Independent-component analysis for hyperspectral remote sensing imagery classification

Qian Du, MEMBER SPIE
Mississippi State University
Department of Electrical and
Computer Engineering
Mississippi State, Mississippi 39762

Ivica Kopriva, MEMBER SPIE
The George Washington University
Department of Electrical and
Computer Engineering
Washington, DC 20052

Harold Szu, FELLOW SPIE
Office of Naval Research
Arlington, Virginia 22217

Abstract. We investigate the application of independent-component analysis (ICA) to remotely sensed hyperspectral image classification. We focus on the performance of two well-known and frequently used ICA algorithms: joint approximate diagonalization of eigenmatrices (JADE) and FastICA; but the proposed method is applicable to other ICA algorithms. The major advantage of using ICA is its ability to classify objects with unknown spectral signatures in an unknown image scene, i.e., unsupervised classification. However, ICA suffers from computational expensiveness, which limits its application to high-dimensional data analysis. In order to make it applicable or reduce the computation time in hyperspectral image classification, a data-preprocessing procedure is employed to reduce the data dimensionality. Instead of using principal-component analysis (PCA), a noise-adjusted principal-components (NAPC) transform is employed for this purpose, which can reorganize the original data with respect to the signal-to-noise ratio, a more appropriate image-ranking criterion than variance in PCA. The experimental results demonstrate that the major principal components from the NAPC transform can better maintain the object information in the original data than those from PCA. As a result, an ICA algorithm can provide better object classification. [DOI: 10.1117/1.2151172]

Subject terms: independent-component analysis; principal-component analysis (PCA); noise-adjusted principal components (NAPC) transform; unsupervised classification; hyperspectral imagery; remote sensing.

Paper 050157RR received Feb. 25, 2005; revised manuscript received May 21, 2005; accepted for publication May 26, 2005; published online Jan. 24, 2006.

1 Introduction

Hyperspectral remote sensing is a new research area that attracts much interest from researchers and practitioners because the high spectral resolution of an acquired image provides the potential of more accurate object detection, classification, and identification than multispectral imagery. A 3-D hyperspectral image cube contains hundreds of coregistered images for the same image scene taken in very narrow spectral bands. But how to efficiently deal with its vast data volume while taking advantage of the optimum amount of spectral information is challenging.

In many practical applications of remote sensing image classification, it may be very difficult or even impossible to get prior information about class signatures, so unsupervised methods need to be applied. The spatial resolution of remote sensing imagery is rather rough. In general, the area covered in each pixel includes different materials and objects. So we have to deal with mixed pixels instead of pure pixels as in conventional digital image processing. Linear spectral unmixing analysis is a popular approach used to handle mixed pixels. This procedure assumes the reflectance of a pixel is a linear mixture of those of all the different materials found in that pixel.[1–4] Let L be the number of spectral bands, and $\mathbf{r}$ a column pixel vector with dimension L in a hyperspectral image. The element r_i in $\mathbf{r}$ is the reflectance collected in the i'th wavelength band. Let $\mathbf{M}$ denote a matrix containing q independent material spectral signatures (referred to as *endmembers* in the linear mixture model[1–4]), i.e., $\mathbf{M}=[\mathbf{m}_1,\mathbf{m}_2,\ldots,\mathbf{m}_q]$. Let $\boldsymbol{\alpha}$ be the unknown abundance column vector of size $q\times 1$ associated with $\mathbf{M}$, which is to be estimated. The i'th item α_i in $\boldsymbol{\alpha}$

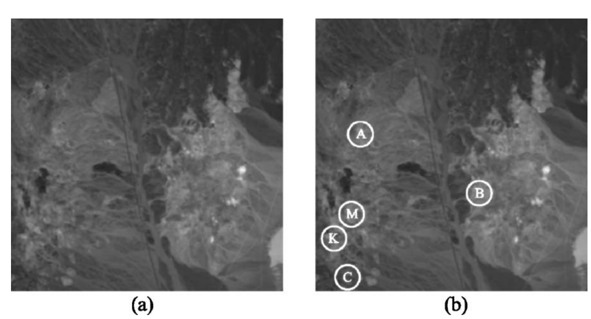

Fig. 1 AVIRIS "Cuprite" image scene: (a) a spectral band image; (b) spatial locations of five pure pixels corresponding to minerals: alunite (A), buddingtonite (B), calcite (C), kaolinite (K), and muscovite (M).

0091-3286/2006/$22.00

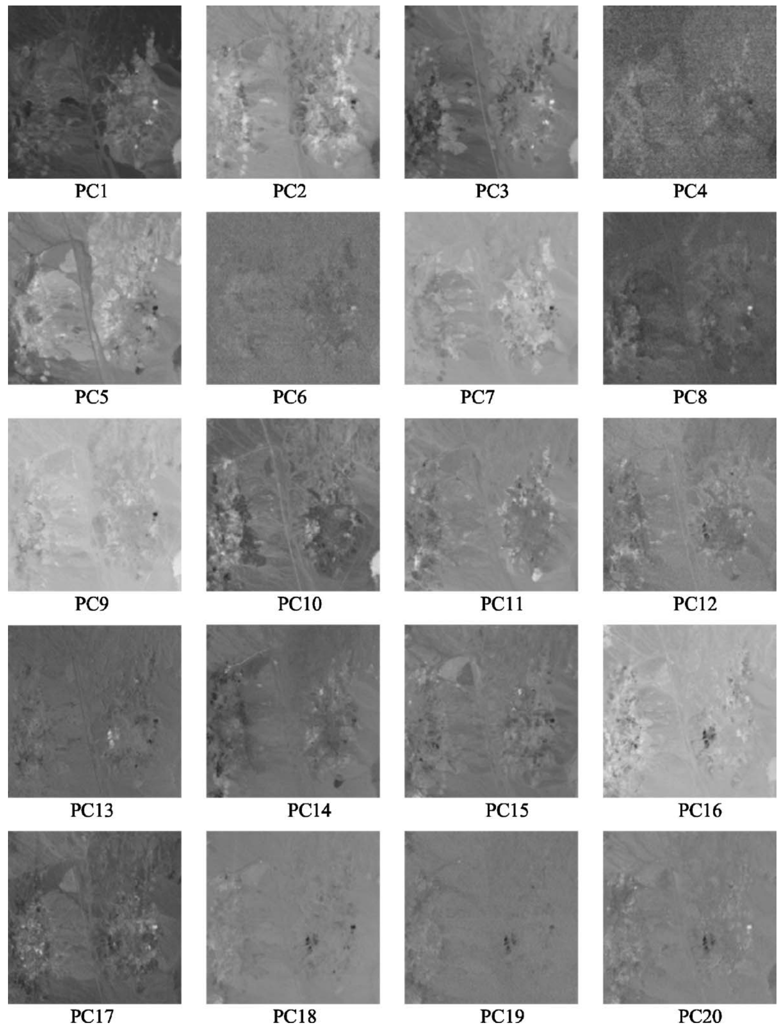

Fig. 2 The first 20 principal components from PCA for the AVIRIS "Cuprite" scene.

represents the abundance fraction of $\mathbf{m}_i$ in pixel $\mathbf{r}$. According to the linear mixture model,[1–4]

$$\mathbf{r} = \mathbf{M}\boldsymbol{\alpha} + \mathbf{n}, \tag{1}$$

where $\mathbf{n}$ is the noise term. When $\mathbf{M}$ is known, the estimation of $\boldsymbol{\alpha}$ can be accomplished by a least-squares approach. But when $\mathbf{M}$ is also unknown, i.e., in unsupervised analysis, the task is much more challenging, since both $\mathbf{M}$ and $\boldsymbol{\alpha}$ need to be estimated.

Independent-component analysis (ICA) is a powerful tool for unsupervised classification, which has been successfully applied to blind source separation.[5–13] The basic idea is to decompose a set of multivariate signals into a basis of statistically independent sources with minimal loss of information content so as to achieve detection and classification. The standard linear ICA-based data model with additive noise is[13]

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n}, \tag{2}$$

where $\mathbf{x}$ is an L-dimensional data vector, $\mathbf{A}$ is an unknown mixing matrix, and $\mathbf{s}$ is an unknown source signal vector. Three assumptions are made on $\mathbf{s}$: (1) each source signal is an independent identically distributed (i.i.d.) stationary random process; (2) the source signals are statistically independent at any time; and (3) at most one among the source signals has a Gaussian distribution. The mixing matrix $\mathbf{A}$, although unknown, is also assumed to be nonsingular. Then the solution to the blind source separation problem is obtained with scale and permutation indeterminacy. Let $\mathbf{W}$ represent the unmixing matrix. It satisfies $\mathbf{WA} = \mathbf{P\Gamma}$, where $\mathbf{P}$ is a generalized permutation matrix and $\mathbf{\Gamma}$ is a diagonal matrix. These requirements ensure the existence and uniqueness of the solution to the blind source separation problem except for the indeterminacy of the ordering, signs, and scaling of the outputs. In contrast with many conventional techniques, which use up to second-order statistics only, ICA exploits higher-order statistics, which makes it more powerful in extracting irregular features in the data.

