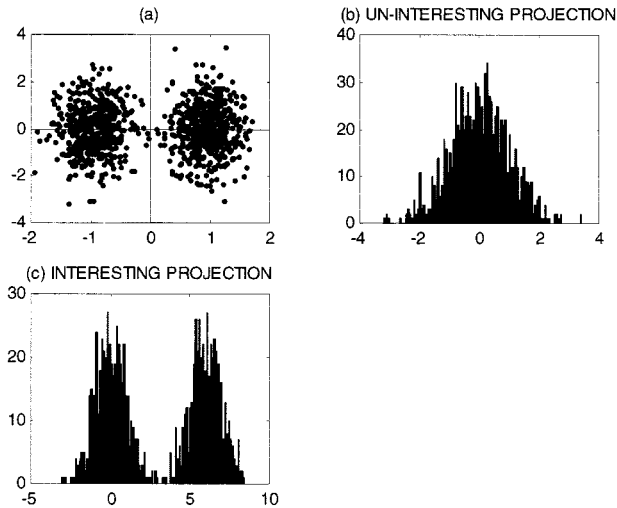


Fig. 2. (a) Scatter plot of samples that have a non-Gaussian distribution, (b) uninteresting vertical projection, and (c) interesting horizontal projection.

where $\|\bullet\|$ indicates the Euclidean norm. In practice, this computation of the projection index is of complexity $O(m)$. Since this operation is done m times per iteration, the total complexity is $O(m^2)$ per iteration. By using a fixed-size sample of the pixels to estimate the probability distribution of the projections, we can reduce the complexity to $O(m)$. We have found in practice that selecting a subset of the pixels to estimate \mathbf{p} yields results that are often indistinguishable from those using all the data. In the implementation of our algorithm, 1000 pixels sampled from the image at uniform intervals were used.

After the best projection \mathbf{w} is obtained by (8), the next step is to project the data matrix \mathbf{Z} onto the subspace orthogonal to \mathbf{w} . This allows us to look for additional projections that are mutually orthogonal by simply repeating our search for the best projection vector described by (7) and (8). In terms of orthogonal projection operators, the procedure for the j th iteration can be carried out by

$$\mathbf{Z}^{(j+1)} = \mathbf{P}_{\mathbf{w}}^{\perp} \mathbf{Z}^{(j)} \quad (9)$$

where

$$\mathbf{P}_{\mathbf{w}}^{\perp} = \mathbf{I} - \mathbf{W}^{(j)} \mathbf{W}^{(j)\#} \quad (10)$$

is the subspace orthogonal to the space spanned by the columns of \mathbf{W} . \mathbf{I} is the identity matrix, and $\mathbf{W}^{(j)\#} = (\mathbf{W}^{(j)T} \mathbf{W}^{(j)})^{-1} \mathbf{W}^{(j)}$ is the pseudo-inverse of $\mathbf{W}^{(j)}$ at the j th iteration. It should be noted that $\mathbf{W}^{(j)}$ contains all the best projection vectors obtained up to iteration j

$$\mathbf{W}^{(j)} = [\mathbf{W}^{(j-1)} \quad \mathbf{w}]. \quad (11)$$

As a result of this step, the rank of the matrix \mathbf{Z} is reduced. This two-step process can be repeated until the rank of the data matrix becomes zero. The end result is an orthogonal matrix \mathbf{W} , where each column is a projection vector.

The main advantage of this approach is its simplicity. Initialization is not required, so that our results do not depend on an initial guess, and there are no gradients to estimate. Although comparison of our strategy with conventional methods is beyond the scope of this paper, our method may be improved by using

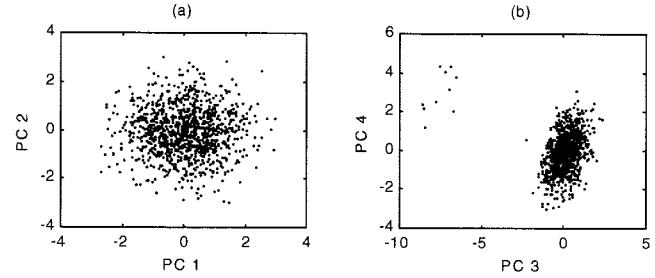


Fig. 3. PCA scatter plot of simulated data. (a) First and second PCA components and (b) Third and fourth PCA components.

a conventional technique to optimize the projection vector with the selected vector from the data as a starting point. This would allow us to find good solutions in areas of the data cloud that are sparsely populated. One can imagine such an approach as producing a sort of “interpolation” between the two best guesses to yield a locally optimal solution.

