CLASSIFICATION IMPROVEMENT BY DIMENSIONALITY REDUCTION BASED ON MULTILINEAR ALGEBRA TOOLS

N. Renard[1], *S. Bourennane*[1] *and J. Blanc-Talon*[2]

[1] Institut Fresnel (CNRS UMR6133), École Centrale de Marseille
DU de Saint Jérôme, F-13013 Marseille Cedex 20, France
phone: +33 4 91 28 82 02, fax: +33 4 91 28 88 13
email: simon.conseil@fresnel.fr, salah.bourennane@fresnel.fr

[2] DGA/D4S/MRIS,
94114 Arcueil, France
phone: +33 1 42 31 92 80 , fax: +33 1 42 31 99 64
email: blanctaj@etca.fr

ABSTRACT

Hyperspectral images (HSI) are multidimensional and multicomponent data with a huge number of spectral bands. To improve classifiers efficiency the principal component analysis (PCA), referred to as PCA_{dr}, the maximum noise fraction (MNF) and more recently the independent component analysis (ICA) are the most commonly used techniques for dimensionality reduction. But to apply those techniques, and in general when dealing with multi-way data, a standard technique consists in vectorizing images provide two-way data. As an alternative, in this paper, we propose to consider HSI as array data or tensor -instead of matrix- which offers multiple ways to decompose data orthogonally. This new method is based on multilinear algebra tools which generalize the PCA to higher order. We show that the result of classification is improved by taking advantage of jointly spatial and spectral information and by performing simultaneously a dimensionality reduction on the spectral way and a projection onto a lower dimensional subspace of the two spatial ways.

1. INTRODUCTION

The emergence of hyperspectral images (HSI) implies the exploration and the collection of a huge amount of data. Hyperspectral imaging sensors provide a huge number of spectral bands, typically up to several hundreds. It is conceded HSI contains many highly correlated bands providing a considerable amount of spectral redundancy. This unreasonably large dimension not only increases computational complexity but also degrades classification accuracy [6]. Indeed, the estimation of statistical properties of classes in a supervised classification process needs the number of training samples to exponentially increase when the number of data dimensions increases. In HSI training data lack, hence the need for feature selection and reduction of data dimensionality, by extracting features from transformed feature. This previous research has demonstrated that high-dimensional data spaces are mostly empty, indicating that the data structure involved exists primarily in a subspace. Dimensionality reduction (DR) extracts features that maximize the separation between the underlying classes and as a result increases classification and detection efficiency.

Due to its simplicity and ease of use, the most popular DR approach is the PCA, referred to as PCA_{dr}, which maximizes the amount of data variance by orthogonal projection. A refinement of PCA_{dr} is the independent component analysis (ICA), referred to as ICA_{dr} [7] which uses higher-order statistic. But the preliminary step to apply those methods is to vectorize the images. Therefore, they rely on spectral properties of the data only, thus neglecting to the spatial rearrangement. To overcome it, [1] proposes a feature extraction method based on multichannel mathematical morphology operator which incorporates the image representation.

In this paper, we propose to use multilinear algebra tools for the DR problem, while considering the data as multi-way data. As was pointed out in [12] the natural representation of a collection of images is a three-dimensional array, or third-order tensor, rather than a matrix of vectorized images. Hence, instead of adapting data to classical matrix-based algebraic techniques (by rearrangement or splitting), the multilinear algebra, the algebra of higher order tensors proposes a powerful mathematical framework for analyzing the multifactor structure of data. Recently used, Tucker3 tensor decomposition has been developed with the aim of generalizing the matrix singular value decomposition (SVD). Tucker3 thus achieves a multimode PCA, also known as higher order SVD (HOSVD) [5] and lower rank-(K_1,K_2,K_3) tensor approximation (LRTA-(K_1,K_2,K_3)) [5]. These multilinear tools have been recently applied in blind source separation to process lower rank approximation of cumulant tensor in ICA, in seismic wave separation, to noise filtering in color images [10] and to faces recognition [14].

