

INDEPENDENT COMPONENT ANALYSIS FOR CLASSIFYING MULTISPECTRAL IMAGES WITH DIMENSIONALITY LIMITATION

QIAN DU

*Department of Electrical and Computer Engineering
Mississippi State University, MS 39762, USA*

IVICA KOPRIVA and HAROLD SZU

*Department of Electrical and Computer Engineering
The George Washington University, Washington, DC 20052, USA*

Received 21 July 2004

Accepted 22 August 2004

Airborne and spaceborne remote sensors can acquire invaluable information about earth surface, which have many important applications. The acquired information usually is represented as two-dimensional grids, i.e. images. One of techniques to processing such images is Independent Component Analysis (ICA), which is particularly useful for classifying objects with unknown spectral signatures in an unknown image scene, i.e. unsupervised classification. Since the weight matrix in ICA is a square matrix for the purpose of mathematical tractability, the number of objects that can be classified is equal to the data dimensionality, i.e. the number of spectral bands. When the number of sensors (or spectral channels) is very small (e.g. a 3-band CIR photograph and 6-band Landsat image with the thermal band being removed), it is impossible to classify all the different objects present in an image scene using the original data. In order to solve this problem, we present a data dimensionality expansion technique to generate artificial bands. Its basic idea is to use nonlinear functions to capture and highlight the similarity/dissimilarity between original spectral measurements, which can provide more data with additional information for detecting and classifying more objects. The results from such a nonlinear band generation approach are compared with a linear band generation method using cubic spline interpolation of pixel spectral signatures. The experiments demonstrate that nonlinear band generation approach can significantly improve unsupervised classification accuracy, while linear band generation method cannot since no new information can be provided. It is also demonstrated that ICA is more powerful than other frequently used unsupervised classification algorithms such as ISODATA.

Keywords: Independent component analysis; unsupervised classification; data dimensionality; remotely sensed imagery; multispectral imagery; aerial photograph.

1. Introduction

Airborne and spaceborne remote sensors can be used to measure the object properties on the earth's surface and monitor the human activities on the earth. When information is acquired by collecting the solar energy that is reflected from the earth surface and displayed in two-dimensional grids, the resultant products are images in optical remote sensing. In some cases, there are hundreds of sensors operating at contiguous spectral ranges for capturing co-registered images, which is called hyperspectral imaging. While in some other cases, the number of sensors is small and only several spectral bands are available, which is called multispectral imaging. Hyperspectral images have abundant spectral information for object detection, classification, and identification, but image analysis is quite complicated due to the data complexity. The image analysis of multispectral images is relatively simple, but we may meet the data dimensionality limitation in some occasions. When the hardware condition (e.g. the number of sensors) cannot be improved, we may resort to some signal processing techniques to explore additional information from the multispectral data that is original acquired.

As remote sensing and its applications receive lots of interests recently, many algorithms in remotely sensed image analysis have been proposed. While they have achieved a certain level of success, most of them are supervised methods, i.e. the information of the objects to be detected and classified is assumed to be known *a priori*. If such information is unknown, the task will be much more challenging. Since the area covered by a single pixel is very large, the reflectance of a pixel can be considered as the mixture of all the materials resident in the area covered by the pixel. Therefore, we have to deal with mixed pixels instead of pure pixels as in conventional digital image processing. Linear spectral unmixing analysis is a popularly used approach in remote sensing image processing to uncover material distribution in an image scene [Singer & McCord, 1979; Adams & Smith, 1986; Adams, Smith & Gillespie, 1993; Settle & Drake, 1993]. Let N be the number of bands (i.e. spectral channels) and $\mathbf{r} = (r_1 r_2 \cdots r_N)^T$ a column

pixel vector with dimensionality N in an image. An element r_i in the \mathbf{r} is the reflectance collected at the i th band. Assume that there are a set of distinct materials (endmembers) present in the image scene and all the pixel reflectances are assumed to be linear mixture from these endmembers. Let \mathbf{M} denote a signature matrix. Each column vector in \mathbf{M} represents an endmember signature, i.e. $\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_p]$, where p is the total number of endmembers resident in the image scene. Let $\boldsymbol{\alpha} = (\alpha_1 \alpha_2 \cdots \alpha_p)^T$ be the unknown abundance column vector associated with \mathbf{M} , which is to be estimated. The i th element α_i in $\boldsymbol{\alpha}$ represents the abundance fraction of \mathbf{m}_i in pixel \mathbf{r} . According to the linear mixture model,

$$\mathbf{r} = \mathbf{M}\boldsymbol{\alpha} + \mathbf{v}, \quad (1)$$

where \mathbf{v} is the noise term. If the endmember signature matrix \mathbf{M} is known, then $\boldsymbol{\alpha}$ can be approximated by using its least squares estimate $\hat{\boldsymbol{\alpha}}$ to minimize the following estimation error:

$$e = \|\mathbf{r} - \mathbf{M}\hat{\boldsymbol{\alpha}}\|^2. \quad (2)$$

If $\hat{\alpha}_i$ for all the pixels are plotted, a new image is generated to describe the distribution of the i th endmember \mathbf{m}_i , where a bright pixel represents high abundance of the i th material in the image scene and a dark pixel represents low abundance. As a result, soft classification can be achieved.

In practice, it may be difficult to have prior information about the image scene and endmember signatures. Moreover, in-field spectral signatures may be different from those in spectral libraries due to atmospheric and environmental effects. So an unsupervised classification approach is preferred. Independent component analysis (ICA) has been successfully applied to blind source separation [Jutten & Herault, 1991; Cardoso & Souloumiac, 1993; Comon, 1994; Cardoso & Laheld, 1996; Cichocki & Unbehauen, 1996; Szu, 1999, 2003, 2004]. The basic idea of ICA is to decompose a set of multivariate signals into a base of statistically independent sources with the minimal loss of information content so as to achieve

classification. The standard linear ICA-based data model with additive noise is

$$\mathbf{u} = \mathbf{A}\mathbf{s} + \mathbf{v}, \quad (3)$$

where $\mathbf{u} \in R^{N \times M}$ is an N dimensional data vector consisted of M samples (for an image data, M is the number of pixels), $\mathbf{A} \in R^{N \times N}$ is an unknown mixing matrix, and $\mathbf{s} \in R^{N \times M}$ is an unknown source signal vector. Three assumptions are made on the unknown source signals \mathbf{s} :

- (1) each source signal is an independent and 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 Gaussian distribution.

The mixing matrix \mathbf{A} although unknown is also assumed to be non-singular. Then the solution to the blind source separation problem is obtained with the scale and permutation indeterminacy, i.e. $\mathbf{Q} = \mathbf{W}\mathbf{A} = \mathbf{P}\mathbf{\Lambda}$, where \mathbf{W} represents the unmixing matrix, \mathbf{P} is generalized permutation matrix and $\mathbf{\Lambda}$ is a diagonal matrix. These requirements ensure the existence and uniqueness of the solution to the blind source separation problem (except for ordering, sign, and scaling). Comparing to many conventional techniques, which use up to the second order statistics only, ICA exploits high order statistics that makes it a more powerful method in extracting irregular features in the data [Hyvarinen, Karhunen & Oja, 2001].

