International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol. 3, No. 1, March, 2010

About Classification Methods Based on Tensor Modelling for Hyperspectral Images

Salah Bourennane and Caroline Fossati

Ecole Centrale Marseille,
Institut Fresnel-UMR CNRS 61333 13397 Marseille cedex 20 France
Salah.bourennane@fresnel.fr, Caroline.fossati@fresnel.fr

Abstract

Denoising and Dimensionality Reduction (DR) are key issue to improve the classifiers efficiency for Hyper spectral images (HSI). The multi-way Wiener filtering recently developed is used, Principal and independent component analysis (PCA; ICA) and projection pursuit (PP) approaches to DR have been investigated. These matrix algebra methods are applied on vectorized images. Thereof, the spatial rearrangement is lost. To jointly take advantage of the spatial and spectral information, HSI has been recently represented as tensor. Offering multiple ways to decompose data orthogonally, we introduced filtering and DR methods based on multilinear algebra tools. The DR is performed on spectral way using PCA, or PP joint to an orthogonal projection onto a lower subspace dimension of the spatial ways. We show the classification improvement using the introduced methods in function to existing methods. This experiment is exemplified using real-world HYDICE data. Multi-way filtering, Dimensionality reduction, matrix and multilinear algebra tools, tensor processing.

Keywords: *Classification, Dimensionality Reduction, Tensor, ICA.*

1. Introduction

Detection and classification are key issues in processing Hyper Spectral Images (HSI). Spectral identification-based algorithms are sensitive to spectral variability and noise in acquisition. Recently, we have proposed multi-way Wiener algorithms that are robust to noise [1].

The emergence of hyper spectral images (HSI) implies the exploration and the collection of a huge amount of data. Imaging sensors provide typically up to several hundreds of spectral bands. This unreasonably large dimension not only increases computational complexity but also degrades the classification accuracy [2]. There of, dimensionality reduction (DR) is often employed. Due to its simplicity and ease of use, the most popular DR is the principal component analysis (*PCA*), referred to as PCA_{dr}. A refinement of PCA_{dr} is the independent component analysis (ICA_{dr}) [3]. While PCA_{dr} maximizes the amount of data variance by orthogonal projection, ICA_{dr} uses higher order statistics which characterize many subtle materials. On the other hand, projection pursuit (PP_{dr}) method [4-8] is define by using uses an index projection, to find interesting projections. Originally, the projections revealing the least Gaussian distributions have been shown to be the most interesting. Regarding its index

projections, which can be set, PP_{dr} method generalizes the matrix algebra based methods. In opposition with the previous DR methods, PP_{dr} uses a deflective rather than a global procedure by sequentially searching the projection. But all these matrix algebra methods require a preliminary step which consists in vectorizing the images. Therefore, they rely on spectral properties only, neglecting the spatial rearrangement. To overcome these disadvantages, [9-11] recently introduced a new HSI representation based on tensor. This representation involves a powerful mathematical framework for analyzing jointly the spatial and spectral structure of data. For this investigation, several tensor decompositions have been introduced [12-14]. In particular, the Tucker3 tensor decomposition [15,16] yields the generalization of the *PCA* to multi-way data. An extension of this decomposition can generalize the lower rank matrix approximation to tensors. This multilinear algebra tool is known as lower rank-(*K*1, *K2*, *K3*) tensor approximation [17], denoted by *LRTA*-(*K*1, *K*2, *K*3). In [1] multi-way filtering algorithms are developed and the shown results demonstrate that this new way to built the Wiener filter is very efficient for the multidimensional data such as HSI.

Based on this tensor approximation, this paper introduces two multilinear algebra methods for DR approach. These two methods yield a multi-way decorrelation and jointly perform a spatial-spectral processing: *p* spectral component are extracted, jointly with a lower spatial rank-(*K*1, *K*2) approximation. The latter spatial processing yields a projection onto a lower dimensional subspace that permits to spatially whiten the data. For the first method, the *PCAdr* is considered for the extraction of the *p* spectral components, this multiway method has been introduced in [10]. And for the second one, the *PPdr* is investigated. The two introduced multilinear algebra-based methods are referred to as *LRTAdr*-(*K*1, *K*2, *p*) and as hybrid *LRTA-PPdr*-(*K*1, *K*2, *p*). We show that the spatial projection onto a lower orthogonal subspace joint to spectral dimension reduction improves the classification efficiency compared to those obtained when matrix algebra-based DR is used.

The remainder of the paper is organized as follows: Section 2 introduces matrix algebra approaches to DR. Then, Section 3 reviews some tensor properties. Section 4 introduces the multilinear algebra method before drawing comparative classification in Section 5.

The following notations are used in the rest of the paper: scalars are denoted by italic lowercase roman, like a; vectors by boldface lowercase roman, like $\mathbf{a}$; matrices by boldface uppercase roman, like $\mathbf{A}$; tensors by uppercase calligraphic, like $\mathcal{A}$. We distinguish a random vector, like $\mathbf{a}$, from one of its realizations, by using a supplementary index, like $\mathbf{a}_i$.

2 Matrix algebra-based DR methods

2.1 HSI representation

To apply these matrix algebra-based methods, the HSI data are considered as a sampling of spectrum. Suppose that I_3 is the number of spectral bands and $I_1 \times I_2$ is the size of each spectral band image (*i.e.* $I_1 \cdot I_2$ is the number of samples). Each image pixel vector is an I_3-dimensional random variable. Those pixel vectors, referred to as spectral signatures, can be concatenated to yield a matrix $\mathbf{R} = \lfloor \mathbf{r}_1 \mathbf{r}_2 \ \ldots \mathbf{r}_{I_1 I_2} \rfloor$ of size $I_3 \times I_1 I_2$. In other words, the i^{th} row in $\mathbf{R}$ is specified by the i^{th} spectral band.

The DR approach extracts a lower number p of features called components, with $p < I_3$, such as

$$\mathbf{Y} = \mathbf{TR}, \tag{1}$$

where $\mathbf{T}$ is a linear transformation matrix of size $p \times I_3$, and $\mathbf{Y}$ the reduced matrix of size $p \times I_1 I_2$.