Several researchers have explored ICA for remote sensing image classification.[14–19] In general, when the ICA approach is used to classify optical multispectral or hyperspectral images, the linear mixture model in Eq. (1) needs to be reinterpreted to fit the model given by Eq. (2). Specifically, the pixel vector $\mathbf{r}$ is denoted as $\mathbf{x}$ in Eq. (2), the

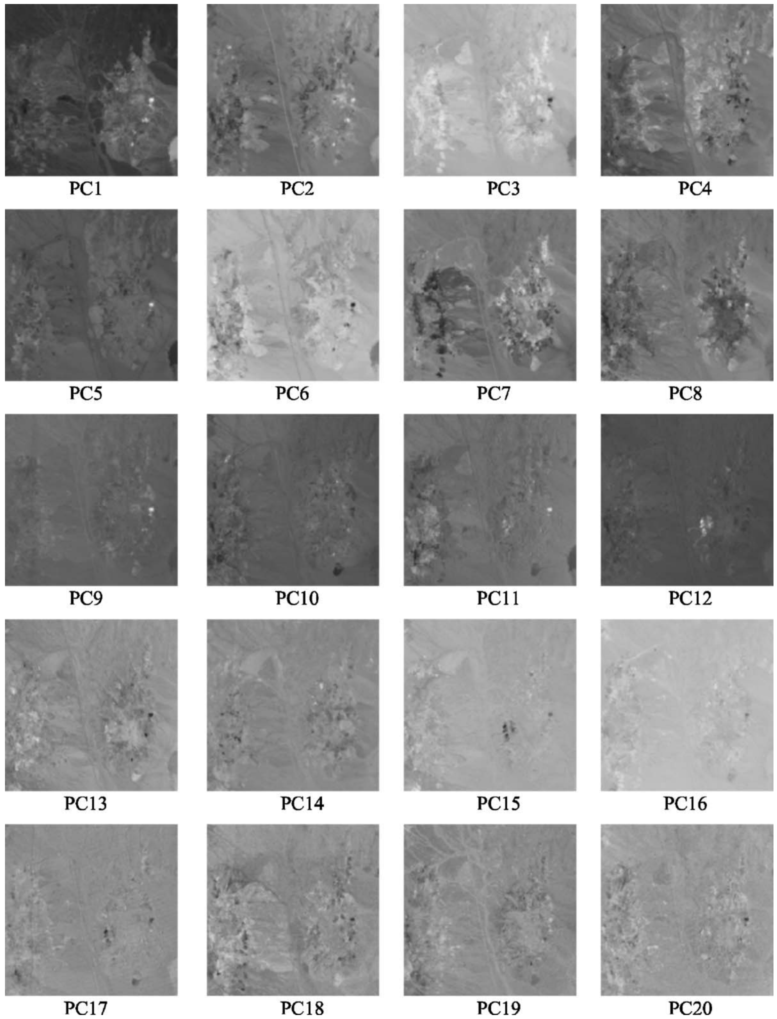

Fig. 3 The first 20 principal components from the NAPC transform for the AVIRIS "Cuprite" scene.

endmember matrix $\mathbf{M}$ in Eq. (1) corresponds to the unknown mixing matrix $\mathbf{A}$ in Eq. (2), and the abundance-fraction vector $\boldsymbol{\alpha}$ in Eq. (1) corresponds to the source signal vector $\mathbf{s}$ in Eq. (2). Moreover, the abundance fractions are considered as unknown random quantities specified by random signal sources in the ICA model (2) rather than unknown deterministic quantities as assumed in the linear mixture mode (1). With these interpretations and the preceding assumptions, we use the model (2) to replace the model (1) hereafter. The advantages offered by using the model (2) in remote sensing image classification are: (1) no prior knowledge of endmembers in the mixing process is required; (2) the spectral variability of endmembers can be accommodated by the unknown mixing matrix $\mathbf{A}$, since the source signals are considered as random scalar quantities; and (3) higher-order statistics can be exploited for better featureextraction and pattern classification.[17]

For mathematical tractability, the mixing matrix $\mathbf{A}$ and unmixing matrix $\mathbf{W}$ in the ICA model are taken to be square matrices of size $L \times L$. The major drawback of ICA is its high computational complexity. For instance, the computational complexity of the joint approximate diagonalization of eigenmatrices (JADE) algorithm is on the order of L^4. When JADE is applied to hyperspectral imagery, where L can be as high as 200, the computation becomes prohibitively expensive. Therefore, in order to make the JADE algorithm applicable to hyperspectral image classification, some data-preprocessing procedure is required to reduce the image data dimensionality. The common approach is to use ordinary principal-component analysis (PCA) for dimension reduction, which reorganizes the original data information in terms of variance. Here, we propose using a noise-adjusted principal components (NAPC) transform for this purpose. NAPC can reorganize the original data information in terms of the signal-to-noise ratio (SNR), which is a more reasonable criterion for handling image data.[20–22] It is demonstrated below that the major principal components (PCs) from the NAPC transform can better represent the original object information than those from PCA. In what follows we introduce an ICA algorithm that can provide better object classification.

Some ICA algorithms, such as the well-known FastICA, can be applied to high-dimensional data. However, the spatial size of a hyperspectral image can be large, algorithm execution can be computationally expensive, and a postprocessing step is required to tease out the interesting objects

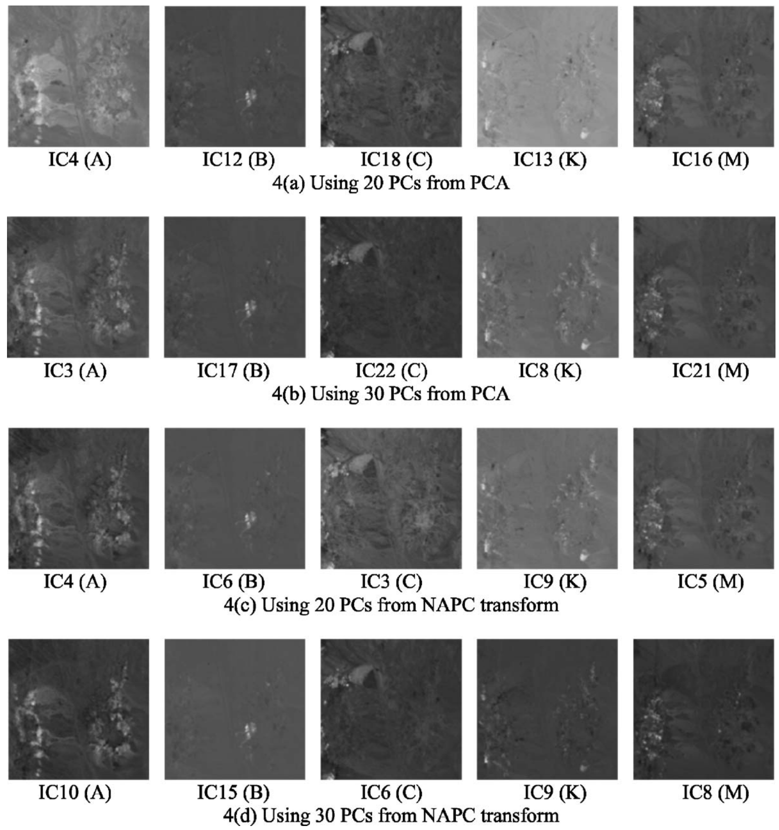

Fig. 4 JADE classification results for the AVIRIS image scene.

from a large number of classification maps. If a preprocessing step using an NAPC transform is applied to reduce the data dimensionality, the ICA classification step and the postprocessing step can be finished much more quickly.

It should be noted that Tu in Ref. 16 was the first to apply the NAPC transform to ICA. In that paper the major purpose for using an NAPC transform was to improve the performance of a Gerschgorin disk approach in the estimation of the number of signals in an image scene, which is the same as the number of independent components to be determined. We have, however, modified the use of the NAPC transform. Firstly, the objective of our research is to employ an NAPC transform to reduce the data dimensionality while maintaining most of the object information. As a result, the ICA algorithms for low-dimensional data (such as JADE) can be applicable to high-dimensional data, while the ICA algorithms feasible for high-dimensional data (such as FastICA) can be made even faster with a comparable classification. Secondly, the noise estimation technique in our NAPC transform is different from the nearest-neighbor difference method used in Refs. 16, 20, and 21, which is introduced in Sec. 3. As demonstrated in Ref. 23, the NAPC transform is sensitive to noise estimation, and the performance of the nearest-neighbor difference method is limited. Thirdly, when the estimation of the number of independent components is necessary, we resort to an approach proposed in Ref. 24, which was proven to be effective in our hyperspectral experiments.

