Estimation of Number of Spectrally Distinct Signal Sources in Hyperspectral Imagery

Chein-I Chang, *Senior Member, IEEE,* and Qian Du, *Member, IEEE*

***Abstract*—With very high spectral resolution, hyperspectral sensors can now uncover many unknown signal sources which cannot be identified by visual inspection or *a priori*. In order to account for such unknown signal sources, we introduce a new definition, referred to as virtual dimensionality (VD) in this paper. It is defined as the minimum number of spectrally distinct signal sources that characterize the hyperspectral data from the perspective view of target detection and classification. It is different from the commonly used intrinsic dimensionality (ID) in the sense that the signal sources are determined by the proposed VD based only on their distinct spectral properties. These signal sources may include unknown interfering sources, which cannot be identified by prior knowledge. With this new definition, three Neyman–Pearson detection theory-based thresholding methods are developed to determine the VD of hyperspectral imagery, where eigenvalues are used to measure signal energies in a detection model. In order to evaluate the performance of the proposed methods, two information criteria, an information criterion (AIC) and minimum description length (MDL), and the factor analysis-based method proposed by Malinowski, are considered for comparative analysis. As demonstrated in computer simulations, all the methods and criteria studied in this paper may work effectively when noise is independent identically distributed. This is, unfortunately, not true when some of them are applied to real image data. Experiments show that all the three eigenthresholding based methods (i.e., the Harsanyi–Farrand–Chang (HFC), the noise-whitened HFC (NWHFC), and the noise subspace projection (NSP) methods) produce more reliable estimates of VD compared to the AIC, MDL, and Malinowski's empirical indicator function, which generally overestimate VD significantly. In summary, three contributions are made in this paper, 1) an introduction of the new definition of VD, 2) three Neyman–Pearson detection theory-based thresholding methods, HFC, NWHFC, and NSP derived for VD estimation, and 3) experiments that show the AIC and MDL commonly used in passive array processing and the second-order statistic-based Malinowski's method are not effective measures in VD estimation.**

***Index Terms*—An information criterion (AIC), Malinowski's emperical indicator function (EIF), minimum description length (MDL), Neyman–Pearson detection, noise subspace projection (NSP), virtual dimensionality (VD).**

I. Introduction

DETERMINATION of intrinsic dimensionality (ID) for remotely sensed imagery is a very challenging problem [1]. According to the definition given in [2], the ID, also referred to as effective dimensionality, is the minimum number of parameters required to account for the observed properties of the data. The true dimensionality of multivariate data is difficult to determine in practice, since its ID cannot be simply determined by the dimensionality of a data sample vector, referred to as component dimensionality in this paper, which is defined by the number of components in a data vector. In particular, when very high-dimensional data are well structured, data tend to be distributed in a low-dimensional space. In this case, the ID is expected to be much smaller than the component dimensionality. Several well-known methods have been proposed such as principal components analysis (PCA) [3] and factor analysis [4], [5], which make use of the eigenvalue distribution to determine the ID. Such approaches may be suitable to multispectral imagery, since only a small number of bands are used to acquire multispectral imagery and the resulting ID is expected to be small. However, a direct application of these methods to hyperspectral imagery can be difficult. More importantly, it may not be effective even if it is applicable [6]. This is particularly true for hyperspectral imagery, where the component dimensionality is the dimensionality of a pixel vector specified by hundreds of spectral channels. The ID of the image is considerably smaller than the component dimensionality. Accordingly, accurately determining the ID is crucial and also critical to the success of hyperspectral image analysis, such as linear unmixing methods, which require the knowledge of how many image endmembers are in image data. Compared to multispectral imagery, hyperspectral imagery has very high component dimensionality, and determining the ID could be more problematic for hyperspectral imagery than for multispectral imagery. This is because a high spectral resolution hyperspectral sensor has the capability of uncovering many unknown target sources spectrally, which cannot be identified by visual inspection or *a priori*.

In order to account for such unknown signal sources, and also to avoid confusion with the ID, we introduce a new definition, referred to as virtual dimensionality (VD) in this paper. It is defined as the minimum number of spectrally distinct signal sources that characterize the hyperspectral data from the perspective view of target detection and classification rather than the image endmembers defined in [7], which are idealized pure signatures. It is different from the ID in the sense that signal sources are considered to be different if they are spectrally distinct. It may not be necessarily measured by the same parameters used to measure the ID. Such signal sources may include known and unknown image endmembers, natural signatures, anomalies, and interferers [8]. It should be noted that the definition of the VD used here is our preference and is by no means a standard terminology. It has a narrow implication driven by the requirements of target detection and classification.

Manuscript received September 20, 2002; revised September 5, 2003. This work was supported by the Office of Naval Research under Contract N00014-01-1-0359.

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.

Q. Du is with the Department of Electrical Engineering and Computer Science, Texas A&M University, Kingsville, TX 78363 USA.

Digital Object Identifier 10.1109/TGRS.2003.819189

A practical issue encountered in PCA-based approaches to VD estimation is that it is very difficult to determine the cutoff threshold between the eigenvalues caused by signals and noise, in particular, when a change between two adjacent eigenvalues is not significant. Malinowski [4], [5] developed an empirical indicator function (EIF) that was used to determine the threshold. He used eignevalues to reconstruct the data correlation matrix and measured the difference between the original and its reconstructed matrices, which was used to estimate the ID. As shown in the experiments, it does not work well for real data in remote sensing. In passive array processing, two information criteria, an information criterion (AIC) [9] and minimum description length (MDL) [10], [11], were used to estimate the number of signals impinging upon the array. It seemed that these criteria might be applicable to VD estimation. Unfortunately, as will be demonstrated in our experiments, this turns out not to be the case.

Recently, Harsanyi, Farrand, and Chang developed an eigenthresholding method, referred to as Harsanyi–Farrand–Chang (HFC) method, to determine the number of spectral endmembers in Airborne Visible Infrared Imaging Spectrometer (AVIRIS) data [12]. It was derived from the concept of the Neyman–Pearson detection theory [13]. Their idea can be briefly described as follows. Let the eigenvalues generated by the sample correlation matrix and the sample covariance matrix be denoted by correlation eigenvaules and covariance eigenvalues, respectively. Since the component dimensionality is equal to the total number of eigenvalues, each eigenvalue specifies a component dimension and provides an indication of the significance of that particular component in terms of energy or variance. If there is no signal source contained in a particular component, the corresponding correlation eigenvalue and covariance eigenvalue in this component should reflect only the noise energy, in which case, correlation eigenvalue and covariance eigenvalue are equal. This fact provides us with a base from which we can formulate the difference between the correlation eigenvalue and its corresponding covariance eigenvalue as a binary composite hypothesis testing problem. The null hypothesis represents the case of the zero difference, while the alternative hypothesis indicates the case that the difference is greater than zero. When the Neyman–Pearson test is applied to each pair of correlation eigenvalue and its corresponding covariance eigenvalue, the number of times the test fails indicates how many signal sources are present in the image. In other words, a failure of the Neyman–Pearson test in a component indicates a truth of the alternative hypothesis, which implies that there is a signal source in this particular component. Using this approach, we can estimate the VD with the receiver operating characteristic (ROC) analysis for evaluating the effectiveness of the decision. Since the HFC method does not have a noise-whitening process that decorrelates noise with signal sources to enhance signal detection performance, an alternative approach is to include a noise-whitening process in the HFC method to remove noise effects on eigenvalues. The resulting method will be referred to as noise-whitened HFC (NWHFC) method. In this case, a noise estimation is required for the NWHFC method.

Since both the HFC and NWHFC methods model the correlation eigenvalue and the corresponding covariance eigenvalue as random parameters in [14], the sample size must be sufficiently large to ensure that the covariance between these two types of eigenvalues is asymptotically zero. However, this may not be valid for a small sample size. In order to alleviate this problem, a third method, referred to as noise subspace projection (NSP) approach, was recently proposed in [15], which used the covariance eigenvalues only. It uses the NWHFC method as a transition to derive an approach that can take advantage of strengths from both HFC and NWHFC. It first estimates the noise covariance matrix of image data, then uses it to whiten the noise variances to unity. Consequently, the binary composite hypothesis problem used in the HFC method can be simplified and reduced to the one that the null hypothesis represents the case of unity variance while the alternative hypothesis being the case of variance greater than unity. Accordingly, calculating the difference between correlation eigenvaules and covariance eigenvalues can be avoided. An immediate benefit from the NSP method is that the sample size is not necessarily large as required by the HFC and NWHFC methods. In order to estimate the noise covariance matrix, a method suggested by Roger and Arnold [16], [17] and Du and Chang [18] will be used in this paper.

The remainder of this paper is organized as follows. Section II derives the HFC method using a Neyman–Pearson detection theory-based eigenthresholding approach. Section III reviews a method proposed by Roger and Arnold for noise covariance matrix estimation. Section IV presents two noise-whitened eigenthresholding methods, the noise-whitened HFC method and the NSP method, plus two information criteria (an AIC and MDL) and Malinowski's method, which will be used for comparative analysis. Section V conducts computer simulations and real data experiments to evaluate the effectiveness of the methods proposed in this paper. Finally, a conclusion is drawn in Section VI.

