Independent Component Analysis-Based Dimensionality Reduction With Applications in Hyperspectral Image Analysis

Jing Wang, *Student Member, IEEE,* and Chein-I Chang, *Senior Member, IEEE*

***Abstract*—In hyperspectral image analysis, the principal components analysis (PCA) and the maximum noise fraction (MNF) are most commonly used techniques for dimensionality reduction (DR), referred to as PCA-DR and MNF-DR, respectively. The criteria used by the PCA-DR and the MNF-DR are data variance and signal-to-noise ratio (SNR) which are designed to measure data second-order statistics. This paper presents an independent component analysis (ICA) approach to DR, to be called ICA-DR which uses mutual information as a criterion to measure data statistical independency that exceeds second-order statistics. As a result, the ICA-DR can capture information that cannot be retained or preserved by second-order statistics-based DR techniques. In order for the ICA-DR to perform effectively, the virtual dimensionality (VD) is introduced to estimate number of dimensions needed to be retained as opposed to the energy percentage that has been used by the PCA-DR and MNF-DR to determine energies contributed by signal sources and noise. Since there is no prioritization among components generated by the ICA-DR due to the use of random initial projection vectors, we further develop criteria and algorithms to measure the significance of information contained in each of ICA-generated components for component prioritization. Finally, a comparative study and analysis is conducted among the three DR techniques, PCA-DR, MNF-DR, and ICA-DR in two applications, endmember extraction and data compression where the proposed ICA-DR has been shown to provide advantages over the PCA-DR and MNF-DR.**

***Index Terms*—Dimensionality reduction (DR), ICA-DR/MNF-DR, independent component analysis (ICA), maximum noise fraction (MNF), PCA-DR, principal components analysis (PCA), virtual dimensionality (VD).**

I. Introduction

PRINCIPAL components analysis (PCA) is a widely used technique for dimensionality reduction (DR) and data compression [1]. It uses eigenvalues to determine the significance of principal components (PCs) so that DR is accomplished by selecting PCs in accordance with magnitude of their associated eigenvalues. Unfortunately, such PCA-DR may not be effective or appropriate for hyperspectral image analysis. A similar approach, called maximum noise fraction (MNF) [2] or noise-adjusted principal components (NAPC) transform [3] which was developed based on signal-to-noise ratio (SNR) also suffers from the same drawbacks as the PCA does. One major issue for both PCA and MNF is that many subtle material substances that are uncovered by very high spectral resolution hyperspectral imaging sensors cannot be characterized by second-order statistics. This may be due to the fact that the samples of such targets are relatively small and not sufficient to constitute reliable statistics. In this case, these targets may not be captured by the second-order statistics-based PCA/MNF in its PCs. Another key issue arising in the PCA/MNF is the determination of number of dimensions to be retained. A common criterion to resolving this problem is to calculate the accumulated sum of eigenvalues that represents a certain percentage of energy needed to be preserved. Unfortunately, as demonstrated in [4]–[6], this was not an effective measure. In order to address these two issues, this paper proposes an independent component analysis (ICA)-based approach to DR, referred to as ICA-DR for hyperspectral image analysis. Interestingly, using the ICA to perform DR has received little attention in the past due to the fact that the ICA was not developed for this purpose. When the ICA-DR is implemented, one immediate issue is how to rank independent components (ICs) in terms of significance since these ICs are generated by random initial projection vectors. Consequently, the ICs generated earlier are not necessarily more significant than those generated later. Unlike the PCA/MNF which prioritizes principal components in accordance with magnitude of eigenvalues or SNR, the ICA does not have such a criterion to prioritize the order that ICs are generated by the ICA. So, it must find a criterion higher than variance to measure the significance of each independent component (IC). The selection of an IC is then based on its score produced by the measure. Two measures, skewness and kurtosis are of interest and can be used to produce such score for each IC. As a special case, when the significance of an IC is measured by variance, the ICA is reduced to the PCA/MNF and the score of a component for significance is the eigenvalue associated with that particular component. Like the PCA/MNF, the ICA also encounters the same issue as does the PCA/MNF, that is, how many ICs are required to be retained without loss of significant information. Obviously, using the accumulated sum of eigenvalues as the PCA does is no longer an option for the ICA. In order to mitigate this dilemma, two approaches are proposed. One is to use a recently developed concept, virtual dimensionality (VD) developed in [5] and [6] to estimate the number of ICs for reduction. Another is to take advantage of the nature in randomness caused by the use of random initial projection vectors in the ICA. As a result, three algorithms are developed for the ICA-DR. One is called ICA-DR1 which

Manuscript received February 5, 2005; revised November 15, 2005.

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

Digital Object Identifier 10.1109/TGRS.2005.863297

is designed by using the VD in conjunction with a criterion for component prioritization and selection. Another is called ICA-DR2 which implements the ICA as a random algorithm with randomness characterized by random initial projection vectors. Accordingly, the ICA-DR2 automatically determines a desired set of ICs for DR without appealing for any criteria. A third algorithm is called ICA-DR3 which makes use of a custom-designed initialization algorithm in conjunction with the VD to generate an appropriate set of initial projection vectors to replace random projection vectors used by the ICA to produce each of ICs. As a consequence of the particularly selected initial projection vectors, the order of ICs generated by the ICA is no longer random. All the ICs are prioritized in accordance with the order of the selected initial projection vectors used by the ICA. Interestingly, each of the proposed three algorithms has its own advantages and all of them perform better than the second-order statistics-based DR such as the PCA-DR and MNF-DR. Additionally, according to conducted experiments, these three algorithms seem to produce nearly the same results.

In order to demonstrate the utility of the ICA in DR, two applications in hyperspectral image analysis are investigated. The first application is endmember extraction where the ICA replaces the PCA or MNF which has been used in two well-known endmember extraction algorithms, the pixel purity index (PPI) [7] and the N-finder algorithm [8] for DR. A second application is data compression where the PCA or MNF is replaced by the ICA to compress spectral information for hyperspectral imagery where the issues of subpixel detection and mixed pixel classification are investigated for DR. As shown in our experiments, the ICA-DR generally performs better than the PCA-DR and MNF-DR.

The remainder of this paper is organized as follows. Section II briefly reviews DR and the ICA to be used in this paper. Section III presents three new ICA-DR algorithms. Section IV conducts two application-based experiments to demonstrate the performance of the ICA-DR in comparison with the PCA-DR and MNF-DR. Section V summarizes our contributions and concludes some remarks.

II. Independent Component Analysis (ICA)

ICA has received considerable interest in recent years because of its versatile applications ranging from source separation, channel equalization to speech recognition and functional magnetic resonance imaging [9]. Its applications to linear mixture analysis for remote sensing images have been also found in [10]–[13]. The key idea of the ICA assumes that data are linearly mixed by a set of separate independent sources and demix these signal sources according to their statistical independency measured by mutual information. In order to validate its approach, an underlying assumption is that at most one source in the mixture model can be allowed to be a Gaussian source. This is due to the fact that a linear mixture of Gaussian sources is still a Gaussian source. More precisely, let $\mathbf{x}$ be a mixed signal source vector expressed by

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{1}$$

where $\mathbf{A}$ is an $L \times p$ mixing matrix and $\mathbf{s}$ is a p-dimensional signal source vector with p signal sources needed to be separated. The purpose of the ICA is to find a demixing matrix $\mathbf{W}$ that separates the signal source vector $\mathbf{s}$ into a set of sources which are statistically independent. Several different criteria have been proposed to measure source independency [9]. Nevertheless, they all originated from the concept of mutual information which is a criterion to measure the discrepancy between two random sources [14].

As a special case of (1), suppose that both $\mathbf{x}$ and $\mathbf{y}$ are zero-mean p-dimensional column random signal source vectors with covariance matrices $\mathbf{\Sigma_x} = (1/p)[\mathbf{x}\mathbf{x}^T]$ and $\mathbf{\Sigma_y} = (1/p)[\mathbf{y}\mathbf{y}^T]$, respectively. In order to decorrelate $\mathbf{x}$ in a similar fashion that the $\mathbf{x}$ is demixed in (1), a whitening matrix $\mathbf{\Sigma_x}^{-1/2}$ defined by the inverse of the square-root of the covariance matrix, $\mathbf{\Sigma_x}$ can be used to whiten the signal source vector $\mathbf{x}$. As a consequence, $\mathbf{\Sigma_y} = (\mathbf{\Sigma_x}^{-1/2})\mathbf{\Sigma_x}(\mathbf{\Sigma_x}^{-1/2})^T = \mathbf{I}$ and the resulting source vector $\mathbf{y}$ becomes an uncorrelated signal source vector in analogy with the signal source vector $\mathbf{s}$ in (1) to become a statistical independent source vector by a demixing matrix $\mathbf{W}$ found by the ICA via (1). In the light of this interpretation, (1) is reduced to

$$\mathbf{x} = \mathbf{\Sigma_x}^{\frac{1}{2}}\mathbf{y} \Rightarrow \mathbf{y} = \mathbf{\Sigma_x}^{-\frac{1}{2}}\mathbf{x} \tag{2}$$

where the mixing matrix $\mathbf{A}$ and the signal source vector $\mathbf{s}$ in (1) are replaced with the square root of the covariance matrix, $\mathbf{\Sigma_x}^{1/2}$ and an uncorrelated random source vector $\mathbf{y}$, respectively. By virtue of (2), the statistical independency measured by the ICA is reduced to the second-order statistics decorrelation by PCA. Accordingly, the ICA actually performs the PCA on the p-dimensional correlated signal source vector $\mathbf{x}$ via a whitening matrix $\mathbf{\Sigma_x}^{-1/2}$ to produce p uncorrelated PCs represented by the uncorrelated signal source vector $\mathbf{y}$ in terms of second-order statistics. The process of decorrelating the second-order statistics source vector $\mathbf{x}$ into an uncorrelated signal source vector $\mathbf{y}$ using (2) is generally referred to as whitening in signal processing and communications [15] with $\mathbf{\Sigma_x}^{-1/2}$ being used as a whitened matrix. The only difference between (1) and (2) is that the mixing matrix $\mathbf{A}$ in (1) is unknown as opposed to the covariance matrix, $\mathbf{\Sigma_x}^{1/2}$ in (2) which can be calculated directly from the observed signal source vector $\mathbf{x}$.

