A Joint Band Prioritization and Band-Decorrelation Approach to Band Selection for Hyperspectral Image Classification

Chein-I Chang, *Senior Member, IEEE*, Qian Du, *Student Member, IEEE*, Tzu-Lung Sun, and Mark L. G. Althouse, *Member, IEEE*

Abstract—Band selection for remotely sensed image data is an effective means to mitigate the curse of dimensionality. Many criteria have been suggested in the past for optimal band selection. In this paper, a joint band-prioritization and band-decorrelation approach to band selection is considered for hyperspectral image classification. The proposed band prioritization is a method based on the eigen (spectral) decomposition of a matrix from which a loading-factors matrix can be constructed for band prioritization via the corresponding eigenvalues and eigenvectors. Two approaches are presented, principal components analysis (PCA)-based criteria and classification-based criteria. The former includes the maximum-variance PCA and maximum SNR PCA, whereas the latter derives the minimum misclassification canonical analysis (MMCA) (i.e., Fisher's discriminant analysis) and subspace projection-based criteria. Since the band prioritization does not take spectral correlation into account, an information-theoretic criterion called divergence is used for band decorrelation. Finally, the band selection can then be done by an eigenanalysis-based band prioritization in conjunction with a divergence-based band decorrelation. It is shown that the proposed band-selection method effectively eliminates a great number of insignificant bands. Surprisingly, the experiments show that with a proper band selection, less than 0.1 of the total number of bands can achieve comparable performance using the number of full bands. This further demonstrates that the band selection can significantly reduce data volume so as to achieve data compression.

Index Terms— Band decorrelation, band prioritization, band selection, divergence, eigenanalysis, hyperspectral classification, orthogonal-subspace projection (OSP), principal-components analysis (PCA).

I. INTRODUCTION

HYPERSPECTRAL sensors can image an area with hundreds of different wavelength ranges for identification of the composition of various materials. As a result, each image scene is represented by an image cube with the third dimension specified by spectral range. Such three-dimensional (3-D) representations create enormous amounts of data for computer processing and data transmission. In order to mitigate this problem, data compression is generally used to reduce data volume. Taking advantage of spectral correlation to achieve data compression is one of the unique features in multispectral/hyperspectral images. A general approach of this kind is band selection to achieve dimensionality reduction. In the past, many criteria have been proposed for band selection [1]–[5] to find bands that are crucial and significant in terms of information conservation. For instance, distance measures (Bhattacharyya distance, Jeffreys–Matusita distance), information-theoretic approaches (divergence, transformed divergence, mutual information) and eigenanalysis [principal components analysis (PCA)] have been applied to multispectral images for optimal band selection. In particular, the use of the divergence measure for band selection has received considerable interest in multispectral imagery. More recently, the divergence was used as a band-selection criterion for hyperspectral-pixel classification [4]. However, it requires computing divergences for all the possible combinations of band subsets. When the divergence measure is applied to hyperspectral imagery, which is generally acquired with more than 200 bands, such a direct calculation becomes formidable. In order to alleviate this problem, an alternative was reported in [4].

In this paper, we present a new band-selection method that comprises a band prioritization and a band decorrelation. The band prioritization prioritizes all bands according to the contained information to be used for classification. Bands are then selected on the basis of their associated priorities. Since the band prioritization does not consider the spectral correlation [6], a band decorrelation using the divergence is used to decorrelate prioritized bands. The proposed band prioritization is an eigenanalysis-based method that was used for fast classification [7]. Although the concept of eigenanalysis has been suggested in [1] and [2], its usefulness was not fully exploited. In [7], Tu *et al.* constructed a loading-factors matrix via the eigen (spectral) decomposition of an appropriate matrix, in which the loading factors were used to rank the priority of each band. These bands were selected in decreasing order of priority to achieve a certain level of classification accuracy. In this paper, we revisit this band-prioritization approach and propose two eigenanalysis-based criteria for band prioritization, the PCA-based criteria, and the classification-based criteria. Two PCA-based criteria are to be considered: maximum-variance PCA (MVPCA) and maximum-SNR PCA (MSNRPCA). Two classification-based criteria are also de-

rived: minimum-misclassification canonical analysis (MMCA) (Fisher's discriminant analysis) and orthogonal-subspace projection (OSP) criteria. The classification power of a band subset is the sum of the loading factors produced by the bands in the selected band subset. A band-power ratio for a band selection is then defined as the classification power of the selected band subset divided by that of all bands being used. After all bands are prioritized, a divergence-based band decorrelation is further used to remove either redundant or insignificant bands. If the divergence between two bands is below a prescribed threshold, the band with lower priority will be removed. Finally, the desired band-selection method can be designed by coupling an eigenanalysis-based band prioritization with a divergence-based band decorrelation.

A comparative study for the four proposed band-prioritization criteria will be conducted using hyperspectral digital imagery collection experiment (HYDICE) image data and will be evaluated by the effectiveness of mixed-pixel classification. The conducted experiments show that the OSP-based band prioritization, coupled with the divergence-based band decorrelation, produced the best results and can effectively reduce large dimensionality at the expense of a slight loss of information for classification. More surprisingly, nearly 94% of the total bands are shown to be unnecessary in terms of classification and thus can be eliminated with negligible loss of classification accuracy. This shows that a joint band prioritization and band decorrelation can achieve significant band reduction while preserving most of the information for classification.

The remainder of this paper is organized as follows. Section II presents two PCA-based eigenanalysis band-prioritization criteria, an MVPCA, and an MSNR PCA, while Section III proposes two classification-based band-prioritization criteria, MMCA, and orthogonal subspace projection (OSP). Since band prioritization does not necessarily decorrelate spectral bands, a divergence-based band-decorrelation approach is also proposed in Section IV. In order to evaluate the performance of the proposed joint band prioritization and band decorrelation, a series of HYDICE experiments are conducted in Section V. Finally, a brief conclusion is given in Section VI.

II. PCA-BASED BAND PRIORITIZATION

Eigenanalysis has been used for band selection in the past [1], [2]. However, it does not take full advantage of the relationship between eigenvalues and eigenvectors. Tu *et al.* in [7] introduced the concept of band prioritization into canonical analysis (or Fisher's discriminant analysis) to prioritize bands so as to achieve band reduction in which the bands were prioritized by their associated loading factors. In order to obtain these loading factors, a loading-factors matrix was constructed from the eigenvalues and eigenvectors of the spectral- or eigendecomposition of an appropriate data matrix. In this section, two PCA-based transforms will be investigated to form different data matrices from which their corresponding loading matrices can be constructed to derive two criteria used for band prioritization.

