Parameters Selection Of Morphological Scale-Space Decomposition For Hyperspectral Images Using Tensor Modeling

Santiago Velasco-Forero, Jesus Angulo

Parameters selection of morphological scale-space decomposition for hyperspectral images using tensor modeling

Santiago Velasco-Forero and Jesús Angulo

Mines ParisTech, Center de Morphologie Mathématique, Fontainebleau, France

ABSTRACT

Dimensionality reduction (DR) using tensor structures in morphological scale-space decomposition (MSSD) for HSI has been investigated in order to incorporate spatial information in DR. We present results of a comprehensive investigation of two issues underlying DR in MSSD. Firstly, information contained in MSSD is reduced using HOSVD but its nonconvex formulation implicates that in some cases a large number of local solutions can be found. For all experiments, HOSVD always reach an unique global solution in the parameter region suitable to practical applications. Secondly, scale parameters in MSSD are presented in relation to connected components size and the influence of scale parameters in DR and subsequent classification is studied.

Keywords: U nsupervised feature extraction, mathematical morphology, tensor analysis, dimensional reduction, classification, principal component, hyperspectral imagery.

1. INTRODUCTION

Hyperspectral images (HSI) are acquired by a sensor that captures radiant flux data in typically hundreds of contiguous bands, producing a signature for each pixel in the image. It is accepted that HSI includes many highly correlated bands presenting considerable amounts of spectral redundancy. Consequently, reducing the dimensionality of hyperspectral data without losing important information is one of the main subject of interest for the remote sensing community. The classification in HSI has the difficult that pixels are located in a high dimensional space increasing computational complexity and degrading classification accuracy.[1] The difficulties in HSI using conventional supervised classification methods originally developed for low-dimensional multispectral has been illustrated by Jimenez et al.[2] Additionally, identification of relatively small objects incorporate issues because high spatial resolution is necessary for accurate classification. Due to its simplicity, principal component analysis (PCA) is the most popular approach to dimensional reduction in HSI using singular value decomposition (SVD) over zero mean covariance matrix. Unfortunately, PCA requires a preliminary step to vectorize the images retaining spectral information and neglecting the spatial information presented in the original array. Recently, several authors have pointed out that the natural representation for hyperspectral images is a three dimensional array instead of a matrix of vectorized images.[3] Three dimensional array or third-order tensor preserves usual image representation and band continuity is represented as the third tensor dimension. Alternatively, some authors incorporate morphological profile in HSI analysis because it considers spatial relationship among pixels producing results that preserve original contours.[4–6] The approach in[5] is based on using some principal components (PCs) from HSI and from each PC, a morphological profile is built. More recently,[6] extends from[5] using data fusion between morphological profile on PCs and the original data. They presented excellent accuracies and improvements compared to those obtained with pixel-based classifiers and supervised feature selection classifiers.

More recently, it was presented in[16] a nonlinear scale-space representation (NSSR) based on morphological levelings, which include openings and closings by reconstruction. NSSR creates a fourth dimension in HSI which is analyzed as a fourth tensor dimension. The analysis of spatial and spectral patterns simultaneously utilizes high-order or multilinear singular value decomposition (HO-SVD) proposed for tensors[7] and the classification is performed using Support Vector Machines (SVM). That shows good results in the situation of small size training

set.[16] In this paper parameter selection for NSSR is explored taking in consideration theoretical issues and classification performance for real HSI.

The paper is organized as follows: Section 2 introduces tensor notation and basic concepts being used later on. Section 3 presents nonlinear scale-space representation with morphological levelings. Section 4 proposes a formulation of classification for reduced tensors using Support Vector Machines (SVM). Section 5 shows the effectiveness of the modified approach via practical examples with comparison with standard approach. Analysis of scale parameters and dimension of spatial filtering are also presented. Section 6 concludes the paper.

2. TENSOR MODELING

2.1 Notation

A HSI has a natural representation as a tensor, denoted by $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$. Each index is called *mode*: the first two are spatial and the third is spectral. Thus, $\mathcal{X}$ is a sample of I_3 images of size $I_1 \times I_2$. Our approach is based on applying multilinear algebra on the whole tensor structure instead of adapting the data tensor to classical matrix-based algebraic techniques by rearrangement.

We introduce the notation and basic definitions of multilinear algebra. Scalars are denoted by lower case letters $(a, b, \ldots)$, vectors by bold lower case letters $(\mathbf{a}, \mathbf{b}, \ldots)$, matrices by bold upper-case letters $(\mathbf{X}, \mathbf{Y}, \ldots)$, and higher-order tensors by calligraphic upper-case letters $(\mathcal{X}, \mathcal{Y}, \ldots)$. The order of tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \ldots \times I_N}$ is N. An element of $\mathcal{X}$ is denoted as $\mathcal{X}_{i_1 \ldots i_n \ldots i_N}$.

2.2 Tensor Decomposition

A matrix $\mathbf{X} \in \mathbb{R}^{I_1 \times I_2}$ is a two-mode mathematical object that has two associated vector spaces, a row space and a column space. SVD orthogonalizes these two spaces and decomposes the matrix as $\mathbf{X} = \mathbf{U}_1 \mathbf{\Sigma} \mathbf{U}_2^T$, where $\mathbf{U}_1$ and $\mathbf{U}_2^T$ represent orthogonal column space and $\mathbf{\Sigma}$ is a diagonal singular value matrix. In terms of the n-mode products, this SVD decomposition can be rewritten as $\mathbf{X} = \mathbf{\Sigma} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2$, where $\times_n$ is the n-mode product[7].[9] Extension to an N-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3 \times \ldots \times I_N}$ were presented by[7] orthogonalizing N spaces and expresses the tensor as the n-mode product of N-orthogonal spaces

$$\mathcal{X} = \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \ldots \times_N \mathbf{U}_N \tag{1}$$