More interestingly, if we further interpret the mixed signal source vector $\mathbf{x}$, the mixing matrix $\mathbf{A}$, and $\mathbf{s}$ in (1) as a hyperspectral image pixel vector $\mathbf{r}$, an image endmember matrix $\mathbf{M}$ and an abundance vector α, respectively, (1) becomes

$$\mathbf{r} = \mathbf{M}\boldsymbol{\alpha} \tag{3}$$

which is exactly the linear mixture model used in hyperspectral image analysis with no noise term. An approach based on (3), called orthogonal subspace projection (OSP) was recently developed for DR in hyperspectral image classification [16]. In this particular case, the DR is performed by reducing the original data dimensionality to endmember dimensionality where each component is specified by a particular image endmember with appropriated estimated abundance fractions for hyperspectral image pixels.

III. ICA-DR

Over the past years, DR is generally performed by the PCA via (2). Interestingly, to the authors' best knowledge, there is little work of applying ICA to DR reported in the literature. One possible reason is that the ICA was not originally developed for the purpose of DR. A second reason may be that the similarity and relationship among the three equations, (1)–(3) have not been recognized. A third one is that the mixing matrices, $\mathbf{\Sigma}_{\mathbf{x}}^{1/2}$ in (2) and $\mathbf{M}$ in (3) are assumed to be known or can be generated directly from the data compared to the mixing matrix $\mathbf{A}$ in (1) which is totally unknown. Finally, unlike the PCA which prioritizes its generated principal components according to the magnitude of eigenvalues, there is no specific criterion to rank components produced by the ICA. Since the ICA is a well-established technique, we will only focus on the issues described above that arise in DR.

In order to implement the ICA, the FastICA algorithm developed by Hyvarinen and Oja [9] was used to find ICs where the deflation approach was applied to generate ICs one by one sequentially and each of ICs is produced by maximizing the negentropy measured by kurtosis. Typically, non-Gaussianity can be measured by the absolute value of kurtosis. But, the kurtosis also has some drawbacks as well. It is very sensitive to outliers [9, p. 182] where a single sample can make the kurtosis very large. To alleviate this problem, some other criteria such as negentropy is introduced as a measure for non-Gaussianity. There are several approximations to negentropy using various other nonlinear functions, such as skewness, tanh etc. According to our applications in hyperspectral target detection, the ICs of major interest are generally super-Gaussian which is usually caused by outliers. In this case, the kurtosis seems to be an appropriate criterion to be used to generate ICs. There is another symmetric approach [9, p. 194] that can be used to find ICs. However, this approach does not offer any advantage over the deflation approach in our applications. Therefore, it is not considered in this paper.

For each spectral band image, it was converted to a vector. More specifically, assume that a hyperspectral image cube has size of $M \times N \times L$ where L is the number of spectral bands and MN is the size of each spectral band image. The hyperspectral image cube can then be represented by a data matrix $\mathbf{X}$ of size $L \times MN$ with L rows and MN columns. In other words, each row in the data matrix $\mathbf{X}$ is specified by a particular spectral band image. As a result, a total of L ICs can be generated by the FastICA. However, as noted in (3), a hyperspectral image pixel can be generally considered as a linear mixture of a set of known image endmembers where the number of endmembers, p is generally much smaller than L, the number of spectral bands. In this case, when the DR is performed, only p ICs are required and there is no need of producing all ICs for image analysis. But, it also gives rise to an issue on that which p ICs must be selected for DR.

According to the introduction, there is a need for the ICA-DR to address problems that cannot be resolved by either the PCA-DR or MNF-DR. However, there are some major issues to implement the ICA for DR, in our case, the FastICA for DR. First of all, the FastICA-generated ICs are not necessarily in order of information significance as the way that PCs are generated by the PCA or the MNF in accordance with decreasing magnitude of eigenvalues or SNRs. Another is that ICs generated by the FastICA in different runs do not necessarily appear in the same order. These issues are primarily due to the nature that the initial projection unit vectors used to produce ICs by the FastICA are randomly generated. Therefore, an IC generated earlier by the FastICA is not necessarily more significant than one generated later.

In order to resolve the issue on the use of the random initial projection unit vectors, three algorithms, called ICA-DR1, ICA-DR2 and ICA-DR3 are developed for DR using the FastICA. In ICA-DR1, we consider each generated IC as a random variable. In light of this interpretation, we assume that the ith component, IC_i can be described by a random variable ζ_i with values taken by the gray level value of the nth pixel in the IC_i, denoted by z_n^i. In this case, the FastICA-generated ICs can be ranked and prioritized by high-order statistics-based criteria.

Unlike the ICA-DR1, the ICA-DR2 considers the FastICA as a random algorithm with randomness caused by nature of random initial unit projection vectors used by the FastICA. The idea of the ICA-DR2 is to run the FastICA a number of times to produce sample average of all ICs where the ICs common in all runs will be considered as significant ICs and used for DR. By contrast, the ICR-DR3 takes a complete opposite approach to the ICA-DR1 and the ICA-DR2. In order to remove the random nature caused by the initial projection unit vectors used by the FastICA, the ICA-DR3 custom-designs a set of initial vectors to initialize the FastICA to produce each of ICs. Consequently, the ICs are always generated in a certain order and never appear in a random order as generated by the ICA-DR1 and the ICA-DR2.

A. ICA-DR1

The idea of the ICA-DR1 is to first determine the number of ICs needed to be retained, p which can be estimated by the VD. It then prioritizes the FastICA-generated ICs using a high-order statistics criterion to select the first p prioritized ICs.

ICA-DR1 Algorithm:

1) Use the VD to determine the number of dimensions p required to be retained.
2) Use the FastICA to find $2p$ ICs, $\{\mathrm{IC}_i\}_{i=1}^{2p}$. It should be noted that for each IC the FastICA randomly generates a unit vector as an initial projection vector to produce the final desired projection vector for that particular component.
3) Calculate the following criterion for IC_i that combines third and fourth orders of statistics for ζ_i

$$J(\mathrm{IC}_i) = \left(\frac{1}{12}\right)\left[\kappa_i^3\right]^2 + \left(\frac{1}{48}\right)\left[\kappa_i^4 - 3\right]^2 \tag{4}$$

where $\kappa_i^3 = E[\zeta_i^3] = (1/MN)\sum_{n=1}^{MN}(z_n^i)^3$ and $\kappa_i^4 = E[\zeta_i^4] = (1/MN)\sum_{n=1}^{MN}(z_n^i)^4$ are sample means of third and fourth orders of statistics in the IC_i. It should be note that (4) is taken from [9, Eq. (5.35), p. 115],

which is used to measure the negentropy by high-order statistics.

4) Prioritize the $\{\mathrm{IC}_i\}_{i=1}^{2p}$ in accordance with the magnitude of $J(\mathrm{IC}_i)$.
5) Select those ICs with the first p largest $J(\mathrm{IC}_i)$ to perform DR.

It should be noted that the ICA-DR1 is supposed to run and prioritize all the ICs, then selects the first p prioritized ICs. However, in practice this is not necessary. According to our experiments, the VD can be used to set an upper bound on the number of ICs required to be generated by the FastICA without having the FastICA run through all ICs. A good upper bound is empirically shown to be twice the VD to avoid small targets being left out.

B. ICA-DR2

The idea of ICA-DR2 is to run the FastICA a number of times. Here we define a run by running the FastICA a single time. Since the initial projection vector randomly generated by the FastICA for each run is different, the order of the generated ICs is also different. Nevertheless, if the information contained in an IC is significant, such an IC will always appear in each run. With this assumption, if we keep running the FastICA to find common ICs over all the runs until common ICs remain unchanged, in which case the process is terminated. The ICs that are common in all runs are the desired ICs for DR. The detailed implementation of the ICA-DR2 is summarized as follows.

ICA-DR2 Algorithm:

1) Initialization: Set $n = 0$ and $\mathrm{VD} = p$.
2) At each n, run the FastICA to find $2p$ ICs, $\{\mathrm{IC}_i^{(n)}\}_{i=1}^{2p}$ where each IC, $\mathrm{IC}_i^{(n)}$ can be formed as a vector, denoted by $\mathbf{v}_i^{(n)}$. It should be noted that the FastICA randomly generates a unit vector as an initial projection vector.
3) If $n < 1$, $n \leftarrow n + 1$ and go to step 2. Otherwise, continue.
4) Find common ICs for all runs up to nth run. Two ICs for different runs, $\mathrm{IC}_i^{(\overline{n})}$ and $\mathrm{IC}_j^{(\widetilde{n})}$ are considered to be distinct if the spectral angle mapper (SAM) between their corresponding vectors, $\mathrm{v}_i^{\overline{n}}$ and $\mathrm{v}_j^{\widetilde{n}}$ is greater than a prescribed threshold ε. Let $\cap_{m=0}^{n}\{\mathrm{IC}_k^{(m)}\}_{k=1}^{2p}$ denote the common ICs obtained for all runs, $0 \le m \le n$.
5) $\cap_{m=0}^{n-1}\{\mathrm{IC}_k^{(m)}\}_{k=1}^{2p} \neq \cap_{m=0}^{n}\{\mathrm{IC}_k^{(m)}\}_{k=1}^{2p}$, go to step 2. Otherwise, the algorithm is terminated and $\cap_{m=0}^{n}\{\mathrm{IC}_k^{(m)}\}_{k=1}^{2p}$ is the desired set of ICs for DR.

It is worth noting that ICA-DR2 does not really need the VD in step 1 as was required by the ICA-DR1 since the ICA-DR2 generally becomes stable and converges very rapidly after a few runs. Like the ICA-DR1, the ICA-DR2 does not have to run through all the ICs to find the common ICs. So, a similar strategy used for the ICA-DR1 can be also applied to the ICA-DR2 by setting an upper bound on the number of ICs required to be generated. A good upper bound for the ICA-DR2 is also twice the VD which is empirically shown by our experiments.

C. ICA-DR3

In ICA-DR1 and ICA-DR2, the initial projection unit vector used by the FastICA to produce each of ICs is generated randomly. Therefore, the ICs produced by the FastICA in different runs generally appear in different orders. When it comes to DR, this becomes a serious issue because an IC appear earlier is not necessarily more important or significant than an IC produced later. In order to resolve this issue, a third ICA-DR algorithm, ICA-DR3 is developed where the initial projection unit vector used to produce each of ICs are selected in a specific manner so that all the ICs will always appear in a fixed order rather than a random order as produced by the ICA-DR1 or the ICA-DR2. As a consequence, there is no need of using (4) to prioritize ICs as the way the ICA-DR 1 does or it requires the FastICA to run a number of times with different random orders as the way the ICA-DR2 does to find the common ICs. A major advantage of using the ICA-DR3 is that all ICs always appear in the same order regardless of how many runs the FastICA is implemented. An algorithm proposed to be used in the ICA-DR3 to generate a set of initial projector unit vectors is called automatic target generation process (ATGP) which is derived from the automatic target detection and classification algorithm in [4] and [17].

