Chapter 14
Projection Pursuit-Based Dimensionality Reduction for Hyperspectral Analysis

Haleh Safavi, Chein-I Chang, and Antonio J. Plaza

Abstract Dimensionality Reduction (DR) has found many applications in hyperspectral image processing. This book chapter investigates Projection Pursuit (PP)-based Dimensionality Reduction, (PP-DR) which includes both Principal Components Analysis (PCA) and Independent Component Analysis (ICA) as special cases. Three approaches are developed for PP-DR. One is to use a Projection Index (PI) to produce projection vectors to generate Projection Index Components (PICs). Since PP generally uses random initial conditions to produce PICs, when the same PP is performed in different times or by different users at the same time, the resulting PICs are generally different in terms of components and appearing orders. To resolve this issue, a second approach is called PI-based PRioritized PP (PI-PRPP) which uses a PI as a criterion to prioritize PICs. A third approach proposed as an alternative to PI-PRPP is called Initialization-Driven PP (ID-PIPP) which specifies an appropriate set of initial conditions that allows PP to produce the same PICs as well as in the same order regardless of how PP is run. As shown by experimental results, the three PP-DR techniques can perform not only DR but also separate various targets in different PICs so as to achieve unsupervised target detection.

H. Safavi (✉)
Remote Sensing Signal and Image Processing Laboratory, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore, MD, USA
e-mail: haleh1@umbc.edu

C.-I Chang
Remote Sensing Signal and Image Processing Laboratory, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore, MD, USA

Department of Electrical Engineering, National Chung Hsing University, Taichung, Taiwan
e-mail: cchang@umbc.edu

A.J. Plaza
Department of Technology of Computers and Communications, University of Extremadura, Escuela Politecnica de Caceres, Caceres, SPAIN
e-mail: aplaza@unex.es

B. Huang (ed.), *Satellite Data Compression*, DOI 10.1007/978-1-4614-1183-3_14,

1 Introduction

One of great challenging issues in hyperspectral analysis is how to deal with enormous data volumes acquired by hyperspectral imaging sensors' hundreds of contiguous spectral bands. A general approach is to perform Dimensionality Reduction (DR) as a pre-processing step to represent the original data in a manageable low dimensional data space prior to data processing. The Principal Components Analysis (PCA) is probably the most commonly used DR technique that reduces data dimensionality by representing the original data via a small set of Principal Components (PCs) in accordance with data variances specified by eigenvalues of the data sample covariance matrix. However, the PCA can only capture information characterized by second-order statistics as demonstrated in [1] where small targets may not be preserved in PCs. In order to retain the information preserved by statistically independent statistics, Independent Component Analysis (ICA) [2] was suggested for DR in [1]. This evidence was further confirmed by [3] where High-Order Statistics (HOS) were able to detect subtle targets such as anomalies, small targets. Since the independent statistics can be measured by mutual information [2], it will be interesting to see if PCA (i.e., 2nd order statistics), HOS and ICA can be integrated into a general setting so that each of these three cases can be considered as a special circumstance of this framework. However, this is easier said than done because several issues need to be addressed. First of all, both PCA and ICA have different ways to generate their projection vectors. For the PCA it first calculate the characteristic polynomial to find eigenvalues from which their associated eigenvectors can be generated as projection vectors to produce Principal Components (PCs) ranked by sample data variances. On the other hand, unlike the PCA, the ICA does not have a similar characteristic polynomial that allows it to find solutions from which projection vectors can be generated. Instead, it must rely on numerical algorithms to find these projection vectors. In doing so it makes use of random initial vectors to generate projection vectors in a random order so that the projection vector-produced Independent Components (ICs) also appear randomly. As a result of using different sets of random initial conditions, the generated ICs not only appear in a random order, but also are different even if they appear in the same order. Consequently, the results produced by different runs using different sets of random initial conditions or by different users will be also different. This same issue is also encountered in the ISODATA (K-means) clustering method in [4]. Therefore, the first challenging issue is how to rank components such as ICs in an appropriate order like PCs ranked by data variances. Recently, this issue has been addressed for ICA in [1] and for HOS-based components in [3] where a concept of using priority scores to rank components was developed to prioritize components according to the significance of the information contained in each component measured by a specific criterion. The goal of this chapter is to extend the ICA-based DR in [1] and the work in [3] by unifying these approaches in context of a more general framework, Projection Pursuit (PP) [5]. In the mean time it also generalizes the prioritization criteria in [3, 4] to the Projection Index (PI)

used by PP where the components generated by the PP using a specific PI are referred to as Projection Index Components (PICs). For example, the PI used to rank PCA-generated components is reduced to data variance with PICs being PCs, while the PI used to generate components by the FastICA in [2] turns out to be the neg-entropy and PICs become ICs.

Three approaches are proposed in this chapter to implement the PP with the PI specified by a particular prioritization criterion. The first approach is commonly used in the literature which makes use of the PI to produce components, referred to as PIPP. With this interpretation when the PI is specified by data variance, the resulting PP becomes PCA. On the other hand, if the PI is specified by mutual information to measure statistical independence, the resulting PP turns out to be ICA. While the first approach is focused on component generation, the second and third approaches can be considered as component prioritization. More specifically, the second approach, referred to as PI-PRioritized PP (PI-PR PP) utilizes the PI as a criterion to prioritize PICs produced by PIPP. In other words, due to the use of random initial conditions the PIPP-generated PICs generally appear in a random order. Using a PI-PRPP allows users to rank and prioritize PICs regardless of what random initial condition is used. Despite that the PI-PRPP resolves the issue of random order in which PICs appear, it does not necessarily imply that PICs ranked by the same order are identical. To further remedy this problem, the third approach, referred to as Initialization-Driven PP (ID-PIPP) is proposed to specify an appropriate set of initial conditions for PIPP so that as long as the PIPP uses the same initial condition, it always produces identical PICs which are ranked by the same order. By means of the 2nd and 3rd approaches the PP-DR can be accomplished by retaining a small number of components whose priorities are ranked by PI-PRPP using a specific prioritization criterion or the first few components produced by ID-PIPP. In order to evaluate the three different versions of the PP, PIPP, PI-PRPP and ID-PIPP, real hyperspectral image experiments are conducted for performance analysis.

2 Projection Pursuit-Based Component Analysis for Dimensionality Reduction

Dimensionality reduction is an important preprocessing technique that represents multi-dimensional data in a lower dimensional data space without significant loss of desired data information. A common approach which has been widely used in many applications such as data compression is the PCA which represents data to be processed in a new data space whose dimensions are specified by eigenvectors in descending order of eigenvalues. Another example is the ICA which represents data to be processed in a new data space whose dimensions are specified by a set of statistically independent projection vectors. This section presents a Projection Index (PI)-based dimensionality reduction technique, referred to as PI-based Project Pursuit (PIPP) which uses a PI as a criterion to find directions of interestingness

of data to be processed and then represents the data in the data space specified by these new interesting directions. Within the context of PIPP the PCA and ICA can be considered as special cases of PIPP in the sense that PCA uses data variance as a PI to produce eigenvectors while the ICA uses mutual information as a PI to produce statistically independent projection vectors.