Tensor $\mathcal{C}$, known as the core tensor, is analogous to the diagonal singular value matrix in conventional matrix SVD. It is important to realize, however, that the core tensor does not have a diagonal structure; rather, $\mathcal{C}$ is in general a full tensor. The core tensor governs the interaction between the mode matrices $\mathbf{U}_n$, for $n = 1, \ldots, N$. Mode matrix $\mathbf{U}_n$ contains the orthonormal vectors spanning the column space of the matrix $\mathbf{X}_n$ that results from the mode-n flattening of $\mathcal{X}$. This method is know as higher-order SVD (HOSVD) from the work of De Lathauwer, De Moor, and Vandewalle,[7] who showed that the HOSVD is a convincing generalization of the matrix SVD and discussed ways to more efficiently compute the leading left singular vectors of $\mathbf{X}_n$. An excellent compendium about tensor decomposition is presented in.[9] The HOSVD is usually performed using Alternative Least Square algorithm used to jointly find n-mode matrices $\mathbf{U}_n$, but recently other approaches has been presented.[18] In the case of three mode tensors $\mathcal{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_{I_3}]$, the objective of HOSVD is to select subspace $\mathbf{U}_1, \mathbf{U}_2$ and $\mathbf{U}_3$ and the core tensor $\mathcal{C}$ such that the L_2-norm reconstruction error is minimized,[9]

$$\min_{\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \mathcal{C}} E_1 = ||\mathcal{X} - \mathcal{C} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3||^2 \tag{2}$$

where $\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3$ are required to be orthogonal, i.e., $\mathbf{U}_1^T \mathbf{U}_1 = \mathbf{U}_2^T \mathbf{U}_2 = \mathbf{U}_3^T \mathbf{U}_3$. With the orthonormality condition, we can obtain $\mathcal{S} = \mathcal{X} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \times_3 \mathbf{U}_3^T$, and 2 can be written as:

$$\min_{\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3} E_1 \quad = \quad ||\mathcal{X}||^2 - ||\mathcal{S}||^2 \tag{3}$$

$$\Leftrightarrow \max_{\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3} E_2 \quad = \quad ||\mathcal{S}||_F^2 \tag{4}$$

As it was presented by Huang,[14] the equation 3 is equivalent to maximize:

$$\max_{\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3} E_2 = \text{Trace}(\mathbf{U}_1^T \mathbf{F} \mathbf{U}_1) = \text{Trace}(\mathbf{U}_2^T \mathbf{G} \mathbf{U}_2) = \text{Trace}(\mathbf{U}_3^T \mathbf{H} \mathbf{U}_3) \tag{5}$$

where:

$$
\begin{aligned}
\mathbf{F}_{ii'} &= \sum_{ll'} (\mathbf{X}_{(l)} \mathbf{U}_2 \mathbf{U}_2^T \mathbf{X}_{(l')}^T)_{ii'} (\mathbf{U}_3 \mathbf{U}_3^T)_{ll'} \\
\mathbf{G}_{jj'} &= \sum_{ll'} (\mathbf{X}_{(l)} \mathbf{U}_1 \mathbf{U}_1^T \mathbf{X}_{(l')}^T)_{jj'} (\mathbf{U}_3 \mathbf{U}_3^T)_{ll'} \\
\mathbf{H}_{ll'} &= \sum_{ii'jj'} \mathcal{X}_{ijl} \mathcal{X}_{i'j'l'} (\mathbf{U}_1 \mathbf{U}_1^T)_{ii'} (\mathbf{U}_2 \mathbf{U}_2^T)_{jj'}
\end{aligned}
$$

Since $\mathbf{F}, \mathbf{G}, \mathbf{H}$ is semipositive definite, $||\mathcal{S}^2||$ are monotonically increasing, therefore HOSVD algorithm converges to a local optimal. Thus theoretically, the solutions HOSVD are not unique. That issue was already pointed in[9] and studied in detail for[15] in real life databases concluding that the convergence depends on the eigenvalue distribution for the matrix $\mathbf{F}, \mathbf{G}$ and $\mathbf{H}$. In section 5.1., we include an analysis of the convergence problem in real HSIs.

2.3 Tensor Principal Component Analysis

In high-dimensional problem as HSI is a great interest to reduce the spectral dimension in order to exceed problems as "Curse of Dimensionality" in distance-based or nonparametric analysis and "Hughes phenomenon" in linear classifiers. Commonly a pre-processing step consist in performing a PCA to reduce the dimensional space. We present a tensor version for PCA based on 1. Thus, the best lower rank tensor approximation of $\mathcal{X}$,[7] denoted by $\tilde{\mathcal{X}}$ is:

$$\tilde{\mathcal{X}} = \mathcal{X} \times_1 \mathbf{P}_1 \times_2 \mathbf{P}_2 \times_3 \ldots \times_N \mathbf{P}_N \tag{6}$$

where $\mathbf{P}_n = \mathbf{U}_n \mathbf{U}_n^T$, and $\mathbf{U}_n$ is found using expression 1. Our motivation is to reduce the high dimension correspondent with some mode-n product, i.e. yield a projection $\mathbb{R}^{I_1 \times I_2 \ldots \times I_N} \to \mathbb{R}^{I_1 \times I_2 \ldots \times k}$. Thus, we present the *Tensor PCA* as follows:

$$\widehat{\mathcal{X}}_{I_1, I_2, \ldots, k} = \mathcal{X} \times_1 \mathbf{P}_1 \times_2 \mathbf{P}_2 \times_3 \ldots \times_N \widehat{\mathbf{U}}_N^T \tag{7}$$

where $\widehat{\mathbf{U}}_N^T$ contains the k eigenvectors associated with the k largest eigenvalues holding of the unfolding matrix $\mathbf{X}_N$.[7] Tensor PCA was previously presented for HSIs[3] including transformation for sphericity. This representation allows to include spatial filtering per order if the largest eigenvector are considered in the dimensions 1 (rows) and 2 (columns), i.e.,

$$\widehat{\mathcal{X}}_{I_1, I_2, \ldots, k} = \mathcal{X} \times_1 \widetilde{\mathbf{U}}_1 \widetilde{\mathbf{U}}_1^T \times_2 \widetilde{\mathbf{U}}_2 \widetilde{\mathbf{U}}_2^T \times_3 \ldots \times_N \widehat{\mathbf{U}}_N^T \tag{8}$$

Remark that we are using different notation for spatial filtering in the order 1 and 2, and dimensional reduction for the others orders. In the case of order-3 tensors, $\widehat{\mathcal{X}}_{I_1, I_2, k} = \mathcal{X} \times_1 \widetilde{\mathbf{U}}_1 \widetilde{\mathbf{U}}_1^T \times_2 \widetilde{\mathbf{U}}_2 \widetilde{\mathbf{U}}_2^T \times_3 \widehat{\mathbf{U}}_3^T$, where $\widetilde{\mathbf{U}}_1$ and $\widetilde{\mathbf{U}}_2$ has the n_1 and n_2 largest eigenvectors associated of the unfolding matrix $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively. In this representation is clear the equivalence with standard SVD if $n_1 = I_1$ and $n_2 = I_2$, due to $\widetilde{\mathbf{U}}_1 \widetilde{\mathbf{U}}_1^T = \mathbf{I}_{I_1 \times I_1}$ and $\widetilde{\mathbf{U}}_2 \widetilde{\mathbf{U}}_2^T = \mathbf{I}_{I_2 \times I_2}$ in that case. In the section 5.1., the influence of n_1 and n_2 in the tensor solution is presented and we propose an heuristic to select these parameters based on cumulative eigenvalues distribution.