Several researchers have explored the ICA to remote sensing image classification. Szu and Buss [2000], and Yoshida and Omatu [2000] investigated ICA classifiers for multispectral images. Tu [2000] and Chang *et al.* [2002] developed their own ICA-type approaches to hyperspectral images. Chen and Zhang [1999] and Fiori [2003] presented their ICA results for SAR image processing. In general, when we intend to apply the ICA approach to classify optical multispectral/hyperspectral images, the linear mixture model in Eq. (1) needs to be reinterpreted so as to fit in the ICA model

in Eq. (3). Specifically, the endmember matrix \mathbf{M} in Eq. (1) corresponds to the unknown mixing matrix \mathbf{A} in Eq. (3) and abundance fractions in $\boldsymbol{\alpha}$ in Eq. (1) correspond to source signals in \mathbf{s} in Eq. (3). Moreover, the p abundance fractions $\alpha_1, \alpha_2, \dots, \alpha_p$ are considered as unknown random quantities specified by random signal sources rather than unknown deterministic quantities as assumed in the original linear mixture model (1). With these interpretations and above-mentioned assumptions, we will use model (3) to replace model (1) thereafter. The advantages offered by using model (3) in remote sensing image classification are:

- (1) no prior knowledge of the endmembers in the mixing process is required;
- (2) the spectral variability of the endmembers can be accommodated by the unknown mixing matrix of \mathbf{A} since the source signals are considered as scalar and random quantities; and
- (3) higher order statistics can be exploited for better feature extraction and pattern classification.

In the ICA model (3), \mathbf{A} is generally set to be a square matrix for mathematical tractability. Then for an N -band image, we can separate and classify N materials/objects/classes, i.e. $p = N$ in model (1). When dealt with a multispectral image or a color composite image (3 RGB bands or CIR bands), this brings about a limitation to classification performance. Because in most of practical applications such as land cover mapping, the number of land cover patterns in an image scene is usually not a small value. In order to let ICA technique be applicable to such cases and satisfying classification be achievable, we will introduce a data dimensionality expansion approach to resolve this problem, which generates artificial spectral measurements to explore additional information in the original data. To ensure such new measurements are linearly independent of the original data, a nonlinear function is to be adopted. In our experiments, an ICA method referred to as Joint Approximate Diagonalization of Eigenmatrices (JADE) algorithm is used, although any other ICA algorithm

such as FastICA [Hyvarinen & Oja, 1997] can be used as well.

The remainder of this paper is organized as follows. Section 2 describes the JADE ICA algorithm. Section 3 introduces a dimensionality expansion technique for the application of the JADE algorithm to remote sensing images. Section 4 presents the qualitative and quantitative experimental results with classification performance comparison. And Sec. 5 presents some concluding remarks.

2. ICA and JADE

The strategy of ICA algorithms is to find a linear transform \mathbf{W} (i.e. the unmixing matrix of size $N \times N$):

$$\mathbf{z} = \mathbf{W}\mathbf{u} = \mathbf{W}\mathbf{A}\mathbf{s} + \mathbf{W}\mathbf{v} = \mathbf{Q}\mathbf{s} + \mathbf{W}\mathbf{v}, \quad (4)$$

such that components of the vector \mathbf{z} are as statistically independent as possible. Based on the assumption that source signals \mathbf{s} are mutually statistically independent and non-Gaussian (except one that is allowed to be Gaussian), the vector \mathbf{z} will represent vector of the source signals \mathbf{s} up to the permutation, scale and sign factors.

There are several different types of ICA algorithms developed recently, such as non-Gaussianity maximization based ICA [Hyvarinen & Oja, 1997], maximum likelihood estimation based ICA [Pham, 1997; Bell & Sejnowski, 1996; Lee, Girolami & Sejnowski, 1999], mutual information minimization based ICA [Comon, 1994], nonlinear decorrelation based ICA [Cichocki & Unbehauen, 1994, 1996], and nonlinear ICA [Oja, 1997]. Here, we select the popular Joint Approximate Diagonalization of Eigenmatrices (JADE) algorithm [Cardoso & Souloumiac, 1993] which is based on the usage of the fourth order statistics (cumulants). The higher order statistical dependence among data samples is measured by the higher order cross-cumulants. Smaller values of the cross-cumulants represent less dependent samples. In JADE the fourth order statistical independence is achieved through minimization of the squares of the fourth order cross-cumulants between the

components z_i of \mathbf{z} in (4). The fourth-order cross-cumulants can be computed as

$$\begin{aligned} C_4(z_i, z_j, z_k, z_l) \\ = E[z_i z_j z_k z_l] - E[z_i z_j]E[z_k z_l] \\ - E[z_i z_k]E[z_j z_l] - E[z_i z_l]E[z_j z_k] \end{aligned} \quad (5)$$

for $1 \leq i, j, k, l \leq N$, where $E[\cdot]$ denotes mathematical expectation operator. Equation (5) can be expressed as an $N^2 \times N^2$ Hermitian matrix. The optimal unmixing matrix \mathbf{W}^* is the one that satisfies

$$\mathbf{W}^* = \arg \min \sum_{i,j,k,l} \text{off}(\mathbf{W}^T C_4(z_i, z_j, z_k, z_l) \mathbf{W}), \quad (6)$$

where $\text{off}(\cdot)$ is a measure for the off-diagonality of a matrix defined as

$$\text{off}(\mathbf{X}) = \sum_{1 \leq i \neq j \leq N} |x_{ij}|^2, \quad (7)$$

where x_{ij} is the ij th element of a matrix \mathbf{X} . An intuitive way to solving the optimization problem expressed in Eq. (6) is to jointly diagonalize the eigenmatrices of $C_4(z_i, z_j, z_k, z_l)$ in Eq. (5) via Givens rotation sweeps. In order for the first and second order statistics not to affect the results, the data \mathbf{u} in Eqs. (3) and (4) should be pre-whitened (mean is zero and covariance matrix is the identity matrix).

One of the advantages of using such a fourth order cumulant based ICA algorithm is its capability to suppress additive Gaussian noise, because the fourth order cumulants asymptotically vanish for Gaussian processes, i.e. they are blind w.r.t. Gaussian noise. Based on the multilinearity property of the cumulants, we can rewrite the fourth order cross-cumulants of the \mathbf{z} in Eq. (4) as [McCullagh, 1995]:

$$\begin{aligned} C_{kl}(z_i, z_j) = \sum_{n=1}^N q_{in}^k q_{jn}^l C_4(s_n) \\ + \sum_{n=1}^N w_{in}^k w_{jn}^l C_4(v_n), \end{aligned} \quad (8)$$

where q and w are the corresponding elements of the matrices \mathbf{Q} and \mathbf{W} in C_4 . Now thanks to the fact that the fourth order cumulants are

blind w.r.t. Gaussian noise, i.e. $C_4(\mathbf{v}) = 0$, Eq. (8) becomes

$$C_{kl}(z_i, z_j) \cong \sum_{n=l}^N q_{in}^k q_{jn}^l C_4(s_n) \quad \text{for } k, l \in \{1, 2, 3\} \text{ and } k + l = 4. \quad (9)$$

Therefore, in addition to recovering the unknown source signals, the JADE algorithm can successfully suppress additive Gaussian noise.

A critical factor in the ICA applications is the level of singularity/non-singularity of the mixing matrix \mathbf{A} . Although \mathbf{A} is unknown, we have assumed it to be non-singular and invertible. Hence, in the context of our application of ICA in remote sensing image analysis, the information taken by different spectral channels must be distinct enough in order to make measurements linearly independent and sources to be blindly separable. This requirement can be met when we use a data dimensionality expansion process to generate artificial spectral measurements, which is to be discussed in the following section.