A. Principal Components Analysis (PCA)

Principal components analysis (PCA), also known as the Karhunen–Loeve Transform, is a decorrelation technique that is widely used for data compression and interpretation [2]. It transforms data coordinates in such a fashion that the first principal component vector is along the direction of maximum variance. It then maximizes the variance in successive components. Therefore, in this paper, it will be considered as an MVPCA transformation.

Let $\Sigma = 1/N \sum_{k=1}^{N} (\mathbf{x}_k - \mathbf{m})(\mathbf{x}_k - \mathbf{m})^T$ be a data-sample covariance matrix where $\mathbf{x}_k$ is the kth l-dimensional pixel vector in a hyperspectral image, $\mathbf{m}$ is the sample-mean vector, N is the total number of pixel vectors, and l is the data dimensionality (i.e., total number of bands). Since Σ is symmetric and nonnegative definite, all its eigenvalues $\{\lambda_i\}_{i=1}^{l}$ are real and nonnegative, and its corresponding l-dimensional eigenvectors $\mathbf{v}_i = (v_{i1}, \cdots, v_{il})^T$ for $i = 1, 2, \cdots, l$ can be chosen to be orthonormal, i.e., $\|\mathbf{v}_i\| = (\sum_{j=1}^{l} v_{ij}^2)^{1/2} = 1$. We can define loading factors r_{ik} associated with $\mathbf{v}_i$ for MVPCA transformation as

$$r_{ik} = \sqrt{\lambda_i} v_{ik} \quad \text{for} \quad k = 1, 2, \cdots, l. \tag{1}$$

It is easy to show that for each $k = 1, 2, \cdots, l$ ρ_k defined by

$$\rho_k = \sum_{i=1}^{l} r_{ik}^2 \tag{2}$$

is indeed the variance σ_{kk}^2 of the kth band image. Summing up r_{ik}^2 over k in (2) for each $i = 1, 2, \cdots, l$ also yields

$$\lambda_i = \sum_{k=1}^{l} r_{ik}^2. \tag{3}$$

So from (2) and (3), $\sum_{k=1}^{l} \rho_k = \sum_{i=1}^{l} \lambda_i$.

Since (2) also represents variances of band images, we can define the sum $\sum_{k=1}^{m} \rho_k$ to be the power produced by m bands with the first m largest variances. Without loss of generality, we assume that $\rho_1 \geq \rho_2 \geq \cdots \geq \rho_l$. In this case, we can further define the variance-based band-power ratio of these m bands to be that of the total number of bands, denoted by $R_{\text{var}}(m)$, as

$$R_{\text{var}}(m) = \frac{\sum_{k=1}^{m} \rho_k}{E} = \frac{\sum_{k=1}^{m} \sigma_{kk}^2}{E} \tag{4}$$

where $E = \sum_{k=1}^{l} \rho_k = \sum_{k=1}^{l} \sigma_{kk}^2$ is the sum of variances of all band images. So $(\rho_1/E) \leq R_{\text{var}}(m) \leq 1$ for m ranging from one to l, where the upper and lower bounds are achieved by using all bands [i.e., $R_{\text{var}}(l) = 1$] and only one band [i.e., $R_{\text{var}}(1)$], respectively. Therefore, using (4), we can determine how many bands are required to achieve a desired variance band-power ratio anywhere between one and ρ_1/E. It is worth noting that the results outlined in (2)–(4) can be found in [8] as well, where the selection of m highest-variance bands presented in this paper is equivalent to finding m principal variables using the criterion $\sum_{k=1}^{m} \sigma_{kk}^2$ defined in (4) with $\sigma_{11}^2 \geq \sigma_{22}^2 \geq \cdots \geq \sigma_{mm}^2$. As will be shown in the

experiments, the measure $R_{\text{var}}(m)$ given by (4) may not be a good measure, since band images are generally correlated, and the bands selected based on (2) may still correlate to those removed bands, a fact also noted in [8]. To mitigate this problem, an SNR-based PCA will be introduced in the following section.

B. Noise Adjusted Principal Component Analysis (NAPC)

The goal of PCA is to find the principal components with their directions along the maximum variances of a data matrix. However, as shown in [9], variance does not necessarily reflect real SNR due to unequal noise variances incurred in different bands. Because of that, Green *et al.* developed a maximum noise fraction (MNF) transform in [9], so that the transformed principal components are evaluated by SNR rather than variance, as used in a PCA. In light of this interpretation, the MNF transform can be thought of as a transformation maximizing SNR. The MNF was further reinterpreted in [10] as an NAPC. The idea of the NAPC is to design a matrix to whiten the covariance matrix so that the noise-adjusted covariance matrix becomes an identity matrix. Therefore, the resulting variances can be interpreted as SNR's. By taking advantage of the NAPC, the possible correlation with removed bands created by the measure $R_{\text{var}}(m)$ defined by (4) for band selection, can be resolved. In this section, this approach is referred to as an MSNRPCA transformation, which will be based on the fast NAPC transform [11] with the noise variances estimated by the nearest-neighbor difference described in [9] and [10].

In the NAPC approach, a noise estimate is required for the observation model

$$\mathbf{z} = \mathbf{s} + \mathbf{n} \tag{5}$$

where $\mathbf{z}$ is an observation vector with the covariance matrix denoted by Σ, $\mathbf{s}$ is a signal vector, and $\mathbf{n}$ is the noise vector independent of $\mathbf{s}$ with the covariance matrix denoted by $\Sigma_{\mathbf{n}}$. The noise estimate will then be used to find a whitening matrix F to orthonormalize $\Sigma_{\mathbf{n}}$ such that

$$F^T \Sigma_{\mathbf{n}} F = I \quad \text{and} \quad F^T F = \Delta_{\mathbf{n}}^{-1}. \tag{6}$$

$\Delta_{\mathbf{n}}$ in (6) is the diagonal matrix of the eigenvalues of $\Sigma_{\mathbf{n}}$. The resulting noise-adjusted data covariance matrix is given by $\Sigma_{adj} = F^T \Sigma_{\mathbf{z}} F$. Let G be the eigenvector matrix resulting from PCA based on Σ_{adj}. Then we obtain

$$G^T \Sigma_{adj} G = \Lambda_{adj} \quad \text{and} \quad G^T G = I \tag{7}$$

where $\Lambda_{adj} = \text{diag}\{\lambda_{adj,i}\}$ is the diagonal matrix of the eigenvalues $\{\lambda_{adj,i}\}_{i=1}^{l}$ of Σ_{adj}. Finally, the desired NAPC transform can be derived from

$$H = GF. \tag{8}$$

Let $\{\mathbf{v}_i\}_i^l$ be the orthonormal eigenvectors associated with $\{\lambda_{adj,i}\}_{i=1}^{l}$. We can define the loading factors in a similar manner to (1) for an MSNRPCA transformation by

$$r_{ik} = \sqrt{\lambda_{adj,i}} v_{ik} \quad \text{for} \quad k = 1, 2, \cdots, l. \tag{9}$$