The term of "Projection Pursuit (PP)" was first coined by Friedman and Tukey [5] which was used as a technique for exploratory analysis of multivariate data. The idea is to project a high dimensional data set into a low dimensional data space while retaining the information of interest. It designs a PI to explore projections of interestingness. Assume that there are N data points $\{\mathbf{X}_n\}_{n-1}^{N}$ each with dimensionality K and $\mathbf{X} = [\mathbf{r}_1\mathbf{r}_2 \cdots \mathbf{r}_N]$ is a $K \times N$ data matrix and $\mathbf{a}$ is a K-dimensional column vector which serves as a desired projection. Then $\mathbf{a}^T\mathbf{X}$ represents an N-dimensional row vector that is the orthogonal projections of all sample data points mapped onto the direction $\mathbf{a}$. Now if we let $H(\cdot)$ is a function measuring the degree of the interestingness of the projection $\mathbf{a}^T\mathbf{X}$ for a fixed data matrix $\mathbf{X}$, a Projection Index (PI) is a real-valued function of $\mathbf{a}$, $I(\mathbf{a}) : R^K \rightarrow R$ defined by

$$I(\mathbf{a}) = H(\mathbf{a}^T\mathbf{X}) \tag{14.1}$$

The PI can be easily extended to multiple directions, $\{\mathbf{a}_j\}_{j-1}^{J}$. In this case, $\mathbf{A} = [\mathbf{a}_1\mathbf{a}_2 \cdots \mathbf{a}_J]$ is a $K \times J$ projection direction matrix and the corresponding projection index is also a real-valued function, $I(\mathbf{A}) : R^{K\times J} \rightarrow R$ is given by

$$I(\mathbf{A}) = H(\mathbf{A}^T\mathbf{X}) \tag{14.2}$$

The choice of the $H(\cdot)$ in (14.1) and (14.2) is application-dependent. Its purpose is to reveal interesting structures within data sets such as clustering. However, finding an optimal projection matrix $\mathbf{A}$ in (14.2) is not a simple matter [6]. In this chapter, we focus on PIs which are specified by statistics of high orders such as skewness, kurtosis, etc. [7].

Assume that the ith projection index-projected component can be described by a random variable ζ_i with values taken by the gray level value of the nth pixel denoted by z_n^i. In what follows, we present a general form for the kth-order orders of statistics: kth moment by solving the following eigen-problem [7, 8]

$$\left(E\left[\mathbf{r}_i\left(\mathbf{r}_i^T\mathbf{w}\right)^{k-2}\mathbf{r}_i^T\right] - \lambda'\mathbf{I}\right)\mathbf{w} = 0 \tag{14.3}$$

It should be noted that when $k = 2, 3, 4$ in (14.3) is then reduced to variance, skewness and kurtosis respectively.

An algorithm for finding a sequence of projection vectors to solve (14.3) can be described as follows [7, 8]

Projection-Index Projection Pursuit (PIPP)

1. Initially, assume that $\mathbf{X} = [\mathbf{r}_1\mathbf{r}_2 \cdots \mathbf{r}_N]$ is data matrix and a PI is specified.

2. Find the first projection vector $\mathbf{w}_1^*$ by maximizing the PI.
3. Using the found $\mathbf{w}_1^*$, generate the first projection image $\mathbf{Z}^1 = \left(\mathbf{w}_1^*\right)^T \mathbf{X} = \left\{\mathbf{z}_i^1 | \mathbf{z}_i^1 = \left(\mathbf{w}_1^*\right)^T \mathbf{r}_i\right\}$ which can be used to detect the first endmember.
4. Apply the orthogonal subspace projector (OSP) specified by $P_{\mathbf{w}_1}^{\perp} = \mathbf{I} - \mathbf{w}_1 \left(\mathbf{w}_1^T \mathbf{w}_1\right)^{-1} \mathbf{w}_1^T$ to the data set $\mathbf{X}$ to produce the first OSP-projected data set denoted by $\mathbf{X}^1$, $\mathbf{X}^1 = P_{\mathbf{w}_1}^{\perp} \mathbf{X}$.
5. Use the data set $\mathbf{X}^1$ and find the second projection vector $\mathbf{w}_2^*$ by maximizing the same PI again.
6. Apply $P_{\mathbf{w}_2}^{\perp} = \mathbf{I} - \mathbf{w}_2 (\mathbf{w}_2^T \mathbf{w}_2)^{-1} \mathbf{w}_2^T$ to the data set $\mathbf{X}^1$ to produce the second OSP-projected data set denoted by $\mathbf{X}^2$, $\mathbf{X}^2 = P_{\mathbf{w}_2}^{\perp} \mathbf{X}^1$ which can be used to produce the third projection vector $\mathbf{w}_3^*$ by maximizing the same PI again. Or equivalently, we define a matrix projection matrix $\mathbf{W}^2 = [\mathbf{w}_1 \mathbf{w}_2]$ and apply $P_{\mathbf{W}^2}^{\perp} = \mathbf{I} - \mathbf{W}^2 \left(\left(\mathbf{W}^2\right)^T \mathbf{W}^2\right)^{-1} \left(\mathbf{W}^2\right)^T$ to the data set $\mathbf{X}$ to obtain $\mathbf{X}^2 = P_{\mathbf{W}^2}^{\perp} \mathbf{X}$.
7. Repeat the procedure of steps 5 and 6 over and over again to produce $\mathbf{w}_3^*, \cdots, \mathbf{w}_k^*$ until a stopping criterion is met. It should be noted that a stopping criterion can be either a predetermined number of projection vectors required to be generated or a predetermined threshold for the difference between two consecutive projection vectors.

3 Projection Index-Based Prioritized PP

According to the PIPP described in Sect. 2 a vector is randomly generated as an initial condition to produce projection vectors that are used to generate PICs. Accordingly, different initial condition may produce different projection vectors and so are their generated PICs. In other words, if the PIPP is performed in different times by different sets of random initial vectors or different users who will use different sets of random vectors to run the PIPP, the resulting final PICs will also be different. In order to correct this problem, this section presents a PI-based Prioritized PP (PI-PRPP) which uses a PI as a prioritization criterion to rank PIPP-generated PICs so that all PICs will be prioritized in accordance with the priorities measured by the PI. In this case, the PICs will be always ranked and prioritized by the PI in the same order regardless of what initial vectors are used to produce projection vectors. It should be noted that there is a major distinction between PIPP and PI-PRPP. While the PIPP uses a PI as criterion to produce a desired projection vector for each of PICs, the PI-PRPP uses a PI to prioritize PICs and this PI may not the same PI used by the PIPP to generate PICs. Therefore, the PI used in both PP and PI-PRPP is not necessarily the same and can be different. As a matter of fact, on many occasions, different PIs can be used in various applications. In what follows, we describe various criteria that can be used to define PI. These criteria are statistics-based measures that go beyond the 2nd order statistics.

Projection Index (PI)-Based Criteria

1. Sample mean of 3rd order statistics: skewness for ζ_j.

$$\mathrm{PI}_{\mathrm{skewness}}(\mathrm{PIC}_j) = \left[\kappa_j^3\right]^2 \tag{14.4}$$

where $\kappa_j^3 = E\left[\zeta_j^3\right] = (1/MN)\sum_{n=1}^{MN}\left(z_n^j\right)^3$ is the sample mean of the 3rd order of statistics in the PIC_j.

2. Sample mean of 4th order statistics: kurtosis for ζ_i.

$$\mathrm{PI}_{\mathrm{kurtosis}}(\mathrm{PIC}_j) = \left[\kappa_j^4\right]^2 \tag{14.5}$$

where $\kappa_j^4 = E\left[\zeta_j^4\right] = (1/MN)\sum_{n=1}^{MN}\left(z_n^j\right)^4$ is the sample mean of the 4th order of statistics in the PIC_j.

3. Sample mean of k-th order statistics: k-th moments for ζ_j.

$$\mathrm{PI}_{\mathrm{k-moment}}(\mathrm{PIC}_j) = \left[\kappa_j^k\right]^2 \tag{14.6}$$

where $\kappa_j^k = E\left[\zeta_j^k\right] = (1/MN)\sum_{n=1}^{MN}\left(z_n^j\right)^k$ is the sample mean of the k-th moment of statistics in the PIC_j.

4. Negentropy: combination of 3rd and 4th orders of statistics for ζ_j.

$$\mathrm{PI}_{\mathrm{negentropy}}(\mathrm{PIC}_j) = (1/12)\left[\kappa_j^3\right]^2 + (1/48)\left[\kappa_j^4 - 3\right]^2 \tag{14.7}$$

It should be note that (14.7) is taken from (5.35) in Hyvarinen and Oja [2, p. 115], which is used to measure the negentropy by high-order statistics.

5. Entropy

$$\mathrm{PI}_{\mathrm{entropy}}(\mathrm{PIC}_j) = -\sum_{j=1}^{MN} p_{ji}\log p_j \tag{14.8}$$

where $p_j = \left(p_{j1}, p_{j2}, \cdots, p_{jMN}\right)^T$ is the probability distribution derived from the image histogram of PIC_i.