Our aim is to adapt the LRTA-(K_1,K_2,K_3) to the DR problem, referred to as LRTA$_{dr_3}$-(K_1,K_2,D_3), to improve classification efficiency in hyperspectral context. It performs jointly a dimensionality reduction of the spectral way (by extracting D_3 uncorrelated spectral components) and transforms the two spatial ways into a lower dimension subspaces equal to K_1 and K_2. As a result, this multimodal DR method takes advantage of spatial and spectral informations.

The remainder of the paper is organized as follows: Section 2 presents the multi-way model and a short overview of its major properties. Section 3 introduces the multimode PCA. Section 4 shows how the LRTA-(K_1,K_2,K_3) is derived to reduce the dimensionality, while reviewing the PCA_{dr} method. Section 5 contains some comparative results of classification performance after dimensionality reduction of hyperspectral images. Finally, concluding remarks are given in Section 6.

2. MULTI-WAY MODELLING AND PROPERTIES

We define a tensor of order 3 as N-way data, the entries of which are accessed via 3 indices. It is denoted by $\mathscr{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, with elements arranged as $x_{i_1 i_2 i_3}$, $i_1 =$

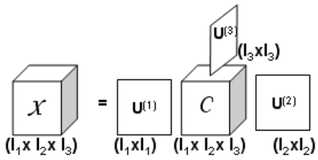


Figure 1: Tucker3 decomposition model.

$1,\ldots,I_1$; $i_2 = 1,\ldots,I_2$; $i_3 = 1,\ldots,I_3$, $\mathbb{R}$ being the real manifold. Each index is called way or mode and the number of levels in the mode is called dimension of that mode. The mode is built on vector space $E^{(n)}$ of dimension I_n, which is the number of data sampled in the physical way associated with mode n. Each way of this multidimensional array is associated with a physical quantity. For instance, in multivariate image analysis, an HSI is a sample of I_3 images of size $I_1 \times I_2$. Each element has three indices and data can be geometrically arranged in a box of dimension $I_1 \times I_2 \times I_3$. HSI data can be modelled as a three-way array: two modes for rows and columns and one mode for spectral channel.

Foremost, let us give a brief review of tensor rank definitions which can be found in [5]. The n-mode rank of tensor data $\mathscr{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, denoted by $Rank_n(\mathscr{X})$, is the dimension of its n-mode vector space $E^{(n)}$ composed of the I_n-dimensional vectors obtained from $\mathscr{X}$ varying index i_n and keeping the other indices fixed. $\mathscr{X}$ is called a rank$-(K_1,K_2,K_3)$ if $Rank_n(\mathscr{X}) = K_n$ whatever $n = 1,2,3$.

This multi-way, or tensor modelling permits to consider multivariate data as inseparable whole data which involves a joint processing on each mode without separability assumption rather than splitting data or processing only the vectorized images. This model naturally implies processing technics based on multilinear algebra. The Tucker3 model [13] is the commonly used tensor decomposition model. This Tucker3 model permits the approximation of a lower rank $-(K_1,K_2,K_3)$ tensor, LRTA-(K_1,K_2,K_3).

3. REVIEW ON LRTA$-(K_1,K_2,K_3)$

The LRTA$-(K_1,K_2,K_3)$ is the high-order generalization of the PCA. In the Tucker3 decomposition model, any 3-way data $\mathscr{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ can be expressed as :

$$\mathscr{X} = \mathscr{C} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)} \tag{1}$$

where $\mathbf{U}^{(n)}$ are orthogonal matrix holding the K_n eigenvectors associated with the K_n largest eigenvalues, $\mathscr{C} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ is the core tensor and $\times_n$ is the n-mode product, properties which can all be found in [5]. When $K_n = I_n$, Tucker3 decomposition is called HOSVD, and when $K_n < I_n$, it is called LRTA-(K_1,K_2,K_3). An example of the Tucker3 three-way decomposition model is illustrated in Fig. 1.