3. ICA for Multispectral Image Classification

For the purposes of simplicity and mathematical tractability, the mixing and unmixing matrices \mathbf{A} and \mathbf{W} are assumed to be square, i.e. the number of source signals to be classified p has to be equal to the data dimensionality N . For instance, if there are six classes present in an image scene, we need to have six spectral channels (sensors) to collect measurement data. For many remote sensing images for commercial purposes, such as color-infrared (CIR) photograph and multispectral images (e.g. 3-band SPOT, 4-band QuickBird, 4-Band Ikonos, 6-band Landsat with the thermal band being removed), it is difficult for the ICA technique to achieve fine classification since the number of classes present in an image scene is generally larger than the number of spectral bands. When the hardware condition (the number of sensors) is limited, we may have

to resort to signal processing approaches to increase the data dimensionality so as to provide additional information for classification improvement.

3.1. Nonlinear band generation

A simplest solution to relaxing the data dimensionality is to adopt some nonlinear functions to generate a set of artificial images as additional linearly independent spectral measurements [Ren and Chang, 2000]. Although theoretically any nonlinear function can be used, some specific rules need to be set for nonlinear function selection. Based on our experimental studies, we find out that a nonlinear function that can enlarge or emphasize the discrepancy between original spectral measurements will help to improve classification performance, since the techniques applied here use spectral information. A simplest but effective choice is multiplication. When two original images \mathbf{B}_i and \mathbf{B}_j are multiplied together (each pair of corresponding pixels at the same location is multiplied), a new image is generated, i.e. $\{\mathbf{B}_i \mathbf{B}_j\}_{i,j=1, i \neq j}^N$. Here, multiplication acts as match filtering. When the multiplicand and multiplier are equal, the product is the maximum (at the quantity level of the multiplicand and multiplier). So multiplication can emphasize the spectral similarity between two spectral measurements of the same pixel, which is equivalent to emphasizing their dissimilarity or discrepancy. Multiplication can be also used for a single band, i.e. $\{\mathbf{B}_i \mathbf{B}_i\}_{i=1}^N$. Then it emphasizes a single spectral measurement itself, which is also equivalent to enlarging the spectral difference from other spectral measurements of this pixel. Adding these two sets of artificial bands, there are totally $N + N + \binom{N}{2} = N^2/2 + 3N/2$ bands available for processing. According to our experience, these two sets of spectral bands can provide useful additional information for detection and classification. Another advantage of using multiplication is the great simplicity of chip design for practical implementation [Du and Nekovei, 2005].

Let us assume there are four three-dimensional pixel vectors as listed in Table 1, which represent four class means. The first pixel

Table 1. Four three-dimensional pixel vectors with six nonlinearly generated bands.

	\mathbf{B}_1	\mathbf{B}_2	\mathbf{B}_3	$\mathbf{B}_1\mathbf{B}_2$	$\mathbf{B}_2\mathbf{B}_3$	$\mathbf{B}_1\mathbf{B}_3$	$(\mathbf{B}_1)^2$	$(\mathbf{B}_2)^2$	$(\mathbf{B}_3)^2$
\mathbf{r}_1	0.2500	0.5000	0.7500	0.1250	0.3750	0.1875	0.0625	0.2500	0.5625
\mathbf{r}_2	0.3400	0.4462	0.7714	0.1517	0.3442	0.2623	0.1156	0.1991	0.5950
\mathbf{r}_3	0.2472	0.5783	0.8024	0.1429	0.4640	0.1984	0.0611	0.3344	0.6439
\mathbf{r}_4	0.2413	0.4037	0.8143	0.0974	0.3287	0.1965	0.0582	0.1630	0.6631

vector $\mathbf{r}_1 = (0.25, 0.5, 0.75)^T$, and \mathbf{r}_2 , \mathbf{r}_3 , \mathbf{r}_4 are generated by adding random values within $(-0.1, +0.1)$ to the elements of \mathbf{r}_1 . So the spectral signatures of these four vectors are very close, and it is difficult to differentiate them from each other. Since there are only three bands, we need to generate more bands for their classification. The above-mentioned techniques are used to generate additional six bands, and the resulting vectors are nine-dimensional. The spectral difference can be quantified by using spectral angle mapper (SAM) [Kruse *et al.*, 1993], which calculates the angle between two unit vectors, i.e. $d(\mathbf{r}_i, \mathbf{r}_j) = \cos^{-1}\left(\frac{\mathbf{r}_i^T \mathbf{r}_j}{|\mathbf{r}_i| \cdot |\mathbf{r}_j|}\right)$. If $\mathbf{r}_i = \mathbf{r}_j$, then $d(\mathbf{r}_i, \mathbf{r}_j) = 0$; if $\mathbf{r}_i \perp \mathbf{r}_j$, then $d(\mathbf{r}_i, \mathbf{r}_j) = \frac{\pi}{2}$. For all other cases, $0 < d(\mathbf{r}_i, \mathbf{r}_j) < \frac{\pi}{2}$. A large $d(\mathbf{r}_i, \mathbf{r}_j)$ means \mathbf{r}_i and \mathbf{r}_j are very dissimilar and easy to be differentiated from each other. Tables 2 and 3 list the spectral distances between each pair of vectors before and after data dimensionality expansion. We can see that the distances of the constructed

nine-dimensional vectors in Table 3 are larger than their counterparts in Table 2, which means that using a simple nonlinear function such as multiplication the spectral differences between these four classes can be increased. In other words, their separability is increased. As a result, not only the data dimensionality is expanded to make an ICA algorithm applicable, but also the classification results can be improved.

3.2. Band interpolation

An intuitive method to generate artificial bands is to interpolate the spectral signature of each pixel. Let $\mathbf{r} = (r_1 r_2 r_3)^T$ denote a pixel vector in a 3-band image, and r_1 , r_2 , and r_3 are reflectance in band \mathbf{B}_1 , \mathbf{B}_2 , and \mathbf{B}_3 , respectively. In order to expand the data dimensionality, we may interpolate an element r_4 using r_1 and r_2 , r_5 using r_2 and r_3 . Then the r_4 of all the pixels construct a new band denoted as $I(\mathbf{B}_1\mathbf{B}_2)$ and r_5 as $I(\mathbf{B}_2\mathbf{B}_3)$. These five bands can be used for classification. If more bands are needed, more elements will be interpolated between $[r_1, r_2]$ and $[r_2, r_3]$. A classical cubic spline interpolation can be used for this purpose [Unser, 1999]. However, because of the linear nature of such interpolation, the generated bands do not contain more information. So the classification results may not be improved, as to be demonstrated in the experiments.

Table 4 lists the same four 3D vectors as in Table 1. For each pixel, three measurements are generated by cubic spline interpolation between its values in \mathbf{B}_1 and \mathbf{B}_2 , and another three measurements are interpolated between the values in \mathbf{B}_2 and \mathbf{B}_3 . The spectral distances of these 9D vectors are calculated in Table 5. Compared to Table 2, they are not larger than the original distances between the four 3D vectors. This

Table 2. SAM-based spectral distances between original 3D vectors.

	\mathbf{r}_1	\mathbf{r}_2	\mathbf{r}_3	\mathbf{r}_4
\mathbf{r}_1	0	0.1117	0.0436	0.1238
\mathbf{r}_2		0	0.1529	0.1215
\mathbf{r}_3			0	0.1596
\mathbf{r}_4				0

Table 3. SAM-based spectral distances between 9D vectors (nonlinearly generated).