3. NONLINEAR SCALE-SPACE REPRESENTATION WITH MORPHOLOGICAL LEVELINGS (NSS-ML)

Mainly motivated by the need of taking into account the spatial information in DR approaches, different authors[4, 16] have proposed to use morphological leveling per bands in the analysis of HSI. NSS-ML generates a multiscale representation where transformation from finer scale to a coarse scale is defined using morphological levelings.[10] The application of leveling for multiscale imaging structure decomposition in remote sensing has

been initially studied[4] using thresholds to segment the image in convex/concaves classes for the different scales. Recently, NSS-ML was formulated to include an additional order in the tensor decomposition based on different scales affecting the marker used in the leveling decomposition per band.[16] Figure 1 illustrates the idea of how to extend the three-order tensor to an fourth-order tensor using an exact decomposition extracting spatial information by scale. As levelings decompose simultaneously the bright/dark image structure preserving contours, the NSS-ML contains spatial object in various scales.
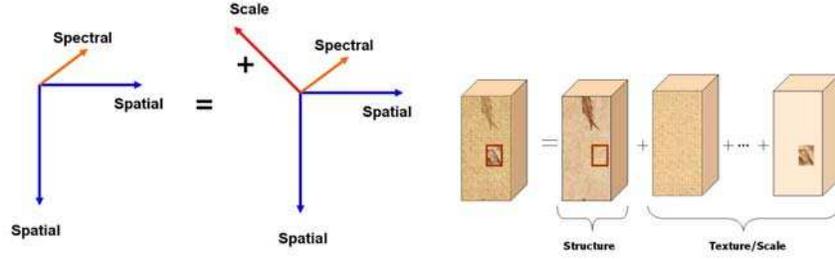

Figure 1. Original three-order tensor is decomposed in structure+texture tensor without loss of information. It is represented by a fourth-order tensor.

3.1 Morphological Levelings

Given a grey scale image $\mathbf{X} = f(x, y)$ and a marker image $\mathbf{M} = m(x, y)$, where the marker is typically a rough simplification of reference image which determines the structures to be reconstructed and those to be simplified during the geodesic erosion/dilation procedure. The leveling $\lambda(f, m)$ can be obtained by the following iterative algorithm:

$$\lambda(f, m) = \lambda^i(f, m) = [f \wedge \delta_f^i(m)] \vee \varepsilon_f^i(m)$$

such that $\lambda^i(f, m) = \lambda^{i+1}(f, m)$ (convergence until idempotency).

The geodesic dilation of size i is given by $\delta_f^i(m) = \delta_f^1 \delta_f^{i-1}(m)$ where the unitary conditional dilation is $\delta_f^1(m) = \delta_B(m) \wedge f$ with $\delta_B(m)$ an unitary dilation. The geodesic erosion of size i, $\varepsilon_f^i(m)$ can be obtained by the property of duality by complement $\varepsilon_f^i(m) = [\delta_{f^c}^i(m^c)]^c$, where f^c is the negative of image f.
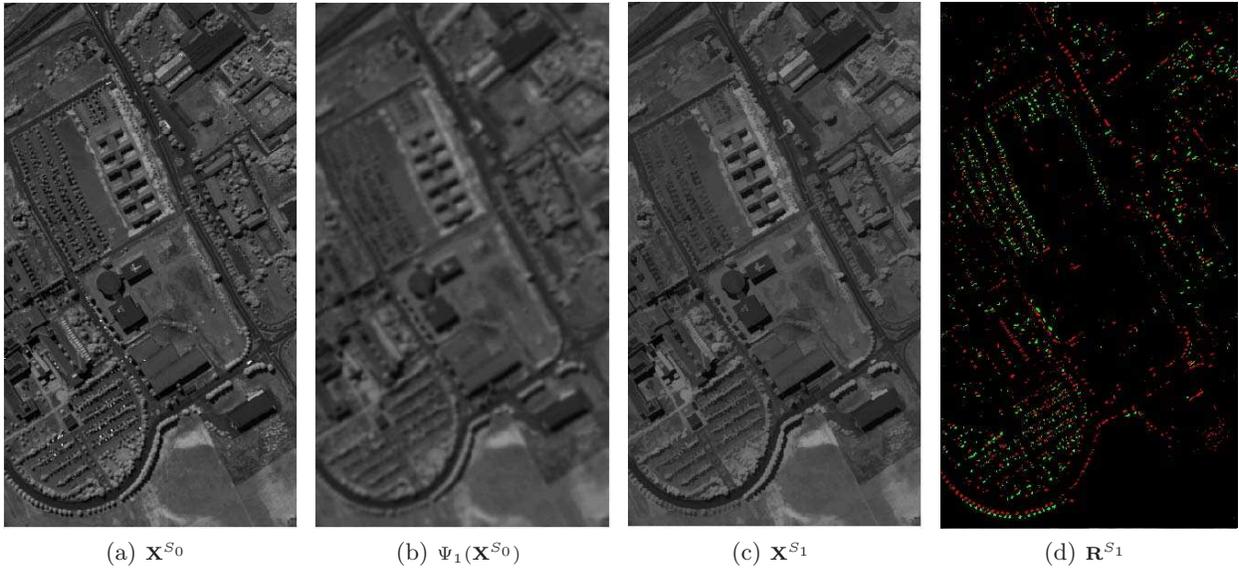

(a) $\mathbf{X}^{S_0}$ (b) $\Psi_1(\mathbf{X}^{S_0})$ (c) $\mathbf{X}^{S_1}$ (d) $\mathbf{R}^{S_1}$

Figure 2. Morphological decomposition steps using $\boldsymbol{\sigma} = (\sigma_1) = 3$

3.2 Morphological Multiscale Decomposition

The idea is to start from a family of transformations indexed by a scale parameter, i.e.,$\{\Psi_n(\mathbf{X})\}_{n=0}^N$, such that $\Psi_0(\mathbf{X}) = \mathbf{X}$. In this study, we use typically a family of Gaussian filters, $\Psi_n(\mathbf{X}) = G_{\sigma_n} * \mathbf{X}$ such that $\sigma_n > \sigma_{n-1}$ for all n.