Given real 3-way data $\mathscr{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, the LRTA-$(K_1,K_2,K_3)$ problem consists in finding the lower rank-(K_1,K_2,K_3) multi-way data $\widehat{\mathscr{X}}$, with $K_n < I_n$, $\forall n = 1,2,3$, which minimizes the following quadratic Frobenius norm:

$$\left\| \mathscr{X} - \widehat{\mathscr{X}} \right\|_F^2. \tag{2}$$

Thus the best lower rank-(K_1,K_2,K_3) multi-way approximation of $\mathscr{X}$ is:

$$\widehat{\mathscr{X}} = \mathscr{X} \times_1 \mathbf{P}^{(1)} \times_2 \mathbf{P}^{(2)} \times_3 \mathbf{P}^{(3)}, \tag{3}$$

and:

$$\mathbf{P}^{(n)} = \mathbf{U}^{(n)} \mathbf{U}^{(n)^T}, \tag{4}$$

is the projector on the K_n-dimensional subspace of $E^{(n)}$ which minimizes (2).

In a vector or matrix formulation, the definition of the projector on the signal subspace is based on the eigenvectors associated with the largest eigenvalues of the covariance matrix of the observation vector set. By extension, in the tensor formulation, the projectors on the n-mode vector spaces are estimated by computing the best LRTA-(K_1,K_2,K_3), in the least-squares sense. $\widehat{\mathscr{X}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ is achieved after an alternating least squares (ALS) algorithm convergence. This ALS algorithm can be summarized in the following steps:

1. initialisation $k = 0$: Perform HOSVD [4] to initialize the projectors $\forall n$=1 to 3, $\mathbf{P}_0^{(n)} = \mathbf{U}_0^{(n)} \mathbf{U}_0^{(n)^T}$. $U_0^{(n)}$ contains the K_n eigenvectors associated with the K_n largest eigenvalues of the unfolding matrix $\mathbf{X}_n$ [9].
2. ALS loop: while $\left\| \mathscr{X} - \widehat{\mathscr{X}_k} \right\|_F^2 > 10^{-4}$,
   (a) for n=1 to 3 :
      i. $\widehat{\mathscr{X}_k} = \mathscr{X} \times_q \mathbf{P}_{k+1}^{(q)} \times_r \mathbf{P}_{k+1}^{(r)}$, with $q \neq r \neq n$;
      ii. n-mode unfold $\widehat{\mathscr{X}_k}$ into matrix $\widehat{\mathbf{X}}_{n,k}$
      iii. compute matrix $C_k^{(n)} = \widehat{\mathbf{X}}_{n,k} \mathbf{X}_{n,k}^T$;
      iv. process $\mathbf{C}_k^{(n)}$ SVD, and $\mathbf{U}_{k+1}^{(n)} \in \mathbb{R}^{I_n \times K_n}$ contains the K_n eigenvectors associated with the K_n largest eigenvalues;
      v. compute $\mathbf{P}_{k+1}^{(n)} = \mathbf{U}_{k+1}^{(n)} \mathbf{U}_{k+1}^{(n)^T}$;

   (b) compute $\widehat{\mathscr{X}_{k+1}} = \mathscr{X} \times_1 \mathbf{P}_{k+1}^{(1)} \times_2 \mathbf{P}_{k+1}^{(2)} \times_3 \mathbf{P}_{k+1}^{(3)}$
3. output: $\widehat{\mathscr{X}_{k_{stop}}} = \mathscr{X} \times_1 \mathbf{P}_{k_{stop}}^{(1)} \times_2 \mathbf{P}_{k_{stop}}^{(2)} \times_3 \mathbf{P}_{k_{stop}}^{(3)}$, the best lower rank-$(K_1,K_2,K_3)$ approximation of $\mathscr{X}$.

