Linear Spectral Random Mixture Analysis for Hyperspectral Imagery

Chein-I Chang, *Senior Member, IEEE*, Shao-Shan Chiang, *Member, IEEE*, James A. Smith, *Fellow, IEEE*, and
Irving W. Ginsberg

Abstract—Independent component analysis (ICA) has shown success in blind source separation and channel equalization. Its applications to remotely sensed images have been investigated in recent years. Linear spectral mixture analysis (LSMA) has been widely used for subpixel detection and mixed pixel classification. It models an image pixel as a linear mixture of materials present in an image where the material abundance fractions are assumed to be *unknown* and *nonrandom parameters*. This paper considers an application of ICA to the LSMA, referred to as ICA-based linear spectral random mixture analysis (LSRMA), which describes an image pixel as a random source resulting from a random composition of multiple spectral signatures of distinct materials in the image. It differs from the LSMA in that the abundance fractions of the material spectral signatures in the LSRMA are now considered to be *unknown* but *random independent signal sources*. Two major advantages result from the LSRMA. First, it does not require prior knowledge of the materials to be used in the linear mixture model, as required for the LSMA. Second, and most importantly, the LSRMA models the abundance fraction of each material spectral signature as an independent random signal source so that the spectral variability of materials can be described by their corresponding abundance fractions and captured more effectively in a stochastic manner. The experimental results demonstrate that the proposed LSRMA provides an effective unsupervised technique for target detection and image classification in hyperspectral imagery.

Index Terms—Hyperspectral image classification, independent component analysis (ICA), linear spectral mixture analysis (LSMA), linear spectral random mixture analysis (LSRMA).

I. INTRODUCTION

OVER the past years, linear spectral mixture analysis (LSMA) has been widely used for hyperspectral image analysis such as detection and classification [1], [2]. It assumes that an image pixel [i.e., digital number (DN)] is linearly mixed by materials with relative abundance fractions present in the image. To be more specific, materials will be referred to as targets in this paper. Two restrictions are generally applied to

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.

S.-S. Chiang is with the Department of Electrical Engineering, Lunghua University of Science and Technology, Taoyuan, Taiwan, R.O.C.

J. A. Smith is with the Laboratory for Terrestrial Physics, NASA Goddard Space Flight Center, Greenbelt, MD 20771 USA.

I. W. Ginsberg was with the the Remote Sensing Laboratory, Department of Energy, Las Vegas, NV 89191 USA. He is now with Equinox Scientific Research, Farmington Hill, MI 48336 (e-mail: equinoxsr@aol.com).

the LSMA. One is that a complete target knowledge must be given *a priori*. In many practical applications, obtaining such *a priori* information is usually very difficult if not impossible. To relax this requirement, several unsupervised methods were proposed to generate target information directly from the image data, e.g., unsupervised vector quantization [3], target generation process [4], and least squares error [5]. A second restriction is that the abundance fractions of targets present in an image are generally unknown and need to be estimated. When they are considered as unknown and nonrandom (i.e., deterministic) quantities, they can be estimated by methods such as least squares estimation. On the other hand, it may be more realistic to assume that the abundance fractions of targets in a pixel are random quantities rather than deterministic quantities so that targets can be described by their corresponding random abundance fractions, which can capture their spectral variability more effectively in a stochastic manner. In order to appropriately represent such a random linear mixture, the abundance fraction of each target must be considered as a random signal source. For example, if the abundance fraction α is modeled as a random parameter in the range [0,1], it means that α can occur anywhere in [0,1] with a certain probability. By contrast, if α is a nonrandom parameter in [0,1], it implies that α must be a number somewhere in [0,1] with probability one. In other words, a nonrandom parameter is merely a number opposed to a random parameter which must be defined in a probability space.

Independent component analysis (ICA) [6]–[12], which has shown great success in blind source separation and other applications, provides a feasible approach to solving the random abundance mixture problem described above. Its application to remotely sensed imagery has been recently investigated in [13]–[17]. The ICA is an unsupervised source separation process. It is different from principal components analysis (PCA) [18], [19] in many aspects. The PCA decorrelates the sample data covariance matrix in such a manner that the data set can be decomposed into a set of uncorrelated and orthogonal components where each component is oriented by an eigenvector. Unlike the PCA, the ICA looks for components which are *statistically independent* rather than *uncorrelated*; thus, it requires statistics of orders higher than the second order. The idea of the ICA makes use of a linear model to describe a mixture of a set of unknown random signal sources, then demixes them in separate components so as to achieve signal detection and classification. If we assume that the abundance fraction of each target in the LSMA is an unknown and independent random signal source, the source mixing model

considered in ICA becomes directly applied to the LSMA, in which case ICA can be used to solve for random abundance fractions for the linear mixture model used in the LSMA. Such an ICA-based LSMA can be viewed as a random version of the LSMA and will be called linear spectral random mixture analysis (LSRMA) hereafter in this paper.

In order for the LSRMA to be effective, two major assumptions must be made. One is that the source components must mutually statistically independent. This implies that the spectral signatures of targets present in an image must be distinct. A second assumption is that, at most, one source component is allowed to be Gaussian. This is because a sum of Gaussian processes is also Gaussian and the ICA cannot separate Gaussian processes using a linear mixture model.

In remotely sensed imagery, the number of target pixels of interest, such as small man-made targets, anomalies, or rare minerals, is generally small compared to the image background. From this point of view, an interesting structure of an image scene is the one resulting from a small number of target pixels in a large area of unknown background. As a consequence, these target pixels are main causes of outliers of distributions which can be detected by a higher order of statistics such as skewness (third moment) to detect asymmetry of the distribution and kurtosis (fourth comment) to detect the flatness of the distribution, as demonstrated in [20]. Therefore, detecting such small target pixels in an unknown image scene can be reduced to finding the outliers or deviations from the background distribution, in which case the background can be considered a homogeneous region. In addition, due to spectral variability, the background may be made up of a number of homogenous regions. If these regions contain large number of pixels, the background pixels can be assumed to be a Gaussian-like distribution while the target pixels of interest can be viewed as non-Gaussian signal sources that create ripples in the Gaussian tails. In this case, target pixels of interest can be separated by the ICA, as we desire. This is because the ICA can detect target pixels using statistics of order higher than the statistics of second-order such as variance. Of course, if there is a set of background pixels that forms a region failing to satisfy the Gaussian assumption, they will be detected as non-Gaussian signal sources. This phenomenon is demonstrated in the Hyperspectral Digital Image Collection Experiment (HYDICE) experiments. Since small targets are generally susceptible to outliers, they will be more likely to be detected by the ICA. However, due to no prior knowledge, the detected small targets may include man-made targets or natural objects such as trees, grass, rocks, and interferers. The LSRMA does not guarantee the detected targets will be the targets of our interest. This must be done in conjunction with spectral database or verified by ground truth.

The ICA proposed in the LSRMA is slightly different from the commonly used ICA [6]–[16] in two ways. In general, the separating matrix (also referred to as unmixing matrix) $\mathbf{W}$ derived from the ICA is assumed to be a square matrix of full rank. In this case, the number of signal sources, say p, must be equal to data dimensionality, L. As demonstrated in our experiments, this assumption may not be valid for hyperspectral images where the number of signal sources is generally much smaller than the number of bands, i.e., $p < L$. Therefore, the

matrix $\mathbf{W}$ is not of full rank. As a result, the learning rule derived from inverting $\mathbf{W}$ may become unstable and may not converge. The problem of this type is called under-complete ICA and has been investigated in [21]. However, this problem can be resolved by prewhitening the data. Second, our proposed ICA approach is based on the constraint that the covariance matrix of unmixed abundance fractions of targets must be an identity matrix. This advantage allows us to design a learning algorithm to converge to independent components which can separate spectrally similar targets. A similar approach was also investigated in [22]. The experiments show that such a learning algorithm is very useful and suitable for target extraction with similar spectral signatures.

The remainder of this paper is organized as follows. Section II describes the concept of the ICA. Section III develops an ICA approach to hyperspectral image analysis, LSRMA. Section IV presents experimental results. Section V suggests an automatic thresholding method for target detection and conducts a quantitative study for the LSRMA in comparison with two commonly used linear unmixing methods: 1) orthogonal subspace projection (OSP) in [23] and 2) constrained energy minimization (CEM) in [24]. Section VI includes some concluding remarks.

II. INDEPENDENT COMPONENT ANALYSIS (ICA)

LSMA is a widely used approach to determination and quantification of multicomponents in remotely sensed imagery. Since every pixel is acquired by spectral bands at different wavelengths, it can be represented by a column vector and a hyperspectral image is actually an image cube. Suppose that L is the number of spectral bands. Let $\mathbf{r}$ be an $L \times 1$ column pixel vector in a multispectral or hyperspectral image where the boldface is used for vectors. Let $\mathbf{M}$ be an $L \times p$ target signature matrix, denoted by $[\mathbf{m}_1 \ \mathbf{m}_2 \ \ldots \ \mathbf{m}_p]$ where $\mathbf{m}_j$ is an $L \times 1$ column vector represented by the jth target signature and p is the total number of targets in the image. Let $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_p)^T$ be a $p \times 1$ abundance column vector associated with $\mathbf{r}$, where α_j denotes the abundance fraction of the jth target signature present in the pixel vector $\mathbf{r}$. A classical approach to solving such a mixed pixel classification problem is the linear spectral unmixing which assumes that a pixel vector $\mathbf{r}$ is linear mixed by the p targets with an unknown abundance vector $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_p)^T$. In this case, the pixel vector $\mathbf{r}$ can be represented by a linear regression model as follows:

$$\mathbf{r} = \mathbf{M}\alpha + \mathbf{n} \qquad (1)$$

where $\mathbf{n}$ is noise that can be interpreted as measurement error, noise, or model error. Here, without confusion, the $\mathbf{r}$ will be used to represent either the pixel vector $\mathbf{r}$ or its spectral signature (i.e., DNs). A spectral linear unmixing method estimates the unknown abundance fractions $\alpha_1, \alpha_2, \ldots, \alpha_p$ via an inverse of a linear mixture model described in (1). One requirement of the LSMA is that the target signature matrix $\mathbf{M}$ must be known *a priori*. Many approaches have been proposed in the past to obtain $\mathbf{M}$ directly from the image data in an unsupervised fashion, such as [3]–[5]. Here, we present a rather different approach—ICA. It is also based on model (1), but

does not require the prior knowledge of $\mathbf{M}$. Most distinctively, it assumes that the p abundance fractions $\alpha_1, \alpha_2, \ldots, \alpha_p$ are unknown random quantities specified by random signal sources rather than unknown deterministic quantities, as assumed in model (1). However, in this case, we need to make the following three additional assumptions on the random abundance vector $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_p)^T$.

1) The p target signatures $\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_p$ in $\mathbf{M}$ must be spectrally distinct.
2) The p abundance fractions $\alpha_1, \alpha_2, \ldots, \alpha_p$ are mutually statistically independent random sources.
3) Each of the p abundance fractions $\alpha_1, \alpha_2, \ldots, \alpha_p$ must be a zero-mean random source and at most one source is Gaussian.

Except for these three assumptions, no prior knowledge is assumed about the model (1).

From a viewpoint of remotely sensed imagery, the first assumption simply says that there are p distinct types of targets in a scene and each column vector in $\mathbf{M}$ represents the spectral signature of one target in the scene. The second assumption implies that the abundance compositions of p distinct target signatures in a pixel $\mathbf{r}$ are random quantities, one independent of another. The third assumption suggests that the ICA can separate targets of interest from the Gaussian-like background distributions. It classifies different targets by detecting them in separate independent components. Since we are only interested in detecting non-Gaussian signal sources, which are generally described by statistics of order higher than the variance, a prewhitening process can make p random abundance fractions $\alpha_1, \alpha_2, \ldots, \alpha_p$ zero-mean and unit-variance, in which case the first- and second-order of statistics will not play a role in target detection. It should be noted that if a background distribution is not Gaussian distributed, it will be detected and classified into a separate independent component. This may occur when some part of the image background is made up of only a small group of pixels.

III. ICA-BASED LSRMA

In order to implement the ICA using model (1), the mixing matrix used in the blind source separation is replaced with the target signature matrix $\mathbf{M}$ and the unknown signal sources to be separated with the p target random abundance fractions, denoted by $\alpha_1, \alpha_2, \ldots, \alpha_p$. With this interpretation, the ICA finds a $p \times L$ separating matrix $\mathbf{W}$ and applies it to an image pixel $\mathbf{r}$ to unmix the $\alpha_1, \alpha_2, \ldots, \alpha_p$. More specifically, the ICA solves an inverse problem of model (1) for a $p \times L$ separating matrix $\mathbf{W}$ via the following equation:

$$\hat{\alpha}(\mathbf{r}) = \mathbf{W}\mathbf{r} \qquad (2)$$

where $\hat{\alpha}(\mathbf{r}) = (\hat{\alpha}_1(\mathbf{r}), \ldots, \hat{\alpha}_p(\mathbf{r}))^T$ is the estimate of abundance fractional vector $(\alpha_1, \alpha_2, \ldots, \alpha_p)^T$ based on $\mathbf{r}$. Since changing order of components in $\hat{\alpha}(\mathbf{r})$ does not affect their statistical independence, the estimate of the ith abundance fraction α_i may appear as any component $\hat{\alpha}_j(\mathbf{r})$ of $\hat{\alpha}(\mathbf{r})$. Furthermore, because multiplying random variables by nonzero scalar factors does not affect the statistical independence, it is also impossible to determine the true amounts of the abundance fractions

$\alpha_1, \alpha_2, \ldots, \alpha_p$ from model (1) without additional assumptions. Unless we are interested in quantification, the order and the true abundance fractions are generally not crucial in target detection and classification. In this case, we can normalize each abundance source to unit variance so that the covariance matrix of the abundance sources becomes the identity matrix. This can be simply done by a sphering (whitening) process.