Another advantage of our approach is a tendency for the projections to align themselves with specific spectral classes. In fact, the projections can be viewed as a sequence of spectral angle maps [14] based on target spectra chosen by maximizing the projection index. This aids in the physical interpretability of the results, since the projection vectors are derived from physical spectra. However, such spectrum is seldom produced by a pure material (endmember), but rather is a mixture of multiple pure endmember spectra. Optimization could improve physical interpretability by finding a projection vector that is closer to a pure (single endmember) spectrum.

IV. COMPUTER SIMULATION

To illustrate our method we have constructed a simulated data set with a non-Gaussian feature which has a small relative variance so that PCA does not work optimally. Ten zero-mean features with 1000 samples each were simulated using a Gaussian random number generator and assigned geometrically decreasing variance values (1, 1/2, 1/4, ...). Ten of the samples (or 1%) were offset by 10σ in the third and fourth largest variance features, creating a non-normal distribution. The ten features were then converted to sample vectors by an arbitrary orthonormal rotation matrix. The resulting data set is one where the information embodied in the altered samples is not the dominant source of variance, and is spread between two features even though it is one-dimensional (1-D) in nature.

The results of PCA are shown in Fig. 3. The scatter plot of the first two components [Fig. 3(a)] shows no non-normal features. As would be expected, the scatter plot of the third and fourth components shows the information [Fig. 3(b)], which is preserved after an orthogonal transformation. To apply PP, we start by scaling all the PCA scores (see Section III-A) and then apply the method described in Section III-B. As Fig. 4 shows, the non-normal feature is captured in the first projection only. Thus, we have revealed the informative low variance feature as well as compressed the information. The projection index values for the PCA and PP projections are given in Table I. It shows that the third PCA component has the largest index, corresponding to the non-normal feature. The fact that the fourth component

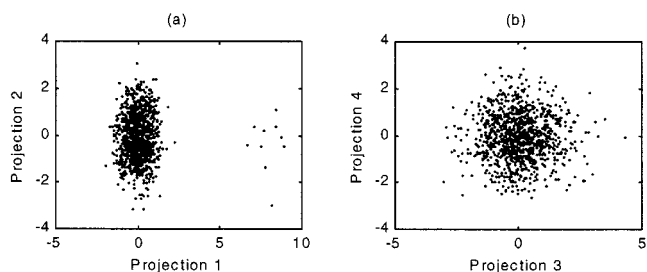


Fig. 4. PP Scatter plot of simulated data. (a) First and second PP projections and (b) Third and fourth PP projections.

TABLE I
 $\max J(\mathbf{p}, \mathbf{q})$ VALUES FOR THE PCA AND PP PROJECTIONS FOR SIMULATED DATA

Component /Projection	$\max J(\mathbf{p}, \mathbf{q})$ PCA	$\max J(\mathbf{p}, \mathbf{q})$ PP
1	0.002	0.505
2	0.006	0.030
3	0.376	0.032
4	0.021	0.027
5	0.012	0.029
6	0.021	0.033
7	0.017	0.029
8	0.015	0.034
9	0.007	0.020
10	0.011	0.005

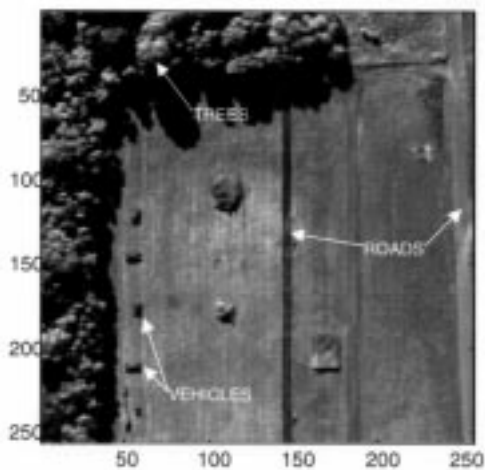


Fig. 5. HYDICE scene.

has a low index value indicates that PCA did remove some of the redundancy that resulted from introducing the variation in two components. PP not only removed the redundancy completely, but also captured the information in the first projection. Also, note that the projection indexes for the PP are not necessarily monotonically decreasing. Since the projections are mutually orthogonal, their ordering is merely a matter of convention (as with PCA). In this paper we have not altered the order produced by the algorithm.