II. Neyman–Pearson Detection Theory-Based Eigenthresholding Method

A Neyman–Pearson detection theory-based eigenthreshold method, referred to as the HFC method, was previously developed in [12] to determine the number of endmembers in AVIRIS data. It first calculated the sample correlation matrix $\mathbf{R}_{L\times L}$ and sample covariance matrix $\mathbf{K}_{L\times L}$ and then found the difference between their corresponding eigenvalues. Let $\{\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_L\}$ and $\{\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L\}$ be two sets of eigenvalues generated by $\mathbf{R}_{L\times L}$ and $\mathbf{K}_{L\times L}$, called correlation eigenvalues and covariance eigenvalues, respectively. By assuming that signal sources are nonrandom unknown positive constants and noise is white with zero mean, we can expect that

$$\hat{\lambda}_l > \lambda_l, \qquad \text{for} \quad l = 1, \ldots, \text{VD} \tag{1}$$

and

$$\hat{\lambda}_l = \lambda_l, \qquad \text{for} \quad l = \text{VD}+1, \ldots, L. \tag{2}$$

More specifically, the eigenvalues in the lth spectral channel can be related by

$$\begin{aligned} \hat{\lambda}_l > \lambda_l > \sigma^2_{n_l}, &\qquad \text{for} \quad l = 1, \ldots, \text{VD} \\ \hat{\lambda}_l = \lambda_l = \sigma^2_{n_l}, &\qquad \text{for} \quad l = \text{VD}+1, \ldots, L \end{aligned} \tag{3}$$

where $\sigma^2_{n_l}$ is the noise variance in the lth spectral channel.

In order to determine the VD, Harsanyi *et al.* [9] formulated the problem of determination of VD as a binary hypothesis problem as follows:

$$H_0 : z_l = \hat{\lambda}_l - \lambda_l = 0 \quad \text{versus} \quad H_1 : z_l = \hat{\lambda}_l - \lambda_l > 0, \quad \text{for} \quad l = 1, 2, \ldots, L \quad (4)$$

where the null hypothesis H_0 and the alternative hypothesis H_1 represent the case that the correlation eigenvalue is equal to its corresponding covariance eigenvalue and the case that the correlation eigenvalue is greater than its corresponding covariance eigenvalue, respectively. In other words, when H_1 is true (i.e., H_0 fails), it implies that there is an endmember contributing to the correlation eigenvalue in addition to noise, since the noise energy represented by the eigenvalue of $\mathbf{R}_{L\times L}$ in that particular component is the same as the one represented by the eigenvalue of $\mathbf{K}_{L\times L}$ in its corresponding component.

Despite the fact that the $\hat{\lambda}_l$ and λ_l in (1)–(3) are unknown constants, according to [14], we can model each pair of eigenvalues $\hat{\lambda}_l$ and λ_l under hypotheses H_0 and H_1 as random variables by the asymptotic conditional probability densities given by

$$p_0(z_l) = p(z_l|H_0) \cong N\left(0, \sigma_{z_l}^2\right), \quad \text{for} \quad l = 1, 2, \ldots, L \quad (5)$$

and

$$p_1(z_l) = p(z_l|H_1) \cong N\left(\mu_l, \sigma_{z_l}^2\right), \quad \text{for} \quad l = 1, 2, \ldots, L \quad (6)$$

respectively, where μ_l is an unknown constant and the variance $\sigma_{z_l}^2$ is given by

$$\sigma_{z_l}^2 = \text{Var}[\hat{\lambda}_l - \lambda_l] = \text{Var}[\hat{\lambda}_l] + Var[\lambda_l] - 2\,\text{Cov}(\hat{\lambda}_l, \lambda_l), \quad \text{for} \quad l = 1, 2, \ldots, L. \quad (7)$$

It is shown in [14] that when the total number of samples N is sufficiently large, $\text{Var}[\hat{\lambda}_l] \cong (2\hat{\lambda}_l^2/N)$ and $\text{Var}[\lambda_l] \cong (2\lambda_l^2/N)$. Therefore, the noise variance $\sigma_{z_l}^2$ in (5) and (6) can be estimated and approximated using (7).

Now, we use Schwarz's inequality to bound $\text{Cov}(\hat{\lambda}_l, \lambda_l)$ in (7) as follows:

$$\text{Cov}(\hat{\lambda}_l, \lambda_l) \le \sqrt{\text{Var}[\hat{\lambda}_l]\text{Var}[\lambda_l]} \cong \frac{2}{N}(\hat{\lambda}_l \cdot \lambda_l). \quad (8)$$

If we further assume that the estimators $\hat{\lambda}_l$ and λ_l are consistent in mean square, the variances of $\hat{\lambda}_l$ and λ_l are asymptotically zero. In other words, $\text{Var}[\hat{\lambda}_l] \cong (2\hat{\lambda}_l^2/N)$ and $\text{Var}[\lambda_l] \cong (2\lambda_l^2/N)$ converge to zero as $N \to \infty$. This further implies that $\text{Cov}(\hat{\lambda}_l, \lambda_l) \to 0$ as well as

$$\sigma_{z_l}^2 = \text{Var}[\hat{\lambda}] + \text{Var}[\lambda] \approx \frac{2\hat{\lambda}_l^2}{N} + \frac{2\lambda_l^2}{N} \to 0, \quad \text{for} \quad l = 1, 2, \ldots, L \text{ as } N \to \infty. \quad (9)$$

From (5), (6), and (9), we define the false-alarm probability and detection power (i.e., detection probability) [13] as follows:

$$P_F = \int_{\tau_l}^{\infty} p_0(z)dz \quad (10)$$

$$P_D = \int_{\tau_l}^{\infty} p_1(z)dz. \quad (11)$$

A Neyman–Pearson detector for $\hat{\lambda}_l - \lambda_l$, denoted by $\delta_{\text{NP}}(\hat{\lambda}_l - \lambda_l)$, in the binary composite hypothesis testing problem specified by (4) can be obtained by maximizing the detection power P_D in (11), while the false-alarm probability P_F in (10) is fixed at a specific given value, α, which determines the threshold value τ_l in (10) and (11). So, a case of $\hat{\lambda}_l - \lambda_l > \tau_l$ indicating that $\delta_{\text{NP}}(\hat{\lambda}_l - \lambda_l)$ fails the test, in which case there is signal energy assumed to contribute to the eigenvalue $\hat{\lambda}_l$ in the lth data dimension. It should be noted that the test for (4) must be performed for each of L data dimensions. Therefore, for each pair of $\hat{\lambda}_l - \lambda_l$, the threshold τ is different and should be l-dependent, i.e., τ_l.

To conclude this section, a comment is noteworthy. In order to show that $\text{Cov}(\hat{\lambda}_l - \lambda_l) \to 0$ as $N \to \infty$, we have assumed that both $\hat{\lambda}_l$ and λ_l are consistent in mean square. However, this assumption is not crucial and does not affect the HFC method. Since $\hat{\lambda}_l$ and λ_l are maximum-likelihood estimators, they are asymptotically unbiased and consistent [13]. In other words, $\hat{\lambda}_l$ and λ_l are always consistent estimators in the sense that $\hat{\lambda}_l$ and λ_l converge in probability to real correlation eigenvalues and real covariance eigenvalues, respectively, regardless of whether $\text{Cov}(\hat{\lambda}_l, \lambda_l) \to 0$ as $N \to \infty$. Therefore, the binary composite hypothesis testing problem described by (4) is always valid. The only issue is that the variance of z_l may or may not be approximated via (9), depending upon the factor that $\text{Cov}(\hat{\lambda}_l, \lambda_l) \to 0$ as $N \to \infty$. Should (9) fail, the estimate of the variance of z_l used in (10) and (11) may not be accurate. This is the reason that we introduce the NWHFC and NSP approach in the following sections.

III. Estimation of Noise Covariance Matrix

As noted in (7), $\text{Cov}(\hat{\lambda}_l, \lambda_l)$ is not necessarily zero, if the $\hat{\lambda}_l$ and λ_l are not asymptotically consistent in mean square. However, $\text{Cov}(\hat{\lambda}_l, \lambda_l)$ can be reduced if a noise-whitened process is included prior to performing the hypothesis testing described by (4). In order to take advantage of this, two approaches will be developed in Section IV, which are the NWHFC method in Section IV-A and the NSP method in Section IV-B. Of particular interest is the NSP method, which requires the computation of either sample covariance matrix $\mathbf{K}_{L\times L}$ or sample correlation matrix $\mathbf{R}_{L\times L}$, but not both as do the HFC and NWHFC methods. The only issue needed to be resolved is to estimate the second-order statistics of the noise.

There exist many methods to estimate the noise covariance matrix, such as residual-based estimation [16] taking into account intraband correlation, nearest neighbor difference (NND) [19], [20] using interband correlation, and linear regression model-based prediction [17] taking advantage of intra/interband correlation. Since NND has shown to be a poor estimation [6], only the residual analysis method developed by Roger and his colleagues will be discussed in this paper for noise covariance matrix estimation.