A. Relative Entropy-Based Measure for ICA

In order to use the ICA, a criterion is required to measure the statistical independence among the estimated abundance fractions $\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_p$. According to information theory [25], relative entropy or Kullback–Leibler information distance function is an appropriate measure. Let $p(\hat{\alpha}(\mathbf{r})) = p(\hat{\alpha}_1(\mathbf{r}), \hat{\alpha}_2(\mathbf{r}), \ldots, \hat{\alpha}_p(\mathbf{r}))$ be the joint probability density function (pdf) of the estimated random abundance vector $\hat{\alpha}(\mathbf{r})$ obtained from (2), and $p(\hat{\alpha}_j)$ be the marginal pdf of the jth abundance fraction $\hat{\alpha}_j$ for $1 \leq j \leq p$ given by (2). Since $\alpha_1, \alpha_2, \ldots, \alpha_p$ in model (1) are assumed to be independent, $p(\alpha) = \prod_{j=1}^{p} p(\alpha_j)$. If we assume that α is a source vector and $\hat{\alpha}(\mathbf{r})$ is the estimate of _ from the observation vector $\mathbf{r}$, then the entropy of α relative to $\hat{\alpha}(\mathbf{r})$ (or relative entropy between α relative to $\hat{\alpha}(\mathbf{r})$), denoted by $D(p(\hat{\alpha}(\mathbf{r}))\|p(\alpha))$, is defined in [25] by

$$
\begin{aligned}
D &\left(p(\hat{\alpha}(\mathbf{r}))\|p(\alpha) = \prod_{j=1}^{p} p(\alpha_j) \right) \\
&= \sum p(\hat{\alpha}(\mathbf{r})) \log \left(\frac{p(\hat{\alpha}(\mathbf{r}))}{p(\alpha)} \right) \\
&= \sum p(\hat{\alpha}(\mathbf{r})) \log \left(\frac{p(\hat{\alpha}(\mathbf{r}))}{\prod_{j=1}^{p} p(\alpha_j)} \right) \\
&= -H(\hat{\alpha}(\mathbf{r})) - \sum p(\hat{\alpha}(\mathbf{r})) \left[\sum_{j=1}^{p} \log(p(\alpha_j)) \right] \quad (3)
\end{aligned}
$$

where $H(\hat{\alpha}(\mathbf{r}))$ is the entropy of the estimated abundance vector $\hat{\alpha}(\mathbf{r})$. Consequently, minimization of (3) over $\hat{\alpha}(\mathbf{r})$ through $\mathbf{W}$ in (2) implies that the smaller the $D(p(\hat{\alpha}(\mathbf{r}))\|p(\alpha))$ is, the less the discrepancy between two pdfs $p(\alpha)$ and $p(\hat{\alpha}(\mathbf{r}))$ is; thus, the more likely to be independent the $\hat{\alpha}(\mathbf{r})$ is. Because the pdf of α is generally unknown and needs to be estimated, the $p(\alpha) = \prod_{j=1}^{p} p(\alpha_j)$ in (3) must be replaced by its estimate $p(\hat{\alpha}(\mathbf{r})) = \prod_{j=1}^{p} p(\hat{\alpha}_j(\mathbf{r}))$. Substituting this estimate into (3) results in

$$
\begin{aligned}
D &\left(p(\hat{\alpha}(\mathbf{r}))\|p(\hat{\alpha}(\mathbf{r})) = \prod_{j=1}^{p} p(\hat{\alpha}_j(\mathbf{r})) \right) \\
&= E \left[-\sum_{j=1}^{P} \log(\hat{\alpha}_j(\mathbf{r})) \right] - H(\hat{\alpha}(\mathbf{r})) \\
&= \sum_{j=1}^{p} H(\hat{\alpha}_j(\mathbf{r})) - H(\hat{\alpha}(\mathbf{r})) \quad (4)
\end{aligned}
$$

where $E[\cdot]$ is the expectation with respect to $p(\hat{\alpha}(\mathbf{r}))$ and $H(\hat{\alpha}_j(\mathbf{r}))$ is the entropy of the jth estimated source $\hat{\alpha}_j(\mathbf{r})$. Unfortunately, even in this case, finding the pdf of $\hat{\alpha}(\mathbf{r})$ is also difficult in practice. In order to mitigate this dilemma, Comon introduced an alternative criterion in [7] that approximates
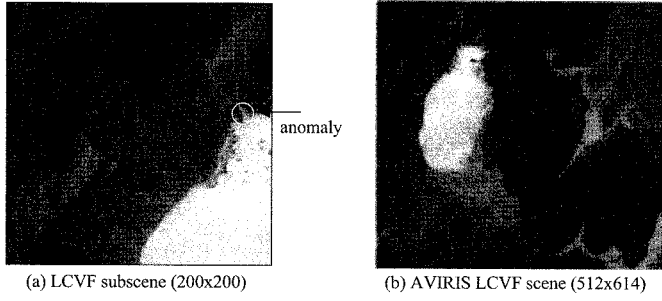
(a) LCVF subscene (200x200)   (b) AVIRIS LCVF scene (512x614)

Fig. 1. (a) LCVF subscene of size 200×200 extracted from Fig. 1(b). (b) AVIRS LCVF scene of size 512×614.

$D\left(p(\hat{\alpha}(\mathbf{r})) \| \prod_{j=1}^{p} p(\hat{\alpha}_j(\mathbf{r}))\right)$ in (4). Instead of minimizing $D\left(p(\hat{\alpha}(\mathbf{r})) \| \prod_{j=1}^{p} p(\hat{\alpha}_j(\mathbf{r}))\right)$, Comon suggested to maximize the higher order statistics of the data, called contrast function of $\mathbf{W}$, denoted by $\psi(\mathbf{W})$ which is defined in [7] by

$$\psi(\mathbf{W}) = \sum_{j=1}^{p} 4\kappa_{jjj}^2 + \kappa_{jjjj}^2 + 7\kappa_{jjj}^4 - 6\kappa_{jjj}^2 \kappa_{jjjj} \quad (5)$$

where κ_{jjj} is the third-order standardized cumulant of the jth $\hat{\alpha}_j(\mathbf{r})$ for $1 \le j \le p$ representing the skewness and κ_{jjjj} is the fourth standardized cumulant of the jth $\hat{\alpha}_j(\mathbf{r})$ for $1 \le j \le p$ representing the kurtosis. If the skewness of $\hat{\alpha}_j$ is sufficiently large, (5) can be further approximated by $7\sum_{j=1}^{p} \kappa_{jjj}^2$. On the other hand, if the kurtosis of $\hat{\alpha}_j(\mathbf{r})$ is sufficiently large, a good approximation of (5) is $\sum_{j=1}^{p} \kappa_{jjjj}^2$. In either case, (5) is reduced to a much simpler criterion.

B. Learning Algorithm to Find $\mathbf{W}$

Since the second-order statistics can be removed by decorrelation, the data vectors $\mathbf{r}$ are first prewhitened prior to separation. In this case, data are completely characterized by statistics with orders higher than 2. Therefore, for simplicity we assume that the data vectors have been prewhitened. In order to derive a learning algorithm, we impose a constraint that the covariance matrix of the estimated abundance vector $\hat{\alpha}(\mathbf{r})$ in (2) must be an identity matrix. To further simplify notations, we denote $\hat{\alpha}(\mathbf{r})$ by $\mathbf{y}$ with $y_j = \hat{\alpha}_j(\mathbf{r})$. The learning algorithm to be developed must solve the following constrained optimization problem:

$$\text{maximize} \quad \psi(\mathbf{W}) = \sum_{j=1}^{p} E\left[y_j^m\right] \text{ over } \mathbf{W} \text{ for } m \ge 3$$

$$\text{subject to} \quad E\left[\mathbf{y}\mathbf{y}^T\right] = \mathbf{I} \quad (6)$$

where $\mathbf{I}$ is the $p \times p$ identity matrix. For such constrained problems, we use exterior penalty methods discussed in [26] to eliminate some or all of the constraints. The idea is to add to the objective function specified by (6) so-called penalty function terms which assign a higher cost to infeasible points. In our case, the penalty function terms imposed on the constraints in (6) are defined by

$$U(h_{ij}(\mathbf{y})) = \begin{cases} 0; & \text{if } h_{ij}(\mathbf{y}) = 0 \\ >0; & \text{if } h_{ij}(\mathbf{y}) \ne 0 \end{cases} \quad (7)$$

where $h_{ij}(\mathbf{y}) = E[y_i y_j] - \delta_{ij}$ and $\delta_{ij} = \begin{cases} 1; & \text{if } i = j \\ 0; & \text{if } i \ne j. \end{cases}$
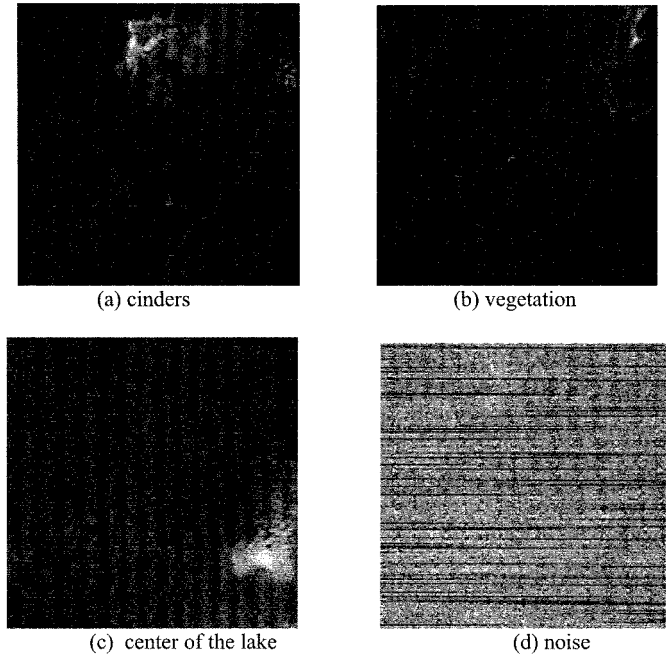

(a) cinders   (b) vegetation

(c) center of the lake   (d) noise

Fig. 2. Detection and classification results produced by the LSRMA with $p = 4$ using skewness: (a) cinders; (b) vegetation; (c) center of the playa; and (d) noise.

(a) anomaly   (b) vegetation

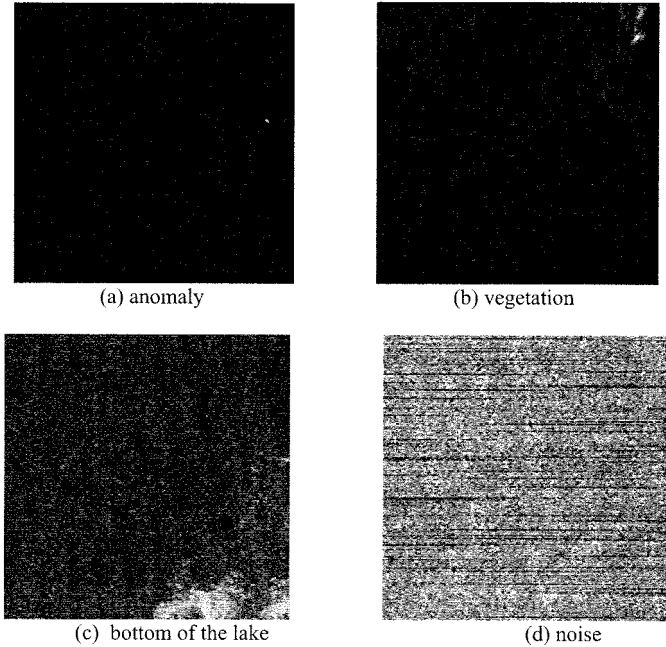(c) bottom of the lake   (d) noise

Fig. 3. Detection and classification results produced by the LSRMA with $p = 4$ using kurtosis: (a) anomaly; (b) vegetation; (c) bottom of the playa; and (d) noise.