6. Information Divergence (ID)

$$\mathrm{PI}_{\mathrm{ID}}(\mathrm{PIC}_j) = \sum_{j=1}^{MN} p_{ji}\log\left(p_{ji}/q_i\right) \tag{14.9}$$

where $p_j = \left(p_{j1}, p_{j2}, \cdots, p_{jMN}\right)^T$ is the probability distribution derived from the image histogram of PIC_i and $\mathbf{q}_j = \left(q_{j1}, q_{j2}, \cdots, q_{jMN}\right)^T$ is the Gaussian probability distribution with the mean and variance calculated from PIC_i.

4 Initialization-Driven PIPP

The PI-PRPP in Sect. 3 intended to remedy the issue that PICs can appear in a random order due to the use of randomly generated initial vectors. The PI-PRPP allows users to prioritize PICs according to information significance measured by a specific PI. Despite the fact that the PICs ranked by PI-PRPP may appear in the same order independent of different sets of random initial conditions they are not necessarily identical because the slight discrepancy in two corresponding PICs at the same appearing order may be caused by randomness introduced by their used initial conditions. Although such a variation may be minor compared to different appearing orders of PICs without prioritization, the inconsistency may still cause difficulty in data analysis. Therefore, this section further develops a new approach, called Initialization-Driven PP (ID-PIPP) which custom-designs an initialization algorithm to produce a specific set of initial conditions for PIPP so that the same initial condition is used all along whenever PIPP is implemented. Therefore, the ID-PIPP-generated PICs are always identical. When a particular initial algorithm, say X, is used to produce an initial set of vectors for the ID-PIPP to converge to projection vectors to produce PICs, the resulting PIPP is referred to as X-PIPP.

One such initialization algorithm described above that can be used for the ID-PIPP is the Automatic Target Generation Process (ATGP) previously developed in [10]. It makes use of an orthogonal subspace projector defined in [11] by

$$P_{\mathbf{U}}^{\perp} = \mathbf{I} - \mathbf{U}\mathbf{U}^{\#} \tag{14.10}$$

where $\mathbf{U}^{\#} = \left(\mathbf{U}^T\mathbf{U}\right)^{-1}\mathbf{U}^T$ is the pseudo inverse of the **U,** repeatedly to find target pixel vectors of interest from the data without prior knowledge regardless of what types of pixels are these targets. Details of implementing the ATGP are provided in the following steps.

Automatic Target Generation Process (ATGP)

1. Initial condition: Let p be the number of target pixels needed to be generated. Select an initial target pixel vector of interest denoted by $\mathbf{t}_0$. In order to initialize the ATGP without knowing $\mathbf{t}_0$, we select a target pixel vector with the maximum length as the initial target $\mathbf{t}_0$, namely, $\mathbf{t}_0 = \arg\{\max_{\mathbf{r}} \mathbf{r}^T\mathbf{r}\}$, which has the highest intensity, i.e., the brightest pixel vector in the image scene. Set $n = 1$ and$\mathbf{U}_0 = [\mathbf{t}_0]$. (It is worth noting that this selection may not be necessarily the best selection. However, according to our experiments it was found that the brightest pixel vector was always extracted later on, if it was not used as an initial target pixel vector in the initialization).

2. At nth iteration, apply $P_{\mathbf{t}_0}^{\perp}$ via (14.10) to all image pixels $\mathbf{r}$ in the image and find the nth target $\mathbf{t}_n$ generated at the n-th stage which has the maximum orthogonal projection as follows.

$$\mathbf{t}_n = \arg\left\{ \max_{\mathbf{r}} \left[\left(P_{[\mathbf{U}_{n-1}\mathbf{t}_n]}^{\perp} \mathbf{r} \right)^T \left(P_{[\mathbf{U}_{n-1}\mathbf{t}_n]}^{\perp} \mathbf{r} \right) \right] \right\} \tag{14.11}$$

where $\mathbf{U}_{n-1} = [\mathbf{t}_1 \mathbf{t}_2 \cdots \mathbf{t}_{n-1}]$is the target matrix generated at the $(n-1)$st stage.

3. Stopping rule: If $n<p-1$, let $\mathbf{U}_n = [\mathbf{U}_{n-1}\mathbf{t}_n] = [\mathbf{t}_1 \mathbf{t}_2 \cdots \mathbf{t}_n]$ be the n-th target matrix, go to step 2. Otherwise, continue.
4. At this stage, the ATGP is terminated. At this point, the target matrix is $\mathbf{U}_{p-1}$, which contains $p-1$ target pixel vectors as its column vectors, which do not include the initial target pixel vector $\mathbf{t}_0$.

As a result of the ATGP, the final set of target pixel vectors produced by the ATGP at step 4 is the final target set which comprises p target pixel vectors, $\{\mathbf{t}_0, \mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_{p-1}\} = \{\mathbf{t}_0\} \cup \{\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_{p-1}\}$ which were found by repeatedly using (14.11). It should be noted that the stopping rule used in the above ATGP was set by a pre-determined number of targets that should be generated. Of course this stopping rule can be replaced by any other rule such as the one used in [10]. Finally, a PIPP using the ATGP as its initialization algorithm is called ATGP-PIPP.

5 Real Hyperspectral Image Experiments

The image scene to be studied for experiments is a real image scene collected by HYperspectral Digital Imagery Collection Experiments (HYDICE) sensor shown in Fig. 14.1a, which has a size of 64 × 64pixel vectors with 15 panels in the scene and the ground truth map in Fig. 14.1b. It was acquired by 210 spectral bands with a spectral coverage from 0.4 μm to 2.5 μm. Low signal/high noise bands: bands 1–3 and bands 202–210; and water vapor absorption bands: bands 101–112 and bands 137–153 were removed. So, a total of 169 bands were used in experiments. The spatial resolution is 1.56 m and spectral resolution is 10 nm.

Within the scene in Fig. 14.1a, there is a large grass field background, and a forest on the left edge. Each element in this matrix is a square panel and denoted by p_{ij} with rows indexed by i and columns indexed by $j = 1, 2, 3$. For each row $i = 1, 2, \cdots, 5$, there are three panels p_{i1}, p_{i2}, p_{i3}, painted by the same paint but with three different sizes. The sizes of the panels in the first, second and third columns are 3m × 3m and 2m × 2m and 1m × 1m respectively. Since the size of the panels in the third column is1m × 1m, they cannot be seen visually from Fig. 14.1a due to the fact that its size is less than the 1.56 m pixel resolution. For

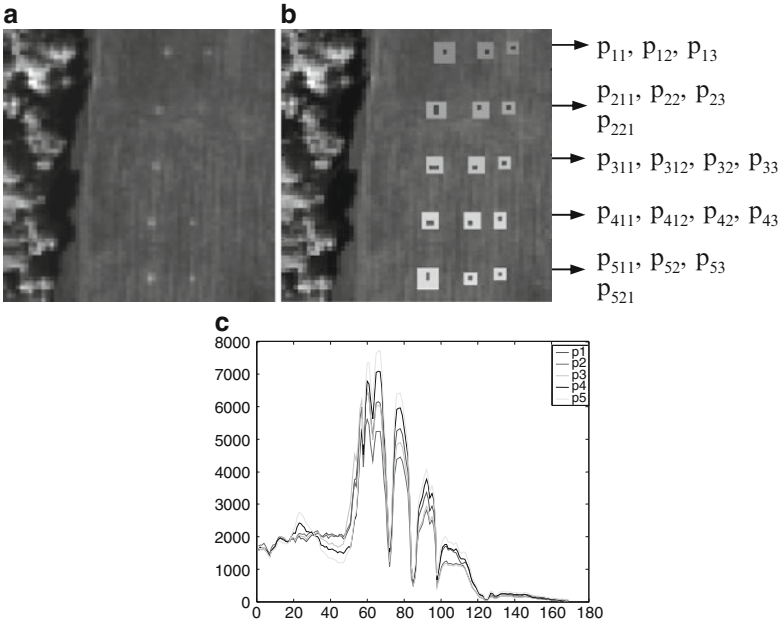

Fig. 14.1 (**a**) A HYDICE panel scene which contains 15 panels; (**b**) Ground truth map of spatial locations of the 15 panels; (**c**) Spectral signatures of $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$, $\mathbf{p}_4$ and $\mathbf{p}_5$