For any given m, we can use (9) and (1)–(4) to define the SNR-based band-power ratio produced by the first m bands to the total number of bands, denoted by $R_{\text{SNR}}(m)$

$$R_{\text{SNR}}(m) = \frac{\sum_{k=1}^{m} \rho_k}{E} \tag{10}$$

where $E = \sum_{k=1}^{l} \rho_k$ is the total variance of Σ_{adj}. Since the noise variances in $\Sigma_{\mathbf{n}}$ have been normalized to one by the whitening matrix F in (6), for each $k = 1, 2, \cdots, l$, the variance of the kth band image ρ_k defined by $\rho_k = \sum_{i=1}^{l} r_{ik}^2$ via (9) becomes the signal energy. Consequently, the ρ_k also represents the SNR of the k-band image. Similarly, E is the total energy of signals in all l-band images. As a result of the NAPC transform specified by (8), (10) is indeed the SNR band-power ratio for selection m bands in accordance with $\rho_1 \rho_2 \cdots \rho_l$. The advantage of using MSNRPCA over MVPCA is clearly demonstrated through operations implemented by (6) and (7), where the noise covariance matrix $\Sigma_{\mathbf{n}}$ has been whitening before the PCA transform is applied. So, the correlation problem with between-band noise variances that results in a poor measure $R_{\text{var}}(m)$ of (4) is resolved by (6). Analogous to (4), (10) is also upper-bounded by one (i.e., all bands are used) and lower-bounded by ρ_1/E (only the first band is used). It can be used to determine how many bands are required to achieve a certain desired level of SNR between $(\rho_1/E) \leq R_{\text{SNR}}(m) \leq 1$.

III. CLASSIFICATION-BASED EIGENANALYSIS BAND PRIORITIZATION

Unlike the MVPCA and MSNRPCA, which are based on maximization of energy, the classification-based eigenanalysis-band prioritization is designed by ranking bands according to the effectiveness of their classification abilities. Two criteria will be derived from this approach, the MMCA and the OSP. In this case, we need to know how many classes will be considered. For illustrative purposes, the class membership of samples is assumed to be known *a priori*. Nonetheless, this constraint is not rigorous and can be relaxed by unsupervised learning methods such as nearest-neighbor clustering [12], Kohonen self-organization feature maps [13], etc.

A. Minimum Misclassification Canonical Analysis (MMCA) or Fisher's Discriminant Analysis

MMCA is derived from Fisher's discriminant function and was used in [7] to minimize the misclassification error.

Let $\{\omega_1, \omega_2, \cdots, \omega_C\}$ be the set of classes of interest and C be the number of classes to be classified. Assume that $\mathbf{x}_{ij}$ is the jth sample vector in class i and $\Xi = \{\mathbf{x}_{ij}\}_{i=1, j=1}^{C, N_i}$ is the set of sample vectors to be used for classification, where N_i is the number of sample vectors in the ith class and $N = N_1 + \cdots + N_C$ is the total number of sample vectors. Let $\mathbf{m} = (1/N) \sum_{i=1}^{C} \sum_{j=1}^{N_i} \mathbf{x}_{ij}$ be the mean of Ξ, and $\mathbf{m}_i = (1/N_i) \sum_{j=1}^{N_i} \mathbf{x}_{ij}$ be the mean of class i. From Fisher's discriminant analysis [12], we can form total, between-class,

and within-class scatter matrices as follows:

$$S_{\mathrm{T}} = \frac{1}{N} \sum_{i=1}^{C} \sum_{j=1}^{N_i} (\mathbf{x}_{ij} - \mathbf{m})(\mathbf{x}_{ij} - \mathbf{m})^T \tag{11}$$

$$S_{\mathrm{W}} = \frac{1}{N} \sum_{i=1}^{C} \sum_{\mathbf{x}_{ij} \in \omega_i} \frac{1}{N} (\mathbf{x}_{ij} - \mathbf{m}_i)(\mathbf{x}_{ij} - \mathbf{m}_i)^T \tag{12}$$

$$S_{\mathrm{B}} = \sum_{i=1}^{C} \frac{N_i}{N} (\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^T. \tag{13}$$

From (9)–(11):

$$S_{\mathrm{T}} = S_{\mathrm{W}} + S_{\mathrm{B}}. \tag{14}$$

In order to minimize the misclassification error, we maximize the Rayleigh quotient

$$J(Z) = \frac{Z^T S_{\mathrm{B}} Z}{Z^T S_{\mathrm{W}} Z} \quad \text{over } Z. \tag{15}$$

The solution to (15) is equivalent to solving the following generalized eigenvalue problem

$$S_{\mathrm{B}} \mathbf{v}_i = \lambda_i S_{\mathrm{W}} \mathbf{v}_i \tag{16}$$

or equivalently

$$S_{\mathrm{W}}^{-1} S_{\mathrm{B}} \mathbf{v}_i = \lambda_i \mathbf{v}_i. \tag{17}$$

For any given band number m, we can use (1)–(3) with eigenvalues and normalized eigenvectors obtained by (17) to define $\sum_{k=1}^{m} \rho_k$ as the discriminant power of m bands to be used for discrimination and also define the total discriminant power to be $E = \sum_{k=1}^{l} \rho_k$. The band-power ratio, denoted by $R_{\mathrm{MMCA}}(m)$, is

$$R_{\mathrm{MMCA}}(m) = \frac{\sum_{k=1}^{m} \rho_k}{E}. \tag{18}$$

The band-power ratio given by (18) is a measure for the misclassification rate of m bands to be used for classification, compared to that of the total number of l bands.

B. Orthogonal Subspace Projection (OSP) Criterion

Like the MMCA, the OSP eigenanalysis-band prioritization is also a classification-based approach. The approach described in this section will be derived from the concept of OSP developed in [14]. For more details and other OSP-based approaches, we refer to [15] and [16]. Before we proceed, a review of linear-mixture models used in these criteria is necessary.