Now, using the penalty function terms specified by (6) and (7), we can define a penalty function $P(\mathbf{y})$ by

$$P(\mathbf{y}) = \sum_{i,j=1}^{p} \beta_{ij} U(h_{ij}(\mathbf{y})) = \frac{1}{2} \sum_{i,j=1}^{p} \beta_{ij} \left(E\left[[y_i y_i] - \delta_{ij}\right]\right)^2 \quad (8)$$

where $\beta_{ij} > 0$ are called penalty parameters or penalty multipliers (typically all $\beta_{ij} = \beta$). Thus, maximizing the constrained

(a) cinders

(b) vegetation

(c) center of the lake
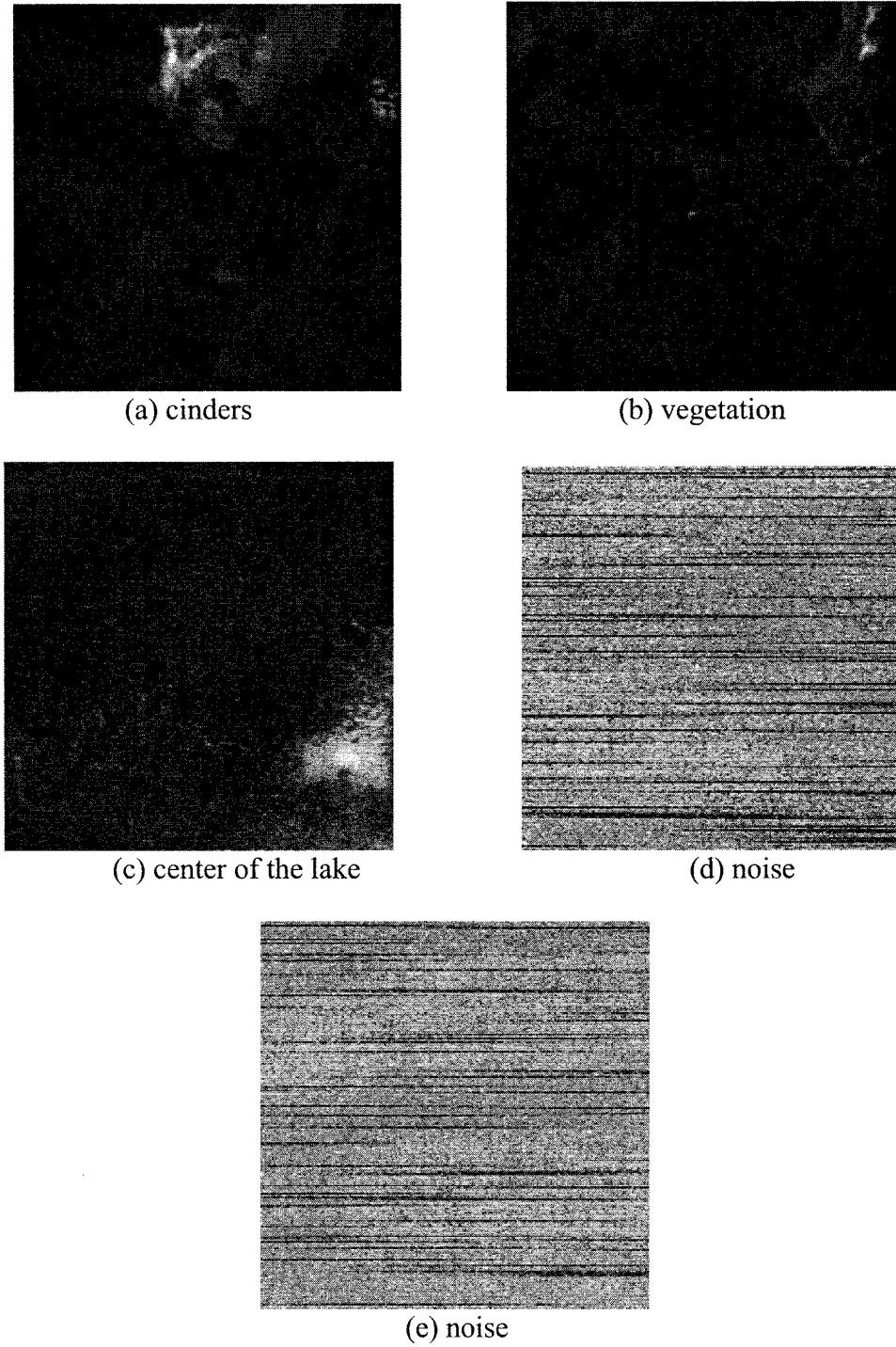

(d) noise

(e) noise

Fig. 4. Detection and classification results produced by the LSRMA $p = 5$ using skewness: (a) cinders; (b) vegetation; (c) center of the playa; and (d)–(e) noise.

problem described by (6) is equivalent to maximizing the following cost function:

$$J(\mathbf{W}) = \psi(\mathbf{W}) - P(\mathbf{y})$$
$$= \psi(\mathbf{W}) - \frac{\beta}{2} \sum_{i,j=1}^{p} (E[[y_i y_i] - \delta_{ij}])^2 \qquad (9)$$

which subtracts the penalty function given by (8) from the objective function in (6). In order to find the separating matrix $\mathbf{W} = [w_{st}]_{p \times l}$ where w_{st} is the (s,t)th matrix element in $\mathbf{W}$, we calculate gradient of the cost function $J(\mathbf{W})$ by differentiating $\psi(\mathbf{W})$ and $P(\mathbf{y})$ with respect to w_{st}, respectively

$$\frac{\partial \psi(\mathbf{W})}{\partial w_{st}} = \frac{\partial}{\partial w_{st}} \sum_{j=1}^{p} E\left[y_j^m\right]^2 = 2m E\left[y_s^m\right] E\left[y_s^{m-1} r_t\right] \qquad (10)$$

$$\frac{\partial P(\mathbf{y})}{\partial w_{st}} = \frac{\beta}{2} \frac{\partial}{\partial w_{st}} \sum_{i,j=1}^{p} (E[y_i y_j] - \delta_{ij})^2$$
$$= 2\beta \sum_{i=1}^{p} (E[y_i y_s] - \delta_{is}) E[y_i r_t]. \qquad (11)$$
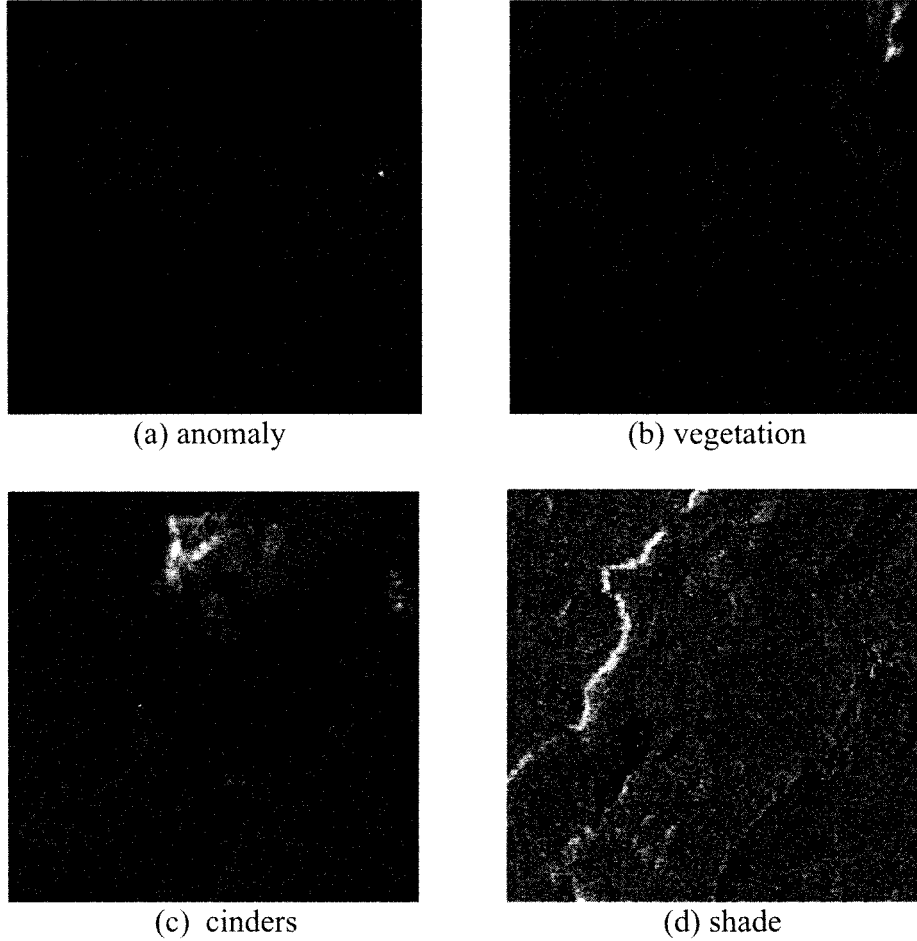
Fig. 5.   Detection and classification results produced by the LSRMA with $p = 5$ using kurtosis: (a) anomaly; (b) vegetation; (c) cinders; and (d) shade.

Equations (10) and (11) can be expressed in terms of matrix forms as follows:

$$\nabla_{\mathbf{W}}\psi(\mathbf{W}) = 2m\Lambda E[g(\mathbf{y})\mathbf{r}^T] \tag{12}$$

$$\nabla_{\mathbf{W}}P(\mathbf{y}) = 2\beta(E[\mathbf{y}\mathbf{y}^T] - \mathbf{I})E[\mathbf{y}\mathbf{r}^T] \tag{13}$$

where $g(\mathbf{y}) = \left(y_1^{m-1}, y_2^{m-1}, \ldots, y_p^{m-1}\right)^T$ and $\Lambda = \text{diag}\{E[y_i^m]\}$ is the diagonal matrix with the ith diagonal element given by $E[y_i^m]$. From (12) and (13), a learning algorithm to generate the separating matrix $\mathbf{W}$ can be designed by

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mu\Lambda E[g(\mathbf{y})\mathbf{r}^T] - \eta(E[\mathbf{y}\mathbf{y}^T] - \mathbf{I})E[\mathbf{y}\mathbf{r}^T] \tag{14}$$

where μ and η are learning parameters for (12) and (13), respectively. It is found empirically that μ controls the convergence speed and should be less than 1 while η controls the constraint and should be greater than μ. In this paper, the η and μ are set to $\eta = 1$ and $\mu = 0.5$ for all the experiments conducted in Section IV. In order for the proposed learning algorithm to converge, a stopping rule is used to terminate the algorithm. It measures the normalized difference between two consecutive $J(\mathbf{W}_k)$ and $J(\mathbf{W}_{k-1})$ resulting from (6), denoted by $\tilde{J}(W_k)$ which is defined by

$$\tilde{J}(\mathbf{W}_k) = \frac{J(\mathbf{W}_k) - J(\mathbf{W}_{k-1})}{J(\mathbf{W}_{k-1})}. \tag{15}$$

If $\tilde{J}(\mathbf{W}_k)$ is less than a prescribed threshold, the algorithm stops; otherwise, it continues. In all the experiments conducted in Section IV, the threshold was set to 10^{-5}.

IV. EXPERIMENTS

In this section, two sets of real hyperspectral image data, airborne visible/infrared imaging spectrometer (AVIRIS), and HYDICE were used for experiments to evaluate the performance of the LSRMA. In addition, two criteria, skewness [$m = 3$ in (6)] and kurtosis [$m = 4$ in (6)] were also used for performance analysis.

A. AVIRIS Data

The data to be used were obtained from an AVIRIS scene of 200×200 pixels shown in Fig. 1(a) and extracted from the lunar crater volcanic field (LCVF) in Northern Nye County, NV, shown in Fig. 1(b) which has been studied extensively [23] and provides a good case for comparative analysis. It was acquired by 224 spectral channels ranging from 0.4 m to 2.5 m with 10-nm spectral resolution and 20-m spatial resolution. After water bands and low signal-to-noise ratio (SNR) bands are removed, only 158 bands remain (i.e., $L = 158$).