each column $j = 1, 2, 3$, the five panels, p_{1j}, p_{2j}, p_{3j}, p_{4j}, p_{5j} have the same size but with five different paints. However, it should be noted that the panels in rows 2 and 3 were made by the same material with two different paints. Similarly, it is also the case for panels in rows 4 and 5. Nevertheless, they were still considered as different panels but our experiments will demonstrate that detecting panels in row 5 (row 4) may also have effect on detection of panels in row 2 (row 3). The 1.56 m-spatial resolution of the image scene suggests that most of the 15 panels are one pixel in size except that $\mathbf{p}_{21}$, $\mathbf{p}_{31}$, $\mathbf{p}_{41}$, $\mathbf{p}_{51}$ which are two-pixel panels, denoted by p_{211}, p_{221}, p_{311}, p_{312}, p_{411}, p_{412}, p_{511}, p_{521}. Since the size of the panels in the third column is 1m $\times$ 1m, they cannot be seen visually from Fig. 14.1a due to the fact that its size is less than the 1.56 m pixel resolution. Figure 14.1b shows the precise spatial locations of these 15 panels where red pixels (R pixels) are the panel center pixels and the pixels in yellow (Y pixels) are panel pixels mixed with the background. Figure 14.1c plots the five panel spectral signatures $\mathbf{p}_i$ for $i = 1, 2, \cdots, 5$ obtained by averaging R pixels in the 3m $\times$ 3mand 2m $\times$ 2m panels in row i in Fig. 14.1b. It should be noted the R pixels in the 1m $\times$ 1m panels are not included because they are not pure pixels, mainly due to that fact that the spatial resolution of the R pixels in the 1m $\times$ 1m panels is 1 m smaller than the pixel resolution is

1.56 m. These panel signatures along with the R pixels in the3m × 3mand 2m × 2m panels were used as required prior target knowledge for the following comparative studies.

In order to perform dimensionality reduction we must know how many PICs needed to be retained, denoted by p after PIPP. Over the past years this knowledge has been obtained by preserving a certain level of energy percentage based on accumulative sum of eignevlaues. Unfortunately, it has been shown in [6, 12] that this was ineffective. Instead, a new concept, called Virtual Dimensionality (VD) was proposed to address this issue and has shown success and promise in [1, 3] where the VD estimated for the HYDICE scene in Fig. 14.1a was 9 with false alarm probability P_F greater than or equal to 10^{-4} . So, in the following experiments, the value of p was set to $p = 9$ where three versions of PIPP were evaluated for applications in dimensionality reduction and endmember extraction. Due to limited space including all experimental results are nearly impossible. In this case, only representatives are included in this chapter, which are PI = skewness (3rd order statistics), kurtosis (4th order statistics) and negentropy (infinite order statistics, i.e., statistical independence).

5.1 PIPP with Random Initial Conditions

In order to demonstrate inconsistent results from using two different sets of random initial vectors by the PIPP Figs. 14.2–14.4 show nine PICs resulting from performing PIPP with PI = skewness, kurtosis and negentropy respectively where it should be noted that the PIPP with PI = negentropy was carried out by the FastICA developed by Hyvarinen and Oja in [9] to produce PICs.

As we can see from these figures, due to the use of random initial conditions the appearing orders of interesting PICs generally are not the same in each run. In particular, some PICs which did not appear in one run actually appeared in another run. In addition, some of the PICs that contained little information showed up among the first a few PICs.

Now, these nine components obtained in Figs. 14.2–14.4 can then be used for endmember extraction performed by the well-known algorithm developed by Winter, referred to as N-FINDR [13] and the results of nine endmember extracted by the N-FINDR are shown in Figs. 14.5–14.7 where it is clear to see that endmembers were extracted in different orders by the PIPP with the same PI using two different sets of random initial vectors.

According to the results in Figs. 14.5–14.7, the best performance was given by the PIPP using PI = negentropy where all five panel signatures were extracted as endmembers.

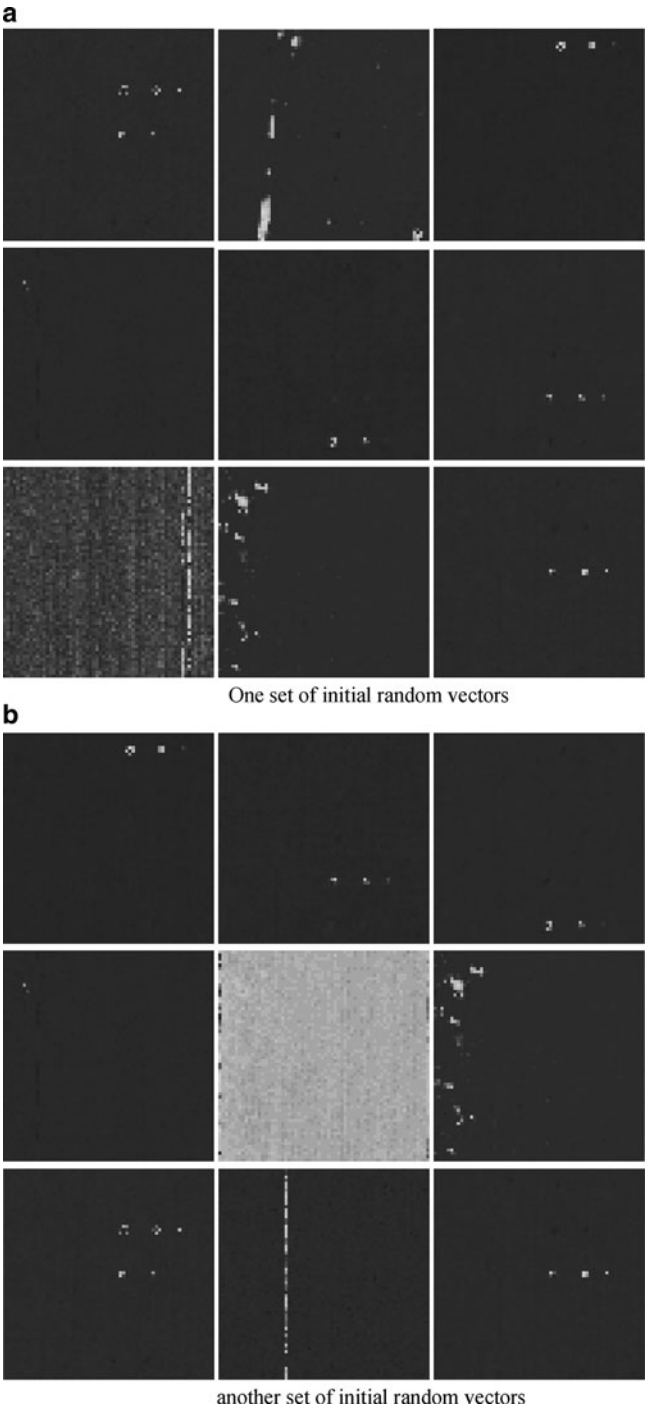

Fig. 14.2 First nine PICs extracted by PI-PP with PI = skewnes using random initial conditions. (**a**) One set of initial random vectors. (**b**) Another set of initial random vectors