The remainder of this paper is organized as follows. Section 2 briefly describes the JADE and FastICA algorithms. Section 3 introduces the NAPC transform for dimension reduction of hyperspectral image data. Section 4 presents experiments using data from the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and Hyperspectral Digital Imagery Collection Experiment (HYDICE) to demonstrate the performance of ICA-based unsupervised classification in conjunction with an NAPC-transform-based dimension reduction. Section 5 concludes with brief remarks.

2 JADE and FastICA Algorithms

The strategy of ICA is to find a linear transform $\mathbf{W}$ (i.e., an unmixing matrix of size $L \times L$) such that the components in the vector $\mathbf{z}$ in the following equation are as statistically independent as possible:

$$\mathbf{z} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s} + \mathbf{W}\mathbf{n} = \mathbf{Q}\mathbf{s} + \mathbf{W}\mathbf{n}. \tag{3}$$

Based on the assumption that source signals in vector $\mathbf{s}$ are mutually statistically independent and non-Gaussian (except one that is allowed to be Gaussian), $\mathbf{z}$ can represent the source signal vector $\mathbf{s}$ up to the permutation, scale, and sign factors.

Several different types of ICA algorithms have been developed recently, such as non-Gaussianity-maximization-based ICA, maximum-likelihood-estimation-based ICA, mutual-information-minimization-based ICA, and nonlinear-decorrelation-based ICA.[13] Among them, the JADE algorithm[6] and FastICA[10,11] are most popular and

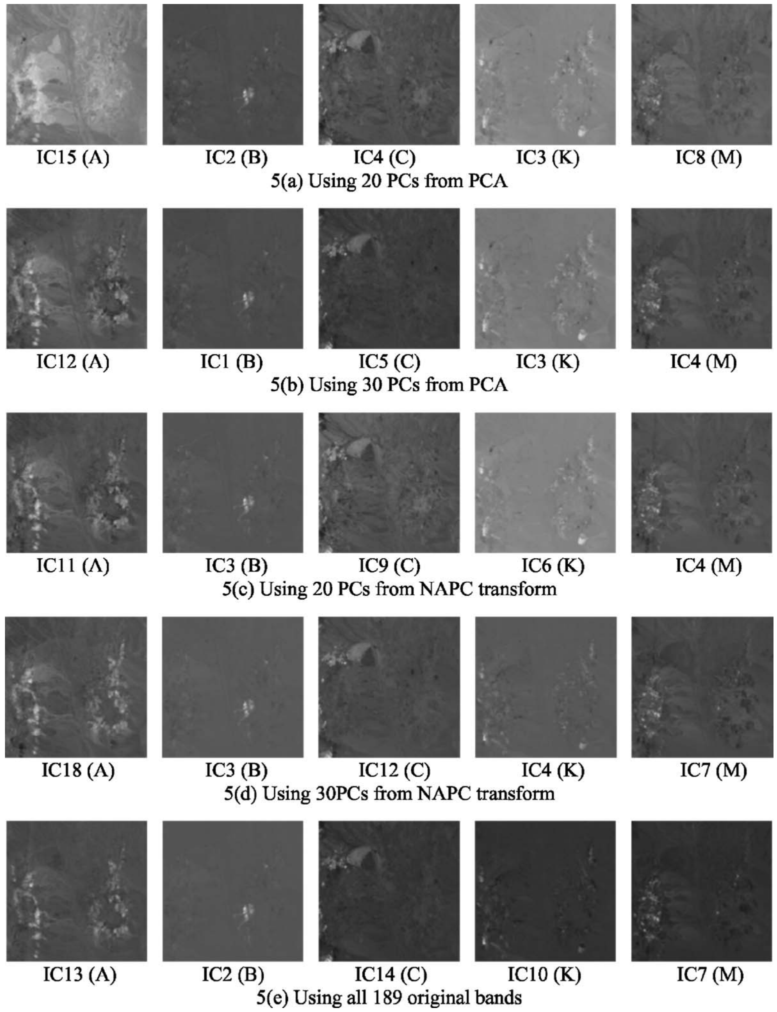

Fig. 5 FastICA classification results for the AVIRIS image scene.

often used. In this paper, we investigate their performance when an NAPC transform is used for dimension reduction.

JADE is based on the use of fourth-order statistics (cumulants). The higher-order statistical dependence among data samples is measured by the higher-order cross-cumulants. Smaller values of cross-cumulants represent less dependent samples. In JADE fourth-order statistical independence is achieved through minimization of the squares of fourth-order cross-cumulants between the components z_i of $\mathbf{z}$ in Eq. (3). Then fourth-order cross-cumulants can be computed as

$$C_4(z_i,z_j,z_k,z_l) = E[z_iz_jz_kz_l] - E[z_iz_j]E[z_kz_l] - E[z_iz_k]E[z_jz_l] - E[z_iz_l]E[z_jz_k] \quad (4)$$

for $1 \leqslant i,j,k,l \leqslant L$, where $E[\cdot]$ denotes the statistical expectation operator. Equation (4) can be expressed as an $L^2 \times L^2$ Hermitian matrix. The optimal unmixing matrix $\mathbf{W}^*$ is the one that satisfies

$$\mathbf{W}^* = \arg\min \sum_{i,j,k,l} \mathrm{off}(\mathbf{W}^T C_4(z_i,z_j,z_k,z_l)\mathbf{W}), \quad (5)$$

where off $(\cdot)$ is a measure of the off-diagonality of a matrix, defined as

$$\mathrm{off}(\mathbf{X}) = \sum_{1 \leqslant i \neq j \leqslant L} |x_{ij}|^2, \quad (6)$$

where x_{ij} denotes the ij'th element of a matrix $\mathbf{X}$. An intuitive way to solving the optimization problem in Eq. (5) is to jointly diagonalize the eigenmatrices of $C_4(z_i,z_j,z_k,z_l)$ in Eq. (4) via Givens rotation.[25] In order for first- and second-order statistics not to affect the results, data should be pre-whitened (so that the mean is zero and the covariance matrix is an identity matrix).

FastICA is based on a fixed-point iteration scheme for finding an optimal transform $\mathbf{w}$ (one of the vectors in the unmixing matrix $\mathbf{W}$) that can maximize the non-Gaussianity of $E[z]=E[\mathbf{w}^T\mathbf{x}]$, i.e., the objective function is

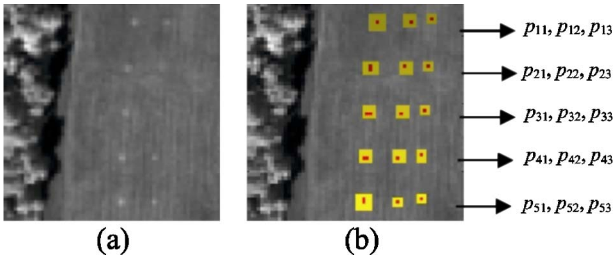

Fig. 6 (a) A HYDICE image scene that contains 15 panels; (b) ground truth map of spatial locations of the 15 panels.

$$\max_{\mathbf{w}} J(\mathbf{w}) = E[G(\mathbf{w}^T\mathbf{x})] \quad \text{subject to } \|\mathbf{w}\| = 1, \tag{7}$$

where $G(\cdot)$ is a nonlinear function measuring non-Gaussianity that can be chosen as $G(u)=(1/a)\log\cosh au$, $G(u)=-(1/a)\exp(-au^2/2)$, or $G(u)=(\frac{1}{4})u^4$. According to the Karush-Kuhn-Tucker condition,[26] the solution of the constrained problem in Eq. (7) can be obtained by solving

$$E[\mathbf{x}g(\mathbf{w}^T\mathbf{x})] - \beta\mathbf{w} = 0, \tag{8}$$

where $g(\cdot)$ is the derivative of $G(\cdot)$, i.e., $g(\cdot)=G'(\cdot)$, and β is a constant that can be evaluated as $\beta=E[\mathbf{w}_0^T\mathbf{x}g(\mathbf{w}_0^T\mathbf{x})]$. Here, $\mathbf{w}_0$ is the value of $\mathbf{w}$ at the optimum. Newton's method can be used to solve Eq. (8).[27] After simplification, a fixed-point iterative algorithm is obtained as $\mathbf{w} \leftarrow E[\mathbf{x}g(\mathbf{w}^T\mathbf{x})] - E[g'(\mathbf{w}^T\mathbf{x})]\mathbf{w}$ and $\mathbf{w} \leftarrow \mathbf{w}/\|\mathbf{w}\|$.