2.2 Principal component analysis based DR approach

In PCAdr context, the extracted components are called principal components (PCs). Each PC is generated by projecting the data spaced onto the i^{th} eigenvector associated with the i^{th} largest eigenvalue of the covariance matrix. This orthogonal projection maximizes the amount of data variance. Therefore the $p-$ spectral PCs generate a reduced matrix $\mathbf{Y}_{PCs}$ of size $p \times I_1 I_2$. Let $\mathbf{U}$ be the $p \times p$ matrix holding the p eigenvectors associated to the p first largest eigenvalues, the PCs are given by:

$$\mathbf{Y}_{PC} = \mathbf{U}^T \mathbf{X}, \tag{2}$$

In this case the matrix $\mathbf{U}^T$ is the transformation matrix referred to as $\mathbf{T}$ in Eq. (1). Sometimes, this transformation is associated to the sphering, given by:

$$\mathbf{Y} = \mathbf{\Lambda}^{-1/2} \mathbf{U}^T \mathbf{X}$$

with, $\mathbf{\Lambda}$ the $p \times p$ eigenvalue diagonal matrix of the covariance matrix.

2.3 Independent component analysis based DR approach

ICA [18] is an unsupervised source separation process , that has been applied to linear blind separation problem [19]. Its application to linear mixture analysis for HSI has been found in [20] and to DR approach in [3]. *ICA* assumes that data are linearly mixed and separates them into a set of statistically in dependent components (ICs). Since *ICA* requires higher-order statistics, many subtle materials or rare targets are more easily characterized. *ICA* finds a $p \times I_3$ separating matrix $\mathbf{W}$ to generate p ICs (with $p < I_3$) such that:

$$\mathbf{Y}_{IC} = \mathbf{WR}. \tag{3}$$

Commonly, a pre-processing step performs a *PCA* to sphere and reduce the samples. To generate p ICs, *FastICA* algorithm is selected using the absolute value of kurtosis as a measure of non-gaussianity. But, while PCs are generated by *PCAdr* in accordance with decreasing magnitude of eigenvalues, the *FastICA* algorithm does not necessarily generate ICs in order of information significance.

This is emphasized since this method estimate simultaneously all the $\mathbf{w}$ vectors of the matrix $\mathbf{W}$ (see Eq. (3)).

2.4 Projection pursuit based DR approach

Projection pursuit is a statistical analysis developed by [6] that searches the direction of the *interesting* projections. The *interesting* concept is defined by a projection index. Therefore,

projection pursuit-based DR method (PP_{dr}) selects a lower dimensional projection from high dimensional data by maximizing (or minimizing) the index projection. This technique is the generalization of the DR approach: if the index projection is the variance, PP_{dr} is similar to the PCA_{dr} , if the projection index is the kurtosis, PP_{dr} is similar to the ICA_{dr}. But, instead of being a global method like PCA_{dr} and ICA_{dr}, the PP_{dr} is a deflective method. Thus, after finding the first direction which maximizes the projection index, the data are projected onto the orthogonal subspace of this direction and so forth. In this paper, we select the non-gaussianity index. For this investigation, the *FastICA* algorithm in deflective mode estimates only one projection, $\mathbf{w}^T\mathbf{r}$ at a time. When i vectors $\mathbf{w}_1...\mathbf{w}_i$ components are estimated, *FastICA* is used to estimate the $\mathbf{w}_{i+1}$ vector and projection contribution of the i previous vectors are deducted, such as:

$$\mathbf{w}_{i+1} \leftarrow \mathbf{w}_{i+1} - \sum\left(\mathbf{w}_{i+1}^T\mathbf{w}_j\right)\mathbf{w}_j . \tag{4}$$

The reduced matrix $\mathbf{Y}_p$, holding the p projections, is equivalent to $\mathbf{Y}_{IC}$ (Eq. (3)) using the non-gaussianity index but generated the vectors $\mathbf{w}$ using the previous process. PP_{dr} approach finds the p first projection pursuit directions to attain a lower p -dimensional space.

3. Tensor representation and some properties

PCA_{dr}, ICA_{dr} and PP_{dr} are matrix algebra-based methods which require data rearrangement. In this paper, we use multilinear algebra tools to consider the whole data. Our methods are based on tensor representation to keep the initial spatial structure and insure the neighborhood effects. Tensor processing have proven its efficiency in several domains, telecommunications [21], image processing [1, 22, 23] and, more recently in hyperspectral image analysis [9-11].

In this representation, the whole HSI data is considered as a *third*-order tensor, the entries of which are accessed via three indices. It is denoted by $\mathcal{R} \in \mathbf{R}^{I_1 \times I_2 \times I_3}$, with elements arranged as $r_{i_1 i_2 i_3}, i_1 = 1,...,I_1; i_2 = 1,...,I_2; i_3 = 1,...,I_3$ and $\mathbf{R}$ is the real manifold. Each index is called mode: two spatial and one spectral modes characterize the HSI tensor.

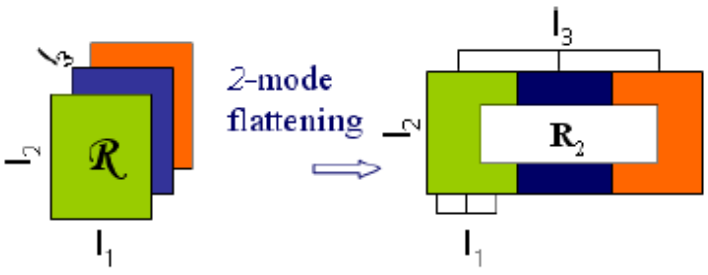


Figure 1. 2-mode flattening of tensor $\mathcal{R}$.