V. HYPERSPECTRAL IMAGING EXPERIMENTS

A. HYDICE Experiments

The hyperspectral digital imagery collection experiment (HYDICE) sensor [15], [16] is a pushbroom imaging spec-

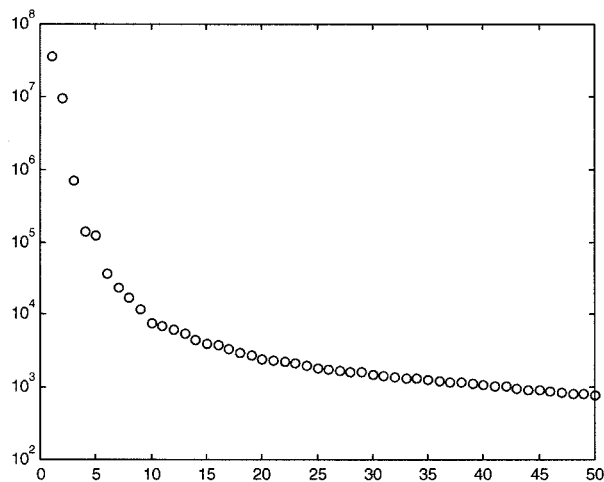


Fig. 6. Eigenvalue spectrum from PCA analysis of HYDICE scene.

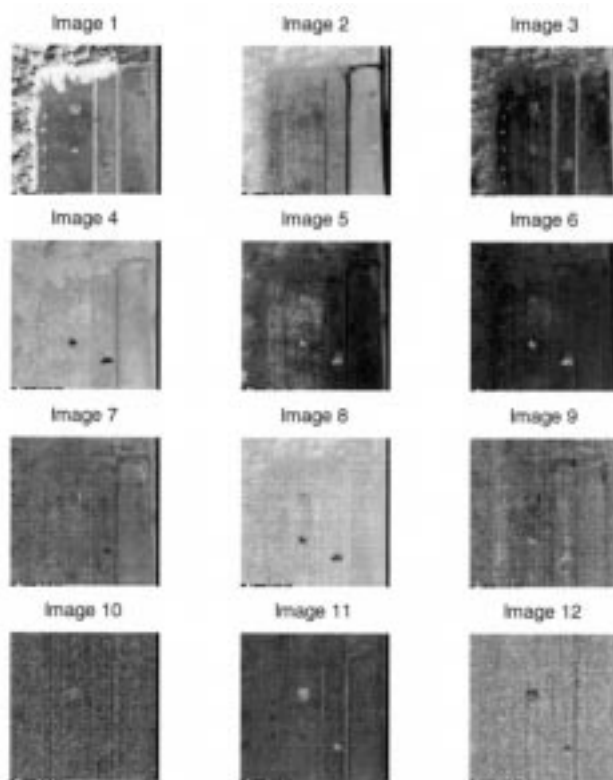


Fig. 7. Component images from PCA analysis of HYDICE scene.

trometer operating in the visible to short-wave infrared spectral region. The data was delivered with the radiometric calibration already done, so it is in units of $W/m^2/Sr/\mu m$. Spectral calibration, nonuniformity correction, and bad pixel compensation were also done prior to delivery.

To illustrate the usefulness of our PP approach, we use a 256×256 image (Fig. 5), which includes the smaller scene used in a previous paper [5]. It contains objects of several types, including vehicles, trees, roads, and other features. Because PP analysis is performed using the image data only, conversion from radiance to apparent reflectance is not necessary. It is assumed that the atmospheric variability is negligible over the scene. Depending

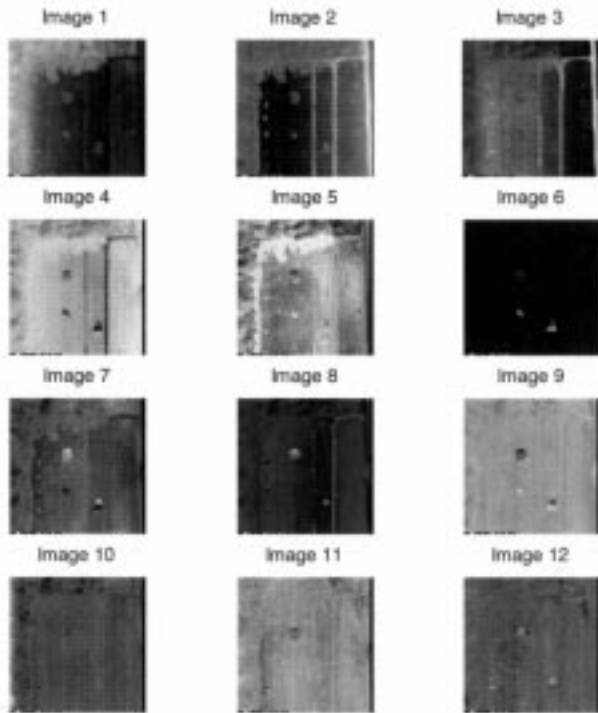


Fig. 8. Component images from MNF analysis of HYDICE scene.