To find a second transform vector $\mathbf{w}$, the algorithm is reexecuted. To prevent different $\mathbf{w}$ from converging to the same maxima, we can simply decorrelate the output z after each iteration by using a Gram-Schmidt-like decorrelation step.[11]

FastICA can generate the final result much more quickly than a stochastic gradient descent method and does not need to carefully adjust the learning rate. In contrast with JADE, which is very successful for low-dimensional data, FastICA can be applied to high-dimensional data.

3 PCA and NAPC Transforms and Their Implementation

3.1 *PCA and NAPC Transforms*

Consider an observation model

$$\mathbf{x} = \boldsymbol{\theta} + \mathbf{n}, \tag{9}$$

where $\{\mathbf{x}\}$ is a set of observation vectors with data dimensionality L, which contains a signal vector $\boldsymbol{\theta}$ and an uncorrelated additive noise term $\mathbf{n}$. The sample mean of $\{\mathbf{x}\}$ is $\mathbf{m}$. The objective of PCA is to find a transformation vector $\mathbf{v}$ such that the variance of the transformed data $\{\mathbf{v}^T(\mathbf{x}-\mathbf{m})\}$ is maximized. It is assumed that the transformed data keep the most information of $\{\mathbf{x}\}$ if its variance is maximal. In order not to let $\mathbf{v}$ affect the variance of the transformed data, a constraint is imposed as $\mathbf{v}^T\mathbf{v}=1$. So the objective function is

$$J(\mathbf{v}) = \mathbf{v}^T\boldsymbol{\Sigma}\mathbf{v} + \lambda(\mathbf{v}^T\mathbf{v} - 1), \tag{10}$$

where $\boldsymbol{\Sigma}$ is the sample covariance matrix and λ is a Lagrange multiplier. Taking the partial derivative of Eq. (10) with respect to $\mathbf{v}$ and setting it equal to 0, we obtain

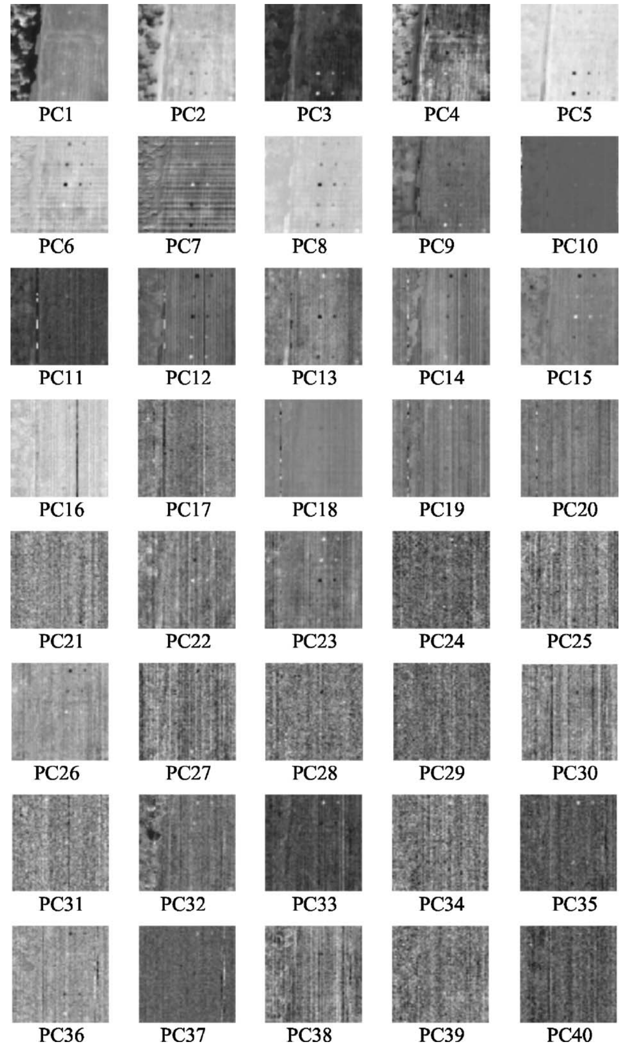

Fig. 7 The first 40 principal components from PCA for the HYDICE image scene.

$$\boldsymbol{\Sigma}\mathbf{v} = \lambda\mathbf{v}. \tag{11}$$

Obviously, this is an eigenproblem with L roots for $\mathbf{v}$ and λ, which are eigenvectors and eigenvalues of $\boldsymbol{\Sigma}$, denoted as $\mathbf{V}=[\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_L]$ and $\boldsymbol{\Lambda}=\mathrm{diag}\{\lambda_1,\lambda_2,\ldots,\lambda_L\}$, respectively. Here, $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_L$ are L eigenvectors, and $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_L$ are their corresponding eigenvalues. The matrices $\mathbf{V}$ and $\boldsymbol{\Lambda}$ can be related as

$$\mathbf{V}^T\boldsymbol{\Sigma}\mathbf{V} = \boldsymbol{\Lambda}. \tag{12}$$

Then the principal components (PCs) from PCA can be calculated by

$$\mathbf{x}_{\mathrm{PCA}} = \mathbf{V}^T(\mathbf{x} - \mathbf{m}) \tag{13}$$

with the variance of the i'th PC being λ_i.

PCA uses variance as the ranking criterion for PCs, and the first several PCs (major PCs) have larger variances. It is assumed that most of the object information is included in major PCs. However, variance can be contributed by both

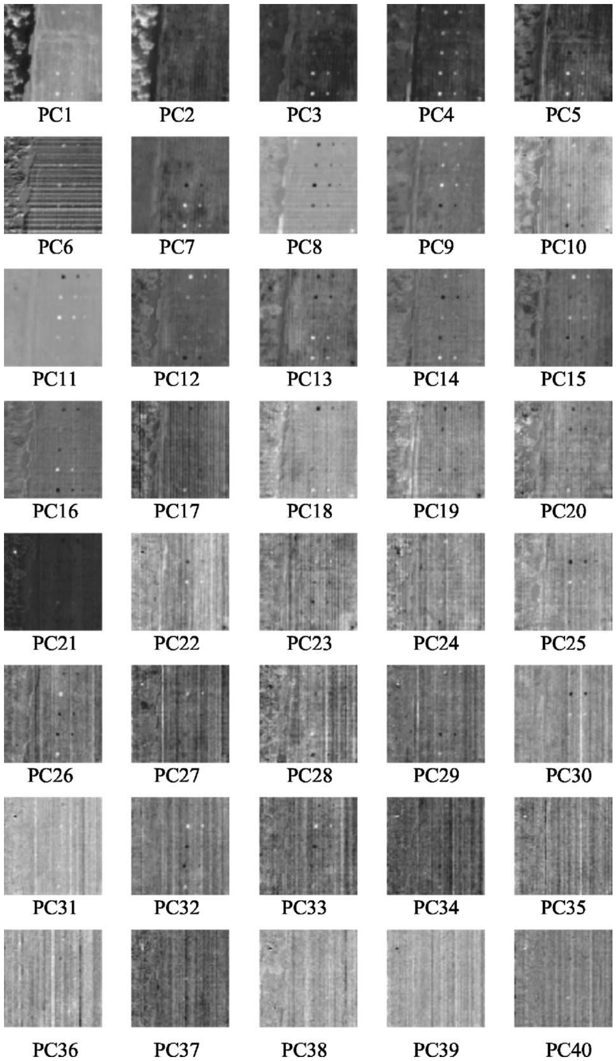

Fig. 8 The first 40 principal components from the NAPC transform for the HYDICE image scene.

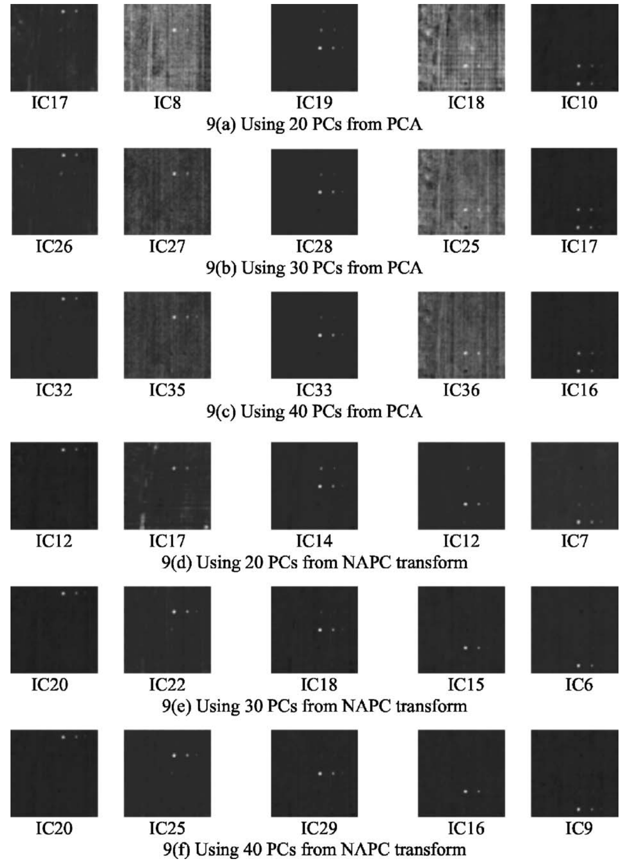

Fig. 9 JADE classification results for the HYDICE image scene.

signals and noise. In particular, the contribution to the variance from small objects may be even smaller than that from noise, and then these objects will appear in minor PCs. As a result, objects cannot be well compacted into major PCs, and some major PCs may contain noise only. In other words, object information is spread into more PCs.