Now, the schema of image structure decomposition was defined in[16] as follow,

.	Image $\equiv$ Structure	Residual $\equiv$ Texture
Scale 0	$\mathbf{X}^{S_0} = \mathbf{X}$	$\mathbf{R}^{S_0} = 0$
Scale 1	$\mathbf{X}^{S_1} = \lambda(\mathbf{X}^{S_0}; \Psi_1(\mathbf{X}^{S_0}))$	$\mathbf{R}^{S_1} = \mathbf{X}^{S_0} - \mathbf{X}^{S_1}$
Scale 2	$\mathbf{X}^{S_2} = \lambda(\mathbf{X}^{S_1}; \Psi_2(\mathbf{X}^{S_1}))$	$\mathbf{R}^{S_2} = \mathbf{X}^{S_1} - \mathbf{X}^{S_2}$
$\vdots$	$\vdots$	$\vdots$
Scale N	$\mathbf{X}^{S_N} = \lambda(\mathbf{X}^{S_{N-1}}; \Psi_N(\mathbf{X}^{S_{N-1}}))$	$\mathbf{R}^{S_N} = \mathbf{X}^{S_{N-1}} - \mathbf{X}^{S_N}$

Therefore, morphological multiscale decomposition of a grey scale image $\mathbf{X}$ as a tensor of order 3 denoted by $\mathcal{X}_{ijk}^{\boldsymbol{\sigma}}$ is defined by:

$$M(\mathbf{X}_{ij}, \boldsymbol{\sigma}) = \mathcal{X}_{ijn}^{\boldsymbol{\sigma}} = \left[\left\{ \mathbf{R}_{ij}^{S_n} \right\}_{n=1}^N, \mathbf{X}_{ij}^{S_N} \right] = \mathcal{X}^{\boldsymbol{\sigma}} \tag{9}$$

Additionally, $\mathcal{X}_{ijn}^{\boldsymbol{\sigma}}$ has the following structure+texture property:

$$\bigoplus_{n=1}^{N+1} \mathcal{X}_{ijn}^{\boldsymbol{\sigma}} = \bigoplus_{n=1}^{N} \mathbf{R}_{ij}^{S_n} + \mathbf{X}_{ij}^{S_N} = \mathbf{X}_{ij} = \mathbf{X}$$

where $\bigoplus$ represents the direct pixel sum in the indicated index. To illustrate the NSSR based on morphological leveling, we utilize "Pavia university" hyperspectral image. The data used were collected using the optical sensor ROSIS 03 on the campus at the University of Pavia, Italy. Originally, there were 115 bands of ROSIS03 sensor covering from 0.43 to 0.86 μm, and the image size is 610×340. The original hyperspectral image showed using bands 80,70,60 is given in figure 3(a). The structure tensor of $\mathcal{X}$ is in $\mathcal{X}^{S_N}$ as is showed in 3(b). Over that image the different textures can be directly superposed as it is presented in 3(c) and 3(d).

In the multivariate remote sensing scenario, given a HSI $\mathcal{X}$ with K bands, our approach use a marginal framework obtaining for each spectral band denoted by $\mathbf{X}_k$ for $k = 1, \ldots, K$ a decomposition using expression 9 as follow:

$$M(\mathcal{X}_{ijk}, \sigma) = \{\mathcal{X}_{ijk}^{\boldsymbol{\sigma}}\}_{k=1}^K = \mathcal{X}_{ijkn}^{\boldsymbol{\sigma}} = \mathcal{X}^{\boldsymbol{\sigma}} \tag{10}$$

The selection of parameter vector $\boldsymbol{\sigma}$ is important but not critical as it is presented in section 5.2. Fig. 4 illustrates morphological decomposition in Indian Pine HSI which is used in the experiments section.

4. CLASSIFICATION BASED ON REDUCED TENSOR

Our motivation of NSS-ML for this study is to incorporate spatial information in the HSI classification. Previous researches have shown that mathematical morphology operators as opening and closing by reconstruction help us to set up spatial information in analysis.[6] Our approach utilizes also mathematical morphology to yield a nonlinear decomposition as it was presented in Section 3. For a hyperspectral image $\mathcal{X}$, the operator $M(\mathcal{X}, \boldsymbol{\sigma}) = \mathcal{X}^{\boldsymbol{\sigma}}$ produces a fourth dimension corresponding to different scales in the image, with the property that summing by scale is exactly the original hyperspectral image ($\bigoplus \mathcal{X}^{\boldsymbol{\sigma}} = \mathcal{X}$). Using tensor reduction presented in section 2 we obtain a smaller tensor preserving information both in bands and scales $\tilde{\mathcal{X}}_{I_1, I_2, k_b, k_s}^{\boldsymbol{\sigma}}$, where k_b and k_s are the number of feature to reduce in bands and scale respectively. Fig. 5 illustrates tensor reduction approach in comparison with standard PCA in the first nine dimensions. As it is presented in fig. 5, tensor PCA (TPCA) produces a reduction that, according to the spatial scale dimensions considered, has associated an object size, i.e., the fig. 5(a) has basically the stored information of structure in the morphological scale-space decomposition.

(a) $\mathcal{X}$ (b) $\mathcal{X}^{S_N}$ (c) $\mathcal{X}^{S_N}+\bigoplus_{n=1}^{2}\mathcal{R}^{S_n}$ (d) $\mathcal{X}^{S_N}+\bigoplus_{n=2}^{3}\mathcal{R}^{S_n}$
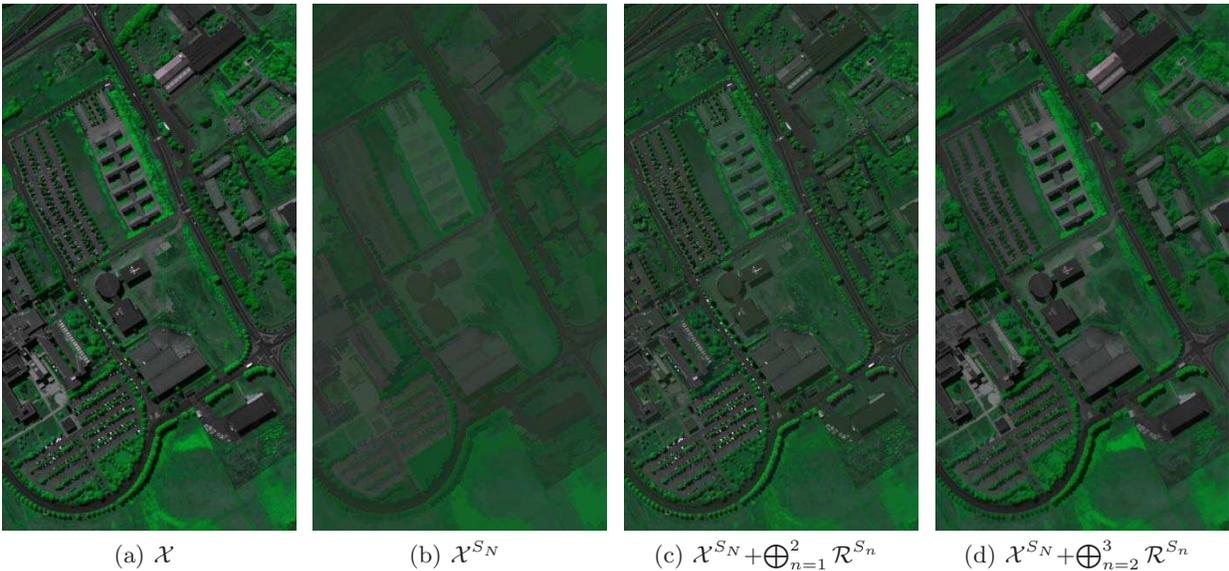
Figure 3. Example NNS-ML in Pavia University. Images show bands [80,70,60]