Tensor representation is mathematically grounded in multilinear algebra which studies the properties of data tensor $\mathcal{R}$ in a given n-mode. Let us define $E^{(n)}$, the n-mode vector space of dimension I_n, associated with the n-mode of tensor $\mathcal{R}$. By definition, $E^{(n)}$ is generated by the column vectors of the n-mode flattened matrix. The n-mode flattened matrix $\mathbf{R}_n$ of tensor

$\mathcal{R} \in \mathbf{R}^{I_1 \times I_2 \times I_3}$ is defined as a matrix from $\mathbf{R}^{I_n \times M_n}$, with: $M_n = I_p I_q$ and $p, q \neq n$. An illustration of the *n*-mode flattening of a *third*-order tensor is represented in Fig. 1. Tensor representation yields to process the whole data from spatial and spectral perspectives.

This representation naturally implies the use of multilinear algebraic tools and especially tensor decomposition and approximation methods. The most commonly used is the Tucker3 [15] decomposition, which generalizes to higher order, the singular value decomposition. This tensor decomposition is expressed as follows:

$$\mathcal{R} = C \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)}, \tag{5}$$

where C is the core tensor, $\mathbf{U}^{(n)}$ is the matrix of eigen vectors associated with the *n*-mode covariance matrix $\mathbf{R}_n \mathbf{R}_n^T$.

4 Multilinear algebra-based DR method

4.1 Tensor formulation of PCA_{dr} and PP_{dr}

Here, we write the equivalent of the Eqs. (1)-(3) in tensor formulation [12-14,17]. et $\mathcal{R} \in \mathbf{R}^{I_1 \times I_2 \times I_3}$ be the HSI data. The previously obtained matrix $\mathbf{R}$ (Section 2.1) is equivalent to the *3*-mode flattened matrix of $\mathcal{R}$ denoted by $\mathbf{R}_3$. Thus in tensor formulation, the DR approach (Eq. (1)) including the image reshaping can be written as follows:

$$\mathcal{Y} = \mathcal{R} \times_3 \mathbf{T}, \tag{6}$$

where $\mathcal{Y}$ is the reduced *three*-order tensor $\in \mathbf{R}^{I_1 \times I_2 \times p}$ holding the p components. $\times_n$ is the *n*-mode product [12,17] generalizing the product between a tensor and a matrix along an *n*-mode.

In the same way, the equivalent to Eq. (2) including the image reshaping is formulated:

$$\mathcal{Y}_{PC} = \mathcal{R} \times_3 \mathbf{U}^{(3)^T}, \tag{7}$$

where $\mathcal{Y}_{PC}$ is a *three*-order tensor $\in \mathbf{R}^{I_1 \times I_2 \times p}$ holding the *p* PCs and $\mathbf{\Lambda}, \mathbf{U}^{(3)}$ the eigenvalues and eigenvectors of the covariance matrix of $\mathbf{R}_3$. $\mathbf{U}^{(3)}$ is the equivalent to $\mathbf{U}$ defined in Eq. (2).

And the equivalent to Eq. (3) including the image reshaping is formulated:

$$\mathcal{Y}_P = \mathcal{R} \times_3 W, \tag{8}$$

where, $\mathcal{Y}_P$ is a *three*-order tensor $\in \mathbf{R}^{I_1 \times I_2 \times p}$ holding the p projections. Eqs. (6)-(8) highlight the spectral or *3*-mode processing in the traditional DR method, neglecting the spatial modes.

4.2 Multilinear algebra and *PCA*-based DR method

The multilinear algebra-based DR method has two objectives: (i) estimate the matrix $\mathbf{U}^{(3)}$ of Eq. (7) using spatial information; (ii) make a joint spatial-spectral processing with the aim of whitening and compressing the spatial and spectral modes. Consequently, the proposed

method performs simultaneously a dimensionality reduction of the spectral mode $(p < I_3)$, and a projection onto a lower (K_1, K_2)– dimensional subspace of the two spatial modes.

To attain these two objectives, the *PCA*-based DR proposed multilinear algebra tool has two approaches [10]: (i) find the lower rank-(K_1, K_2, K_3) approximation tensor of $\mathcal{R}$, this method is referred to as the *LRTA*-(K_1, K_2, K_3) and the resulting tensor as $\mathcal{Y}_{(K_1,K_2,K_3)}$; (ii) keep only the p first principal spectral components. Thereof, the proposed method that is the association of these two approaches is referred to as the *LRTA*$_{dr}$-(K_1, K_2, p) and the resulting reduced and approximated tensor to as $\mathcal{Y}_{PC\text{-}(K_1,K_2,p)}$. Contrary to the *LRTA*-(K_1, K_2, K_3), the *LRTA*$_{dr}$- (K_1, K_2, p) extracts p principal components in the spectral mode. The spatial subspace dimension is the K_n -value, for n = 1, 2 and $p\,(< I_3)$ the number of retained features. For convenience, the K_n-value is denoted by $K_{1,2}$ in the following.

The *LRTA*-(K_1, K_2, K_3) minimizes the following quadratic Frobenius norm: $\left\|\mathcal{R} - \mathcal{Y}\right\|_F^2$. Finding the lower rank approximation, using the Tucker3 decomposition, consists in estimating the $\mathbf{U}^{(n)}$ orthogonal matrix. [17] shows that minimizing the quadratic Frobenius norm with respect to $\mathcal{Y}$ amount to maximize with respect to $\mathbf{U}^{(n)}$ matrix the quadratic function:

$$g\left(\mathbf{U}^{(1)}, \mathbf{U}^{(2)}, \mathbf{U}^{(3)}\right) = \left\|\mathcal{R} \times_1 \mathbf{U}^{(1)T} \times_2 \mathbf{U}^{(2)T} \times_3 \mathbf{U}^{(3)T}\right\|_F^2, \tag{9}$$

The least square solution involves the *LRTA*-(K_1, K_2, K_3) expression:

$$\mathcal{Y}_{(K_1,K_2,K_3)} = \mathcal{R} \times_1 \mathbf{P}^{(1)} \times_2 \mathbf{P}^{(2)} \times_3 \mathbf{P}^{(3)}, \tag{10}$$

where $\mathbf{P}^{(n)} = \mathbf{U}^{(n)}_{1\ldots K_n} \mathbf{U}^{(n)\ T}_{1\ldots K_n}$, n = 1; 2; 3.