The NAPC transform was proposed to solve this problem by ranking PCs in terms of image quality, i.e., SNR, so that most of the object information can be represented in major PCs.[20–22] The objective is to find a transformation vector $\mathbf{v}$ such that the SNR in the transformed data can be maximized. The objective function is

$$J(\mathbf{v}) = \mathrm{SNR} = \frac{\mathbf{v}^T\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\mathbf{v}}{\mathbf{v}^T\boldsymbol{\Sigma}_{\mathbf{n}}\mathbf{v}} = \frac{\mathbf{v}^T\boldsymbol{\Sigma}\mathbf{v}}{\mathbf{v}^T\boldsymbol{\Sigma}_{\mathbf{n}}\mathbf{v}} - 1, \tag{14}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ and $\boldsymbol{\Sigma}_{\mathbf{n}}$ are the signal and noise covariance matrices, respectively. In Eq. (14) the relationship $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_{\boldsymbol{\theta}}+\boldsymbol{\Sigma}_{\mathbf{n}}$ is used. Let $\mathbf{F}=\mathbf{K}_{\mathbf{n}}\boldsymbol{\Delta}_{\mathbf{n}}^{-1/2}$ be the noise-whitening matrix such that

$$\mathbf{F}^T\boldsymbol{\Sigma}_{\mathbf{n}}\mathbf{F} = \mathbf{I} \quad \text{and} \quad \mathbf{F}^T\mathbf{F} = \boldsymbol{\Delta}_{\mathbf{n}}^{-1}, \tag{15}$$

where $\mathbf{K}_{\mathbf{n}}$ and $\boldsymbol{\Delta}_{\mathbf{n}}$ are the eigenvector and eigenvalue matrices of $\boldsymbol{\Sigma}_{\mathbf{n}}$, respectively, and $\mathbf{I}$ is an identity matrix. Let $\mathbf{h}=\mathbf{F}^{-1}\mathbf{v}$. Equation (14) becomes

$$J(\mathbf{h}) = \frac{\mathbf{h}^T(\mathbf{F}^T\boldsymbol{\Sigma}\mathbf{F})\mathbf{h}}{\mathbf{h}^T\mathbf{h}} - 1. \tag{16}$$

Using the same technique in the derivation of PCA, it can be easily shown that the L roots of $\mathbf{h}$ in $\mathbf{H}=[\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_L]$ are the eigenvectors of the noise-whitened covariance matrix $\boldsymbol{\Sigma}_{n\_\mathrm{adj}}$, defined as

$$\boldsymbol{\Sigma}_{n\_\mathrm{adj}} = \mathbf{F}^T\boldsymbol{\Sigma}\mathbf{F}. \tag{17}$$

Then the PCs from the NAPC transform can be calculated by

$$\mathbf{x}_{\mathrm{NAPC}} = \mathbf{H}^T\mathbf{F}^T(\mathbf{x}-\mathbf{m}) \tag{18}$$

with the variance of the i'th PC being the i'th eigenvalue of $\boldsymbol{\Sigma}_{n\_\mathrm{adj}}$, which is equal to SNR+1 in the i'th PC.

We can see that an NAPC transform includes two steps: noise whitening and ordinary PCA. Because noise is whitened, the noise variance in each PC is the same. The PCs are actually ranked in terms of signal variances. As a result, object information can be better compacted into the first several PCs.

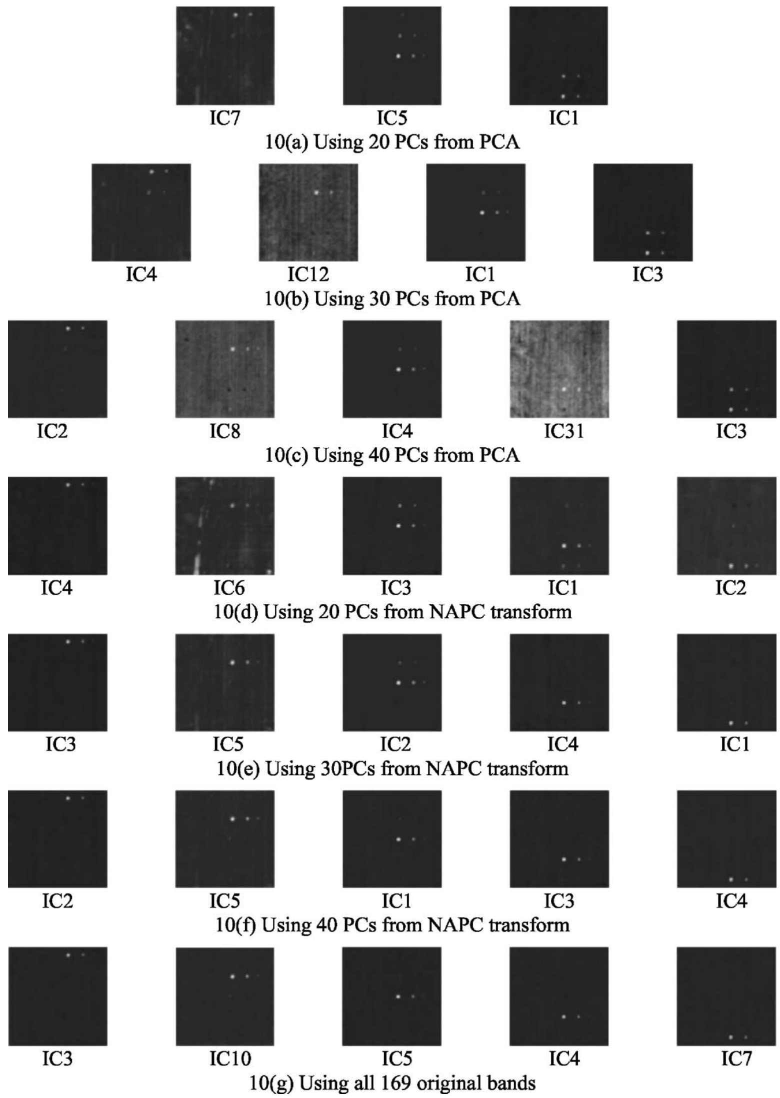

Fig. 10 FastICA classification results for the HYDICE image scene.

3.2 *Noise Estimation*

The major difficulty in performing an NAPC transformation is to have an accurate noise covariance matrix $\mathbf{\Sigma_n}$. The following method based on interband correlation is adopted in our research for its simplicity and effectiveness.[28]

Let the sample covariance matrix $\mathbf{\Sigma}$ be decomposed as

$$\mathbf{\Sigma} = \mathbf{DED} \tag{19}$$

where $\mathbf{D}=\mathrm{diag}\{\sigma_1,\sigma_2,\ldots,\sigma_L\}$ is a diagonal matrix with σ_l^2 (the variance of the l'th original band) being the diagonal elements of $\mathbf{\Sigma}$, and $\mathbf{E}$ is the correlation coefficient matrix, whose mn'th element represents the correlation coefficient between the m'th and n'th bands. Similarly, in analogy with the decomposition of $\mathbf{\Sigma}$, its inverse $\mathbf{\Sigma}^{-1}$ can also be decomposed as

$$\mathbf{\Sigma}^{-1} = \mathbf{D}_{\mathbf{\Sigma}^{-1}}\mathbf{E}_{\mathbf{\Sigma}^{-1}}\mathbf{D}_{\mathbf{\Sigma}^{-1}}, \tag{20}$$

where $\mathbf{D}_{\mathbf{\Sigma}^{-1}}=\mathrm{diag}\{\zeta_1,\zeta_2,\ldots,\zeta_L\}$ is a diagonal matrix with ζ_l^2 the diagonal elements of $\mathbf{\Sigma}^{-1}$, and $\mathbf{E}_{\mathbf{\Sigma}^{-1}}$ is a matrix similar to $\mathbf{E}$ with the diagonal elements equal to 1 and all off-diagonal elements equal to -1 or 1. It turns out that ζ_l can be related to σ_l as

$$\zeta_l = \frac{1}{[\sigma_l^2(1-\rho_{L-l}^2)]^{1/2}}, \tag{21}$$

where ρ_{L-l}^2 is the multiple correlation coefficient of the l'th band, on the other $L-1$ bands, obtained by using multiple regression theory.[29] So ζ_l^2 is the reciprocal of a good noise variance estimate of the l'th band. Therefore, the noise covariance matrix $\mathbf{\Sigma}_n$ can be estimated by $\mathbf{\Sigma}_n=\mathrm{diag}\{\zeta_1^{-2},\zeta_2^{-2},\ldots,\zeta_L^{-2}\}$, which is a diagonal matrix.