Directly from tensor reduction we have a matrix $k_b \times k_s$ for each pixel. Taking advantage of flexibility in support vector machine problem formulation, we solve the classification problem using gaussian kernel in matrix formulation. We define the classification problem as follow:

Given a training set $S = \{(\mathbf{X}_1, y_1), \ldots, (\mathbf{X}_s, y_s)\} \in \mathbb{R}^{k_b} \times \mathbb{R}^{k_\sigma} \times \{-1, 1\}$, the decision function is found by solving the convex optimization problem:

$$\max_{\alpha} \quad h(\alpha) \;\; = \;\; \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j k\left(\mathbf{X}_i, \mathbf{X}_j\right)$$

$$\text{subject} \quad \text{to} \quad 0 \;\; \leq \;\; \alpha_i \leq C \quad \text{and} \quad \sum_{i=1}^{l}\alpha_i y_i = 0$$

where C is a constant to penalize the training errors, α's are the Lagrange coefficients, and h is the kernel function. We use a Gaussian kernel for matrices defined by:

$$h_{\widetilde{\sigma}}\left(\mathbf{X}_i, \mathbf{X}_j\right) = \exp\left(-\frac{\|\mathbf{X}_i - \mathbf{X}_j\|^2}{2\widetilde{\rho}^2}\right)$$

where the kernel requires tuning for the proper value $\widetilde{\rho} \in \mathbb{R}^{+}$ and the norm is the Euclidean norm.

5. EXPERIMENTS

In this section, we start analyzing the influence of the number of eigenvalues in the spatial filtering induced by TPCA reduction. Additionally, comparison in the scale parameters for the tensor decomposition are presented, and an illustration of this influence is showed. The proposal method is compared with classical dimensional reduction in Indian Pine HSI. This image contains 145x145 pixels and 200 spectral bands in the 400-2500 nm range.

5.1 Parameter in Tensor-PCA

Tensor PCA is based on HOSVD as it was presented in section 2.3. HOSVD is a nonconvex problem and it can be obtain a local optimal solution. Inspired in,[15] we analysis now if the solution of a tensor decomposition is unique taking an experimental approach for HSIs. As the algorithm to solve HOSVD is based on seed matrix,
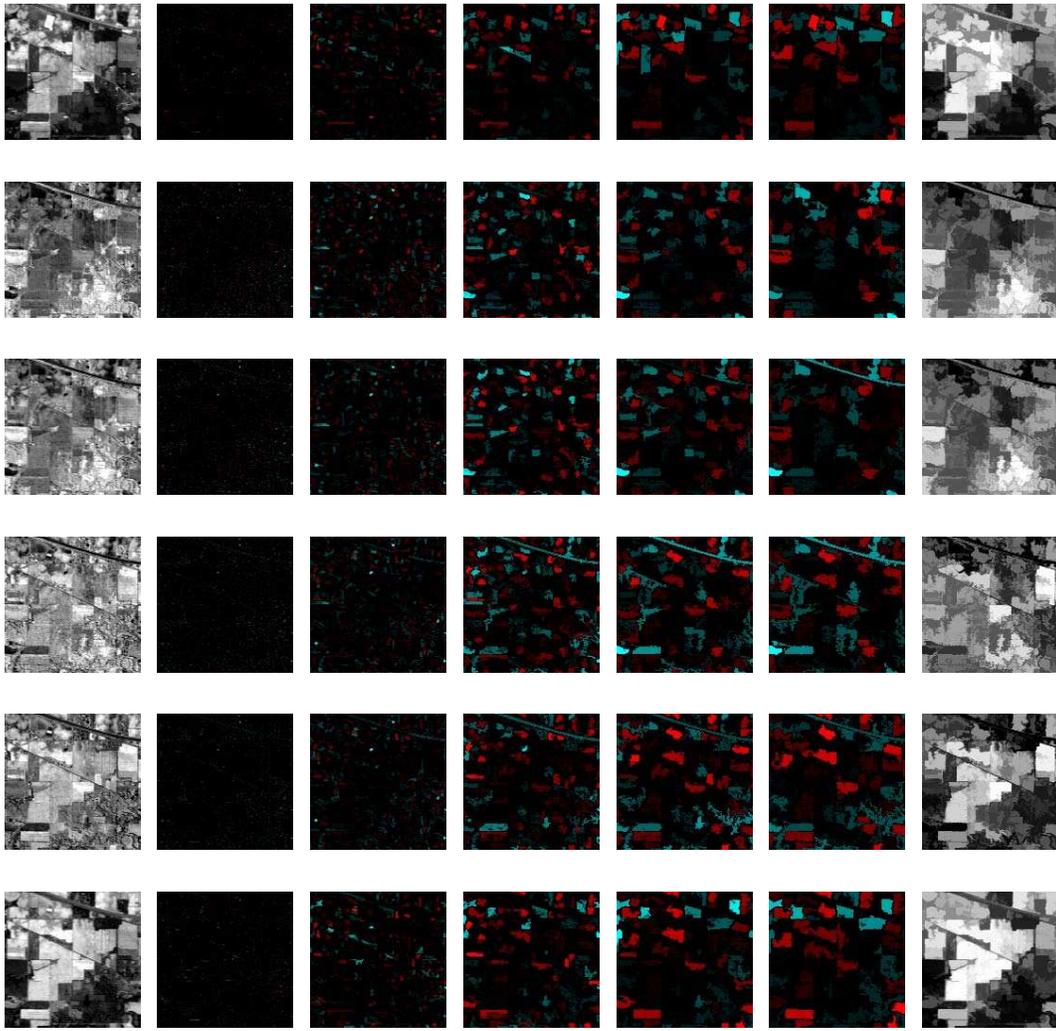
Figure 4. Morphological decomposition steps using $\boldsymbol{\sigma} = [1, 3, 5, 7, 9]$ for six spectral bands. NSS-ML contains spatial information per band preserving original contours.