The reduced tensor, $\mathcal{Y}_{PC\text{-}(K_1,K_2,p)}$, is obtained by simultaneously performing the previous approximation and by extracting the p PCs using *PCA*$_{dr}$. This statement can be written as follows [10]:

$$\mathcal{Y}_{PC\text{-}(K_1,K_2,p)} = \mathcal{R} \times_1 \mathbf{P}^{(1)} \times_2 \mathbf{P}^{(2)} \times_3 \mathbf{\Lambda}^{-1/2} \mathbf{U}^{(3)T}, \tag{11}$$

The joint estimation of the orthogonal matrix $\mathbf{U}^{(n)}$, $\forall n$, which solve the Eq. (9), is feasible using an alternating least squares algorithm (ALS). The algorithm 1 summarizes the *LRTA*$_{dr}$-(K_1, K_2, p) method.

Algorithm 1 : *LRTA*$_{dr}$-(K_1, K_2, p)

input: $\mathcal{R} \in \mathbf{R}^{I_1 \times I_2 \times I_3}$, K_1, K_2, p -values

output: $\mathcal{Y}_{PC\text{-}(K_1,K_2,p)} \in \mathbf{R}^{I_1 \times I_2 \times p}$

initialization: $i = 0, \forall n = 1,2$

$U^{(n),0} \leftarrow$ first K_n eigenvectors of matrix $\mathrm{E}\left[\mathbf{R}_n \mathbf{R}_n^T\right]$

repeat

ALS loop :
For all $n = 1,2,3$ **do**
$\hat{\mathcal{R}}^{i} \leftarrow \mathcal{R} \times_{q} \mathbf{U}^{(q),i^T} \times_{r} \mathbf{U}^{(r),i^T}$ with $q, r \neq n$
$\mathrm{E}\left[\hat{\mathbf{R}}_n^i \hat{\mathbf{R}}_n^{i\,T}\right]$ eigenvalue decomposition
$U^{(n),i+1} \leftarrow$ first K_n or p eigenvectors
$i \leftarrow i+1$
end
until *convergence of Eq. (9)*
$\mathbf{P}^{(n)} \leftarrow \mathbf{U}^{(n)}\mathbf{U}^{(n)T}$
$\mathcal{Y}_{PC\text{-}(K_1,K_2,p)} = \mathcal{R} \times_1 \mathbf{P}^{(1)} \times_2 \mathbf{P}^{(2)} \times_3 \mathbf{\Lambda}^{-1/2}\mathbf{U}^{(3)T}$

4.3 Multilinear algebra and *PP*-based DR method

To pursue this previous work, we introduce a generalized multilinear algebra approach to DR using projection pursuit. The same objectives, presented in the previous section, are considered. This multilinear algebra-based method is referred to as the hybrid *LRTA-PP*$_{dr}$-$\left(K_1, K_2, p\right)$ and the resulting approximated and reduced tensor as $\mathcal{Y}_{P\text{-}(K_1,K_2,p)}$. Finding $\mathcal{Y}_{P\text{-}(K_1,K_2,p)}$ consists in minimizing the Frobenius norm: $\left\|\mathcal{Y}_P - \mathcal{Y}_{P\text{-}(K_1,K_2,p)}\right\|_F^2$ with $\mathcal{Y}_P$ define in Eq. (8)

Using the Tucker3 representation, the reduced tensor $\mathcal{Y}_P$ (Eq. (8)) can be decomposed as follows:

$$\mathcal{Y}_P = C \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)}, \tag{12}$$

The lower rank-$\left(K_1, K_2\right)$ approximated tensor of $\mathcal{Y}_P$ to be estimated, can be also decomposed following the Tucker3 representation as:

$$\mathcal{Y}_{P-(K_1,K_2,p)} = \mathcal{D} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)}, \tag{13}$$

Following the minimizing criterion, we can express the core tensor $\mathcal{D}$ such as:

$$\mathcal{D} = \mathcal{Y}_P \times_1 \mathbf{U}^{(1)T} \times_2 \mathbf{U}^{(2)T} \times_3 \mathbf{U}^{(3)T}$$

As a result:

$$\mathcal{D} = \mathcal{R} \times_1 \mathbf{U}^{(1)T} \times_2 \mathbf{U}^{(2)T} \times_3 \mathbf{U}^{(3)T}\mathbf{W}, \tag{14}$$

Finally, the *LRTA-PP*$_{dr}$-$\left(K_1, K_2, p\right)$ method is expressed as:

$$\mathcal{Y}_{P-(K_1,K_2,p)} = \mathcal{R} \times_1 \mathbf{P}^{(1)} \times_2 \mathbf{P}^{(2)} \times_3 \mathbf{W}, \tag{15}$$

As previously, the joint estimation of the orthogonal matrix $\mathbf{U}^{(n)}$, $\forall n$, and the demixing matrix $\mathbf{W}$ is feasible using an alternating least squares algorithm (ALS). The algorithm 2 summarizes the hybrid *LRTA-PP*$_{dr}$-$\left(K_1, K_2, p\right)$ method.

Algorithm 2 : hybrid *LRTA-PP$_{dr}$-* (K_1, K_2, p)

input: $\mathcal{R} \in \mathbf{R}^{I_1 \times I_2 \times I_3}$, K_1, K_2, p -values

output: $\mathcal{Y}_{P\text{-}(K_1,K_2,p)} \in \mathbf{R}^{I_1 \times I_2 \times p}$

initialization: $i = 0, \forall n = 1,2$

$U^{(n),0} \leftarrow$ first K_n eigenvectors of matrix $\mathrm{E}\left[\mathbf{R}_n \mathbf{R}_n^T\right]$

while $\left\| \mathcal{Y}^{i+1}_{P\text{-}(K_1,K_2,p)} - \mathcal{Y}^{i}_{P\text{-}(K_1,K_2,p)} \right\|_F^2 >$ *threshold* do

ALS loop :

for $n = 3$ **do**

$\hat{\mathcal{R}}^{i} \leftarrow \mathcal{R} \times_1 \mathbf{U}^{(1),i+1^T} \times_2 \mathbf{U}^{(2),i+1^T}$