3.3 *Estimation of the Number of Major Principal Components*

In practice it is difficult to decide where to truncate the eigenspectrum, as the data information may be distributed into many PCs. A hypothesis-testing-based estimation technique in Ref. 24, called noise subspace projection (NSP), was used to estimate the number of distinct signal sources present in the image scene, referred to as the virtual dimensionality (VD). This number can be used as a reference for us to decide how many PCs to keep.

The sample covariance matrix $\mathbf{\Sigma}$ can be whitened using Eq. (17). As a result, the noise variance of each band in the whitened $\mathbf{\Sigma}_{n\_\mathrm{adj}}$ is reduced to unity. Let $\{\bar{\mathbf{v}}_l\}_{l=1}^{L}$ be a set of eigenvalues for $\mathbf{\Sigma}_{n\_\mathrm{adj}}$. We can express $\mathbf{\Sigma}_{n\_\mathrm{adj}}$ as

$$\mathbf{\Sigma}_{n\_\mathrm{adj}} = \sum_{l=1}^{\mathrm{VD}} \bar{\lambda}_l \bar{\mathbf{v}}_l \bar{\mathbf{v}}_l^T + \sum_{l=\mathrm{VD}+1}^{L} \bar{\lambda}_l \bar{\mathbf{v}}_l \bar{\mathbf{v}}_l^T, \tag{22}$$

where $\{\bar{\mathbf{v}}_l\}_{l=1}^{\mathrm{VD}}$ and $\{\bar{\mathbf{v}}_l\}_{l=\mathrm{VD}+1}^{L}$ span the signal subspace and noise subspace, respectively. The variances of the noise components in the second term of Eq. (22) have been whitened and normalized to unity, i.e., $\bar{\lambda}_l=1$ for $l=\mathrm{VD}+1,\ldots,L$, and $\bar{\lambda}_l>1$ for $l=1,\ldots,\mathrm{VD}$. Then the problem of VD estimation can be formulated as the following binary hypothesis-testing problem:

$$\left.\begin{array}{l} H_0{:}y_l=\bar{\lambda}_l=1 \\ \text{versus} \\ H_1{:}y_l=\bar{\lambda}_l>1 \end{array}\right\} \quad \text{for } l=1,\ldots,L. \tag{23}$$

Each eigenvalue $\bar{\lambda}_l$ under hypotheses H_0 and H_1 can be modeled as a random variable y_l, which has asymptotic conditional Gaussian distributions $\mathcal{N}$ specified by[30]

$$p_0(y_l)=p(y_l|H_0)\sim\mathcal{N}(1,\sigma_{y_l}^2) \quad \text{for } l=1,2,\ldots,L \tag{24}$$

and

$$p_1(y_l)=p(y_l|H_1)\sim\mathcal{N}(\mu_1,\sigma_{y_l}^2) \quad \text{for } l=1,2,\ldots,L, \tag{25}$$

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

$$\sigma_{y_l}^2=\mathrm{Var}[\bar{\lambda}_l]\approx\frac{2\bar{\lambda}_l^2}{N} \tag{26}$$

with N the number of pixel samples.

Now, using Eqs. (23) to (26), we can find the Neyman-Pearson detector δ_{NP} to determine the VD.[24] The major advantage of hypothesis-testing-based VD estimation techniques is the introduction of Neyman-Pearson detection theory for the estimation of the number of distinctive signals, instead of subjectively selecting the "large" eigenvalues or "large" Gerschgorin disks as in Ref. 16.

4 Experiments

When applying an ICA algorithm to hyperspectral image classification, the first several PCs (major PCs) are kept for classification, and the rest are discarded: we assume that most of the object information is presented in major PCs. If some discarded minor PCs contain more information than major PCs, then the final classification results will be degraded. So it is important to compact the important data information into major PCs. We show below that the NAPC transform is a better choice than the PCA for this task.

Two sets of real hyperspectral image data—the AVIRIS "Cuprite" scene and the HYDICE "Forest" scene—were used in the experiments. The former was used for qualitative demonstration of the performance of the NAPC transform for ICA classification, while the latter was used for quantitative evaluation because its pixel-level ground truth is available.

4.1 *AVIRIS Data Experiment—a Qualitative Study*

The AVIRIS "Cuprite" subimage scene, of size 350×350, shown in Fig. 1, was collected in Nevada in 1997. The spatial resolution is about 20 m. Originally it had 224 bands with 0.4- to 2.5-μm spectral range. After water absorption bands and low-SNR bands were removed, 189 bands were used in the experiment. This image scene is well understood mineralogically, and a spectral library of pure minerals is available.[31] The N-FINDER algorithm[32] was used to locate the endmember signatures from the image scene itself, which were compared with the spectral library to get the locations and distribution of pure pixels (endmembers). We found out that five minerals were present: alunite (A), buddingtonite (B), calcite (C), kaolinite (K), and muscovite (M). Their approximate spatial locations of these minerals are marked in Fig. 1(b).

Figure 2 shows the first 20 PCs from PCA. We can see that PC4 and PC6 are very noisy, but they have pretty high ranks in terms of variance. We checked all the PC images, and found out that the data information was spread in many PCs, and even PC71 contained some information. Figure 3 shows the first 20 PCs from the NAPC transform. Now the PC images were ordered in terms of SNR, so a PC image with low image quality was ranked lower. All the data information was spread over the first 37 PC images. Obviously, the NAPC transform can better compact the original data information into major PCs.

The NSP method in Sec. 3.3 was used to estimate the number of distinct signals. The estimate was 23 (for $P_F=0.001$), which was the reference for the number of PCs that ought to be kept. For the purpose of comparison, the first 20 and 30 PCs generated from PCA and the NAPC transform were selected, respectively, and the JADE and FastICA algorithms were applied. The classification results are shown in Figs. 4 and 5, where only the independent components (ICs) related to the five minerals of interest are presented. The classification results by using 30 PCs were better than using 20 PCs, since the classification maps had a higher contrast and pixels belonging to the background were better suppressed. When using 20 PCs from PCA, alunite (A) and kaolinite (K) were not well classified, as shown in Fig. 4(a). But when using 20 PCs from the NAPC transform in Fig. 4(c), the minerals could be better classified. Using 30 PC from the NAPC transform in Fig. 4(d) was also better than using 30 PCs from PCA in Fig. 4(b), particularly when classifying alunite and kaolinite. It is noteworthy that JADE is inapplicable to the original 189-band image data.

FastICA classification is shown in Fig. 5. When 20 PCs from PCA were used for classification in Fig. 5(a), alunite was not classified. When 30 PCs from PCA were used in Fig. 5(b), the alunite classification was improved, although there were still some background pixels being picked up. Figure 5(c), using 20 PCs from the NAPC transform, was

Table 1 The performance of JADE using PCA and the NAPC transform for dimension reduction in the AVIRIS experiment (Corr: correlation coefficient between a classified image using major PCs and the one using all original bands).

	Corr			
Mineral	PCA20	PCA30	NAPC20	NAPC30
Alunite (A)	0.4288	0.6587	0.6950	0.8102
Buddingtonite (B)	0.8396	0.8825	0.8917	0.9611
Calcite (C)	0.7784	0.9237	0.7667	0.9231
Kaolinite (K)	0.4147	0.5368	0.5376	0.8988
Muscovite (M)	0.7219	0.7841	0.7893	0.9179
Average	0.6367	0.7572	0.7361	0.9022

Table 2 The performance of FastICA using PCA and the NAPC transform for dimension reduction in the AVIRIS experiment (Corr: correlation coefficient between a classified image using major PCs with the one using all original bands).