ICA-DR3 Algorithm:

1) Use the VD to determine the number of dimensions, p, required to be retained.
2) Perform sphereing on the data matrix $\mathbf{X}$ and let the resulting sphered data matrix be denoted by $\hat{\mathbf{X}}$.
3) Apply the ATGP to $\hat{\mathbf{X}}$ to find p target pixel vector, $\{\mathbf{t}_i\}_i^p$.
4) Use the FastICA to find p ICs, $\{\mathrm{IC}_i\}_{i=1}^{p}$ where the ith IC_i is generated by the FastICA with the ith target pixel vector chosen to be the initial projection vector instead of being generated randomly.

Two comments on the ICA-DR3 are noteworthy.

a) There is a good reason to choose the ATGP to produce initial projection unit vectors for each of ICs. This is because the ATGP generates target pixels by a sequence of OSP which is also used in the FastICA to produce a sequence of ICs. Therefore, the ATGP-generated target pixels, $\{\mathbf{t}_i\}_i^p$, are mutual orthogonal each other. This implies that one target pixel used as an initial projection vector to generate an IC will not be used again as an initial projection vector to generate other ICs.
b) It may seem to be intuitive to use eigenvectors as initial projection vectors to generate ICs. Unfortunately, according to our experiments this approach does not always work and results were not consistent. On the other hand, the ATGP used in the ICA-DR3 worked consistently well and better than the use of eigenvectors. Therefore, the ATGP was chosen in the ICA-DR3.

As a concluding remark for Section III, it is worthwhile having a discussion on the three algorithms developed in this section. First of all, the three developed algorithms are completely unsupervised with no need of human intervention. In

TABLE I
COMPARISON AMONG THREE ALGORITHMS, ICA-DR1, ICA-DR2, AND ICA-DR3

	ICA-DR1	ICA-DR2	ICA-DR3
Number of ICs required to generate	$2p$	$2p$	p
Number of ICs required to select	p	automatic	p
Initial condition	random	random	ATGP
Criterion for IC prioritization	High-order statistics	no	ATGP
Criterion for IC selection	priority	automatic	ATGP
Computation	low	moderate-high	low
Algorithm used to generate ICA	FastICA	FastICA	FastICA

particular, there is no try-out by users. These three algorithms are also based on completely different design rationales and each of them deserves its own merit. The ICA-DR1 prioritizes ICs according to a high-order statistics-based criterion which is a combination of the third order statistics, skewness and the fourth order statistics, kurtosis. On the other hand, the ICA-DR2 does not rely on any criterion. Instead, it runs itself as if a random algorithm with an arbitrary set of random initial projection vectors in a given run. The algorithm is then terminated when different runs of the FastICA produce common ICs. As a result of such a random nature, the ICA-DR2 requires the FastICA to run in a number of times with different sets of randomly generated projection vectors. In this case, the VD is only used to estimate an upper bound on the number of ICs that the FastICA must generate in each run. Compared to the ICA-DR1 and ICA-DR2, both of which use randomly generated vectors as initial projection vectors, the ICA-DR3 takes a complete opposite approach. It makes use of an initialization algorithm, called ATGP to produce an appropriate set of initial projection vectors to be used for the FastICA. Since the targets generated by the ATGP are spectrally distinct in terms of OSP, each of these ATGP-generated targets represents one type of signal sources. The FastICA then uses these ATGP-generated targets as initial projection vectors to make sure that these ATGP-generated target sources are separated in individual ICs so and no two or more ATGP-generated target sources are present in one single IC. Additionally, all the ICs will be also prioritized in a simple way by the same order that the ATGP-generated targets are generated. One salient difference among the three ICA-DR algorithms is the number of components required by the FastICA to generate. The ICA-DR1 and the ICA-DR2 generally require to generate all the ICs before IC selection. However, it is found empirically that twice the VD, $2p$ provides a good upper bound on the number of ICs required to be generated. Compared to the ICA-DR1 and the ICA-DR2, the ICA-DR3 only has to generate p ICs without performing IC selection as do the ICA-DR1 and ICA-DR2. So, from a computational complexity point view, the ICA-DR3 has least computing time and the ICA-DR2 may suffer from highest computation since it must run the FastICA in a number of runs before the algorithm is terminated. Moreover, the ICA-DR1 produces and then selects the first p ICs prioritized by (4). Like the ICA-DR1, the ICA-DR2 also produces $2p$ ICs. Its difference from the ICA-DR1 is that the ICA-DR2 repeatedly runs the FastICA with different sets of random initial projection vectors and in the mean time it also finds their common set of ICs in all runs until the two consecutive runs produce the same common set of ICs. So, there is no need for ICA-DR2 to select or prioritize ICs. To the contrary, the ICA-DR3 generates p ICs by the FastICA using a specific set of the p targets generated by the ATGP as its initial projection vectors. In this case, the ICs are prioritized and selected by the order that the p ATGP-generated targets appear. As expected, the running time required for the ICA-DR3 is generally less than that for the ICA-DR1 and the ICA-DR2 because the latter must generate $2p$ ICs, while the former always takes advantage of generating ICs one by one in sequence. In next section, we will show by experiments that both ICA-DR1 and ICA-DR3 are very suitable to DR with applications in data compression and endmember extraction. Finally, we summarize comparison among these three algorithms in Table I where the value of the p is estimated by the VD.

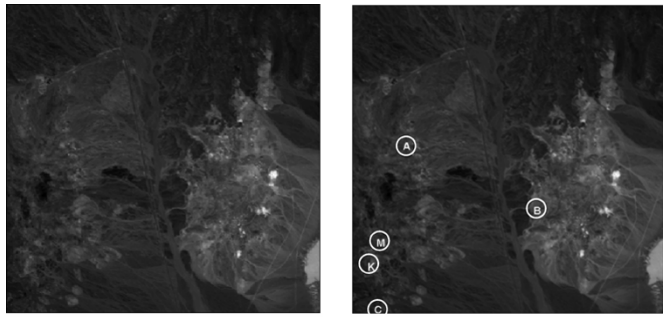


Fig. 1. (a) Spectral band number 50 (827 nm) of the Cuprite AVIRIS image scene. (b) Spatial positions of five pure pixels corresponding to minerals: alunite (A), buddingtonite (B), calcite (C), kaolinite (K), and muscovite (M).

IV. EXPERIMENTS

In order to demonstrate the utility of the ICA-DR in hyperspectral image analysis, two applications are considered, endmember extraction and data compression. Two sets of real hyperspectral image data were used for experiments, which are Airborne Visible Infrared Imaging Spectrometer (AVIRIS) Cuprite data and HYperspectral Digital Image Collection Experiment (HYDICE) data.

A. Endmember Extraction

The first application is endmember extraction which is one of fundamental tasks in hyperspectral image analysis. It finds and identifies the purest signatures in image data. Since it requires intensive computing process, DR is generally performed prior to endmember extraction. For example, two most widely used

TABLE II
VD ESTIMATES FOR THE CUPRITE SCENE IN FIG. 1 WITH VARIOUS FALSE ALARM PROBABILITIES

	$P_F=10^{-1}$	$P_F=10^{-2}$	$P_F=10^{-3}$	$P_F=10^{-4}$	$P_F=10^{-5}$
VD	34	30	24	22	20

TABLE III
(a) THREE ENDMEMBERS EXTRACTED BY PPI WITH PCA-DR. (b) THREE ENDMEMBERS EXTRACTED BY PPI WITH MNF-DR. (c) FIVE ENDMEMBERS EXTRACTED BY PPI WITH ICA-DR1. (d) FOUR ENDMEMBERS EXTRACTED BY PPI WITH ICA-DR2. (e) FOUR ENDMEMBERS EXTRACTED BY PPI WITH ICA-DR3

	A (62,161)	B (209,234)	C (30,347)	K (22,298)	M (33,271)
A' (291,234)	0.0395 (1)	0.1675	0.2054	0.1076	0.1463
C' (42,217)	0.2184	0.1025	0.0458 (1)	0.2296	0.1277
K' (22,299)	0.1076	0.1789	0.2177	0.022 (2)	0.1305

(a)

	A (62,161)	B (209,234)	C (30,347)	K (22,298)	M (33,271)
B' (68,135)	0.1643	0.0726 (1)	0.1035	0.1928	0.1243
C' (43,216)	0.2333	0.1157	0.0837 (1)	0.2403	0.1403
M' (19,280)	0.1651	0.0844	0.1162	0.1508	0.0809 (1)

(b)

	A (62,161)	B (209,234)	C (30,347)	K (22,298)	M (33,271)	Order of IC
A' (206,217)	0.0918 (1)	0.0991	0.1504	0.1194	0.1033	18
B' (207,235)	0.1555	0.0203 (1)	0.0933	0.1663	0.0867	3
C' (11,156)	0.2115	0.0967	0.0350 (1)	0.2136	0.1170	10
K' (24,305)	0.0967	0.1695	0.2138	0.0314 (1)	0.1309	5
M' (265,137)	0.1047	0.0952	0.1375	0.0996	0.0589 (1)	7

(c)

	A (62,161)	B (209,234)	C (30,347)	K (22,298)	M (33,271)	Order of IC
B' (205,216)	0.1190	0.0571 (1)	0.1171	0.1419	0.0876	4
C' (18,350)	0.2087	0.1059	0.0280 (1)	0.2109	0.1104	9
K' (22,305)	0.0977	0.1801	0.2212	0.0240 (1)	0.1377	8
M' (23,280)	0.0975	0.1110	0.1313	0.1045	0.0765 (1)	10

(d)

	A (62,161)	B (209,234)	C (30,347)	K (22,298)	M (33,271)	Order of IC
B' (209,235)	0.1324	0.0383 (1)	0.1212	0.1529	0.0962	4
C' (30,349)	0.2096	0.1043	0.0317 (1)	0.2110	0.1119	18
K' (22,304)	0.1030	0.1619	0.2026	0.0300 (2)	0.1183	6
M' (264,138)	0.1008	0.1093	0.1538	0.0936	0.0692 (2)	8

(e)

endmember extraction methods, PPI and N-FINDR algorithm implement either PCA or MNF for DR to reduce computational complexity. In this section, these two algorithms were used as benchmark comparative analysis where the five DR techniques, PCA-DR, MNF-DR, ICA-DR1, ICA-DR2 and ICA-DR3 are evaluated for performance.