1) Linear Spectral Mixture Model: Linear spectral mixture is a widely used model in remotely sensed imagery to determine and quantify multicomponents. Suppose that l is the number of spectral bands. Let $\mathbf{r}_i$ be an $l \times 1$ column vector and denote the ith pixel in a multispectral or hyperspectral image. In this case, each pixel is viewed as a pixel vector with dimension l. Assume that M is an $l \times p$ signature matrix denoted by $[\mathbf{m}_1 \, \mathbf{m}_2 \cdots \mathbf{m}_p]$ where $\mathbf{m}_j$ is an $l \times 1$ column vector represented by the jth signature (substance) resident in

the pixel $\mathbf{r}_i$ and p is the number of signatures of interest. Let $\alpha_i = (\alpha_{i1} \, \alpha_{i2} \, \cdots \, \alpha_{ip})^T$ be a $p \times 1$ abundance-column vector associated with $\mathbf{r}_i$ where α_{ij} denotes the fraction of the jth signature in the pixel $\mathbf{r}_i$.

A classical approach to solving the mixed-pixel classification problem is linear unmixing, which assumes that the materials within a pixel vector are linearly mixed and can be represented by a linear-regression model

$$\mathbf{r}_i = M \alpha_i + \mathbf{n}_i \tag{19}$$

where $\mathbf{n}_i$ is an $l \times 1$ column vector representing an additive white Gaussian noise with zero mean and variance $\sigma^2 I$, and I is the $l \times l$ identity matrix.

2) Orthogonal Subspace Projection (OSP): Now we assume that there are C signatures that need to be classified with $C < p$. In this case, we rewrite model (19) to separate the undesired signatures U from the desired signature D as follows:

$$\mathbf{r} = D\mathbf{A} + U\mathbf{B} + \mathbf{n} \tag{20}$$

where the subscript i is suppressed, the desired signature matrix is denoted by $D = [\mathbf{m}_1, \cdots, \mathbf{m}_C]$ associated with the abundance vector $\mathbf{A} = (\alpha_1, \cdots, \alpha_C)^T$ and the undesired spectral-signature matrix is denoted by $U = [\mathbf{m}_{C+1}, \cdots, \mathbf{m}_p]$ along with its corresponding abundance vector $\mathbf{B} = (\alpha_{C+1}, \cdots, \alpha_p)^T$. A generalized orthogonal-subspace projection classifier can be derived easily by the procedure given in [14] to yield

$$P_U^\perp \mathbf{r} = P_U^\perp D\mathbf{A} + P_U^\perp \mathbf{n} \tag{21}$$

where $P_U^\perp = I - UU^\#$ and $U^\# = (U^T U)^{-1} U^T$ is the pseudo-inverse of U, and the notation $\perp$ in $P_U^\perp$ indicates that the projector $P_U^\perp$ maps the observed pixel $\mathbf{r}$ into the range space $\langle U \rangle^\perp$ (the orthogonal complement of $\langle U \rangle$). Equation (21) represents a standard signal-detection problem. If the maximization of the SNR matrix given by

$$\mathrm{SNR} = \frac{[W^T P_U^\perp D\mathbf{A}][\mathbf{A}^T D^T P_U^\perp W]}{W^T P_U^\perp E[\mathbf{n}\mathbf{n}^T] P_U^\perp W} \quad \text{over} \quad W \tag{22}$$

is chosen to be the optimal criterion for the signal model (21), the maximum SNR of (20) is equivalent to finding the maximum eigenmatrix of the following generalized eigenvalue problem

$$[\sigma^2 P_U^\perp]^{-1} [P_U^\perp D\mathbf{A}(D\mathbf{A})^T P_U^\perp] \mathbf{x} = \Delta \mathbf{x}. \tag{23}$$

The maximum eigenmatrix for (23) is given by

$$\begin{aligned} \Delta_{\max} &= \mathrm{SNR}_{\max} \\ &= (\sigma^2 D^T P_U^\perp D)^{-1} [D^T P_U^\perp D\mathbf{A}][(D\mathbf{A})^T P_U^\perp D] \\ &= \frac{\mathbf{A}\mathbf{A}^T}{\sigma^2} [D^T P_U^\perp D]. \end{aligned} \tag{24}$$

Based on the approach outlined by (20)–(24), a mixed-pixel classification can be carried out by an OSP classifier $P_{\mathrm{OSP}} = D^T P_U^\perp$, that is,

$$P_{\mathrm{OSP}}\mathbf{r} = D^T P_U^\perp \mathbf{r} = D^T P_U^\perp \mathbf{A} + D^T P_U^\perp \mathbf{n}. \tag{25}$$

In other words, we first apply $P_U^\perp$ to model (25) to eliminate U, then use the matched filter M_D to extract the D from the signal model (21). A short script of the program using MAT-LAB to run the OSP on matrix computations, is given in the appendix for reference. Then, following the same arguments given for (15)–(17), and using (22)–(24), we can define the sum $\sum_{k=1}^m \rho_k$ to be the classification power produced by the m-band selection and further define the band-power ratio of the D-class classification power rendered by m bands to the total D-class classification power produced by l bands, denoted by $R_{\mathrm{OSP},D}(m)$ to be

$$R_{\mathrm{OSP},D}(m) = \frac{\displaystyle\sum_{k=1}^m \rho_k}{\displaystyle\sum_{k=1}^l \rho_k}. \tag{26}$$

Before concluding this section, three remarks are noteworthy.

1) Similar equations to (21)–(25), described in the OSP approach developed in [14], can be found in [17] and [18] as well, where the concept of OSP is very closely related to that of the simultaneous diagonalization and dimensionality reduction used in magnetic-resonance imaging applications.

2) Comparing (22) to (15), their solutions can be obtained by maximizing the Raleigh quotient. However, this is equivalent to solving a generalize eigenvalue problem that further results in (23) and (17). Nevertheless, keep in mind that they were derived from two different optimal criteria. Equation (23) is based on maximization of SNR, while (15) uses Fisher's discriminant distance to minimize the misclassification error.

3) As noted by the first remark, the OSP-based eigenanalysis-band prioritization is also an MSNR criterion. Comparing to the MSNRPCA criterion, they are different in the sense that the former makes use of subspace projection to eliminate undesired signatures and suppress noise before the MSNR criterion is applied, whereas the latter simply ranks principal components in decreasing order of magnitude of SNR. More specifically, the orthogonal subspace projection-based eigenanalysis-band prioritization has done more than what the MSNRPCA has. It not only maximizes the SNR that the MSNRPCA does, but also enhances the desired signature because of the undesired signatures annihilation and noise suppression, a task which the MSNRPCA does not do. On the other hand, the advantage of MSNRPCA over the orthogonal subspace projection-based eigenanalysis-band prioritization is that it is derived independent of classification. Therefore, it has applications other than classification.

