AN ICA-BASED MULTILINEAR ALGEBRA TOOLS FOR DIMENSIONALITY REDUCTION IN HYPERSPECTRAL IMAGERY

N. Renard and S. Bourennane

Institut Fresnel / UMR-CNRS, D. U. de Saint-Jérôme, 13397 Marseille cedex 20 France

ABSTRACT

Dimensionality reduction (DR) is a major issue to improve the efficiency of the classifiers in Hyperspectral images (HSI). Recently, the independent component analysis (ICA) approach to DR has been investigated. But, this signal processing is applied on vectorized images, losing spatial rearrangement. 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 develop a new DR method based on multilinear algebra tools and on ICA. The DR is performed on spectral way using ICA jointly to an orthogonal projection onto a lower subspace dimension of the spatial ways. We show the Maximum Likelihood classifier improvement using the proposed method.

Index Terms— Dimensionality reduction, independent component analysis, multilinear algebra tools, tensor processing.

1. INTRODUCTION

The emergence of hyperspectral 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 classification accuracy [1]. Dimensionality reduction (DR) is often employed. Due to its simplicity and ease of use, the most popular DR method is the PCA, referred to as PCA_{dr}. A refinement of PCA_{dr} is the independent component analysis (ICA), referred to as ICA_{dr} [2]. While PCA_{dr} maximizes the amount of data variance by orthogonal projection, ICA_{dr} uses higher order statistics. But these matrix algebra methods requires a preliminary step which consists in vectorizing the images. Therefore, they rely on spectral properties only, neglecting to the spatial rearrangement.

To overcome this weakness, [3] recently introduced a new HSI representation based on tensor in DR context. This representation involves a powerful mathematical framework for analyzing jointly the spatial and spectral structure of data. [3] proposes an multilinear algebra-based method yielding a multi-way decorrelation. Joint spatial-spectral processing is performed : p spectral components are extracted using PCA_{dr}, 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 data. Referred to as lower rank-(K_1, K_2, p) tensor approximation based DR method ($LRTA_{dr}$-(K_1, K_2, p)), [3] has shown the classification efficiency improvement compared to when PCA_{dr} is used.

In this paper, we proposed a multilinear algebra tool when ICA_{dr} is considered for the extraction of the p spectral components. We show that the spatial decorrelation joint to spectral dimension reduction based on ICA_{dr} improves the classification efficiency compared to those obtained when ICA_{dr} is used. The proposed method is referred to as hybrid $LRTA$-ICA_{dr}-(K_1, K_2, p).

Section 2 introduces the ICA_{dr} method for HSI. Then in Section 3, the tensor representation and the $LRTA_{dr}$-(K_1, K_2, p) is briefly reviewed before outlying the proposed hybrid $LRTA$-ICA_{dr}-(K_1, K_2, p) in Section 4. Section 5 presents comparative classification result.

2. INDEPENDENT COMPONENT ANALYSIS FOR DR, ICA_{DR}

ICA [4] is an unsupervised source separation process, that has been applied to linear blind separation problem [5]. Its application to linear mixture analysis for HSI has been found in [6]. ICA assumes that data are linearly mixed and separates them into a set of statistically independent components (ICs). Since ICA requires higher-orders, many subtle materials or rare targets are more easily characterized.

To apply this signal processing, 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. Each image pixel vector is an I_3-dimensional random variable. Those pixel vectors, referred to as spectral signatures, are concatenated to yield a matrix $\mathbf{R} = [\mathbf{r}_1\ \mathbf{r}_2\ \cdots\ \mathbf{r}_{I_3}]$ 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. Commonly, each vector pixel is considered as a linear mixture of a set of p known endmembers, as follows :

$$\mathbf{r} = \mathbf{M}\mathbf{s}, \tag{1}$$

where $\mathbf{M}$ is an $I_3 \times p$ target signature matrix where $\mathbf{m}_j$ is the j^{th} target signature and p the number of target in the image, and $\mathbf{s}$=$[s_1\ s_2\ \cdots\ s_p]^T$ is the abundance column vector. ICA

finds a $p \times I_3$ separating matrix $\mathbf{W}$ to generate p ICs (with $p < I_3$) such that :

$$\mathbf{y}_{IC} = \mathbf{W}\mathbf{r}, \qquad (2)$$

where, $\mathbf{y}_{IC}$ is a p-dimensional vector. To generate p ICs, *fastica* algorithm is selected using the absolute value of kurtosis as a measure of nongaussianity. Commonly a pre-processing step consists in performing a PCA to sphere and reduce the samples. ICA is applied on a $p \times I_1 I_2$ matrix $\mathbf{Y}_{PC}$:

$$\mathbf{Y}_{PC} = \mathbf{D}^{-1/2}\mathbf{U}^T\mathbf{R}, \qquad (3)$$

where $\mathbf{D}$ is an eigenvalue matrix and $\mathbf{U}$ an eigenvector matrix.