1) AVIRIS Cuprite Data: The first image data was collected over the Cuprite mining site, Nevada, in 1997 and shown in Fig. 1(a). It is a 224-band AVIRIS image scene with size of 350 $\times$ 350 pixels and well understood mineralogically, and has reliable ground truth in the form of a library of mineral spectra, collected at the site by USGS available at website[1] where the five minerals, alunite (A), buddingtonite (B), calcite (C), kaolinite (K), and muscovite (M) are specified by pixels white-circled and labeled by A, B, C, K, and M in Fig. 1(b) according to the ground truth provided by the USGS. This fact has made this scene a standard test site for endmember extraction. It should be noted that bands 1–3, 105–115, and 150–170 have been removed prior to the analysis due to water absorption and low SNR in those bands. As a result, a total of 189 bands were used for experiments.

The VD estimated for this image scene with different values of false alarm probability P_F is given in Table II where the Harsanyi–Farrand–Chang (HFC) method developed in [5] and [6] was used for VD estimation. For those who are interested in the HFC method, a brief description of it is provided in the Appendix.

In the following experiments, the VD was chosen to be 22 with the false alarm probability set to $P_F = 10^{-4}$. In order to demonstrate the performance of ICA-DR in comparison with the PCA-DR and the MNF-DR in applications of endmember extraction, the two well-known endmember extraction algorithms, PPI [7] and N-FINDR algorithm [8] were implemented with DR to 22 components. Using the ground truth provided by Fig. 1(b) and the spectral angle mapper (SAM) as the spectral similarity measure for signature identification, Table III(a)–(e)

[1][Online]. Available: http://speclab.cr.usgs.gov/cuprite.html.

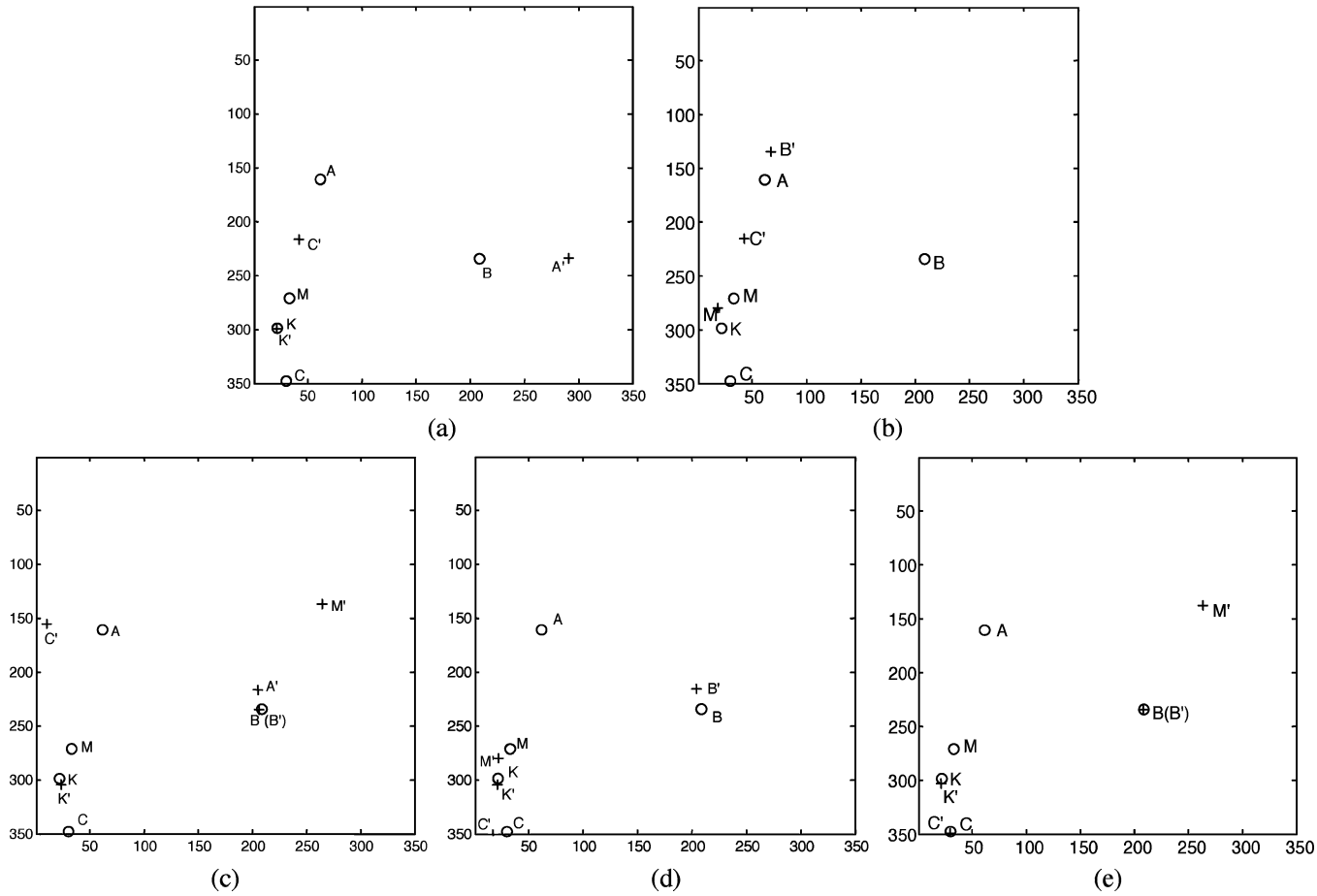


Fig. 2. Spatial locations of ground truth endmembers and PPI-extracted endmembers. (a) PCA-DR. (b) MNF-DR. (c) ICA-DR1. (d) ICA-DR2. (e) ICA-DR3.

tabulates endmembers extracted by the PPI with five different DR techniques, PCA-DR, MNF-DR, ICA-DR1, ICA-DR2, and ICA-DR3, where the first row lists five ground truth endmembers with their spatial coordinates and the first column lists the PPI-extracted endmembers with their spatial coordinates. The values in the tables were produced by SAM between the PPI-extracted endmembers and ground truth endmembers, and the numbers in parentheses are the scores produced by the PPI where the shade is used to highlight the identification results. Additionally, the numbers in the last column of Table III(c)–(e) indicate the order of the IC that extracts its corresponding endmembers in the first column and the primes "′" were used in the first column to indicate that the found mineral pixels were not the same ground truth pixels marked by white circles in Fig. 1(b).

According to the identification results in Table III(a)–(e), ICA-DR1 performed the best and was the only one successfully extracted all the five mineral signatures, while the PCA-DR and MNF-DR were the worst and extracted only three endmembers. The ICA-DR2 and ICA-DR3 were right in between and extracted four endmembers with missing the mineral "A." It is also worth noting that the ICA-DR generally extracted pixels more pure than PCA-DR and MNF-DR did except the alunite in Table III(a) and Table III(c).

In order to compare the locations of PPI-extracted endmembers against that of the ground truth endmembers, Fig. 2(a)–(e) shows their respective spatial locations produced by the PPI using the five DR techniques where the locations of the ground truth endmembers are marked by circles and the locations of the PPI-extracted endmembers are marked by crosses. From Fig. 2(a)–(e) and Table III(a)–(e), the endmembers extracted by the PPI using the five DR techniques were generally not the same pixels specified by the ground truth, but their signatures were very close in terms of SAM.

Finally, for the purpose of endmember detection and classification, Figs. 3 and 4 show all the 22 components produced by the PCA-DR and MNF-DR, and Figs. 5–7 only show the components produced by ICA-DR1, ICA-DR2, and ICA-DR3, respectively, in which the minerals corresponding to Table III(a)–(e) were found to be present where the number underneath each of components indicates the order of components generated by DR techniques.

Since the PCA-DR and MNF-DR are generally designed for information preservation, not designed for detection and classification, it is nearly impossible to conduct such analysis by visual inspection without appealing for a spectral measure. Therefore, Figs. 3 and 4 include all the 22 components to demonstrate the difficulty with identifying endmembers of interest. On the other hand, Figs. 5–7 show otherwise. The ICA-generated components not only can be used for ednember extraction, but also can be used for endmember detection and classification where different minerals were detected and extracted in individual and

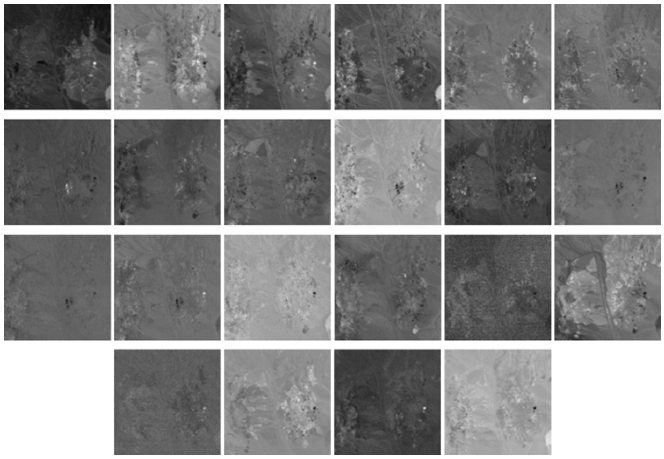


Fig. 3. Twenty-two PCs produced by the PCA-DR.

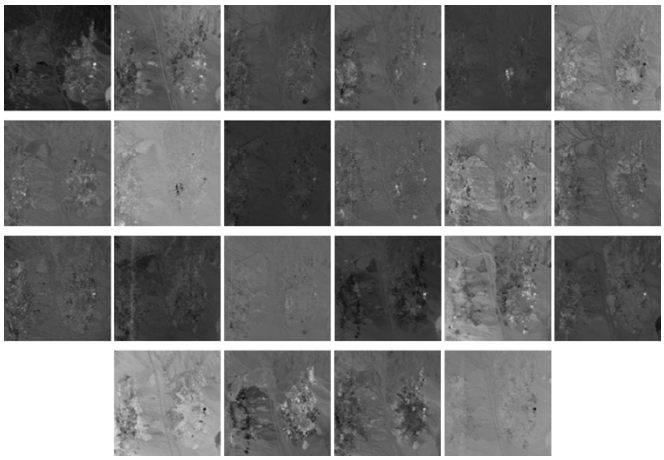


Fig. 4. Twenty-two component images produced by the MNF-DR.

separate components for classification. It should be noted that Fig. 7 included one extra IC labeled by (a) that was found to contain the mineral "A" for which the PPI failed to extract according to Table III(e). Interestingly, this mineral "A" could be extracted by the N-FINDR algorithm as shown below.