The $\mathbf{K}_{L\times L}$ can be decomposed as follows:

$$\mathbf{K}_{L\times L} = \mathbf{D}_{\mathbf{K}}\mathbf{E}_{\mathbf{K}}\mathbf{D}_{\mathbf{K}} \quad (12)$$

where $\mathbf{D}_{\mathbf{K}}$ is a diagonal matrix given by $\mathbf{D}_{\mathbf{K}} = \text{diag}\{\sigma_1, \sigma_2, \ldots, \sigma_L\}$ with $\{\sigma_l^2\}_{l=1}^{L}$ being diagonal elements of $\mathbf{K}_{L\times L}$, which is the variance of B_l, and

$$\mathbf{E}_{\mathbf{K}} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \cdots & \rho_{1L} \\ \rho_{21} & 1 & \rho_{23} & \cdots & \rho_{2L} \\ \rho_{31} & \rho_{32} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \rho_{(L-1)L} \\ \rho_{L1} & \rho_{L2} & \cdots & \rho_{L(L-1)} & 1 \end{bmatrix} \quad (13)$$

with ρ_{mn} being the correlation coefficient at the (m,n)th entry of $\mathbf{K}_{L\times L}$ and $m \neq n$.

Similarly, in analogy with the decomposition of $\mathbf{K}_{L\times L}$, its inverse $\mathbf{K}^{-1}_{L\times L}$ can be also decomposed as

$$\mathbf{K}^{-1}_{L\times L} = \mathbf{D}_{\mathbf{K}^{-1}}\mathbf{E}_{\mathbf{K}^{-1}}\mathbf{D}_{\mathbf{K}^{-1}} \tag{14}$$

where $\mathbf{D}_{\mathbf{K}^{-1}}$ is a diagonal matrix given by $\mathbf{D}_{\mathbf{K}^{-1}} = \{\varsigma_1, \varsigma_2, \ldots, \varsigma_L\}$ with $\{\varsigma_l^2\}_{l=1}^{L}$ being the diagonal elements of $\mathbf{K}^{-1}_{L\times L}$, and

$$\mathbf{E}_{\mathbf{K}^{-1}} = \begin{bmatrix} 1 & \xi_{12} & \xi_{13} & \cdots & \xi_{1L} \\ \xi_{21} & 1 & \xi_{23} & \cdots & \xi_{2L} \\ \xi_{31} & \xi_{32} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \xi_{(L-1)L} \\ \xi_{L1} & \xi_{L2} & \cdots & \xi_{L(L-1)} & 1 \end{bmatrix} \tag{15}$$

with ξ_{mn} being the correlation coefficient at the (m,n)th entry of $\mathbf{K}^{-1}_{L\times L}$ and $m \neq n$. It turns out that the ς_l in $\mathbf{D}_{\mathbf{K}^{-1}}$ can be related to the σ_l by the following formula:

$$\varsigma_l = \sigma_l^{-1}\left(1 - r^2_{L-l}\right)^{-\frac{1}{2}} = \frac{1}{\sqrt{\sigma_l^2\left(1 - r^2_{L-l}\right)}} \tag{16}$$

where r^2_{L-l} is the multiple correlation coefficients of band B_l on the other $L-1$ bands $\{B_k\}_{k=1,k\neq l}^{L}$ obtained by using the multiple regression theory. So, ς_l^2 is the reciprocal of a good noise variance estimate of band B_l. It should be noted that the $\mathbf{D}_{\mathbf{K}^{-1}}$ in (14) is not an inverse of the $\mathbf{D}_{\mathbf{K}}$ in (12), and nor is $\mathbf{E}_{\mathbf{K}^{-1}}$ an inverse of the $\mathbf{E}_{\mathbf{K}}$. The major advantage of using ς_l over σ_l is that ς_l removes its correlation on other ς_k's for $k \neq l$, while σ_l does not.

Therefore, the noise covariance matrix $\mathbf{K}_n$ can be estimated by $\mathbf{K}_n = \text{diag}\{(1/\varsigma_1^2), (1/\varsigma_2^2), \ldots, (1/\varsigma_L^2)\}$, which is a diagonal matrix.

IV. Techniques for VD Estimation

In order for a noise-whitening process to be effective, the noise second-order statistics must be estimated accurately. The techniques introduced in the previous section have shown success in estimation of the noise covariance matrix of AVIRIS data and can be used for this purpose.

A. NWHFC Method

Since the HFC method does not have a noise-whitening process, an alternative is to modify the HFC method by including a noise-whitening process as preprocessing to remove the second-order statistical correlation such that the noise variance in the corresponding correlation eigenvalue and covariance eigenvalue will be the same. As a result, the VD estimate can be more accurate due to the fact that the noise variances have been decorrelated and do not have effects on the eigenvalue comparison. The resulting HFC method will be referred to as noise-whitened HFC method. More specifically, the reason that we introduce the NWHFC method is threefold. One is that the correlation between $\hat{\lambda}_l$ and λ_l [i.e.,Cov$(\hat{\lambda}_l, \lambda_l)$] can be reduced. Another is that a noise-whitening process makes the noise components in the corresponding eigenvalues $\hat{\lambda}_l$ and λ_l equal to 1 in (3). As a result, when there is no signal present in the lth dimension, $\hat{\lambda}_l = \lambda_l = 1$. Such a noise-whitening process generally improves the accuracy of the VD estimation

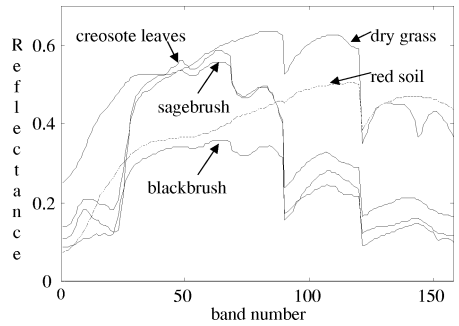


Fig. 1. Five AVIRIS reflectance spectra.

and can be viewed as a hybrid of the noise-adjust PCA in [19] and the HFC method. The third is to use it as a transition from the HFC method to the NSP-based eigenthreshloding method described in the following section.

B. NSP

The effectiveness of the HFC and NWHFC methods is based on the assumption that Cov$(\hat{\lambda}_l, \lambda_l)$ is asymptotically zero. However, for the case that sample size N is not sufficiently large, this may not be true. So, in order to mitigate this dilemma, a noise subspace projection method is developed in this section to use only $\mathbf{K}_{L\times L}$. If the noise estimate is accurate, the NSP method can outperform the HFC and NWHFC, particularly when the sample size is small. It can be viewed as a hybrid of the noise-adjusted PCA, the HFC, and NWHFC methods.

The sample covariance matrix $\mathbf{K}_{L\times L}$ can be whitened using the estimated $\mathbf{K}_n$ in Section III by $\mathbf{K}_{\mathbf{n}}^{-(1/2)}\mathbf{K}\mathbf{K}_{\mathbf{n}}^{-(1/2)}$. As a result, the noise variance of each band in the whitened $\overline{\mathbf{K}}$ is reduced to unity. Let $\{\mathbf{u}_l\}_{l=1}^{L}$ be a set of eigenvalues generated by $\overline{\mathbf{K}}$. We can express $\overline{\mathbf{K}}$ as

$$\overline{\mathbf{K}} = \sum_{l=1}^{\text{VD}} \overline{\lambda}_l \mathbf{u}_l\mathbf{u}_l^T + \sum_{l=\text{VD}+1}^{L} \overline{\lambda}_l \mathbf{u}_l\mathbf{u}_l^T \tag{17}$$

where $\{\mathbf{u}_l\}_{l=1}^{\text{VD}}$ and $\{\mathbf{u}_l\}_{l=\text{VD}+1}^{L}$ span signal subspace and noise subspace, respectively. The variances of the noise components in (17) have been whitened and normalized to unity, so $\overline{\lambda}_l = 1$ for $l = \text{VD}+1, \ldots, L$, and $\overline{\lambda}_l > 1$ for $l = 1, \ldots, \text{VD}$. Then, the problem of VD estimation can be formulated as the following binary hypothesis testing problem:

$$H_0 : y_l = \overline{\lambda}_l = 1 \quad H_1 : y_l = \overline{\lambda}_l > 1, \qquad \text{for} \quad l = 1, 2, \ldots, L \tag{18}$$

where

$$p_0(y_l) = p(y_l|H_0) \cong N\left(1, \sigma^2_{y_l}\right), \qquad \text{for} \quad l = 1, 2, \ldots, L \tag{19}$$

and

$$p_1(y_l) = p(y_l|H_1) \cong N\left(\mu_l, \sigma^2_{y_l}\right), \qquad \text{for} \quad l = 1, 2, \ldots, L \tag{20}$$

where μ_l is an unknown constant and $\sigma^2_{y_l}$ is given by

$$\sigma^2_{y_l} = \text{Var}[\overline{\lambda}_l] \cong \frac{2\overline{\lambda}_l^2}{N}. \tag{21}$$