$\mathrm{E}\left[\hat{\mathbf{R}}_n^i \hat{\mathbf{R}}_n^{i\,T}\right]$ eigenvalue decomposition

$U^{(3),i+1} \leftarrow p$ first eigenvectors

$\hat{\mathcal{R}}^{i} \leftarrow \hat{\mathcal{R}}^{i} \times_3 \mathbf{U}^{(3),i+1^T}$

$\mathbf{W} \leftarrow$ *fastica*, deflective approach from $\hat{\mathcal{R}}^{i}$

end

for $n = 1,2$ **do**

$\hat{\mathcal{R}}^{i} \leftarrow \mathcal{R} \times_q \mathbf{U}^{(q),i^T} \times_3 \mathbf{W}^i$ with $q \neq n$

$\mathrm{E}\left[\hat{\mathbf{R}}_n^i \hat{\mathbf{R}}_n^{i\,T}\right]$ eigenvalue decomposition

$U^{(n),i+1} \leftarrow$ first K_n

end

$\mathbf{P}^{(n)} \leftarrow \mathbf{U}^{(n),i+1} \mathbf{U}^{(n),i+1^T}$

$\mathcal{Y}^{i+1}_{P\text{-}(K_1,K_2,p)} \leftarrow \mathcal{R} \times_1 \mathbf{P}^{(1)} \times_2 \mathbf{P}^{(2)} \times_3 \mathbf{W}$

$i \leftarrow i + 1$

5. Experimental results

In this section, two experiments are introduced. First, the interest of the multi-linear algebra-based filtering and DR methods are highlighted using simulated data and showing the scatter plots. Secondly, the classification improvement is exemplified using two real-world HYDICE data.

5.1 Experiment on simulated data

In order to show the interest of multi-way filtering on the classification, we consider a simulated data shown Fig.5, where for the original image a Gaussian noise with variable variance is added. To quantify the benefits of our filter, for each SNR we calculate the overall(OA) classification rate exhibits after multi-way Wiener filtering. For example, the Spectral-Angle Mapper (SAM) is investigated. For this experiment, a simulated data is considered (see Fig. 2a). The spatial size of this simulated data is 50×50 with 150 spectral channels. Three endmembers are present associated with a each class. The endmembers

spectral signature are introduced in Fig. 2a to appreciate the similarity. A Gaussian random number generator simulated 2100 samples associated with the first class, 300 with the second class and 100 with the third one. The scatter plot in Figs. 2b show that close spectral channels are more correlated as farther ones. The set of dots, that portray the samples, highlights that the three classes are sometimes not distinct enough.

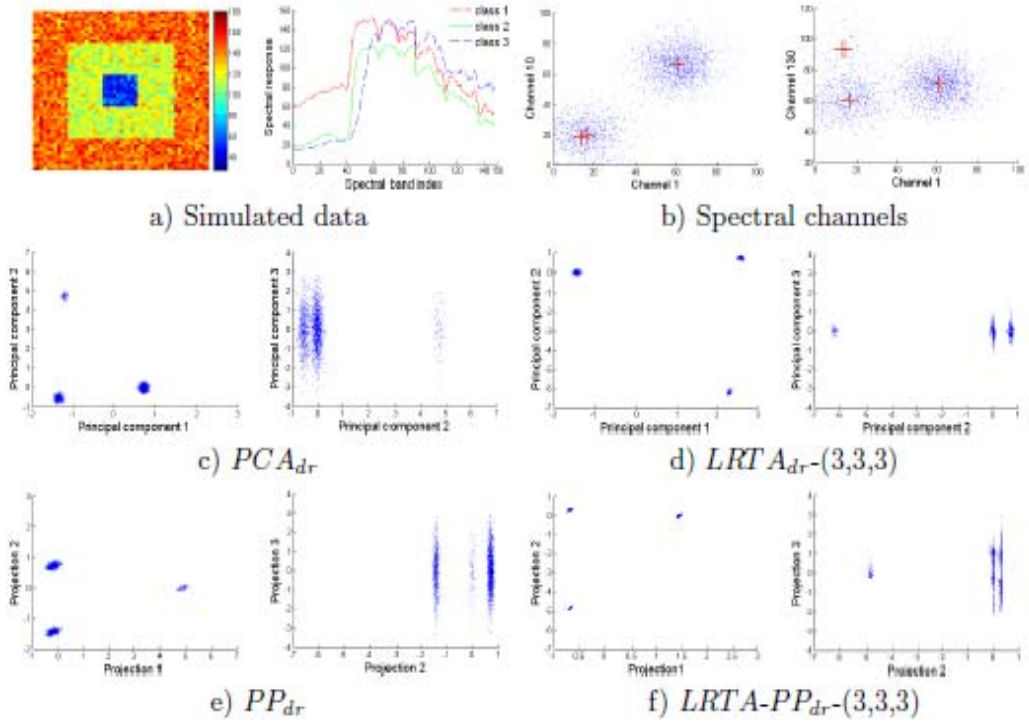


Figure 2: a) simulated tensor and scatter plots of, b) close and farther spectral channels (red crosses are endmembers), c) PCs, d) PCs obtained using LRTA_{dr}-(3,3,3), e) projections, f) projections obtained using hybrid $LRTA$-PP_{dr}-(3,3,3).

DR methods are applied to extract three components. Figures 2c and 2e respectively show the scatter plot of the spectral components obtained using PCA_{dr} and ICA_{dr}. Two scatter plots are represented: first and second components then the second and third components. This very basic classification context does not allow to compare the PCA_{dr} and ICA_{dr}-based spectral DR, but yields to show the interest of the multilinear algebra based methods. Then, the scatter plot obtained using PCA_{dr} and ICA_{dr} can be compared with those obtained using the $LRTA_{dr}$-(K_1, K_2, p) and hybrid $LRTA$-PP_{dr}-(K_1, K_2, p) in Figs. 2d and 2f respectively. For this experiment, the $K_{1,2} - value$ are set to 3.Those comparisons denoted that the sets of dots are more concentrated and more distinct in Figs. 2d and 2f than those in Figs. 2c and 2e respectively. This is due to the spatial decorrelation of the components by projecting the data onto an orthogonal lower subspace of the $LRTA_{dr}$-(K_1, K_2, p) and the hybrid $LRTA$-PP_{dr}-(K_1, K_2, p) DR methods. This interesting remark forecasts classification improvement, that is exemplified, in the next section, using real-world HYDICE data.