	\mathbf{r}_1	\mathbf{r}_2	\mathbf{r}_3	\mathbf{r}_4
\mathbf{r}_1	0	0.1275	0.0663	0.1507
\mathbf{r}_2		0	0.1800	0.1409
\mathbf{r}_3			0	0.1962
\mathbf{r}_4				0

Table 4. Four three-dimensional pixel vectors with six linearly generated bands.

	\mathbf{B}_1	\mathbf{B}_2	\mathbf{B}_3	$\mathbf{I}(\mathbf{B}_1\mathbf{B}_2)_1$	$\mathbf{I}(\mathbf{B}_1\mathbf{B}_2)_2$	$\mathbf{I}(\mathbf{B}_1\mathbf{B}_2)_3$	$\mathbf{I}(\mathbf{B}_2\mathbf{B}_3)_1$	$\mathbf{I}(\mathbf{B}_2\mathbf{B}_3)_2$	$\mathbf{I}(\mathbf{B}_2\mathbf{B}_3)_3$
\mathbf{r}_1^*	0.2500	0.5000	0.7500	0.3125	0.3750	0.4375	0.5625	0.6250	0.6875
\mathbf{r}_2^*	0.3400	0.4462	0.7714	0.3491	0.3731	0.4071	0.4992	0.5745	0.6670
\mathbf{r}_3^*	0.2472	0.5783	0.8024	0.3405	0.4274	0.5070	0.6429	0.7024	0.7559
\mathbf{r}_4^*	0.2413	0.4037	0.8143	0.2611	0.2982	0.3474	0.4755	0.5713	0.6859

Table 5. SAM-based spectral distances between original 9D vectors (linearly generated).

	\mathbf{r}_1^*	\mathbf{r}_2^*	\mathbf{r}_3^*	\mathbf{r}_4^*
\mathbf{r}_1^*	0	0.0889	0.0335	0.1154
\mathbf{r}_2^*		0	0.1189	0.1108
\mathbf{r}_3^*			0	0.1440
\mathbf{r}_4^*				0

means using the measurements generated by a linear method cannot improve the classification because no more independent spectral information is carried out, no matter how sophisticated such a linear method is. Interestingly, the spectral distances in Table 5 are even smaller than the original in Table 2. This is due to the fact that more training samples are required to separate two similar classes when the data dimensionality is increased [Hughes, 1968; Landgrebe, 2002], which makes it more difficult to achieve comparable class separation.

4. Experiments

4.1. Qualitative study

Example 1: CIR image experiment

Figure 1 shows a CIR composite image about the urban/suburban area of Houghton, Michigan, which has three bands. There are buildings, vegetation, and water bodies present in this image scene. ICA classification results using the original 3-bands were shown in Fig. 2. The urban area with part of vegetation (e.g. grassland) was classified in independent component 1 (IC1), lake and river were classified in IC2, and other vegetation (e.g. broadleaf trees) was classified in IC3. Three new bands ($\mathbf{B}_1\mathbf{B}_2, \mathbf{B}_2\mathbf{B}_3, \mathbf{B}_1\mathbf{B}_3$) were nonlinearly generated, which were used together

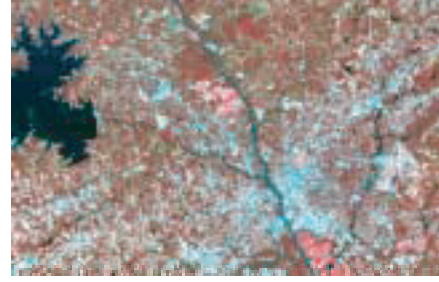


Fig. 1. An original CIR image.

with three original bands for ICA classification. As shown in Fig. 3, finer classification results were provided, where the urban area with high residential density was classified in IC2, buildings (with another type of roofing) were classified in IC1, vegetation (e.g. grassland) was classified in IC3 and vegetation (e.g. broadleaf trees) was classified in IC4. It is very interesting that both lake and river were classified into two different ICs: IC5 and IC6. The reason is that they have different clarity and depth. The river flowing through the urban area is more turbid with suspended matters, and water is relatively shallow. Part of the lake was also displayed with gray shades in the classified river image in IC6, which means the water clarity and depth in different lake areas is different. For instance, the water close to the shore is more turbid and shallow. It should be noted that no more classes were produced if more artificial bands were generated.

Figure 4 shows the ICA classification results using original 3-band with two cubic spline interpolated bands. We can see that IC2, IC3 and IC4 presented the same class information, which corresponds to IC1 in Fig. 2 using the original 3-band image, IC1 is the same as IC2 in Fig. 2, and IC5 is the same as IC3 in Fig. 2. Using the original 3-band image, only part of vegetation was classified, river and lake were claimed as a single class, and all others were grouped

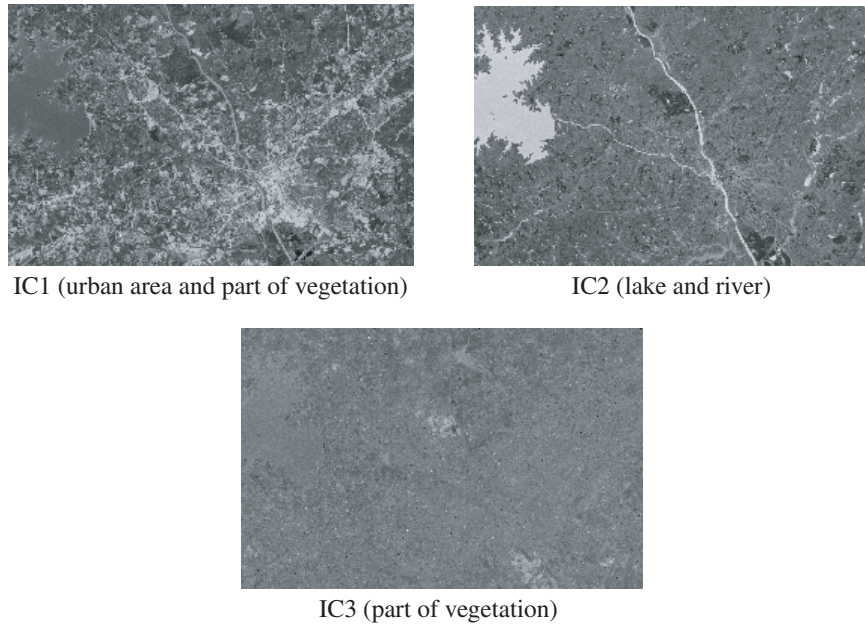


Fig. 2. Classification using original 3-band CIR image.