Under the hypothesis H_0, (21) can be further reduced to

$$\sigma^2_{y_l} \cong \frac{2}{N}. \tag{22}$$

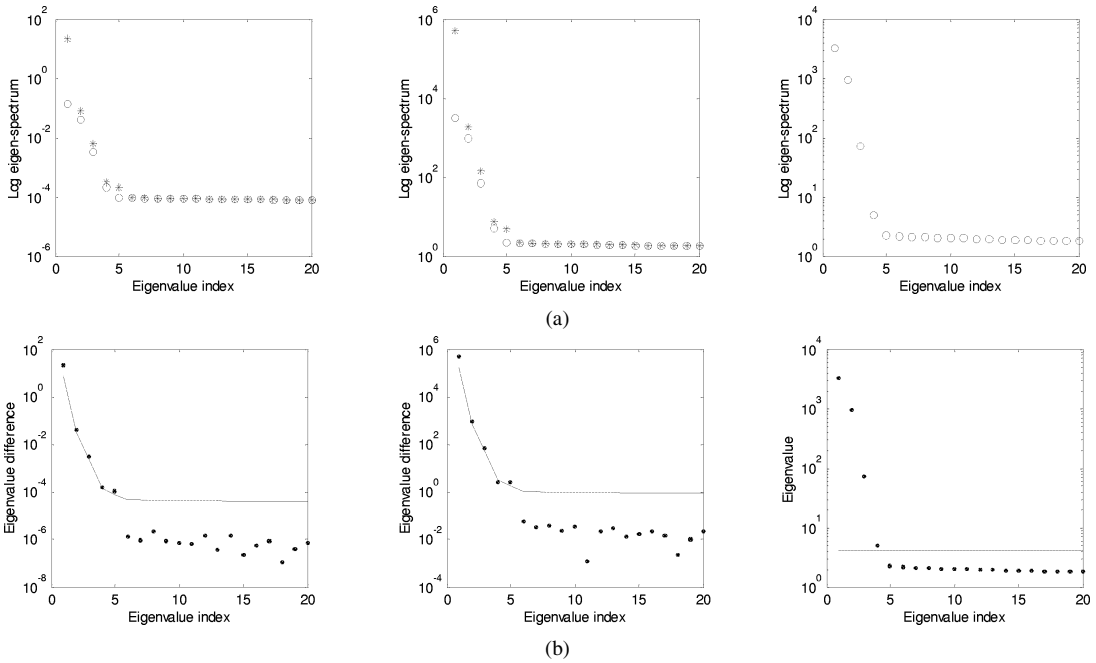


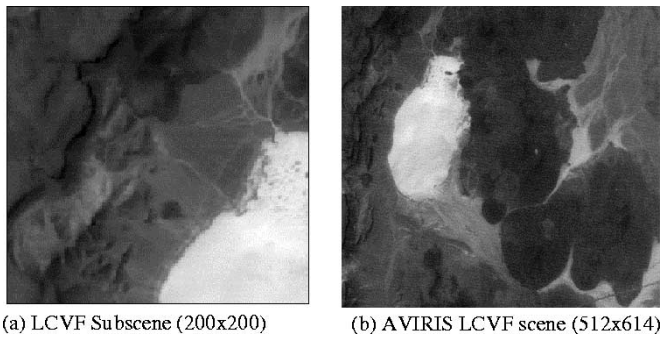
Fig. 2. Computer simulations using five AVIRIS signatures with additive i.i.d. Gaussian noise. (a) Eigenvalue distributions. The asterisk represents correlation eigenvalues. The circles represent covariance eigenvalues. (b) Eigenvalue differences and thresholds. Dots represent eigenvalue differences. Solid line indicates thresholds. (Left) HFC. (Middle) NWHFC. (Right) NSP.

TABLE I
VD ESTIMATES OBTAINED BY THE HFC, THE NWHFC, THE NSP-BASED EIGENTHRESHOLDING METHODS, THE AIC-BASED, THE MDL-BASED, AND THE EIF METHODS FOR COMPUTER SIMULATIONS FOR VARIOUS SNRS

	HFC	NWHFC	NSP	AIC	MDL	EIF
SNR = 10:1	3	3	3	3	3	4
SNR = 15:1	3	3	3	4	3	4
SNR = 20:1	4	3	3	4	3	4
SNR = 25:1	5	4	4	4	4	5
SNR = 30:1	5	4	4	4	4	5
SNR = 35:1	5	4	4	4	4	5
SNR = 40:1	5	4	4	4	4	5

TABLE II
VD ESTIMATES OBTAINED BY THE HFC, THE NWHFC, THE NSP-BASED EIGENTHRESHOLDING METHODS, THE AIC-BASED, THE MDL-BASED, AND THE EIF METHODS FOR TWO EXPERIMENTS OF COMPUTER SIMULATIONS WITH THE SAME SNR $= 25 : 1$

	HFC	NWHFC	NSP	AIC	MDL	EIF
i.i.d. Gaussian noise (1st scenario)	5	4	4	4	4	4
Non-i.i.d. Guassian noise (2nd scenario)	5	4	4	139	110	4



(a) LCVF Subscene (200x200) (b) AVIRIS LCVF scene (512x614)

Fig. 3. AVIRIS image scene.

Now using (19)–(22), we can find the Neyman–Pearson detector δ_{NP} for (18) to determine VD. It is noteworthy that if (22) is used, the threshold will be the same for all the eigenvalues, since (19) is independent of the eigenvalue index.

C. AIC and MDL

In passive sensor array processing, a similar problem to VD estimation also arises from the estimation of the number of signal sources impinging upon the array. This problem is related to how to select an appropriate model for a parameterized family of probability density functions used to best fit the sensor array data. It is known that two commonly used methods, an information-theoretic criterion (AIC) suggested by Akaike [9], and an MDL proposed by Schwartz [10] and Rissanen [11], can be used for model selection. They can be derived from the following formulas that were obtained by Wax and Kailath [21]:

$$\mathrm{AIC}(p) = -2\log\left(\frac{\prod_{j=p+1}^{L}\lambda_j^{\frac{1}{(L-p)}}}{\frac{1}{L-p}\sum_{j=p+1}^{L}\lambda_j}\right)^{(L-p)N} + 2p(2L-p) \tag{23}$$

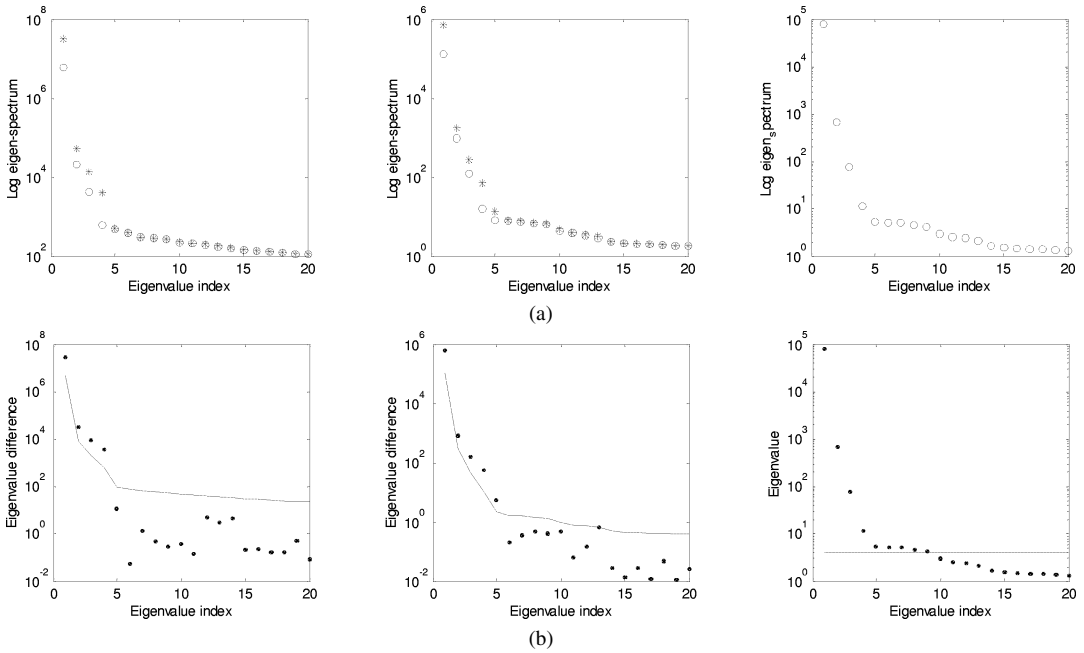


Fig. 4. Results produced for the AVIRIS image scene in Fig. 3. (a) Eigenvalue distributions. The asterisk represents correlation eigenvalues. The circles represent covariance eigenvalues. (b) Eigenvalue differences and thresholds. Dots represent eigenvalue differences. Solid line indicates thresholds. (Left) HFC. (Middle) NWHFC. (Right) NSP.

$$\mathrm{MDL}(p) = -\log\left(\frac{\prod_{j=p+1}^{L}\lambda_j^{\frac{1}{(L-p)}}}{\frac{1}{L-p}\sum_{j=p+1}^{L}\lambda_j}\right)^{(L-p)N} + \left(\frac{1}{2}\right)p(2L-p)\log N \tag{24}$$

where p is the number of free parameters that specifies a family of probability density functions, and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$ are eigenvalues generated by $\mathbf{K}_{L\times L}$ as in (1)–(3).