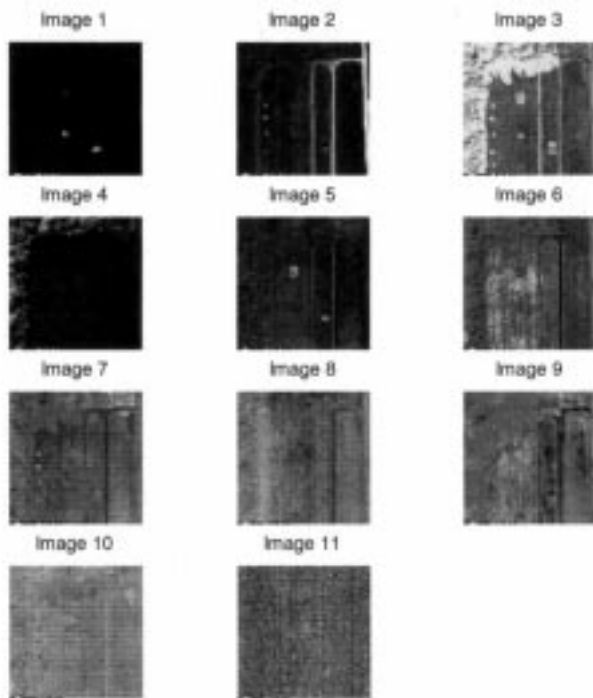


Fig. 9. Projection images from PP analysis of HYDICE scene.

on the application, however, it may be desirable to normalize the pixel spectra prior to PCA. In this example, we use the calibrated spectra as it is. The only alteration to the data is the removal of bands 1–4, which contain image artifacts that can affect the results. The water bands were kept unaltered.

Fig. 6 shows eigenvalue spectrum for the PCA of this scene (first 50 eigenvalues). It is used to help estimate the number of components that contain any significant amount of variance.

TABLE II
max $J(p, q)$ VALUES FOR THE PCA, MNF, AND PP PROJECTIONS FOR HYDICE SCENE

Image Number	max $J(p,q)$ PCA	max $J(p,q)$ MNF	max $J(p,q)$ PP
1	0.415	0.712	1.329
2	0.272	0.350	1.952
3	0.341	0.342	1.019
4	0.132	0.440	1.131
5	0.105	0.224	0.078
6	0.105	1.265	0.182
7	0.042	0.068	0.062
8	0.109	0.193	0.009
9	0.015	0.342	0.058
10	0.009	0.307	0.024
11	0.067	0.110	0.008

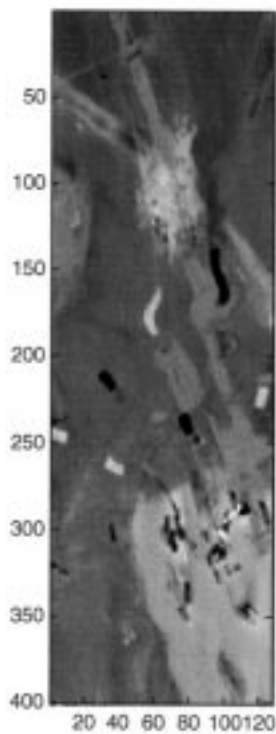


Fig. 10. LWIR SEBASS scene.

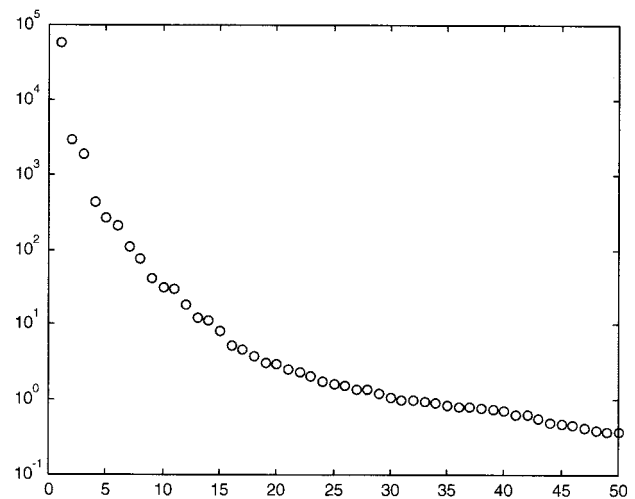


Fig. 11. Eigenvalue spectrum from PCA analysis of SEBASS scene.