Similarly, the N-FINDR algorithm was also implemented for endmember extraction via the five DR techniques. Table IV(a)–(e) tabulates endmembers extracted by the N-FINDR algorithm with five different DR techniques, PCA-DR, MNF-DR, ICA-DR1, ICA-DR2, and ICA-DR3, where the first row lists five ground truth endmembers with their spatial coordinates and the first column lists the N-FINDR algorithm-extracted endmembers with their spatial coordinates. The values in the tables were produced by SAM between the N-FINDR algorithm-extracted endmembers and ground truth endmembers and the shade is used to highlight the identification results. The numbers in the last column of Table IV(c)–(e) indicate the order of the IC that extracts its corresponding endmembers in the first column. Also, primes "′" were used in the first column to indicate the found mineral pixels were not the same ground truth pixels marked by white circles in Fig. 1(b).

According to Table IV, the best ones were PCA-DR, ICA-DR1 and ICA-DR3 which extracted all the five minerals, while the other two, MNF-DR and ICA-DR2 failed to extract one mineral. Comparing Table IV to Table III, the N-FINDR algorithm using the ICA-DR2 missed the same mineral as did the PPI using the ICA-DR2. However, this was not the case for the MNF-DR which missed the minerals "A" and "K" with the PPI and missed only mineral "B" with the N-FINDR algorithm. Our experiments demonstrated that the PCA-DR and MNF-DR may not be consistent if different endmember extraction algorithms are used. Additionally, the experiments also showed an interesting, yet surprising finding that MNF-DR was generally not as good as PCA-DR for endmember extraction.

Like Fig. 2(a)–(e), Fig. 8(a)–(e) plots the spatial locations produced by the N-FINDR algorithm using the five DR techniques where the locations of the ground truth endmembers are marked by circles and the locations of the PPI-extracted endmembers are marked by crosses.

Fig. 8(a)–(e) and Table IV(a)–(e) also concluded that the endmembers extracted by the N-FINDR algorithm using the five DR techniques were generally not the same pixels as those specified by the ground truth, but their signatures were very close in terms of SAM. It is worth noting that the same ICs in Figs. 5–7 were used for the N-FINDR algorithm for endmember extraction.

2) HYDICE Data: The second data used for endmember extraction was the HYDICE image shown in Fig. 9(a), which has size of 64 × 64 pixels with 15 panels in the scene. Within the scene there has a large grass field background, a forest on the left edge and a barely visible road running on the right edge of the scene. It was acquired by 210 spectral bands with a spectral coverage from 0.4 to 2.5 μm. 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. The spatial resolution is 1.56 m and spectral resolution is 10 nm.

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 $i = 1, 2, \ldots, 5$, the three panels p_{i1}, p_{i2}, p_{i3} were painted by the same material but have three different sizes. For each column $j = 1, 2, 3$, the five panels p_{1j}, p_{2j}, p_{3j}, p_{4j}, p_{5j} have the same size but were painted by five different materials. It should be noted that the panels in rows 2 and 3 are made by the same material with different paints, so did the panels in rows 4 and 5. Nevertheless, they were still considered as different materials. The sizes of the panels in the first, second and third columns are 3 × 3 m, 2 × 2 m, and 1 × 1 m, respectively. So, the 15 panels have five different materials and three different sizes. Fig. 9(b) shows the precise spatial locations of these 15 panels where red pixels (R pixels) are the panel center pixels and the pixels in yellow (Y pixels) are panel pixels mixed with background. The 1.56-m spatial resolution of the image scene suggests that most of the 15 panels are one pixel in size except that p_{21}, p_{31}, p_{41}, p_{51} which are two-pixel panels. Since the size of the panels in the third column is 1 × 1 m, they cannot be seen visually from Fig. 9(a) due to the fact that its size is less than the 1.56-m pixel resolution. With the ground truth in Fig. 9(b), this 15-panel HYDICE image scene provides another excellent example for experiments where the signatures of the pure R pixels in first and second columns can

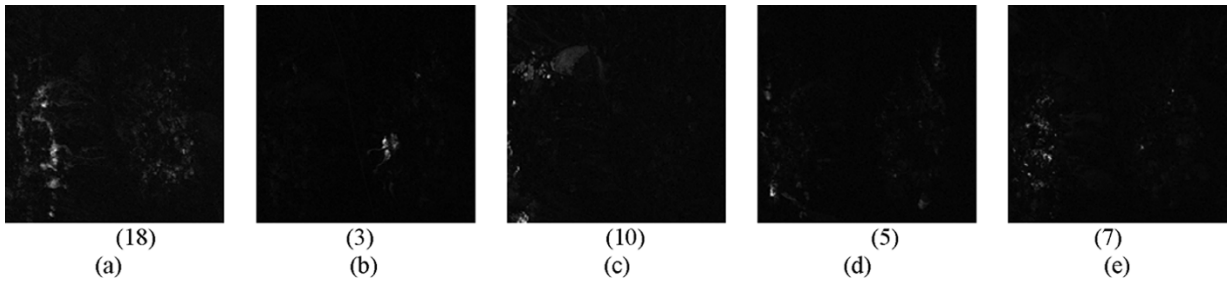


Fig. 5. ICs produced by the ICA-DR1, in which the five minerals, A, B, C, K, and M were present. (a) A (alunite). (b) B (buddingtonite). (c) C (calcite). (d) K (kaolinite). (e) M (muscovite).

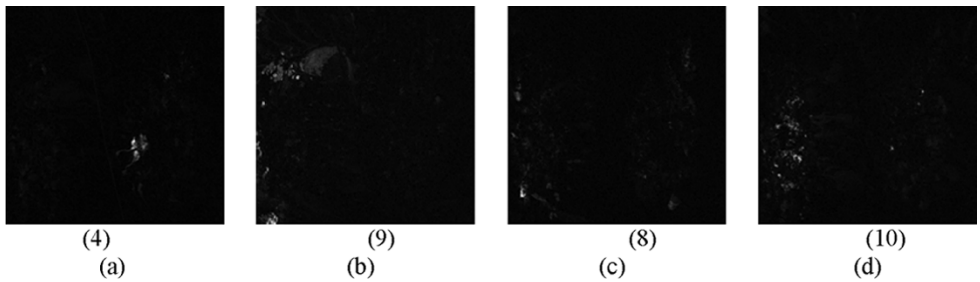


Fig. 6. ICs produced by the ICA-DR2 in which four minerals, B, C, K, and M were present. (a) B (buddingtonite). (b) C (calcite). (c) K (kaolinite). (d) M (muscovite).

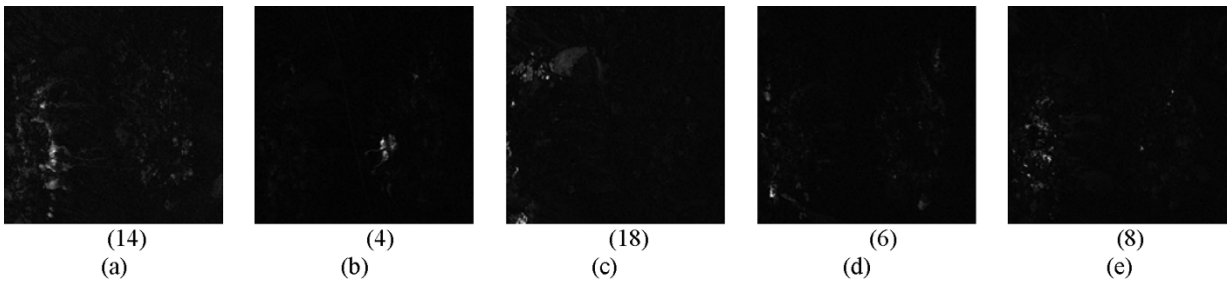


Fig. 7. ICs produced by the ICA-DR3 in which five minerals A, B, C, K, and M were present. (a) A (alunite). (b) B (buddingtonite). (c) C (calcite). (d) K (kaolinite. (e) M (muscovite).

be considered as endmembers. The VD for this image scene was estimated in Table V.

For our experiments, the VD was chosen to be 9 with the false alarm probability set to $P_F = 10^{-4}$. Figs. 10–13 show the nine components obtained by the PCA, MNF, ICA-DR1, ICA-DR2, and ICA-DR3, respectively, where the upper bound on the number of ICs was set to 18. Since both ICA-DR1 and ICA-DR2 produced nearly the same nine ICs shown in Fig. 12, we used the notation "ICA-DR1(2)" to indicate both ICA-DR1 and ICA-DR2.

The results in Figs. 10–13 provided more clear evidence than those in Figs. 3–7 in that the second order statistics-based DR techniques, PCA and MNF preserved most of the image background as opposed to the statistical independence-based ICA which retained panels of interest while discarding the image background. The reason for this is largely due to the fact that the image background is generally characterized by second order statistics rather than high-order statistics. Since the results of implementing the PPI using the nine components in Figs. 10–13 are available in [18], Figs. 14–17 only show endmembers extracted by the N-FINDR algorithm using the 9 components in Figs. 10–13, respectively.

It is worth noting that instead of using tables (Tables III and IV) as we did for the Cuprite data, we have used images to better demonstrate the experimental results for visual inspection where the endmembers extracted by the N-FINDR algorithm in these figures were exactly R pixels in the first column of the ground truth map in Fig. 9(b). Comparing Figs. 15–17 to Figs. 13 and 14, it clearly shows that using the N-FINDR algorithm with the ICA-DR performed significantly better than using the N-FINDR algorithm with the PCA-DR and the MNF-DR in the sense that the former extracted all the five distinct R panel pixels compared to the latter only extracted three and two distinct R panel pixels, respectively. Similar results were also obtained in [18] by the PPI with the ICA-DR, PCA-DR, and MNF-DR.