5.2 Experiment on real-world data

Two real-world images collected by HYDICE [24] imaging are considered for this investigation, with a 1.5 m spatial and 10 nm spectral resolution. The first one, referred to as HSI02, has 150 spectral bands (from 435 to 2326 nm), 310 rows and 220 columns. HSI02 can be represented as a *three*-order tensor, referred to as $\mathcal{R} \in \mathbf{R}^{310\times220\times150}$. The second one, referred to as HSI03, has 164 spectral bands (from 412 to 2390 nm), 170 rows and 250 columns. HSI03 can be represented as a *three*-order tensor, referred to as $\mathcal{R} \in \mathbf{R}^{170\times250\times164}$.

Table 1. Information classes and samples

HSI02			HSI03		
Classes	Training samples	Test samples	Classes	Training samples	Test samples
field	1 002	40 811	house	65	613
trees	1 367	5 537	road	210	2364
road	139	3 226	trees	148	11482
shadow	372	5 036	field	325	5353
target 1	128	519			
target 2	78	285			
target 3	37	223			

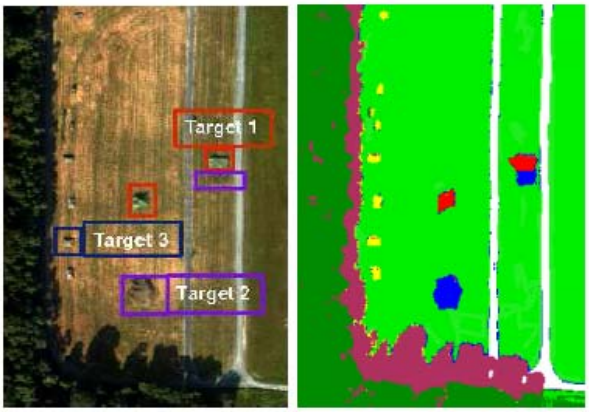
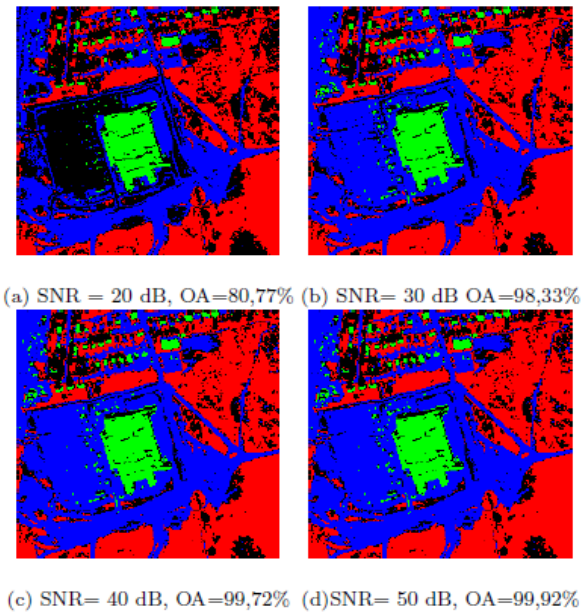


Figure 3. Classes in the HYDICE image HSI02 and its ground truth.

Two experiments are introduced in this section. The first one shows the efficiency of the ALS algorithm and, the $K_{1,2}-value$ influence is highlighted in the second experiment. To quantify the improvement, we focus on the overall classification (OA) rate exhibits after DR methods. Three classifiers are investigated: *(i)* maximum likelihood (ML), *(ii)* Mahalanobis distance (MD) and *(iii)* spectral angle mapper (SAM).

Figure 4. Classes in the HYDICE image HSI03 and its ground truth.



(a) SNR = 20 dB, OA=80,77% (b) SNR= 30 dB OA=98,33%

(c) SNR= 40 dB, OA=99,72% (d)SNR= 50 dB, OA=99,92%

Figure 5. OA value evolution in function of the increasing ((a)-(d)) SNR .

Efficiency of the ALS algorithm. The ALS algorithm guarantees the cross dependency between the spatial and spectral processing realizing by the proposed multilinear algebra-based methods. To exemplified the ALS efficiency, Fig. 6 shows the OA evolution in function of the iteration index of the ALS algorithm, when ML, SAM and, MD classifiers are considered.

For this simulation the $K_{1,2}$ and p values are set to (40, 40, 20) and (60, 60, 20) according to the use of HSI02 and HSI03 respectively. It is worth noting that the OA collected values at the iteration 0 are the classification results obtained on the real-world image without DR processing. This reading highlights the noteworthy improvement from the first iteration of the ALS when $LRTA_{dr}$-(K_1, K_2, p) and hybrid $LRTA\text{-}PP_{dr}$-(K_1, K_2, p) are used. Figure 6 shows that the classification result is iteratively improved until a maximum value is reached

whatever the classifier and HYDICE data used. This state is attained after a few iterations, less than 10.