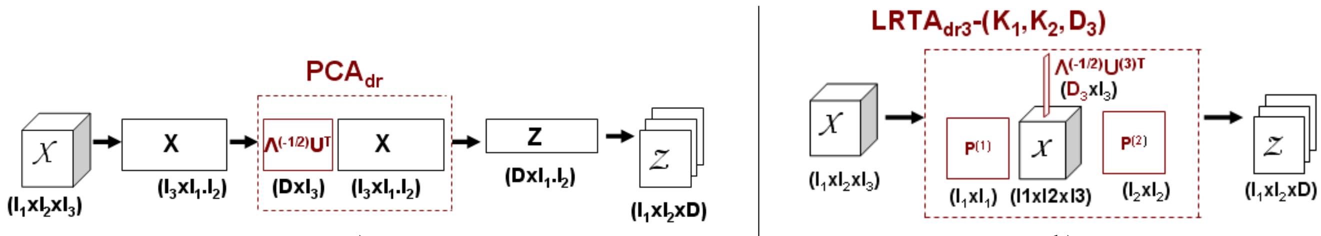
The LRTA-(K_1,K_2,K_3) uses intact multi-way structure to derive jointly the n-mode projectors. Indeed, the LRTA-(K_1,K_2,K_3) takes into account the cross-dependency of information contained in each mode thanks to the ALS algorithm. The next section shows how the LRTA-(K_1,K_2,K_3) can be an interesting tool for hyperspectral images.

4. *LRTA$_{DR_3}$*-(K_1,K_2,D_3), A DIMENSIONALITY REDUCTION TOOL

4.1 Principal component analysis for DR

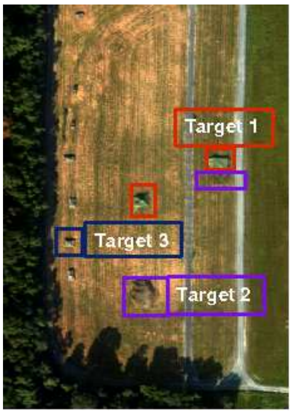
In hyperspectral context, there is a great interest in reducing the spectral ways by selecting the more significants spectral features in order to maximize the separation between classes.

Suppose that we collect I_3 images of full size $I_1 \times I_2$. Each of the I_3 images $\mathbf{X}$ is transformed into a vector $\mathbf{x}^T$ by row concatenation. The tensor $\mathscr{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ becomes a matrix $\mathbf{X} \in \mathbb{R}^{I_3 \times p}$ where $p = I_1 \cdot I_2$. The aim of the DR is to extract a small number D of features with $D < I_3$, called principal component (PC). Each PC is generated by projecting the data spaced onto the nth eigenvector associated with the nth largest eigenvalue. Therefore, the D PCs generate a reducing matrix $\mathbf{Z} \in \mathbb{R}^{D \times p}$. Figure 2 a) illustrates the PCA_{dr} strategy, and the processus to define the PCs is :



a) b)

Figure 2: Dimensionality reduction strategy : a) PCA_{dr}. b) $LRTA_{dr_3}$-(K_1, K_2, D_3)



a)

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

b)

Figure 3: Classes in the HYDICE image RGB (a), information classes and samples (b).

1. Perform the PCA on $\mathbf{X}$ to find eigenvalues λ_i and their corresponding eigenvectors $\mathbf{u_i}$, for $i = 1, \ldots, I_3$.
2. Define the eigenvalue diagonal matrix $\mathbf{\Lambda} \in \mathbb{R}^{D \times D}$ holding the D largest λ_i, for $i = 1, \ldots, D$ and their associated eigenvectors in the matrix $\mathbf{U} \in \mathbb{R}^{p \times D}$.
3. Sphere the $\mathbf{X}$ matrix : $\mathbf{Z} = \mathbf{\Lambda}^{-1/2}\mathbf{U}^T\mathbf{X}$.

The data can be reshape to an multivariate images $\mathscr{Z} \in \mathbb{R}^{I_1 \times I_2 \times D}$.

4.2 Multimodal DR

By the way described above, we can turn the well-known LRTA-(K_1, K_2, K_3) into a spectral dimensionality reduction tool, referred to as $LRTA_{dr_3}$-(K_1, K_2, D_3). It extracts D_3 spectral PCs in order to derive the tensor $\mathscr{Z} \in \mathbb{R}^{I_1 \times I_2 \times D_3}$. The challenge carried out thanks to the $LRTA_{dr_3}$-(K_1, K_2, D_3) is to jointly reduce the dimensionality of the spectral mode and to transform the spatial modes into a lower dimensional subspace, different number of components -(K_1, K_2, D_3) can be retained for each mode. The $LRTA_{dr_3}$-(K_1, K_2, D_3) model can be written:

$$\mathscr{Z} = \mathscr{X} \times_1 \mathbf{P}^{(1)} \times_2 \mathbf{P}^{(2)} \times_3 \mathbf{\Lambda}^{-1/2}\mathbf{U}^{(3)^T}, \qquad (5)$$

Where $\mathbf{U}$ is the matrix holding the D_3 eigenvectors associated with the D_3 largest eigenvalues, $\mathbf{\Lambda}$ is the diagonal eigenvalues matrix holding the D_3 largest eigenvalues and $\mathbf{P}^n$ are the n-mode projectors defined in the section 3.