In [22], a linear unmixing problem could be formulated as a passive sensor array problem, where signal sources impinging on a sensor array were interpreted as desired target sources to be detected from a set of coregistered images acquired by a bank of spectral channels. With this interpretation, a linear mixing problem can be solved by techniques that are used to solve sensor array problems. In order to take advantage of these two criteria and apply them to the estimation of the VD, we need to examine their underlying assumptions, i.e., 1) the noise must be independent identically distributed (i.i.d.), and 2) the observation process is a zero-mean Gaussian random process. When the noise is Gaussian, the first assumption can be taken care of by the whitening technique proposed in Section III, in which case the noise variance will be normalized to unity. Then, the resulting sample covariance matrix can be used in the Gaussian process made in the second assumption. Now, the problem of estimating VD can be solved by minimizing (23) and (24) as follows:

$$\mathrm{VD}_{\mathrm{AIC}} = \arg\left\{\min_{p}\mathrm{AIC}(p)\right\} \tag{25}$$

$$\mathrm{VD}_{\mathrm{MDL}} = \arg\left\{\min_{p}\mathrm{MDL}(p)\right\}. \tag{26}$$

D. Malinowski's Method

In absorption spectroscopy, mass spectra, and chromatography, the determination of the number of factors in a data matrix is also an important issue. Malinowski used the PCA to develop a factor analysis-based indication function to estimate this number, referred to as an empirical indicator function in [5]. Since this problem of determining the number of factors and experimental error in a data matrix is similar to that of estimating the VD, Malinowski's method will be also implemented for comparison. The criterion used by Malinowski's method is the EIF defined by

$$\mathrm{EIF}(q) = \frac{\left(\sum_{l=q+1}^{L}\lambda_l\right)^{\frac{1}{2}}}{N^{\frac{1}{2}}(L-q)^{\frac{3}{2}}}. \tag{27}$$

Using (27), the VD can be estimated via Malinowski's EIF by

$$\mathrm{VD}_{\mathrm{EIF}} = \arg\left\{\min_{q}\mathrm{EIF}(q)\right\}. \tag{28}$$

V. Experiments

The experiments conducted in this section consist of computer simulations and real hyperspectral image experiments. The purpose of these experiments is to demonstrate the ability of the three methods presented in this paper by comparing their performance to AIC, MDL, and Malinowski's EIF method in VD estimation.

A. Computer Simulations

Five AVIRIS reflectance spectra (dry grass, red soil, creosote leaves, blackbrush, and sagebrush), as shown in Fig. 1, were used for simulations, where 158 bands were used after water bands were removed. Each of these five spectral signatures was used to simulate 1000 pixels with abundance generated by random numbers in the range (0,1) for computer simulations. Additive

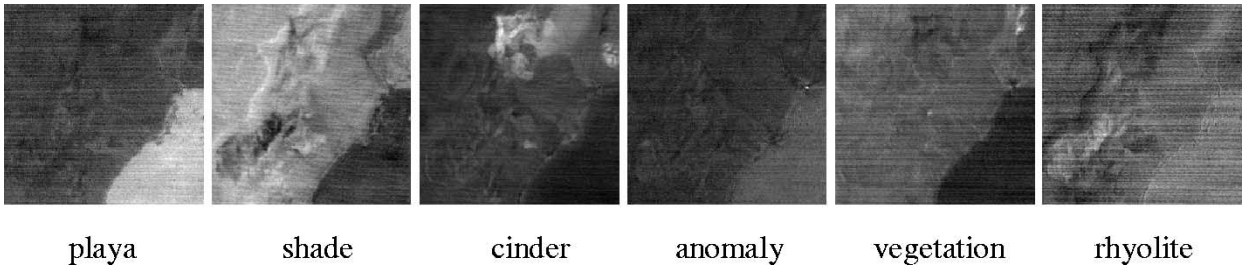


Fig. 5. Unsupervised classification for the AVIRIS image produced by the UNCLS method.

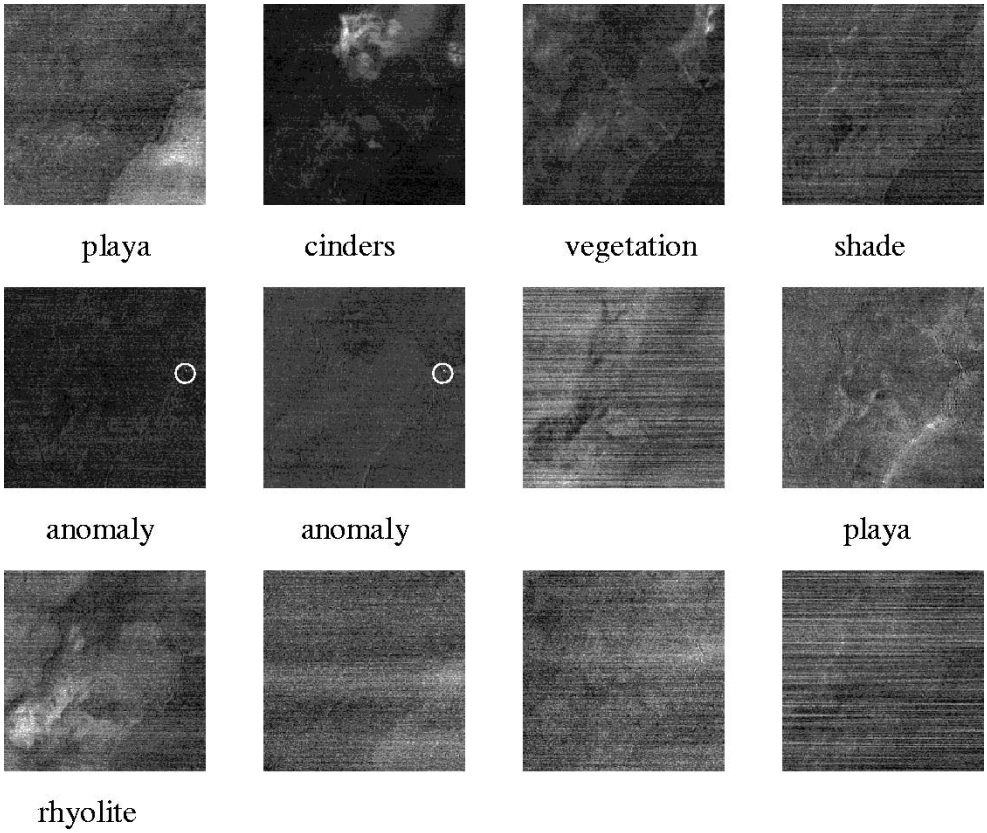


Fig. 6. Unsupervised classification for the AVIRIS image produced by the ATDCA method.

Gaussian noise was added to these 1000 mixed pixels to achieve SNR 25 : 1 for the entire data. In order to simulate an i.i.d. case, the same Gaussian noise with the same variance was added to each band. Fig. 2(a) shows the log eigenspectra generated by the HFC, the NWHFC, and NSP-based eigenthresholding methods, respectively, where the eigenvalues produced by $\mathbf{R}_{L\times L}$ and $\mathbf{K}_{L\times L}$ are denoted by * and °, respectively. The noise estimation method for NWHFC is the residue analysis in Section III because of its simplicity and efficiency. As noted in Fig. 2(a), when the eigenvalue index increases, the correlation eigenvalues start to overlay the covariance eigenvalues. This phenomenon can be explained by the fact that noise is the only signal sources contributed to these eigenvalues. With this interpretation, the VD was estimated by the HFC, the NWHFC, and NSP-based eigenthresholding methods to be 5, 4, and 4, respectively, using the Neyman–Pearson test with the false-alarm probability P_F set to $P_F = 0.001$. In Fig. 2(b), the eigenvalue differences of HFC and NWHFC (denoted by dots) as well as thresholds used for HFC, NWHFC, and NSP-based methods (denoted as curves) were plotted. The HFC eigenthresholding method correctly

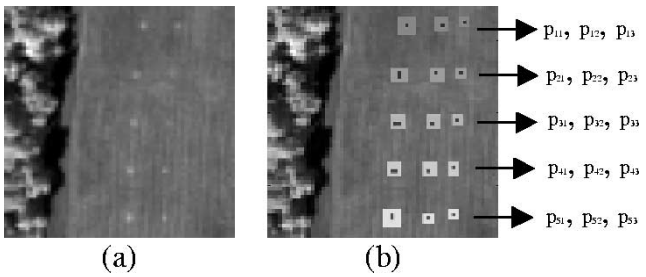


Fig. 7. Fifteen-panel HYDICE image scene.

predicted the true data dimensionality, 5, while both NWHFC and NSP-based eigenthresholding methods predicting 4. Since the spectra of creosote leaves and sagebrush in Fig. 1 are shown in [23] to be very close, they were considered to be the same signature by the NWHFC and NSP methods. This is also true for the AIC and MDL, where the VD was estimated to be 4. The result produced by Malinowski's EIF was 5. Similar experiments were also conducted for various SNRs. Table I tabulates the values of VD estimated by the six methods for various SNRs. It was