IV. DIVERGENCE-BASED BAND DECORRELATION

As mentioned previously, band prioritization does not take care of spectral correlation. It is always possible that two near bands share so much information that they may result in similar priorities [6]. In this case, one band can well represent the other. Consequently, there is no significant loss of information to select one band while removing the other. This is particularly true for hyperspectral images. The proposed divergence-based band decorrelation is an information-theoretic criterion to measure the correlation between two band images. If the divergence between two band images is below a specified threshold, this implies that these two band images are highly correlated so that the band image with lower priority will be removed, and the band image with higher priority will be retained to represent it in the band-selection process.

From information theory [19], there is a criterion called "divergence," which can be used to measure the discrepancy between any two probability distributions. Let $\boldsymbol{p} = \{p_i\}_{i=1}^l$ and $\boldsymbol{q} = \{q_i\}_{i=1}^l$ be the image gray-level histograms of any two band images. The divergence, denoted by $D(\boldsymbol{p}, \boldsymbol{q})$, can be defined as

$$D(\boldsymbol{p}, \boldsymbol{q}) = L(\boldsymbol{p}; \boldsymbol{q}) + L(\boldsymbol{q}; \boldsymbol{p}) \tag{27}$$

where $L(\boldsymbol{p}; \boldsymbol{q})$ is the relative entropy of $\boldsymbol{p}$ with respect to $\boldsymbol{q}$ defined by

$$L(\boldsymbol{p}; \boldsymbol{q}) = \sum_{i=1}^l p_i \log \frac{p_i}{q_i} \tag{28}$$

and $L(\boldsymbol{q}; \boldsymbol{p})$ is the relative entropy $\boldsymbol{q}$ with respect to $\boldsymbol{p}$ defined by

$$L(\boldsymbol{q}; \boldsymbol{p}) = \sum_{i=1}^l q_i \log \frac{q_i}{p_i}. \tag{29}$$

Note that $L(\boldsymbol{p}; \boldsymbol{q})$ is also referred to as directed divergence, cross entropy, or Kullback–Leibler distance between $\boldsymbol{p}$ and $\boldsymbol{q}$. While (27) is symmetric [i.e., $D(\boldsymbol{p}, \boldsymbol{q}) = D(\boldsymbol{q}, \boldsymbol{p})$], (28) and (29) are generally not symmetric [i.e., $L(\boldsymbol{p}; \boldsymbol{q}) \neq L(\boldsymbol{q}; \boldsymbol{p})$].

Equation (27) can be used to measure the similarity between two images. Namely, it can measure the overlapped information contained in any pair of images. If the divergence is below a prescribed threshold, the band with lower priority will be removed. The implementation of the proposed band decorrelation is described as follows.

Algorithm for Divergence-Based Band Decorrelation: Assume that all bands have been prioritized by band prioritization $\Omega = \{B_j\}_{j=1}^l$ where $B_1 \geq B_2 \geq \cdots \geq B_l, \geq$ is the notation of priority order, and l is the total number of bands. Let $\mathbf{p}_j$ be the image gray-level histogram of the band image specified by B_j. ε is a prescribed threshold.

1) Initialization: Let counter j be set to 1 (i.e., $j = 1$ and the initial band set $\Omega_1 = \{B_1\}$).

2) At the j-stage for $j > 1$, the current visited-band image is B_j.
   For each $B_i \in \Omega_{j-1}$, compute $D(\mathbf{p}_i, \mathbf{p}_j)$.

3) If $D(\mathbf{p}_i, \mathbf{p}_j) < \varepsilon$, B_j will be removed. Let $\Omega_j = \Omega_{j-1}$, and go to step 5.

4) Otherwise, $D(\mathbf{p}_i, \mathbf{p}_j) \geq \varepsilon$ and check if the B_i is the last band in Ω_{j-1}.
   a) If it is the last band in Ω_{j-1}, then $D(\mathbf{p}_i : \mathbf{p}_j) \geq \varepsilon$ for all $B_i \in \Omega_{j-1}$. This implies that the B_j must

TABLE I
BAND NUMBERS AND BAND-POWER RATIOS REQUIRED FOR MVPCA, MSNRPCA, MMCA, AND OSP CRITERIA

	MVPCA	MSNRPCA	MMCA	SP	UBS
Number of Total Bands	10	11	10	12	12
Selected Band Numbers	87, 57, 86, 100, 99, 83, 90, 102, 70, 126	93, 70, 72, 118, 73, 95, 91, 87, 90, 135, 15	97, 79, 128, 64, 100, 78, 95, 101, 90, 14	64, 67, 85, 78, 34, 95, 31, 22, 118, 14, 101, 7	17, 34, 51, 68, 85, 102, 119, 136, 153, 170, 187, 204
Band Power Ratio	0.0896	0.2260	0.2859	0.1400	0.2399

contain some information that cannot be represented by one of the bands in Ω_{j-1}. Thus, it must be added to Ω_{j-1}. Let $\Omega_j = \Omega_{j-1} \cup \{B_j\}$, and go to step 5.

   b) If it is not the last band in Ω_{j-1}, continue to check next B_i in Ω_{j-1}, and go to step 3.

5) If $j = l$, stop algorithm. Otherwise, let $j = j + 1$, and go to step 2.

It should be noted that in steps 2–4, all B_i's in Ω_{j-1} must be checked, examined, and compared to B_j. The band set Ω_l, produced by the above band-decorrelation algorithm, is the final desired band set that will be used for classification.

The band-selection problem considered in this section is very similar to a best-subset selection for multiple linear-regression analysis in [20], where various computational methods are proposed for minimizing the sum of squared residuals (i.e., least-squares error). However, there are two differences. One is that the band-decorrelation algorithm described above is based on the criterion of divergence rather than least-squares error. It is a simple exhaustive process. Another difference is that the methods described in [20] assume that the size of the selected subset must be known *a priori* (in our case, the number of bands needing to be selected must be known in advance), whereas our algorithm does not require this assumption. It terminates as long as a prescribed threshold is met. This is a significant advantage because in reality, we do not know how many bands are required for band selection. As will be shown in the experiments, if the number of selected bands is known, a simple uniform-band selection without finding specific bands can achieve good results. In this case, there is no need for band-selection algorithms.
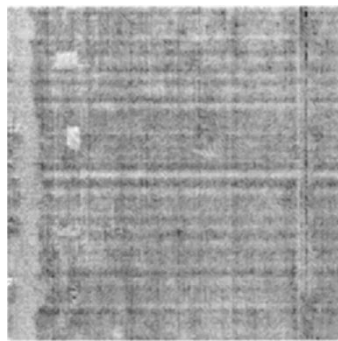
V. EXPERIMENTS