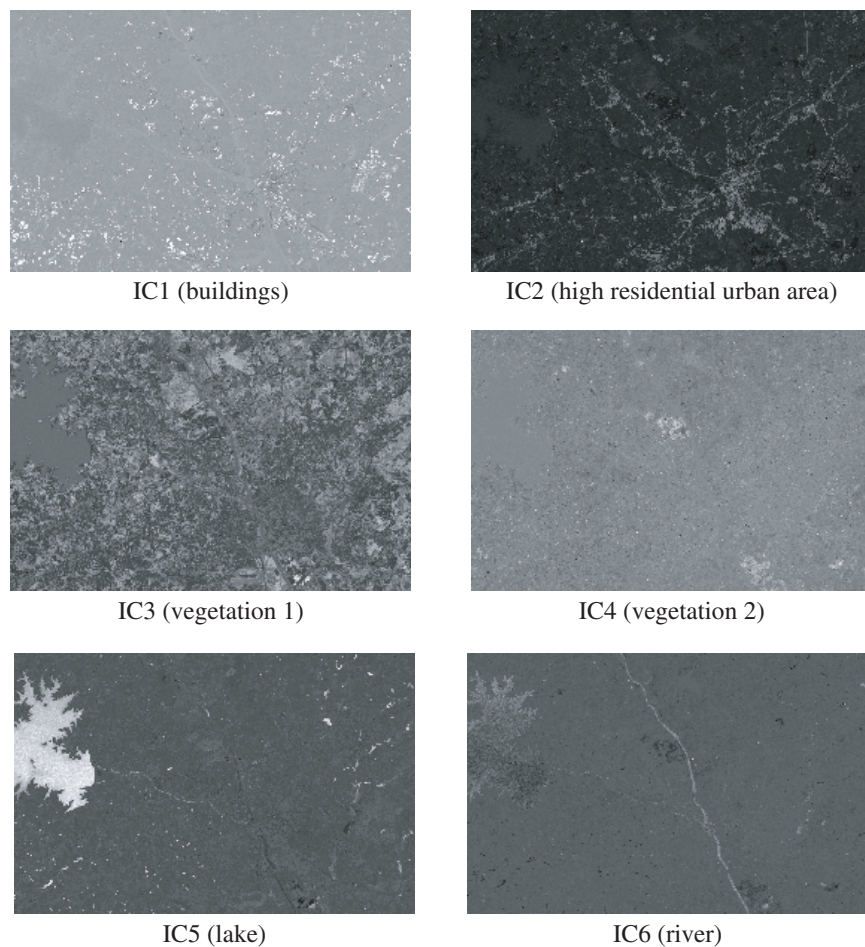


Fig. 3. Classification using three original bands plus three nonlinearly generated bands in the CIR image experiment.

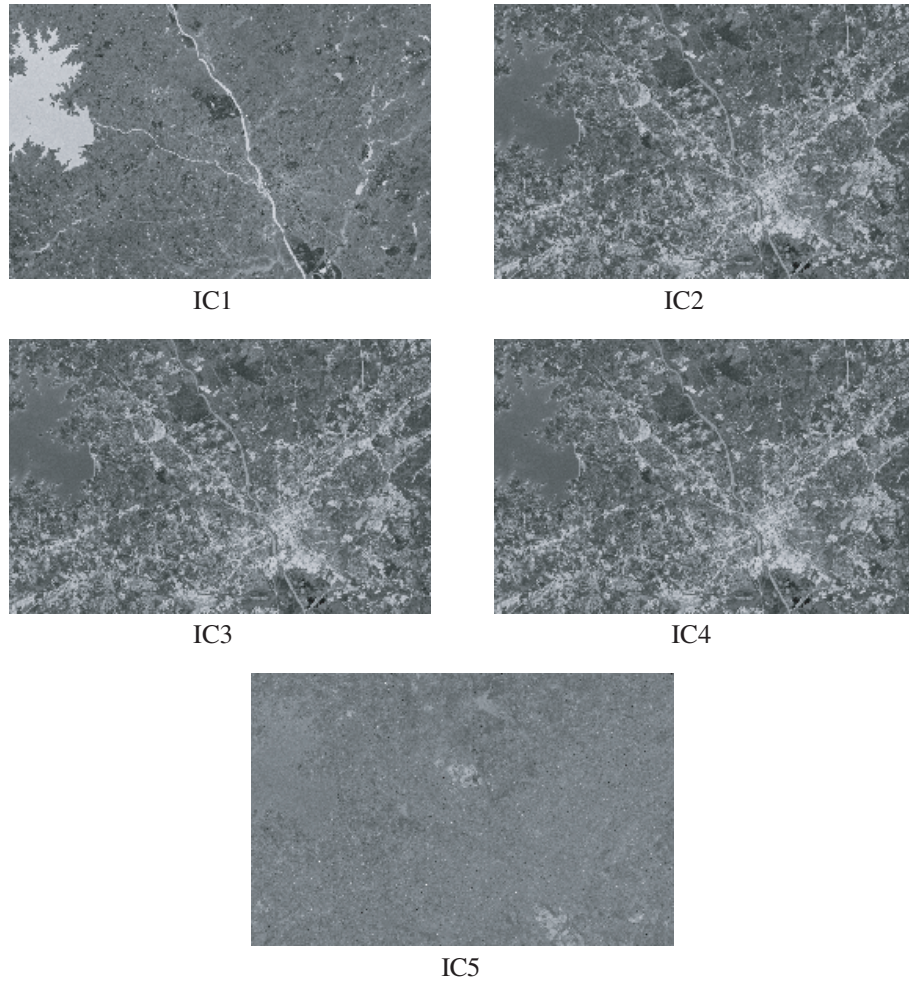


Fig. 4. Classification using data generated by cubic spline band interpolation in the CIR image experiment.

together. This demonstrates that linear methods such as band interpolation cannot provide new data information, so no improvement in classification can be brought about although data dimensionality is increased.

Example 2: SPOT image experiment

The image used in the second experiment was a 3-band SPOT multispectral image about a Washington DC suburban area, as shown in Fig. 5. According to prior information, there are five classes present in this image scene: Buildings, roads, vegetation, soil, and water. Figure 6 shows the classification using the original 3 bands, where IC1 contained all the classes, and IC2 and IC3 were for different types of buildings. Obviously, these five objects were not

well classified, because data dimensionality was too limited to separate them from each other. Then we generated three new bands by multiplying each pair of bands ($\mathbf{B}_1\mathbf{B}_2, \mathbf{B}_2\mathbf{B}_3, \mathbf{B}_1\mathbf{B}_3$). These three new bands together with original three bands enabled us to classify 6 classes, which were shown in Fig. 7: Roads and some buildings in IC1, vegetation in IC2, soil in IC3, different types of buildings in IC5 and IC6, and water in IC4. Buildings have diverse spectral signatures, mainly depending upon the materials covering the roofs. That is why they were classified into several classes. Some buildings were made up of the same material as the roads, so they were classified into the same class in IC1. Compared to Fig. 6, classification was improved by using nonlinearly generated bands. If we generate more artificial bands, we do not find

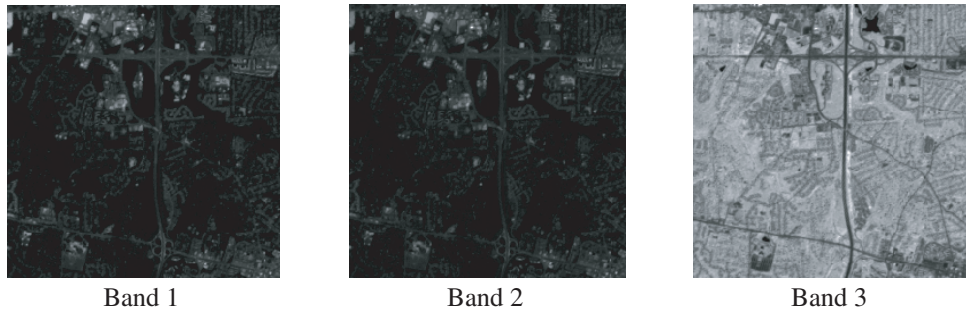


Fig. 5. An original 3-band SPOT image.