Figure 2 b) illustrates the $LRTA_{dr_3}$-(K_1, K_2, D_3) scheme. The major $LRTA_{dr_3}$-(K_1, K_2, D_3) attribute in relation to the PCA_{dr} is the use of the spatial information in order to select the PCs. Indeed, thanks to the ALS loop, the spectral features are estimated iteratively like the spatial n-mode projectors. They are *optimal* in the sense of the mean square error.

To estimate the D_3-dimension [2] introduces some criteria which determine the virtual dimensionality defining the minimum number of spectrally distinct signal sources that characterize the hyperspectral data. Concerning the (K_1, K_2)-dimensional subspace, [11] proposes to extend the Akaike information criterion (AIC) to estimate the signal subspace in the case of Gaussian additive noise. In this paper, we focus on introducing multimodal tools in hyperspectral context and all (K_1, K_2, D_3)-dimensions are fixed empirically.

5. RESULTS

The data used in the experiments are real data collected by HYDICE imaging, with a 1.5 m spatial and 10 nm spectral resolution. The full scene consists of 310 lines and 220 samples with 148 spectral bands. The absorption bands have been preliminary removed. This HSI can be represented by a 3-order tensor, noted by $\mathscr{X} \in \mathbb{R}^{310 \times 220 \times 148}$.

Figure 3 a) shows the entire scene used for experiments. The land cover classes are : field, trees, road, shadow and 3 different targets. The resulting number of training and testing pixels for each class are given in Fig. 3 b). For convenience, a preprocessing remove the mean of each vector pixels of the initial multi-way data $\mathscr{X}$.

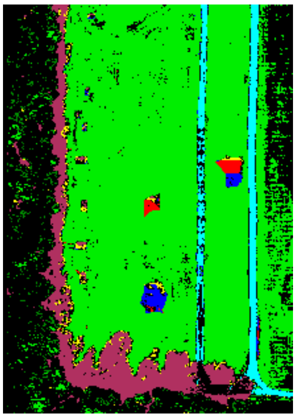
To highlight the advantage of multimodal DR method we compare the classification result after applying the $LRTA_{dr_3}$-(K_1, K_2, D_3) and the PCA_{dr}. The classification is performed thanks to a well-known and largely used algorithm in hyperspectral, the spectral angle mapper (SAM) [8] algorithm. Figure 4 shows a visual classification result obtained from the original tensor $\mathscr{X}$ and after the DR methods which select 10 spectral features and where the spatial dimensional subspaces have been fixed to 40 for the $LRTA_{dr_3}$-$(40, 40, 10)$.

Visually 4 a) permits to appreciate the DR usefulness, all black pixels in the classification result represent the unclassified pixels. In comparison with PCA_{dr}, the $LRTA_{dr_3}$-$(40, 40, 10)$ provides classes which are more homogeneous and the means area corresponding to the background and the target more identifiable with less unclassified pixels.

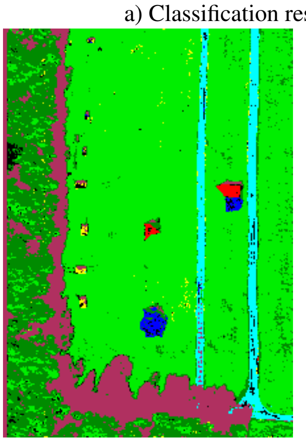
Table 1: Overall (OA) and individual test accuracies in percentage obtained after applying the PCA_{dr}-(D_3) and the $LRTA_{dr}$-(K_1, K_2, D_3).