In this section, a series of experiments using a HYDICE image is presented to illustrate the five eigenanalysis-based band-prioritization criteria used for band selection. A comparative study is also conducted along with uniform band selection (UBS), in which the bands are selected uniformly. The HYDICE image used, shown in Fig. 1 (image of band 30), is radiance data taken in Maryland in August 1995 using 210 bands with 10 nm spectral resolution and spectral coverage 0.4–2.5 μ. The ground-sampling distance (GSD) is approximately 0.78 m. The figure has a size of 128 × 128 and shows a large grass field with tree lines running along the left edge. This field contains a road running along the right edge of
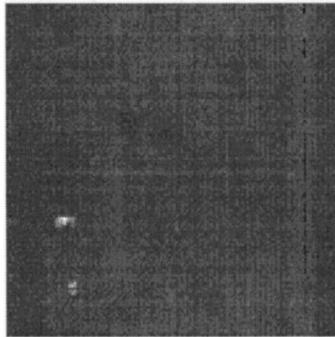

Fig. 1. HYDICE image scene of band 30.

the image. There are four vehicles vertically aligned where the top three are treaded vehicles and the bottom one is a wheeled vehicle. The size of the treaded vehicles is approximately 4 × 8 m, and the size of the wheeled vehicle is about 3 × 6 m. In addition to these vehicles, there is an object located in the center of the scene. The following experiments are designed to demonstrate how the band prioritization and band decorrelation can be used jointly for band selection and how they affect classification results. The classifier used for target classification to conduct a comparative study of various band-selection criteria is specified by (25) (i.e., the OSP classifier derived in [14]).
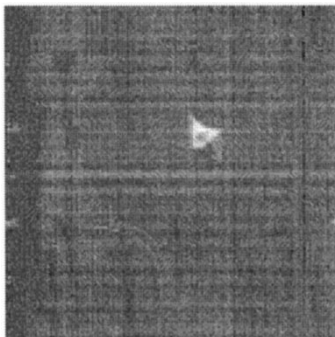
First of all, we assume that the signature matrix M is made up of six signatures {the first treaded vehicle, the wheeled vehicle, object, grass, tree, and road} where the desired signature matrix D consists of three man-made signatures, {treaded vehicle, wheeled vehicle, object}, and the undesired signature matrix is composed of the remaining three natural-background signatures {grass, tree, road}. All the required signature information was directly extracted from the image scene. The experiments were done by first prioritizing all bands using the four-band prioritization criteria proposed in this paper: MVPCA, MSNRPCA, MMCA, and OSP and then following with the divergence-based band decorrelation described in the previous section. Finally, (4), (10), (18), and (26) were used to calculate the band-power ratio R for the selected band sets. For these experiments, the divergence threshold was set to 1.5. It was empirically chosen and seemed appropriate. The numbers of bands required for MVPCA, MSNRPCA, MMCA, and OSP are tabulated in Table I and in the range from 10 to 12. If the threshold was set too high, only a few bands could be selected, and results were not good.
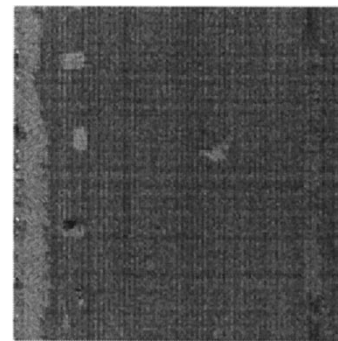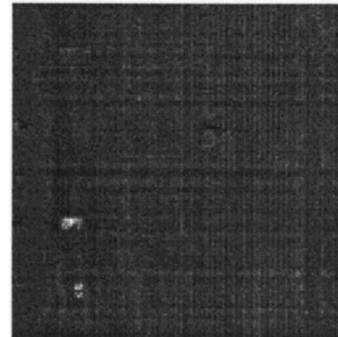
Fig. 2. Classification results produced by MVPCA according to Table I: (a) for treaded vehicles, (b) for wheeled vehicles, and (c) for the object.

Fig. 3. Classification results produced by MSNRPCA according to Table I: (a) for treaded vehicles, (b) for wheeled vehicles, and (c) for the object.

If the threshold was set too low, many unnecessary redundant bands were selected that did not provide much additional information. Also listed in Table I are their corresponding band numbers in order of priority and associated band-power ratios. Figs. 2–5 show the classification results using MVPCA, MSNRPCA, MMCA, and OSP band-prioritization criteria with bands selected according to Table I. In order to evaluate the performance of these four criteria, two more experiments are also included. Fig. 6 is produced by a UBS where the bands were selected uniformly from 210 bands, with the total number of bands chosen to be the maximum number of bands required for any of four criteria. In this case, 12 was the highest number (for OSP) and thus, 12 bands uniformly distributed among the 210 were chosen. Fig. 7 is produced by using the full set of 210 bands. As shown in Figs. 2–7, images labeled by (a) are the classification results for treaded vehicles, images labeled by (b) are for wheeled vehicles, and images labeled by (c) are

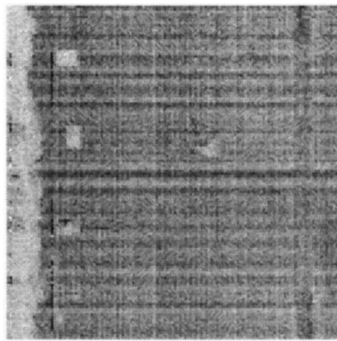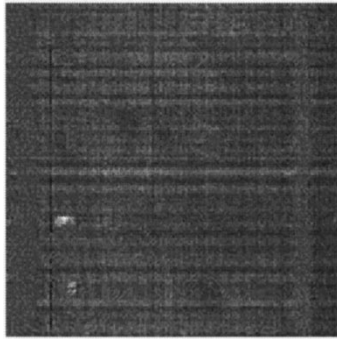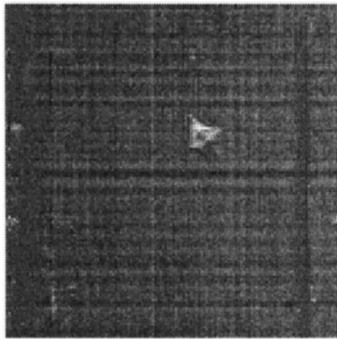for objects. Comparing Figs. 2–6 to Fig. 7, all four criteria along with UBS produced comparable results. In order to conduct a comparative analysis, Fig. 7 was used as a base for comparison. This is because the objective of band selection is to select appropriate bands that can retain as much information contained in Fig. 7 as possible. According to this criterion, OSP produced the best-matched results to Fig. 7, while Fig. 2 may be the worst. This can be explained by the fact that the variance-based band-power ratio of MVPCA might still account for part of the variances contributed by those removed bands that may be correlated with selected bands. In order to resolve this correlation problem, MSNRPCA was introduced to improve MVPCA. As evidenced in Fig. 3, the results produced by MSNRPCA are generally better than those in Fig. 2 in terms of target classification. Fig. 3 looks closer to Fig. 7 than Fig. 2 does to Fig. 7 after the between-band correlation was properly taken care of by a whitening matrix implemented in
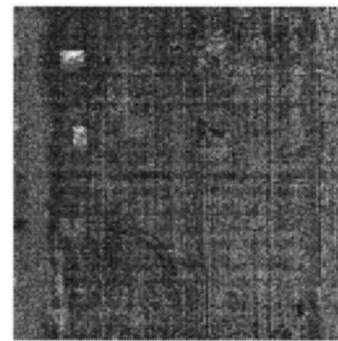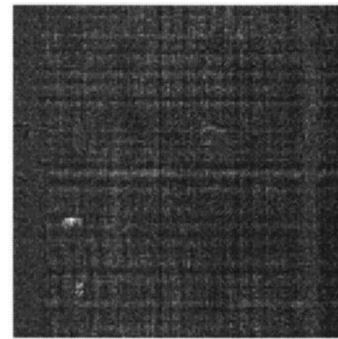
Fig. 4. Classification results produced by MMCA according to Table I: (a) for treaded vehicles, (b) for wheeled vehicles, and (c) for the object.