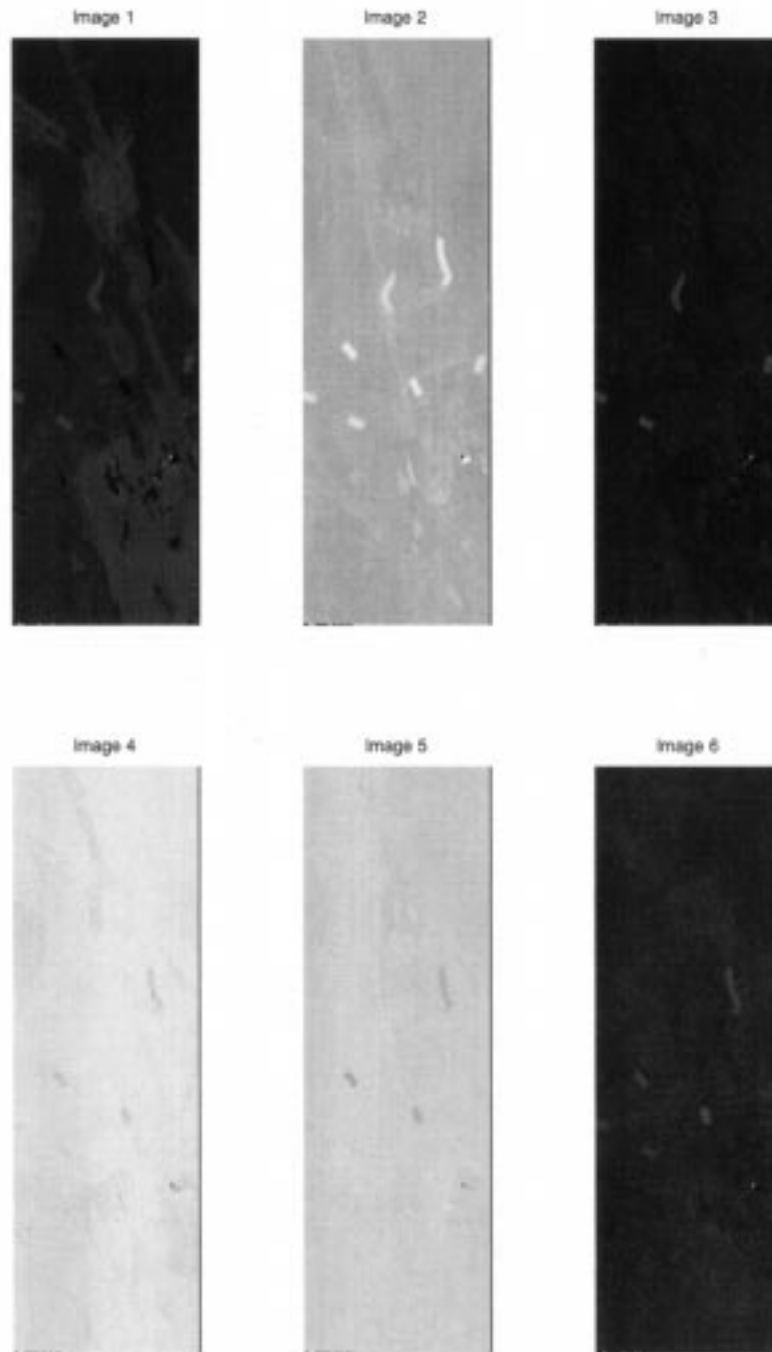


Fig. 12. Component images from PCA analysis of SEBASS scene.

The rest are presumed to contain only noise and can thus be discarded in order to reduce computation. Our experience has shown that finding the exact number of eigenvalues is not critical to successful application of PP. It is preferable to overestimate the number of significant components in order to capture low variance information that may otherwise be lost. In this example we use the largest eleven components, accounting for 99.81% of the signal variance. The component images from the PCA are shown in Fig. 7. A visual analysis reveals that the signal variance is dominated by the shadow features together with the vegetation and roads, followed by the vehicles and the other objects in the scene. The latter have their signatures spread among several

components. Although most of the useful information is captured in the first four components, there is a significant amount of information down to the eleventh component. Additionally, the MNF transform was computed on the same image (Fig. 8) using the ENVI software package [17], with the noise covariance matrix estimated from the image using the “shift difference” method. No improvement over PCA is evident in terms of compression for this scene. The MNF transform in this case contains visually useful information up to the twelfth component.

We now apply our PP algorithm to generate eleven projections shown in Fig. 9, based on (7)–(10). Most of the information has now been compressed into the first seven projections. Fur-