we run it with different random starting points, if the solution of all these runs agree with each other, then we can consider that the decomposition is unique for HSI. The measure that we are using is for each independent test, Difference= $\sum_i \|\mathbf{U}_i^t - \overline{\mathbf{U}}_i\|_F$, where $\overline{\mathbf{U}}_i$ is the solution of HOSVD when the seeds are considered identical matrices. We present some examples over calibrated hyperspectral AVIRIS image "Yellowstone calibrated scene 3" which has been provided by NASA Multispectral and Hyperspectral Data Compression Working Group.[17] Figure 6(a) is the false-color preview for a section of the 200 columns, 200 rows in 224 bands. In natural image, randomization in block scramble produces eigenvalue distribution with an step associated with the block size.[15] In HSIs, the spectral redundance avoid that behavior as it is showed in Fig. 6(d)6(e) for matrices $\mathbf{F}$ and $\mathbf{G}$ associated with HSIs with spatial block illustrated in Figs. 6(b)6(c). The convergence of HOSVD, it is apparently not related with some specific eigenvalues threshold, then we propose to use the cumulative eigenvalue distribution with empirical cut-off of 99%. That value shows convergence in all our experiments. For Indian Pine, we obtain $n_1 = 65$ and $n_2 = 82$, as it is presented in Fig. 7(a). The comparison for two values set is exhibited in Figs. 7(b) and 7(c).
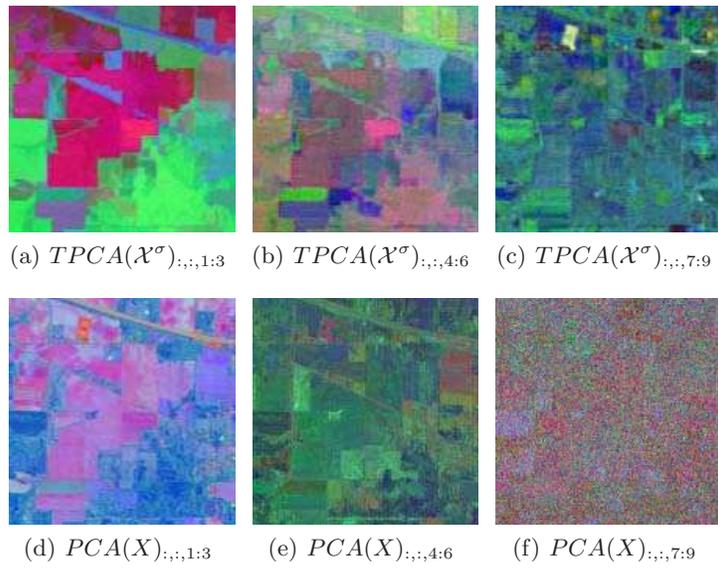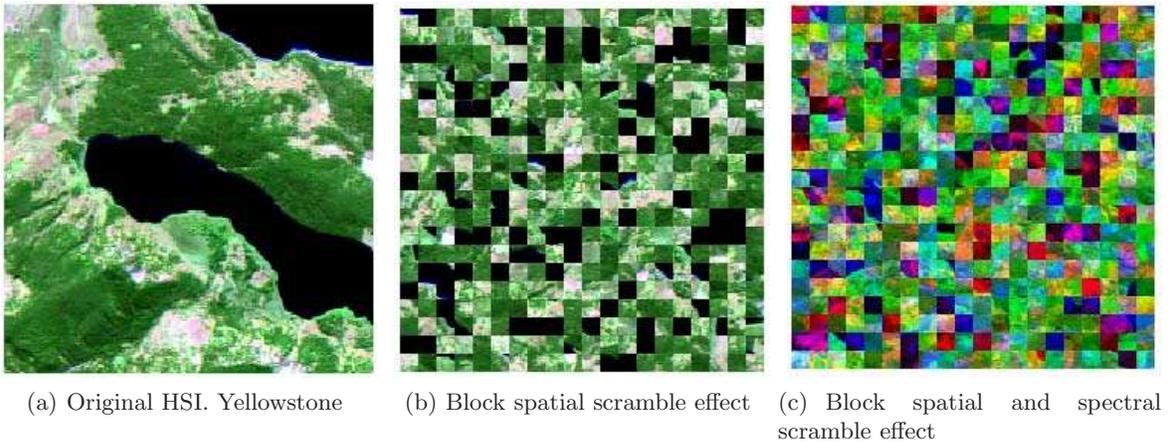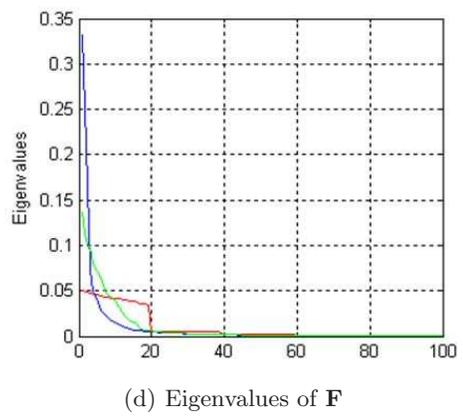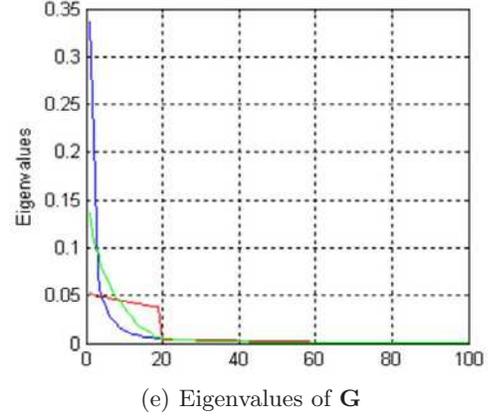
(a) $TPCA(\mathcal{X}^{\sigma})_{:,:,1:3}$ (b) $TPCA(\mathcal{X}^{\sigma})_{:,:,4:6}$ (c) $TPCA(\mathcal{X}^{\sigma})_{:,:,7:9}$

(d) $PCA(X)_{:,:,1:3}$ (e) $PCA(X)_{:,:,4:6}$ (f) $PCA(X)_{:,:,7:9}$

Figure 5. TPCA and PCA reduction using NNS-ML for $\boldsymbol{\sigma} = [1, 3, \ldots, 11]$

(a) Original HSI. Yellowstone   (b) Block spatial scramble effect   (c) Block spatial and spectral scramble effect
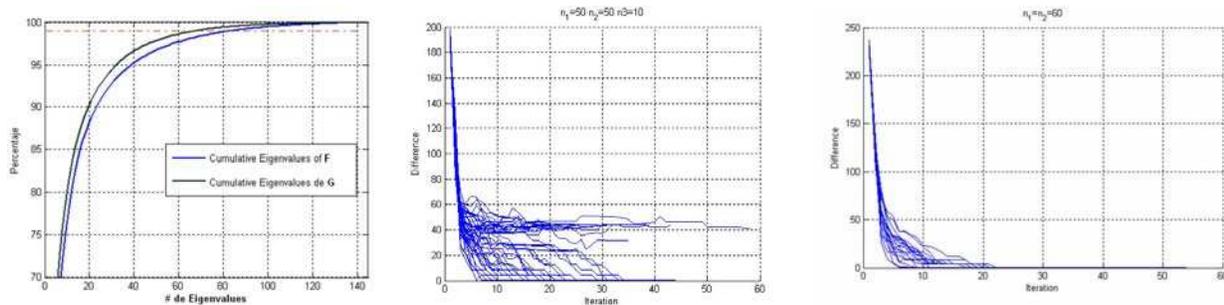

(d) Eigenvalues of $\mathbf{F}$ (e) Eigenvalues of $\mathbf{G}$

Figure 6. Eigenvalues comparison for matrices $\mathbf{F}, \mathbf{G}$ for original HSI in blue (6(a)), spatial block scramble HSI in green (6(b)) and spatial and spectral block scramble HSI in red (6(c)).