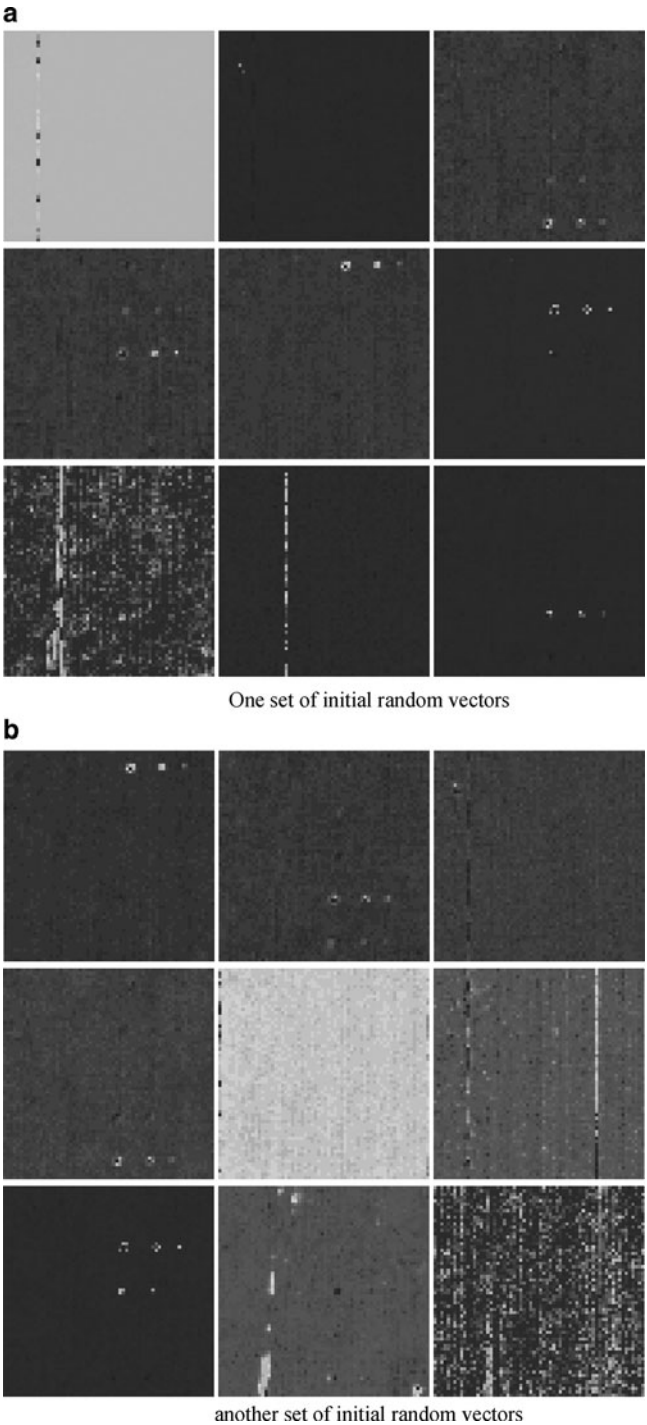

Fig. 14.3 First nine PICs extracted by PI-PP with PI = kurtosis using random initial conditions. (**a**) One set of initial random vectors. (**b**) Another set of initial random vectors

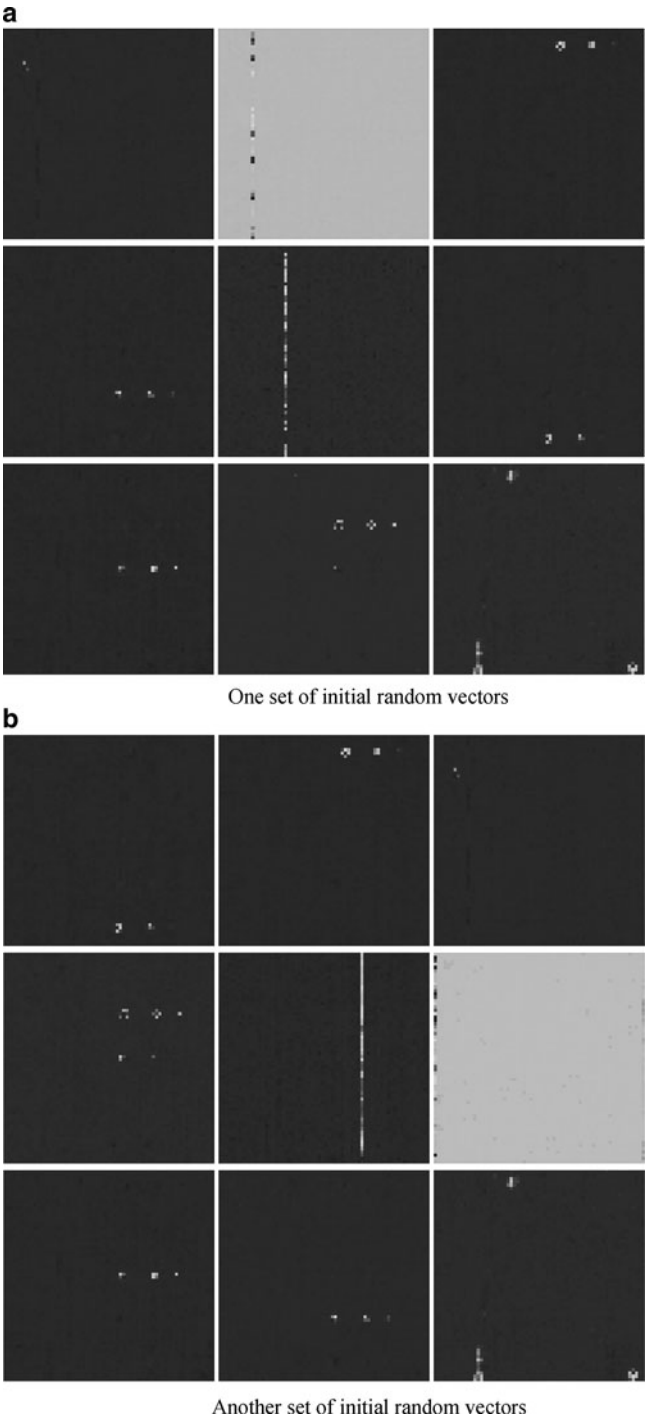

Fig. 14.4 First nine PICs extracted by PI-PP with PI = negentropy PI using random initial conditions. (**a**) One set of initial random vectors. (**b**) Another set of initial random vectors

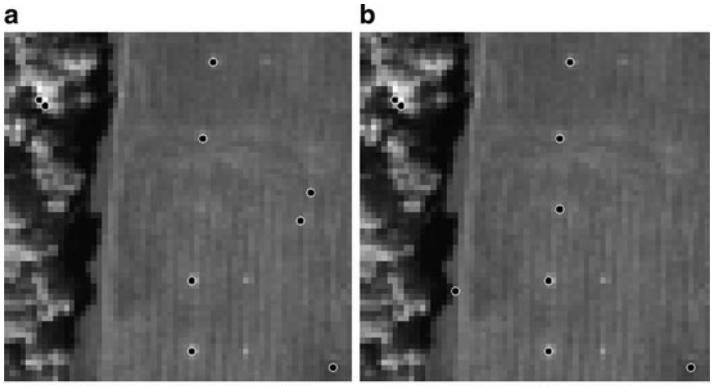

Fig. 14.5 Nine endmembers extracted by PIPP with PI = skewnes using random initial conditions. (**a**) One set of initial random vectors. (**b**) another set of initial random vectors

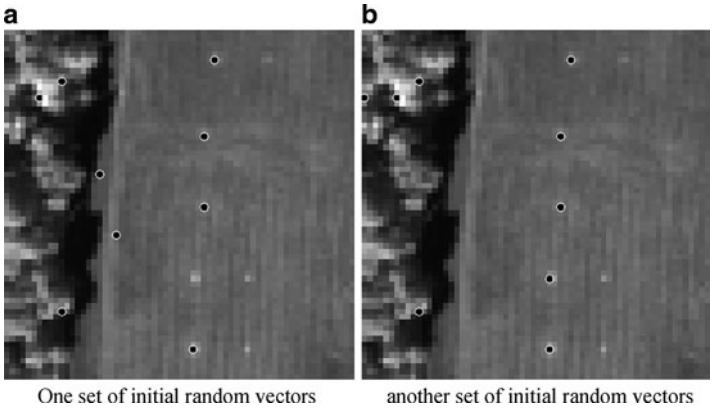

Fig. 14.6 Nine endmembers extracted by PIPP with PI = kurtosis using random initial conditions (**a**) One set of initial random vectors. (**b**) another set of initial random vectors