3. REVIEW ON TENSOR REPRESENTATION AND ON $LRTA_{DR}$-(K_1, K_2, P)

The ICA_{dr} and the PCA_{dr} extracted p components using only spectral information. Recently, [3] overcomes this drawback when PCA_{dr} is considered. This proposed method was referred to as the lower rank-(K_1, K_2, p) tensor approximation DR-based method, $LRTA_{dr}$-(K_1, K_2, p). Based on tensor representation, it keeps the initial spatial structure and insures the neighborhood effects. The whole HSI is considered as a $third$-order tensor, the entries of which are accessed via three indices. It is denoted by $\mathcal{R} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, with elements arranged as $r_{i_1 i_2 i_3}$, $i_1 = 1, \ldots, I_1$; $i_2 = 1, \ldots, I_2$; $i_3 = 1, \ldots, I_3$ and $\mathbb{R}$ is the real manifold. Each index is called mode : two spatial and one spectral modes characterize the HSI tensor. The interest of tensor modelling is the mathematically founded multilinear algebraic tools studying 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. The n-mode flattened matrix $\mathbf{R}_n$ of tensor $\mathcal{R} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ is defined as a matrix from $\mathbb{R}^{I_n \times M_n}$, with : $M_n = I_q I_r$, with $q, r \neq n$. $\mathbf{R}_n$ columns are I_n-dimensional vectors obtained from $\mathcal{R}$ by varying the index i_n and keeping the other indices fixed.

In tensor formulation [7], the obtained matrix $\mathbf{R}$ (section 2) is equivalent to the 3-mode flattened matrix of $\mathcal{R}$. Then, the PCA_{dr}, defined in Eq. (3), is applied on the 3-mode flattened $\mathbf{R}_3$. We write this statement, in tensor formula, as follows :

$$\mathcal{Y}_{PC} = \mathcal{R} \times_3 \mathbf{D}^{-1/2}\mathbf{U}^{(3)T}, \qquad (4)$$

where $\mathcal{Y}_{PC} \in \mathbb{R}^{I_1 \times I_2 \times p}$ is a $three$-order tensor holding the p principal components (PCs). With $\mathbf{U}^{(3)}$ being the orthogonal 3-mode matrix holding the p eigenvectors of the flattened matrix $\mathbf{R}_3$. $\times_n$ is the n-mode product [7, 8] generalizing the product between a tensor and a matrix along an n-mode.

Eq. (4) is equivalent to Eq. (3) including the image reshaping and highlights the spectral or 3-mode processing in the traditional DR method. The proposed tensor-based DR method has two objectives. First, estimate the matrix $\mathbf{U}^{(3)}$ in Eq. (4) (denoted by $\mathbf{U}$ in Eq. (3)) using spatial information. Secondly, make a joint spatial-spectral processing with the aim of whitening the spatial and spectral modes.

Indeed, any *three*-order tensor $\mathcal{R}$ can be decomposed following the Tucker3 tensor decomposition [8] as :

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

where $\mathbf{U}^{(n)}$ is the orthogonal n-mode matrix holding the K_n eigenvectors associated with the K_n largest eigenvalues of the flattened matrix $\mathbf{R}_n$, $\mathcal{C} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$.

The purpose is to find the lower rank-(K_1, K_2, p) tensor $\mathcal{Y}$, with $K_n < I_n$, $n = 1, 2$, and $p < I_3$, which minimizes the following quadratic Frobenius norm :

$$\|\mathcal{R} - \mathcal{Y}\|_F^2 . \qquad (6)$$

The $three$-order tensor $\mathcal{Y}$ holds p principal components (PC) extracted ($p < I_3$) which are spatially approximated. Finding the lower rank approximation consist in estimating the $\mathbf{U}^{(n)}$ orthogonal matrix. And [7] shows that minimizing Eq. (6) with respect to $\mathcal{Y}$ amount to maximize with respect to $\mathbf{U}_n$ matrix the quadratic function :

$$g(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3) = \left\| \mathcal{R} \times_1 \mathbf{U}^{(1)T} \times_2 \mathbf{U}^{(2)T} \times_3 \mathbf{U}^{(3)T} \right\|^2 . \qquad (7)$$