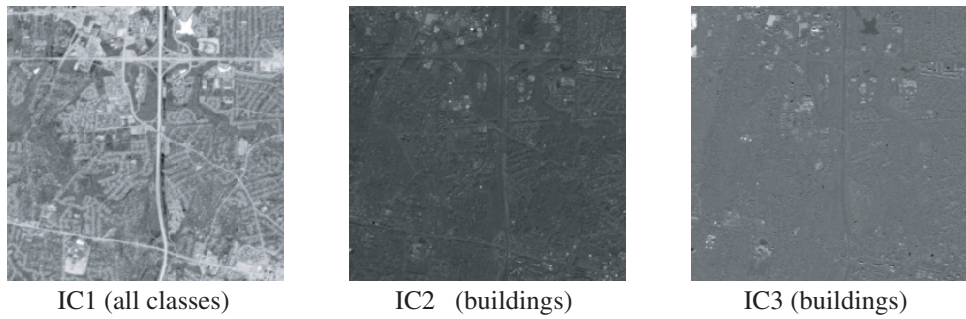


Fig. 6. Classification using original 3-band SPOT image.

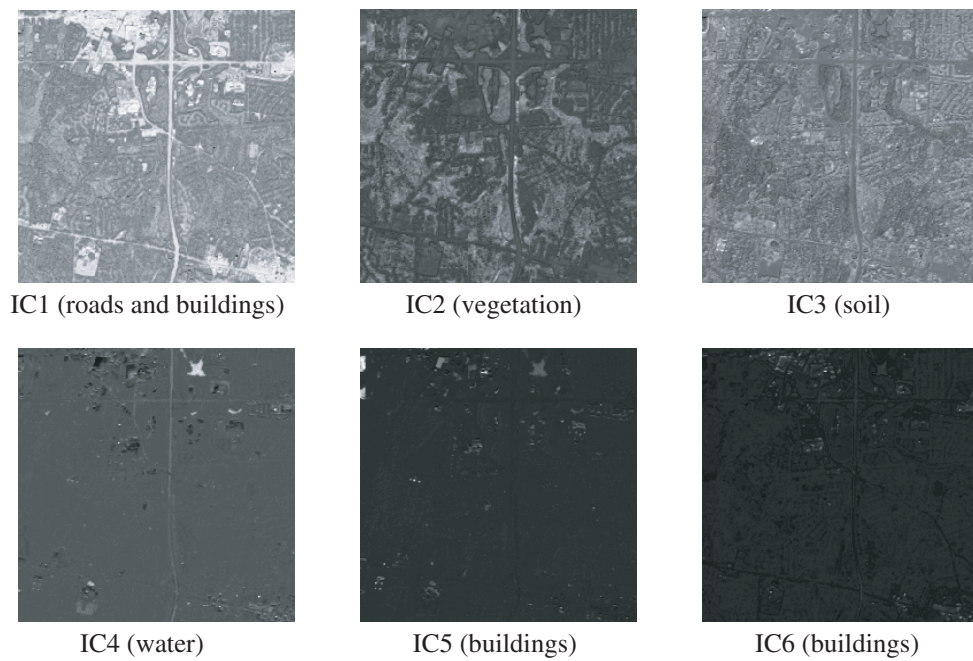


Fig. 7. Classification using three original bands plus three nonlinearly generated bands in the SPOT image experiment.

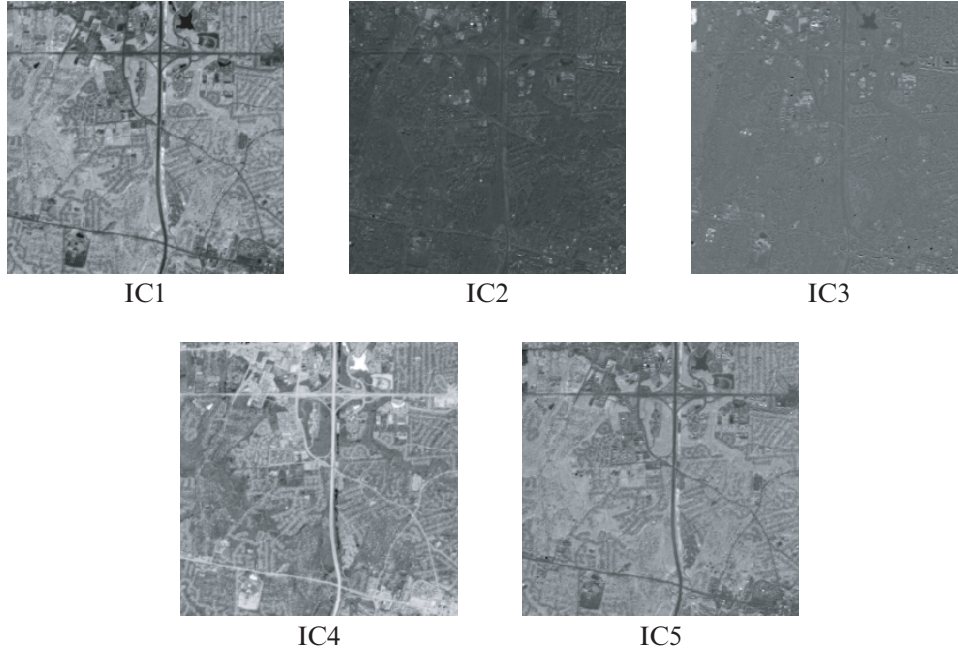


Fig. 8. Classification using data generated by cubic spline interpolation in the SPOT image experiment.

more meaningful classes and buildings and roads cannot be further separated. This also means the classification accuracy is upper-bounded by the information included in the original data. If there is no spectral difference (or difference is too subtle) between two classes, they will not be separated no matter how many nonlinear measurements are generated. So the nonlinear approach presented here can only “explore” information embedded in the original data, but not “create” brand-new information.

When using the band interpolation, the classification could not be improved as shown in Fig. 8. IC4, IC2, and IC3 in Fig. 8 are the same as IC1, IC2 and IC3 in Fig. 6 using three original bands, respectively. IC1 and IC5 in Fig. 8 present the similar information, which were very close to the negative image of IC4. This experiment further demonstrates that linearly generated bands cannot help to improve classification.

4.2. Quantitative study

In order to quantitatively assess the performance of the JADE algorithm in conjunction with the dimensionality expansion

approach and compare with other frequently used unsupervised classification algorithm in remote sensing such as ISODATA, an HYDICE image scene with precise pixel-level ground truth is used. The original image has 210 bands. After water absorption bands are removed, only 169 bands are remained. In order to simulate a multispectral case, only 17 bands are uniformly selected from these 169 bands and used in the experiment. As shown in Fig. 9(a), there are 15 panels present in the image scene which are arranged in a 5×3 matrix. Each element in this matrix is denoted by p_{ij} with row indexed by i and column indexed by j . The three panels in the same row were made from the same material and are of size $3\text{ m} \times 3\text{ m}$, $2\text{ m} \times 2\text{ m}$,

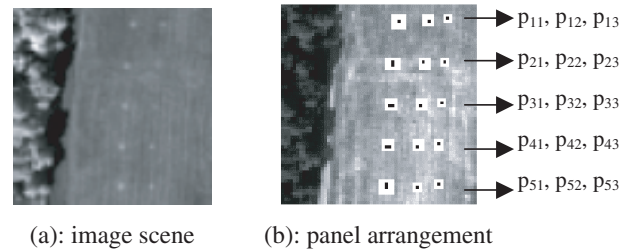


Fig. 9. The HYDICE image scene used in the experiment.

and $1\text{ m} \times 1\text{ m}$, respectively, and they are considered as a single class. The ground truth map was plotted in Fig. 9(b) with the precise spatial locations of panel pixels, where the black pixels are referred to as panel center pixels and white pixels are considered as panel pixels mixed with background pixels. The panels in each row are in the same class, and the signatures of these five panel classes are very similar. The grassland and trees to the left of the panels are the two background classes.