	Corr			
Mineral	PCA20	PCA30	NAPC20	NAPC30
Alunite (A)	0.5530	0.6249	0.7015	0.7738
Buddingtonite (B)	0.8374	0.8822	0.8915	0.9576
Calcite (C)	0.7754	0.9184	0.7651	0.9080
Kaolinite (K)	0.4575	0.5257	0.5122	0.9052
Muscovite (M)	0.7144	0.7912	0.7920	0.8104
Average	0.6676	0.7485	0.7326	0.8710

better than Fig. 5(a), and Fig. 5(d), using 30 PCs from the NAPC transform, was better than Fig. 5(b), because the background in the alunite and kaolinite classification maps was better suppressed. Figure 5(e) is the classification result when all the 189 original bands were used, for FastICA was applicable to the original data set. We can see that the difference between Figs. 5(e), 5(d), and 4(d) is minor. Figure 4(d), using JADE on 30 NAPC components, looks closer to Fig. 5(e) than does Fig. 5(d), using FastICA on 30 NAPC components.

Due to the lack of pixel-level ground truth, we are unable to quantify the classification accuracy. However, quantitative comparison was made between a classified image using major PCs and its counterpart using all 189 original bands (from the FastICA method). Here we assume that the best classification result is provided by using all the bands. The correlation coefficient (Corr) between two images is adopted as the similarity metric for this purpose. The larger the correlation coefficient is, the more similar two images are—in this case, also, the better the classification result. Tables 1 and 2 list the correlation coefficients when JADE and FastICA were applied to classify the five minerals, respectively. We can see that with the same number of PCs being selected, using PCs from the NAPC transform always yielded a more similar result to the one using all original bands, except that the correlation coefficients were slightly smaller when classifying calcite. But the overall performance (average Corr) of the NAPC-based technique is still much better. In addition, the classified images using only 30 PCs from the NAPC transform have an average Corr as great as 0.9 (the highest Corr is 1) with respect to those using all 189 bands. This indicates these 30 PCs from the NAPC transform contain the primary object information.

This experiment provides qualitative evaluation of the NAPC transform for ICA classification: (1) the selection of the same number of PCs from the NAPC transform allows a better ICA classification than from PCA; (2) for achieving similar classification performance, the NAPC transform requires a smaller number of PCs; (3) when an ICA algorithm is applicable to the original high-dimensional data, a comparable classification can be generated by using a smaller number of PC images from the NAPC transform. Compared to the same number of major PCs from PCA, the major PCs from the NAPC transform contain more object information with less noise, which enables an ICA algorithm to distinguish different objects from each other. For each resultant classification map, pixels from other objects and background can be better suppressed.

4.2 *HYDICE Data Experiment—a Quantitative Study*

The HYDICE "Forest" subimage scene, of size 64×64, shown in Fig. 6(a) was collected in Maryland in 1995 from a flight altitude of 10,000 ft with about 1.5-m spatial resolution. The spectral coverage is 0.4 to 2.5 μm. The water absorption bands and low-SNR bands were removed, reducing the data dimensionality from 210 to 169. This scene includes 15 panels arranged in a 5×3 matrix. Each element in this matrix is denoted by p_{ij} with row indexed by $i = 1, \ldots, 5$ and column indexed by $j = a, b, c$. The three panels in the same row (p_{ia}, p_{ib}, p_{ic}) were made from the same material of sizes 3×3, 2×2, and 1×1, respectively, which can be considered as one class, P_i. The ground truth map in Fig. 6(b) shows the precise locations of the panel center pixels. These panel classes have very close spectral signatures, and it is difficult to discriminate them from each other.

Figure 7 shows the first 40 PCs from PCA. We can see that they were not ordered in terms of image quality. For instance, PC21 had noise only and PC35 had information about P_1 and P_2, but PC21 had higher rank than PC35. The same situation happened when ranking minor PC32, PC33, PC35, and PC37, which contained panel information. The original data information was distributed among many PCs, and even PC77 contained some information for panels. Figure 8 shows the first 40 principal components from the NAPC transform. All the first 35 PCs had some information, and the PCs starting from 36 were noisy. Once again, the NAPC transform provided better performance in compacting data information into major PCs.

The NSP method in Sec. 3.3 was used to estimate the number of distinct signals. The estimate was 20 (for P_F

Table 3 The performance of JADE using PCA and the NAPC transform for dimension reduction in the HYDICE experiment (N_C: number of correctly classified pixels; N_F: number of false-alarm pixels; R_{oc}: rate of overall classification).

		PCA20		PCA30		PCA40		NAPC20		NAPC30		NAPC40	
	N_P	N_C	N_F	N_C	N_F	N_C	N_F	N_C	N_F	N_C	N_F	N_C	N_F
P_1	3	2	0	2	0	2	0	2	0	2	0	2	0
P_2	4	4	893	3	0	3	0	3	18	3	0	3	0
P_3	4	3	0	3	0	3	0	3	0	3	0	3	0
P_4	4	4	901	4	243	4	18	3	0	2	0	3	0
P_5	4	3	3	3	3	3	3	3	0	3	0	3	0
Total	19	16	1797	15	246	15	21	14	18	14	0	14	0
R_{oc}		0.3553		0.5147		0.5496		0.6077		0.7368		0.7368	

=0.001), which was the reference for the number of PCs that ought to be kept. For comparison, we selected the first 20, 30, and 40 PCs from PCA and the NAPC transform for ICA classification. In Figs. 9(a)–9(c), the JADE classification results using PCA for dimensional reduction are presented. Only the ICs related to panels are shown here. Obviously, using more PCs could improve the classification. However, even if the first 40 PCs were used in Fig. 9(c), neither P_2 and P_3 nor P_4 and P_5 could be well separated from each other, because their spectral signatures are similar and the data information contained in the first 40 PCs from PCA is not enough for accurate panel discrimination. Figures 9(d)–9(f) show the JADE classification using the PCs from the NAPC transform, where, using only 30 PCs, all the five panel classes could be correctly classified. This means the first 30 PCs from the NAPC transform contain all the panel information for accurate panel discrimination.

FastICA classification results are shown in Fig. 10. Using 20 or 30 PCs from PCA, P_2 and P_3, P_4, and P_5 were not separated, as shown in Fig. 10. Even when 40 PCs were used, in Fig. 10(c), P_4 and P_5 still were not separated. When 20 PCs from the NAPC transform were used, P_2 and P_3 were not separated, as seen in Fig. 10(d); classification was significantly improved in Fig. 10(e), when 30 PCs from the NAPC transform were used; all panels were cor-

Table 4 The performance of FastICA using PCA and the NAPC transform for dimension reduction in the HYDICE experiment (N_C: number of correctly classified pixels; N_F: number of false-alarm pixels; R_{oc}: rate of overall classification).

		PCA20		PCA30		PCA40		NAPC20		NAPC30		NAPC40		All bands	
	N_P	N_C	N_F	N_C	N_F	N_C	N_F	N_C	N_F	N_C	N_F	N_C	N_F	N_C	N_F
P_1	3	2	0	2	0	2	0	2	0	2	0	2	0	2	0
P_2	4	—	—	3	0	3	0	3	31	3	0	3	0	3	0
P_3	4	—	—	3	0	3	0	3	3	3	0	3	0	3	0
P_4	4	—	—	—	—	4	574	3	0	2	0	3	0	3	0
P_5	4	—	—	—	—	3	3	3	0	3	0	3	0	3	0
Total	19	—	—	—	—	15	577	14	34	14	0	14	0	14	0
R_{oc}		—		—		0.5127		0.5293		0.7368		0.7368		0.7368	

rectly classified in Fig. 10(f), when 40 PCs from the NAPC transform were used. In Fig. 10(g), FastICA was applied to all the 169 original bands; the 40-NAPC results in Fig. 10(f) were comparable.

Since the pixel-level ground truth is available for this HYDICE scene, quantitative study was conducted as shown in Tables 3 and 4. Table 3 lists the results for the JADE classification, where N_C denotes the number of correctly classified pure panel pixels, and N_F denotes the number of false-alarm pixels. The numbers of pure pixels for the five panel classes, N_{P_i}, are 3, 4, 4, 4, and 4, respectively, and the total number of pure pixels, N_P, is 19. Before being compared with the ground truth, each gray scale classification map was converted into binary by setting the threshold at the central value of the grayscale range. As listed in Table 3, 14 out of 19 panel pixels were correctly classified with no false alarms when using the first 30 or 40 PCs from the NAPC transform. But when using PCA, even if the first 40 PCs were selected, there were still some false-alarm pixels (N_F=14 in this case). It should be noted that in this image scene the five panels in the rightmost column were smaller than the area covered by a single pixel. This is why these five panel pixels were missed. The overall classification rate R_{oc} was calculated using the definition in Ref. 33: $R_{\mathrm{oc}}=\Sigma_{i=1}^{5}p(P_i)R_C(P_i)$, where $p(P_i)=\Sigma_{i=1}^{5}N_{P_i}/N_P$ is the occurrence rate of P_i, and $R_C(P_i)=N_C(P_i)/[N_{P_i}+N_F(P_i)]$ is the classification rate of P_i. We can see that the R_{oc} from an NAPC-based technique is always greater than its counterpart from a PCA-based technique. It should be noted that in this image scene the five panels in the rightmost column were smaller than the area covered by a single pixel (i.e., the spatial resolution). This is why these five panel pixels were missed and the largest R_{oc} with N_F=0 is 0.7368.