B. Data Compression

One of major applications for DR is data compression. The PCA has been commonly used for DR. Until recently, the MNF began to emerge as another alternative for DR in hyperspectral image analysis. Both the PCA and MNF are considered as second order statistics-based transforms. Unfortunately, in many applications, preserving information of second-order

TABLE IV
(a) FIVE ENDMEMBERS EXTRACTED BY N-FINDR ALGORITHM WITH PCA-DR. (b) FOUR ENDMEMBERS EXTRACTED BY N-FINDR ALGORITHM WITH MNF-DR. (c) FIVE ENDMEMBERS EXTRACTED BY N-FINDR ALGORITHM WITH ICA-DR1. (d) FOUR ENDMEMBERS EXTRACTED BY N-FINDR ALGORITHM WITH ICA-DR2. (e) FIVE ENDMEMBERS EXTRACTED BY N-FINDR ALGORITHM WITH ICA-DR3

	A (62,161)	B (209,234)	C (30,347)	K (22,298)	M (33,271)
A' (286,237)	0.0981	0.1884	0.2124	0.1387	0.1667
B' (206,228)	0.1247	0.0422	0.1143	0.1395	0.0845
C' (10,342)	0.2490	0.1314	0.0725	0.2426	0.1381
K' (23,298)	0.1098	0.1524	0.1918	0.0613	0.1079
M' (19,281)	0.1458	0.0781	0.1141	0.1347	0.0706

(a)

	A (62,161)	B (209,234)	C (30,347)	K (22,298)	M (33,271)
A '(80,234)	0.0235	0.1665	0.2143	0.1012	0.1542
C' (19,104)	0.2235	0.1003	0.0511	0.2276	0.1222
K' (17,148)	0.0812	0.1435	0.1771	0.0418	0.1011
M' (33,274)	0.1675	0.0934	0.0971	0.1483	0.0382

(b)

	A (62,161)	B (209,234)	C (30,347)	K (22,298)	M (33,271)	Order of IC
A' (62,162)	0.0434	0.1495	0.1961	0.0865	0.1320	18
B'(234,139)	0.1591	0.0750	0.0793	0.1768	0.0969	3
C' (25,348)	0.2192	0.1069	0.0362	0.2196	0.1174	10
K' (23,305)	0.0889	0.1835	0.2284	0.0341	0.1453	5
M' (19,280)	0.1651	0.0844	0.1162	0.1508	0.0809	7

(c)

	A (62,161)	B (209,234)	C (30,347)	K (22,298)	M (33,271)	Order of IC
B' (118,134)	0.1853	0.0671	0.0731	0.1904	0.0931	4
C' (25,348)	0.2192	0.1069	0.0362	0.2196	0.1174	9
K' (23,305)	0.0889	0.1835	0.2284	0.0341	0.1453	8
M' (19,280)	0.1651	0.0844	0.1162	0.1508	0.0809	10

(d)

	A (62,161)	B (209,234)	C (30,347)	K (22,298)	M (33,271)	Order of IC
A' (80,234)	0.0235	0.1665	0.2143	0.1012	0.1542	14
B' (65,124)	0.1385	0.0623	0.0790	0.1584	0.0826	4
C' (25,348)	0.2192	0.1069	0.0362	0.2196	0.1174	18
K' (22,304)	0.1030	0.1619	0.2026	0.0300	0.1183	6
M' (19,280)	0.1651	0.0844	0.1162	0.1508	0.0809	8

(e)

statistics is generally not sufficient in subtle signature characterization, such as small or rare targets, anomalies which cannot be generally captured by second-order statistics. Under such a circumstance, the second-order statistics-based DR may be very likely to sacrifice or compromise these targets during data compression. In order to resolve this dilemma, ICA-based DR was developed to cope with this problem. Since a detailed study and analysis was conducted in [4] and [19], many results available in [4] and [19] will not be included here. Instead, we demonstrate the superior performance of target detection performed on the ICA-DR compressed images compared to the PCA-DR and MNF-DR compressed images. The constrained energy minimization (CEM) developed in [3] was used for detection. Figs. 18–21 show the detection results produced by the CEM based on images obtained by the PCA-DR, MNF-DR, ICA-DR1(2), and ICA-DR3, respectively.

As we can see from Figs. 18–21, the results by the ICA-DR in Figs. 20 and 21 were significantly better than those in Figs. 18 and 19 by the PCA-DR and the MNF-DR where the former extracted all the pure panel R pixels even including some subpixels, while the latter could not separate the panels in the first three rows from the panels in the last two rows even all panel pixels were detected. In order to make further comparison, the CEM was applied to the original uncompressed image and the results are shown in Fig. 22.

Compared to Fig. 22, the results in Figs. 20 and 21 were comparable to those in Fig. 22. This implies that the ICA-DR preserves the critical information that the PCA-DR and the MNF-DR cannot in panel detection and classification.

As final remarks, several conclusions are noteworthy.

1) As demonstrated in our experiments, ICA-DR generally performed significantly better than PCA-DR or MNF-DR in the sense that the former preserves crucial and critical information such as endmembers, anomalies, small targets which generally contribute little to second-order statistics such as variance compared to the latter which preserves second-order statistics such as image background that accounts for most of variance.

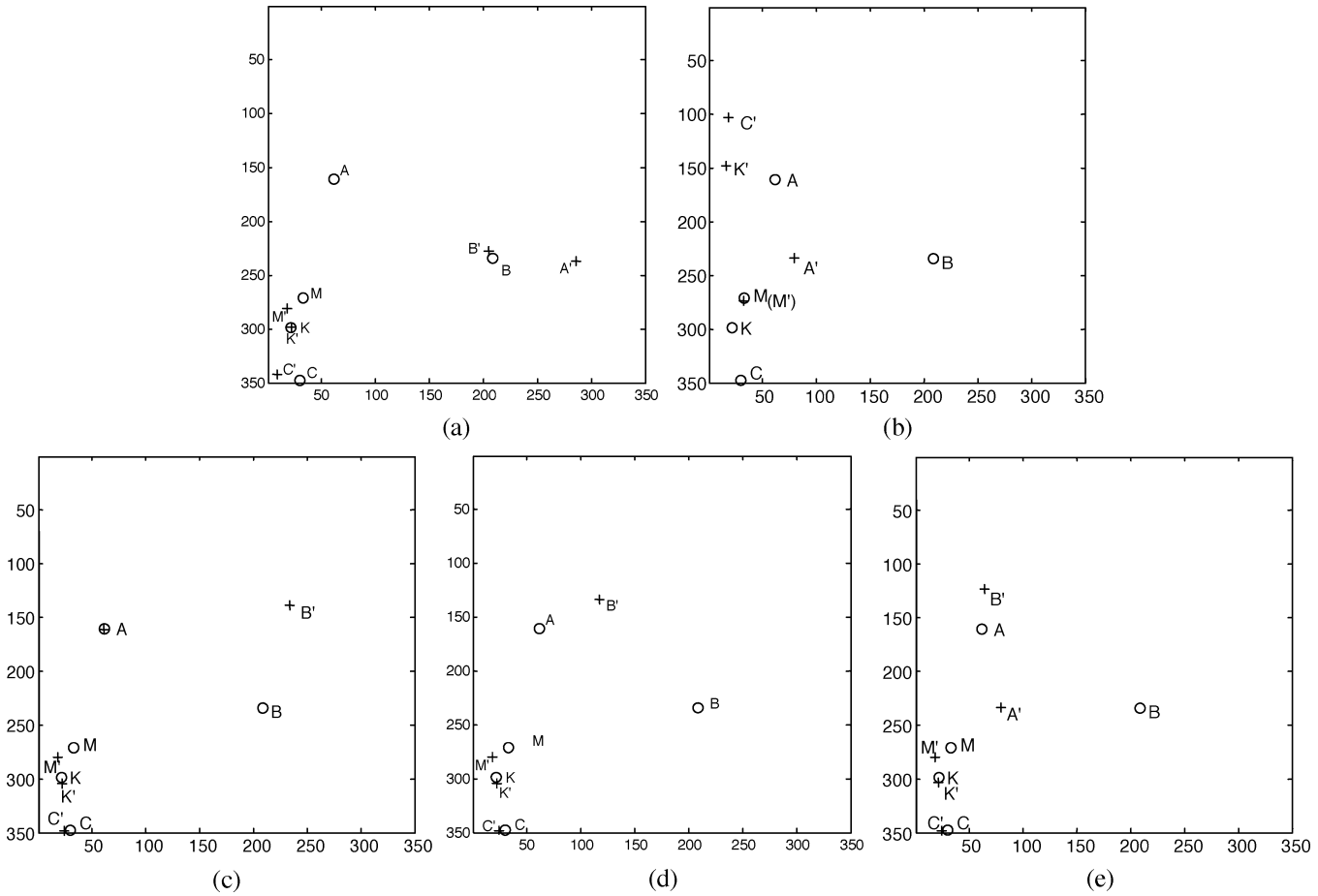


Fig. 8. Spatial locations of ground truth endmembers and N-FINDR-extracted endmembers. (a) PCA-DR. (b) MNF-DR. (c) ICA-DR1. (d) ICA-DR2. (e) ICA-DR3.

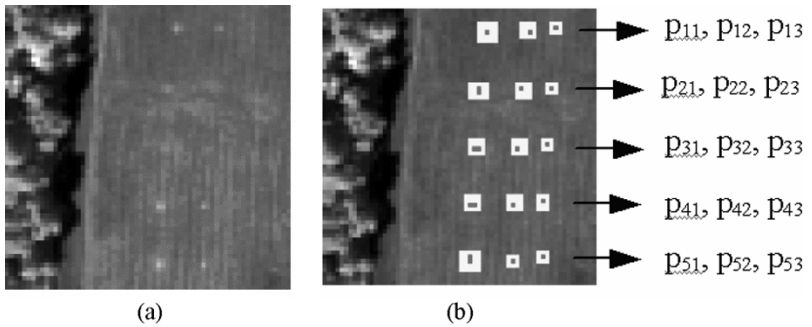


Fig. 9. Fifteen-panel HYDICE image. (a) Fifteen-panel image scene. (b) Ground truth map of 15 panels.