In this case, 17 bands were enough for classifying the 7 classes and the JADE algorithm was directly applied. The five independent components containing panels were shown in Fig. 10(a), where the five panel classes could be separated although some background pixels were misclassified, particularly when classifying P2 and P3. The ISODATA results were shown in Fig. 10(b), where only two classes contained panels. We can see that the five panel classes could not be well separated because their spectral signatures were very close and ISODATA would only cluster them together. Here, the ISODATA algorithm was initiated by selecting the pixel with the largest norm as the first class centroid, and the pixel with the greatest distance from this pixel as the second class centroid. The algorithm was stopped when the same number of classes as in JADE algorithm was generated. Actually, if the ISODATA continues running, the background will be split into more classes, but panels cannot be further separated. We

also randomly picked up a pixel as the initial class. Then the final results were slightly different with panels still being misclassified into two classes. It should be noted that the ISODATA images are black and white instead of grayscale through the use of ICA technique, because it performs 0–1 membership assignment (i.e. hard classification).

In order to further improve the performance, 16 bands were added which were nonlinearly generated by multiplying all the pairs of adjacent bands. These 16 bands with original 17 bands were used for re-classification. As shown in Fig. 11, both ISODATA and JADE would provide better results in terms of better background suppression. But the ISODATA still could not separate the five panel classes from each other. Table 6 lists the numbers of correctly detected pure panel pixels (N_D) and false alarm alarmed pixel (N_F) after comparing with the pixel-level ground truth. Here, the grayscale classification maps generated by JADE were converted into black and white images by setting the threshold to be the middle point of the gray levels in each classification map. We can see that JADE significantly outperforms ISODATA since its related N_D is larger and N_F is much smaller. When 33 bands were used, JADE could correctly classify 14 out of 19 pure panel pixels without false alarm. The missed five panel pixels are in the rightmost column and their size is $1\text{ m} \times 1\text{ m}$. The spatial resolution of this image scene is 1.5 m , so these panel pixels are

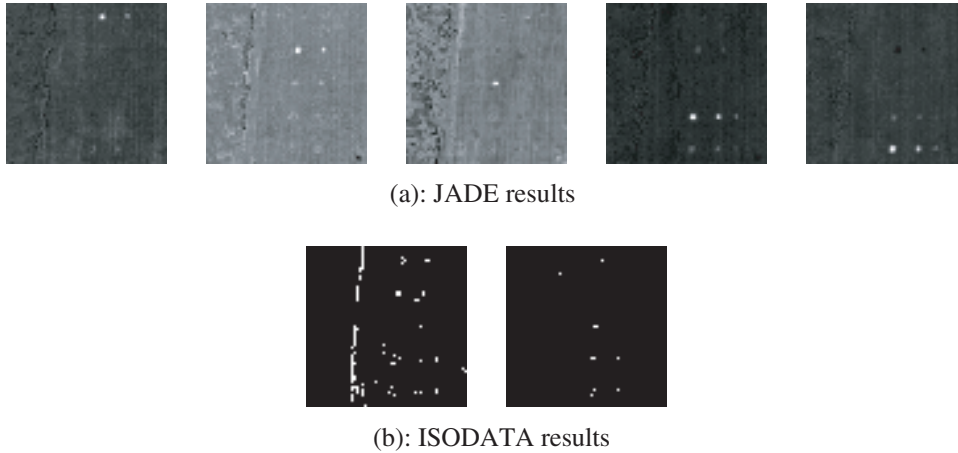


Fig. 10. Panel classification results using original 17 HYDICE bands.

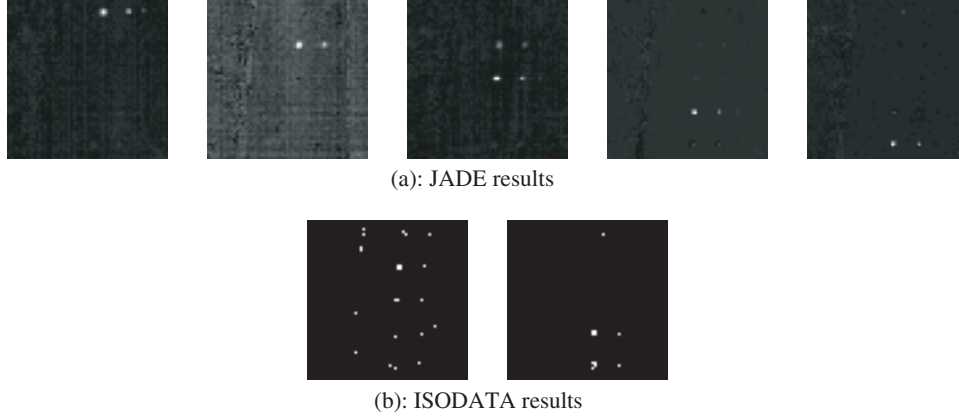


Fig. 11. Panel classification results using 33 HYDICE bands (original 17 bands plus generated 16 bands).

Table 6. Classification performance comparison.

The Number of Pure Pixels		JADE (33 bands)		JADE (17 bands)		ISODATA (33 bands)		ISODATA (17 bands)	
		N_D	N_F	N_D	N_F	N_D	N_F	N_D	N_F
P1	3	2	0	1	0	1	47	1	91
P2	4	3	0	3	47	3	40	3	90
P3	4	3	0	3	28	3	46	2	10
P4	4	3	0	3	0	3	39	2	13
P5	4	3	0	3	0	2	38	3	13
Total	19	14	0	13	75	12	210	11	217

embedded at the subpixel level and prone to be missed. The major reason to why JADE is more powerful than ISODATA is that JADE as well as other ICA algorithms takes advantage of higher order statistics in the data (the fourth order cumulants in JADE) while the clustering-based ISODATA uses only the second order statistics.

This experiment demonstrates that ICA is a more powerful classification approach than other commonly used unsupervised classification technique such as ISODATA. It also shows that even when the data dimensionality is larger than the number of classes and an ICA algorithm can be directly applied, by using nonlinearly generated bands the classification accuracy can be improved.

5. Conclusions and Discussions

Independent component analysis (ICA) can provide unsupervised classification for remotely

sensed images, which is a very useful approach when no prior information is available about an image scene and its classes. But when applied to multispectral images or CIR photograph, its performance is limited by the data dimensionality, i.e. the number of classes (e.g. objects or materials) to be classified cannot be greater than the number of spectral bands. In order to relax this limitation, we present a nonlinear band generation method to produce additional spectral measurements. We find out that a simple multiplication operation can explore the spectral discrepancy between original spectral measurements and improve the classification performance. We also demonstrate that ICA in conjunction with such a nonlinear band generation method outperforms ISODATA, a frequently used unsupervised classification algorithm in remote sensing, because it uses higher order statistics in the data that is more sensitive to the features of objects/classes.

Several comments are noteworthy.