Class	Initial Image	D_3 =5 bands PCA$_{dr}$	D_3 =5 bands LRTA$_{dr}$ K$_1$ K$_2$ 310 220	D_3 =5 bands LRTA$_{dr}$ K$_1$ K$_2$ 60 60	D_3 =5 bands LRTA$_{dr}$ K$_1$ K$_2$ 40 40	D_3 =10 bands PCA$_{dr}$	D_3 =10 bands LRTA$_{dr}$ K$_1$ K$_2$ 310 220	D_3 =10 bands LRTA$_{dr}$ K$_1$ K$_2$ 60 60	D_3 =10 bands LRTA$_{dr}$ K$_1$ K$_2$ 40 40
field	88.9	93.2	88.3	95.4	97.7	97.5	98.3	99.7	100
forest	4.4	41.0	54.3	62.3	72.9	64.1	69.5	75.8	82.5
road	85.3	98.7	83.3	94.7	95.5	89.6	95.6	91.1	97.7
shadow	80.1	95.1	95.7	97.4	96.8	93.8	96.1	93.5	95.7
target 1	64.9	67.0	54.9	72.7	76.7	63.6	57.6	79.6	81.3
target 2	80.7	77.9	75.4	67.7	66.3	68.4	73.7	83.2	84.2
target 3	31.5	39.6	44.9	65.8	78.0	42.8	38.5	63.6	51.3
OA	78.98	87.96	84.76	91.83	**94.61**	92.73	95.36	95.83	**97.32**

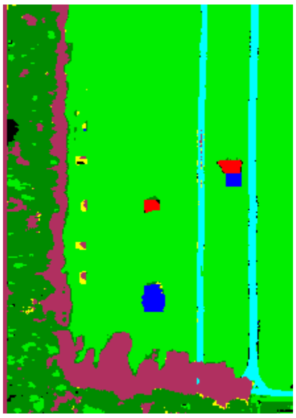
Class		D_3 =20 bands PCA$_{dr}$	D_3 =20 bands LRTA$_{dr}$ K$_1$ K$_2$ 310 220	D_3 =20 bands LRTA$_{dr}$ K$_1$ K$_2$ 60 60	D_3 =20 bands LRTA$_{dr}$ K$_1$ K$_2$ 40 40	D_3 =30 bands PCA$_{dr}$	D_3 =30 bands LRTA$_{dr}$ K$_1$ K$_2$ 310 220	D_3 =30 bands LRTA$_{dr}$ K$_1$ K$_2$ 60 60	D_3 =30 bands LRTA$_{dr}$ K$_1$ K$_2$ 40 40
field		98.4	99.1	99.9	99.8	98.6	99.2	99.9	99.9
forest		81.6	74.3	77.2	79.6	83.0	72.4	76.6	81.0
road		90.8	96.2	96.9	98.0	91.9	96.5	98.3	97.9
shadow		93.9	96.2	95.9	95.8	93.6	96.1	96.1	95.4
target 1		56.8	76.1	83.0	80.5	50.7	77.6	81.3	80.9
target 2		59.3	75.4	83.5	85.9	57.5	84.6	83.9	85.9
target 3		45.9	38.5	45.4	43.8	47.0	41.7	48.1	53.5
OA		95.15	95.64	96.75	**96.90**	95.43	95.61	96.78	**97.15**



a) Classification result from initial data.



b) Classification result from $\mathscr{X} \in \mathbb{R}^{310\times 220\times 10}$ extracted thanks to PCA_{dr}



c) Classification result from $\mathscr{X} \in \mathbb{R}^{310\times 220\times 10}$ extracted thanks to $LRTA_{dr_3}$-$(40, 40, 10)$.

Figure 4: Dimensionality reduction outcome for classification, 10 spectral features are extracted.

To appreciate quantifiable comparisons between the two DR methods, Table 1 gives overall (OA) and individual test accuracies in percentage exhibited by SAM classifier. OA is defined as follows :

$$OA = \frac{1}{M}\sum_{i=1}^{i=P} a_{ii}, \tag{6}$$

where M is the total number of samples, P is the number of classes C_i for $i = 1, \ldots,$ P and a_{ij} is the number of test samples that actually belong to class C_i and are classified into C_j for $i, j = 1, \ldots,$ P.