Fig. 5. Classification results produced by OSP according to Table I: (a) for treaded vehicles, (b) for wheeled vehicles, and (c) for the object.
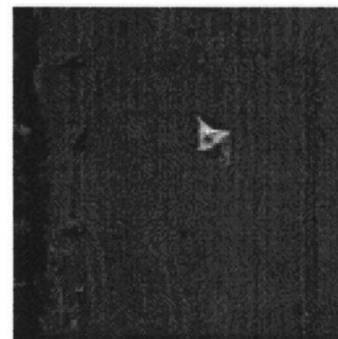
MSNRPCA. Figs. 2–6 show different performances one way or another. For instance, Fig. 2 shows the best detection of the wheeled vehicle but may be the worst detection of the treaded vehicles and the object. Fig. 3 produced by MSNRPCA shows a better detection for the wheeled vehicle and the object but not for treaded vehicles. On the contrary, Fig. 6, produced by UBS, shows a good detection of treaded vehicles and the object but a bad detection of the wheeled vehicle because it also detected the object as well. Fig. 4, produced by MMCA, can be ranked between Figs. 3 and 6, where the detection performance of vehicles is between MSNRPCA and UBS, but the detection of the object turns out to be the worst of the three. It is interesting to note that Fig. 7, produced by the entire 210 bands, shows a bright broken line along the vehicles. This line is caused by a strong interferer in the scene, which is only shown in a few band images among all 210 bands. No criteria detected it.

It is also noted that in all figures, the third treaded vehicle was missed when the treaded vehicles were classified. But, it was picked up in the wheeled-vehicle classification. This occurrence is not surprising, because the spectrum of the third treaded vehicle is much more similar to that of the wheeled vehicle than to those of the first two treaded vehicles. For more details, we refer to [16]. As a result, classifying one will detect the other. In this case, spatial information such as shape or size may be useful to help us separate these two vehicles, because one is larger than the other.

Table I also shows some interesting findings. In order to produce classification results comparable to Fig. 7, no more than 12 bands (i.e., approximately 6% of 210 bands) are needed for the proposed band-selection method. More surprisingly, their corresponding band-power ratios are even less than a quarter of total band energy. This implies that most of the 210

Fig. 6. Classification results produced by the UBS: (a) for treaded vehicles, (b) for wheeled vehicles, and (c) for the object.

Fig. 7. Classification results produced by using all 210 bands: (a) for treaded vehicles, (b) for wheeled vehicles, and (c) for the object.

bands are either redundant or insignificant. So, the proposed joint-band prioritization and band-decorrelation approach to band selection can effectively eliminate a great number of unnecessary bands and achieve good classification results. The band reduction rate can be as high as 94%. This is a tremendous advantage that offers substantial saving in storage and computation. Nevertheless, these bands cannot be randomly selected. This must be done by a careful selection of desired bands using an appropriate criterion. In our proposed UBS, it is determined by the maximum number of bands required for any of the four band-prioritization criteria. Furthermore, despite no band content-based selection criterion for UBS, a value for the total number of bands to select is needed. This value cannot be randomly determined.

Although the experimental results presented in this paper were conducted based on only one image scene (which was also studied in [16]), several other HYDICE images were also
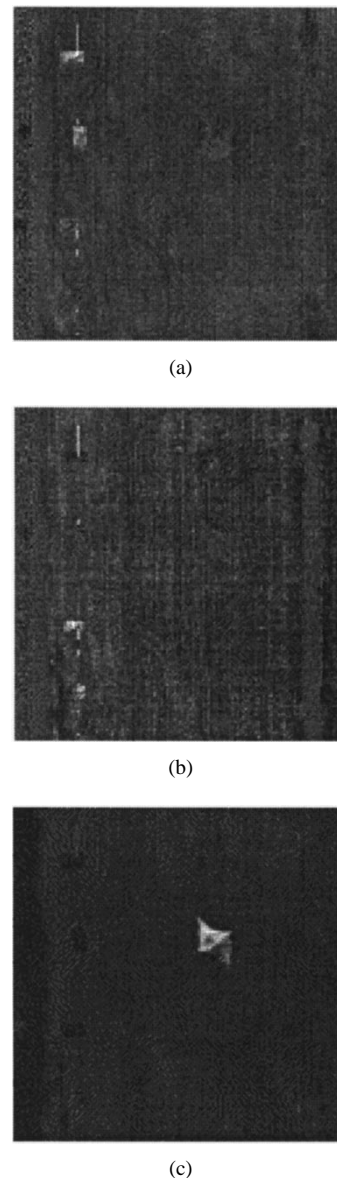
tested for comparative analysis. They all resulted in similar conclusions [6].

VI. CONCLUSION

This paper presented a joint band-prioritization and band-decorrelation approach to band selection. The band prioritization was based on an eigenanalysis and decomposed a matrix into an eigenform matrix from which a loading-factors matrix could be constructed and used to prioritize bands. The loading factors determined the priority of each band and ranked all bands in accordance with their associated priorities. The band prioritization was then followed by a divergence-based band decorrelation that used the divergence measure to remove redundant or insignificant bands. As shown by HYDICE data, the proposed band-selection method could effectively reduce band dimensionality with very little loss of information in hyperspectral image classification. Recently,

it has been shown in [16] that interference played a significant role in hyperspectral target detection and image classification. By considering interference as an unknown separate source, an interference and noise-adjusted PCA (INAPCA), derived from NAPC, was developed in [21]. It considerably improved PCA performance via interference annihilation. However, in this case, finding potential interfering signatures for annihilation (as well as accurately estimating noise variances) is crucial to satisfactory performance [21]–[23]. If both the noise and interference can be reliably estimated, then an error analysis based on SNR or receiver-operating characteristics (ROC) for the band selection may become possible. A further study on this issue may be worth pursuing.