- (1) Although the JADE algorithm is used in our experiments, the data dimensionality expansion approach is applicable to any ICA algorithm that has the data dimensionality limitation problem.
- (2) Here we deal with unsupervised classification problem, so we do not know how many classes ought to be classified in a real application. But when enough spectral bands are generated, we will find that some classification maps may only contain noise and some just repeat similar classification result as others. Some techniques are available for dimensionality estimation [Chang & Du, 2004], but it is out of the scope of this paper.
- (3) The computational complexity of the ICA approach generally is increased on the order of N^4 , which is its major drawback. Hence in real applications we do not prefer adding too many artificial bands in order to achieve the balance between classification accuracy and computational complexity.
- (4) The number of classification maps to be generated is equal to the final data dimensionality. Classification maps are in a random order, so their interpretation has to be made carefully. The ICA model is scaling and sign indeterminate since any scalar multiplier β (sign is included) in one of the sources s_i can always be canceled by dividing the corresponding column \mathbf{a}_i of \mathbf{A} by $1/\beta$. By displaying the classification maps according to their relative gray levels (i.e. the pixel at the minimum value is in black, the pixel at the maximum value is in white, and others in a shade between black and white), the (positive) scalar factor $|\beta|$ can be easily offset. As for the sign factor, displaying the negative image of a classification map and then examining both positive and negative images can help to select the one that is more meaningful.
- (5) The success of the data dimensionality expansion approach depends upon the information contained in the original data. For

instance, such a technique may not be able to help a RGB photograph since the information contained in the three original color bands is too limited. But it can usually help a CIR image because the near-infrared band contains important information that is different from color bands. In other words, this technique can only explore the information embedded in the original data, but not create completely new information, which is reasonable.

As future work, we will investigate if multiplication using multiple bands can provide further improvement and conduct more tests on other data sets.

Acknowledgments

The authors wish to acknowledge the support from Office of Naval Research (ONR) for this study. The authors also thank Professor Chein-I Chang at University of Maryland for providing the SPOT and HYDICE data used in the experiments.

References

- Adams, J. B. and Smith, M. O. [1986] "Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander 1 suite," *Journal of Geophysical Research* **91**(B8), 8098–8112.
- Adams, J. B., Smith, M. O. and Gillespie, A. R. [1993] "Image spectroscopy: Interpretation based on spectral mixture analysis," In: C. M. Pieters and P. A. Englert (eds.) *Remote Geochemical Analysis: Elemental and Mineralogical Composition*, Cambridge University Press, pp. 145–166.
- Bell, A. J. and Sejnowski, T. J. [1996] "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation* **7**(6), 1129–1159.
- Cardoso, J. F. and Souloumiac, A. [1993] "Blind beamforming for non gaussian signals," *IEEE Proceedings-F* **140**(6), 362–370.
- Cardoso, J. F. and Laheld, B. [1996] "Equivalent adaptive source separation," *IEEE Transactions on Signal Processing* **44**(12), 3017–3030.

- Chang, C.-I., Chiang, S.-S., Smith, J. A. and Ginsberg, I. W. [2002] "Linear spectral random mixture analysis for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing* **40**(2), 375–392.
- Chang, C.-I. and Du, Q. [2004] "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing* **42**(3), 608–619.
- Chen, C. H. and Zhang, X. [1999] "Independent component analysis for remote sensing study," *Proceedings of SPIE* **387**, 150–158.
- Cihocki, A., Unbehauen, R. and Rummert, E. [1994] "Robust learning algorithm for blind separation of signals," *Electronic Letters* **28**(21), 1986–1987.
- Cichocki, A. and Unbehauen, R. [1996] "Robust neural networks with online learning for blind identification and blind separation of sources," *IEEE Transactions on Circuit and Systems* **43**(11), 894–906.
- Comon, P. [1994] "Independent component analysis — A new concept?" *Signal Processing* **36**, 287–314.
- Du, Q. and Nekovei, R. [2005] "Implementation of real-time constrained linear discriminant analysis to remote sensing image classification," *Pattern Recognition* (in print).
- Fiori, S. [2003] "Overview of independent component analysis technique with an application to synthetic aperture radar (SAR) imagery processing," *Neural Networks* **16**, 453–467.
- Hughes, G. F. [1968] "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory* IT-14, 55–63.
- Hyvarinen, A. and Oja, E. [1997] "A fast fixed-point algorithm for independent component analysis," *Neural Computation* **9**(7), 1483–1492.
- Hyvarinen, A., Karhunen, J. and Oja, E. [2001] *Independent Component Analysis*, John Wiley & Sons.
- Jutten, C. and Herault, J. [1991] "Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing* **24**(1), 1–10.
- Kruse, F. A., Lefkoff, A. B., Boardman, J. W., Heidebrecht, K. B., Shapiro, A. T., Barloon, P. J., and Goetz, A. F. H. [1993] "The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data," *Remote Sensing of Environment* **44**, 145–163.
- Landgrebe, D. [2002] "Hyperspectral image data analysis," *IEEE Signal Processing Magazine* **19**(1), 17–28.
- Lee, T. W., Girolami, M. and Sejnowski, T. J. [1999] "Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources," *Neural Computation* **11**(2), 417–441.
- Lotsch, A., Friedl, M. A. and Pinzon, J. [2003] "Spatio-temporal deconvolution of NDVI image sequences using independent component analysis," *IEEE Transactions on Geoscience and Remote Sensing* **41**(12), 2938–2942.
- McCullagh, P. [1995] *Tensor Methods in Statistics*, London, Chapman & Hall.
- Oja, E. [1997] "The nonlinear PCA learning rule in independent component analysis," *Neurocomputing* **17**(1), 25–45.
- Pham, D. T. [1997] "Blind separation of mixtures of independent sources through a quasimaximum likelihood approach," *IEEE Trans. Signal Processing* **45**, 1712–1725.
- Ren, H. and Chang, C.-I. [2000] "A generalized orthogonal subspace projection approach to unsupervised multispectral image classification," *IEEE Transactions on Geosciences and Remote Sensing* **38**(6), 2515–2528.
- Settle, J. J. and Drake, N. A. [1993] "Linear mixing and estimation of ground cover proportions," *International Journal of Remote Sensing* **14**, 1159–1177.
- Singer, R. B. and McCord, T. B. [1979] "Mars: Large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance," *Proc. 10th Lunar Planet. Sci. Conf.*, pp. 1835–1848.
- Szu, H. [1999] "Independent component analysis (ICA): An enabling technology for intelligent information/image technology (IIT)," *IEEE Circuits Devices Magazine* **10**, 14–37.
- Szu, H. and Buss, J. [2000] "ICA neural net to refine remote sensing with multiple labels," *Proceedings of SPIE* **4056**, 32–49.
- Szu, H., Noel, S., Yim, S.-B., Willey, J. and Landa, J. [2003] "Multimedia authenticity protection with ICA watermarking and digital bacteria vaccination," *Neural Networks* **16**, 907–914.
- Szu, H. [2004] "Geometric topology for information acquisition by means of smart sensor web," *International Journal of Information Acquisition* **1**(1), 1–22.

- Tu, T. M. [2000] "Unsupervised signature extraction and separation in hyperspectral images: A noise-adjusted fast independent component analysis approach," *Optical Engineering* **39**(4), 897–906.
- Unser, M. [1999] "Splines: A perfect fit for signal and image processing," *IEEE Signal Processing Magazine* **16**(6), 22–38.
- Yoshida, T. and Omatu, S. [2000] "Pattern recognition with neural networks," *Proceedings of International Geoscience and Remote Sensing Symposium*, pp. 699–701.