The least square solution involves the $LRTA_{dr}$-(K_1, K_2, p) expression [3] :

$$\mathcal{Y}_{PC} = \mathcal{R} \times_1 \mathbf{P}^{(1)} \times_2 \mathbf{P}^{(2)} \times_3 \mathbf{D}^{-1/2}\mathbf{U}^{(3)T}_{1\cdots p}, \qquad (8)$$

where $\mathbf{P}^{(n)} = \mathbf{U}^{(n)}_{1\cdots K_n} \mathbf{U}^{(n)T}_{1\cdots K_n}$, $n = 1, 2$, and $\mathbf{D}$, $\mathbf{U}^{(3)}_{1\cdots p}$ are the eigenvalue matrix and the corresponding first p eigenvectors of the 3-mode flattened of $\mathcal{R}$ respectively. $\mathbf{U}^{(n)}$ is achieved using an alternating least squares (ALS) algorithm convergence [7] guaranteed the cross-dependency of the spectral and spatial processing. [3] shows the classification efficiency improvement when the $LRTA_{dr}$-(K_1, K_2, p) is used that includes a joint spatial and spectral whitening.

4. THE PROPOSED MULTILINEAR ALGEBRA AND ICA-BASED METHOD

In this paper, we pursue this previous work by adapting it for ICA_{dr} approach. As a result, in tensor formula, Eq. (2) including the image reshaping is equivalent to :

$$\mathcal{Y}_{IC} = \mathcal{R} \times_3 \mathbf{W}, \qquad (9)$$

where, $\mathcal{Y}_{IC}$ is a $three$-order tensor $\in \mathbb{R}^{I_1 \times I_2 \times p}$ holding the p ICs. Considering that the ICs are extracted from the whitened and reduced data, two solutions are developed. The first proposition considers the $LRTA_{dr}$ prior to the ICA_{dr} and the second one considers an hybrid $LRTA$-ICA_{dr}-(K_1, K_2, p) approach to DR.

• $LRTA_{dr}$-(K_1, K_2, p) prior to ICA_{dr}.

We propose to substitute the spectral whitening preprocessing (Eq. 3) by the $LRTA_{dr}$-(K_1, K_2, p) (Eq. 4). As a result, ICA_{dr} is applied on a spatial/spectral whitened data rather than spectral whitened only. With the guarantee that the $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$ and $\mathbf{U}^{(3)}$ orthogonal matrix are cross-dependent using ALS algorithm. We summarize this corresponding algorithm as follows :

1.**Initialization:** $k = 0$, $\forall\ n = 1, 2, 3$
$\mathbf{U}^{(n),0}$contains the first K_n eigenvectors of the flattened matrix $\mathbf{R}_n$.
2. **ALS loop:** until convergence of Eq.(7)
$\forall\ n = 1, 2, 3$
a.$\widehat{\mathcal{R}}^k = \mathcal{R} \times_q \mathbf{U}^{(q),k^T} \times_r \mathbf{U}^{(r),k^T}$ with q,r≠n.
b.$\widehat{\mathbf{R}}_n^k \widehat{\mathbf{R}}_n^{k^T}$ eigenvalue decomposition
c.$\mathbf{U}^{(n),k+1}$ holds the K_n first eigenvectors
d.$\mathbf{P}^{(n),k+1} = \mathbf{U}^{(n),k+1} \mathbf{U}^{(n),k+1^T}$
3. **Lower rank approximation and reduction**
$\mathcal{Y}_{PC} = \mathcal{R} \times_1 \mathbf{P}^{(1),ks} \times_2 \mathbf{P}^{(2),ks} \times_3 \mathbf{D}^{-1/2} \mathbf{U}^{(3),ks^T}$ Eq.(8)
ks : final iteration
4. **ICs extraction**
a.$\mathbf{W}$ estimated using *fastica* from $\widehat{\mathcal{R}}_{ks}$
b.$\mathcal{Y}_{IC} = \widehat{\mathcal{R}}^{ks} \times_3 \mathbf{W}$

But in the latter method, the $\mathbf{W}$ demixing matrix estimation does not take advantage of the spatial information. The second proposition makes jointly the spatial decorrelation and the ICA_{dr}-based spectral reduction.

• Hybrid $LRTA$-ICA_{dr}-(K_1, K_2, p).