According to the ground truth in [23], there were five targets of interest: cinders, rhyolite, playa (dry lake), vegetation, and shade. Therefore, at least five components, $p \geq 5$, are needed

(a) cinders (b) vegetation (c) center of the lake
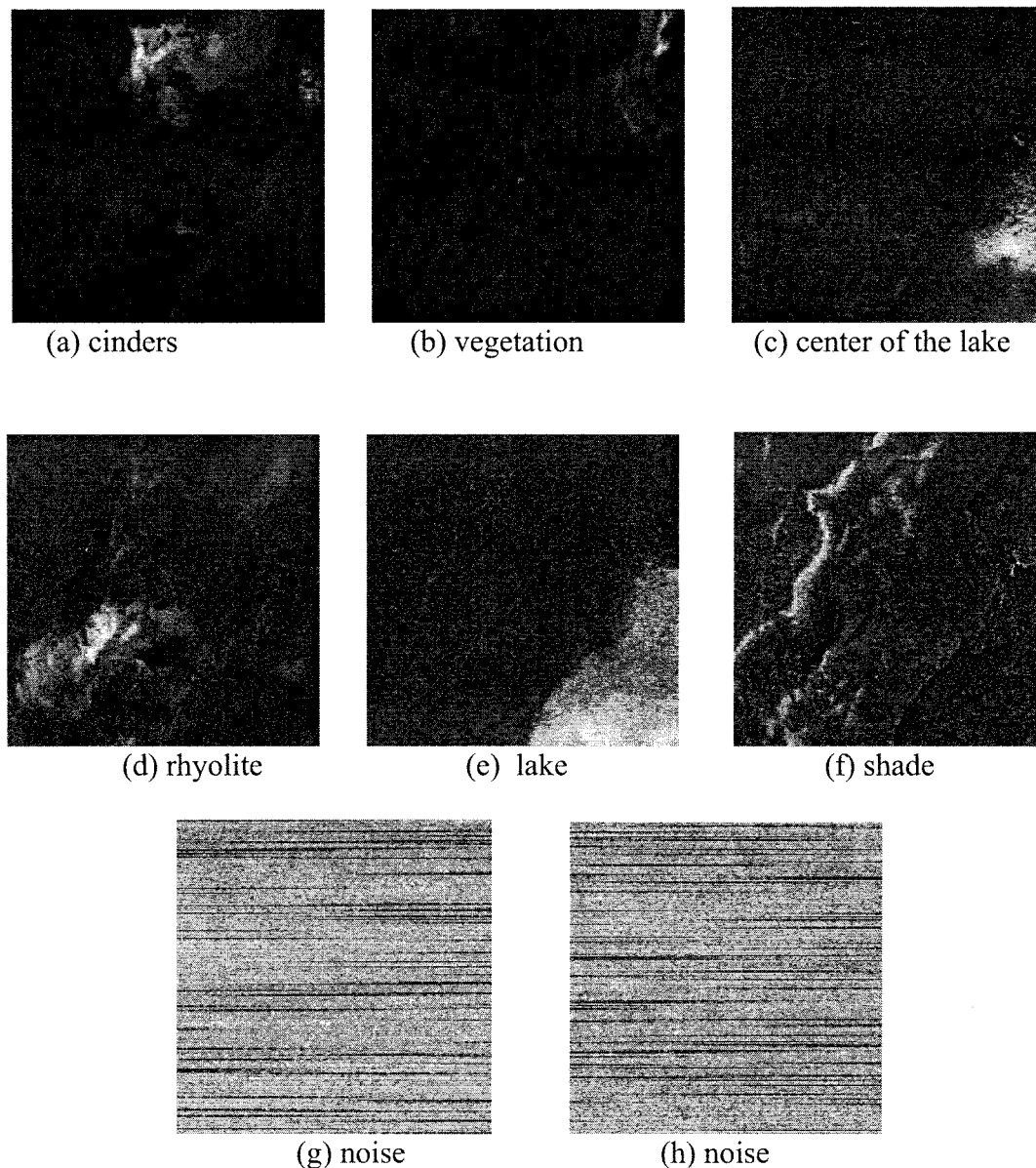
(d) rhyolite (e) lake (f) shade

(g) noise (h) noise

Fig. 6. Detection and classification results produced by the LSRMA $p = 9$ using skewness: (a) cinders; (b) vegetation; (c) center of playa; (d) rhyolite; (e) playa; (f) shade; and (g)–(h) noise.

to classify and separate these five targets. If we assume that no prior knowledge is available about the image scene, the number of targets in an image scene p must be estimated from the data. As shown in [5], there was a single two-pixel anomaly located at the top edge of the lake shown inside a circle marked in Fig. 1(a). This anomaly cannot be seen or detected visually from the scene and was not identified by the ground truth. It was extracted by an unsupervised constraint subpixel detection method [5] and also detected by an anomaly detector, RX algorithm in [27] and [28]. In order for the LSRMA to detect this anomaly in addition to the five targets of interest, the p must be greater than five, so that the five targets (cinders, rhyolite, playa dry lake, vegetation, and shade) and the anomaly can be detected and classified in separate components. In [16], Tu proposed a noise adjusted transformed Gershgorin disk (NATGD) to estimate the p for the same identical image scene in Fig. 1(a), which was four. Unfortunately, this number was underestimated. With $p = 4$,

the shade and the anomaly were not detected in [16]. However, here we use the number produced by a Neyman–Pearson detector-based eigen-thresholding method described in [28] and [29], which was estimated to be eight. Therefore, in this case, $p = 9$ with one extra component included to accommodate the noise. It should be noted that P$p = 9$ is only an estimate. It does not imply that there were exactly eight targets present in the image scene. In order to demonstrate the impact of different values of p on the performance, we conducted experiments with $p = 4$ estimated by NATGD, $p = 5$ estimated by a method developed by Harsanyi *et al.* in [28]–[30], and $p = 9$. The learning parameters μ and η used in (14) for the following experiments were empirically set to $\mu = 0.5$ and $\eta = 1$.

Example 1 ($p = 4$): Fig. 2 shows the results generated by the LSRMA using skewness as a criterion. From the ground truth in [23], only the cinders, vegetation, and center of playa were classified in Fig. 2(a) and (c) and the fourth component

(a) anomaly

(b) vegetation

(c) cinders

(a) center of the lake

(b) edge of the lake

(c) bottom of the lake
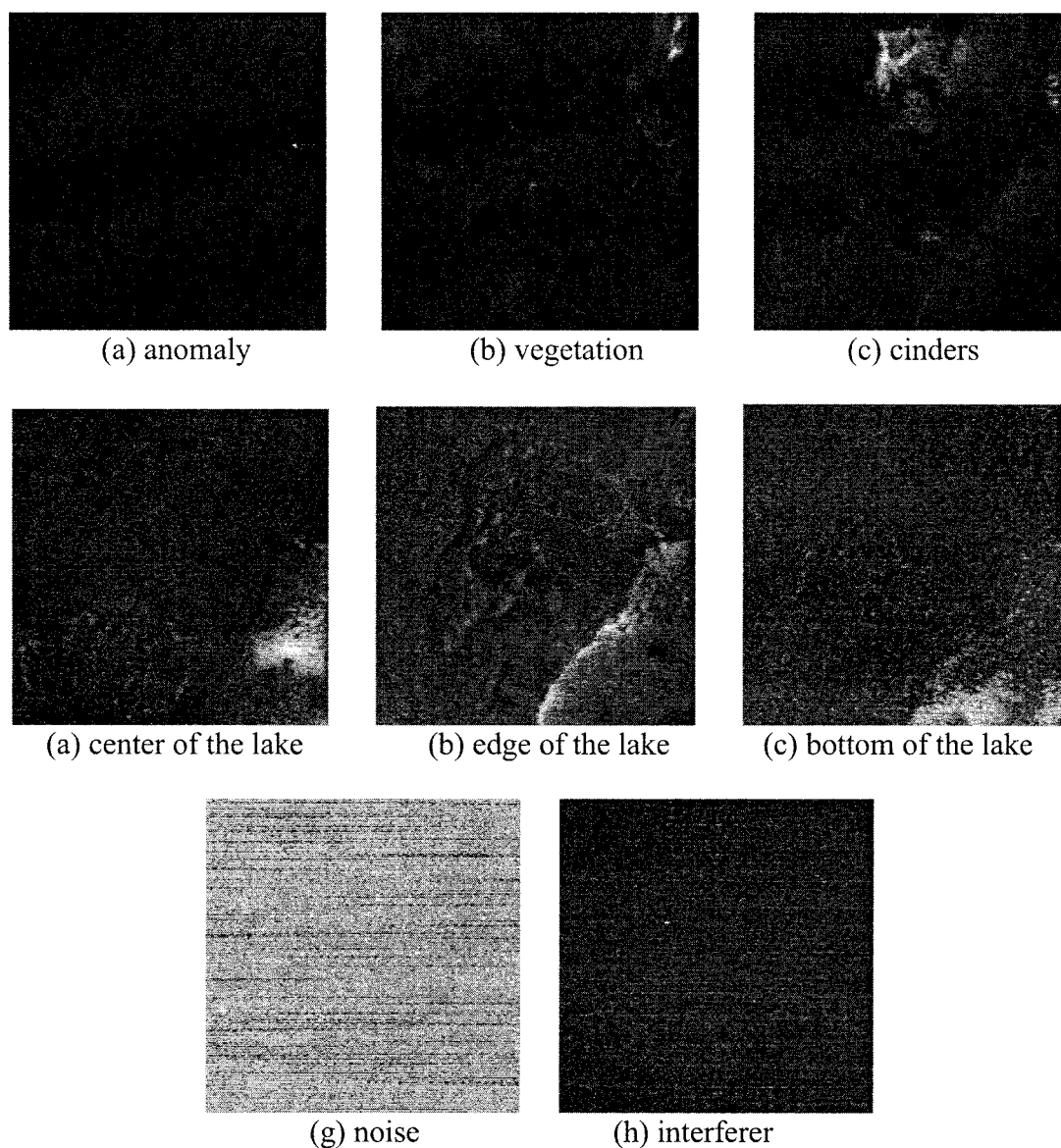

(g) noise

(h) interferer

Fig. 7. Detection and classification results produced by the LSRMA with $p = 9$ using kurtosis: (a) anomaly; (b) vegetation; (c) cinders; (d) center of playa; (e) the edge of playa; (f) bottom of playa; and (g)–(h) noise.

showed very little information in the image. It missed the detection of the rhyolite, shade, and part of playa. Fig. 3(a)–(d) shows the results produced by the LSRMA using kurtosis where only vegetation and the bottom part of playa were classified in Fig. 3(b) and (c). Once again, the fourth component did not detect any meaningful targets. It also missed the detection of the cinders, rhyolite, shade, and a large portion of playa. Interestingly, Fig. 3(a) detected the anomaly observed in [5]. This experiment demonstrated that skewness and kurtosis performed very differently, but the kurtosis seemed more effective to detect small targets. However, more experiments are required to substantiate our conclusion.

Example 2 $(p = 5)$: In this example, we conducted an experiment similar to Example 1 with $p = 5$. Figs. 4 and 5 were results obtained for skewness and kurtosis, respectively. Comparing Fig. 4 to Fig. 2, there was not much changed in results for skewness except that the fifth component showed nothing but noise. However, when the kurtosis was used, there was a drastic

change between Figs. 3 and 5. Both detected the anomaly in their first components shown in Figs. 3(a) and 5(a). However, the cinders and shade which were extracted in Fig. 5(c)–(d), respectively, were missed in Fig. 3. Unfortunately, it still also missed the detection of the rhyolite and the playa.

Example 3 $(p = 9)$: A similar experiment was also conducted for the case of $p = 9$. Figs. 6 and 7 show the results produced by the LSRMA using skewness and kurtosis, respectively, where the skewness and kurtosis demonstrated different strengths in target extraction. The cinders, the vegetation, the playa, the rhyolite, and the shade were detected by skewness in the first six components shown by Fig. 6(a)–(f). Interestingly, the playa was detected in two separate components in Fig. 6(c) and (e) and the anomaly was detected along with the shade in Fig. 6(f). By contrast, kurtosis extracted the anomaly, vegetation, and cinders in the first three components shown by Fig. 7(a)–(c), while the playa was extracted in the next three separate components in Fig. 7(d)–(f). Unfortunately, the rhy-

olite and the shade which were extracted in Fig. 6(c) and (f) by skewness were not detected by the kurtosis. Since the playa covers a very large area of the image scene, i.e., dry lake on the crater, we may expect that its spectral variability could be relatively large. Therefore, the playa was classified in three separate components with the center detected in Figs. 6(c) and 7(d), the edge detected in Fig. 7(e), and the bottom detected in Figs. 6(e) and 7(f). This phenomenon was also observed in [5] and [31]. Such subtle spectral variations were overlooked in [23] because the playa signature was obtained by averaging most part of the dry lake area using visual inspection. Using this averaged spectral signature as the prior knowledge, the entire playa was extracted, and the single two-pixel anomaly was averaged out and could not be detected. A similar problem was also found in [16] where the lake was detected as an entity, and the shade and anomaly were not detected. This resulted from the fact that the estimated p was too small and the learning rule was derived by the orthogonality of $\mathbf{W}$, not $\hat{\alpha}(\mathbf{r})$. These experiments demonstrate that skewness may be a good criterion for classification of large areas while the kurtosis may be effective in extracting small targets or insignificant targets.

One comment is noteworthy. In general, the number of independent components should be equal to the number of spectral dimensionality, which is $L = 158$. However, after the p exceeds a certain number, there is very little change in the projection vectors of the components that are beyond this number. In this case, we did not include these components. Therefore, in order to determine how many components should be generated, the Euclidean distance difference between the projection vectors of two consecutive components is compared. If it is less than a prescribed value—in our case, it was set to 10^{-5}—then the LSRMA stops generating new components. That is why there were various numbers of components generated for different values of p. However it should be noted that different values of p will generate different numbers of components.

To conclude the AVIRIS experiments, variable values of p were also evaluated by skewness and kurtosis. The results for $p = 4, 5, 6, 7$, and 8 were not as good as those obtained by $p = 9$. For the results of $p = L = 158$ and $p = 8, 12$, we refer the reader to [17]. In order to find which number is appropriate for p, we ran additional experiments for $10 \leq p \leq 40$. There was little change in the first few components and the performance of both criteria was not significantly improved except that more components were used to detect different portions of image scene. In this case, $p = 9$ may be a good estimate.