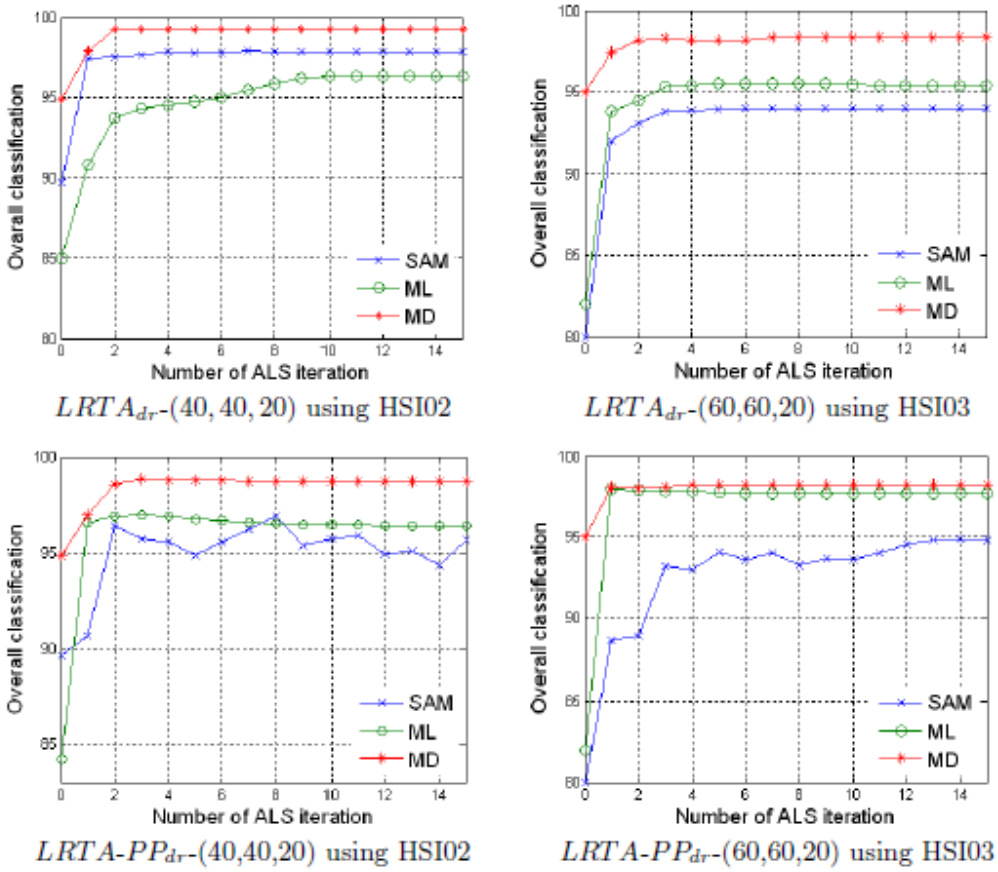


Figure 6. OA value evolution in function of the number of ALS iteration.

However, when hybrid *LRTA-PP*$_{dr}$-(K_1, K_2, p) DR method is considered, the OA value obtained from the SAM classifier, after notable increase, oscillate around a value. The SAM classifier results seem to have more variations than the others classifiers in function of the hybrid *LRTA-PP*$_{dr}$-(K_1, K_2, p)iteration index. This is due probably to the fact that the hybrid *LRTA-PP*$_{dr}$-(K_1, K_2, p) performs a demixing matrix $(\mathbf{W})$ estimation at each iteration. The estimation of **W** using *FastICA* algorithm is not unique since it crucially depends of the initialization step. Moreover, the SAM classifier has an unclassified class for the pixel which has a spectral angle measure, with all classes, superior to a thresh-old (fixed to 0.15 in the experiment). Then, from one iteration to another, some samples with spectral angle measure close to the threshold can be unclassified or well-classified, this can explain the iterative fluctuation of the SAM classifier. Note that these classification results are obtained for a specific $K_{1,2}$ and p values and, the maximum attained OA value depends on these values.

Influence of the ($K_{1,2}$)-value. The influence of the number p of retained components is conceded. In this section, we assess the influence of the spatial dimension **(** $K_{1,2}$ **)** of the projection subspace which is required by the proposed multilinear algebra-based DR methods. The $K_{1,2}$ value is the number of first eigenvalues of the $\mathbf{R}_n$ covariance matrix, kept in the Algorithms 1 and 2, for n= 1, 2.

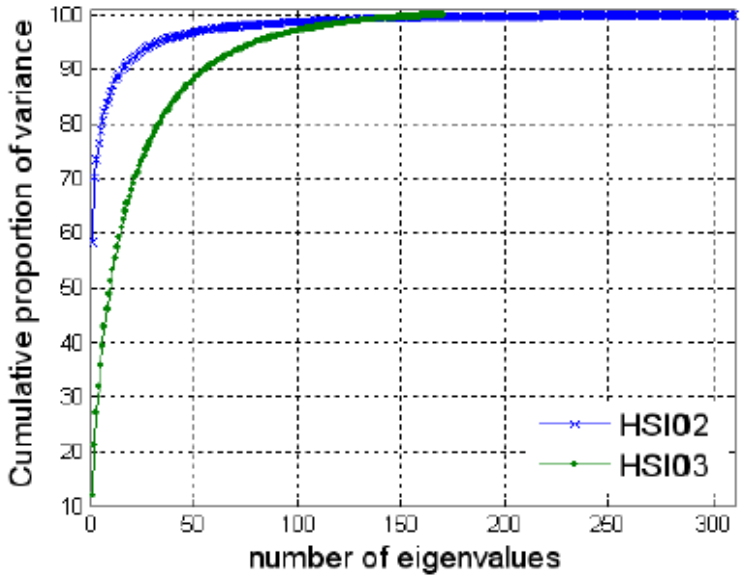


Figure 7. Cumulative proportion of the variance in function of the eigenvalue number $K_{1,2}$ of the 1-mode covariance matrix.

In this experiment, the classification results are assessed as a function of the ($K_{1,2}$)-value. For sake of clarity, only the OA exhibits by the Mahalanobis distance classifier are illustrated. Figure 7 specifies the cumulative proportion variance in function of the number of eigenvalues kept. It concerns the eigenvalue of the covariance matrix of flattened $\mathbf{R}_n$, for n = 1, 2. Thereby, we appreciate at what time this gain is negligible and, we consider that the others eigenvalues providing no more information, can be suppressed. For example, concerning HSI02, we consider that from $K_{1,2}$ value equal to 50, the gain is less than 1% and becomes insignificant. Concerning HSI03, it is reached when $K_{1,2}$ value equal to 90. At the following, the $K_{1,2}$ is set to 50 when HSI02 is used and 90 when HSI03 is used.