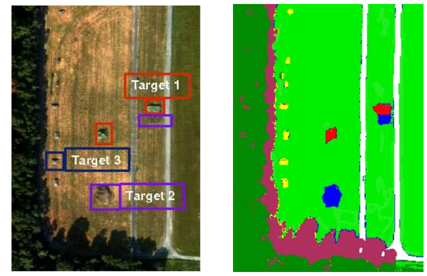
In this proposition, the orthogonal matrix $\mathbf{U}^{(1)}$, $\mathbf{U}^{(2)}$, $\mathbf{U}^{(3)}$ and the demixing matrix $\mathbf{W}$ are jointly estimated and cross-dependent using the ALS algorithm. This statement is written, from Eqs.(8) and (9) as :

$$\mathcal{Y}_{IC} = \mathcal{R} \times_1 \mathbf{P}^{(1)} \times_2 \mathbf{P}^{(2)} \times_3 \mathbf{W}. \tag{10}$$

We summarize this corresponding algorithm as follows :

1.**Initialization:** $k = 0$, $\forall\ n = 1, 2, 3$
$\mathbf{U}^{(n),0}$ holds the first K_n eigenvectors of the flattened matrix $\mathbf{R}_n$.
2.**ALS loop:** until convergence of Eq.(7)
i.*for* $n = 1, 2$
a.$\widehat{\mathcal{R}}^k = \mathcal{R} \times_q \mathbf{U}^{(q),k^T} \times_3 \mathbf{W}^k$ with q≠n.
b.$\widehat{\mathbf{R}}_n^k \widehat{\mathbf{R}}_n^{k^T}$ eigenvalue decomposition
c.$\mathbf{U}^{(n),k+1}$ holds the K_n first eigenvectors
d.$\mathbf{P}^{(n),k+1} = \mathbf{U}^{(n),k+1} \mathbf{U}^{(n),k+1^T}$
ii.*for* $n = 3$
a.$\widehat{\mathcal{R}}^k = \mathcal{R} \times_1 \mathbf{U}^{(1),k^T} \times_2 \mathbf{U}^{(2),k^T}$
b.$\widehat{\mathbf{R}}_n^k \widehat{\mathbf{R}}_n^{k^T}$ eigenvalue decomposition
c.$\widehat{\mathcal{R}}^k = \widehat{\mathcal{R}} \times_3 \mathbf{U}^{(3),k^T}$
d.$\mathbf{W}$ estimated using *fastica* from $\widehat{\mathcal{R}}^k$
3.**hybrid spatial lower rank/ICA reduction**
$\mathcal{Y}_{IC} = \mathcal{R} \times_1 \mathbf{P}^{(1),ks} \times_2 \mathbf{P}^{(2),ks} \times_3 \mathbf{W}^{ks}$ Eq.(10)
ks : final iteration

The next section shows the advantage of using multilinear algebra for DR on the classification result.



a) b)

Fig. 1. a) Classes in the HYDICE image, b) ground truth.

Table 1. Information classes and samples

Classes	Training samples	Test samples	Color
field	1 002	40 811	green 1
trees	1 367	5 537	green 2
road	139	3 226	white
shadow	372	5 036	pink
target 1	128	519	red
target 2	78	285	blue
target 3	37	223	yellow

5. EXPERIMENTS

Real-world data collected by HYDICE imaging are considered for this investigation with a 1.5 m spatial and 10 nm spectral resolution. Including 148 spectral bands (from 435 to 2326 nm), 310 rows and 220 columns, this HSI can be represented as a *three*-order tensor, referred to as $\mathcal{R} \in \mathbb{R}^{310\times220\times148}$. Figure 1 shows the entire scene. The land cover classes are : field, trees, road, shadow and 3 different targets. The resulting number of training and testing pixels for the seven classes are given on Table 1. Classification is performed using the Maximum Likelihood algorithm.

For sake of clarity, the (K_1, K_2)-values are empirically set to 150, for all experiments from this real-world images.

The first experiment compares the classification result obtained after ICA-based DR methods introduced in this paper. Figure 2 shows the visual classification result when ICA_{dr} is used, when $LRTA_{dr}$ is applied prior to ICA_{dr} and when the hybrid $LRTA$-ICA_{dr}-(K_1, K_2, p) is considered. It permits to visually appreciate the ICA-based DR usefulness. The hybrid $LRTA$-ICA_{dr}-(K_1, K_2, p) yields more homogeneous classes with mean area corresponding to the background and the target more identifiable with less unclassified pixels. This tendency is confirmed in Fig. 3 which shows the overall classification evolution with respect to the number of retained ICs. For all p-values our DR method exhibits better classification result, more significantly in comparison to those obtained when $LRTA_{dr}$ is applied prior to ICA_{dr}. The joint estimation of the spatial projectors ($\mathbf{P}^{(n)}$) and the demixing matrix