(a) Cumulative Eigenvalues of **F** and **G**

(b) Difference in convergence for $n_1 = 50$ and $n_2 = 50$

(c) Difference in convergence for $n_1 = 70$ and $n_2 = 70$

Figure 7. Comparison in convergence between two parameters sets in HOSVD tensor decomposition in Indian Pine HSI. The dimension utilized for HOSVD in the experiments is associated with the 99% of cumulate variance, in this case $n_1 = 65$ and $n_2 = 82$.

5.2 Scales in Morphological Decomposition

The parameter vector $\boldsymbol{\sigma}$ is determining the number of scales considered in the fourth-order tensor decomposition that extract structure and texture/scale decomposition as it was presented in the section 3.2. The structure of this decomposition exhibits that $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, \ldots, \sigma_M]$ is associated with the spatial-size of object that we are interested in. Leveling is reconstructing perfectly contours of objects which have a size bigger than that associated to the scale-size, the objects of size smaller that the scale are suppressed. To analysis the scale affect in the proposed decomposition, we include a series of experiments in supervised approach. For different values of $\boldsymbol{\sigma}$, changing the step and σ_M are presented in Figs. 8(a) 8(b). It clear that the scale parameters are not perturbing drastically the performance in supervised classification. Fig. 8(a) illustrate that σ_M too big are not necessary and it even can affect lightly the performance. Fig. 8(b) shows the smallest scale-step value has the worse performance. That is due to NSS-ML is based on differences, i.e., $\mathbf{R}^{S_1} = \mathbf{X}^{S_0} - \mathbf{X}^{S_1}$, so if $S_1 - S_0 \to 0 \Rightarrow ||\mathbf{R}^{S_1}||_F^2 \to = 0$.
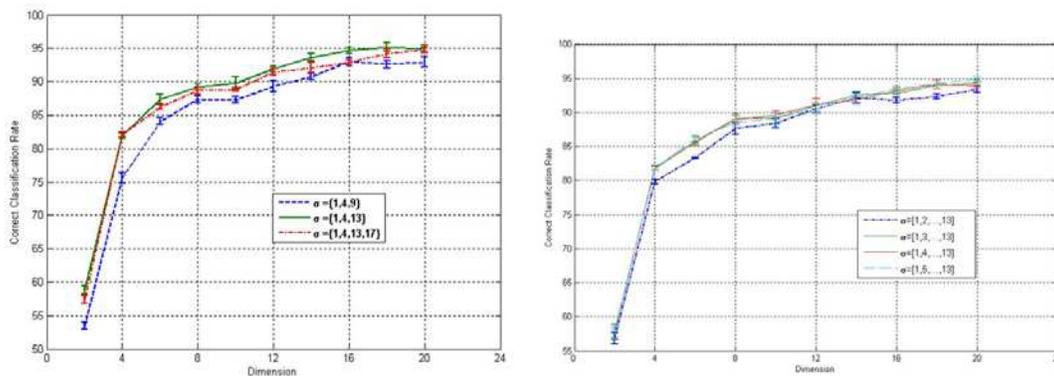

(a) Classification using 10% in 20 repetitions for different values of σ_M

(b) Classification using 10% in 20 repetitions for different values of step in $\boldsymbol{\sigma}$

Figure 8. Indian Pine Classification using SVM for different scale parameters $\boldsymbol{\sigma}$

5.3 Classification

Previous to classification using SVM-KM,[11] each feature was standardized. The kernel parameter $\widetilde{\rho}$ were tuned in range $\widetilde{\rho} = \{1, 2, \ldots, 10\}$ and the regularization parameter for SVM was varied in $C = \{10^0, \ldots, 10^5\}$. In considerate scenario, we followed the procedure presented in.[12] The finally selected classes were: "Corn-no till"

(1434), "Corn-min till" (834), "Grass/Pasture" (497), "Grass/Trees" (747), "Hay-windrowed" (489), "Soybean-no till" (968), "Soybean-min till" (2468), "Soybean-clean till" (614), and "Woods" (1294). Therefore, we present classification for 9 classes with 5 and 10 labeled samples per class.
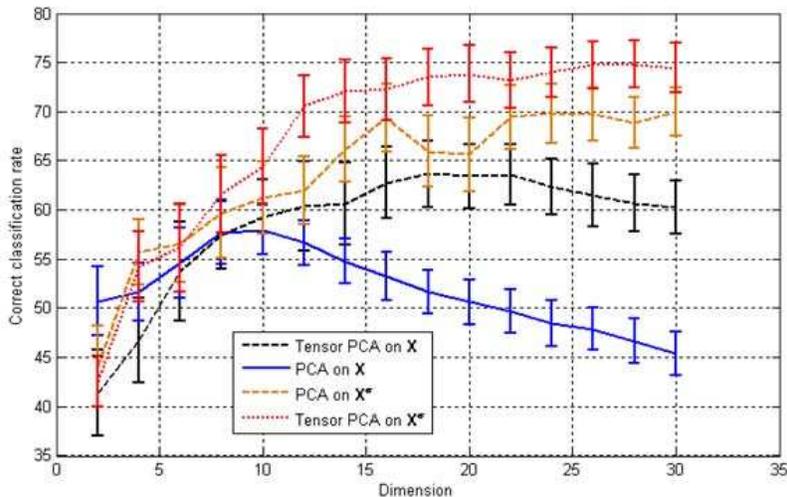
(a) 5 training samples per class in 20 random repetitions for 9 classes.
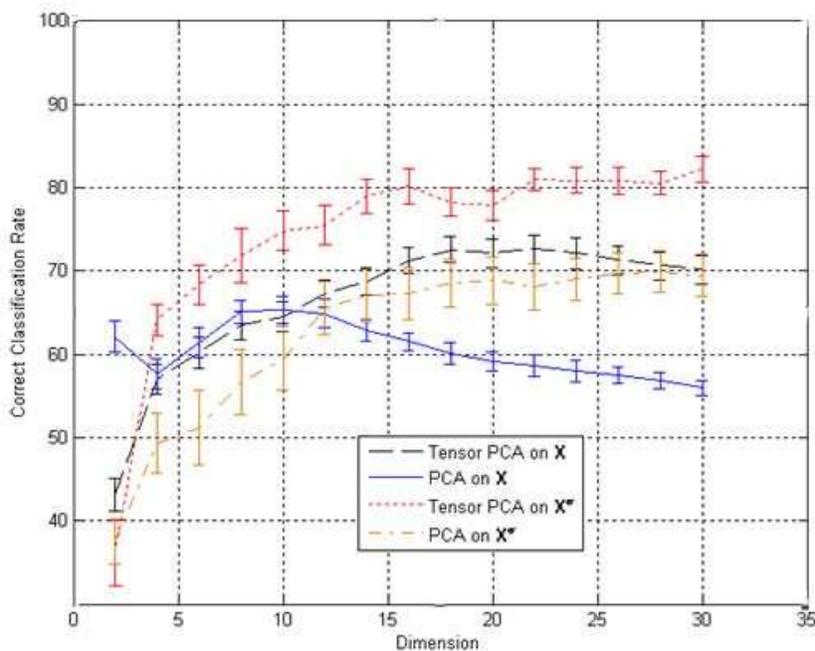
(b) 20 training samples in 9 classes in 20 random repetitions.