6. Conclusion

In this paper, we show the interest of multi-way filtering in terms of classification. We also introduced two multilinear algebra-based DR methods, the $LRTA_{dr}$-(K_1, K_2, p) and the hybrid $LRTA\text{-}PP_{dr}$-(K_1, K_2, p). These techniques make a joint spatial/spectral processing with the aim of decorrelated and compressing the spatial and spectral dimension simultaneously. The

cross-dependency of the spatial and spectral processing is guaranteed using an alternating least squares (ALS) algorithm. The spatial/spectral processing concerns: (i) a lower spatial rank-(K_1, K_2) approximation is performed to project the spatial dimension onto a lower subspace dimension $(K_{1,2} < I_{1,2})$; (ii) a spectral dimension reduction is performed to extract p components $(p < I_3)$. For this latter issue, when $LRTA_{dr}$-(K_1, K_2, p) is considered, PCA_{dr} is used and, when hybrid $LRTA$-PP_{dr}-(K_1, K_2, p) is considered, PP_{dr} is used.

Classification results are exemplified using two real-world HYDICE data and highlight the overall classification improvement when multilinear algebra-based methods are used rather than the traditional matrix algebra-based methods.

Acknowledgments

The authors thank Dr. N. Renard and A. Cailly for their helpful for the computer simulations.

References

[1] D. Muti and S. Bourennane, "Survey on tensor signal algebraic filtering", Signal Proceesing Journal, Elsevier., vol. 87, pp. 237-249, Feb. 2007.

[2] D. Landgrebe, "Hyperspectral image data analysis as a high dimensional signal processing problem", IEEE Signal Process. Magazine, vol. 19, no. 1, pp. 17-28, Jan. 02.

[3] J. Wang and C. Chang, "Independent component analysis based dimensionality reduction with applications in hyperspectral image analysis", IEEE Trans. On Geosc. and Remote Sens., vol. 44, no. 6, pp. 1586-1588, June 2006.

[4] A. Ifarraguerri and C. Chang, "Unsupervised hyperspectral image analysis with projection pursuit", IEEE Trans. on Geoscience and Remote Sensing, vol. 38, no. 6, pp. 2529-2538, Nov. 2000.

[5] M. Jones and R. Sibson, "What is projection pursuit," J.R. Statist. Soc. A, vol. 150, no. 1, pp. 1-36, 1987.

[6] J. Friedman and J. tuckey, "A projection pursuit algorithm for exploratory data analysis," IEEE Transactionson Computers, vol. 23, no. 9, pp. 881-889, 1974.

[7] L. Jimenez and D. Landgrebe, "Hyperspectral data analysis and supervised feature reduction via projection pursuit", IEEE Trans. on Geoscience and Remote Sensing, vol. 37, no. 6, pp. 2653-2667, 1999.

[8] L. Jimenez-Rodriguez, E. Arzuaga-Cruz, and M. Velez-Reyes, "Unsupervised linear feature-extraction methods and their effects in the classification of high-dimensional data", IEEE Transactions on Geoscience and Remote Sensing, vol. 45, no. 2, pp. 469-483, 2007.

[9] N. Renard and S. Bourennane, "Improvement of target detection methods by multiway filtering", IEEE Trans. on Geoscience and Remote Sensing, vol. 46, no. 8, pp. 2407-2417, Aug. 2008.

[10] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images", IEEE Geoscience and Remote Sensing Letters, vol. 5, no. 2, pp. 138-142, April 2008.

[11] D. Letexier, S. Bourennane, and J. Blanc-Talon, "Nonorthogonal tensor matricization for hyperspectral image filtering", IEEE Geoscience and Remote Sensing Letters, vol. 5, no. 1, pp. 3-7, Janv. 2008.

[12] L. Tucker, "Some mathematical notes on three-mode factor analysis", Psychometrika, vol. 31, pp. 279-311, 66.

[13] T. Kolda, "Orthogonal tensor decomposition", SIAM Joural on Matrix Analysis and Applications, vol. 23, no. 1, pp. 243-55, Mar. 01.

[14] R. Harshman and M. Lundy, “Research methods for multilmode data analysis”, H. Law, J. Snyder, J. Hattie, and R. McDonald, Eds. Praeger, New York, 70, pp. 122-215.

[15] L. Tucker, “The extension of factor analysis to three-dimensional matrices”. Holt, Rinehart and Winston, NY, 1964, pp. 109-127.

[16] L. De Lathauwer, B. De Moor, and J. Vandewalle, “A multilinear singular value decomposition”, SIAM Journal on Matrix Analysis and Applications, vol. 21, pp. 1253-1278, April 2000.

[17] “On the best rank-(r1; : : : ; rN) approximation of higher-order tensors”, SIAM Journal on Matrix Analysis and Applications, vol. 21, pp. 1324-{42, Apr. 00.

[18] P. Comon, “Independent component analysis: A new concept”, Signal Process., vol. 36, pp. 287-314, 1994.

[19] A. Hyvarinen, J. Karhunen, and E. Oja, Independent Component Analysis. John Wiley Sons, 2001.

[20] C. Chang, S. Chiang, J. Smith, and I. Ginsberg, “Linear spectrum random mixture analysis for hyperspectral imagery”, IEEE Trans. on Geoscience and Remote Sensing, vol. 40, no. 2, pp. 375-392, Feb. 2002.

[21] N. Sidiropoulos, G. Giannakis, and R. Bro, “Blind parafac receivers for DS-CDMA systems”, IEEE Trans. on Signal Processing, vol. 48, no. 3, pp. 810-823, 2000.

[22] T. Hazan, S. Polak, and A. Shashua, “Sparse image coding using a 3d nonnegative tensor factorization”, in Computer Vision, IEEE International Conference on Computer Vision (ICCV), vol. 4179, no. 1, Beijing, (China), Oct 2005, pp.50-57.

[23] M. Vasilescu and D. Terzopoulos, “Multilinear independent components analysis”, Proc. of IEEE CVPR'05., vol. 1, pp. 547-53, 05.

[24] R. Basedow, W. Aldrich, J. Colwell, and W. Kinder, “Hydice system performance an update”, in SPIE Proc., vol. 2821, Denver, CO, August 1996, pp. 76-84.