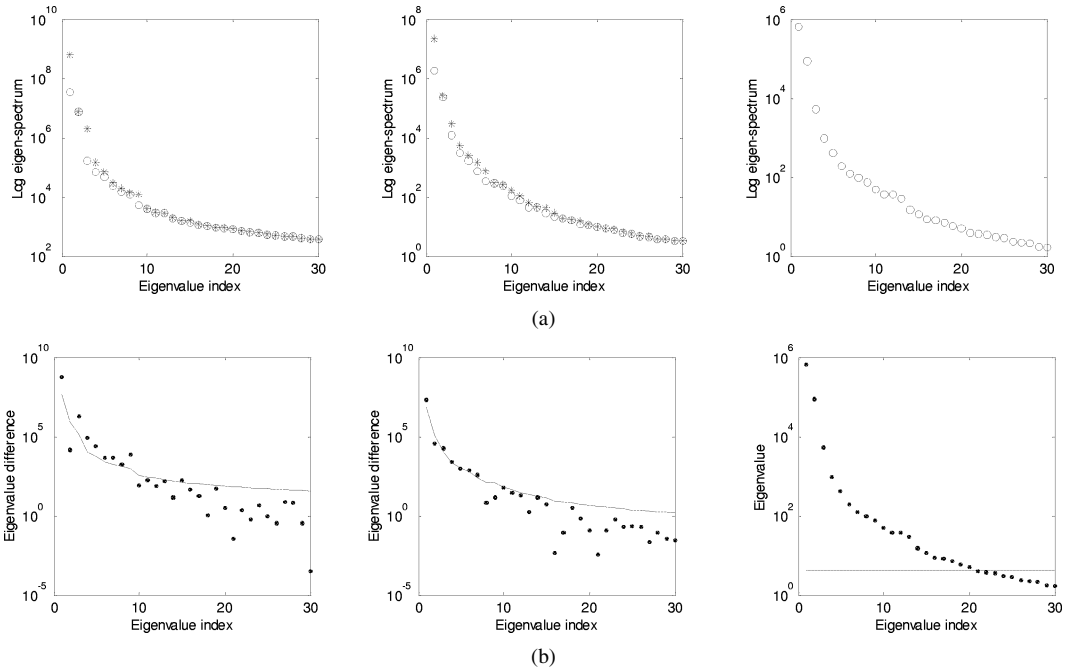


Fig. 8. Results produced for the HYDICE image scene in Fig. 7. (a) Eigenvalue distributions. The asterisk represents correlation eigenvalues. The circles represent covariance eigenvalues. (b) Eigenvalue differences and thresholds. Dots represent eigenvalue differences. Solid line indicates thresholds. (Left) HFC. (Middle) NWHFC. (Right) NSP.

found that the estimated VD for the SNR higher than 25 : 1 by HFC, NWHFC, NSP, AIC, MDL, and Malinowski's EIF stayed the same, 5, 4, 4, 4, 4, 5, respectively. However, when the SNR was lower than 25 : 1, the VD estimated by the HFC, NWHFC, NSP, and MDL was reduced and stayed 3 except 4 estimated by Malinowski's EIF (for SNR $= 10 : 1, 15 : 1, 20 : 1$), by HFC (for SNR $= 20 : 1$), and by AIC (for SNR $= 15 : 1, 20 : 1$). These experiments made sense. Due to increased noise interference, the blackbrush becomes indiscernible from creosote leaves and sagebrush, as shown in [23]. In this case, the three signatures will be considered to belong to one class, and this class along with dry grass and red soil makes up three distinct classes in computer simulations.

Unlike the above simulation with an equal noise variance added to each individual band, we simulated another scenario in such a way that additive Gaussian noise with a different variance was added to each individual band to achieve the band SNR 25 : 1. In this case, the noise is not i.i.d., since the variance is varied from band to band. Interestingly, the values of VD estimated by the three eigenthresholding based methods and Malinowski's EIF method still remained unchanged, but the AIC and MDL completely failed, in which their estimated VD were 139 and 110, respectively.

The second scenario demonstrates that the i.i.d. assumption is crucial to the success of the AIC and MDL, while it has very little effect on the eigenthresholding based methods. Table II summarizes the VD results estimated by the HFC, NWHFC, NSP, AIC, MDL, and EIF for the two scenarios of simulated non-i.i.d. Gaussian noise with SNR $= 25 : 1$.

B. AVIRIS and HYDICE Experiments

The 224-band AVIRIS image shown in Fig. 3 was used in the experiments. This was also the same image considered in [24]. Fig. 3(a) is a subscene of 200 $\times$ 200 pixels cropped from the left upper corner of the Lunar Crater Volcanic Field in Northern Nye County, NV in Fig. 3(b), where five signatures of interest in these images demonstrated in [24] were "red oxidized basaltic cinders," "rhyolite," "playa (dry lakebed)," "shade," and "vegetation." The log eigenspectra generated by the HFC, NWHFC, and NSP-based eigenthresholding methods are shown in Fig. 4(a), where * and $\circ$ denoted the eigenvalues produced by $\mathbf{R}_{L\times L}$ and $\mathbf{K}_{L\times L}$, respectively. The VD estimated by the HFC, NWHFC and NSP-based eigenthresholding methods was 4, 6, and 9, respectively, using the Neyman–Pearson test with the false-alarm probability P_F set to $P_F = 0.001$, as shown in Fig. 4(b). In order to see if these estimated values are appropriate, we applied the unsupervised nonnegatively constrained least squares (UNCLS) method proposed in [25] to classify the AVIRIS image in Fig. 3. It should be noted that since no prior knowledge was assumed, the target detection and classification must be conducted in an unsupervised manner. Fig. 5 shows the unsupervised classification results produced by the UNCLS method, where at least five target signatures (i.e., cinders, playa, rhyolite, shade, and vegetation) in the image scene were identified plus an anomaly detected at the upper edge of the dry lake. This implies that the VD must be equal to or greater than 6. As a matter of fact, due to a wide coverage of the dry lake, the playa signature varied significantly from its top to bottom. So, if we further applied the automatic target detection and classification (ATDCA) proposed in [26] to the AVIRIS scene, it required nine target signatures to separate the rhyolite from cinders. It was because the dry lake was split into two regions, as noted in Fig. 6. This is also close to the result produced by the ICA-based approach in [27], which required at least eight components to discriminate these five materials. These mean that the results using NWHFC and NSP are more reasonable. When the AIC and the MDL were applied to the AVIRIS image scene, the estimated VD was 157 for both of

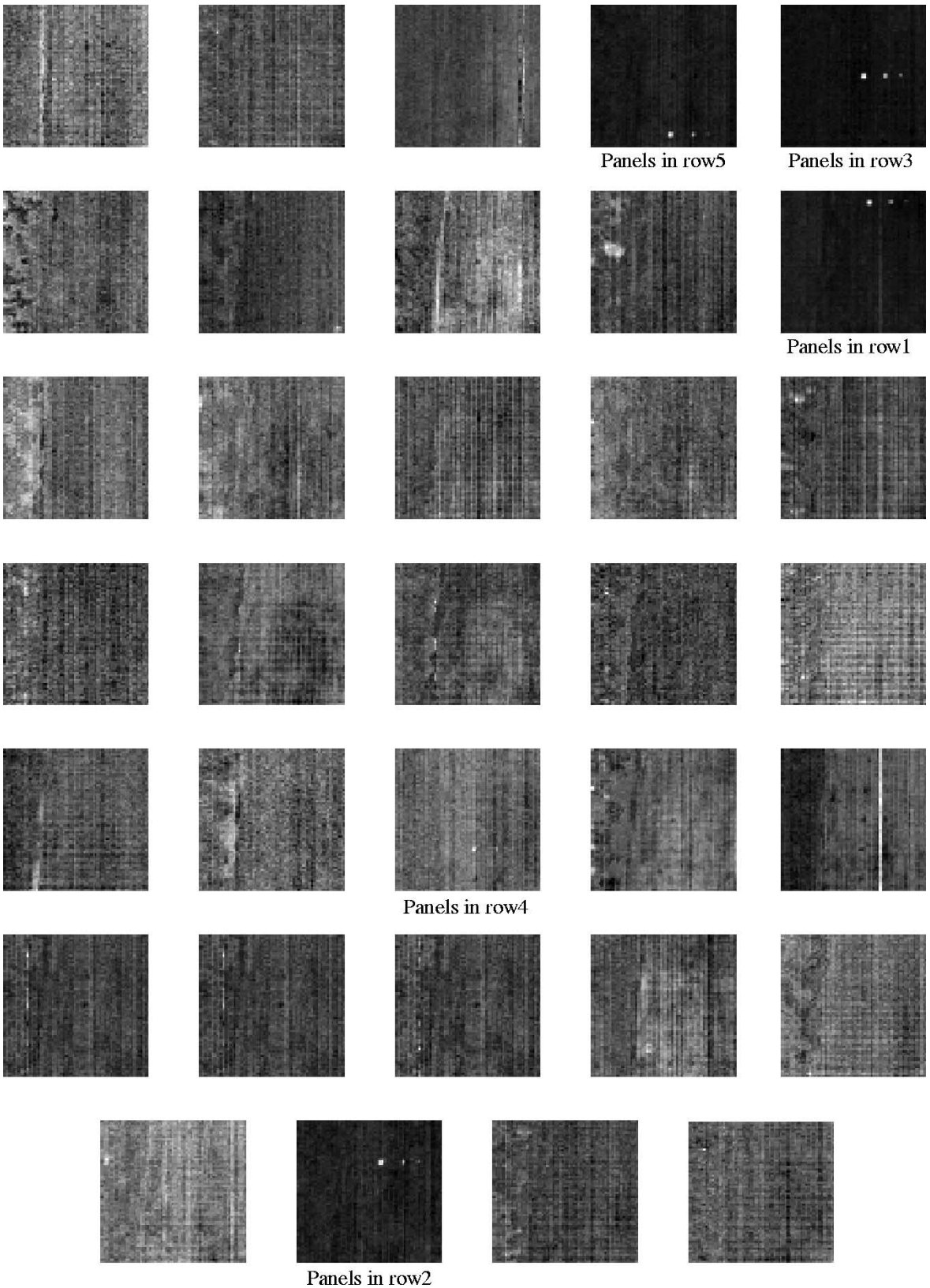


Fig. 9. Unsupervised classification for the HYDICE image produced by the UNCLS method.