TABLE V
VD ESTIMATES FOR THE HYDICE SCENE IN FIG. 9 WITH VARIOUS FALSE ALARM PROBABILITIES

	$P_F = 10^{-1}$	$P_F = 10^{-2}$	$P_F = 10^{-3}$	$P_F = 10^{-4}$	$P_F = 10^{-5}$
VD	14	11	9	9	7

2) We did not include experiments using the percentage of accumulated eigenvalues as a criterion for DR due to the fact that the eigen-analysis is also a second-order statistics approach. It has been shown in [4]–[6], [19], and [20] that such a criterion was not effective.
3) In the application of hyperspectral data compression, it has been shown in [19] and [20] that commonly used objective measures for compression, mean squared error (MSE) or SNR were not effective in preserving targets with subtle information since missing these types of targets can only result in very small MSE or SNR. Therefore, exploitation-based criteria for compression are generally preferred in applications such as endmember extraction, target detection and classification. ICA-based DR for data compression is proposed particularly to address this issue. However, when both second-order

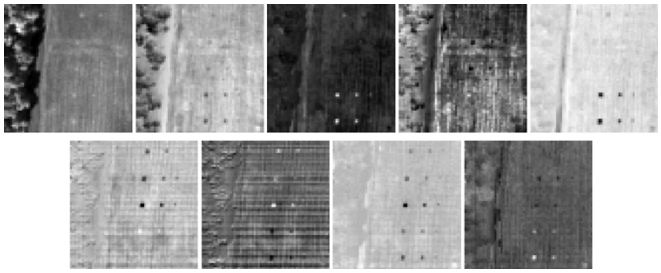


Fig. 10. Nine PCs produced by PCA.

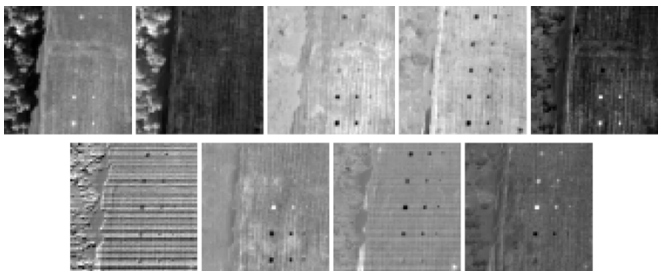


Fig. 11. Nine components produced by MNF.

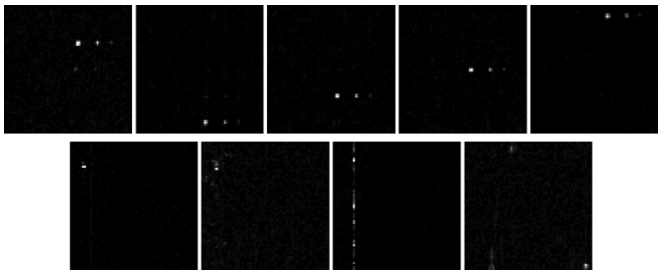


Fig. 12. Nine ICs produced by ICA-DR1(2).

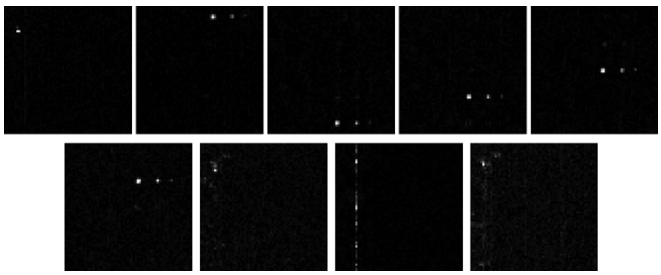


Fig. 13. Nine ICs produced by ICA-DR3.

statistics and high-order statistics are required to be preserved during data compression, a mixed PCA/ICA compression was recently developed for this purpose in [21].

4) Among all the three ICA-DR algorithms the ICA-DR1 and ICA-DR3 were shown to be most promising in applications. However, due to the use of different initial projection vectors (i.e., random vectors for the ICA-DR1 and ATGP-generated target vectors for the ICA-DR3) both may produce different results. Interestingly, all needed information for designated applications is preserved in the p prioritized ICs by the ICA-DR1 using the criterion (4) and the p ICs generated by the ICA-DR3 using the p ATGP-generated as initial projection vectors. Because of that, both performed similarly on many cases and also well in our experiments.
5) The measure used to evaluate the ICA-DR1 and ICA-DR3 is quite different. The performance of the ICA-DR1 is completely determined by the criterion given by (4). It can be also extended by any other high-order statistics, but may not have much advantage according to our experiments [22]. On the other hand, the ICA-DR3 depends heavily on its initial projection vectors produced by its initialization algorithm, ATGP. Fortunately, the ATGP has

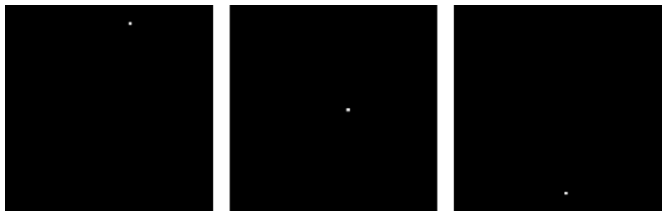


Fig. 14. Panel pixels extracted by N-FINDR algorithm using PCA-DR.

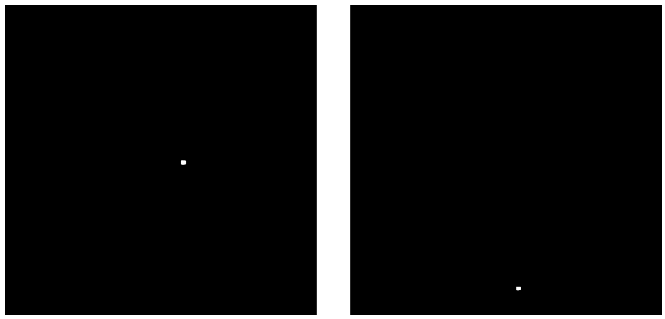


Fig. 15. Panel pixels extracted by N-FINDR algorithm using MNF-DR.

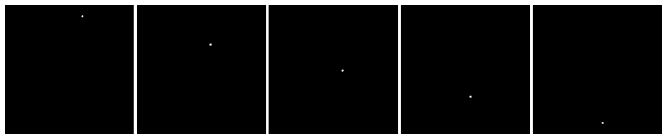


Fig. 16. Panel pixels extracted by N-FINDR algorithm using ICA-DR1(2).

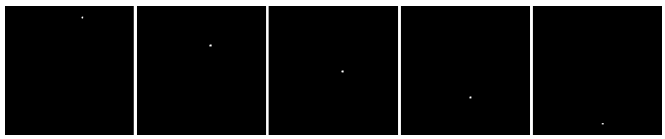


Fig. 17. Panel pixels extracted by N-FINDR algorithm using ICA-DR3.

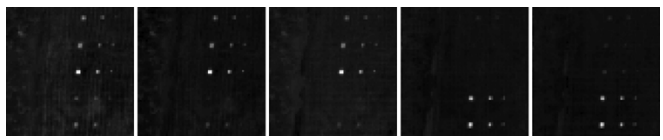


Fig. 18. CEM detection results in PCA-compressed image.

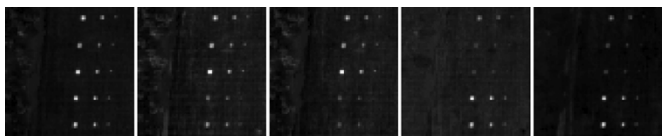


Fig. 19. CEM detection results in MNF-compressed image.

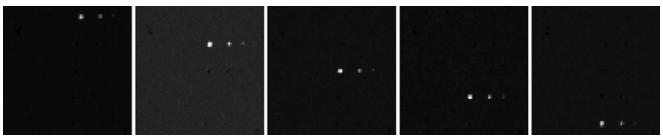


Fig. 20. Detection results of CEM in ICA-DR1(2)-compressed image.

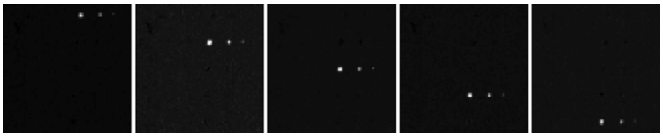


Fig. 21. CEM detection results in ICA-DR3-compressed image.

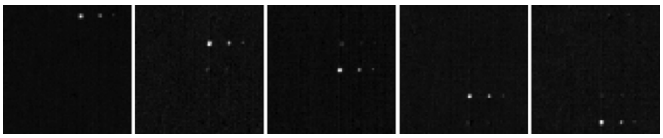


Fig. 22. CEM detection results in the original uncompressed image.

been shown in various applications to be very effective in capturing targets of interest such as unsupervised linear spectral mixture analysis, unsupervised target detection and classification [5] and endmember extraction [23].

6) Finally, it should be noted that the proposed ICA-DR was evaluated based on two particular image scenes, AVIRIS Cuprite data for endmember extraction and HYDICE 15-panel data for data compression in target detection and classification. For these specific applications, the ICA-DR was shown to be a very effective and promising technique. Many more applications are yet to be investigated for different image data to explore the potential of the ICA-DR.

V. Conclusions

This paper presents an ICA-DR. Theoretically, we can generate all ICs and examine all of them to select which components that we would like to retain. Practically, this is not realistic, particularly for hyperspectral data which have hundreds of components. The issue is how do we know and select which components are really desired for our applications? To the authors' best knowledge, there has been no such work reported on how to prioritize and select ICs based on exploitation criteria. Although the eigenvectors have been used for this purpose, it has been shown in [5] and [6] that it was not effective because eigen-analysis is limited to second-order statistics. Furthermore, it is a common practice that the PCA or MNF has been widely used for DR to avoid the issue of prioritizing components. Once again, if the transform used for DR is second-order statistics like PCA or MNF, it also runs into the same issue encountered in the eigen-analysis. This paper resolves these challenging issues described above with four contributions. First of all, the concept of VD which was originally developed for estimating number of spectrally distinct signatures is suggested to estimate number of dimensions needed to be retained. This is quite different from a common approach which is the use of eigenvalues to calculate percentage of energy as a criterion to determine how many PCs required to be retained by the PCA or MNF. Second, despite that the PCA and MNF also use eigenvalues to prioritize their PCs, there is no similar guide available for ICA to prioritize ICA-generated ICs. This paper introduces three different criteria for IC prioritization and selection. Third, according to these three different criteria, three algorithms, ICA-DR1, ICA-DR2, and ICA-DR3 are developed to select a set of desired ICs to achieve DR. Finally, a fourth contribution is to conduct a comprehensive study via two sets of different real hyperspectral images to evaluate the performance of the three proposed ICA-DR techniques in comparison with commonly used the variance-based PCA-DR, SNR-based MNF-DR in two major applications, endmember extraction and data compression. The experimental results demonstrate that the ICA-DR algorithms generally outperformed second-order statistics-based transforms such as PCA, MNF to perform DR.