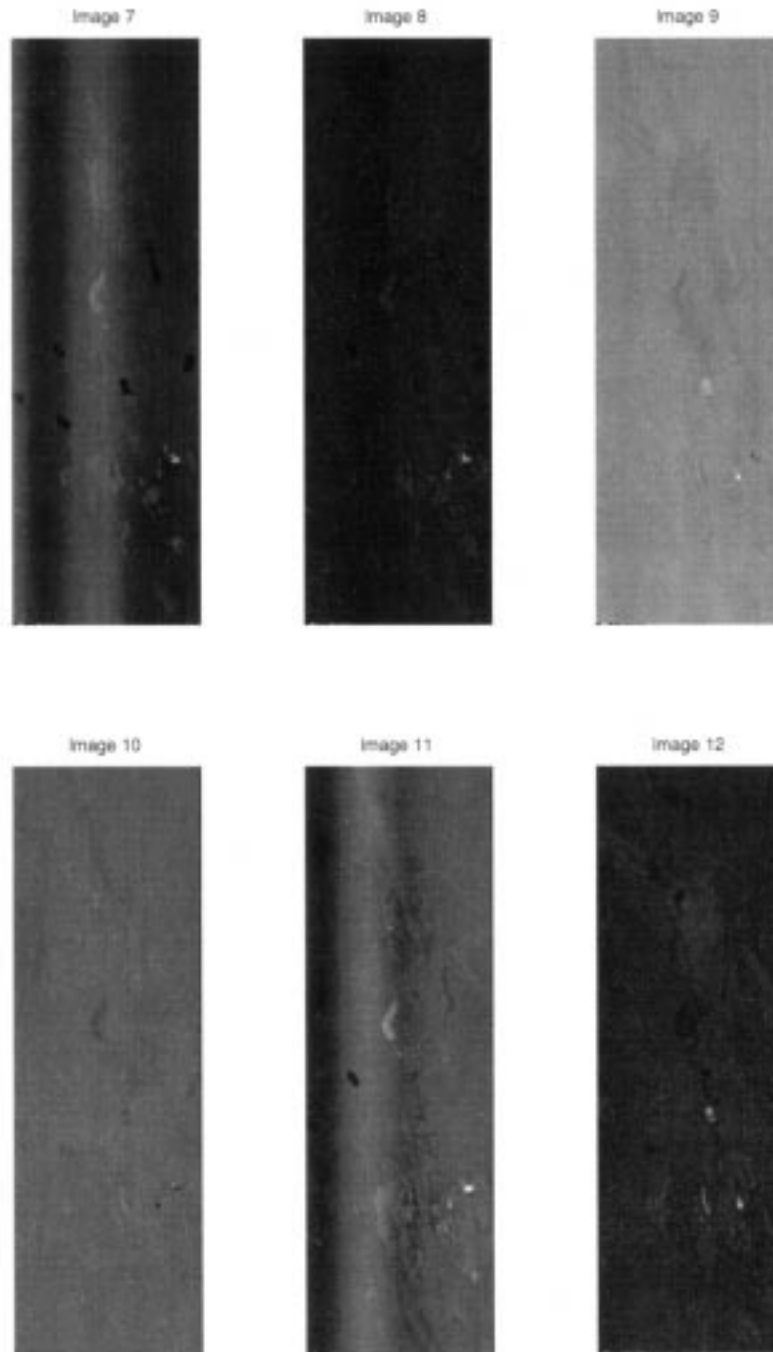


Fig. 12. (Continued.) Component images from PCA analysis of SEBASS scene.

thermore, the components tend to correspond to different types of objects. The first captures the information related to two of the objects while the second projection contains information on the roads and some of the vehicles. The third and fourth projections correspond mainly to the shadows and trees respectively. The next three projections contain information on the remaining objects and vehicles, as well as some variability of the background, which consists of both grass and dirt. Note that the remaining images (8–11) are mostly noise.

Table II lists the projection index values for the PCA, MNF, and PP projections. For PP, the weaker features in images 5–7 are reflected by the relatively low value of the projection index.

The PCA components have generally lower index values compared to PP. The MNF components show higher values of the projection index than PCA, but lower than PP. Another interesting pattern to note is that in the case of PP, the index values start out higher and then drop significantly after the fourth component, indicating a higher degree of information compression by PP than by PCA or MNF, which show a more steady decline in the index values (with the exception of MNF component 6). Because each transform computed a different set of projections, the reader should be aware of that when examining the values not to compare the numbers line-by-line, but rather consider them as a whole.