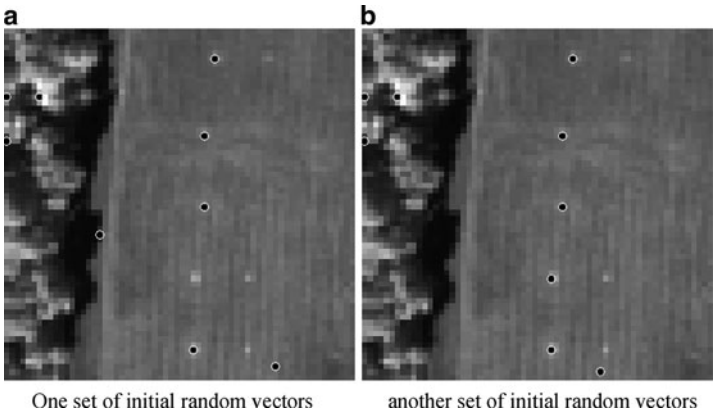

Fig. 14.7 Nine endmembers extracted by PIPP with PI = negentropy using random initial conditions. (**a**) One set of initial random vectors. (**b**) another set of initial random vectors

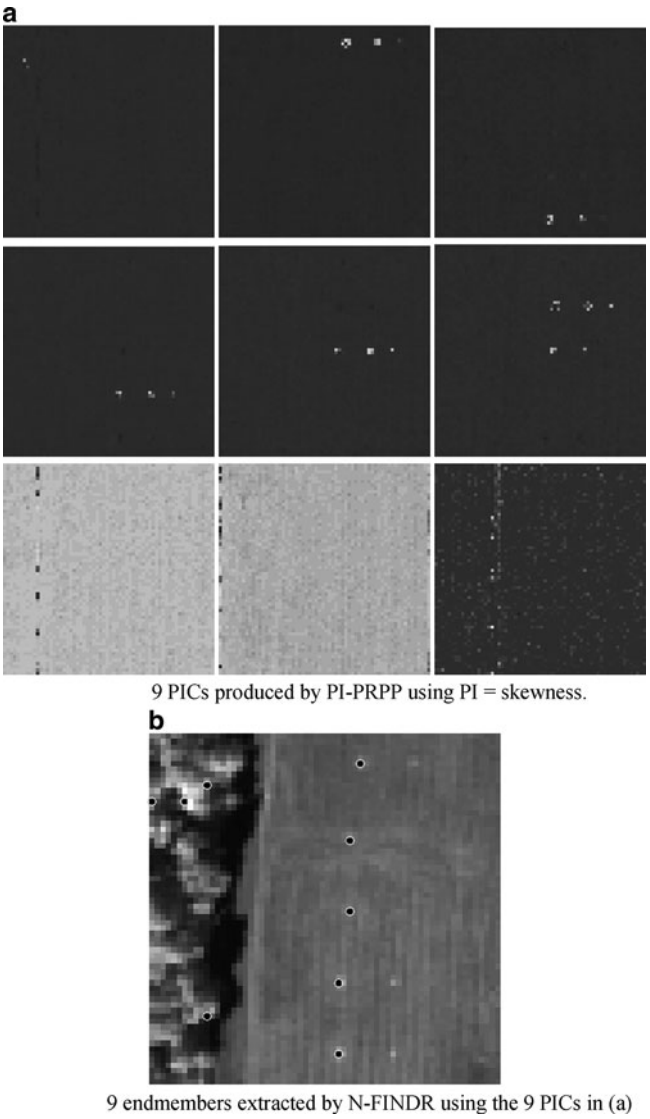

Fig. 14.8 PI-PRPP with PI = skewness. (**a**) 9 PICs produced by PI-PRPP using PI = skewness. (**b**) Nine endmembers extracted by N-FINDR using the nine PICs in (**a**)

5.2 PI-PRPP

As noted, the first nine PIPP-generated components in each of Figs. 14.2–14.4 were different not only in order but also in information contained in components. Accordingly, the endmember extraction results were also different in Figs. 14.5–14.7. The PI-PRPP was developed to remedy this problem. In order to make comparison with the results in Figs. 14.2–14.4, the PI-PRPP used the same three PIs, skewness, kurtosis and negentroy to produce PICs which were further prioritized by the same PI = negentropy. Figures 14.8a, 14.9a and 14.10a show the

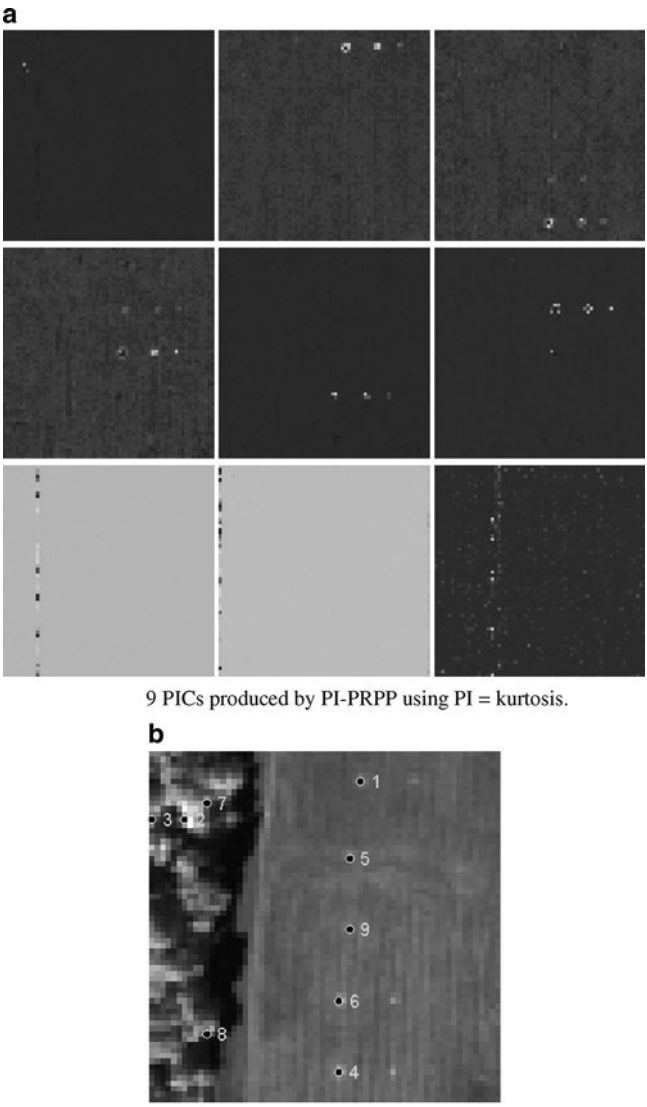

Fig. 14.9 PI-PRPP with PI = kurtosis. (**a**) Nine PICs produced by PI-PRPP using PI = kurtosis. (**b**) Nine endmembers extracted by N-FINDR using the nine PICs in (**a**)

nine negentropy-prioritized PICs produced by PI = skewness, kurtosis and negentropy respectively. Figures 14.8b, 14.9b and 14.10b also show the endmember extraction results by applying the N-FINDR to the nine prioritized PICs as image cubes where five panel pixels corresponding to all the five endmembers were successfully extracted compared to only four endmembers