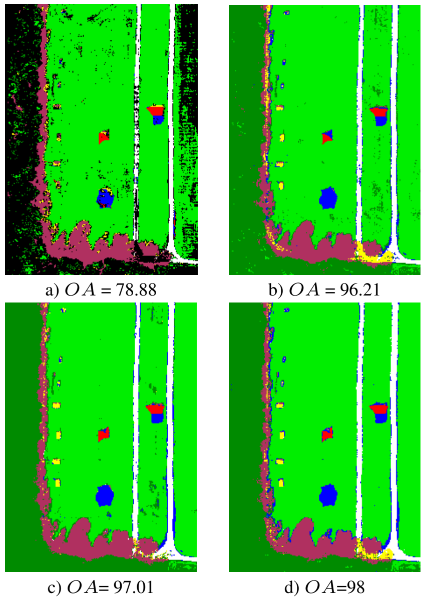


Fig. 2. Classification obtained when a) ICA_{dr} is used, b) $LRTA_{dr}$ is used before ICA_{dr} c) hybrid $LRTA$-ICA_{dr}-$(150, 150, 20)$ is considered.

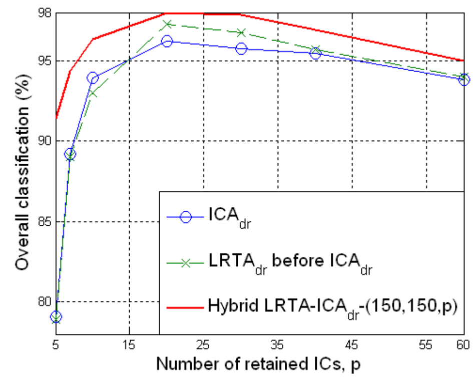


Fig. 3. Overall classification with respect to number of ICs retained : the p-value.

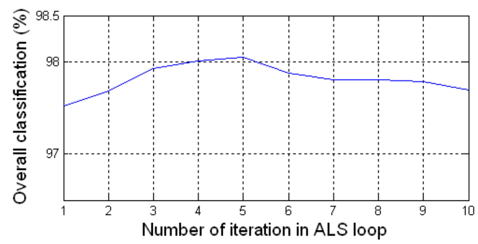


Fig. 4. Overall classification with respect to the number of iteration performed in the ALS loop when the hybrid $LRTA$-ICA_{dr}-$(150, 150, 20)$ algorithm is considered.

($\mathbf{W}$) seems to have interesting impact on classification.

The second experiment illustrates the ALS convergence of the hybrid $LRTA$-ICA_{dr}-$(150, 150, 20)$ algorithm, when 20 ICs are retained. Figure 4 shows that the fitting of the orthogonal matrix $\mathbf{U}^{(n)}$, and the demixing matrix $\mathbf{W}$ iteratively improves the classification efficiency until 5 iterations for this real-world data.

6. CONCLUSION

In this paper, we have introduced a new multilinear algebra and ICA based DR method, referred to as hybrid $LRTA$-ICA_{dr}-(K_1, K_2, p). This method makes jointly an ICA_{dr}-based spectral reduction and a spatial decorrelation. As a result, the p ICs are projected into a lower (K_1, K_2)-dimensional space. The cross-dependence of the spatial projection matrix and the demixing matrix is guaranteed using an alternating least squares algorithm. Experiments from real-world HYDICE show the classification improvement when our proposed method is used.

7. REFERENCES

[1] D. Landgrebe, "Hyperspectral image data analysis as a high dimensional signal processing problem," *Special issue of the IEEE Signal Process. Magazine*, vol. 19, no. 1, pp. 17–28, Jan. 02.

[2] J. Wang and C.I. 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.

[3] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Geoscience and Remote Sensing Letters, accepted for future publication*, vol. 1, no. 1, pp. 1–1, 2008.

[4] P. Comon, "Independent component analysis : A new concept," *Signal Process.*, vol. 36, pp. 287–314, 1994.

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

[6] C.-I. Chang, S.S. Chiang, J.A. Smith, and I.W Ginsberg, "Linear spectrum random mixture analysis for hyperspectral imagery," *IEEE Geoscience and Remote Sensing*, vol. 40, no. 2, pp. 375–392, Feb. 2002.

[7] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-$(r_1, \ldots, r_N)$ approximation of higher-order tensors," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, pp. 1324–42, Apr. 00.

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