APPENDIX

In what follows, a short script of the program using MAT-LAB to run the OSP for matrix computations is provided for reference.

```
% load desired D and undesired U signatures
load U
load D
% generate the OSP projector
P = eye(210) − U ∗ inv(U' ∗ U) ∗ U';
IP = inv(P);
% input parameters of the data set
head=512;
column=320;
totalband=210;
bpp = 2;
frame=column*totalband*bpp;
begindot=41;
enddot=168;
beginline=191;
endline=318;
dotnum=enddot-begindot+1;
linenum=endline-beginline+1;
% calculate the matrix S for OSP
S=zeros(totalband);
fid = fopen('filename', 'r');
fseek(fid, head, 'bof');
fseek(fid, (beginline − 1) ∗ frame, 'cof');
fseek(fid, (begindot − 1) ∗ bpp, 'cof');
for i = 1:linenum
    block=zeros(totalband,dotnum);
    for j = 1:totalband
        block(j,:)=fread(fid,[1,dotnum],'int16');
        fseek(fid, (column-dotnum) ∗ bpp, 'cof');
    end
    block=P*block;
    for j = 1:dotnum
        S = S + block(:, j) ∗ block(:, j)';
    end
end
fclose(fid);
S = IP ∗ S/16 384;
% the end
```
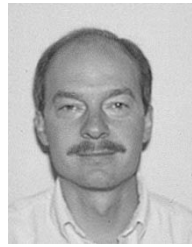
REFERENCES

[1] P. W. Mausel, W. J. Kramber, and J. K. Lee, "Optimum band selection for supervised classification of multispectral data," *Photogramm. Eng. Remote Sens.*, vol. 56, no. 1, pp. 55–60, Jan. 1990.

[2] J. A. Richards, *Remote Sensing Digital Image Analysis*, 2nd ed. New York: Springer-Verlag, 1993.

[3] C. Conese and F. Maselli, "Selection of optimum bands from TM scenes through mutual information analysis," *ISPRS J. Photogramm. Remote Sensing*, vol. 48, no. 3, pp. 2–11, 1993.

[4] S. D. Stearns, B. E. Wilson, and J. R. Peterson, "Dimensionality reduction by optimal band selection for pixel classification of hyperspectral imagery," *Applicat. Digital Image Processing XVI*, vol. 2028, pp. 118–127, 1993.

[5] M. L. G. Althouse, "Vapor cloud detection in multispectral infrared image sequences using co-occurrence matrix methods," Ph.D. dissertation, Dep. Comp. Sci. Elect. Eng., Univ. Maryland, Baltimore County, Baltimore, Aug. 1994.

[6] C.-I. Chang, T.-L. Sun, and Q. Du, "Eigen-analysis based band prioritization approach to band selection for hyperspectral image classification," Lab. Rep., Remote Sensing Signal and Image Processing Lab., Dep. Comp. Sci. Elect. Eng., Univ. Maryland, Baltimore County, Baltimore, 1997.

[7] T. M. Tu, C.-H. Chen, J.-L. Wu and C.-I. Chang, "A fast two-stage classification method for high dimensional remote sensing data," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, pp. 182–191, Jan. 1998.

[8] G. P. McCabe, "Principal variables," *Technometrics*, vol. 25, no. 2, pp. 137–144, 1984.

[9] 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, Jan. 1988.

[10] 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.

[11] R. E. Roger, "A fast way to compute the noise-adjusted principal components transform matrix," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 1194–1196, Nov. 1994.

[12] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

[13] S. Haykin, *Neural Networks*. New York: Macmillan, 1994.

[14] J. Harsanyi and C.-I Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sensing*, vol. 32, pp. 779–785, July 1994.

[15] C.-I Chang, X. Zhao, M. L. G. Althouse, and J.-J. Pan, "Least squares subspace projection approach to mixed pixel classification in hyperspectral images," *IEEE Trans. Geosci. Remote Sensing*, vol. 36, pp. 898–912, May 1998.

[16] 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, pp. 735–743, Mar. 1998.

[17] J. W. V. Miller, J. P. Windham, and S. C. Kwatra, "Optimal filtering of radiographic image sequences using simultaneous diagonalization," *IEEE Trans. Med. Imag.*, vol. MI-3, pp. 116–123, 1984.

[18] J. W. V. Miller, J. B. Farison, and Y. Shin, "Spatially invariant image sequences," *IEEE Trans. Image Processing*, vol. 1, pp. 148–161, Apr. 1992.

[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[20] W. J. Kennedy, Jr. and J. E. Gentle, *Statistical Computing*. New York: Marcel-Dekker, 1980.

[21] C.-I. Chang and Q. Du, "Interference and noise adjusted principal components analysis," *IEEE Trans. Geosci. Remote Sensing*, vol. 37, pp. 2387–2396, Sept. 1999.

[22] R. E. Roger, "Principal components transform with simple, automatic noise adjustment," *Int. J. Remote Sensing*, vol. 17, no. 14, pp. 2719–2727, 1996.

[23] R. E. Roger and J. F. Arnold, "Reliability estimating the noise in AVIRIS hyperspectral images," *Int. J. Remote Sensing*, vol. 17, no. 10, pp. 1951–1962, 1996.

Chein-I Chang (S'81–M'87–SM'92), for a photograph and biography, see p. 2396 of the September 1999 issue of this TRANSACTIONS.

Qian Du (S'98), for a photograph and biography, see p. 2396 of the September 1999 issue of this TRANSACTIONS.

Tzu-Lung Sun, photograph and biography not available at the time of publication.

Mark L. G. Althouse (S'90–M'92) received the B.S. degree in physics from Pennsylvania State University, State College, and the M.S. and Ph.D degrees in electrical engineering from Johns Hopkins University, Baltimore, MD and the University of Maryland, Baltimore County (UMBC), Baltimore, respectively.

Since 1981, he has been with the U.S. Army Chemical and Biological Defense Command, Aberdeen Proving Ground, Aberdeen, MD, working in the area of detection and identification of chemical and biological agent clouds. Recently, he has concentrated on remote infrared sensors for chemical vapor detection, both optical design and signal processing methods. He is a part-time Faculty Member in the Department of Computer Science and Electrical Engineering, UMBC.

Dr. Althouse is a member of Tau Beta Pi, Sigma Xi, OSA, and SPIE.