B. HYDICE Data

The HYDICE data used in the following experiments were directly extracted from the HYDICE image scene of size 64×64 shown in Fig. 8(a). The image data were acquired by the airborne HYDICE sensor in August 1995 from a flight altitude of 10 000 ft with the ground sampling distance approximately 1.5 m. It has 210 spectral channels ranging from 0.4 μm to 2.5 μm with spectral resolution 10 nm. 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. Therefore, a total of 169 bands were used for the experiments. There are 15 panels located on a grass field and arranged in a
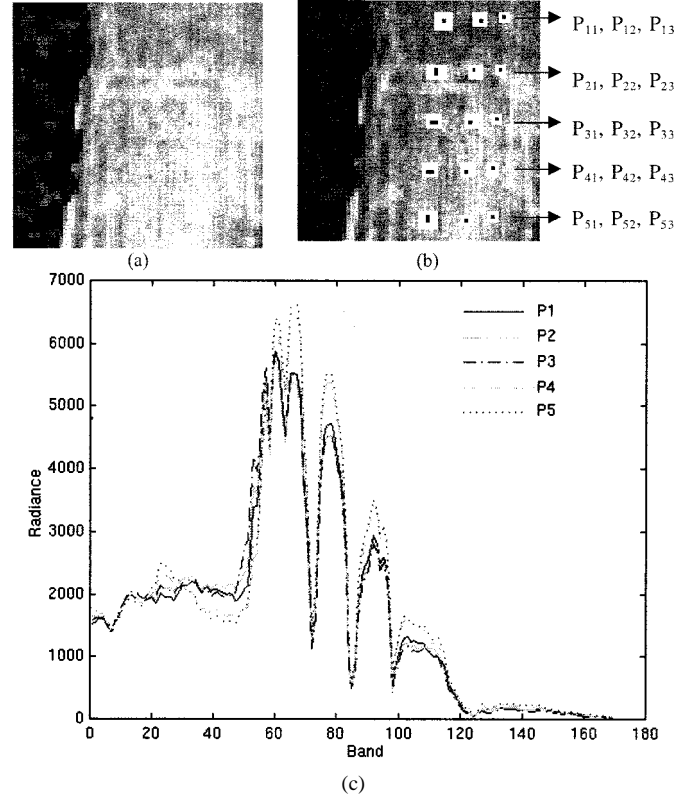

Fig. 8. (a) A 15-panel HYDICE scene; (b) ground truth map of Fig. 8(a); and (c) five panel signatures obtained by averaging B pixels in the 15 panels.

5×3 matrix where there is a forest on the left edge of the scene and a road on the right edge of the scene. A ground truth map of this 15-panel scene is shown in Fig. 8(b) and provides the precise spatial locations of these 15 panels. Black pixels are panel center pixels, considered to be pure pixels, and the pixels in the white masks are panel boundary pixels mixed with background pixels, considered to be mixed pixels. Each element in this matrix is a square panel and is 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}, and 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}, and p_{5j} have the same size but were made from 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_{21}, p_{31}, p_{41}, and p_{51}, which are two-pixel panels, all of the remaining panels are only one pixel in size. Five spectral signatures were obtained from the center pixels of the 15 panels in Fig. 8(b) to represent five different panel signatures. They were denoted by P_1, P_2, P_3, P_4, and P_5 where P_i is the ith panel signature obtained by averaging the black panel center pixels in row i and their spectra are shown in Fig. 8(c).

Example 4 (Computer Simulations): In this experiment, we simulated a scene with size of 50×50, which is similar to Fig. 8(a). It consists of 25 single-pixel panels simulated by P_1, P_2, P_3, P_4, and P_5 in Fig. 8(c). These 25 pixels are arranged in five rows with five pixels in each row, and were generated to simulate the 15-panel HYDICE scene in Fig. 8(a). The 25 pixels were simulated by the panel signature P_i, shown in Fig. 8(c),
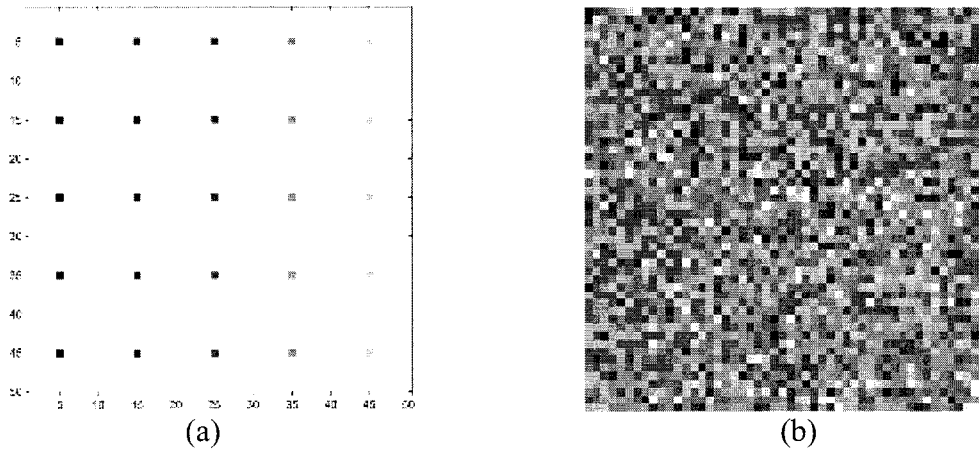
Fig. 9.   (a) Abundance fractions map assigned to the 25 simulated pixels. (b) Background scene simulated by tree and grass signatures uniformly.
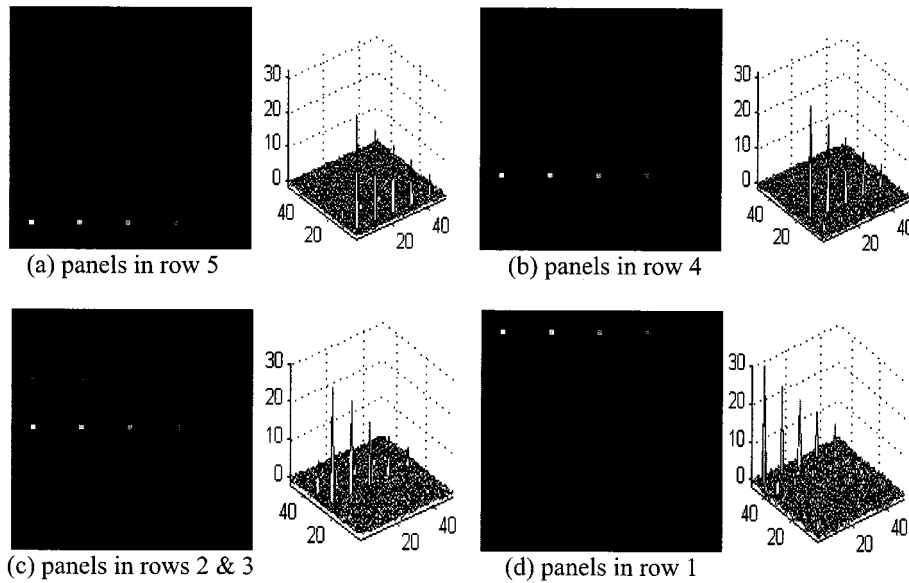

Fig. 10.   Detection and classification results produced by the LSRMA with $\mathrm{SNR} = 5$ dB using skewness: (a) panels in row 5; (b) panels in row 4; (c) panels in rows 2 and 3; and (d) panels in row 1.

with the assigned abundance fraction map shown in Fig. 9(a). The pixels in the same column have the same abundance fraction contributed by five different panel signatures P_1, P_2, P_3, P_4, and P_5. The pixels in the same row contain the same panel signature with abundance fractions assigned by 1.0, 0.8, 0.6, 0.4, and 0.2, respectively, i.e., $\alpha = (1.0, 0.8, 0.6, 0.4, 0.2)^T$. In this case, the mixing matrix $\mathbf{M}$ is formed by $\mathbf{M} = [P_1 \, P_2 \, P_3 \, P_4 \, P_5]$. In addition, a background scene shown in Fig. 9(b) was also added to these 25 simulated pixels. It was simulated by two background signatures, tree and grass uniformly with their abundance fractions being positive and summed to one. In addition, a zero-mean Gaussian noise with different variances is added to achieve different levels of SNR. In the following simulations, six scenarios were simulated to achieve different SNR, with $\mathrm{SNR} = 30$ dB, 25 dB, 20 dB, 15 dB, 10 dB, and 5 dB were conducted to evaluate the performance of the LSRMA for skewness and kurtosis. In the skewness case, the ICA converged to five components for $\mathrm{SNR} = 30$ dB, 25 dB, 20 dB, 15 dB, and 10 dB, and all 25 panel pixels were detected and accurately classified into their own classes in these five components. For the case of $\mathrm{SNR} = 5$ dB,

the LSRMA converged to only four components and the panel pixels in rows 2 and 3 were detected but forced to be classified into one class. This is because the spectra of P_2 and P_3 are very similar, as demonstrated in [32, Table IV]. For comparison, only results for $\mathrm{SNR} = 5$ dB and 10 dB are shown in Figs. 10 and 11, respectively, where the amounts of detected abundance fractions are also plotted for reference. Unlike the case of skewness, the LSRMA using kurtosis produced a very large number of components for $\mathrm{SNR} = 30$ dB, 25 dB, 20 dB, 15 dB, 10 dB, and 5 dB, but all 25 panel pixels were detected and accurately classified into their own classes in the first five components. Since the results for all the cases were similar except for different amounts of abundance fractions detected by different SNRs, only results for the case of $\mathrm{SNR} = 5$ dB is shown in Fig. 12. It is worth noting that there were also interferers and background signatures with relatively small abundance fractions detected in the components beyond the first five components. Furthermore, as the SNR was increased, the number of converged components was also increased, but the amount of detected abundance fractions are more accurate and close to the true abundance fractions. This
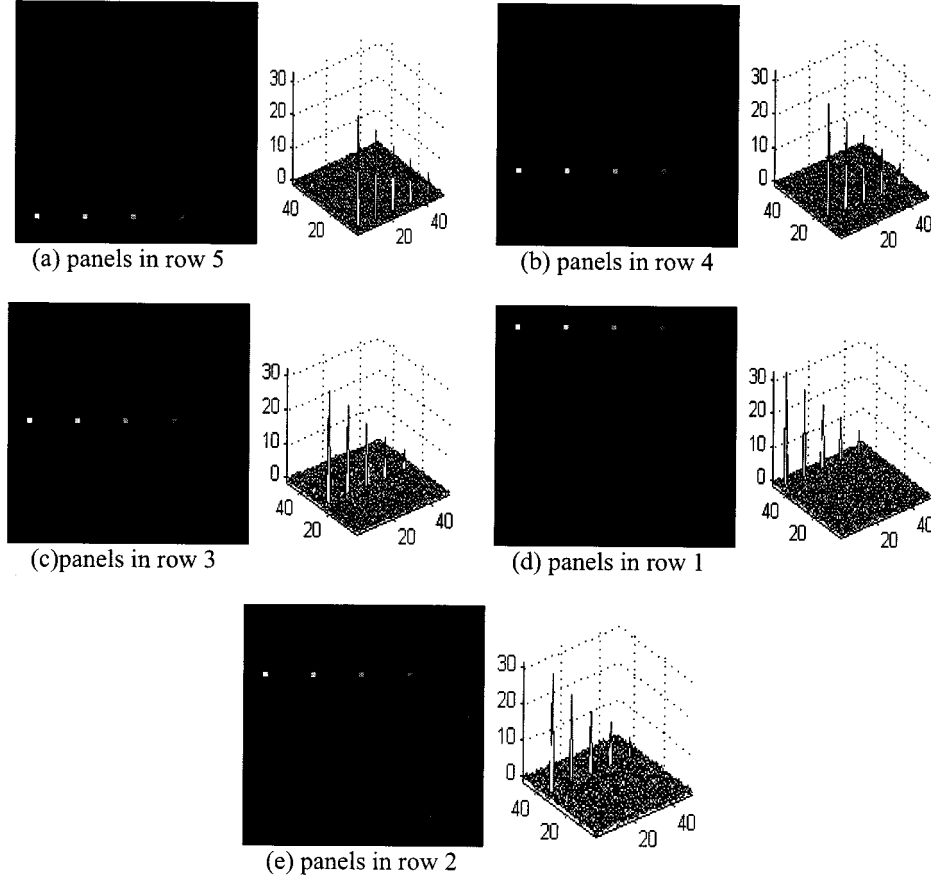
(a) panels in row 5

(b) panels in row 4

(c) panels in row 3

(d) panels in row 1

(e) panels in row 2

Fig. 11.   Detection and classification results produced by the LSRMA with $\mathrm{SNR} = 10$ dB using skewness: (a) panels in row 5; (b) panels in row 4; (c) panels in row 3; (d) panels in row 1; and (e) panels in row 2.

implies that more interferers background signatures were also detected, each of which was classified into a separate and individual component. These computer simulations demonstrated that kurtosis is very effective but also sensitive to target detection. In addition, the performance of both criteria is also proportional to the strength of SNR.