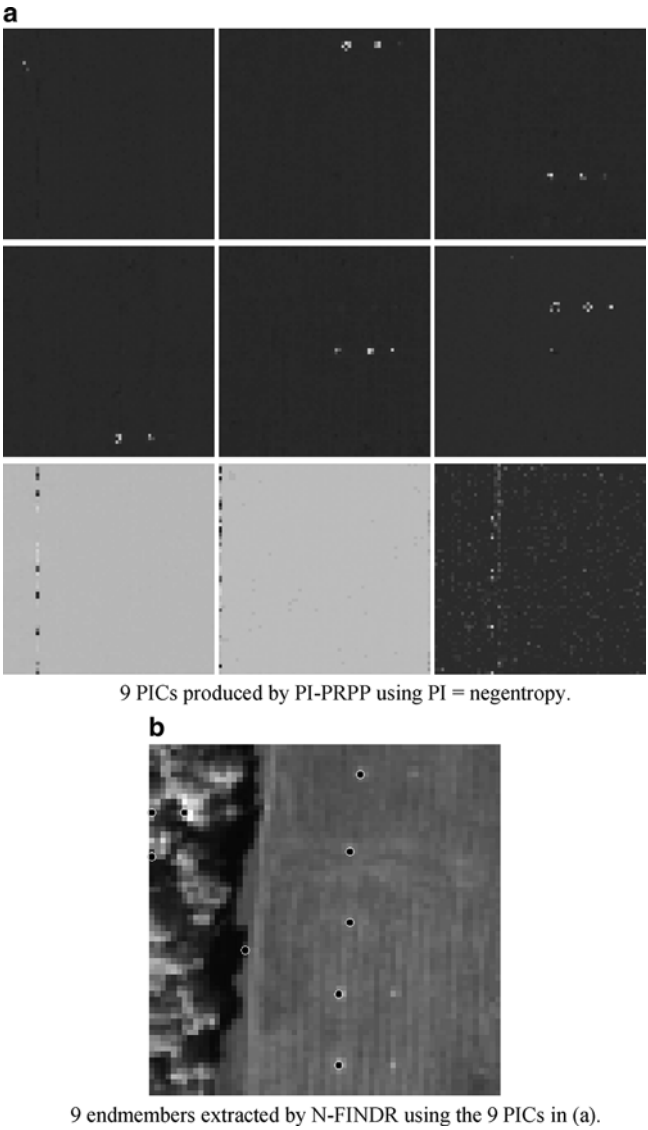

Fig. 14.10 PI-PRPP with PI = negentropy. (**a**) Nine PICs produced by PI-PRPP using PI = negentropy. (**b**) Nine endmembers extracted by N-FINDR using the nine PICs in (**a**)

extracted by the PIPP in one run shown in Figs. 14.5a, 14.6a and 14.7a. These experiments clearly demonstrated advantages of using the PR-PIPP over the PIPP.

It should be noted that similar results were also obtained for nine PICs prioritized by PI = skewness and kurtosis. Thus, their results are not included here.

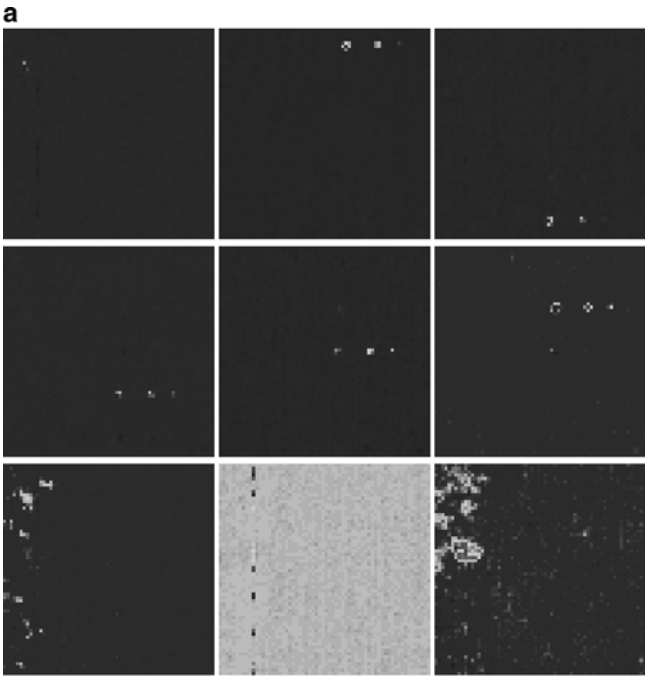

Fig. 14.11 ATGP-PIPP with PI = skewness. (**a**) Nine PICs produced by ID-PIPP using PI = skewness. (**b**) Nine endmembers extracted by N-FINDR using the nine PICs in (**a**)

5.3 ID-PIPP

In the experiments of the PIPP and PI-PRPP, the initial conditions were generated by a random generator. This section investigates the ID-PIPP and compares its performance against the PIPP and PI-PRPP. The ATGP was the initialization algorithm used to produce a set of initial vectors to initialize the PIPP. Figures 14.11–14.13 show the results of 9 PICs produced by the ATGP-PIPP and

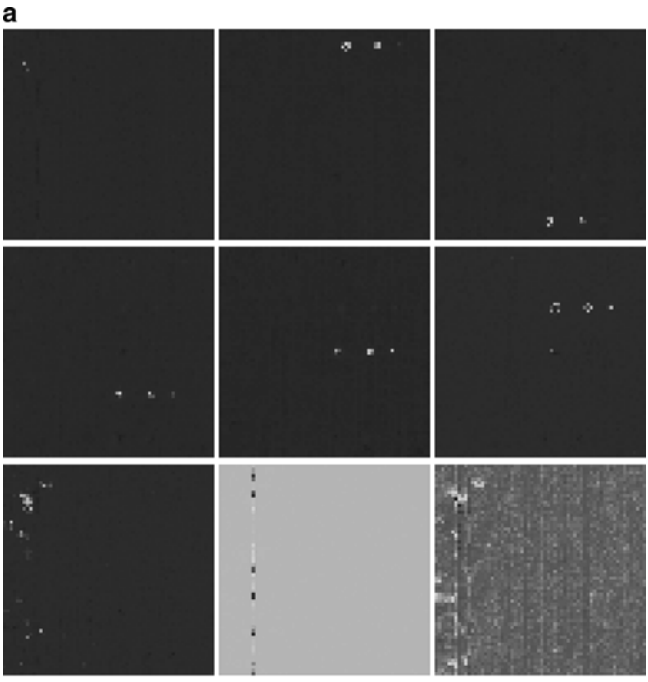

Fig. 14.12 ATGP-PIPP with PI = kurtosis. (**a**) Nine PICs produced by ATGP-PIPP using PI = kurtosis. (**b**) Nine endmembers extracted by N-FINDR using the Nine PICs in (**a**)

endmember extraction by the N-FINDR using the nine ATGP-PIPP-generated PICs using PI = skewness, kurtosis and negentropy respectively where like the PI-PRPP there were five panel pixels corresponding to five endmembers which were successfully extracted.

Finally, in order to complete our comparative study and analysis, we also included experiments performed by the PCA which is the 2nd order statistics.

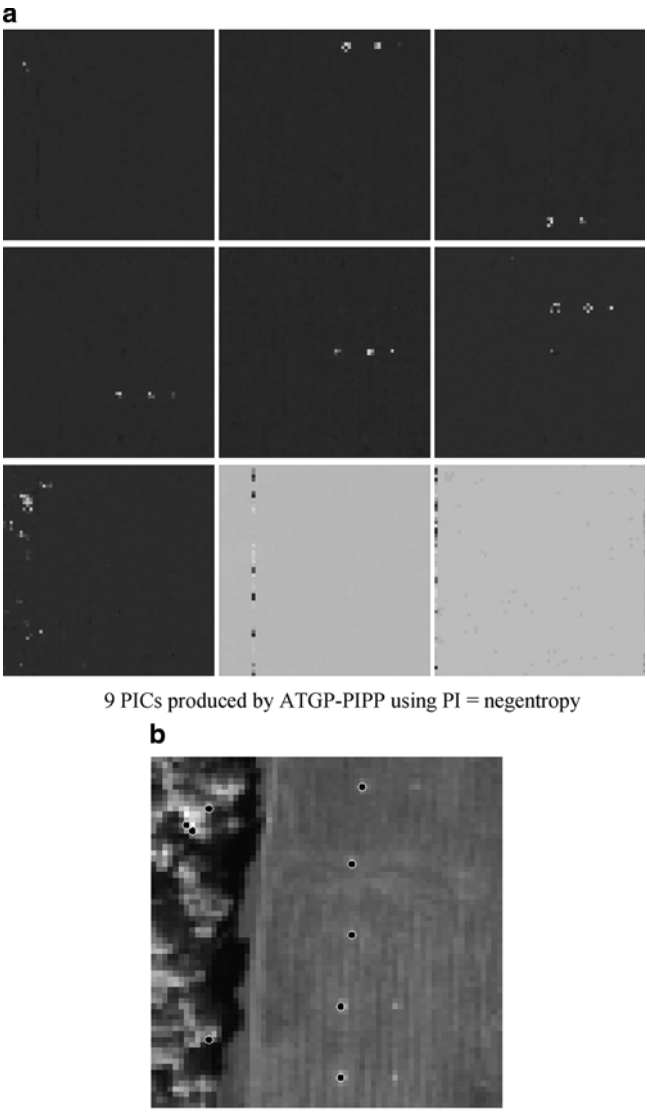

Fig. 14.13 ATGP-PIPP with PI = negentropy. (**a**) Nine PICs produced by ATGP-PIPP using PI = negentropy. (**b**) Nine endmembers extracted by N-FINDR using the nine PICs in (**a**)

Figure 14.14a shows the nine PCs produced by the PCA. These nine PCs were then formed as an image cube to be processed by the N-FINDR to extract nine endmembers shown in Fig. 14.14b where only three panel pixels in rows 1, 3 and 5 corresponding to three endmembers were extracted. These experiments provided simple evidence that the PCA was ineffective in preserving endmember information in its PCs.