The quantitative study results about the FastICA classification are listed in Table 4, which shows that 30 and 40 PCs from the NAPC transform provided comparable classification to the one using all the original bands. But using 40 PCs from PCA, the number of false-alarm pixels, N_F, was as large as 577. If the threshold is changed, then the number of correctly classified pixels (N_C) and false-alarm pixels (N_F) will also be changed. But the NAPC-based technique always provides a comparable N_C with a much smaller N_F than the PCA-based technique using the same number of PCs. Therefore, it provided larger classification rates.

This HYDICE experiment further demonstrates that the NAPC transform is a better preprocessing method than PCA when ICA is applied to high-dimensional image data, where a smaller number of PCs permits better classification performance for the following ICA-based unsupervised classification. In this case, 30 PCs from the NAPC transform provide the same classification rate as do all the bands.

5 Conclusion

Independent-component analysis (ICA) is a popular tool for unsupervised classification. But its very high computational complexity impedes its application to high-dimensional data analysis. The common approach is to use principal-component analysis (PCA) to reduce the data dimensionality before applying the ICA classification. When dealing with image data, we find that it may be more appropriate to use a noise-adjusted principal-components (NAPC) transform for this preprocessing step. The principal components from the NAPC transform are ranked in terms of image quality. As a result, object information can be better compacted into major principal components (PCs). When an ICA algorithm is executed on these major PCs, classification can be improved. The classification result can be comparable or even identical to that when all the original bands are used. The major difficulty with the NAPC transform is the estimation of the noise covariance matrix. The discussed estimation method based on interband correlation is very simple and seems to be effective in our real-data experiments.

References

1. R. B. Singer and T. B. McCord, “Mars: large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance,” in *Proc. 10th Lunar and Planetary Science Conf.*, pp. 1835–1848, Lunar and Planetary Inst., Houston (1979).
2. J. B. Adams and M. O. Smith, “Spectral mixture modeling: a new analysis of rock and soil types at the Viking lander 1 suite, ” *J. Geophys. Res.* **91**(B8), 8098–8112 (1986).
3. J. J. Settle and N. A. Drake, “Linear mixing and estimation of ground cover proportions,” *Int. J. Remote Sens.* **14**, 1159–1177 (1993).
4. 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., pp. 145–166, Cambridge University Press (1993).
5. C. Jutten and J. Herault, “Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture,” *Signal Process.* **24**(1), 1–10 (1991).
6. J. F. Cardoso and A. Souloumiac, “Blind beamforming for non-gaussian signals,” *IEE Proc. F, Radar Signal Process.* **140**(6), 362–370 (1993).
7. P. Comon, “Independent component analysis—a new concept?” *Signal Process.* **36**, 287–314 (1994).
8. J. F. Cardoso and B. Laheld, “Equivalent adaptive source separation,” *IEEE Trans. Signal Process.* **44**(12), 3017–3030 (1996).
9. A. Cichocki and R. Unbehauen, “Robust neural networks with online learning for blind identification and blind separation of sources,” *IEEE Trans. Circuits Syst., I: Fundam. Theory Appl.* **43**(11), 894–906 (1996).
10. A. Hyvarinen and E. Oja, “A fast fixed-point algorithm for independent component analysis,” *Neural Comput.* **9**(7), 1483–1492 (1997).
11. A. Hyvarinen, “Fast and robust fixed-point algorithms for independent component analysis,” *IEEE Trans. Neural Netw.* **10**(3), 626–634 (1999).
12. H. Szu, “Independent component analysis (ICA): an enabling technology for intelligent information/image technology (IIT),” *IEEE Circuits Devices Mag.* **10**, 14–37 (1999).
13. A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons (2001).
14. H. Szu and J. Buss, “ICA neural net to refine remote sensing with multiple labels,” *Proc. SPIE* **4056**, 32–49 (2000).
15. T. Yoshida and S. Omatu, “Pattern recognition with neural networks,” in *Proc. Int. Geoscience and Remote Sensing Symp.*, pp. 699–701, IEEE (2000).
16. T. Tu, “Unsupervised signature extraction and separation in hyperspectral images: a noise-adjusted fast independent component analysis approach,” *Opt. Eng.* **39**(4), 897–906 (2000).
17. 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.* **GE–40**(2), 375–392 (2002).
18. X. Zhang and C. H. Chen, “New independent component analysis method using high order statistics with application to remote sensing images,” *Opt. Eng.* **41**(7), 1717–1728 (2002).
19. S. Fiori, “Overview of independent component analysis technique with an application to synthetic aperture radar (SAR) imagery processing,” *Neural Networks* **16**, 453–467 (2003).
20. 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.* **GE–26**, 65–74 (1988).
21. 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.* **GE–28**(3), 295–304 (1990).

22. R. E. Roger, "A faster way to compute the noise-adjusted principal components transform matrix," *IEEE Trans. Geosci. Remote Sens.* **GE–32**(6), 1194–1196 (1994).
23. C.-I Chang and Q. Du, "Interference and noise adjusted principal components analysis," *IEEE Trans. Geosci. Remote Sens.* **GE–37**(9), 2387–2396 (1999).
24. C.-I Chang and Q. Du, "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.* **GE–42**, 608–619 (2004).
25. G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins Univ. Press (1996).
26. M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, 2nd ed., Wiley (1993).
27. A. Ralston and P. Rabinowitz, *A First Course in Numerical Analysis*, 2nd ed., McGraw-Hill (1978).
28. R. E. Roger and J. F. Arnold, "Reliably estimating the noise in AVIRIS hyperspectral imagers," *Int. J. Remote Sens.* **17**(10), 1951–1962 (1996).
29. R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, Wiley (1982).
30. T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 3rd. ed., Wiley-Interscience (2003).
31. http://speclab.cr.usgs.gov/spectral.lib04/spectral-lib04.html.
32. M. E. Winter, "N-FINDER: an algorithm for fast autonomous spectral endmember determination in hyperspectral data," *Proc. SPIE* **3753**, 266–275 (1999).
33. C.-I Chang and H. Ren, "An experiment-based quantitative and comparative analysis of target detection and image classification algorithms for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.* **GE–38**(2), 1044–1063 (2000).

Qian Du received her PhD degree in electrical engineering from University of Maryland Baltimore County in 2000. She was an assistant professor in the Department of Electrical Engineering and Computer Science at Texas A&M University-Kingsville from 2000 to 2004. Since fall 2004, she has been with the Department of Electrical and Computer Engineering at Mississippi State University as an assistant professor. Dr. Du has been working on remote sensing image analysis for many years, with expertise on hyperspectral imaging. She is a member of SPIE, IEEE, ASPRS, and ASEE.

Ivica Kopriva received his BS degree in electrical engineering from Military Technical Faculty, Zagreb, Croatia, in 1987, and MS and PhD degrees in electrical engineering from the Faculty of Electrical Engineering and Computing, Zagreb, Croatia, in 1990 and 1998, respectively. Currently, he is a senior research scientist at George Washington University, Department of Electrical and Computer Engineering. His work is focused on the theory of unsupervised learning with application to solving blind imaging problems. He also works on second- and higher-order-statistics-based electromagnetic array signal processing with applications to direction finding.

Harold Szu received his PhD degree in physics from Rockefeller University in 1971. He was with the Naval Research Lab from 1977 to 1990. From 1990 to 1996 he led the Information Science Group at Naval Surface Warfare Center. He returned to serve as a program officer at the Office of Naval Research since 1997. He is also affiliated with the George Washington University as a research professor and director of the Digital Media RF Lab. Dr. Szu has been a champion of brain-style computing for decades. He is a founder, former president, and current governor of the International Neural Network Society (INNS). He received the INNS D. Gabor Award for his outstanding contribution to neural network applications in information sciences and pioneer implementations of fast simulated-annealing search in 1997, and the Italy Academy Eduardo R. Caianiello Award for elucidating and implementing a chaotic neural net as a dynamic realization for fuzzy logic membership function in 1999. Recently, he has contributed to the unsupervised-learning theory of the thermodynamic free energy of sensory pairs for fusion. He has published numerous journal papers, books, and book chapters. Dr. Szu is a fellow of SPIE, OSA, IEEE, and AIMBE.