Example 5 (Real HYDICE Scene):   In this example, we conducted experiments based on the real scene in Fig. 8(a) where the number of mixing targets was estimated to be $p = 20 + 1 = 21$ by the Neyman–Pearson eigen-thresholding method proposed in [28] and [29]. Figs. 13 and 14 show the results generated by the LSRMA using skewness and kurtosis, respectively. As we can see from these figures, the LSRMA was able to detect all 15 panels in the second through sixth components in both cases where the panels in the third column were barely visible. This is because these panels are of size 1 m × 1 m, while the spatial resolution is 1.5 m × 1.5 m. In this case, the detection of the panels of 1 m × 1 m is actually subpixel detection. Interestingly, a strong interferer located in the left upper corner in the forest of Fig. 8(a) was detected by both skewness and kurtosis in their first component. By visual inspection of the scene in Fig. 8(a), there is no way to identify this interferer. The detection of the unidentified interferer in Fig. 8 and the anomalous target in Fig. 1(a) demonstrates that the LSRMA can be used as an anomaly detector to detect unknown targets. However, it was unable to discriminate panels in row 2 from those in row 3. This also was witnessed in Fig. 10(a) by the computer

simulation using the skewness with $\mathrm{SNR} = 5$ dB. In order to make a comparison, we also conducted a similar experiment for $p = L = 169$. The results for skewness and kurtosis are shown in Figs. 15 and 16, respectively, where the interferer and all 15 panels were detected and classified into six separate components. Like previous experiments, both skewness and kurtosis have the same difficulty with discriminating panels in row 2 from those in row 3. It should be noted that only 6 and 13 components were produced by the skewness and the kurtosis, respectively. This implies that p is greater than the number of targets (six in this example), and the number of the components will be eventually reduced no matter how large the value of p is.

According to our experiments running p from 5 to 40, it turned out that when $p \geq 18$, the LSRMA using skewness and kurtosis became stable in the sense that all 15 panels and the interferer were detected in the first six independent components. The remaining components beyond six contain mostly noise and background signatures such as grass, forest, tree, road, etc. In this case, our estimate $p = 21$ produced by the proposed eigen-thresholding method in [28] and [29] was reasonably close to 18.

V. AUTOMATIC THRESHOLDING METHOD

In view of the fact that the images generated by LSRMA are generally grayscale, the detection is usually carried out by visual inspection. The interpretation of detection and classifica-
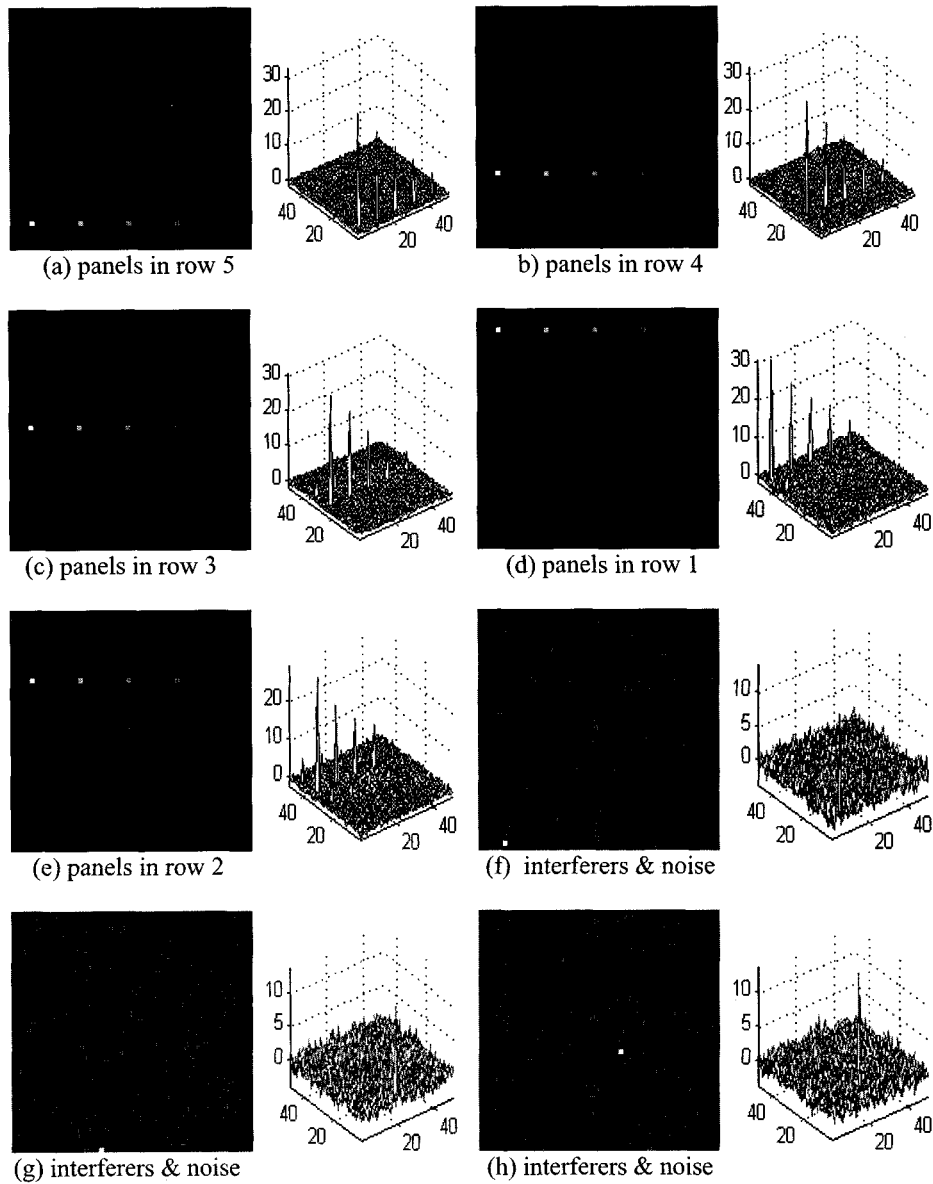
(a) panels in row 5   b) panels in row 4

(c) panels in row 3   (d) panels in row 1

(e) panels in row 2   (f) interferers & noise

(g) interferers & noise   (h) interferers & noise

Fig. 12. Detection and classification results produced by the LSRMA with $\mathrm{SNR} = 5$ dB using kurtosis: (a) panels in row 5; (b) panels in row 4; (c) panels in row 3; (d) panels in row 1; (e) panels in row 2; and (f)–(h) noise or interferers.

tion results can be subjective and vary with human judgment. In order to avoid such human intervention and to make an objective assessment, we develop a computer-automated thresholding method in this section. The suggested threshold criterion was proposed in [33], and can be used to automatically extract the anomalous target pixels and segment them from the background.

Let $\delta_{\mathrm{LSRMA}}(\mathbf{r})$ be the abundance fraction of an image pixel $\mathbf{r}$ resulting from LSRMA, which will represent gray level value of $\mathbf{r}$. It should be noted that $\delta_{\mathrm{LSRMA}}(\mathbf{r})$ is generally a real number and does not necessarily lie in the range [0,1]. For a given gray level value α, we define a rejection region, denoted by $R(\alpha) = \{\mathbf{r}|\delta_{\mathrm{LSRMA}}(\mathbf{r}) < \alpha\}$, by a set made up of all the image pixels in the LSRMA-generated image whose gray level values less than the α. We then use the histogram of the LSRMA-generated image to define the rejection probability $P(\alpha)$ as

$$P(\alpha) = \mathrm{Pr}(R(\alpha)). \qquad (16)$$

A threshold value α_0 can be used to segment targets out from the background by a prescribed confidence coefficient γ such that

$$P(\alpha_0) = \gamma. \qquad (17)$$

More precisely, assume that the confidence coefficient is set to γ with $P(\alpha_0) = \gamma$. If $\delta_{\mathrm{LSRMA}}(\mathbf{r}) > \alpha_0$, $\mathbf{r}$ will be detected as a target pixel and a background pixel, otherwise. It should be noted that the confidence coefficient can be adjusted and is determined by target size. For small targets such as panels in Fig. 8(a), a panel size of 3 m $\times$ 3 m with pixel resolution 1.5 m would have at most four pixels. Therefore, the ratio of a panel with four pixels to the entire image size with $64 \times 64 = 4096$ pixels is no more than 0.001. In this case, a reasonable estimate of the confidence coefficient γ would be approximately $\gamma = 1 - 0.001 = 0.999$. Tables I and II tally the number of panel B pixels detected in Figs. 13 and 14 with $\gamma = 0.997$, 0.998, and 0.999, respectively. When $\gamma = 0.999$, both skewness and kurtosis can achieve 0% false alarm rate, they also missed detection of six B pixels. On the other
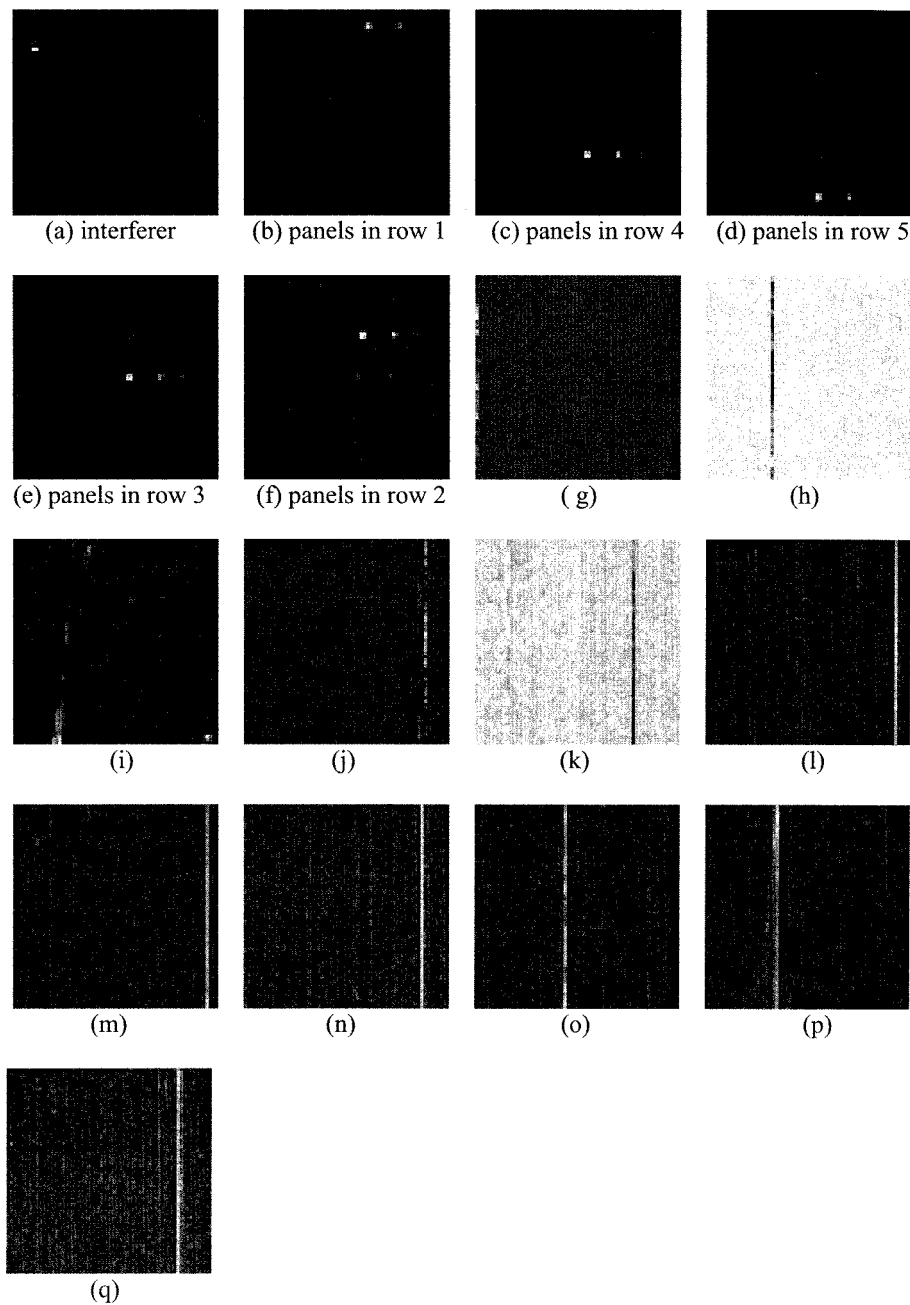
Fig. 13. Detection and classification results produced by the LSRMA with $p = 21$ using skewness: (a) interferer; (b) panels in row 1; (c) panels in row 4; (d) panels in row 5; (e) panels in row 3; (f) panels in row 2; and (g)–(r) noise or interferers.

hand, when γ was decreased to 0.997, both skewness and kurtosis detected all the 19 B pixels of the 15 panels at the expense of detecting nine and seven false alarm pixels, respectively. Fig. 17(a) and (b) shows the binary images resulting from thresholding the images in Figs. 13(b)–(f) and 14(b)–(f) using $\gamma = 0.997$ as the confidence coefficient in (17). As we can see from these images, the five B pixels of panels in the third column (i.e., $p_{13}, p_{23}, p_{33}, p_{43}$, and p_{53} in Fig. 8(b), one for each row) were effectively detected in Fig. 17(a) and (b) by both skewness and kurtosis with $\gamma = 0.997$. It should be noted that these five B panel pixels $p_{13}, p_{23}, p_{33}, p_{43}$, and p_{53} cannot be seen from visual inspection of Fig. 8(a) because their size is merely of 1 m × 1 m and smaller than the 1.5 m pixel resolution. This example further demonstrates

ability of the LSRMA in subpixel detection. More importantly, the experiments also suggest that the three assumptions made for the LSRMA in the introduction seem to be reasonable and acceptable in our applications.