Figure 9. OA accuracy for different dimensional reduction approaches in Indian Pine.

Results for different dimensional reduction approaches are presented in Figs. 9(a) and 9(a). This extreme scenario shows the importance to incorporate spatial information previously to do feature reduction. Our approach involving morphological information are clearly better that its equivalent spectral. Additionally, our proposal involving tensor reduction has the best performance using 2 spatial dimension (in the reduction with more than 70% in overall classification for 5 labeled samples per class. Other results are presented in Fig. 9(b).

Our approach finds an excellent performance in comparison with state of the art.[13] Optimum gaussian kernel parameter was obtained cross-validation $\widetilde{\rho} = 3$ and regularization parameter for SVM, $C = 10^5$.

6. CONCLUSION

The paper proposed a framework integrating structural/spatial information in unsupervised dimensional exploration and feature extraction for hyperspectral images. Nonlinear Scale-Space representation with morphological levelings incorporates favorably the spatial information. Tensor structure integrates structural/spatial information, together with the spectral one, in feature extraction causing drastic dimension reductions without detrimental effect to classifier performance. We notice that complex objects are not defined by single level sets and consequently their structures appear in several scales of the decomposition. Results in real hyperspectral images shows as tensor approach incorporates more usefully spatial information in dimensional reduction stage in comparison with its matrix equivalent version. Results present a better performance that state of the art in classification using spatial and spectral information.[13] In addition, we have shown that HOSVD obtain an unique solution taking a sufficient number of dimensions in the spatial filtering introduced by HOSVD.

REFERENCES

[1] D. A. Landgrebe, Hyperspectral Image data analysis as a high dimensional signal processing problem. Special issue of the IEEE Signal Processing Magazine, 19 (1), 17-28, Jan. 2002.

[2] L.O. Jiménez, E. Arzuaga, and M. Vélez, Unsupervised Linear Feature-Extraction Methods and their Effects in the Classification of High-Dim. Data, IEEE Trans. Geosci. Remote Sens., 45(2), 469-483, Feb. 2007.

[3] N.Renard and S. Bourennane, Improvement of Target Detection Methods by Multiway Filtering, IEEE Trans. Geosci. Remote Sens., 46(8), Aug. 2008.

[4] M. Pesaresi and J.A. Benediktsson, A new approach for the morphological segmentation of high-resolution satellite imagery, IEEE Trans. Geosci. Remote Sens., 39(2), 309-320, Feb. 2001.

[5] J. A. Palmason, J.A. Benediktsson, J.R. Sviensson, and J. Chanussot, Classification of hypersepctral data from urban areas using morphological preprocessing and ICA, in proc. IGARSS,176-179, Jul. 2005.

[6] M. Fauvel, J.A. Benediktsson, J. Chanussot, J. Sveinsson, Spectral and Spatial Classification of Hypersepctral Data Using SVMs and Morphological Profiles, IEEE Trans. Geosci. Remote Sens., 46(11), Nov. 2008.

[7] L. Lathauwer, B. Moor, and J.Vandewalle, A Multilinear Singular Value decomposition, SIAM J. Matrix Anal. Appl.,21(4), 1253-1278, 2000.

[8] G.M. Foody and A. Mathur, A relative evaluation of multiclass image classification by support vector machines, IEEE Trans. Geosci. Remote Sens., 42(6), 1335-1343, Jun. 2004.

[9] T. G. Kolda and B. W. Bader, Tensor Decompositions and Applications, SIAM Review, 51(3), Sept. 2009 (to appear)

[10] F. Meyer and Petros Maragos, Nonlinear Scale-Space Representation with Morphological Levelings, Journal of Visual Communication and Image Representation 11(2), 245265,2000.

[11] S. Canu, Y. Grandvalet, V. Guigue and A. Rakotomamonjy, SVM and Kernel Methods Matlab Toolbox, INSA de Rouen, France, 2005.

[12] G. Camps-Valls, T. Bandos, and D. Zhou., Semi-supervised graph-based hyperspectral image classification, IEEE Trans. Geosci. Remote Sens., 53(10), 3044 - 3054, Oct. 2007.

[13] A. Plaza, J. A. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, J.C. Tilton and G. Triani, Recent Advances in Techniques for Hyperspectral Image Processing, Remote Sensing Environment. Accepted for publication, 2009.

[14] H. Huang, C. Ding, D. Luo and T. Li, Simultaneous tensor subspace selection and clustering: the equivalence of high order svd and k-means clustering, KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, 200, 327-335, Las Vegas, Nevada, USA, 2008.

[15] D. Luo, H. Huang and C. Ding, Are Tensor Decomposition Solutions Unique? On the global convergence of HOSVD and ParaFac algorithms, CoRR, 2009.

[16] S. Velasco-Forero and J. Angulo, Morphological scale-space for hypersepctral images and dimensionality exploration using tensor modeling, First IEEE Workshop on Hyperspectral Image and Signal Processing: Emerging Remote Sensing (WHISPERS 2009), IEEE, 08/2009.

[17] A.B. Kiely and M.A. Klimesh, Exploiting Calibration-Induced Artifacts in Lossless Compression of Hyperspectral Imagery, IEEE Trans. on GeoRS, 47(8), 2672-2678, 2009.

[18] L. Eldén and B. Savas, A Newton–Grassmann method for computing the Best Multi-Linear Rank-(r_1, r_2, r_3) Approximation of a Tensor, SIAM J. Matrix Anal. Appl., 31,248-271,2009.