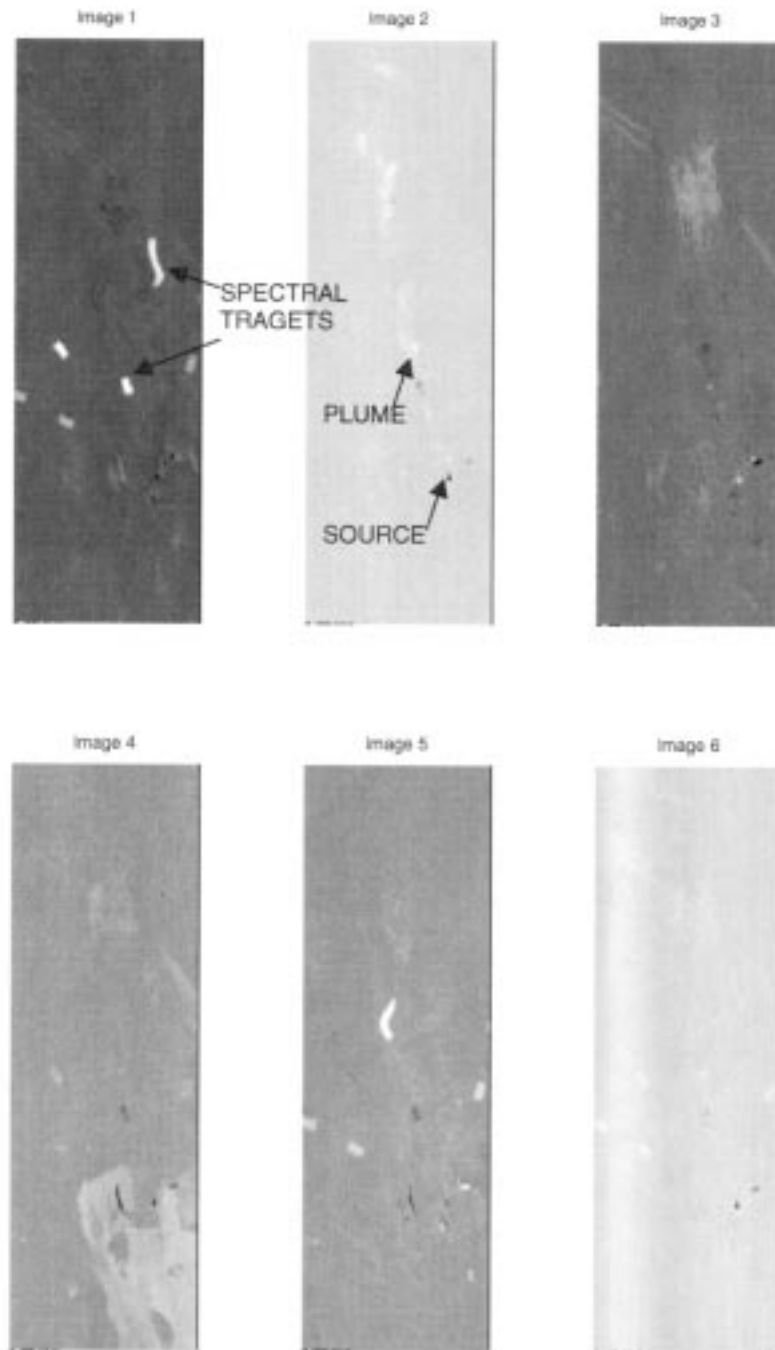


Fig. 13. Projection images from PP analysis of SEBASS scene.

B. SEBASS LWIR Experiments

The spatially enhanced broadband array spectrograph system (SEBASS) [18] is a dual-band mid-wave/long-wave infrared pushbroom imaging spectrometer. As with the HYDICE data, the radiometric calibration and bad pixel correction were done prior to delivery. We use the LWIR SEBASS data here to demonstrate the ability of PP to uncover important image features that have a low variance relative to the overall image variability. Our example image is a 127×400 pixel scene that contains a source of Sulfur Hexafluoride (SF_6) gas (Fig. 10). The infrared absorption spectrum of SF_6 has a strong spectral feature near the 10.5

μm band, which can be used to detect the presence of the gas in the atmosphere. The gas source, although quite hot, is very small relative to the image, and so contributes little variance. PCA (Figs. 11 and 12) provides little visual information on the gas plume. An analysis of the eigenvectors shows that the gas-related information is contained in the eighth and ninth components, but the noise level is too high to visually discern a plume in the projection images.

Fifteen components were used for PP. Fig. 13 shows the results of the PP analysis for projections. The first projection locates some spectral targets which had been placed on the field. The second projection clearly aligns itself with the SF_6 gas in-

formation, revealing the plume. We can see the hot source toward the bottom of the image (in this case hotter pixels are darker). At this point the gas is emitting more energy than it is absorbing from the background behind it. As it moves away from the source, it cools down below the background temperature and begins to absorb more energy than it emits, so the image contrast is reversed. Note that as the plume drifts away from the source, it passes over the colder spectral target, briefly producing a net emission. Projections 3 through 5 locate other objects and natural features in the scene. The remaining projections carry little information, illustrating again the information compression capabilities of this method.

VI. CONCLUSIONS

Although PP has been part of the data analyst's toolbox for over two decades, it has been slow to gain acceptance in certain areas, including remote sensing. With the increased use of imaging spectrometry, however, it becomes necessary to exploit techniques that are capable of quickly reducing the massive volume of data while simultaneously preserving as much information as possible. It has been demonstrated in this paper that a significant reduction can be achieved by using PP with an information divergence projection index. In addition, the component images resulting from our analysis can be further processed using image segmentation or unmixing methods so as to produce a single classified image containing all the information that by a given projection index.

ACKNOWLEDGMENT

The authors would like to thank the HYDICE Program Office, Washington, DC, for providing HYDICE data and the Aerospace Corporation, Los Angeles, CA, for providing the SEBASS data.