In order to evaluate the performance of LSRMA, two linear unmixing methods, OSP, in [23] and CEM in [24] were used for comparative analysis. Such selections were made based on two reasons. One is that both OSP and CEM have shown success in target detection and classification and have been used in hyperspectral image analysis. Another is from a target knowledge viewpoint. The OSP requires complete target knowledge, whereas the CEM only needs the knowledge of the desired target of interest. Compared to the OSP and the CEM, the LSRMA
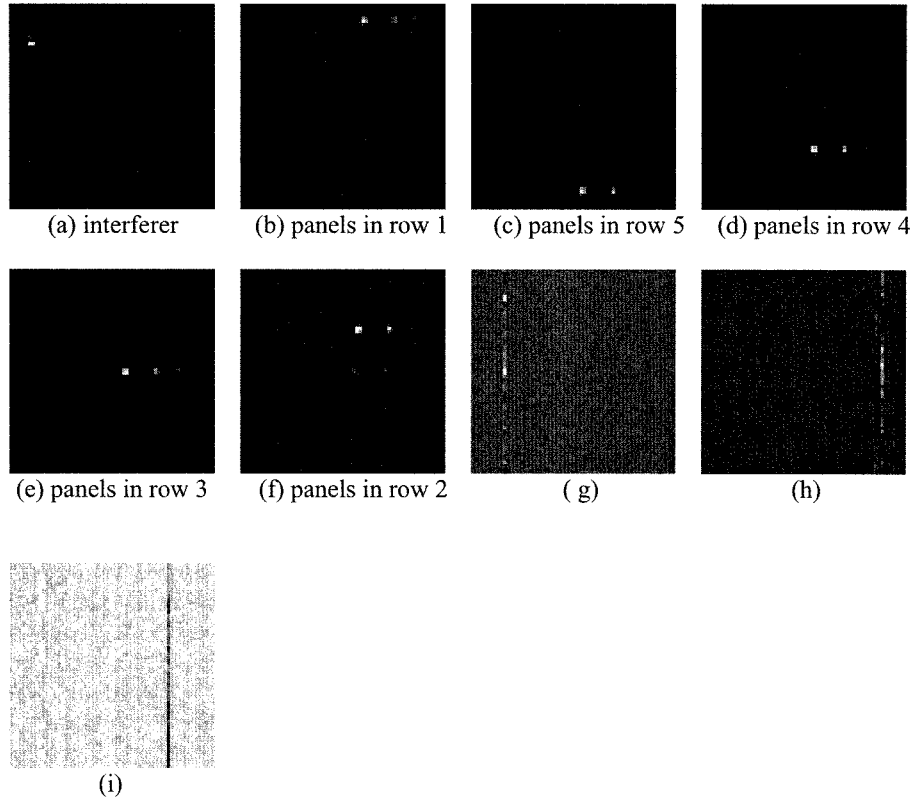
Fig. 14. Detection and classification results produced by the LSRMA with $p = 21$ using kurosis: (a) interferer; (b) panels in row 1; (c) panels in row 5; (d) panels in row 4; (e) panels in row 3; (f) panels in row 2; and (g)–(i) noise or interferers.
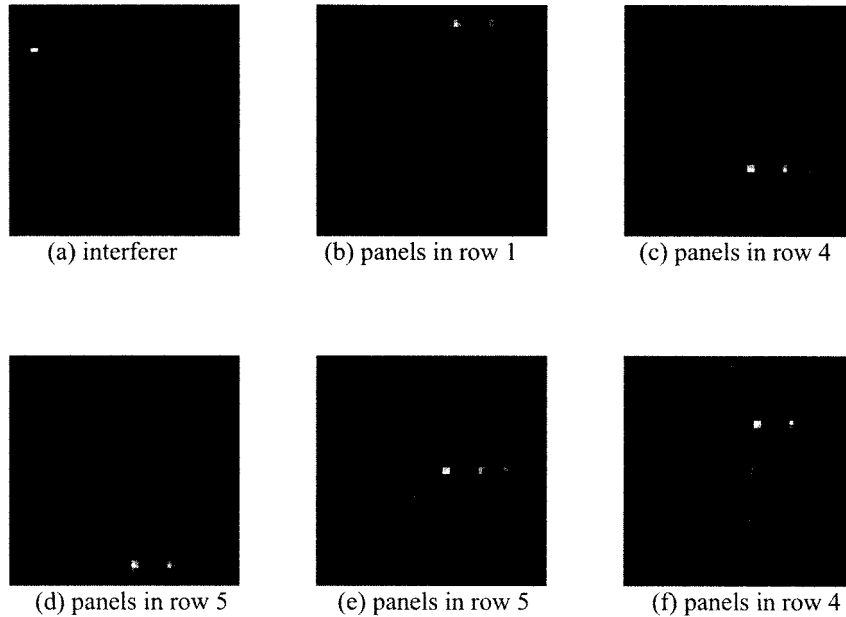

Fig. 15. Detection and classification results produced by the LSRMA with $p = 169$ using skewness: (a) interferer; (b) panels in row 1; (c) panels in row 4; (d) panels in row 5; (e) panels in row 3; and (f) panels in row 2.

does not require any prior target knowledge except p, which is the number of targets assumed to be present in the image scene. However, if p can be estimated reliably, in fact, the LSRMA does not need any information at all.

Tables III and IV tally the results produced by the OSP and the CEM using the same confidence coefficients $\gamma = 0.997, 0.998,$ and 0.999 where the OSP used the five panel signatures in Fig. 8(c) as its complete target knowledge while the CEM only used P_i as its desired target signature to detect panels in row i. Interestingly, if we compare Table IV to Tables I and II, the LSRMA performed as well as the CEM and their results were nearly the same, even if the LSRMA did
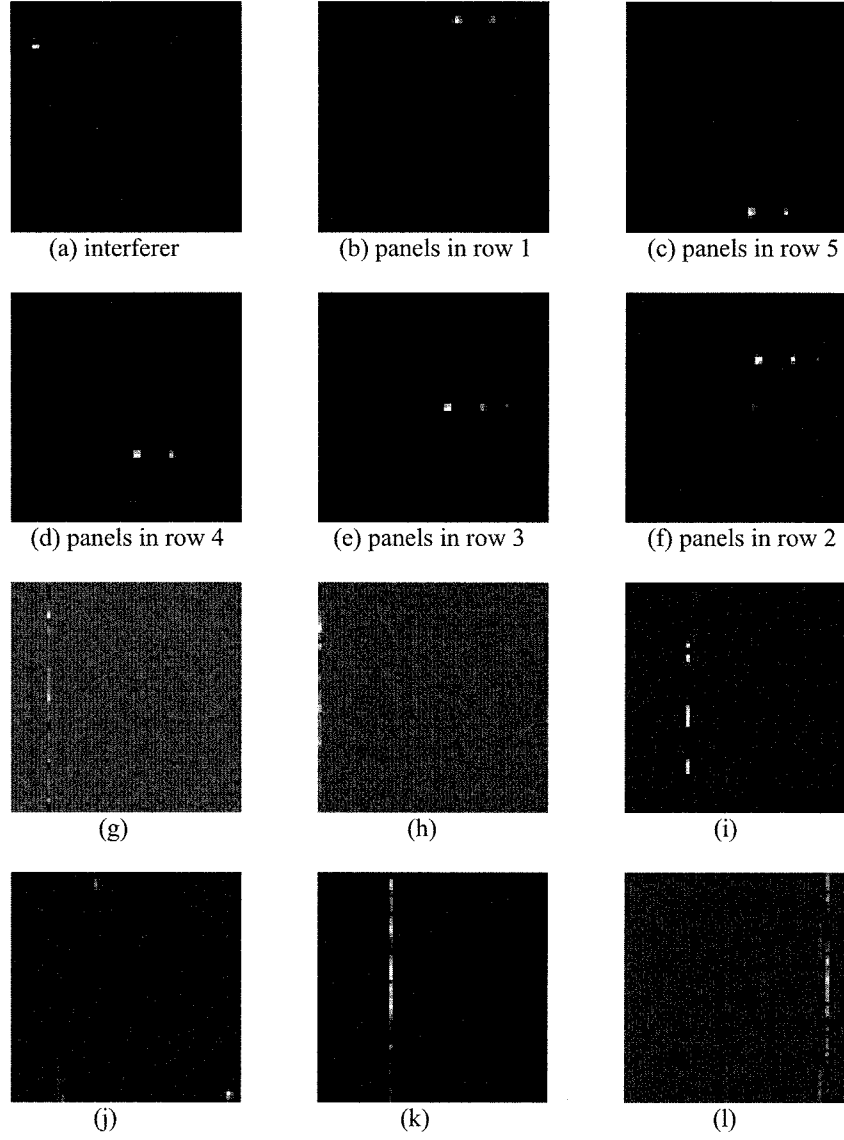
Fig. 16. Detection and classification results produced by the LSRMA with $p = 169$ using kurosis: (a) interferer; (b) panels in row 1; (c) panels in row 5; (d) panels in row 4; (e) panels in row 3; (f) panels in row 2; and (g)–(l) noise or interferers.

TABLE I
TALLY OF NUMBER OF PANEL B PIXELS
DETECTED IN FIG. 13(B)–(F) WITH $\gamma = 0.997, 0.998$, AND 0.999

T	N_B	$\gamma = 0.997$		$\gamma = 0.998$		$\gamma = 0.999$	
		N_{BD}	N_F	N_{BD}	N_F	N_{BD}	N_F
P1	3	3	0	2	0	2	0
P2	4	4	3	3	2	3	0
P3	4	4	4	4	0	2	0
P4	4	4	1	4	0	3	0
P5	4	4	1	3	0	3	0
Total	19	19	9	16	2	13	0

TABLE II
TALLY OF NUMBER OF PANEL B PIXELS DETECTED IN FIG. 14(B)–(F) WITH
$\gamma = 0.997, 0.998$, AND 0.999

T	N_B	$\gamma = 0.997$		$\gamma = 0.998$		$\gamma = 0.999$	
		N_{BD}	N_F	N_{BD}	N_F	N_{BD}	N_F
P1	3	3	0	2	0	2	0
P2	4	4	2	4	1	3	0
P3	4	4	3	4	0	2	0
P4	4	4	1	4	0	3	0
P5	4	4	1	3	0	3	0
Total	19	19	7	17	1	13	0

not assume any target knowledge. On the other hand, Table III shows that the OSP performed very poorly. This is because the target knowledge used in the OSP did not well represent the image scene. It was made up of only the five panel signatures $\{P_1, P_2, P_3, P_4, P_5\}$ and did not include background signatures such as the large grass field, the forest on the left edge and the road on the right edge of the scene.