Appendix

The purpose of this appendix is to provide a brief introduction of the concept of the VD and a method, called Harsanyi–Farrand–Chang (HFC) method developed in [24] to estimate the VD. The details about the VD can be found in [5] and [6]. The name of VD was originally coined in [5] and later in [6]. It was designed to determine the number of spectrally distinct signatures. If a component such as PC or IC is used to accommodate a spectrally distinct signature for classification and identification, the number of required components happens to be the number of spectrally distinct signatures, which is the VD. Despite several methods were developed in [2], the method developed by Harsanyi *et al.* [24], referred to as HFC method is selected for two reasons. One is simple to implement. Another is that it was shown to be effective in determining the number of spectrally signatures for AVIRIS data [24]. Its idea is very simple. It first calculates the sample correlation matrix, $\mathbf{R}$, and sample covariance matrix, $\mathbf{K}$, then finds the difference between their corresponding eigenvalues.

More specifically, let $\{\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_L\}$ and $\{\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_L\}$ be two sets of eigenvalues generated by $\mathbf{R}$ and $\mathbf{K}$, called correlation eigenvalues and covariance eigenvalues, respectively, where the L is the number of spectral channels. By assuming that signal sources are nonrandom unknown positive constants and noise is white with zero mean, we can expect that

$$\hat{\lambda}_l > \lambda_l, \quad \text{for } l = 1, \ldots, \text{VD} \tag{A1}$$

and

$$\hat{\lambda}_l = \lambda_l, \quad \text{for } l = \text{VD}+1, \ldots, L. \tag{A2}$$

Using (A-1) and (A-2), the eigenvalues in the lth spectral channel can be related by

$$\begin{aligned} &\hat{\lambda}_l > \lambda_l > \sigma^2_{n_l}, \quad \text{for } l = 1, \ldots, \text{VD} \\ &\text{and} \\ &\hat{\lambda}_l = \lambda_l = \sigma^2_{n_l}, \quad \text{for } l = \text{VD}+1, \ldots, L \end{aligned} \tag{A3}$$

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

In order to determine the VD, Harsanyi *et al.* [24] formulated the VD determination problem as a binary hypothesis problem as follows:

$$\begin{aligned} &H_0: \ z_l = \hat{\lambda}_l - \lambda_l = 0 \\ &\text{versus} \qquad\qquad\qquad\qquad \text{for } l = 1,2,\ldots,L \\ &H_1: \ z_l = \hat{\lambda}_l - \lambda_l > 0, \end{aligned} \tag{A4}$$

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 $\mathbf{R}$ in that particular component is the same as the one represented by the eigenvalue of $\mathbf{K}$ in its corresponding component.

Despite the fact that the $\hat{\lambda}_l$ and λ_l in (A1)–(A3) are unknown constants, according to [25], 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

$$p_0(z_l) = p(z_l|H_0) \cong N\left(0, \sigma^2_{z_l}\right), \quad \text{for } l = 1,2,\ldots,L \tag{A5}$$

and

$$p_1(z_l) = p(z_l|H_1) \cong N\left(\mu_l, \sigma^2_{z_l}\right), \quad \text{for } l = 1,2,\ldots,L \tag{A6}$$

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

$$\begin{aligned} \sigma^2_{z_l} &= \mathrm{Var}[\hat{\lambda}_l - \lambda_l] \\ &= \mathrm{Var}[\hat{\lambda}_l] + \mathrm{Var}[\lambda_l]\mathrm{Cov}(\hat{\lambda}_l, \lambda_l) \\ &\quad \text{for } l = 1,2,\ldots,L. \end{aligned} \tag{A7}$$

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

From (A5), (A6), and (A9), we define the false alarm probability and detection power (i.e., detection probability) as follows:

$$P_F = \int_{\tau_l}^{\infty} p_0(z)dz \tag{A8}$$

$$P_D = \int_{\tau_l}^{\infty} p_1(z)dz. \tag{A9}$$

A Neyman–Pearson detector for $\hat{\lambda}_l - \lambda_l$, denoted by $\delta_{NP}(\hat{\lambda}_l - \lambda_l)$ for the binary composite hypothesis testing problem specified by (A4) can be obtained by maximizing the detection power P_D in (A9), while the false alarm probability P_F in (A8) is fixed at a specific given value, α which determines the threshold value τ_l in (A8) and (A9). So a case of $\hat{\lambda}_l - \lambda_l > \tau_l$ indicating that $\delta_{NP}(\hat{\lambda}_l - \lambda_l)$ 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 (A4) must be performed for each of L spectral dimensions. Therefore, for each pair of $\hat{\lambda}_l - \lambda_l$, the threshold τ is different and should be l-dependent, that is τ_l.

REFERENCES

[1] J. Richards and X. Jia, *Remote Sensing Digital Image Analysis*, third ed. New York: Springer-Verlag, 1999.

[2] 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 Sens.*, vol. 26, no. 1, pp. 65–74, Jan. 1988.

[3] 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 Sens.*, vol. 28, no. 3, pp. 295–304, May 1990.

[4] B. Ramakishna, A. Plaza, C.-I Chang, H. Ren, Q. Du, and C.-C. Chang, "Spectral/spatial hyperspectral image compression," in *Hyperspectral Data Compression*, G. Motta and J. Storer, Eds. New York: Springer-Verlag, 2005.

[5] C.-I Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Norwell, MA: Kluwer, 2003.

[6] C.-I Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, Mar. 2004.

[7] J. W. Boardman, F. A. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of AVIRIS data," presented at the *Summaries JPL Airborne Earth Science Workshop*, Pasadena, CA, 1995.

[8] M. E. Winter, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *Proc. SPIE Conf. Imaging Spectrometry V*, vol. 3753, Denver, Co, 1999, pp. 266–275.

[9] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.

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

[11] 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.

[12] 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 Sens.*, vol. 40, no. 2, pp. 375–392, Feb. 2002.

[13] X. Zhang and C. H. Chen, "New independent component analysis method using higher order statistics with applications to remote sensing images," *Opt. Eng.*, vol. 41, pp. 1717–1728, Jul. 2002.

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

[15] H. V. Poor, *An Introduction to Signal Detection and Estimation Theory*. New York: Springer-Verlag, 1994.

[16] J. C. Harsanyi and C.-I Chang, "Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, Jul. 1994.

[17] H. Ren and C.-I Chang, "Automatic spectral target recognition in hyperspectral imagery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1232–1249, Oct. 2003.

[18] J. Wang and C.-I Chang, "Dimensionality reduction by independent component analysis for hyperspectral image analysis," presented at the *IEEE Int. Geosci. Remote Sens. Symp.*, Seoul, Korea, Jul. 2005.

[19] B. Ramakrishna, J. Wang, A. Plaza, and C.-I Chang, "Spectral/spatial hyperspectral image compression in conjunction with virtual dimensionality," presented at the *SPIE Conf. Algorithms Technol. Multispectral, Hyperspectral, Ultraspectral Imagery XI*, vol. 5806, Orlando, FL, 2005.

[20] B. Ramakishna, A. Plaza, C.-I Chang, H. Ren, Q. Du, and C.-C. Chang, "Spectral/spatial hyperspectral image compression," in *Hyperspectral Data Compression*, G. Motta and J. Storer, Eds. New York: Springer-Verlag, 2005, ch. 11, pp. 309–346.

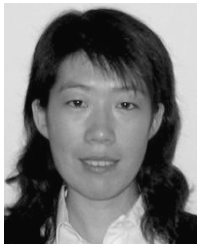
[21] J. Wang and C.-I Chang, "Mixed PCA/ICA spectral/spatial compression for hyperspectral imagery," presented at the *OpticsEast, Chem. Biol. Standoff Detection III (SA103)*, Boston, MA, Oct. 2005.

[22] H. Ren, Q. Du, J. Wang, C.-I Chang, and J. Jensen, "Automatic target recognition hyperspectral imagery using high order statistics," *IEEE Trans. Aerosp. Electron. Syst.*, to be published.

[23] C.-I Chang and A. Plaza, "Fast iterative algorithm for implementation of pixel purity index," *IEEE Trans. Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 63–67, Jan. 2006.

[24] J. C. Harsanyi, W. Farrand, and C.-I Chang, "Detection of subpixel spectral signatures in hyperspectral image sequences," in *Proc. Amer. Soc. Photogrammetry Remote Sens. Annu. Meeting*, 1994, pp. 236–247.

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



Jing Wang received the B.S. degree in electrical engineering and the M.S degree in computer engineering from the Beijing University of Post and Telecommunications, Beijing, China, in 1998 and 2001. She also received the M.S degree in electrical engineering from the University of Maryland, Baltimore County (UMBC), Baltimore, in 2005, where she is currently pursuing the Ph.D. degree.

She is currently a Research Assistant in the Remote Sensing, Signal and Image Processing Laboratory, UMBC. Her research interests include signal and image processing, pattern recognition and data compression.

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

He has been with the University of Maryland, Baltimore County (UMBC), Baltimore, since 1987 and is currently Professor in the Department of Computer Science and Electrical Engineering. He was a Visiting Research Specialist in the Institute of Information Engineering at the National Cheng Kung University, Tainan, Taiwan, from 1994 to 1995. His research interests include multispectral/hyperspectral image processing, automatic target recognition, medical imaging, information theory and coding, signal detection and estimation and neural networks. He has authored a book *Hyperspectral Imaging: Techniques for Spectral Detection and Classification* (Kluwer, 2003). He received a National Research Council (NRC) senior research associateship award from 2002 to 2003 sponsored by the U.S. Army Soldier and Biological Chemical Command, Edgewood Chemical and Biological Center, Aberdeen Proving Ground, MD. He has three patents and several pending on hyperspectral image processing. He is on the Editorial Board of the *Journal of High Speed Networks* and was the Guest Editor of a special issue of the same journal on telemedicine and applications.

Dr. Chang is an Associate Editor in the area of hyperspectral signal processing for IEEE TRANSACTION ON GEOSCIENCE AND REMOTE SENSING, a Fellow of SPIE, and a member of Phi Kappa Phi and Eta Kappa Nu.