REFERENCES

- [1] C.-I Chang, T.-L. E. Sun, and M. L. G. Althouse, "An unsupervised interference rejection approach to target detection and classification for hyperspectral imagery," *Opt. Eng.*, vol. 37, no. 3, pp. 735–743, 1998.
- [2] C.-I Chang and Q. Du, "Interference and noise-adjusted principal components apal components analysis," *IEEE Trans. Geosci Remote Sensing*, vol. 37, p. 2387, Sept. 1999.
- [3] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "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, 1988.
- [4] 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.
- [5] A. Ifarraguerri and C.-I Chang, "Projection pursuit analysis of hyperspectral scenes," *Proc. SPIE*, vol. 3372, pp. 51–59, 1998.
- [6] J. H. Friedman and J. W. Tukey, "A projection pursuit algorithm for exploratory data analysis," *IEEE Trans. Computers*, vol. 23, pp. 881–889, Sept. 1974.
- [7] M. C. Jones and R. Sibson, "What is projection pursuit?," *J. R. Statist. Soc. A*, vol. 150, no. 1, pp. 1–36, 1987.
- [8] J. H. Friedman, "Exploratory projection pursuit," *J. Amer. Statist. Assoc.*, vol. 82, no. 39, pp. 249–266, 1987.

- [9] P. J. Huber, "Projection pursuit," *Ann. Statist.*, vol. 13, no. 2, pp. 435–475, 1985.
- [10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [11] L. Jimenez and D. A. Landgrebe, "Projection pursuit in high dimensional data reduction: Initial conditions, feature selection and the assumption of normality," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 2653–2667, Nov. 1999.
- [12] S.-S. Chiang and C.-I Chang, "Target subpixel detection for hyperspectral imagery using projection pursuit," in *Proc. Conf. Image and Signal Processing for Remote Sensing V*, Florence, Italy, Sept. 1999, pp. 107–115.
- [13] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.
- [14] R. H. Yuhas, A. F. H. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the Spectral Angle Mapper (SAM) algorithm," in *Summaries 3rd Annu. JPL Airborne Geoscience Workshop*, vol. 1, Pasadena, CA, 1992, pp. 147–149.
- [15] W. S. Aldrich, M. E. Kappus, and R. G. Resmini, "HYDICE post-flight data processing," *Proc. SPIE*, vol. 2758, pp. 354–363, 1996.
- [16] R. W. Basedow, W. S. Aldrich, J. E. Colwell, and W. D. Kinder, "HYDICE system performance update," *Proc. SPIE*, vol. 2821, pp. 76–84, 1996.
- [17] *ENVI User's Guide*, Better Solutions Consulting LLC, Boulder, CO, 1998, p. 320.
- [18] J. A. Hackwell *et al.*, "LWIR/MWIR imaging hyperspectral sensor for airborne and ground-based remote sensing," *Proc. SPIE*, vol. 2819, pp. 102–107, 1996.



Agustin Ifarraguerri (S'95–M'00) received the M.S. degree in bioengineering from the University of Michigan, Ann Arbor, and the Ph.D. in electrical engineering from the University of Maryland, Baltimore County (UMBC), in 2000.

He has been with the U.S. Army Edgewood Chemical Biological Center since 1992, where he has contributed to the development of several systems for the detection of chemical and biological agents and is currently participating in the development of an imaging spectrometer for standoff detection of bat-

tfield chemical agents.



Chein-I Chang (S'81–M'87–SM'92) received the B.S., M.S., and M.A. degrees from Soochow University, Taipei, Taiwan, the Institute of Mathematics at National Tsing Hua University, Hsinchu, Taiwan, and the State University of New York, Stony Brook, respectively, all in mathematics, in 1973, 1975, and 1977. He also received the M.S. and M.S.E.E. degrees from the University of Illinois, Urbana, both in 1982, and the Ph.D. in electrical engineering from the University of Maryland, College Park, in 1987.

Since 1987, he has been with the Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County (UMBC), Baltimore, MD, first as a Visiting Assistant Professor from January 1987 to August 1987, then as an Assistant Professor from 1987 to 1993, and currently as an Associate Professor. He was a Visiting Specialist with the Institute of Information Engineering, National Cheng Kung University, Tainan, Taiwan, from 1994 to 1995. He is on the editorial board of *JOURNAL OF HIGH SPEED NETWORKS* and is the Guest Editor of a special issue on *TELEMEDICINE AND APPLICATIONS*. His research interests include automatic target recognition, multispectral/hyperspectral image processing, medical imaging, information theory and coding, signal detection and estimation, and neural networks.

Dr. Chang is a member of SPIE, INNS, Phi Kappa Phi, and Eta Kappa Nu.