The classification results are evaluated for several numbers of spectral features retained, and for each case we test empirically several spatial (K_1, K_2)-dimensional subspaces for the $LRTA_{dr_3}$-(K_1, K_2, D_3). Table 1 shows that beyond 20 bands the PCA_{dr} and the $LRTA_{dr_3}$-(K_1, K_2, D_3) have more and more related OA and they do not add more classification improvement. The other important remark report on the Table 1, is that the difference between the classification efficiency obtained from the $LRTA_{dr_3}$-(K_1, K_2, D_3) and the PCA_{dr} is all the more significant as the values of the (K_1, K_2)-dimensional subspaces decrease. It is revealed that the $LRTA_{dr_3}$-(K_1, K_2, D_3) permits better classification efficiency by jointly selecting only 10 spectral features and reducing the spatial dimensional subspaces to 40. It is conceded that the number of spectral features retained have an impact on the classification efficiency. The results obtained show that the spatial subspaces dimensions are also important.

6. CONCLUSION

An multivariate data analysis tool referred to as $LRTA_{dr_3}$-(K_1, K_2, D_3) has been proposed. This multimodal dimensionality reduction tool takes into account the spatial and spectral information to select *optimal* spectral features in the sense of the mean square error. It reveals to be quite interesting for classification efficiency of high-dimensional hyperspectral data. Indeed the classification result depends not only on the number of extracted spectral features but also on the dimension of spatial subspaces. Those promising results encourage us to integrate tensorial approach in the ICA_{dr} method with the same proposed strategy. This further work could overcome a major issue for PCA_{dr} [3] which is that many subtle materials or rare targets require higher order statistics to be characterized.

REFERENCES

[1] Plaza A., Martinez P., Plaza J., and Perez R. Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations. *IEEE Transactions on Geoscience and remote sensing*, 43(3):466–479, march 2005.

[2] C.I. Chang and Q. Du. Estimation of number of spectrally distinct signal sources in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 42(2):608–619, March 2004.

[3] A. Cheriyadat, A. Bruce, and L.M. Bruce. Why principal component is not an appropriate feature extraction method for hyperspectral data. *Proceeding of the 2003 IEEE Geoscience and Remote Sensing Symposium*, 6(2):3420–3422, 21-25 July 2003.

[4] L. De Lathauwer, B. De Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, 21:1253–1278, April 2000.

[5] 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*, 21:1324–1342, April 2000.

[6] D. Landgrebe. Hyperspectral image data analysis as a high dimensional signal processing problem. *Special issue of theIEEE Signal Processing Magazine*, 19(1):17–28, January 2002.

[7] D. Lennon, G Mercier, Mouchot M.C., and Hubert-Moy L. Independant component analysis as a tool for the dimension reduction and the representation of hyperspectral images. *In Spie Remote Sensing*, 4541:2893–2895, 2001.

[8] D. Manolakis and G. Shaw. Detection algorithms for hyperspectral imaging applications. *IEEE Signal Processing Magazine*, 19:29–43, 2002.

[9] D. Muti and S. Bourennane. Fast optimal lower-rank tensor approximation. In *IEEE ISSPIT*, pages 621–625, Marrakesh, Morocco, December 18-21, 2002.

[10] D. Muti and S. Bourennane. Survey on tensor signal algebraic filtering. *Signal Proceesing Journal, Elsevier.*, 87:237–249, Feb. 2007.

[11] N. Renard, S. Bourennane, and J. Blanc-Talon. Multiway filtering applied on hyperspectral images. In *Lecture notes in Journal Computer Science*, volume 4179, pages 127–137, September 2006.

[12] A. Shashua and A Levin. Linear images coding for regression and classification using the tensor-rank principle. *In Spie Remote Sensing*, 4541:2893–2895, 2001.

[13] L.R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31:279–311, 1966.

[14] M. A. O. Vasilescu and D. Terzopoulos. Multilinear image analysis for facial recognition. In *IEEE Int. Conf. on Pattern Recognition (ICPR2002)*, volume 2, Quebec city, Canada, 11-15, August 2002.