them. After the noise was whitened, the estimated VD was 150 for AIC and 121 for MDL. These values are significantly higher than the numbers estimated by the three eigenthresholding based methods. As illustrated in computer simulations, due to the fact that the noise in remotely sensed imagery is generally not i.i.d. as usually assumed in array processing, the VDs estimated by the AIC and the MDL for the AVIRIS image were largely affected by this assumption. Consequently, the AIC and MDL performed poorly in the estimation of VD. The VD estimated from Malinowski's EIF was 15, and 10 for

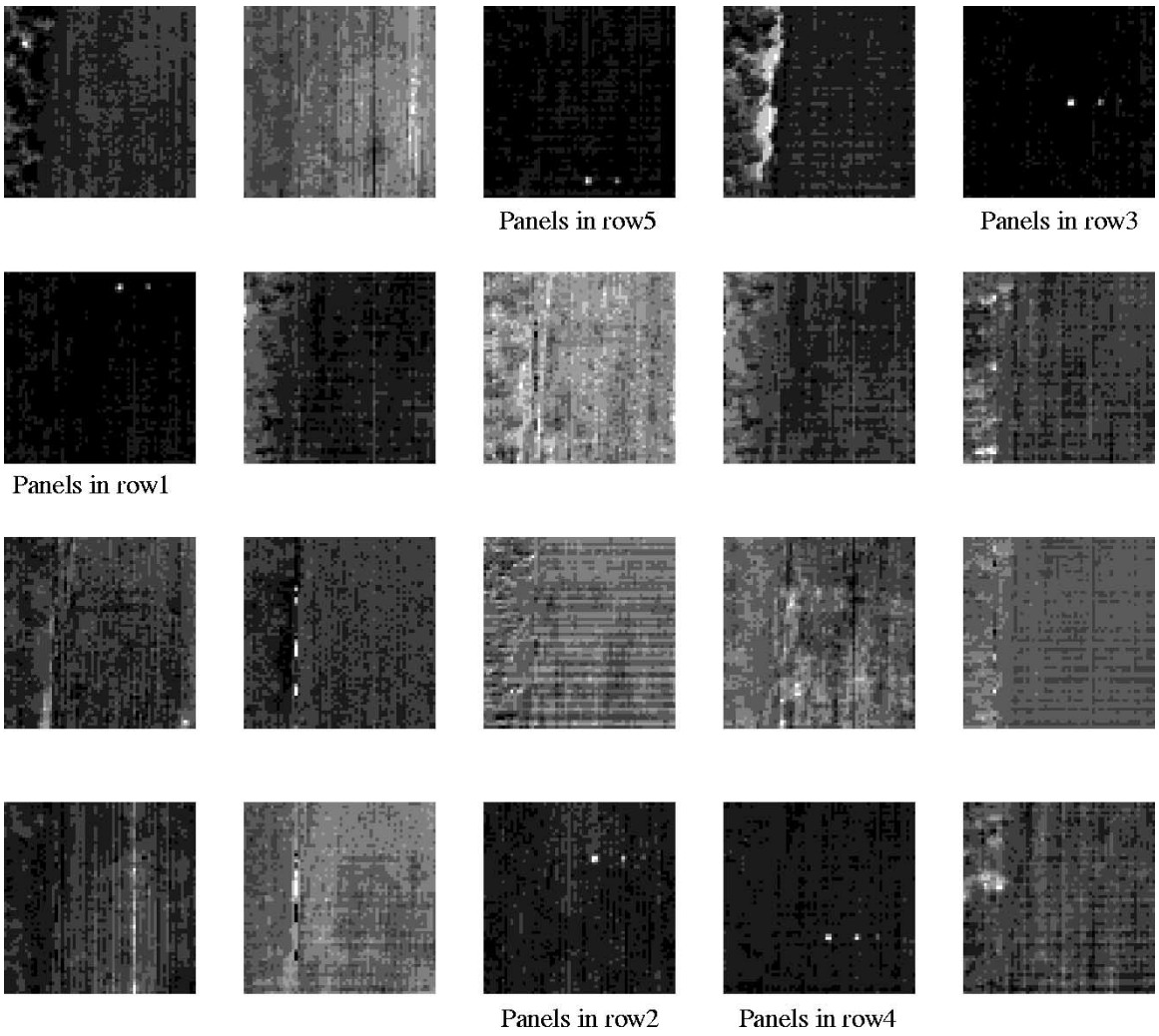


Fig. 10. Unsupervised classification for the HYDICE image results by the ATDCA.

the noise-whitened data, which were also large. According to experiments conducted in [5], Malinowski's EIF worked well when the noise is random and fairly uniform. But this may not be true in real data, where the noise may vary from band to band and even be correlated among bands. As a result, Malinowski's EIF did not perform well as did in the computer simulation.

Similar experiments were also conducted for the 15-panel HYDICE image using the HFC, the NWHFC, and the NSP-based eigenthresholding methods. The image scene shown in Fig. 7(a) has 15 panels located on the field and arranged in a 5×3 matrix. The low-signal/high-noise bands: bands 1–3 and bands 202–210, and water vapor absorption bands: bands 101–112 and bands 137–153, were removed. So, a total of 169 bands were used for the experiments. The spatial resolution is 1.5 m, and spectral resolution is 10 nm. The ground truth of the image scene is shown in Fig. 7(b) and provides the precise spatial locations of these 15 panels. Black pixels are panel center pixels and the pixels in the white masks are panel boundary pixels mixed with background pixels. Each element in this matrix is a square panel and denoted by p_{ij} with row indexed by $i = 1, 2, \ldots, 5$ and column indexed by $j = 1, 2, 3$. For each row i, the three panels p_{i1}, p_{i2}, p_{i3} were made from the same material but have three different sizes. For each column j, the five panels p_{1j}, p_{2j}, p_{3j}, p_{4j}, p_{5j} have the same size but were made by five different materials. The sizes of the panels in the first, second, and third columns are $3\,m \times 3\,m$, $2\,m \times 2\,m$, and $1\,m \times 1\,m$, respectively. The 1.5-m spatial resolution of the image scene suggests that except for p_{1l}, p_{2l}, p_{3l}, p_{4l}, p_{5l}, which are two-pixel panels, all the remaining panels are only one-pixel in size. Apparently, without ground truth it is difficult to locate these panels in the scene. Fig. 8(a) shows the distributions of the log eigenvalues produced by $\mathbf{R}_{L\times L}$ and $\mathbf{K}_{L\times L}$. Once again, the VD estimated by the HFC, NWHFC, and NSP-based eigenthresholding methods using the Neyman–Pearson test with the false-alarm probability P_F set to $P_F = 0.001$ was 9, 18, and 20, as shown in Fig. 8(b) (only the first 30 eigenvalues and their differences were plotted for a clear presentation, and several signals were detected in the eigenvalues higher than 30 in the case of NWHFC). If the UNCLS method was applied to this 15-panel scene, 32 target signatures were required to generate for the UNCLS method to effectively detect and classify these 15 panels into five separate target classes, as shown in Fig. 9. The results in Fig. 9 may lead to a conclusion that a good estimate of the VD should be close to 32. Unfortunately, this is not true. If we applied the ATDCA to the same HYDICE scene in Fig. 7(a), only 19 target signatures were required for the ATDCA to effectively detect and classify these 15 panels into five separate target classes, as shown in Fig. 10. This is further confirmed by the results produced by the ICA-based approach in [27], which

TABLE III
VD ESTIMATES OBTAINED BY THE HFC, THE NWHFC, THE NSP-BASED EIGENTHRESHOLDING METHODS, THE AIC-BASED, THE MDL-BASED, AND THE EIF METHODS FOR AVIRIS AND HYDICE IMAGES

	HFC	NWHFC	NSP	AIC	MDL	EIF
AVIRIS image	4	—	—	157	157	15
AVIRIS image (noise whitened)	—	6	9	150	121	10
HYDICE image	9	—	—	167	128	94
HYDICE image (noise whitened)	—	18	20	124	85	28

TABLE IV
VD ESTIMATES FOR AVIRIS DATA USING DIFFERENT P_F

	HFC	NWHFC	NSP
$P_F = 10^{-5}$	4	4	5
$P_F = 10^{-4}$	4	4	7
$P_F = 10^{-3}$	4	6	9
$P_F = 10^{-2}$	4	9	9
$P_F = 10^{-1}$	4	15	12

TABLE V
VD ESTIMATES FOR HYDICE DATA USING DIFFERENT P_F

	HFC	NWHFC	NSP
$P_F = 10^{-5}$	7	10	19
$P_F = 10^{-4}$	9	14	20
$P_F = 10^{-3}$	9	18	20
$P_F = 10^{-2}$	11	27	23
$P_F = 10^{-1}$	14	37	26

required at least 18 components to discriminate the 15 panels. Therefore, $\mathrm{VD} = 19$ should be a good estimate. According to the ground truth, the panel signatures in rows 2 and 3 are very similar, so are the panel signatures in rows 4 and 5. In order for the UNCLS to effectively separate them, it would require more subtle target signatures for discrimination. As a consequence, some of natural or background signatures belonging to the same class had been forced to split into separate classes.