VI. CONCLUSION

This paper presented an ICA-based LSRMA approach to hyperspectral target detection and classification. It is different from the commonly used ICA approach in that the learning algorithm is derived from the orthogonality constraint imposed on the abundance vector rather than the separating matrix $\mathbf{W}$. In addition, the separating matrix $\mathbf{W}$ is not necessarily

Detection of Panel 1 (α_0 =4.9019)

Detection of Panel 2 (α_0 =5.7652)

Detection of Panel 3 (α_0 =4.0263)

Detection of Panel 4 (α_0 =3.6379)

Detection of Panel 5 (α_0 =4.3647)

Detection of all the 15 panels

(a)

Detection of Panel 1 (α_0 =4.7596)

Detection of Panel 2 (α_0 =5.2888)

Detection of Panel 3 (α_0 =3.4243)

Detection of Panel 4 (α_0 =3.6922)

Detection of Panel 5 (α_0 =4.1389)

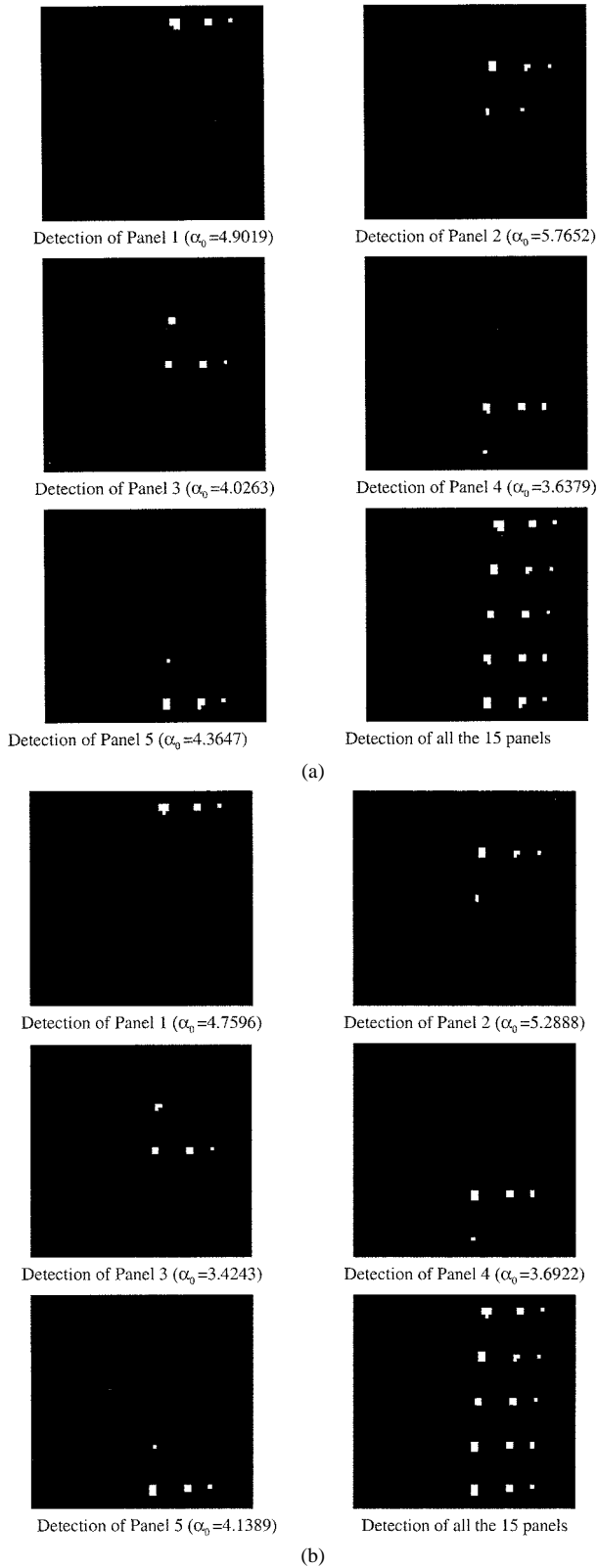Detection of all the 15 panels

(b)

Fig. 17. (a) Image in Fig. 13(b)–(f) thresholded by the values of α_0 that were determined by $\gamma = 0.997$ via (17). (b) Image in Fig. 14(b)–(f) thresholded by the values of α_0 that were determined by $\gamma = 0.997$ via (17).

TABLE III
TALLY OF NUMBER OF PANEL B PIXELS PRODUCED BY THE OSP WITH
$\gamma = 0.997, 0.998,$ AND 0.999

		$\gamma = 0.997$		$\gamma = 0.998$		$\gamma = 0.999$	
T	N_B	N_{BD}	N_F	N_{BD}	N_F	N_{BD}	N_F
P1	3	0	13	0	8	0	5
P2	4	0	15	0	8	0	4
P3	4	0	12	0	8	0	5
P4	4	0	14	0	9	0	4
P5	4	1	12	1	8	0	4
Total	19	1	66	1	41	0	22

TABLE IV
TALLY OF NUMBER OF PANEL B PIXELS PRODUCED BY THE CEM WITH
$\gamma = 0.997, 0.998,$ AND 0.999

		$\gamma = 0.997$		$\gamma = 0.998$		$\gamma = 0.999$	
T	N_B	N_{BD}	N_F	N_{BD}	N_F	N_{BD}	N_F
P1	3	3	0	2	0	2	0
P2	4	4	2	4	1	3	0
P3	4	4	1	4	0	3	0
P4	4	4	2	4	0	3	0
P5	4	4	0	3	0	3	0
Total	19	19	5	17	1	14	0

signatures are generally similar to some degree and not orthogonal. The new designed learning algorithm is able to converge to nonorthogonal independent components. The proposed LSRMA offers several advantages over existing LSMA-based techniques which require complete knowledge for M used in the linear model. It has been shown in [34] that LSMA-based techniques were sensitive to target knowledge M and noise. The proposed LSRMA does not have this sensitivity problem since there is no target knowledge required for the M used in LSRMA. Second, the LSMA-based techniques generally use the least squares error as an optimal criterion which is the second-order statistics. However, the criteria such as relative entropy, skewness and kurtosis used in LSRMA go beyond the second-order statistics and have been shown effective in target detection in [20] and [35]. Third, assuming that abundance fractions as random signal sources seems more appropriate to nonstationary environments. Furthermore, in order to obtain an objective assessment without appealing for human interpretation, an automatic thresholding method is also introduced in this paper. In particular, a quantitative study is also conducted in this paper for comparative analysis with two commonly used methods, OSP which requires complete target knowledge and CEM which only needs partial target knowledge. As shown in experiments, the LSRMA performed as well as CEM and significantly better than OSP. Interestingly, as also demonstrated by experiments, skewness and kurtosis used in the LSRMA have different strengths in classification. Because the skewness measures asymmetry of a distribution, it generally can detect changes in large areas. On the other hand, the kurtosis measures the flatness of a distribution, thus it can detect small targets.

a square matrix of full rank nor is the mixing matrix M orthogonal. These advantages are very useful in hyperspectral image classification. Since hyperspectral sensors are capable of uncovering targets with subtle differences, their spectral

References

[1] J. B. Adams, M. O. Smith, and A. R. Gillespie, "Image spectroscopy: Interpretation based on spectral mixture analysis," in *Remote Geochemical Analysis: Elemental and Mineralogical Composition*, C. M. Pieters and P. A. Englert, Eds. Cambridge, U.K.: Cambridge Univ. Press, 1993, pp. 145–166.

[2] M. O. Smith, J. B. Adams, and D. E. Sabol, "Spectral mixture analysis—New strategies for the analysis of multispectral data," in *Image Spectroscopy—A Tool for Environmental Observations*, J. Hill and J. Mergier, Eds. Brussels and Luxembourg, Belgium: ECSC, EEC, EAEC, 1994, pp. 125–143.

[3] C. Brumbley and C.-I. Chang, "An unsupervised vector quantization-based target signature subspace projection approach to classification and detection in unknown background," *Pattern Recognit.*, vol. 32, pp. 1161–1174, July 1999.

[4] H. Ren and C.-I. Chang, "A generalized orthogonal subspace projection approach to unsupervised multispectral image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, pp. 2515–2528, Nov. 2000.

[5] 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.

[6] A. Hyvarinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Networks*, vol. 13, pp. 411–430, 2000.

[7] P. Comon, "Independent component analysis, a new concept?," *Signal Process.*, vol. 36, pp. 287–314, 1994.

[8] J. Karhunen, E. Oja, L. Wang, R. Vigário, and J. Joutsensalo, "A class of neural networks for independent component analysis," *IEEE Trans. Neural Networks*, vol. 8, pp. 486–504, May 1997.

[9] E. Oja, J. Karhunen, L. Wang, and R. Vigário, "Principal and independent components in neural networks—Recent developments," in *Proc. VII Italian Workshop Neural Networks WIRN*, Italy, May 1995, pp. 16–35.

[10] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, pp. 1129–1160, 1995.

[11] T. W. Lee, *Independent Component Analysis: Theory and Applications*. Norwell, MA: Kluwer, 1998.

[12] M. Girolami, *Self-Organizing Neural Networks: Independent Component Analysis and Blind Source Separation*. New York: Springer-Verlag, 2000.

[13] J. Bayliss, J. A. Gualtieri, and R. F. Cromp, "Analyzing hyperspectral data with independent component analysis," *Proc. SPIE*, vol. 3240, pp. 133–143, 1997.

[14] C. H. Chen and X. Zhang, "Independent component analysis for remote sensing study," in *EOS/SPIE Symp. Remote Sensing, Conference on Image and Signal Processing for Remote Sensing V*, vol. SPIE 3871, Florence, Italy, Sept. 20–24, 1999, pp. 150–158.

[15] H. Szu, "Independent component analysis (ICA): An enabling technology for intelligent information/image technology (IIT)," *IEEE Circuits Devices Mag.*, vol. 10, pp. 14–37, Dec. 1999.

[16] T. M. Tu, "Unsupervised signature extraction and separation in hyperspectral images: A noise-adjusted fast independent component analysis approach," *Opt. Eng.*, vol. 39, no. 4, pp. 897–906, 2000.

[17] S.-S. Chiang, C.-I. Chang, and I. W. Ginsberg, "Unsupervised hyperspectral image analysis using independent components analysis," in Proc. IEEE Int. Geoscience and Remote Sensing Symp., , Hawaii, July 24–28, 2000.

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

[19] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*, 2nd ed. New York: Academic, 1997.

[20] S. S. Chiang, C.-I. Chang, and I. W. Ginsberg, "Unsupervised subpixel target detection for hyperspectral images using projection pursuit," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, pp. 1380–1391, July 2001.

[21] S. Amari, "Natural gradient learning for over-and under-complete bases in ICA," *Neural Comput.*, vol. 11, pp. 1875–1883, 1999.

[22] J.-F. Cardoso and B. H. Laheld, "Equivalent adaptive source separation," *IEEE Trans. Signal Processing*, vol. 44, pp. 3017–3030, Dec. 1996.

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

[24] J. C. Harsanyi, "Detection and classification of subpixel spectral signatures in hyperspectral image Ssequences," Ph.D. dissertation, Dept. Electrical Engineering, Univ. Maryland-Balt. County, Baltimore, 1993.

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

[26] A. Cichocki and R. Unbehauen, *Neural Networks for Optimization and Signal Processing*. New York: Wiley, 1993, pp. 364–367.

[27] C.-I Chang and S.-S. Chiang, "Anomaly detection and classification for hyperspectral imagery," IEEE Trans. Geosci. Remote Sensing, to be published.

[28] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Dordrecht, The Nehterlands: Kluwer, 2002, to be published.

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

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

[31] D. Heinz and C.-I. Chang, "Fully constrained least squares linear mixture analysis for material quantification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 39, pp. 529–545, Mar. 2001.

[32] 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.

[33] S.-S. Chiang, *Automatic Target Detection and Classification in Hyperspectral Imagery*. Baltimore: Department of Computer Science and Electrical Engineering, Univ. Maryland-Balt. County, 2001.

[34] 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.

[35] A. Ifarraguerri and C.-I. Chang, "Multispectral and hyperspectral image analysis with projection pursuit," *IEEE Trans. Geosci. Remote Sensing*, vol. 38, pp. 2529–2538, Nov. 2000.

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 at National Tsing Hua University, Hsinchu, Taiwan, R.O.C., in 1975, and the M.A. degree from the State University of New York at Stony Brook in 1977, all in mathematics. He also 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) 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.

Dr. Chang is a member of SPIE, INNS, Phi Kappa Phi, and Eta Kappa Nu. He is currently Associate Editor of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and also on the editorial board of *Journal of High Speed Networks*, and was the Guest Editor of a special issue of this journal on Telemedicine and Applications.

Shao-Shan Chiang (S'98–M'01) received the B.S. degree in mathematical science from National Chengchi University, Taipei, Taiwan, R.O.C., in 1988, the M.S. degree in mathematics from National Taiwan University, Taipei, Taiwan in 1990, and the M.S. and Ph.D. degrees, both in electrical engineering, from University of Maryland Baltimore County (UMBC), Baltimore, MD, in 1999 and 2001, respectively.

From 1992 to 1996, he was a Software Engineer in Taipei. He was a Research Assistant in Anesthesiology Research Laboratory, the University of Maryland at Baltimore from 1996 to 1998, and was a Research Assistant in the Remote Sensing Signal and Image Processing Laboratory at UMBC from 1999 to 2001. Since 2001, he has been an Assistant Professor with the Department of Electrical Engineering, Lunghwa University of Science and Technology, Taoyuan, Taiwan. His research interests include signal processing, communications, pattern recognition, and remote sensing image processing.

Dr. Chiang is a member of SPIE.

James A. Smith (F'96) received the B.S. and M.S. degrees in mathematics and the Ph.D. degree in physics from the University of Michigan, Ann Arbor, and the M.S. degree in computer science from the Johns Hopkins University, Baltimore, MD.

As a NASA Goddard Space Flight Senior Fellow, he performs research in high-resolution thermal infrared sensing using the DOE Multispectral Thermal Imager and in hyperspectral sensing using NASA's EO-1 ALI and Hyperion systems.

Dr. Smith is a fellow of SPIE and the AAAS and currently serves as the Editor-in-Chief of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

Irving W. Ginsberg received the B.A. and Ph.D. degrees in theoretical physics from Wayne State University, Detroit, MI.

He is a Consultant to government and industry on the application of spectral and infrared technologies and algorithms. He was the Chief Scientist of the Department of Energy's Remote Sensing Laboratory until 2001. Previously, he worked for the University of Michigan's Willow Run Laboratories and the Environmental Research Institute of Michigan. Much of his work has been on the phenomenology of reflectance, and in the exploitation of multispectral and hyperspectral imagery.