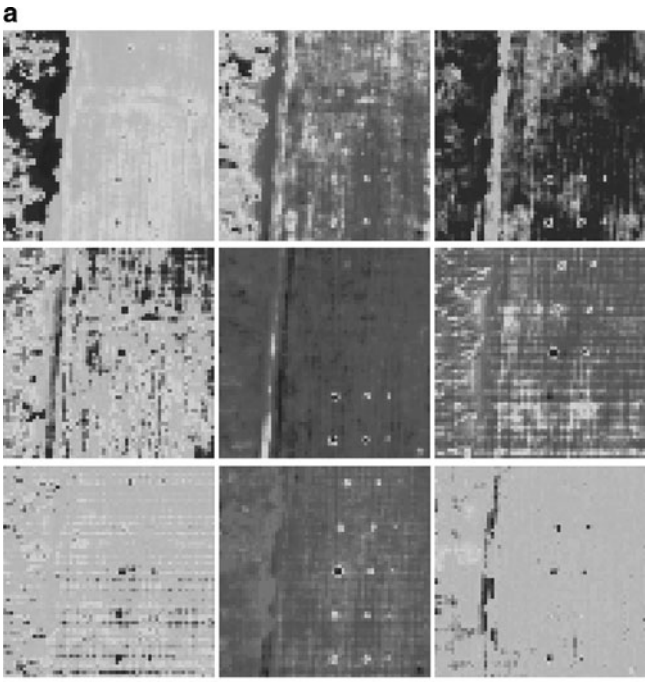

Fig. 14.14 PCA results. (**a**) Nine PCs produced by PCA. (**b**) Nine endmembers extracted by N-FINDR using the nine PCs in (**a**)

As a final concluding remark, the same experiments can also be conducted for other data sets such as various synthetic image-based scenarios [14] and another HYDICE scene in [15]. Similar results and conclusions can be also drawn from these experiments. Therefore, in order to avoid duplicates, these results are not presented here.

6 Conclusions

Dimensionality Reduction (DR) is a general pre-processing technique to reduce the vast amount of data volumes provided by multi-dimensional data while retaining most significant data information in a lower dimensional space. One type of multi-dimensional data involving enormous data volumes is hyperspectral imagery. This chapter presents a new approach to DR for hyperspectral data exploitation, called Projection- Index-based Projection Pursuit-based (PIPP) DR technique which includes the commonly used PCA and a recently developed ICA as its special cases. Three version of PIPP are developed for DR, Projection Index-Based PP (PIPP), PI-Prioritized PP (PI-PRPP) and Initialization-Driven PIPP (ID-PIPP). The PIPP uses a selected PI as a criterion to produce projection vectors that are used to specify components, referred to as Projection Index Components (PICs) for data representation. For example, when PI is used as sample data variance, the PIPP is reduced to PCA where the PICs are considered as PCs. On the other hand, when the PI is used as mutual information, the PIPP becomes ICA and PICs are actually ICs. Unfortunately, the PIPP still suffers from two major drawbacks which prevent it from practical implementation. This may be reasons that very little work done on the use of PIPP to perform DR. One of most serious problems with implementing the PIPP is the use of random initial conditions which result in different appearing orders of PICs when two different sets of random initial conditions are used. Under these circumstances PICs appear earlier do not necessarily imply that they are more significant than those PICs appear later as demonstrated in our experiments (see Figs. 14.2–14.4). Another problem is that when DR is performed, there is no appropriate guideline to be used to determine how many PICs should be selected. This issue is very crucial and closely related to the first problem addressed above. In order to cope with the second problem, a recently developed concept, called Virtual Dimensionality (VD) can be used for this purpose. However, this only solves half a problem. As noted, because the PIPP uses a random generator to generate initial conditions, a PIC generated earlier by the PIPP does not necessarily have more useful information that the one generated later by the PIPP. Once the number of PIC, p is determined by the VD, we must ensure that all desired PICs appear as the first p PICs. To this end, two versions of the PIPP are developed to address this issue. One is referred to as PI-based Prioritized PP (PI-PRPP) which uses a PI to rank appearing order of PICs in accordance with the information contained in PICs prioritized by the PI. It should be noted that the PI used for prioritization is not necessarily the same one used by the PIPP as also illustrated in Figs. 14.8–14.10. Although the PI-PRPP prioritizes PICs by a specific PI, it does not necessarily imply that the PI-PRPP generated PICs with same priorities are identical due to the randomness caused by random initial conditions. Therefore, a second version of the PIPP is further developed to mitigate this inconsistency issue. It is called Initialization-Driven PP (ID-PIPP) which uses a custom-designed initialization algorithm to produce a specific set of initial conditions for the PIPP. Since the initial conditions are always the same, the final PP-generated PICs are

always consistent. Finally, experiments are conducted to demonstrate the utility of these three PP-based DR techniques and results show their potentials in various applications.

References

1. Wang, J. and Chang, C.-I, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," IEEE Trans. on Geoscience and Remote Sensing, 44(6), 1586–1600 (2006).
2. Hyvarinen, A., Karhunen, J. and Oja, E., [Independent Component Analysis], John Wiley & Sons (2001).
3. Ren, H., Du, Q., Wang, J., Chang, C.-I and Jensen, J., "Automatic target recognition hyperspectral imagery using high order statistics," IEEE Trans. on Aerospace and Electronic Systems, 1372–1385 (2006).
4. Duda, R. O. and Hart, P. E., Pattern Classification and Scene analysis, John Wiley & Sons, 1973.
5. Friedman, J. H., and Tukey, J. W., "A projection pursuit algorithm for exploratory data analysis," IEEE Transactions on Computers, c-23(9), 881–889 (1974).
6. Chang, C.-I, Hyperspectral Imaging: Techniques for Spectral Detection and Classification, Kluwer Academic/Plenum Publishers, New York (2003).
7. Ren, H., Du, Q., Wang, J., Chang, C.-I and Jensen, J., "Automatic traget recognition hyperspectral imagery using high order statistics," IEEE Trans. on Aerospace and Electronic Systems, 42(4), 1372–1385 (2006).
8. Chu, S., Ren, H. and Chang, C.-I, "High order statistics-based approaches to endmember extraction for hyperspectral imagery," to be presented in SPIE 6966, (2008).
9. Hyvarinen, A. and Oja, E., "A fast fixed-point for independent component analysis," Neural Comp., 9(7), 1483–1492 (1997).
10. Ren, H. and Chang, C.-I, "Automatic spectral target recognition in hyperspectral imagery," IEEE Trans. on Aerospace and Electronic Sys., 39(4), 1232–1249 (2003).
11. Harsanyi, J. and Chang, C.-I, "Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach," IEEE Trans. on Geoscience and Remote Sensing, 32(4), 779–785 (1994).
12. Chang, C.-I, and Du, Q., "Estimation of number of spectrally distinct signal sources in hyperspectral imagery," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 608–619, March 2004.
13. Winter, M.E., "N-finder: an algorithm for fast autonomous spectral endmember determination in hyperspectral data," *Image Spectrometry V*, Proc. SPIE 3753, pp. 266–277, 1999.
14. Chang, Y.-C., Ren, H., Chang, C.-I and Rand, B., "How to design synthetic images to validate and evaluate hyperspectral imaging algorithms," SPIE Conference on *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIV*, March 16–20, Orlando, Florida, 2008.
15. Safavi, H., and Chang, C.-I , "Projection pursuit-based dimensionality reduction," SPIE Conference on *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIV*, March 16–20, Orlando, Florida, 2008.