Finally, the AIC, the MDL, and Malinowski's EIF were also applied to the 15-panel scene, the estimated VD were 167, 128, and 94, respectively. After the noise was whitened, they were down to 124, 85, and 28, respectively. Once again, they overestimated the VD. This experiment further supports our conclusion that the AIC, the MDL, and Malinowski's EIF may not be appropriate methods for the VD estimation for remotely sensed images. Table III summaries the VD estimation results obtained by the HFC, the NWHFC, the NSP-based eigenthresholding methods, the AIC, the MDL, and Malinowski's EIF for AVIRIS and HYDICE images. Tables IV and V list the VD estimates under different P_F. We can see that the VD estimate is increased as P_F becomes large. HFC is less sensitive to the P_F, while NWHFC is most sensitive to the P_F. In general, $P_F = 0.001$ is a good choice in these two experiments.

If PCA or noise-adjusted PCA is applied to VD estimation, we can select the first several eigenvalues that are significantly larger than others. In our conducted AVIRIS experiment, this number could be 3 or 4 according to the eigenvalue distributions in Fig. 4(a). However, this practice may not work for our HYDICE experiments, since there is no drastic drop in the eigenvalue distributions in Fig. 8(a) that can help to determine the number of eigenvalues to be used. For the noise-adjusted PCA, the same problem remains. The HYDICE experiment provides a good example to demonstrate the problem of using the PCA or noise-adjusted PCA in VD estimation.

VI. SUMMARY AND CONCLUSION

The determination of VD is a difficult and challenging problem. A common approach is to use PCA-based techniques. Unfortunately, the major difficulty lying in such techniques is that the eigenvalues are still mixtures of signal sources present in data and noise. So, a problem occurs when signal sources are weak and have very little energy contributed to eigenvalues, in particular, when there is no drastic drop in the eigenvalue distribution. This paper presents three Neyman–Pearson detection theory-based eigenthresholding methods to resolve such a problem, which model the VD estimation as a binary composite hypothesis testing problem and the VD estimation error can be measured by ROC analysis. The first approach, referred to as HFC method, was developed by Harsanyi, Farrand and Chang [9]. This was later modified and extended to the NWHFC method that includes a whitening process. A third approach, called the noise subspace projection method, takes advantage of the NWHFC method to derive an NSP-based eigenthresholding method. All the three methods develop a Neyman–Pearson detector to estimate the number of signal sources in terms of their energies. On some occasions, weak signal sources with small energies may not contribute significantly to eigenvalues, in which case the false-alarm probability must be set high to be able to detect such signal sources. The ROC analysis resulting from the Neyman–Pearson detection theory provides an effective tool to measure the trade-off between the detection power (i.e., VD estimation error) and false-alarm probability.

The AIC and MDL have been widely used in the passive sensor array processing to estimate the number of signals impinging on the array. In order to determine their utility in estimation of VD for remotely sensed images, they were also included for comparison. Additionally, Malinowski's EIF has been also used in mass spectroscopy to estimate controlling factors in a data matrix. So, its applicability to VD estimation was also studied. As shown in our computer simulations, HFC, NWHFC, NSP, AIC, MDL, and Malinowski's EIF worked effectively when the noise is i.i.d.. Interestingly, this is no longer true when they are applied to real image data. Our experimental results demonstrated that all the three eigenthresholding based methods, the HFC, NWHFC, and NSP methods, produced similar values of VD in contrast to the AIC and MDL which overestimated VD significantly. As for Malinowski's EIF method, its performance was improved by data whitening, but it still overestimated the VD. This is due to the fact that the noise in remote sensing images is generally not i.i.d., which is a crucial condition for the AIC and MDL criteria and Malinowski's EIF method.

As a concluding remark, the methods proposed in this paper provide only an estimate of the number of signal sources in image data that are assumed to be spectrally distinct. Therefore, in most cases, these signal sources may include unknown interferers such as clutters, background signatures, etc. As a result, the VD may overestimate the true number of image endmembers, i.e., intrinsic dimensionality. In real applications, such a VD estimate can be used as a reference rather than the exact value of the intrinsic dimensionality.

ACKNOWLEDGMENT

The authors would like to thank the support received from the Office of Naval Research. In addition, the first author would like to thank the National Research Council Senior Research Associateship sponsored by the U.S. Army Soldier and Biological Command, Edgewood Chemical Biological Center (ECBC), Aberdeen Proving Ground, MD, as well as support from TRW Foundation.

REFERENCES

[1] K. Fukunaga, "Intrinsic dimensionality extraction," in *Classification, Pattern Recognition and Reduction of Dimensionality*, P. R. Krishnaiah and L. N. Kanal, Eds. Amsterdam, The Netherlands: North-Holland, 1982, vol. 2, Handbook of Statistics, pp. 347–360.

[2] ——, *Statistical Pattern Recognition*, 2nd ed. New York: Academic, 1992.

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

[4] E. R. Malinowski, "Theory of error in factor analysis," *Anal. Chem.*, vol. 49, pp. 606–612, 1977.

[5] ——, "Determination of the number of factors and experimental error in a data matrix," *Anal. Chem.*, vol. 49, pp. 612–617, 1977.

[6] 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.

[7] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*. New York: Academic, 1997.

[8] 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, Mar. 1998.

[9] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Control*, vol. AC-19, pp. 716–723, 1974.

[10] G. Schwarz, "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, pp. 461–464, 1978.

[11] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[12] J. Harsanyi, W. Farrand, and C.-I Chang, "Determining the number and identity of spectral endmembers: An integrated approach using Neyman–Pearson eigenthresholding and iterative constrained RMS error minimization," in *Proc. 9th Thematic Conf. Geologic Remote Sensing*, Feb. 1993.

[13] H. V. Poor, *An Introduction to Detection and Estimation Theory*, 2nd ed. New York: Springer-Verlag, 1994.

[14] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 2nd ed. New York: Wiley, 1984.

[15] C.-I Chang and Q. Du, "A noise subspace projection approach to determination of intrinsic dimensionality for hyperspectral imagery," in *Proc. EOS/SPIE Symp. Remote Sensing, Conf. Image and Signal Processing for Remote Sensing V*, vol. 3871, Florence, Italy, Sept. 20–24, 1999, pp. 34–44.

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

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

[18] Q. Du and C.-I Chang, "An interference rejection-based radial basis function neural network approach to hyperspectral image classification," in *Proc. Int. Joint Conf. Neural Networks*, Washington, DC, July 1999, pp. 2698–2703.

[19] 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.

[20] 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.

[21] M. Wax and T. Kailath, "Detection of signals by information criteria," *IEEE Trans. Acoustic, Speech, and Signal Process.*, vol. ASSP-33, pp. 387–392, Apr. 1985.

[22] J. C. Harsanyi, "Detection and classification of subpixel spectral signatures in hyperspectral image sequences," Dept. Elect. Eng., Univ. Maryland Baltimore County, Baltimore, M, Aug. 1993.

[23] C.-I Chang, "An information theoretic-based approach to spectral variability, similarity and discriminability for hyperspectral image analysis," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1927–1932, Aug. 2000.

[24] 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.

[25] C.-I Chang and D. Heinz, "Subpixel spectral detection for remotely sensed images," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, pp. 1144–1159, May 2000.

[26] H. Ren and C.-I Chang, "Automatic spectral target recognition in hyperspectral imagery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, pp. 1232–1249, Oct. 2003.

[27] C.-I Chang, S.-S. Chiang, J. A. Smith, and I. W. Ginsberg, "Linear spectral random mixture analysis for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sensing*, pp. 375–392, Feb. 2002.

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

He has been with the University of Maryland Baltimore County (UMBC), Baltimore, since 1987, as a Visiting Assistant Professor from January 1987 to August 1987, Assistant Professor from 1987 to 1993, Associate Professor from 1993 to 2001, and Professor in the Department of Computer Science and Electrical Engineering since 2001. He was a Visiting Research Specialist in the Institute of Information Engineering at the National Cheng Kung University, Tainan, Taiwan, from 1994 to 1995. He has a patent on automatic pattern recognition and several pending patents on image processing techniques for hyperspectral imaging and detection of microcalcifications. His research interests include automatic target recognition, multispectral/hyperspectral image processing, medical imaging, information theory and coding, signal detection and estimation, and neural networks. He is the author of a book *Hyperspectral Imaging: Techniques for Spectral Detection and Classification* (Norwell, MA: Kluwer). He is on the editorial board and was the Guest Editor of a special issue on telemedicine and applications of the *Journal of High Speed Networks*.

Dr. Chang received a National Research Council Senior Research Associateship Award from 2002–2003 at the U.S. Army Soldier and Biological Chemical Command, Edgewood Chemical and Biological Center, Aberdeen Proving Ground, MD. He is an Associate Editor in the area of hyperspectral signal processing for the IEEE TRANSACTION ON GEOSCIENCE AND REMOTE SENSING. He is a Fellow of SPIE and a member of Phi Kappa Phi and Eta Kappa Nu.

Qian Du (M'00) received the Ph. D degree in electrical engineering from the University of Maryland Baltimore County, Baltimore, in 2000.

She is currently an Assistant Professor in the Department of Electrical Engineering and Computer Science, Texas A&M University, Kingsville. Her research includes image processing, pattern classification, remote sensing, and neural networks.

Dr. Du is a member of SPIE and the honor society of Phi Kappa Phi.